ON THE AARONSON–AMBAINIS CONJECTURE: A QUASI-POLYNOMIAL BOUND, TOP-HEAVINESS, AND GLUING ALONG DECISION TREES

PAATA IVANISVILI

Abstract. The Aaronson–Ambainis conjecture asserts that every polynomial $f\colon \{-1,1\}^n \to [-1,1]$ of degree at most d has a variable of influence at least $(\mathrm{Var} f/d)^{O(1)}$. This paper does *not* prove the conjecture, which remains open. To our knowledge, the best previously known general bound is $\mathrm{Var}(f)^2 e^{-O(d)}$. We prove the general bound $\mathrm{MaxInf}(f) \geq (\mathrm{Var} f/d)^{O(\log d)} \exp(-O(\log^4 d))$, which improves on it asymptotically. Its constants are astronomically large: at $\mathrm{Var} f = 1$ it beats e^{-d} only for $d \gtrsim 1.9 \cdot 10^{13}$. Through the reduction of Aaronson and Ambainis, it implies that the acceptance probability of a t-query quantum algorithm can be approximated within ε on all but a δ fraction of inputs by a classical algorithm making $\exp(O(\log^4 t + \log t \cdot \log(1/(\varepsilon\delta))))$ queries ($t \geq 3$), to our knowledge the first subexponential overhead for general algorithms, again only asymptotically; this bounds the number of queries, not the running time. We propose a quantum analogue of the OSSS inequality, prove it for one query, for two queries with a larger constant, and for nonadaptive algorithms, and show that it would give a classical simulation of quantum query algorithms on most inputs with polynomially many queries. We also constrain counterexamples, reformulate the conjecture exactly, and prove it for several classes. First, counterexamples are top-heavy: $\mathrm{MaxInf}(f) \geq (W^{[1,k+1]}(f)/(4d^2 3^k))^2$ for every $k \geq 0$, so the conjecture holds with polynomial loss unless the Fourier weight of f on levels $1, \ldots, O(\log d)$ is polynomially small; Lipschitz images of a counterexample are top-heavy as well, and its level sets are noise sensitive on average. Second, a gluing lemma, proved through a Hilbert-space-valued OSSS inequality, converts influential variables at the leaves of a depth-t decision tree into an influential variable of f, at a cost polynomial in t and d. It yields an exact decision-tree reformulation of the conjecture for one-block (gradient) forms, shows that the conjecture is stable under shallow adaptive switching, and, with bounds on leaf root sums, settles several selector classes; a subcube cover lemma settles functions with constant-magnitude subcube-supported derivatives. Third, we prove polynomial bounds for symmetric functions with optimal exponents ($\mathrm{Var}^2/(16d^2)$), for functions symmetric within each of b blocks ($\mathrm{Var}^2/(4b^2d^4)$), for wreath-symmetric functions ($\mathrm{Var}^2/(2416\, d^6)$) for all block sizes, and the sharp order Var^2/d^2 in two regimes) and for polynomials of smooth linear statistics; we improve constants for graph properties and cyclically invariant functions, and reduce the case of polynomials bounded on the torus to a spike inequality, which we prove in each fixed degree with a gap exponentially small in d. For graph properties of degree d we prove the polynomial bound $\mathrm{MaxInf}(f) \geq \mathrm{Var}(f)^3/(1.9 \cdot 10^9 d^{16})$, using the clone-and-subsample technique of Zhandry's small-range lemma. We isolate the remaining gap as a question about shallow decision trees that make leaf root sums polynomially small, show that the address function caps several routes at $2^{-\Omega(d)}$, and give a transfer inequality, implied by the conjecture, that with polynomial loss would yield $\mathrm{MaxInf}(f) \geq (\mathrm{Var} f/d)^{O(\log d)}$. The new bound proves the mass form of this transfer with a quasi-polynomial prefactor, through a recursion over chains of derivatives on a grid of heights, in which lightness is measured against a fixed fraction of the node's weight. We also show that the L^∞ Talagrand conjecture of Lovett and Zhang, which would imply the conjecture, is false, propose a degree-free strengthening in terms of real block sensitivity, and show that two natural strategies fail: the analogous bound in terms of real sensitivity is false with an exponential gap, and linear dual certificates for the extremal problem give only the trivial bound up to a factor $O(\log^2 n)$. For a single restriction at rate $1/d$, size-biased tools (random priming, a load identity for adaptive explorations, and a level-1 inequality in spread directions) give polynomial bounds under capture-type hypotheses, which are circular as reductions but hold for structured classes. Along this route the conjecture reduces to an open anti-spike inequality for regular bounded functions, some of whose hard instances are products of nearly Boolean counterexamples of lower degree. Several core results, not including the quasi-polynomial bound, are also verified in Lean 4.

1. Introduction

1.1. **The conjecture.** Let $Q_n = \{-1,1\}^n$ with the uniform probability measure. Every $f\colon Q_n \to \mathbb{R}$ has a unique multilinear (Fourier–Walsh) expansion $f = \sum_{S\subseteq[n]} \widehat{f}(S) x^S$,

$x^S = \prod_{i\in S} x_i$, and $\deg f = \max\{|S| : \widehat{f}(S) \neq 0\}$. Write

$$\mathcal{B}_d(n) = \{f\colon Q_n \to [-1,1] :\ \deg f \le d\}, \qquad \mathrm{Inf}_i(f) = \sum_{S\ni i} \widehat{f}(S)^2, \qquad \mathrm{MaxInf}(f) = \max_i \mathrm{Inf}_i(f),$$

$\mathcal{B}_d = \bigcup_n \mathcal{B}_d(n)$, and $\mathrm{Var} f = \sum_{S\neq\emptyset} \widehat{f}(S)^2$. (Precise conventions are fixed in Section 2.)

Conjecture 1.1 (Aaronson–Ambainis [2, Conjecture 1.7]; Conjecture 6 in arXiv:0911.0996)**.** *There are absolute constants $c, C > 0$ such that for all $n, d \ge 1$ and all $f \in \mathcal{B}_d(n)$,*

$$\mathrm{MaxInf}(f) \ \ge\ c\left(\frac{\mathrm{Var} f}{d}\right)^C.$$

[*open*] In [2] the conjecture is stated for $p\colon \{0,1\}^N \to [0,1]$ of degree d with $\mathrm{Var} p \ge \varepsilon$ and $\mathrm{Inf}_i[p] = \mathbb{E}_X(p(X) - p(X^i))^2$ (no factor 1/4), with conclusion $\mathrm{Inf}_i[p] \ge (\varepsilon/d)^{O(1)}$ for some i. Putting $x_i = 1 - 2X_i$ and $f = 2p - 1$ gives $f \in \mathcal{B}_d(N)$, $\mathrm{Var} f = 4\mathrm{Var} p$ and $\mathbb{E}_X(p(X) - p(X^i))^2 = \mathrm{Inf}_i(f)$; hence the two formulations are equivalent (the factor 4 is absorbed in c). The range $[0,1]$ on $\{-1,1\}^n$, used e.g. in [12], is equivalent in the same way. By [2, Appendix, Proposition 32] the conjecture is also equivalent to its L^1 version (with L^1 influences and L^1 variance). The motivation in [2] is the classical simulation of quantum query algorithms on most inputs: the conjecture implies that the acceptance probability of every T-query quantum algorithm can be approximated on a $1-\delta$ fraction of inputs, within ε, by a deterministic classical algorithm making $\mathrm{poly}(T, 1/\varepsilon, 1/\delta)$ queries [2, Theorem 22 of the arXiv version].

The conjecture is false for L^2-bounded f (take $n^{-1/2}\sum_i x_i$), and the exponent of $\mathrm{Var} f$ must be at least 2 (take $f = n^{-1}\sum_i x_i$, with $\mathrm{Var} f = 1/n$ and $\mathrm{MaxInf} f = 1/n^2$).

1.2. **Known results.** We list the results we use or compare with, with the exact exponents as stated in the sources.

- *Exponential bounds.* The junta theorem of Dinur, Friedgut, Kindler and O'Donnell [16, Theorem 1] shows that every $f \in \mathcal{B}_d$ has a $2^{O(d)}/\varepsilon^2$-junta g with $\|f-g\|_2^2 \le \varepsilon$; as noted in [2] this gives $\mathrm{MaxInf}(f) \ge \mathrm{Var}(f)^3/2^{O(d)}$. Applying the DFKO inequality [16, Theorem 3] to $(f - \mathbb{E}f)/\sqrt{\mathrm{Var} f}$ with $t = 2/\sqrt{\mathrm{Var} f}$ gives directly $\mathrm{MaxInf}(f) > \mathrm{Var}(f)^2/(4C^d)$, since $|f - \mathbb{E}f| < 2$ when $\mathrm{Var} f > 0$. O'Donnell and Zhao, with an observation joint with J. Wright [43, eq. (1)], showed $\mathrm{MaxInf}(F) \ge e^{2-2k}\mathrm{Var}(F)^2$ for one-block decoupled forms $F(y,z) = \sum_i y_i g_i(z)$ of degree k with range $[-1,1]$ (so $\deg g_i \le k-1$; in Section 2 the degree bound of the g_j is called k), and proved that Conjecture 1.1 holds if and only if it holds for such forms [43, Theorem 2.13]; together these give $\mathrm{MaxInf}(f) \ge \mathrm{Var}(f)^2 e^{-O(d)}$ (Proposition 2.5 below).
- *Special classes.* Montanaro [38, Corollary 20] proved that if f is a homogeneous degree-k block-multilinear form on k blocks of n variables with $\|f\|_\infty \le 1$ and *all* n^k coefficients equal to $\pm\alpha$, then every variable has influence $\Omega(\mathrm{Var}(f)^2/k^3)$. (This is narrower than the frequently quoted paraphrase "all nonzero coefficients have equal magnitude"; the address function of Section 2.4 has equal-magnitude coefficients but non-product support.) Bansal, Sinha and de Wolf [7, Theorem 1.3] proved $\mathrm{MaxInf}(f) \ge \mathrm{Var}(f)^2/(e(d+1)^4)$ for block-multilinear f of degree d with $\|f\|_{\mathrm{cb}} \le 1$, and a root-influence inequality $\|f\|_{\mathrm{cb}} \ge (e(d+1))^{-1/2}\sum_i \mathrm{Inf}_{b,i}(f)^{1/2}$ for the first and last blocks $b \in \{1, d\}$ of homogeneous forms [7, Theorem 1.4]. Escudero Gutiérrez [23, Theorems 1.6 and 1.8] improved this to $\mathrm{MaxInf} \ge (\mathrm{Var}/d)^2$ for cb-bounded block-multilinear forms, to Var^2 in the homogeneous case, and proved a degree-free bound for homogeneous Fourier-completely-bounded polynomials. Escudero Gutiérrez, Saucedo and Palazuelos [24, Theorem 1.3] proved the optimal inequality $\|a\|_{\mathrm{cb}} \ge t^{-1}\sum_{s\in[t]}\sum_i \mathrm{Inf}_{x_s(i)}[a]^{1/2}$ for block-multilinear a of degree t. *Caveat:* all these root-influence bounds concern block-multilinear forms whose blocks are *independent* variables. For

the symmetrized ("bias") polynomial of a cb-bounded form the analogous statement fails: the address function is the bias of a $(k+1)$-query algorithm, hence of the form $T(\tilde{x},\dots,\tilde{x})$, $\tilde{x}=(x,1)$, with $\|T\|_{\mathrm{cb}}\le 1$ by [4, Theorem 1.3], yet $\sum_i \mathrm{Inf}_i^{1/2}\approx 2^{k/2}$ (Remark 4.11). For symmetric f the conjecture is known: Bačkurs [5, Theorem 2] proved $\mathrm{Inf}_i(f)=\Omega(\varepsilon^3/(d^4\ln(1/\varepsilon)))$ when $\mathrm{Var}(f)=\varepsilon$ (in his normalization), and Ivanisvili [26] presented this proof in the form $\mathrm{MaxInf}(f)\ge c(\mathrm{Var}(f)/d)^4$. The sharp form $\mathrm{MaxInf}(f)\ge \mathrm{Var}(f)^2/(16d^2)$, tight up to the constant (Proposition 6.2), we have not found in print. For Boolean-valued f it follows from the OSSS inequality [42] and the polynomial relation between decision-tree depth and degree [40, 9, 37]: $D(f)\le \mathrm{bs}(f)\deg(f)$ [37] ($D(f)$ the decision-tree depth, $\mathrm{bs}(f)$ the block sensitivity, see Section 5.3) and $\mathrm{bs}(f)\le \deg(f)^2$ (e.g. from [19, Theorem 3.3] by block substitution) give $D(f)\le d^3$, hence $\mathrm{MaxInf}(f)\ge \mathrm{Var}(f)/D(f)\ge \mathrm{Var}(f)/d^3$.

- *Transitive symmetry.* Filmus, Hatami, Keller and Lifshitz [19, Proposition 3.18] proved that every transitive-invariant $f\in\mathcal{B}_d(n)$ (for all i,j some permutation of the coordinates that fixes f maps i to j) has $\mathrm{Var}(f)\le d^4e^{2d}/n$. All influences of such f are equal, so $\mathrm{MaxInf}(f)\ge \mathrm{Var}(f)/n\ge \mathrm{Var}(f)^2/(d^4e^{2d})$. This covers graph properties, cyclically invariant functions and wreath-symmetric functions: for these classes the conjecture in each fixed degree is known, with $n\mathrm{Var}f\le 16e^4\approx 874$ for $d=2$ and $n\mathrm{Var}f\le 81e^6\approx 3.3\cdot 10^4$ for $d=3$. (The same proof gives $n\mathrm{Var}f\le n^2\mathrm{Inf}_1(f)\le n^2C_{d-1}\|d_1f\|_1^2\le C_{d-1}d^4$, with C_k the best constant in $\|g\|_2^2\le C_k\|g\|_1^2$ for $\deg g\le k$; Szarek's inequality [48], applied to g plus a fresh variable times its constant term, gives $C_1=2$, and Bonami's inequality with Hölder gives $C_2\le 81$, hence 32 and 6561.)
- *L^1 influences.* Bačkurs and Bavarian [6, Theorem 1.2] proved $\sum_i\|d_if\|_1=O(d^3)$ for $f\in\mathcal{B}_d$ (an earlier version had $O(d^3\log d)$, the bound quoted in [2]); Filmus, Hatami, Keller and Lifshitz [19, Theorem 3.3] proved the pointwise bound $\sum_i|d_if(x)|\le d^2$, which is sharp. We prove the weaker $4d^2$ in Lemma 2.2 by the classical symmetrization argument.
- *Random restrictions.* Results of Bhattacharya [12, Theorems 6.2 and 6.4] on random restrictions, and a caveat on the printed form of [12, Theorem 6.2], are recorded in Appendix A.
- *Fractional certificates.* Lovett and Zhang [35, Theorem 13, Theorem 27, Corollary 4] proved that for $f\colon\{0,1\}^n\to[0,1]$ of degree d the fractional block sensitivity satisfies $\mathrm{FBS}_\varepsilon(f)\le O(d^8\log^{16}n/\varepsilon^4)$, and that if $\mathrm{Var}f\ge\Omega(\varepsilon)$ then an ε-fraction of inputs are (r,ε)-sensitive with $r=\mathrm{poly}(d,1/\varepsilon,\log n)$. They formulated a "Talagrand inequality for L^∞" [35, Conjecture 37]: if $X,Y\subseteq\{0,1\}^n$ have L^∞-influences at most τ, then some $x\in X$ satisfies $d_\infty(x,\mathrm{conv}\,Y)\le\mathrm{poly}(\tau)$; and showed that it implies Conjecture 1.1 [35, Claim 38] (see [35] for FBS_ε, (r,ε)-sensitivity, L^∞-influences and d_∞). Tribes shows that this conjecture is false as stated, and a gap version of Tribes refutes its symmetric form as well (Proposition H.4).
- *A withdrawn proof.* Keller and Klein [29] announced a proof of Conjecture 1.1; the preprint was withdrawn (v2) with the comment that "Lemma 5.3 does not prove the assertion it claims to prove and this collapses the entire argument". Several of its lemmas are correct, in particular the low-level inequality [29, Lemma 4.4], which we reprove in a slightly different form in Theorem 3.2.
- *Variants and consequences.* Girish, Mittal, Nehoran and Raz [21, Theorem 4.3] proved that the conjecture for the p-biased product measure is equivalent to the conjecture for the uniform measure on a slice, for all fixed $p,q\in(0,1)$. Blanc, Docter, Strassle and Tan [13, Theorem 1] showed that t-query r-round quantum algorithms can be simulated on a $1-\delta$ fraction of inputs within ε by classical algorithms making $2^{O(r^2)}(t\log(1/\delta)/\varepsilon)^{O(r)}$ queries; this is a simulation statement for a subclass of bounded polynomials, not the influence conjecture. Liu and Mutreja [33] showed that T nonadaptive (parallel) quantum queries can be simulated on most inputs with $\mathrm{poly}(T,1/\varepsilon,1/\delta)$ classical queries (Corollary M.17 gives

$\lceil 4T^2/(\varepsilon^4\delta^3)\rceil$), and r adaptive rounds with $O(CrT/(\varepsilon\sqrt{\delta}))^{c4^r}$ queries; combined with the real-valued OSSS inequality [42, eq. (6)] this gives (our derivation; it is not stated in [33]) $\mathrm{MaxInf} \geq (\mathrm{Var}/(rT))^{O(4^r)}$ for acceptance probabilities of r-round algorithms. A disjoint-support separator of Li, Li, Li and Liu [32, Lemma 3.5] (multilinear p, q of degree at most d with $pq \equiv 0$ on the cube admit a decision tree of depth $O(d^4)$ deciding which of them is nonzero, where $p + q \neq 0$) gives, with OSSS, a polynomial influence bound for the exact two-polynomial zero-pattern problem, as claimed in the unrefereed preprint [31]; this uses exact zeros and does not bear on Conjecture 1.1 directly.

- *The sharp form.* A stronger form is [3, Problem 1.2]: there is an absolute constant C with

$$\mathrm{Var}(f) \leq \|f\|_\infty \deg(f)^C \max_j \|d_j f\|_2$$

(for $\|f\|_\infty \leq 1$ this reads $\mathrm{MaxInf}(f) \geq \mathrm{Var}(f)^2/d^{2C}$). We ask whether $C = 1$ is possible up to a constant factor, i.e. $\mathrm{MaxInf}(f) \geq c\mathrm{Var}(f)^2/d^2$; we know of no published statement of this sharp form. Both exponents would be optimal: $f = n^{-1}\sum_i x_i$ forces the exponent 2 of Var, and for odd d the Chebyshev polynomials $f = T_d(s)$, $s = (Kd^2)^{-1}\sum_{i\leq Kd^2} x_i$, satisfy $\mathrm{Var} f \geq 0.943/K$ and $\mathrm{MaxInf} f \leq 1.61\mathrm{Var}(f)^2/d^2$ for $K \geq 32$, and $\mathrm{Var} f \geq 0.416$, $\mathrm{MaxInf} f \leq 1.36/d^2$ for $K = 1$, $d \geq 9$ (Example 6.3). Recent unrefereed preprints [27, 47] claim polynomial Bohnenblust–Hille constants on the cube; these give the conjecture for (levelwise) flat spectra; the flat case was already derived from Bohnenblust–Hille inequalities in [15, §4.1]. In general, coefficient-based arguments cannot give the conjecture, since the address function has $\max_S \widehat{f}(S)^2 = 4^{-k}$ (Proposition 2.8(iii)).

1.3. **What is proved here.** *Status tags.* Every theorem, lemma, proposition and corollary is tagged with one of the following: [*known*] (with a citation; a proof is sometimes included for completeness), [*standard*] (a standard argument, proof included), [*proved here*] (a complete proof is given), [*proved here, computer-assisted*] (the proof relies on a recorded exact computation), [*proof outline*] (only an outline is given; not counted as proved), [*conditional*] (proved under a stated hypothesis that is not known to hold), [*numerical*] (computation only, not a proof), [*heuristic*], or [*open*]. We also use [*audited*]: the proof has been checked by an independent auditor; and [*unaudited*]: the proof has not been checked independently, and the statement is not counted as proved. These occur mainly in Sections 9–11 and Appendices J–M, also in combinations: [*proved here, audited*] means that a complete proof is given and was checked; [*audited twice in note form, and the text audited twice*] means that the mathematics was audited twice, independently, in the working notes from which the text is transcribed, and that the printed text was then audited twice more, independently, once as a transcription of the notes and once as a referee reads it; [*components as tagged*] means that its parts carry their own tags; [*proof outline; audited*] means that only an outline is printed, so the statement is not counted as proved here, although the full argument was checked. Used alone, [*computer-assisted*] marks a finite evaluation of explicit expressions whose values are printed. Results that are also verified in Lean 4 are listed in Appendix O and marked "Lean-verified" in Tables 1 and 2; none of the results of Section 10 is Lean-verified.

How the results fit together. Section 3 shows that a counterexample has almost no Fourier weight on levels $1, \dots, O(\log d)$; the tool is the level-k inequality applied to the derivatives (Theorem 3.2; the technique is known). Section 4 proves the gluing lemma by applying a Hilbert-space-valued OSSS inequality (Theorem 4.1) to the mass vector $(g_j^2)_j$ of a one-block form $F = \sum_j y_j g_j(z)$; the leaf root sums $R_l = \sum_j \mathbb{E}[g_j^2 \mid l]^{1/2}$ control the gain, and we bound them by register (decoupling) lemmas (Lemmas 4.6–4.9) and by the dimension of the span of the selectors (Proposition 4.12). The address function (Section 2.4) shows which routes

Result	Statement	Status	Sec.
A general bound			
Quasi-polynomial bound (Thm. 10.1)	$\mathrm{MaxInf}(f) \geq (\mathrm{Var} f/d)^{O(\log d)} \exp(-O(\log^4 d))$ for all $f \in \mathcal{B}_d$, $d \geq 2$, via SM-OF with a quasi-polynomial prefactor (Thm. 10.26); asymptotically better than $\mathrm{Var}(f)^2 e^{-O(d)}$, but the constants are astronomical: at $\mathrm{Var} f = 1$ it beats e^{-d} only for $d \gtrsim 1.9 \cdot 10^{13}$	proved here; audited twice in note form, and the text audited twice	10
Classical simulation via [2] (Cor. 10.29)	the acceptance probability of a t-query quantum algorithm is approximated within ε on all but a δ fraction of inputs with $\exp(O(\log^4 t + \log t \cdot \log(1/(\varepsilon\delta))))$ queries for $t \geq 3$ (previously $2^{O(t)}\mathrm{poly}(1/(\varepsilon\delta))$); better only for $t \gtrsim 1.5 \cdot 10^{12}$–$3.3 \cdot 10^{13}$	proved here from Thm. 10.1; audited	10
Structure of counterexamples and the gluing method			
Top-heaviness (Cor. 3.3)	$\mathrm{MaxInf}(f) \geq \left(W^{[1,k+1]}(f)/(4d^2 3^k)\right)^2$ for all $k \geq 0$ (known in substance [29]; Lean: $2d^2$ for $4d^2$)	proved here; Lean-verified	3
Images, level sets (Thms. 3.5, 3.6)	Lipschitz images of a counterexample are top-heavy; its level sets are noise sensitive on average	proved here	3
Gluing lemma (Thm. 4.2)	$\kappa_D - \varepsilon V \leq 8t^2\tau_D$, $\tau_D = \max_{i \in D} \mathrm{Inf}_{z_i}(F)$, for bounded one-block forms F ($\sum_j \lvert g_j\rvert \leq 1$) and z-trees D of depth $t \geq 1$	proved here; Lean-verified	4
Leaf root sums (Thm. 4.14)	$R_l \leq R$ on leaves carrying half of the variance $\Rightarrow$ $\max(\varepsilon, \tau_D) \geq V^3/(128R^2t^2)$	proved here	4
Reformulation (Thm. 4.15)	the conjecture $\iff$ each bounded one-block form has a z-tree of depth $\mathrm{poly}(k/V)$ with, on average, a y-variable of influence $\mathrm{poly}(V/k)$ at the leaves	proved here	4
Gluing for all f (Thms. 4.4, 4.5)	$\mathrm{MaxInf}(f) \geq \mathbb{E}_l[\mathrm{MaxInf}(f\lvert_l)^2]/(144d^5t^2)$; the conjecture is stable under shallow adaptive switching	proved here; Lean-verified	4
Classes with bounds polynomial in $\mathrm{Var} f$ *and* d			
Selector classes (Cors. 4.10, 4.13)	tree-mediated register forms: $\max(\varepsilon, \tau_D) \geq V^3/(16t^2)$ (polynomial if the depth is $t \leq \mathrm{poly}(k/V)$); polynomial single-statistic selectors: $\varepsilon \geq V^2/(k+1)$	proved here	4
Subcube (Cor. 5.2, Thm. 5.6, Cor. 5.12)	constant-magnitude subcube-supported derivatives: $\mathrm{MaxInf} \geq I^2/(128d^6)$; constant-magnitude subcube selectors: $V^5/(17424k^4)$; six derived classes (Cor. 5.12(4)–(9))	proved here; Lean-verified in part (Appendix O)	5
Symmetric (Prop. 6.2)	$\mathrm{MaxInf} \geq \mathrm{Var}^2/(16d^2)$, optimal up to the constant	proved here	6
Block-symmetric (Thm. 6.4)	$\mathrm{MaxInf} \geq \mathrm{Var}^2/(4b^2d^4)$ for b blocks (Lean: 16 for 4)	proved here; Lean-verified	6
Wreath-symmetric (Thms. B.1, 6.5, 6.8, 6.6)	M blocks of size w: $\mathrm{MaxInf} \geq \mathrm{Var}^2/(40d^2w(w+1))$ and $\mathrm{Var}^2/(2416\,d^6)$ always; $\mathrm{Var}^2/(170d^2)$ if $d \geq 2$, $M \geq 4\pi d^3$, $w \geq 8d^6 \ln(2d)$; $\mathrm{Var}^2/(27D^2)$ for composed functions of degree at most $D = ab$ ($D \geq 2$, $w \geq 4D^2$)	proved here; Thm. B.1 Lean-verified (weaker constant)	6
Linear statistics (Thm. 6.9)	$\mathrm{MaxInf} \geq I^2/(324\,d^2r)$ for r statistics at the natural smoothness scale	proved here	6
Graph properties (Thm. 6.16)	$\mathrm{MaxInf} \geq \mathrm{Var}^3/(1.9 \cdot 10^9 d^{16})$ for every graph property of degree d, by Zhandry's clone-and-subsample technique	proved here	6.5
Transitive classes in fixed degree, and torus-bounded polynomials			
Graph properties (Thms. 6.14, 6.15)	sharp limit constant as the number of vertices tends to infinity; $\mathrm{MaxInf} \geq \mathrm{Var}^2/97$ for $d \leq 3$ (Thm. 6.15)	proved here, computer-assisted	6.5
Cyclic (Thms. 6.18–6.20)	$n\mathrm{Var} f \leq 9$ for $d = 2$ and ≤ 1961 for $d = 3$; $nW^1 \leq d^2$, $nW^2 \leq 2d^4$, $nW^3 \leq 54d^6$ for every d	proved here	6.6
Torus-bounded (Thms. E.2, E.8)	block-symmetric: $\mathrm{MaxInf} \geq (\mathrm{Var}/b)^2$, free of d; reduced to the spike inequality (Conj. E.7), which holds in each fixed degree, with a gap $\eta_d \geq c\,2^{-3d/2}/d$ (no gain for the conjecture)	proved here	E

TABLE 1. Main results: a general bound, structure and special classes.

Result	Statement	Status	Sec.
Reformulations, the route to the quasi-polynomial bound, and obstructions			
Equivalent forms (Thms. 4.17, 7.3, 7.4, G.4)	leaf-wise tree statements, nearly Boolean functions, product selectors, one-scale restriction transfer: each is equivalent to the conjecture up to polynomial losses	proved here	4, 7, G
Touching transfer (Thm. 8.7)	conjecture (exponent at most the cap C) $\Rightarrow$ TT with polynomial loss $\Rightarrow$ MaxInf $\geq (\mathrm{Var}/d)^{O(\log d)}$; TT fails for losses $L \leq d^{1-o(1)}$; without the conjecture, TT holds at a θ-dependent loss with a quasi-polynomial prefactor (Prop. 10.3, via Thm. 10.26)	proved here; both implications Lean-verified; the θ-dependent form audited, not Lean-verified	8
Spikes after restriction (Sec. 9, App. J)	priming via real block sensitivity; a captured spike carries little mass (Lemma 9.8), so capture hypotheses are false, equivalent to TT, or imply the small-spiked-mass condition SM (Thm. 9.9); size-biased tools (RL1, the Load Lemma, Thm. 9.14); SM for old-free functions at rate $\frac{1}{2}$ (SM-OF) with a θ-free loss (polynomial prefactor, θ-exponent at most $1/(2T)$, $T = \lceil \log_2(2ed) \rceil$) would give MaxInf $\geq (\mathrm{Var}/d)^{O(\log d)}$; with a quasi-polynomial prefactor it is proved in Sec. 10 and gives Thm. 10.1; at the loss $\mathrm{poly}(d)\,\theta^{-2}$ it holds (Lemma 9.20) and gives nothing new; at a θ-free loss, two of its three pieces are proved	proved here (tools audited); SM-OF with a quasi-polynomial prefactor proved here, audited; open with a polynomial prefactor	9
One restriction at rate $1/d$ (Sec. 11)	spike forcing at a polynomial threshold; a conditional theorem in capture, amplification and bounded-delayed-activation forms, with bounds $(\mathrm{Var}/d)^{2747}m$, $(\mathrm{Var}/d)^{2825+8C_0}$ and $(\mathrm{Var}/d)^{2855+2C_0}$ (Thm. 11.5), circular as a reduction but unconditional for heavy juntas and the class $\mathcal{G}$ (Cor. 11.7); bias triggers fail (address functions, parity gates); the anti-spike inequality ASR implies the conjecture (Prop. 11.12)	proved here (audited; constants of Prop. 11.12 unaudited); ASR open	11
Capped routes (Props. 2.8, H.3, H.4)	juntas, ℓ_q sums ($q < 1$), single coefficients, y-root sums: capped at $2^{-\Omega(d)}$; linear deficit transfer false where needed; [35, Conjecture 37] false, also in symmetric form	proved here	2.4, H
Barriers (Props. 7.11, 7.12, 7.13)	bounded F with $\mathrm{Var}F \geq c$, $\mathrm{MaxInf}(F) \leq e^{-\Omega(s(F))}$ (s the real sensitivity); linear dual certificates give at most MaxInf $\geq 4\mathrm{Var}(\log_2 n)^2/n$	proved here	7.4
Conjecture 1.1	$\mathrm{MaxInf}(f) \geq c(\mathrm{Var}f/d)^C$	open	12
Question 4.16	shallow z-trees with leaf root sums $\mathrm{poly}(k/V)$ on half of the variance (suffices for the conjecture)	open	4
Conjecture 7.9	degree-free: every $g\colon Q_n \to [-1,1]$ is ε-close in L^2 to a decision tree of depth $\mathrm{poly}(\mathrm{rbs}(g)/\varepsilon)$, rbs the real block sensitivity (implies the conjecture)	open	7.3
Conjecture 8.8	degree-free: $L\Delta(g;\theta,L) \leq A\,\mathrm{rbs}(g)^{1/c}$ for bounded g; implies MaxInf $\geq (\mathrm{Var}/d)^{O(C+(1+2/c)\log d)}$ (Prop. 8.9)	open	8
SM-OF, polynomial prefactor	SM-OF (Def. 9.16) at a θ-free loss $(d/V)^{O(1)}\theta^{-c}$, $cT \leq \frac{1}{2}$; would remove the factor $\exp(-O(\log^4 d))$ from Thm. 10.1; implied by the conjecture	open	9
Question 11.11	anti-spike inequality: a relatively regular bounded function of degree d and small height is rarely spiked by one rate-$1/d$ restriction; implied by (LJ), implies the conjecture (Prop. 11.12), not known to follow from it	open	11
Conjecture M.3 (quantum OSSS)	$\mathrm{Var}^2 \leq 4t\sum_k \delta_k \mathrm{Inf}_k$ for biases of t-query algorithms; proved for $t = 1$, for $t = 2$ up to the constant, and for nonadaptive algorithms (constant 4); $\mathrm{poly}(t)$ form $\Rightarrow$ polynomial classical simulation	open	M

TABLE 2. Main results, continued: reformulations, routes, barriers and open statements.

are capped, and the adaptive address function (Section 4.4) that the trees must in general be adaptive. Section 5 combines the Helly property of subcubes with the Motzkin–Straus inequality [39] (Lemma 5.1). Section 6 and Appendices B–F treat the remaining classes of Table 1 and, in addition, disjoint-support and r-sparse separators (Theorems F.1 and F.3, Corollary F.4), functions of one linear form (Theorem 6.12 for regular weights, and with a bound polynomial in m for weights with a regular tail after m leading weights; Proposition 6.10 for general weights at the natural scale), further wreath bounds (Theorems B.1 and B.8, Corollary B.9), local cyclic functions (Theorem D.11) and product selectors on the torus (Theorem E.6). Graph properties, cyclically invariant functions and wreath-symmetric functions are transitive, so the conjecture for them in each fixed degree already follows from [19, Proposition 3.18] (Section 1.2). For graph properties Theorem 6.16 gives a bound polynomial in both $\mathrm{Var} f$ and d; it applies the clone-and-subsample technique of Zhandry's small-range lemma [57, 58] to influences. The graph and cyclic results in the third group of Table 1 are not bounds polynomial in d uniformly in n, except on the first three Fourier levels of cyclically invariant functions and in the limit of Theorem 6.14(ii). Theorems 6.18, 6.19 and 6.15 improve the constants obtained from the proof of [19, Proposition 3.18], while the wreath bounds, Theorem 6.20, the limit in Theorem 6.14(ii) and Theorem D.11 go beyond it. Section 7 collects statements equivalent to the conjecture, a sufficient one (an average-case Nisan lemma, Theorem 7.5), a degree-free strengthening in terms of real block sensitivity (Conjecture 7.9), and two barriers for proofs of the general statement (Section 7.4); Section 8 treats the touching transfer and a degree-free condition that would make it work (Conjecture 8.8). Section 9 analyses the spikes created by one restriction: priming turns many captured spikes into large real block sensitivity, so captured spikes carry little mass, and every capture-type hypothesis is false, equivalent to TT, or implies the small-spiked-mass condition SM (Theorem 9.9). The remaining statement, the mass bound SM-OF for old-free functions at rate $\frac{1}{2}$, holds at the loss $\mathrm{poly}(d)\,\theta^{-2}$; it is reduced, by an equivalence (Theorem 9.30), to its large-jump part, with size-biased tools proved in Appendix J. Section 10 proves SM-OF at a loss of θ-exponent $1/(2T)$ with a quasi-polynomial prefactor (Theorem 10.26). Its proof is a recursion over chains of derivatives on a grid of heights, in which lightness is measured against a fixed fraction of the node's weight, and in which every node uses its own two-stage exploration (Lemma 10.13). Through a chain with a θ-dependent loss (Lemma 10.2), a variant of the chain of Section 8, this gives the quasi-polynomial bound of Theorem 10.1. With a polynomial prefactor SM-OF remains open. Through the greedy reduction of [2], Theorem 10.1 also gives a classical simulation of t-query quantum algorithms on most inputs with $\exp(O(\log^4 t + \log t \cdot \log(1/(\varepsilon\delta))))$ queries (Corollary 10.29). Section 11 takes a single restriction at rate $1/d$ instead: spike forcing at a polynomial threshold, admissible explorations and a small-bias estimate give a conditional theorem with explicit polynomial exponents, which is circular as a reduction but covers structured classes, and the residual is the regular case, some of whose hard instances are products of nearly Boolean functions; there the gap becomes the anti-spike inequality of Question 11.11. The appendices contain remarks on the literature, the long proofs and constant chains of Section 6, torus-bounded polynomials, separators, the one-scale transfer with the spike gadget and a search form (Appendix G), two refuted routes (a deficit transfer, and [35, Conjecture 37]), the limits of the touching transfer, numerical evidence (Appendix N) and the Lean 4 formalization (Appendix O). Appendix M proposes a quantum OSSS conjecture for biases of quantum query algorithms and proves it for one query with the sharp constant 4, for two queries with a larger constant, and for nonadaptive algorithms with the constant 4 in place of $4t$; the last case gives a classical simulation of nonadaptive t-query algorithms on most inputs with $\lceil 4t^2/(\varepsilon^4\delta^3)\rceil$ queries (Corollary M.17).

1.4. **What remains open.** Conjecture 1.1 itself. By Lemma 2.4 and Theorem 4.15 it is equivalent, up to polynomial losses, to the following statement about bounded one-block forms $F = \sum_j y_j g_j(z)$ of degree k and variance V: *some z-decision tree of depth* $\mathrm{poly}(k/V)$ *has, on average over its leaves, a y-variable of influence* $\mathrm{poly}(V/k)$. By the gluing lemma it suffices that a shallow tree make the leaf root sums $\sum_j \|g_j|_l\|_2$ polynomial on half of the variance (Question 4.16). This holds for the selector classes of Sections 4 and 5 (for tree-mediated register forms when the tree has polynomial depth) and, with the trivial tree, for the symmetric, block-symmetric, wreath-symmetric and smooth-statistic classes of Section 6, whose proofs bound the root sum $\sum_i \mathrm{Inf}_i(f)^{1/2}$. For cyclically invariant functions it is open, since it would give polynomial bounds for them; for graph properties it is open as well, although Theorem 6.16 gives a polynomial bound by another route. A positive answer implies the conjecture; we do not know whether the conjecture implies a positive answer. The touching transfer TT with a polynomial loss would give $\mathrm{MaxInf}(f) \geq (\mathrm{Var}f/d)^{O(\log d)}$ (Section 8); Theorem 10.1 reaches this bound up to the factor $\exp(-O(\log^4 d))$. The mass bound SM-OF of Section 9 is the mass form of TT that the methods of that section address. It implies TT. At a θ-free loss, with a polynomial prefactor and θ-exponent at most $1/(2T)$ with $T = \lceil \log_2(2ed) \rceil$ (Conjecture 9.29 at such an exponent), it would remove that factor. Section 10 proves it with a quasi-polynomial prefactor, and Remark 10.27 explains why that scheme gives this prefactor. At the loss $\mathrm{poly}(d)\,\theta^{-2}$ SM-OF holds unconditionally (Lemma 9.20, Proposition 9.21) and gives nothing new; Table 4 lists what each θ-exponent would give. SM-OF at a θ-free loss is implied by the conjecture (Remark 9.18), and it is not known to be equivalent to it. Along the one-shot route of Section 11 the gap is the anti-spike inequality of Question 11.11: it implies the conjecture, it follows from the large-jump bound (LJ), some of its hard instances are products of nearly Boolean counterexamples of lower degree, and it is not known to follow from the conjecture. For biases of quantum query algorithms, the quantum OSSS conjecture (Conjecture M.3) is open in general for adaptive algorithms with $t \geq 2$ queries; with any constant polynomial in t it would give Conjecture 1.1 for these biases, and hence a classical simulation of quantum query algorithms on most inputs with polynomially many queries (Appendix M). Section 12 states the remaining gap and the other open problems.

2. Preliminaries

2.1. **Notation.** For $f\colon Q_n \to \mathbb{R}$ and $i \in [n]$ let $x^{i\to\pm1}$ be x with x_i set to ±1, $x^{\oplus i}$ be x with x_i flipped, and

$$d_i f(x) = \tfrac{1}{2}\big(f(x^{i\to1}) - f(x^{i\to-1})\big) = \sum_{S\ni i} \widehat{f}(S) x^{S\setminus\{i\}},$$

which does not depend on x_i. Then $\mathrm{Inf}_i(f) = \mathbb{E}(d_i f)^2 = \frac{1}{4}\mathbb{E}(f(x) - f(x^{\oplus i}))^2$, the total influence is $I(f) = \sum_i \mathrm{Inf}_i(f) = \sum_S |S|\widehat{f}(S)^2$, and $\mathrm{Var}f \leq I(f) \leq (\deg f)\mathrm{Var}f$. For $0 \leq a \leq b$ let

$$W^{[a,b]}(f) = \sum_{a\leq|S|\leq b} \widehat{f}(S)^2, \qquad I^{[a,b]}(f) = \sum_{a\leq|S|\leq b} |S|\,\widehat{f}(S)^2, \qquad W^k = W^{[k,k]}, \quad W^{\leq k} = W^{[0,k]}.$$

For $p \geq 1$, $\|f\|_p = (\mathbb{E}|f|^p)^{1/p}$. We use the L^1 quantities

$$\Lambda_1(f) = \sum_i \|d_i f\|_1, \qquad \lambda_1(f) = \max_i \|d_i f\|_1, \qquad \lambda_2(f) = \max_i \|d_i f\|_2 = \mathrm{MaxInf}(f)^{1/2}.$$

The noise operator is $T_\rho f = \sum_S \rho^{|S|}\widehat{f}(S)x^S$; for $|\rho| \leq 1$ it is an average of f over a product measure and maps $\mathcal{B}_d(n)$ to itself. Hypercontractivity (Bonami): $\|T_\rho g\|_2 \leq \|g\|_{1+\rho^2}$ for $0 \leq \rho \leq 1$, and $\|h\|_q \leq (q-1)^{k/2}\|h\|_2$ for $\deg h \leq k$ and $q \geq 2$ [41, Chapter 9].

For vector-valued $h\colon Q_n \to H$, H a real Hilbert space, $d_i h$ is defined coordinatewise and $\mathrm{Inf}_i(h) = \mathbb{E}\|d_i h\|^2$.

Decision trees. A decision tree D on Q_n queries a coordinate at each internal node and branches on its value; we always assume that no coordinate is queried twice on a path. A leaf l is a subcube $\{x : x_{i_1} = b_1, \dots, x_{i_s} = b_s\}$; p_l is its probability, and given l the unqueried coordinates are uniform and independent. The depth t is the maximal path length, and $\delta_i = \Pr[D \text{ queries } i]$ is the revealment of i; thus $\sum_i \delta_i \le t$. A function is *leaf-determined* (computed by D) if it is constant on every leaf. For $g\colon Q_n \to \mathbb{R}$ we write $g|_l$ for the restriction of g to l, viewed as a function of the free coordinates with the uniform measure, and $\mathbb{E}[\,\cdot \mid l]$ for the corresponding expectation.

2.2. **The real-sensitivity bound.** We use the following classical lemma of Ehlich–Zeller and Rivlin–Cheney [17, 44], in the form of [40, Theorem 3.3].

Lemma 2.1 ([*known*])**.** *Let q be a real polynomial, $m \ge 1$, and suppose $b_1 \le q(i) \le b_2$ for all integers $0 \le i \le m$ and $|q'(\xi)| \ge c > 0$ for some real $\xi \in [0, m]$. Then $\deg q \ge \sqrt{cm/(c + b_2 - b_1)}$.*

Proof (for completeness). Let $d = \deg q \ge 1$ and $c' = \max_{[0,m]} |q'| \ge c$. Every $x \in [0, m]$ is within $1/2$ of an integer $i \in [0, m]$, so $|q(x) - q(i)| \le c'/2$ and $q([0, m])$ lies in an interval of length $b_2 - b_1 + c'$. Markov's inequality on $[0, m]$ gives $c' \le d^2(b_2 - b_1 + c')/m$, i.e. $d^2 \ge c'm/(c' + b_2 - b_1) \ge cm/(c + b_2 - b_1)$, as $x \mapsto xm/(x + b_2 - b_1)$ is nondecreasing. □

Lemma 2.2 (real sensitivity bound; [*known*], cf. [19, Theorem 3.3], [43, Corollary 2.12])**.** *For every $f \in \mathcal{B}_d(n)$ and every $z \in Q_n$,*

$$\sum_{i=1}^{n} |d_i f(z)| \le 4d^2.$$

Proof. Replacing f by $x \mapsto f(z \odot x)$ (coordinatewise product), which has the same degree and range, we may assume $z = \mathbf{1} = (1, \dots, 1)$. Let $P = \{i : f(\mathbf{1}^{\oplus i}) > f(\mathbf{1})\}$ and $m = |P|$; assume $m \ge 1$. For $T \subseteq P$ let $\mathbf{1}^{\oplus T}$ be $\mathbf{1}$ with the coordinates in T flipped, and for $k = 0, \dots, m$ put $q(k) = \mathbb{E}_{T \subseteq P, |T| = k} f(\mathbf{1}^{\oplus T})$ (T uniform among k-subsets). Since $f(\mathbf{1}^{\oplus T}) = \sum_S \widehat{f}(S)(-1)^{|S \cap T|}$ and

$$(-1)^{|S \cap T|} = \prod_{i \in S \cap P} (1 - 2 \cdot \mathbf{1}[i \in T]) = \sum_{U \subseteq S \cap P} (-2)^{|U|} \mathbf{1}[U \subseteq T],$$

$$\Pr_{|T| = k}[U \subseteq T] = \frac{k(k-1)\cdots(k - |U| + 1)}{m(m-1)\cdots(m - |U| + 1)},$$

q agrees on $\{0, \dots, m\}$ with a real polynomial of degree at most $\max_S |S \cap P| \le d$, and $|q| \le 1$ there. Moreover $q(1) - q(0) = \Delta/m$ with $\Delta = \sum_{i \in P}(f(\mathbf{1}^{\oplus i}) - f(\mathbf{1})) \in (0, 2m]$, so $q'(\xi) = \Delta/m$ for some $\xi \in (0, 1)$. Lemma 2.1 with $c = \Delta/m$ and $b_2 - b_1 = 2$ gives $d^2 \ge \Delta/(\Delta/m + 2) \ge \Delta/4$. Applying the same argument to $-f$ handles $\{i : f(\mathbf{1}^{\oplus i}) < f(\mathbf{1})\}$. Hence $\sum_i |f(\mathbf{1}) - f(\mathbf{1}^{\oplus i})| \le 8d^2$, and $|d_i f(\mathbf{1})| = \frac{1}{2}|f(\mathbf{1}) - f(\mathbf{1}^{\oplus i})|$. □

Remark 2.3. The sharp constant is d^2 [19, Theorem 3.3] (via a Banach-space Markov inequality); it is attained asymptotically by $T_d(n^{-1}\sum_i x_i)$ at $\mathbf{1}$. All constants $4d^2$ below may be replaced by d^2. Averaging over z gives $\Lambda_1(f) \le 4d^2$ for $f \in \mathcal{B}_d$.

2.3. **One-block forms and the gradient form.** A *one-block form* is $F(y, z) = \sum_{j=1}^{m} y_j g_j(z)$ on $Q_m \times Q_n$, with $\deg g_j \le k$. It is *bounded* ($|F| \le 1$) if and only if $\sum_j |g_j(z)| \le 1$ for all z. Since the y_j are orthonormal and independent of z,

$$\mathrm{Var} F = V := \sum_j \mathbb{E} g_j^2, \qquad \mathrm{Inf}_{y_j}(F) = \mathbb{E} g_j^2, \qquad \mathrm{Inf}_{z_i}(F) = \sum_j \mathrm{Inf}_i(g_j). \tag{1}$$

We write $\varepsilon = \varepsilon(F) = \max_j \mathbb{E} g_j^2$ and $\tau = \tau(F) = \max_i \mathrm{Inf}_{z_i}(F)$, so $\mathrm{MaxInf}(F) = \max(\varepsilon, \tau)$; note $V \le \mathbb{E}(\sum_j |g_j|)^2 \le 1$. The *root sum* of the family is $R = \sum_j \|g_j\|_2$.

Lemma 2.4 (gradient form; [*known*] [43, Definition 1.1, Theorem 2.13], explicit constants)**.** *For $f \in \mathcal{B}_d(n)$ let $F(y,z) = (4d^2)^{-1} \sum_{i=1}^n y_i \, d_i f(z)$ on Q_{2n}. Then $F \in \mathcal{B}_d(2n)$, $\mathbb{E}F = 0$, and*

$$\mathrm{Var} F = \frac{I(f)}{16 d^4} \ge \frac{\mathrm{Var} f}{16 d^4}, \qquad \mathrm{Inf}_{y_i}(F) = \frac{\mathrm{Inf}_i(f)}{16 d^4},$$
$$\mathrm{Inf}_{z_j}(F) = \frac{\sum_{S \ni j} (|S|-1) \widehat{f}(S)^2}{16 d^4} \le \frac{(d-1)\mathrm{Inf}_j(f)}{16 d^4}.$$

Consequently, if Conjecture 1.1 holds for one-block forms with constants c, C, then it holds for all $f \in \mathcal{B}_d$ in the form $\mathrm{MaxInf}(f) \ge 16 d^3 c \, (\mathrm{Var} f/(16 d^5))^C$.

Proof. Boundedness is Lemma 2.2. In Fourier terms $F = (4d^2)^{-1} \sum_S \widehat{f}(S) \sum_{i \in S} y_i z^{S \setminus \{i\}}$, and the monomials $y_i z^{S \setminus \{i\}}$ are distinct for distinct pairs (i, S); this gives the degree bound and all formulas. For the consequence, $\mathrm{MaxInf}(F) \le \max(1, d-1)\mathrm{MaxInf}(f)/(16 d^4) \le \mathrm{MaxInf}(f)/(16 d^3)$ and $\mathrm{Var} F \ge \mathrm{Var} f/(16 d^4)$. □

Proposition 2.5 ([*known*] [43, eq. (1) and Theorem 2.13])**.** *For $f \in \mathcal{B}_d$, $\mathrm{MaxInf}(f) \ge e^{2-2d} I(f)^2/(16 d^4)$.*

Proof. By Lemma 2.2, $\sum_i \|d_i f\|_1 \le 4d^2$; since $\deg d_i f \le d-1$, $\|d_i f\|_2 \le e^{d-1} \|d_i f\|_1$ [41, Theorem 9.22]. Hence $I(f) = \sum_i \|d_i f\|_2^2 \le \mathrm{MaxInf}(f)^{1/2} \sum_i \|d_i f\|_2 \le \mathrm{MaxInf}(f)^{1/2} \, 4d^2 e^{d-1}$. □

Remark 2.6 ($\log n$ losses are harmless; [*known*], cf. [35, ECCC version, Comment 1.3]). Let $f \in \mathcal{B}_d(n)$ and $\varepsilon = \mathrm{Var}(f)/2$. By the DFKO junta theorem [16, Theorem 1] there is a set J with $|J| \le 2^{O(d)}/\varepsilon^2$ and a function of x_J within L^2-distance $\varepsilon^{1/2}$ of f. Then $g = \mathbb{E}[f \mid x_J]$ satisfies $g \in \mathcal{B}_d(|J|)$, $\mathrm{Var} f - \mathrm{Var} g = \|f-g\|_2^2 \le \varepsilon$ (so $\mathrm{Var} g \ge \mathrm{Var} f/2$), and $\mathrm{Inf}_i(g) = \sum_{i \in S \subseteq J} \widehat{f}(S)^2 \le \mathrm{Inf}_i(f)$. Hence any bound of the form $\mathrm{MaxInf}(f) \ge \mathrm{poly}(\mathrm{Var} f/d)/\mathrm{polylog}(n)$, valid for all n, implies Conjecture 1.1. (Losses polynomial in n only give back $2^{-O(d)}$.)

2.4. **The address function and three capped routes.** Let $k \ge 1$, $N = 2^k$, and $\iota \colon Q_k \to [N]$ a bijection. The *address function* is

$$\mathrm{Addr}_k(a, x) = x_{\iota(a)}, \qquad a \in Q_k, \ x \in Q_N.$$

Proposition 2.7 ([*standard*])**.** Addr_k *is Boolean,* $\deg \mathrm{Addr}_k = k+1$, $\mathrm{VarAddr}_k = 1$, $\mathrm{Inf}_{x_b} = 2^{-k}$ *for every target b,* $\mathrm{Inf}_{a_l} = 1/2$ *for every address bit, and every Fourier coefficient satisfies* $\widehat{\mathrm{Addr}_k}(S)^2 \in \{0, 4^{-k}\}$.

Proof. With $\beta(b) = \iota^{-1}(b)$, $\mathbf{1}[a = \beta] = \prod_l \frac{1+\beta_l a_l}{2} = 2^{-k} \sum_{T \subseteq [k]} \beta^T a^T$, so $\mathrm{Addr}_k = 2^{-k} \sum_b \sum_T \beta(b)^T x_b a^T$, a sum of 4^k distinct monomials with coefficients $\pm 2^{-k}$. Hence the degree is $k+1$, $\mathbb{E}\mathrm{Addr}_k = 0$, $\mathrm{Var} = 1$, $\mathrm{Inf}_{x_b} = 2^k 4^{-k}$ and $\mathrm{Inf}_{a_l} = 2^k 2^{k-1} 4^{-k} = 1/2$. □

Proposition 2.8 (capped routes; [*proved here*])**.** *Let $f = \mathrm{Addr}_k$, so $d = \deg f = k+1$, $\mathrm{Var} f = 1$ and $\mathrm{MaxInf} f = 1/2$.*

- *(i)* Juntas. *If J is a set of coordinates with $\sum_{S \not\subseteq J} \widehat{f}(S)^2 \le 1/2$, then J contains at least 2^{k-1} targets. Hence the (true) implication "$\|f-g\|_2^2 \le \varepsilon$ for an L-junta $g \Rightarrow \mathrm{MaxInf} f \ge (\mathrm{Var} f - \varepsilon)/L$" cannot give better than 2^{1-k}.*
- *(ii)* ℓ_q sums. *For every $q < 1$, $\sum_i \mathrm{Inf}_i(f)^q \ge 2^{k(1-q)}$; so no bound $\sum_i \mathrm{Inf}_i^q \le \mathrm{poly}(d)$ holds on $\mathcal{B}_d$.*

(iii) Single coefficients. $\max_S \widehat{f}(S)^2 = 4^{-k}$*; so no bound* $\max_{S\neq\emptyset} \widehat{f}(S)^2 \geq \mathrm{poly}(\mathrm{Var}f/d)$ *holds on* $\mathcal{B}_d$.

(iv) y-root sums. *As a one-block form,* $f(a,x) = \sum_b x_b\, g_b(a)$ *with* $g_b = \mathbf{1}[\iota(\cdot) = b]$, $\sum_b |g_b| \equiv 1$, *and* $\sum_b \|g_b\|_2 = 2^{k/2}$. *Hence the one-line argument of Proposition 2.5 (which uses only y-influences) cannot be improved beyond 2^{-k} for one-block forms; any proof must use the influences of the z-variables (the address bits).*

Proof. (i) Each target $b \notin J$ contributes the 2^k monomials $x_b a^T$, of total weight 2^{-k}, to $\sum_{S\not\subseteq J} \widehat{f}(S)^2$; so at most 2^{k-1} targets lie outside J. (ii) The 2^k targets contribute $2^k \cdot 2^{-kq}$. (iii) and (iv) are immediate from Proposition 2.7. □

3. Top-heaviness

In this section $\tau = \mathrm{MaxInf}(f)$.

3.1. The level-k inequality on derivatives.

Lemma 3.1 ([*known*]; Bonami–Hölder, as in [28, 52], [41, §9.6])**.** *Let* $g\colon Q_n \to \mathbb{R}$, $k \geq 0$ *and* $\delta \in (0,1]$. *Then* $W^{\leq k}(g) \leq \delta^{-k}\|g\|_{1+\delta}^2$. *In particular:*

(a) if $|g| \leq 1$ *then* $W^{\leq k}(g) \leq \delta^{-k}\|g\|_1^{2/(1+\delta)}$;
(b) $W^{\leq k}(g) \leq 3^k \|g\|_1\|g\|_2$.

Proof. Let $q = 1 + 1/\delta \geq 2$, with conjugate exponent $1+\delta$, and $g^{\leq k} = \sum_{|S|\leq k} \widehat{g}(S)x^S$. By Hölder and Bonami's inequality ($\|h\|_q \leq (q-1)^{k/2}\|h\|_2$ for $\deg h \leq k$),

$$W^{\leq k}(g) = \mathbb{E}[g\, g^{\leq k}] \leq \|g\|_{1+\delta}\|g^{\leq k}\|_q \leq \|g\|_{1+\delta}\, \delta^{-k/2}\, W^{\leq k}(g)^{1/2}.$$

(a) If $|g| \leq 1$ then $\mathbb{E}|g|^{1+\delta} \leq \mathbb{E}|g|$. (b) Take $\delta = 1/3$; by Hölder with exponents $3/2$ and 3, $\mathbb{E}|g|^{4/3} = \mathbb{E}[|g|^{2/3}|g|^{2/3}] \leq (\mathbb{E}|g|)^{2/3}(\mathbb{E}g^2)^{1/3}$, i.e. $\|g\|_{4/3}^2 \leq \|g\|_1\|g\|_2$. □

Theorem 3.2 ([*standard*]; part (b) is in substance [29, Lemma 4.4])**.** *Let* $G\colon Q_n \to [-1,1]$ *(no degree assumption) and* $k \geq 0$. *Then*

(a) for every $\delta \in (0,1]$*:* $\quad W^{[1,k+1]}(G) \leq I^{[1,k+1]}(G) \leq \delta^{-k}\, \lambda_1(G)^{\frac{1-\delta}{1+\delta}}\, \Lambda_1(G)$;
(b) $I^{[1,k+1]}(G) \leq 3^k \sum_i \|d_iG\|_1\|d_iG\|_2 \leq 3^k\, \lambda_2(G)\, \Lambda_1(G)$.

Since $|d_iG| \leq 1$ *we have* $\lambda_2(G) \leq \lambda_1(G)^{1/2}$, *so (b) is at least as strong as (a) with* $\delta = 1/3$.

Proof. Fix i and let $g = d_iG$, so $|g| \leq 1$. For $T \not\ni i$, $\widehat{g}(T) = \widehat{G}(T\cup\{i\})$, and $\widehat{g}(T) = 0$ if $i \in T$; hence $W^{\leq k}(d_iG) = \sum_{S\ni i,\, |S|\leq k+1} \widehat{G}(S)^2$ and

$$\sum_i W^{\leq k}(d_iG) = \sum_{1\leq|S|\leq k+1} |S|\, \widehat{G}(S)^2 = I^{[1,k+1]}(G).$$

(a) By Lemma 3.1(a), $W^{\leq k}(d_iG) \leq \delta^{-k}\|d_iG\|_1^{(1-\delta)/(1+\delta)}\|d_iG\|_1 \leq \delta^{-k}\lambda_1(G)^{(1-\delta)/(1+\delta)}\|d_iG\|_1$; sum over i. (b) Lemma 3.1(b), summed over i. Finally $\|d_iG\|_2^2 \leq \|d_iG\|_1$ because $|d_iG| \leq 1$. □

The degree of f enters only through $\Lambda_1(f) \leq 4d^2$ (Lemma 2.2) and $\lambda_2(f) = \mathrm{MaxInf}(f)^{1/2}$.

Corollary 3.3 (counterexamples are top-heavy; [*proved here*]; known in substance [29, Lemma 4.4])**.** *For every* $f \in \mathcal{B}_d$ *and every* $k \geq 0$,

$$\mathrm{MaxInf}(f) \;\geq\; \Big(\frac{W^{[1,k+1]}(f)}{4d^2\, 3^k}\Big)^2.$$

Equivalently, $W^{[1,j]}(f) \leq 4d^2\, 3^{j-1}\mathrm{MaxInf}(f)^{1/2}$ *for all* $j \geq 1$. *In quantile form: if* $q \in (0,1]$, $\mathrm{Var}f > 0$, *and* $j_q = \min\{j \geq 1 : W^{[1,j]}(f) \geq q\mathrm{Var}f\}$, *then* $\mathrm{MaxInf}(f) \geq \big(q\mathrm{Var}f/(4d^2\, 3^{j_q-1})\big)^2$.

Proof. Theorem 3.2(b) with $G = f$, $\lambda_2(f) = \mathrm{MaxInf}(f)^{1/2}$ and $\Lambda_1(f) \le 4d^2$. For the quantile form take $k = j_q - 1$ ($j_q \le d$ exists since $W^{[1,d]}(f) = \mathrm{Var} f$). □

In particular, if $W^{[1,j]}(f) \ge \mathrm{Var}(f)/d^A$ for some $j \le 1 + \log_3 d$, then $\mathrm{MaxInf}(f) \ge \mathrm{Var}(f)^2/(16d^{2A+6})$. Thus a counterexample to Conjecture 1.1 must have all but a polynomially small part of its Fourier weight above level $c \log d$. With exponent 2 the corollary is essentially [29, Lemma 4.4]; Theorem 3.2(a) with $\delta = 1/3$ alone gives only the weaker bound $\mathrm{MaxInf}(f) \ge (W^{[1,k+1]}(f)/(4d^2 3^k))^4$.

The relation of Corollary 3.3 to the tail theorem of [12] is discussed in Remark A.1 (Appendix A).

Remark 3.4 (top-heaviness does not control spikiness; [*proved here*]). Let $B_1, \dots, B_{2^k}$ be disjoint blocks of size k and $f(a, x) = \chi_{B_{\iota(a)}}(x)$, $a \in Q_k$, where $\chi_B(x) = x^B$, $\iota \colon Q_k \to [2^k]$ is a bijection, $\beta(b) = \iota^{-1}(b) \in Q_k$ is the address of block b, and $\beta(b)^U = \prod_{i \in U} \beta(b)_i$. Then $\widehat{f}(U \cup B_b) = 2^{-k} \beta(b)^U$, so every monomial contains a full block: $W^{<k}(f) = 0$, $\deg f = 2k$, and f is Boolean. For $j \in B_b$, $d_j f = \mathbf{1}[a = \beta(b)] \chi_{B_b \setminus \{j\}}$, so $\|d_j f\|_2 / \|d_j f\|_1 = 2^{k/2} = 2^{\deg f/4}$, the maximal "spikiness" allowed by hypercontractivity up to the base. Nevertheless every address bit has influence $1/2$. So any improvement over Proposition 2.5 for top-heavy functions must exploit small MaxInf, not top-heaviness alone.

3.2. **Lipschitz images and level sets.**

Theorem 3.5 (Lipschitz images of counterexamples are top-heavy; [*proved here*])**.** *Let $f \in \mathcal{B}_d$, $\tau = \mathrm{MaxInf}(f)$, and let $P \colon [-1, 1] \to [-1, 1]$ be L-Lipschitz (not necessarily a polynomial). Then for every $k \ge 0$,*

$$W^{[1,k+1]}(P \circ f) \le I^{[1,k+1]}(P \circ f) \le 4d^2\, 3^k\, L\, \min\{1, L\sqrt{\tau}\}.$$

Proof. Let $G = P \circ f \colon Q_n \to [-1, 1]$. Pointwise, $|d_i G(x)| = \frac{1}{2} |P(f(x^{i \to 1})) - P(f(x^{i \to -1}))| \le L |d_i f(x)|$. Hence $\Lambda_1(G) \le L \Lambda_1(f) \le 4d^2 L$ and $\lambda_2(G) \le \min\{1, L\lambda_2(f)\} = \min\{1, L\sqrt{\tau}\}$. Apply Theorem 3.2(b). □

The same argument applies to any $G \colon Q_n \to [-1, 1]$ with $|d_i G| \le L |d_i f|$ pointwise, and, since $d_i T_\rho = \rho T_\rho d_i$ and T_ρ is a contraction on every L^p, to $T_\rho(P \circ f)$.

Theorem 3.6 (level sets of counterexamples are noise sensitive on average; [*proved here*])**.** *Let $f \in \mathcal{B}_d$ with $\tau = \mathrm{MaxInf}(f) > 0$, and $h_\theta = \mathbf{1}\{f > \theta\}$. Then for every $k \ge 0$,*

$$\int_{-1}^{1} W^{[1,k+1]}(h_\theta)\, d\theta \;\le\; (16d^2\, 3^k + 2)\, \tau^{1/6}.$$

Proof. Let $L > 0$ and $P_\theta(s) = \min\{1, \max\{0, L(s - \theta)\}\}$, which is L-Lipschitz with values in $[0, 1]$. Then $P_\theta \circ f = h_\theta$ except on $\{\theta < f < \theta + 1/L\}$, where they differ by at most 1; hence $\|P_\theta \circ f - h_\theta\|_2^2 \le \Pr[\theta < f < \theta + 1/L]$ and

$$\int_{-1}^{1} \|P_\theta \circ f - h_\theta\|_2^2\, d\theta \le \mathbb{E}\left|\{\theta : f - 1/L < \theta < f\}\right| \le \frac{1}{L}.$$

Since $W^{[1,k+1]}(\cdot) = \|\Pi \cdot\|_2^2$ for an orthogonal projection Π, $W^{[1,k+1]}(h_\theta) \le 2W^{[1,k+1]}(P_\theta \circ f) + 2\|P_\theta \circ f - h_\theta\|_2^2$. By Theorem 3.5, $W^{[1,k+1]}(P_\theta \circ f) \le 4d^2 3^k L^2 \sqrt{\tau}$. Integrating over $\theta \in [-1, 1]$,

$$\int_{-1}^{1} W^{[1,k+1]}(h_\theta)\, d\theta \le 16d^2\, 3^k L^2 \sqrt{\tau} + \frac{2}{L},$$

and $L = \tau^{-1/6}$ gives the claim. □

Remark 3.7. Since $f+1=\int_{-1}^{1} h_\theta\,d\theta$, Minkowski's and the Cauchy–Schwarz inequality give $W^{[1,K]}(f)\le 2\int_{-1}^{1} W^{[1,K]}(h_\theta)\,d\theta$, so Theorem 3.6 contains a weak form of Corollary 3.3. Writing $\sum_{S\neq\emptyset}\rho^{|S|}\widehat{h}(S)^2 \le W^{[1,m]}(h)+\rho^{m+1}\mathrm{Var}h$ shows that for a counterexample the level sets h_θ have, on average over θ, noise stability at most $(\mathbb{E}h_\theta)^2$ plus a polynomially small quantity, for every fixed noise rate $1-\rho\ge c>0$ (quantitatively, by Corollary 5.12(3), an average excess stability η forces $\mathrm{MaxInf}(f)\ge(\eta/(36d^2))^6\,3^{-6\ln(1/\eta)/(1-\rho)}$; at noise rate $c/\log d$ this is only quasi-polynomial).

Remark 3.8 (images cannot create low-level weight; [*proved here*]). Theorem 3.5 cannot be reversed by choosing a clever P. Let A_1,A_2 be copies of Addr_k on disjoint sets of variables and $f=(A_1+A_2)/2\in\{-1,0,1\}$; then $\mathrm{Var}f=1/2$ and $\mathrm{MaxInf}f=1/8$. Every $P\colon\{-1,0,1\}\to[-1,1]$ satisfies $P\circ f=\alpha+\beta f+\gamma f^2$ with $|\beta|\le 1$, $|\gamma|\le 2$, and $f^2=(1+A_1A_2)/2$. By Proposition 2.7, $W^{[1,m]}(\mathrm{Addr}_k)=2^{-k}\sum_{j<m}\binom{k}{j}\le 2^{-k}k^m=:\omega$ for $k\ge 2$, and $W^{[1,m]}(A_1A_2)\le\omega^2$ because the two factors are independent. Hence, if $\omega\le 1$, $W^{[1,m]}(P\circ f)\le 2\big(\beta^2\frac{\omega}{2}+\frac{\gamma^2}{4}\omega^2\big)\le 3\omega$. For $m=O(\log d)$, $d=k+1$, this is $2^{-k+o(k)}$, while $\mathrm{Var}f=1/2$.

4. Gluing along decision trees

Throughout this section $F(y,z)=\sum_j y_j g_j(z)$ is a bounded one-block form ($\sum_j|g_j|\le 1$ pointwise) with $\deg g_j\le k$, and V,ε,τ are as in (1). For a decision tree D on the z-variables with leaves l we write

$$c_{jl}=\mathbb{E}[g_j^2\mid l]^{1/2},\qquad V_l=\sum_j c_{jl}^2=\mathrm{Var}(F|_l),\qquad R_l=\sum_j c_{jl},\qquad \kappa_D=\sum_l p_l\sum_j c_{jl}^4,$$

so that $V=\sum_l p_l V_l$ and $c_{jl}^2=\mathrm{Inf}_{y_j}(F|_l)$, where $F|_l$ is the restriction of F to the leaf. Let $\tau_D=\max_{i\in D}\mathrm{Inf}_{z_i}(F)$, the maximum over the variables queried by D.

4.1. A vector-valued OSSS inequality.

Theorem 4.1 ([*known*]: [42, Theorems 3.1–3.3, Corollary 3.4] and the proof of [42, eq. (6)]; see the paragraph *Sources* after the proof)**.** *Let D be a decision tree on Q_n of depth t with revealments δ_i, let H be a real Hilbert space, let $f\colon Q_n\to H$ be leaf-determined and $h\colon Q_n\to H$ arbitrary, and* $\mathrm{Cov}(f,h)=\mathbb{E}\langle f,h\rangle-\langle\mathbb{E}f,\mathbb{E}h\rangle$. *Then*

$$\text{(L1)}\qquad |\mathrm{Cov}(f,h)|\le\sup_x\|f(x)\|\ \sum_i\delta_i\,\mathbb{E}\|d_ih\|,$$

$$\text{(L2)}\qquad |\mathrm{Cov}(f,h)|\le\|f-\mathbb{E}f\|_{L^2(H)}\Big(2t\sum_i\delta_i\mathrm{Inf}_i(h)\Big)^{1/2}.$$

Proof. Let x,y be independent and uniform on Q_n. Run D on x; let $s\le t$ be the number of queries and $i_1,\dots,i_s$ the (distinct) queried coordinates, $J=\{i_1,\dots,i_s\}$. For $0\le m\le t$ let $z^{(m)}$ be x with the coordinates $i_1,\dots,i_{\min(m,s)}$ replaced by those of y; thus $z^{(0)}=x$ and $z^{(t)}=u:=(y_J,x_{J^c})$.

Step 1. Conditionally on the leaf of x, which is an event $\{x_{i_1}=b_1,\dots,x_{i_s}=b_s\}$ with fixed i's and b's, the coordinates x_{J^c} are uniform and y is independent and uniform; hence u is uniform and independent of the leaf. Since $f(x)$ depends only on the leaf, $\mathbb{E}\langle f(x),h(u)\rangle=\langle\mathbb{E}f,\mathbb{E}h\rangle$, and since x and u are both uniform,

$$\mathrm{Cov}(f,h)=\mathbb{E}\langle f(x),h(x)-h(u)\rangle=\mathbb{E}\langle f(x)-\mathbb{E}f,h(x)-h(u)\rangle.$$

Step 2. $h(x)-h(u)=\sum_{m=1}^{t}\Delta_m$ with $\Delta_m=h(z^{(m-1)})-h(z^{(m)})$, and $\Delta_m=0$ if $m>s$.

Step 3. Fix a node of D at depth $m-1$, labelled by the coordinate i, and let $E_\pi = \{x_{i_r} = b_r,\ r < m\}$ be the event that x reaches it; E_π depends only on $x_{i_1}, \dots, x_{i_{m-1}}$. Conditionally on E_π, $z^{(m-1)}$ has the independent uniform coordinates $y_{i_1}, \dots, y_{i_{m-1}}$ in those positions and uniform coordinates of x elsewhere, so $z^{(m-1)}$ is uniform; and $z^{(m)}$ is $z^{(m-1)}$ with coordinate $i \notin \{i_1, \dots, i_{m-1}\}$ replaced by the independent uniform bit y_i. With probability $1/2$ nothing changes, and otherwise coordinate i is flipped, where $\|h(w) - h(w^{\oplus i})\| = 2\|d_i h(w)\|$. Hence for $p \in \{1, 2\}$,

$$\mathbb{E}\big[\|\Delta_m\|^p\, \mathbf{1}\{m \le s,\ i_m = i\}\big] = \Pr[m \le s,\ i_m = i]\ 2^{p-1}\, \mathbb{E}\|d_i h\|^p.$$

Since no coordinate is queried twice on a path, $\sum_{m=1}^{t} \Pr[m \le s,\ i_m = i] = \delta_i$.

(L1): $|\mathrm{Cov}(f, h)| \le \sup\|f\| \sum_m \mathbb{E}\|\Delta_m\| = \sup\|f\| \sum_i \delta_i \mathbb{E}\|d_i h\|$. (L2): by Cauchy–Schwarz, $|\mathrm{Cov}(f, h)| \le \|f - \mathbb{E}f\|_2 \sum_{m=1}^{t} (\mathbb{E}\|\Delta_m\|^2)^{1/2} \le \|f - \mathbb{E}f\|_2 \sqrt{t}\, (\sum_m \mathbb{E}\|\Delta_m\|^2)^{1/2}$, and $\sum_m \mathbb{E}\|\Delta_m\|^2 = 2\sum_i \delta_i \mathrm{Inf}_i(h)$. □

Sources. [42, Theorems 3.1–3.2] are stated for functions with values in a metric space, so they include Hilbert space; there the influence re-randomizes a coordinate, which for uniform bits is half the flip version used here. Theorem 3.1 gives (L1) for $h = f$, and Theorem 3.2 a two-function form. The squared-norm version, with the depth factor as in (L2), is [42, Theorem 3.3, Corollary 3.4]. The covariance form with a Hilbert-valued leaf-determined f is not stated in [42]; it follows from the proof of [42, eq. (6)] with Cauchy–Schwarz, which is the proof above.

Numerical remarks on the factor t are in Remark N.1 (Appendix N).

4.2. **The gluing lemma.** *Prior work.* The closest statements in print are [29, Theorem 6.4] (a decision tree yields an influential variable, via OSSS), [35, Claim 33] (an equivalent form of the conjecture via a single restriction of small variance) and the closing remark of [12] (a shallow tree that finds influential coordinates of restrictions would give the conjecture). The inequality $\mathrm{Inf}_i(f) \ge \mathbb{E}_{\mathrm{leaf}} \mathrm{Inf}_i(f|_{\mathrm{leaf}})$ for every i is folklore (flipping a coordinate that is not queried keeps the leaf), and a union bound over the leaves loses a factor 2^{-t}; the new part of Theorems 4.2 and 4.4 is the loss polynomial in t.

Theorem 4.2 (gluing lemma; [*proved here*])**.** *Let F be a bounded one-block form and D a decision tree on the z-variables of depth $t \ge 1$. Then*

$$\kappa_D - \varepsilon V \ \le\ 8t \sum_i \delta_i \mathrm{Inf}_{z_i}(F) \ \le\ 8t^2\, \tau_D.$$

Proof. Let $H = \ell_2$ (indexed by j) and consider the *mass vector* $q(z) = (g_j(z)^2)_j \in H$. Put $m_j = \mathbb{E}g_j^2$. Then $\mathbb{E}[q \mid l] = (c_{jl}^2)_j$ and $\mathbb{E}q = (m_j)_j$. Let $f = \mathbb{E}[q \mid \mathrm{leaf}] - \mathbb{E}q$, which is leaf-determined with $\mathbb{E}f = 0$. Since $\sum_l p_l c_{jl}^2 = m_j$,

$$B := \mathbb{E}\|f\|^2 = \sum_l p_l \sum_j (c_{jl}^2 - m_j)^2 = \kappa_D - \sum_j m_j^2 \ \ge\ \kappa_D - \varepsilon V,$$

using $\sum_j m_j^2 \le (\max_j m_j) \sum_j m_j = \varepsilon V$. As f is leaf-determined, $\mathrm{Cov}(f, q) = \mathbb{E}\langle f, \mathbb{E}[q \mid \mathrm{leaf}]\rangle = \mathbb{E}\langle f, f + \mathbb{E}q\rangle = B$. Theorem 4.1(L2) gives $B \le B^{1/2}(2t \sum_i \delta_i \mathrm{Inf}_i(q))^{1/2}$, i.e. $B \le 2t \sum_i \delta_i \mathrm{Inf}_i(q)$. Finally, with $g_j^{\pm} = g_j(z^{i \to \pm 1})$, $d_i(g_j^2) = (g_j^+ + g_j^-)\, d_i g_j$ and $|g_j^{\pm}| \le 1$, so $\mathrm{Inf}_i(q) = \sum_j \mathbb{E}(d_i(g_j^2))^2 \le 4 \sum_j \mathbb{E}(d_i g_j)^2 = 4\mathrm{Inf}_{z_i}(F)$. The last inequality uses $\delta_i = 0$ for $i \notin D$ and $\sum_i \delta_i \le t$. □

The lemma is "degree-free": k does not appear. Note that $\kappa_D \ge \mathbb{E}_l \max_j \mathrm{Inf}_{y_j}(F|_l)^2$; so the lemma glues heavy y-variables at the leaves of a shallow tree into a heavy variable of F, at the cost $9t^2$ (see Theorem 4.15). Applying Theorem 4.1 to $h = (g_j)_j$ instead of the mass vector can be useless: for $g_j = \mathbf{1}[T(z) = j]\, z_{c(j)}$, where T is a decision tree and the $z_{c(j)}$ are fresh variables, one has $\mathbb{E}[h \mid \text{leaf of } T] = 0$.

Corollary 4.3 (leaf root sums; [*proved here*])**.** *If $R_l \le R$ for every leaf l of D (depth $t \ge 1$), then $\kappa_D \ge V^3/R^2$ and*

$$\max(\varepsilon, \tau_D) \;\ge\; \frac{V^3}{16R^2t^2}.$$

For $t = 0$ (no tree) the same argument gives $\varepsilon \ge V^2/R^2$, where $R = \sum_j \|g_j\|_2$.

Proof. By Hölder with exponents 3/2 and 3, $V_l = \sum_j c_{jl}^{2/3} c_{jl}^{4/3} \le R_l^{2/3} (\sum_j c_{jl}^4)^{1/3}$, so $\sum_j c_{jl}^4 \ge V_l^3/R^2$; by Jensen, $\kappa_D \ge R^{-2} \sum_l p_l V_l^3 \ge V^3/R^2$. By Theorem 4.2, $V^3/R^2 \le \varepsilon V + 8t^2\tau_D$, so $\varepsilon \ge V^2/(2R^2)$ or $\tau_D \ge V^3/(16R^2t^2)$, and $V^2/(2R^2) \ge V^3/(16R^2t^2)$ since $V \le 1 \le t$. For $t = 0$: $V = \sum_j \|g_j\|_2^2 \le \varepsilon^{1/2} R$. □

4.3. **Gluing for general functions and closure under shallow switching.**

Theorem 4.4 (gluing inequality; [*proved here*])**.** *For every $f \in \mathcal{B}_d$ and every decision tree T of depth $t \ge 1$ with leaves l (restrictions $f|_l$),*

$$\mathrm{MaxInf}(f) \;\ge\; \frac{\mathbb{E}_l\big[\mathrm{MaxInf}(f|_l)^2\big]}{144\, d^5\, t^2}.$$

Proof. Apply Theorem 4.2 to the gradient form F of Lemma 2.4, $g_i = d_i f/(4d^2)$, with T regarded as a tree on the z-variables. If i is not queried on the path to l then $(d_i f)|_l = d_i(f|_l)$, so $\mathbb{E}[(d_i f)^2 \mid l] = \mathrm{Inf}_i(f|_l)$, and $\mathrm{Inf}_i(f|_l) = 0$ for queried i; hence $\kappa_T \ge \mathbb{E}_l \max_i \mathbb{E}[g_i^2 \mid l]^2 \ge \mathbb{E}_l \mathrm{MaxInf}(f|_l)^2/(256d^8)$. On the other side, by Lemma 2.4, $\varepsilon = \mathrm{MaxInf}(f)/(16d^4)$, $V = I(f)/(16d^4) \le d/(16d^4)$ and $\tau_T \le (d-1)\mathrm{MaxInf}(f)/(16d^4)$, so

$$\kappa_T \le \varepsilon V + 8t^2\tau_T \le \mathrm{MaxInf}(f)\Big(\frac{1}{256d^7} + \frac{t^2}{2d^3}\Big) \le \frac{9t^2}{16d^3}\mathrm{MaxInf}(f).$$

Comparing, $\mathrm{MaxInf}(f) \ge \frac{16d^3}{9t^2} \cdot \frac{\mathbb{E}_l \mathrm{MaxInf}(f|_l)^2}{256d^8}$. □

The trivial bound is $\mathrm{MaxInf}(f) \ge \max_l p_l \mathrm{MaxInf}(f|_l)$, which is exponentially small in t.

Theorem 4.5 (closure under shallow switching; [*proved here*])**.** *Let $\mathcal{C} \subseteq \mathcal{B}_d$ be a class with $\mathrm{MaxInf}(h) \ge c(\mathrm{Var}h/d)^C$ for all $h \in \mathcal{C}$, where $C \ge \frac{1}{2}$. If $f \in \mathcal{B}_d$ and $f|_l \in \mathcal{C}$ for every leaf l of a decision tree of depth $t \ge 1$, then*

$$\mathrm{MaxInf}(f) \;\ge\; \min\Big(\frac{\mathrm{Var}(f)^2}{4t^2},\; \frac{c^2(\mathrm{Var}f/(2d))^{2C}}{144\, d^5 t^2}\Big).$$

Proof. Let $\varphi = \mathbb{E}[f \mid \text{leaf}]$, which is leaf-determined with $|\varphi| \le 1$. Then $\mathrm{Var}f = B + W$ with $B = \mathrm{Var}\varphi = \mathrm{Cov}(\varphi, f)$ and $W = \mathbb{E}_l \mathrm{Var}(f|_l)$. By Theorem 4.1(L1) with $H = \mathbb{R}$, $B \le \sum_i \delta_i \mathbb{E}|d_i f| \le t\,\mathrm{MaxInf}(f)^{1/2}$; so if $B \ge \mathrm{Var}f/2$ then $\mathrm{MaxInf}(f) \ge \mathrm{Var}(f)^2/(4t^2)$. Otherwise $W \ge \mathrm{Var}f/2$, and by Jensen ($2C \ge 1$) $\mathbb{E}_l \mathrm{MaxInf}(f|_l)^2 \ge c^2 \mathbb{E}_l (\mathrm{Var}(f|_l)/d)^{2C} \ge c^2 (W/d)^{2C}$; apply Theorem 4.4. □

Classes that can be used in Theorem 4.5 include symmetric and block-symmetric functions (Proposition 6.2 and Theorem 6.4), Boolean functions, and functions with root influence $\sum_i \mathrm{Inf}_i^{1/2} \le \mathrm{poly}(d)$.

4.4. **Register lemmas: where complete boundedness is used.**

Lemma 4.6 (register lemma; [*proved here*], the conclusion is known: [23, Remark 4.1] for homogeneous forms, cf. [24, Theorem 1.3])**.** *Let H be a Hilbert space, and for $s = 1, \dots, t$ let $P_0^s, \dots, P_n^s$ be mutually orthogonal projections with $\sum_i P_i^s = I$, and $\pi_s(x) = P_0^s + \sum_{i=1}^n x_i P_i^s$. Let $V_1, \dots, V_{t-1}$ be contractions, $\|u\|, \|v\| \le 1$, and*

$$T(x^1, \dots, x^t) = \langle u, \pi_1(x^1) V_1 \pi_2(x^2) V_2 \cdots V_{t-1} \pi_t(x^t) v \rangle, \qquad (x^1, \dots, x^t) \in Q_n^t.$$

Then for every block s, $\sum_{j=1}^n \mathrm{Inf}_{x_j^s}(T)^{1/2} \le 1$, where $\mathrm{Inf}_{x_j^s}(T) = \mathbb{E}|d_{x_j^s} T|^2$.

Proof. For $x \in Q_n$, $\pi_s(x)$ is a self-adjoint unitary. T is affine in x_j^s, and

$$d_{x_j^s} T = \langle a, P_j^s b \rangle, \qquad a = \big(\pi_1(x^1) V_1 \cdots \pi_{s-1}(x^{s-1}) V_{s-1}\big)^* u, \quad b = V_s \pi_{s+1}(x^{s+1}) \cdots V_{t-1} \pi_t(x^t) v,$$

with $\|a\|, \|b\| \le 1$; a depends only on $x^{<s}$ and b only on $x^{>s}$. Let $\rho_a = \mathbb{E}\, aa^*$ and $\rho_b = \mathbb{E}\, bb^*$, positive operators of trace at most 1. By independence,

$$\mathrm{Inf}_{x_j^s}(T) = \mathbb{E}\,\mathrm{tr}\big(P_j b b^* P_j a a^*\big) = \mathrm{tr}\big((P_j \rho_a P_j)(P_j \rho_b P_j)\big) \le \mathrm{tr}(P_j \rho_a)\,\mathrm{tr}(P_j \rho_b),$$

since $\mathrm{tr}(AB) \le \mathrm{tr}A\, \|B\| \le \mathrm{tr}A\, \mathrm{tr}B$ for positive A, B. By Cauchy–Schwarz and $\sum_{j \ge 1} P_j \le I$, $\sum_j \mathrm{Inf}_{x_j^s}(T)^{1/2} \le (\mathrm{tr}\rho_a\, \mathrm{tr}\rho_b)^{1/2} \le 1$. □

By the Christensen–Sinclair factorization (as used in [4]), a t-linear form on ℓ_∞^{n+1} of cb norm at most 1 has a representation as in Lemma 4.6 in which the coordinate $x_0 = 1$ carries P_0^s (this needs the factorization with unital representations; the proof of Lemma 4.6 uses only $\sum_i P_i^s \le I$); for homogeneous block-multilinear forms ($P_0^s = 0$) Lemma 4.6 is the root-influence inequality of [23, Remark 4.1] for every block, and summing over s gives the bound of [24, Theorem 1.3]. (We have not compared the convention with a free constant coordinate with the normalization of the cb norm in [7, 24] for non-homogeneous forms.) The lemma isolates the mechanism: *a derivative of a cb form is a pairing of two bounded vector fields depending on disjoint sets of variables, localized by orthogonal projections.* The next three lemmas quantify what is lost when the two fields share variables.

Lemma 4.7 (decoupled registers; [*proved here*])**.** *Let $z = (z', z'')$ be a splitting of the variables into two disjoint sets, and $g_j(z) = \langle A_j(z'), B_j(z'') \rangle$ with $\sum_j \mathbb{E}\|A_j\|^2 \le 1$ and $\sum_j \mathbb{E}\|B_j\|^2 \le 1$. Then $\sum_j \|g_j\|_2 \le 1$; consequently every one-block form $F = \sum_j y_j g_j$ (bounded or not) satisfies $\max_j \mathrm{Inf}_{y_j}(F) \ge \mathrm{Var}(F)^2$.*

Proof. With $\rho_j^A = \mathbb{E} A_j A_j^*$ and $\rho_j^B = \mathbb{E} B_j B_j^*$, independence gives $\|g_j\|_2^2 = \mathrm{tr}(\rho_j^A \rho_j^B) \le \mathrm{tr}\rho_j^A\, \mathrm{tr}\rho_j^B$, and Cauchy–Schwarz gives $\sum_j \|g_j\|_2 \le (\sum_j \mathbb{E}\|A_j\|^2)^{1/2} (\sum_j \mathbb{E}\|B_j\|^2)^{1/2} \le 1$. Then $\mathrm{Var}F = \sum_j \|g_j\|_2^2 \le \max_j \|g_j\|_2 \sum_j \|g_j\|_2 \le \max_j \|g_j\|_2$. □

Lemma 4.8 (shared coordinates; [*proved here*], tight)**.** *Let U, W, U' be disjoint sets of variables, $a = a(z_U, z_W)$ and $b = b(z_W, z_{U'})$ with $\|a\|, \|b\| \le 1$ pointwise, $P_1, P_2, \dots$ mutually orthogonal projections, and $g_j = \langle a, P_j b \rangle$. Then $\sum_j \|g_j\|_2 \le 2^{|W|/2}$, with equality for the address function ($a = b = e_{\iota(z_W)}$, $U = U' = \emptyset$, P_j the projection onto e_j).*

Proof. Fix z_W. As in the proof of Lemma 4.6, $\mathbb{E}[g_j^2 \mid z_W] \le \alpha_j(z_W)\beta_j(z_W)$ with $\alpha_j = \mathrm{tr}(P_j \rho_a(z_W))$, $\beta_j = \mathrm{tr}(P_j \rho_b(z_W))$, $\sum_j \alpha_j \le 1$, $\sum_j \beta_j \le 1$. Hence $\|g_j\|_2 \le \big(2^{-|W|} \sum_{z_W} \alpha_j \beta_j\big)^{1/2} \le 2^{-|W|/2} \sum_{z_W} (\alpha_j \beta_j)^{1/2}$, and summing over j (Cauchy–Schwarz for each z_W) gives $\sum_j \|g_j\|_2 \le 2^{-|W|/2} 2^{|W|}$. For the address function $g_j = \mathbf{1}[\iota(z_W) = j]$ and $\sum_j \|g_j\|_2 = 2^k 2^{-k/2}$. □

Thus the number of shared (coupling) coordinates is the exact exponent of the loss between the cb and the merely bounded setting.

Lemma 4.9 (tree-mediated coupling; [*proved here*])**.** *Let a, b ($\|a\|, \|b\| \le 1$), P_j and $g_j = \langle a, P_j b\rangle$ be as above, and let D be a decision tree on z of depth t such that for every leaf l, conditionally on l, a and b depend on disjoint sets of the free coordinates. Then $R_l \le 1$ for every leaf and $\sum_j \|g_j\|_2 \le \sum_l p_l^{1/2} \le 2^{t/2}$.*

Proof. Given l the free coordinates are uniform and independent, so the argument of Lemma 4.6 gives $R_l = \sum_j c_{jl} \le 1$. Then $\|g_j\|_2 = (\sum_l p_l c_{jl}^2)^{1/2} \le \sum_l p_l^{1/2} c_{jl}$; sum over j, and use $\sum_l p_l^{1/2} \le (\#\text{leaves})^{1/2} \le 2^{t/2}$. □

Corollary 4.10 (tree-mediated register forms; [*proved here*])**.** *Let $F = \sum_j y_j g_j$ be a bounded one-block form and D a z-tree of depth $t \ge 1$ such that at every leaf l the restricted family satisfies the hypothesis of Lemma 4.7 (with respect to the conditional measure and some splitting of the free coordinates depending on l). Then $\max(\varepsilon, \tau_D) \ge V^3/(16t^2)$; this gives Conjecture 1.1 for such forms whenever $t \le \mathrm{poly}(k/V)$. This applies in particular to $g_j = \langle a, P_j b\rangle$ as in Lemma 4.9 (take $A_j = P_j a$, $B_j = P_j b$), and these forms are automatically bounded.*

Proof. Lemma 4.7 at each leaf gives $R_l \le 1$; apply Corollary 4.3. Boundedness of $\langle a, P_j b\rangle$: $\sum_j |\langle P_j a, P_j b\rangle| \le \|a\|\|b\| \le 1$. □

For the address function with D the tree of address bits ($t = k$), this gives MaxInf $\ge 1/(16k^2)$ (the truth is $1/2$); the same holds for the "adaptive address" in which the leaves of a complete depth-k decision tree with distinct variables at all nodes select the output. For the latter with $k \ge 3$, no nonadaptive tree of depth $< 2^k - 1$ has $R_l \le 1$ at all leaves (as $R_l \ge 1/\max_j c_{jl}$ and $\sum_j c_{jl}^2 = 1$, such a tree must determine the leaf of the address tree), so the tree in Corollary 4.10 must in general be adaptive.

Remark 4.11 (independent blocks are essential; [*proved here*]). The address function Addr_k is computed by a deterministic $(k+1)$-query algorithm (query the k address bits, then the addressed bit), hence by a $(k+1)$-query quantum algorithm with output bias Addr_k. By [4, Theorem 1.3] there is a form T of degree $2k+2$ with $\|T\|_{\mathrm{cb}} \le 1$ and $T((x,1),\dots,(x,1)) = \mathrm{Addr}_k(x)$. Yet $\sum_i \mathrm{Inf}_i(\mathrm{Addr}_k)^{1/2} = k/\sqrt{2} + 2^{k/2}$. Hence no inequality of the form $\sum_i \mathrm{Inf}_i(p)^{1/2} \le \mathrm{poly}(\deg p)\,\|T\|_{\mathrm{cb}}$ holds for symmetrized (bias) polynomials p; this is consistent with Lemma 4.6, which concerns forms in independent blocks. In the bias representation, the address bits are read by the first and last blocks and the data bit by the middle blocks, which share the address variables: this is exactly the situation of Lemma 4.8.

4.5. **Dimension bounds and single-statistic selectors.**

Proposition 4.12 (dimension bound; [*proved here*])**.** *Let $\mathcal{W} \subseteq L^2(Q_n)$ be a subspace of dimension D with orthonormal basis $e_1, \dots, e_D$ and kernel diagonal $\Lambda_{\mathcal{W}} = \sum_r e_r^2$ (the reciprocal of the Christoffel function in the usual convention; here D is a dimension, not a decision tree). If $g_j \in \mathcal{W}$ for all j, then*

$$\sum_j \|g_j\|_2 \;\le\; \Big(\sup_z \sum_j |g_j(z)|\Big)\, \mathbb{E}\Lambda_{\mathcal{W}}^{1/2} \;\le\; \sqrt{D}\, \sup_z \sum_j |g_j(z)|.$$

Equality holds for the address function ($\mathcal{W}$ = functions of the k address bits, $\Lambda_{\mathcal{W}} \equiv 2^k$).

Proof. For $g \in \mathcal{W}$, $g = \sum_r \langle g, e_r\rangle e_r$, so $|g(z)| \le \|g\|_2 \Lambda_{\mathcal{W}}(z)^{1/2}$ by Cauchy–Schwarz. Hence $\|g\|_2^2 = \mathbb{E}|g|\,|g| \le \|g\|_2 \mathbb{E}[|g|\Lambda_{\mathcal{W}}^{1/2}]$, i.e. $\|g\|_2 \le \mathbb{E}[|g|\Lambda_{\mathcal{W}}^{1/2}]$. Summing over j, $\sum_j \|g_j\|_2 \le \mathbb{E}[(\sum_j |g_j|)\Lambda_{\mathcal{W}}^{1/2}]$, and $\mathbb{E}\Lambda_{\mathcal{W}}^{1/2} \le (\mathbb{E}\Lambda_{\mathcal{W}})^{1/2} = \sqrt{D}$. □

Applied at a leaf (with the conditional measure) this gives $R_l \le (\dim \operatorname{span}\{g_j|_l\})^{1/2}$ for bounded forms.

Corollary 4.13 (polynomial single-statistic selectors; [*proved here*])**.** *Let $s\colon Q_n \to \mathbb{R}$ have degree $e \ge 1$ and let $g_j = \varphi_j \circ s$, where the φ_j are real polynomials of degree at most $\lfloor k/e \rfloor$ and $\sum_j |g_j| \le 1$. Then $\sum_j \|g_j\|_2 \le \sqrt{k+1}$ and $\varepsilon \ge V^2/(k+1)$.*

Proof. All g_j lie in $\operatorname{span}\{s^r : 0 \le r \le \lfloor k/e \rfloor\}$ (powers computed pointwise), of dimension at most $k+1$. Apply Proposition 4.12 and Corollary 4.3 with $t = 0$. □

This covers "slab spikes" and Chebyshev bumps $g_j = \varphi_j(n^{-1/2}\sum_i z_i)$: each such g_j may be exponentially spiky ($\|g_j\|_2 \gg \|g_j\|_1$), but only the dimension matters. For $s = \sum_i z_i$ and symmetric g_j the polynomial hypothesis is automatic, since a symmetric multilinear polynomial of degree $\le k$ is a polynomial of degree $\le k$ in $\sum_i z_i$ [40]. *The polynomial hypothesis cannot be dropped:* $s(a) = \sum_{r=1}^{k} 2^r a_r$ is injective on Q_k, so the address selectors $g_b = \mathbf{1}[a = \beta(b)]$ are of the form $\varphi_b \circ s$ (with φ_b of degree $2^k - 1$) and have cube degree k, but $\sum_b \|g_b\|_2 = 2^{k/2}$ and $\varepsilon = 2^{-k}$.

4.6. **The hybrid statement.**

Theorem 4.14 (hybrid; [*proved here*])**.** *Let F be a bounded one-block form and D a z-tree of depth $t \ge 1$. Suppose there is a set G of leaves with $\sum_{l\in G} p_l V_l \ge V/2$ and $R_l \le R$ for $l \in G$. Then*

$$\max(\varepsilon, \tau_D) \;\ge\; \frac{V^3}{128\, R^2 t^2}.$$

Proof. As in Corollary 4.3, $\kappa_D \ge R^{-2}\sum_{l\in G} p_l V_l^3$, and by Hölder

$$\textstyle\sum_G p_l V_l \le (\sum_G p_l)^{2/3}(\sum_G p_l V_l^3)^{1/3} \le (\sum_G p_l V_l^3)^{1/3};$$

hence $\kappa_D \ge V^3/(8R^2)$. Theorem 4.2 gives $\varepsilon \ge V^2/(16R^2)$ or $\tau_D \ge V^3/(128R^2t^2)$. □

The leaf bounds $R_l \le R$ may come from decoupling (Lemma 4.7: $R_l \le 1$), from dimension (Proposition 4.12: $R_l \le \sqrt{\dim}$), or from sums of such pieces: if $g_j|_l = g_j' + g_j''$ then $R_l \le \sum_j \|g_j'\|_2 + \sum_j \|g_j''\|_2$ by the triangle inequality; here each piece must satisfy the hypothesis of the bound used for it on its own (for instance Proposition 4.12 requires control of $\sup \sum_j |g_j'|$, and pieces of a bounded family need not be bounded). Examples covered: the address and adaptive address functions ($t = k$, $R = 1$), slab spikes (Corollary 4.3 with $t = 0$, $R = \sqrt{k+1}$), polynomial single-statistic selectors applied after an address ($t = k$, $R = \sqrt{k+1}$), and mixtures of these.

4.7. **An exact decision-tree reformulation.** Say that one-block forms satisfy $\mathrm{AA}_1(c, C)$ if $\max(\varepsilon, \tau) \ge c(V/k)^C$ for every $k \ge 1$ and every bounded one-block form with $\deg g_j \le k$; by Lemma 2.4 and [43, Theorem 2.13] this is equivalent to Conjecture 1.1 up to polynomial losses.

Theorem 4.15 ([*proved here*])**.** *The following are equivalent, with polynomial dependence between the constants.*

(i) $\mathrm{AA}_1(c, C)$ *holds for some $c, C > 0$.*

(ii) There are c', C' such that every bounded one-block form with $\deg g_j \le k$ and variance V admits a z-decision tree D of depth $t \le (k/V)^{C'}/c'$ with

$$\kappa_D' := \mathbb{E}_l \max_j \operatorname{Inf}_{y_j}(F|_l)^2 \;\ge\; c'(V/k)^{C'}.$$

Proof. (ii)⇒(i). If $t = 0$ then $\kappa'_D = \varepsilon^2$. If $t \geq 1$, then $\kappa_D \geq \kappa'_D$, and Theorem 4.2 gives $\kappa'_D \leq \varepsilon V + 8t^2\tau \leq 9t^2 \max(\varepsilon, \tau)$, as $V \leq 1 \leq t$. Hence $\max(\varepsilon, \tau) \geq \kappa'_D/(9t^2)$, which is polynomial in V/k.

(i)⇒(ii). Assume $\mathrm{AA}_1(c, C)$ and let $w = c(V/(4k))^C$. Build D by the rule: at a node ρ (a partial assignment of z), if some unqueried z_i has $\mathrm{Inf}_{z_i}(F_\rho) \geq w$, query one such z_i; otherwise stop. Let $I_z(F_\rho) = \sum_i \mathrm{Inf}_{z_i}(F_\rho)$. For $i' \neq i$, $\mathrm{Inf}_{i'}(g) = \mathbb{E}_{z_i}\mathrm{Inf}_{i'}(g|_{z_i})$ (as $d_{i'}$ commutes with fixing z_i), and $\mathrm{Inf}_i(g|_{z_i}) = 0$; hence $\mathbb{E}_{z_i} I_z(F_{\rho,z_i}) = I_z(F_\rho) - \mathrm{Inf}_{z_i}(F_\rho) \leq I_z(F_\rho) - w$ whenever z_i is queried at ρ. Thus $I_z(F_{\rho_m}) + wm$ is a supermartingale along the random path, and for the (bounded) number N of queries, optional stopping gives $w\,\mathbb{E}N \leq I_z(F) = \sum_j I(g_j) \leq k \sum_j \mathrm{Var} g_j \leq kV$. Truncate the tree at depth $T = \lceil 4k/w \rceil$; by Markov's inequality $\Pr[N \geq T] \leq V/4$, and since $V_l \leq 1$ the truncated leaves carry $\sum p_l V_l \leq V/4$. Leaves with $V_l < V/4$ carry less than $V/4$. The remaining leaves l were stopped by the rule and have $V_l \geq V/4$; $F|_l$ is a bounded one-block form of z-degree at most k with all z-influences $< w$, so $\mathrm{AA}_1(c, C)$ gives $\max_j \mathrm{Inf}_{y_j}(F|_l) \geq c(V_l/k)^C \geq w$. These leaves carry $\sum p_l V_l \geq V/2$, hence $\sum p_l \geq V/2$, and $\kappa'_D \geq (V/2)w^2 = (V/2)c^2(V/(4k))^{2C}$, with depth $T = \lceil 4k/w \rceil$. □

Theorem 4.15 is the one-block analogue of the decision-tree form of the conjecture ([2, §3.1] for one direction, [42] for the other; see also [12, §1]), with two differences: only z-variables are queried, and the leaves are only required to have a heavy y-variable rather than small variance.

4.8. **The remaining gap and its leaf-wise variant.**

Question 4.16 (leaf root sums; [*open*])**.** Is it true that every bounded one-block form with $\deg g_j \leq k$ and variance V has a z-decision tree of depth $t \leq \mathrm{poly}(k/V)$ and a set G of leaves with $\sum_{l \in G} p_l V_l \geq V/2$ and $R_l \leq \mathrm{poly}(k/V)$ for $l \in G$?

By Theorem 4.14 and Lemma 2.4, a positive answer implies Conjecture 1.1; the answer is positive in every class treated above. Section 12 discusses how far Question 4.16 is from the conjecture (for instance, $R \leq e^k$ for every bounded family, by hypercontractivity as in Proposition 2.5).

$\mathrm{TR}_{\mathrm{leaf}}(t, R)$ is the statement: there are a z-tree of depth at most t and, at each leaf l, a kept index set S_l, with discarded mass $\sum_l p_l \sum_{j \notin S_l} c_{jl}^2 \leq V/4$ and kept root sums $\sum_{j \in S_l} c_{jl} \leq R$. The *heavy-y exit* at depth t with threshold h is the statement that there is a z-tree of depth at most t whose leaves carrying at least $(1 - \frac{1}{2k})V$ of the mass have some $c_{jl}^2 \geq h$. The *halving lemma with heavy-y exit* $\mathrm{HL}(t, h)$ is the statement: for every bounded one-block form with $\deg g_j \leq k$ and root sum $\sum_j \|g_j\|_2 \leq R$, $R \geq 2$, there is a z-tree of depth at most t such that leaves carrying at least a fraction $1 - \frac{1}{2k}$ of the mass either have some $c_{jl}^2 \geq h$ or have $R_l \leq R/2$.

Theorem 4.17 ([*proved here*])**.**
(a) $\mathrm{TR}_{\mathrm{leaf}}(t, R)$ *with* $t \geq 1$ *implies* $\max(\varepsilon, \tau) \geq 27V^3/(1024t^2R^2)$. (b) $\mathrm{AA}_1(c, C)$ *implies* $\mathrm{TR}_{\mathrm{leaf}}(16k/w, 4w^{-1/2})$ *with* $w = c(V/(8k))^C$. (c) $\mathrm{AA}_1(c, C)$ *implies the heavy-y exit at depth* $16k^2/w'$ *with* $h = Vw'/(128k^2)$, $w' = c(V/(8k^2))^C$. *Hence* AA_1, $\mathrm{TR}_{\mathrm{leaf}}$ *with polynomial parameters, and the halving lemma with heavy-y exit are equivalent up to polynomial losses; in particular these tree statements are not weaker than the conjecture.*

Proof. (a) Apply Lemma 5.9 with $q'_j = g_j^2 \mathbf{1}\{z \in l,\ j \in S_l\}$: $V' \geq 3V/4$ and $\bar{R}'_D \leq R$. (b), (c) *Peeling.* Fix $\theta \in (0, V/2]$, $\beta \in (0, 1)$, $w = c(\theta/k)^C$. Each node carries a remaining index set J' (initially all j) with remaining mass $V' = \sum_{j \in J'} c_{j,\mathrm{node}}^2$. If $V' \leq \theta$, stop (good leaf). Otherwise apply $\mathrm{AA}_1(c, C)$ to the restricted remaining family $G' = \sum_{j \in J'} y_j g_j|_{\mathrm{node}}$: either some $c_{j,\mathrm{node}}^2 \geq w$, and we *peel* j (move it to the kept set, no query), or some $\mathrm{Inf}_{z_i}(G') \geq w$, and we query z_i. Truncate (bad leaf) after $T = 2k/(w\beta)$ queries or $M = 2/(w\beta)$ peels. The

potential $\Phi = \sum_i \mathrm{Inf}_{z_i}(G')$ satisfies $\Phi \le kV'$ at every node and $\Phi \le kV$ at the root; it drops in expectation by at least w per query (for $i' \neq i$, $\mathrm{Inf}_{z_{i'}}(G') = \mathbb{E}_{z_i}\mathrm{Inf}_{z_{i'}}(G'|_{z_i})$) and does not increase under peeling, so by optional stopping $\mathbb{E}[\#\text{queries}] \le kV/w$; the remaining mass is a martingale under queries and drops by at least w per peel, so $\mathbb{E}[\#\text{peels}] \le V/w$. By Markov's inequality the bad leaves have probability, hence mass, at most $kV/(wT) + V/(wM) = \beta V$. A good leaf has remainder at most θ and at most M kept indices, so its kept root sum is at most $\sqrt{M}$. For (b) take $\theta = V/8$, $\beta = 1/8$: the discarded mass is at most $\theta + \beta V = V/4$, $\sqrt{M} = 4w^{-1/2}$, and the depth is $T = 16k/w$. For (c) take $\theta = V/(8k)$, $\beta = 1/(8k)$: a good leaf with $V_l \ge 2\theta$ has kept mass at least θ, hence some $c_{jl}^2 \ge \theta/M = Vw/(128k^2)$, and the remaining leaves carry at most $\beta V + 2\theta < V/(2k)$. *Converse.* Assume $\mathrm{HL}(t,h)$ and iterate it $N = \lceil k \log_2 e \rceil$ times, starting from $R \le e^k$ (Proposition 2.5) and recursing only into the halved leaves (restrictions of bounded forms are bounded forms); the tree has depth at most tN. (Here t and h may depend polynomially on the variance of the form. Recurse only into halved leaves with $V_l \ge V/(16N)$; the dropped leaves carry mass less than $V/(16N)$ per round, hence less than $V/16$ in total, and all forms to which HL is applied have variance at least $V/(16N)$, so t and h stay polynomial in k/V.) If the exited leaves carry probability at least $p_0 = V/8$, then $\kappa_D \ge p_0 h^2$, and Theorem 4.15 ((ii)⇒(i), $\kappa_D \le 9(tN)^2 \max(\varepsilon,\tau)$) gives $\max(\varepsilon,\tau) \ge p_0 h^2/(9t^2N^2)$. Otherwise, since each round loses at most a fraction $\frac{1}{2k}$ of the remaining mass and the exited leaves carry mass at most their probability, the halved leaves at the end keep mass at least $(1-\frac{1}{2k})^N V - p_0 \ge V/4 - V/8 - V/16$, and their root sums are at most $\max(2, e^k 2^{-N}) = 2$; Lemma 5.9 with these leaves gives $\max(\varepsilon,\tau) \ge (V/16)^3/(64t^2N^2)$. □

Any violator of these tree statements therefore contains a restricted subfamily violating the conjecture. Section 12 compares these statements with Question 4.16.

5. Subcube structures

By Proposition 2.5 and its proof, a counterexample must carry most of its influence on coordinates j whose derivative is *spiky*, $\|d_j f\|_2 \gg \|d_j f\|_1$, i.e. concentrated on a small set. Lemma 2.2 says that these spikes overlap with multiplicity at most $4d^2$ at every point. For the address function the spike sets are subcubes carved by the same (address) coordinates, which therefore have large influence. In this section we prove that this mechanism is forced when the spike sets are exact subcubes. A subcube is a set $C = \{x : x_K = a\}$ with $K \subseteq [n]$, $a \in \{-1,1\}^K$; we call K its set of fixed coordinates and $|K|$ its codimension, and $\mu(C) = 2^{-|K|}$.

5.1. The subcube cover lemma.

Lemma 5.1 (subcube cover lemma; [*proved here*])**.** *Let $C_j = \{x : x_{K_j} = a_j\}$, $1 \le j \le N$, be subcubes with $|K_j| \le r$ and multiplicity at most M, i.e. $\sum_j \mathbf{1}_{C_j}(x) \le M$ for all x. Let $\mu_j = \mu(C_j)$, $\Sigma = \sum_j \mu_j$ and $\mathrm{load}_i = \sum_{j:\, i \in K_j} \mu_j$. Then*

$$\max_i \mathrm{load}_i \;\ge\; \frac{1}{r}\Big(\frac{\Sigma}{M} - \frac{\sum_j \mu_j^2}{\Sigma}\Big) \;\ge\; \frac{1}{r}\Big(\frac{\Sigma}{M} - \max_j \mu_j\Big).$$

Proof. Helly property. If $C_{j_1}, \dots, C_{j_m}$ intersect pairwise, the assignments $a_{j_1}, \dots, a_{j_m}$ agree on every common fixed coordinate, so their union is a consistent partial assignment, and any completion of it lies in all C_{j_s}. Hence a pairwise intersecting subfamily has at most M members: the intersection graph H on $[N]$ ($j \sim l$ iff $j \neq l$ and $C_j \cap C_l \neq \emptyset$) has clique number at most M.

Motzkin–Straus. For a graph with clique number ω and a probability vector p, $\sum_{(j,l)} p_j p_l \le 1 - 1/\omega$, the sum over ordered pairs of adjacent vertices [39]. With $p = \mu/\Sigma$ the ordered intersecting

pairs carry $\sum_{j\sim l}\mu_j\mu_l \le (1-1/M)\Sigma^2$, so the ordered *disjoint* pairs ($j\neq l$, $C_j\cap C_l=\emptyset$) carry at least $\Sigma^2-\sum_j\mu_j^2-(1-1/M)\Sigma^2=\Sigma^2/M-\sum_j\mu_j^2$.

Attribution. Two disjoint subcubes disagree on some $i\in K_j\cap K_l$. Hence the disjoint pairs carry at most $\sum_i\sum_{j,l:\,i\in K_j\cap K_l}\mu_j\mu_l=\sum_i \mathrm{load}_i^2\le \max_i \mathrm{load}_i\sum_i \mathrm{load}_i$, and $\sum_i \mathrm{load}_i=\sum_j|K_j|\mu_j\le r\Sigma$. □

The Helly step is essential: the second-moment bound alone gives nothing, since $\Sigma=\mathbb{E}\sum_j \mathbf{1}_{C_j}\le M$ always. For a decision-tree partition ($M=1$) the root coordinate has load 1, while the lemma guarantees only about $1/r$; we do not know whether the factor $1/r$ is necessary.

Corollary 5.2 (subcube-supported derivatives; [*proved here*])**.** *Let $f\in\mathcal{B}_d(n)$, $d\ge 2$, and suppose there is $c>0$ such that for every j either $d_jf\equiv 0$ or $d_jf=c\,s_j\mathbf{1}_{C_j}$ with $s_j\colon Q_n\to\{-1,1\}$ arbitrary and C_j a subcube. Then* $\mathrm{MaxInf}(f)\ge c\,I(f)/(32d^4)\ge I(f)^2/(128d^6)$.

Proof. Let $C_j=\{x_{K_j}=a_j\}$. Since $\mathbf{1}_{C_j}=(d_jf)^2/c^2$ has degree at most $2d-2$ and a subcube indicator has degree equal to its codimension, $r:=\max_j|K_j|\le 2d$. By Lemma 2.2, $c\,\#\{j: x\in C_j\}=\sum_j|d_jf(x)|\le 4d^2$, so the multiplicity is at most $M:=4d^2/c$. Moreover $\mathrm{Inf}_j(f)=c^2\mu_j$, so $\Sigma=I(f)/c^2$, and $\Sigma\le M$ gives $c\ge I(f)/(4d^2)$. If $i\in K_j$, flipping x_i moves every point of C_j out of C_j, so $|d_jf(x)-d_jf(x^{\oplus i})|=c$ on a set of measure $2\mu_j$ and $\mathrm{Inf}_i(d_jf)\ge\frac14c^2\cdot 2\mu_j$. Since $\sum_j\mathrm{Inf}_i(d_jf)=\sum_{S\ni i}(|S|-1)\widehat{f}(S)^2\le(d-1)\mathrm{Inf}_i(f)$, we get $\mathrm{Inf}_i(f)\ge c^2\mathrm{load}_i/(2(d-1))$. If $\max_j\mu_j\ge\Sigma/(2M)$, then $\mathrm{MaxInf}(f)\ge c^2\max_j\mu_j\ge I/(2M)=cI/(8d^2)$. Otherwise Lemma 5.1 gives $\max_i\mathrm{load}_i\ge\Sigma/(4dM)$, and $\mathrm{MaxInf}(f)\ge c^2\Sigma/(8d^2M)=I/(8d^2M)=cI/(32d^4)$. In both cases $\mathrm{MaxInf}(f)\ge cI/(32d^4)\ge I^2/(128d^6)$. □

The class is narrow (the sensitive set of a coordinate of a Boolean function is in general a union of several subcubes, not a single subcube, and the derivatives of the address bits in Section 2.4 are not subcube-supported), but it isolates the mechanism. The soft analogue, in which the subcubes are replaced by the superlevel sets of spiky low-degree derivatives, is open; see Section 7.

5.2. **Root-weighted covers and subcube selectors.** The gluing lemma only needs the *average* leaf root sum $\bar{R}_D=\sum_l p_lR_l$.

Corollary 5.3 (average root sum; [*proved here*])**.** *In the setting of Section 4, for every decision tree D on the z-variables of depth $t\ge 1$,*

$$\max(\varepsilon,\tau_D)\;\ge\;\frac{V^3}{16\,t^2\,\bar{R}_D^2}.$$

Proof. As in Corollary 4.3, $\sum_j c_{jl}^4\ge V_l^3/R_l^2$. By Hölder with exponents 3 and 3/2, $V=\sum_l p_l(V_l^3/R_l^2)^{1/3}R_l^{2/3}\le\kappa_D^{1/3}\bar{R}_D^{2/3}$, so $\kappa_D\ge V^3/\bar{R}_D^2$; conclude with Theorem 4.2 as in Corollary 4.3. □

Lemma 5.4 (root-weighted subcube cover lemma; [*proved here*])**.** *In the setting of Lemma 5.1, let $\Phi=\sum_j\mu_j^{1/2}$ and $\mathrm{rl}_i=\sum_{j:\,i\in K_j}\mu_j^{1/2}$. Then $\max_i\mathrm{rl}_i\ge(\Phi^2/M-\Sigma)/(r\Phi)$.*

Proof. Repeat the proof of Lemma 5.1 with $p_j=\mu_j^{1/2}/\Phi$: the ordered disjoint pairs carry $\sum\mu_j^{1/2}\mu_l^{1/2}\ge\Phi^2/M-\sum_j\mu_j$; they carry at most $\sum_i\mathrm{rl}_i^2\le\max_i\mathrm{rl}_i\sum_i\mathrm{rl}_i$; and $\sum_i\mathrm{rl}_i=\sum_j|K_j|\mu_j^{1/2}\le r\Phi$. □

Lemma 5.5 (subcube potential; [*proved here*])**.** *Let C_j be subcubes of codimension at most r and multiplicity at most M. For a node v of a decision tree let μ_v be the uniform measure on the subcube of inputs reaching v and $\Phi(v) = \sum_j \mu_v(C_j)^{1/2}$. Build a tree greedily: at v, if $\Phi(v) \geq 2M$ query a coordinate maximizing $\mathrm{rl}_i(v) = \sum_{j:\, i \in K_j(v)} \mu_v(C_j)^{1/2}$, where for $\mu_v(C_j) > 0$ we write $K_j(v)$ for the set of coordinates of K_j not queried on the path to v (the fixed coordinates of the restricted subcube), and otherwise stop. Let Φ_s be the expectation of Φ over the nodes at depth s (or their stopped ancestors). Then the tree never queries a coordinate twice and $\Phi_s - 2M \leq (1 - \frac{1}{5rM})^s(\Phi_0 - 2M)$.*

Proof. Restricting to the subcube of v maps each C_j either to the empty set or to a subcube of codimension at most r (fixing the unqueried coordinates of K_j), with multiplicity at most M; in particular $\sum_j \mu_v(C_j) \leq M$. Querying i: a piece with i among its fixed coordinates has relative measure $2\mu_v(C_j)$ on one side and 0 on the other, so its expected term drops from $\mu_v^{1/2}$ to $2^{-1/2}\mu_v^{1/2}$, and other pieces are unchanged; thus $\mathbb{E}\Phi(\text{children}) = \Phi(v) - (1 - 2^{-1/2})\mathrm{rl}_i(v)$. If $\Phi(v) \geq 2M$, Lemma 5.4 gives $\max_i \mathrm{rl}_i(v) \geq (\Phi^2/M - M)/(r\Phi) \geq \frac{3}{4}\Phi/(rM)$, and since $(1 - 2^{-1/2})\frac{3}{4} \geq \frac{1}{5}$, $\mathbb{E}\Phi(\text{children}) \leq (1 - \frac{1}{5rM})\Phi(v)$. Writing $\alpha = 1/(5rM)$, at a queried node $\mathbb{E}\Phi(\text{children}) - 2M \leq (1-\alpha)(\Phi(v) - 2M)$, and at a stopped node $\Phi(v) - 2M < 0$, so $\Phi(v) - 2M \leq (1 - \alpha)(\Phi(v) - 2M)$; averaging gives the recursion. A queried coordinate has $\mathrm{rl} = 0$ afterwards, so it is never queried again; and a node with $\Phi(v) \geq 2M$ has a piece with a nonempty set of unqueried fixed coordinates (otherwise all pieces contain the whole subcube and $\Phi(v) \leq M$), so the maximizing query is a genuine coordinate. □

Theorem 5.6 (subcube selectors; [*proved here*])**.** *Let $F = \sum_j y_j g_j$ be a bounded one-block form with $\deg g_j \leq k$, $k \geq 1$, where $g_j = c\, s_j \mathbf{1}_{C_j}$ with a common $c > 0$, arbitrary $s_j \colon Q_n \to \{-1, 1\}$ and subcubes C_j. Then some decision tree D on the z-variables of depth $t \leq 11k^2/V$ has $\bar{R}_D \leq 3$, and*

$$\mathrm{MaxInf}(F) \;\geq\; \max(\varepsilon, \tau) \;\geq\; \frac{V^5}{17424\, k^4}.$$

Proof. Fixing the coordinates outside K_j turns g_j into $\pm c$ times a point mass on Q_{K_j}, which has degree $|K_j|$; since restriction does not raise degree, $r := \max_j |K_j| \leq k$. The multiplicity is at most $\sum_j |g_j|/c \leq 1/c$, and being an integer it is at most $M := \lfloor 1/c \rfloor$; as $c \leq 1$, $\frac{1}{2} \leq cM \leq 1$. Also $V = c^2 \sum_j \mu(C_j) \leq c^2 M \leq c$, so $M \leq 1/V$. At the root, $\Phi = \sum_j \|g_j\|_2/c \leq e^k \sum_j \|g_j\|_1/c \leq e^k/c \leq 2e^k M$, by the $(2,1)$-hypercontractive estimate $\|g\|_2 \leq e^k \|g\|_1$ for $\deg g \leq k$ [41, Theorem 9.22]. Run Lemma 5.5 for $t = \lceil 5rM(k+1) \rceil$ steps: $\Phi_t < 2M + e^{-(k+1)} 2e^k M < 3M$. Since $R_l = c\,\Phi(l)$ at each leaf, $\bar{R}_D = c\Phi_t < 3cM \leq 3$, and $t \leq 5k(k+1)/V + 1 \leq 11k^2/V$. Corollary 5.3 gives $\max(\varepsilon, \tau_D) \geq V^3/(144t^2) \geq V^5/(144 \cdot 121k^4)$. If the tree is trivial ($t = 0$ effectively, i.e. $\Phi_0 < 2M$), then $R = \sum_j \|g_j\|_2 < 2cM \leq 2$ and $\varepsilon \geq V^2/R^2 \geq V^2/4$ by Corollary 4.3. □

The tree of Lemma 5.5 is a *halving* procedure: it reduces the expected root sum geometrically until it is $O(1)$; with depth $O(rM \log(R/V))$ one can check that the same procedure gives a second proof of Corollary 5.2 with a $\log^2$ loss (we do not use or prove this here). For the gradient form (Lemma 2.4) Theorem 5.6 applies when the nonzero derivatives of f have a common constant magnitude on subcubes, which is the setting of Corollary 5.2.

5.3. **Banded derivatives and the classes they cover.** We need three facts about decision-tree complexity of Boolean functions $h \colon Q_n \to \{0, 1\}$. For $S \subseteq [n]$, x^S is x with the coordinates in S flipped; $\mathrm{bs}(h, x)$ is the maximal number of pairwise disjoint blocks S_i with $h(x^{S_i}) \neq h(x)$, $\mathrm{bs}(h) = \max_x \mathrm{bs}(h, x)$, and $s(h)$ is the same with $|S_i| = 1$; $C(h, x)$ is the minimal size of a

certificate for x (a set S such that $y_S = x_S$ implies $h(y) = h(x)$), $C^{(1)}(h) = \max_{h(x)=1} C(h,x)$; and $D(h)$ is the minimal depth of a decision tree computing h.

Lemma 5.7 ([*known*] [40, 9]; proofs included)**.** (a) $C(h,x) \le s(h)\,\mathrm{bs}(h,x)$. (b) $D(h) \le C^{(1)}(h)\,\mathrm{bs}(h) \le \mathrm{bs}(h)^3$.

Proof. (a) Take a maximal family of disjoint minimal sensitive blocks $S_1, \dots, S_b$ for x. By minimality x^{S_i} is sensitive to every $l \in S_i$, so $|S_i| \le s(h)$; the union U is a certificate, since otherwise the difference set of a violating y would contain a minimal sensitive block disjoint from U. (b) While the current restriction h' is non-constant, pick x with $h'(x) = 1$, query a minimal certificate of x for h', and output 1 if the answers agree with it. If rounds $1, \dots, s$ all fail and h' is still non-constant, take y with $h'(y) = 0$ consistent with all answers; round i yields the nonempty set B_i of certificate coordinates on which y disagrees with the round-i certificate; the B_i are disjoint and sensitive for y, and non-constancy gives one more disjoint sensitive block among the unqueried coordinates, so $s + 1 \le \mathrm{bs}(h)$. Finally $C^{(1)} \le C \le s \cdot \mathrm{bs} \le \mathrm{bs}^2$ by (a). □

Lemma 5.8 (gapped supports; [*proved here*])**.** *If* $\deg g \le k$ *and* $|g| \in \{0\} \cup [c, Ac]$ *on* Q_n*, then* $E = \{g \ne 0\}$ *satisfies* $\mathrm{bs}(\mathbf{1}_E) \le 4k^2(1 + A^2)$*, hence* $D(\mathbf{1}_E) \le 64k^6(1 + A^2)^3$.

Proof. Fix x and disjoint sensitive blocks $S_1, \dots, S_b$. For $y \in \{0,1\}^b$ let $x(y)$ flip the blocks with $y_i = 1$; then $Q(y) = g(x(y))^2/c^2$ is a polynomial of degree at most $2k$ in y, with values in $\{0\} \cup [1, A^2]$. Its symmetrization $P(s) = \mathbb{E}_{|y|=s} Q(y)$ is a univariate polynomial of degree at most $2k$ [40] with $0 \le P(i) \le A^2$ for $i = 0, \dots, b$, and $|P(1) - P(0)| \ge 1$ (if $x \notin E$ then $P(0) = 0$ and $P(1) \ge 1$; if $x \in E$ then $P(0) \ge 1$ and $P(1) = 0$). By the mean value theorem $|P'(\xi)| \ge 1$ for some $\xi \in [0,1]$, and [40, Theorem 3.3] (if $b_1 \le P(i) \le b_2$ for all integers $0 \le i \le N$ and $|P'(\xi)| \ge \gamma$ for some $\xi \in [0, N]$, then $\deg P \ge \sqrt{\gamma N/(\gamma + b_2 - b_1)}$) gives $2k \ge \sqrt{b/(1 + A^2)}$. The last claim follows from Lemma 5.7(b). □

The sub-mass version of the gluing corollary lets us ignore small selector values.

Lemma 5.9 (sub-masses; [*proved here*])**.** *In the setting of Section 4, let* $0 \le q'_j \le g_j^2$ *pointwise,* $V' = \sum_j \mathbb{E} q'_j > 0$*, and* $\bar{R}'_D = \sum_l p_l \sum_j \mathbb{E}[q'_j \mid l]^{1/2}$*. Then for every tree* D *of depth* $t \ge 1$*,* $\max(\varepsilon, \tau_D) \ge V'^3/(16 t^2 \bar{R}'^2_D)$.

Proof. With $m'_{jl} = \mathbb{E}[q'_j \mid l] \le c_{jl}^2$, $\kappa_D \ge \sum_l p_l \sum_j m'^2_{jl}$. Per leaf, Hölder with exponents $\frac{3}{2}, 3$ gives $\sum_j m'_{jl} \le (\sum_j m'^{1/2}_{jl})^{2/3} (\sum_j m'^2_{jl})^{1/3}$, and Hölder on the leaf measure as in Corollary 5.3 gives $\kappa_D \ge V'^3/\bar{R}'^2_D$. If $\varepsilon V \ge \kappa_D/2$ then $\varepsilon \ge \kappa_D/2 \ge V'^3/(2\bar{R}'^2_D)$; otherwise $8t^2\tau_D \ge \kappa_D/2$ by Theorem 4.2. □

Theorem 5.10 (banded selectors; [*proved here*])**.** *Let* $F = \sum_j y_j g_j$ *be a bounded one-block form with* $\deg g_j \le k$ *and* $|g_j(z)| \in \{0\} \cup [c, Ac]$ *for all* j, z*. Let* $E_j = \{g_j \ne 0\}$ *and* $r'' = \max(1, \max_j D(\mathbf{1}_{E_j}))$*. Then* $r'' \le 64k^6(1 + A^2)^3$ *and* $\max(\varepsilon, \tau) \ge V^5/(1082\, A^6 r''^4)$*. For* $A = 1$ *(constant magnitude, arbitrary supports)* $r'' \le 512k^6$.

Proof. The bound on r'' is Lemma 5.8. Split each E_j into the 1-leaves $C_{j,a}$ of an optimal tree for $\mathbf{1}_{E_j}$: disjoint subcubes of codimension at most r''. The pieces through a point z have distinct j and $|g_j(z)| \ge c$, so the multiplicity is at most $M := \lfloor 1/c \rfloor$; also $V \le A^2 c^2 \sum_j \mu(E_j) \le A^2 c^2 M \le A^2 c$, so $M \le A^2/V$. Every piece has measure at least $2^{-r''}$, so $\Phi_0 = \sum_{j,a} \mu(C_{j,a})^{1/2} \le 2^{r''/2} M$. If $\Phi_0 < 2M$, then $R \le Ac\Phi_0 < 2A$ and $\varepsilon \ge V^2/(4A^2)$ by Corollary 4.3. Otherwise run Lemma 5.5 for $t = \lceil 1.74 r''^2 M \rceil \le 2.74 r''^2 A^2/V$ levels: $\Phi_t - 2M \le e^{-1.74 r''/5} 2^{r''/2} M \le M$, since $\frac{5}{2} \ln 2 < 1.74$. At a leaf l, $\mathbb{E}[g_j^2 \mid l]^{1/2} \le \sum_a \mathbb{E}[g_j^2 \mathbf{1}_{C_{j,a}} \mid l]^{1/2} \le Ac \sum_a \mu_l(C_{j,a})^{1/2}$, so $\bar{R}_D \le$

$Ac\Phi_t \le 3AcM \le 3A$. Corollary 5.3 gives $\max(\varepsilon,\tau) \ge V^3/(144A^2t^2) \ge V^5/(144\cdot 2.74^2 A^6 r''^4)$, and $144\cdot 2.74^2 < 1082$. □

Theorem 5.11 (banded decision-tree families; [*proved here*])**.** *Let $F=\sum_j y_j g_j$ be a bounded one-block form (no degree assumption), $L=\lceil \log_2(2/V)\rceil$, $B_{j,l}=\{2^{-l-1}<|g_j|\le 2^{-l}\}$ for $0\le l\le L$, and $r=\max(1,\max_{j,l\le L} D(\mathbf{1}_{B_{j,l}}))$. Then*

$$\max(\varepsilon,\tau) \;\ge\; \frac{V^5}{1.32\cdot 10^6\, r^4 \log_2(16/V)^2}.$$

Proof. Split each $B_{j,l}$ into the 1-leaves $C_{j,l,a}$ of an optimal tree (subcubes of codimension at most r). The band-l pieces through a point have distinct j and $|g_j|>2^{-l-1}$, so fewer than $M_l := 2^{l+1}$ of them; this persists under restriction, and since alive pieces have conditional measure at least 2^{-r}, the band potential $\Phi_l(v)=\sum_{j,a}\mu_v(C_{j,l,a})^{1/2}$ satisfies $\Phi_l(v)\le 2^{r/2}M_l$ at every node. Run phases $l=0,\dots,L$: starting from every leaf of the previous phase, run Lemma 5.5 on band l for $t_l=\lceil 1.74r^2M_l\rceil$ levels; at the end $\mathbb{E}\Phi_l\le 3M_l$ as in the proof of Theorem 5.10. By concavity of the square root a query never increases the expectation of any $\Phi_{l'}$, so $\mathbb{E}\Phi_{l'}\le 3M_{l'}$ survives the later phases. Let $q'_j=g_j^2\mathbf{1}\{|g_j|>2^{-L-1}\}$. At a final leaf $\sum_j \mathbb{E}[q'_j \mid l]^{1/2} \le \sum_{l'\le L} 2^{-l'}\Phi_{l'}(l)$, so $\bar R'_D \le \sum_{l'\le L} 2^{-l'}\cdot 3\cdot 2^{l'+1} = 6(L+1)$. The discarded mass is $\sum_j \mathbb{E}g_j^2\mathbf{1}\{|g_j|\le 2^{-L-1}\} \le 2^{-L-1}\sum_j\mathbb{E}|g_j| \le V/4$, so $V'\ge 3V/4$. The depth is $t\le \sum_l(1.74r^2\,2^{l+1}+1)\le 3.48r^2\cdot 8/V+\log_2(8/V)\le 31r^2/V$, using $2^L<4/V$ and $\log_2(8/V)\le 3/V$. Lemma 5.9 gives $\max(\varepsilon,\tau)\ge (3V/4)^3V^2/(16\cdot 31^2 r^4\cdot 36(L+1)^2)$, and $16\cdot 961\cdot 36\cdot 64/27 < 1.32\cdot 10^6$, $L+1\le\log_2(16/V)$. (If the tree is empty, then $\sum_j m_j^2\le \varepsilon V$ and Hölder give $\varepsilon\ge V'^3/\bar R'^2_D$, a stronger bound.) □

Corollary 5.12 (classes of functions; [*proved here*])**.** *Let $f\in\mathcal{B}_d$, $v=\mathrm{Var}(f)/(16d^4)$, and $g_i=d_if/(4d^2)$ (so $V=I(f)/(16d^4)\ge v$ and $\mathrm{MaxInf}(f)\ge 16d^3\max(\varepsilon,\tau)$ for the gradient form, by Lemma 2.4). Then:*

1. *(low-level weight)* $\mathrm{MaxInf}(f)\ge (W^{[1,k+1]}(f)/(4d^2 3^k))^2$ *for all $k\ge 0$ (Corollary 3.3);*
2. *(min form)* $\mathrm{MaxInf}(f)\ge \min\{e^{-10(j-1)},(W^{[1,j]}(f)/(4d^2))^3\}$ *for all $j\ge 1$;*
3. *(noise-stable level sets) if $\rho=1-1/K$ and $\int_{-1}^{1}[\mathrm{Stab}_\rho(h_\theta)-(\mathbb{E}h_\theta)^2]\,d\theta\ge\eta$, then* $\mathrm{MaxInf}(f)\ge (\eta/(36d^2))^6\,3^{-6K\ln(1/\eta)}$*;*
4. *(unit-step derivatives) if $|d_if(z)|\in\{0,c_0\}$ for all i,z, then* $\mathrm{MaxInf}(f)\ge 16d^3v^5/(1082\,r'^4)$ *with $r'=\max(1,\max_i D(\mathbf{1}_{\{d_if\neq 0\}}))\le 512d^6$; this includes Boolean f and lattice-valued edge-Lipschitz f;*
5. *(banded derivatives) if every nonzero $|d_if(z)|$ lies in $[c_0,Ac_0]$, then* $\mathrm{MaxInf}(f)\ge 16d^3v^5/(1082A^6r''^4)$ *with $r''=\max(1,\max_i D(\mathbf{1}_{\{d_if\neq0\}}))\le 64d^6(1+A^2)^3$; e.g. $f=\frac{1}{m}\sum_b G_b$ with Boolean G_b and each variable in at most A of the G_b;*
6. *(bounded interaction degree) if each d_if depends on at most r coordinates, then* $\mathrm{MaxInf}(f)\ge 16d^3v^5/(1.32\cdot 10^6 r^4\log_2(16/v)^2)$*; e.g. $f=\sum_b c_bG_b(x_{S_b})$ with $|S_b|\le s$ and each variable in at most A sets S_b, where $r\le As$;*
7. *(tree-banded derivatives) the same bound if every band set of the g_i at levels $l\le\lceil\log_2(2/v)\rceil$ has decision-tree depth at most r;*
8. *(one-statistic derivatives) if $d_if=\varphi_i(s(x))$ for one statistic s and real polynomials φ_i with $\deg\varphi_i\deg s\le d-1$, then* $\mathrm{MaxInf}(f)\ge 16d^3v^2$*;*
9. *(mixtures of (6)/(7) with (8)) if $[n]=I_T\sqcup I_S$ (disjoint union), where $\{g_i\}_{i\in I_T}$ satisfies the hypothesis of (7) and $\{g_i\}_{i\in I_S}$ spans a space of dimension at most D_0, then (with $r\ge 1$)* $\mathrm{MaxInf}(f)\ge 16d^3(3v/4)^3v^2/(16\cdot 961\,r^4(6(L+1)+\sqrt{D_0})^2)$ *and $L=\lceil\log_2(2/v)\rceil$.*

Proof. All one-block bounds used are increasing in V, so V may be replaced by v (in (7) and (9) the band levels then refer to v, a stronger hypothesis). (2): if $\tau := \mathrm{MaxInf}(f) < e^{-10(j-1)}$ then $3^{j-1} < \tau^{-(\ln 3)/10} \le \tau^{-1/6}$, and Corollary 3.3 gives $W^{[1,j]}(f) \le 4d^2\tau^{1/2-1/6}$. (3): with $m = \lceil K\ln(1/\eta)\rceil$, $\sum_{S\neq\emptyset}\rho^{|S|}\widehat{h_\theta}(S)^2 \le W^{[1,m]}(h_\theta) + \rho^{m+1}\mathrm{Var}h_\theta$, $\rho^{m+1} \le \eta$ and $\int \mathrm{Var}h_\theta \le \frac12$, so $\int W^{[1,m]}(h_\theta) \ge \eta/2$; Theorem 3.6 with $k = m-1$ and $16d^2 3^{m-1} + 2 \le 18d^2 3^{m-1}$ gives the claim. (4), (5): Theorem 5.10 with $k = d-1$ and $c = c_0/(4d^2)$. (6), (7): every band set of g_i is a function of at most r coordinates, resp. has depth at most r; apply Theorem 5.11. (8): the g_i lie in the span of $1, s, \dots, s^{d-1}$, of dimension at most d, so $\varepsilon \ge V^2/d$ by Proposition 4.12 and Corollary 4.3 at $t = 0$, and $\mathrm{MaxInf}(f) = 16d^4\varepsilon$. (9): run the phases of Theorem 5.11 on I_T (the band multiplicities are unchanged since $\sum_{i\in I_T}|g_i| \le 1$). For $i \in I_S$ take $q_i' = g_i^2$ (no truncation). The I_S-root sum $\sum_{i\in I_S}\mathbb{E}[g_i^2 \mid l]^{1/2}$ is at most $\sqrt{D_0}$ at the root (Proposition 4.12), and it is a supermartingale along the tree by Jensen's inequality (concavity of the square root), so its expectation over the leaves is at most $\sqrt{D_0}$; root sums add, so $\bar R_D' \le 6(L+1) + \sqrt{D_0}$, and Lemma 5.9 (with the truncation applied to I_T only, discarded mass at most $v/4$) concludes. If the tree is empty, $\sum_i m_i^2 \le \varepsilon V$ and Hölder give $\varepsilon \ge V'^3/\bar R_D'^2$, which is stronger. □

6. Symmetric and structured classes

6.1. Symmetric and block-symmetric functions.

Remark 6.1 (symmetric functions; [*known*]). If $f \in \mathcal{B}_d(n)$ is invariant under all permutations of coordinates, then $\mathrm{MaxInf}(f) \ge \mathrm{Var}(f)^2/(4d^4)$. Indeed $f(x) = q(K)$ where K is the number of -1's and $\deg q \le d$ [40], and $\mathrm{MaxInf}(f) = I(f)/n \ge \mathrm{Var}(f)/n$. If $n \le 2d^2$ we are done since $\mathrm{Var}f \le 1$. If $n > 2d^2$, Lemma 2.1 (with $c = \max_{[0,n]}|q'|$, $b_2 - b_1 \le 2$) gives $\max_{[0,n]}|q'| \le 2d^2/(n-d^2) < 4d^2/n$, hence $\mathrm{Var}f \le \mathbb{E}(q(K) - q(n/2))^2 \le (4d^2/n)^2\, n/4$, i.e. $1/n \ge \mathrm{Var}(f)/(4d^4)$.

Using Bernstein's inequality in the bulk, where the measure concentrates, and Markov's inequality only on an exponentially small tail, one gains d^2.

Proposition 6.2 (symmetric functions, sharp exponents; [*proved here*])**.** *If $f \in \mathcal{B}_d(n)$ is invariant under all permutations of coordinates, then* $\mathrm{MaxInf}(f) \ge \mathrm{Var}(f)^2/(16d^2)$.

Proof. Write $f(x) = P(s)$ with $s = n^{-1}\sum_i x_i$, $\deg P \le d$ and $|P| \le 1$ on the grid $G_n = \{-1, -1+2/n, \dots, 1\}$ [40]. All influences are equal, so $\mathrm{MaxInf}(f) = I(f)/n \ge \mathrm{Var}(f)/n$, and the claim is immediate if $n \le 16d^2$. Let $n > 16d^2$ and $M = \max_{[-1,1]}|P|$. Every $t \in [-1,1]$ is within $1/n$ of G_n, so Markov's inequality $\max|P'| \le d^2 M$ gives $M \le 1 + d^2M/n$, i.e. $M \le n/(n-d^2) \le 16/15$. Flipping x_1 changes s by $2/n$, so $d_1 f(x) = n^{-1}P'(\xi)$ for some ξ with $|\xi - s| \le 2/n$, and

$$\mathrm{Var}f \;\le\; I(f) \;=\; n\,\mathbb{E}(d_1 f)^2 \;\le\; n^{-1}\,\mathbb{E}\max_{|t-s|\le 2/n} P'(t)^2.$$

Let $a = \frac12 - \frac{2}{n} \ge \frac38$. If $|s| \le a$ the maximum is over $|t| \le \frac12$, where Bernstein's inequality $|P'(t)| \le dM(1-t^2)^{-1/2}$ gives $P'(t)^2 \le \frac43 d^2M^2$; otherwise we use $P'(t)^2 \le d^4M^2$, on an event of probability at most $2e^{-na^2/2} \le 2e^{-9d^2/8}$ by Hoeffding's inequality. Hence $\mathrm{Var}f \le n^{-1}d^2M^2(\frac43 + 2d^2e^{-9d^2/8}) \le 2.26\,d^2/n$, since $2d^2e^{-9d^2/8} \le 2e^{-9/8} < 0.65$ for $d \ge 1$. Therefore $\mathrm{MaxInf}(f) \ge \mathrm{Var}(f)/n \ge \mathrm{Var}(f)^2/(2.26\,d^2)$. □

Example 6.3 (the Chebyshev family; [*proved here*], constants checked by exact binomial sums)**.** Let d be odd, $n = Kd^2$, $s = n^{-1}\sum_i x_i$ and $f = T_d(s) \in \mathcal{B}_d(n)$ (bounded since $|s| \le 1$), so $\mathbb{E}f = 0$. Since $T_d^2 = (1 + T_{2d})/2$ and $T_{2d}(s) = -\cos(2d\arcsin s)$, $\mathrm{Var}f = \frac12(1 - \mathbb{E}\cos(2d\arcsin s))$; by independence $\mathbb{E}\cos(2ds) = \cos(2/(Kd))^{Kd^2} \le e^{-2/K}$, and $|\arcsin s - s| \le (\frac{\pi}{2} - 1)|s|^3$ with $\mathbb{E}|s|^3 \le (3/n^2)^{3/4}$, so $\mathrm{Var}f \ge \frac12(1 - e^{-2/K}) - 1.31K^{-3/2}d^{-2}$ for $d \ge 3$. As in the proof of

Proposition 6.2, $d_1 f = n^{-1} T_d'(\xi)$ with $|\xi - s| \le 2/n$, Bernstein's inequality on $|s| \le \frac{1}{2} - \frac{2}{n}$ and Markov's inequality elsewhere give $\mathrm{MaxInf} f \le (\frac{4}{3} + 2d^2 e^{-Kd^2a^2/2})/(K^2 d^2)$ with $a = \frac{1}{2} - \frac{2}{n}$. This yields the bounds quoted in Section 1.2: for $K \ge 32$, $0.943/K \le \mathrm{Var} f \le 2.47/K$ and $\mathrm{MaxInf} f \le 1.61 \mathrm{Var}(f)^2/d^2$; for $K = 1$ and $d \ge 9$, $\mathrm{Var} f \ge 0.416$ and $\mathrm{MaxInf} f \le 1.36/d^2$. Thus, using also $f = n^{-1}\sum_i x_i$ ($d = 1$, $\mathrm{MaxInf} f = \mathrm{Var}(f)^2$), the best constant c in $\mathrm{MaxInf} \ge c\mathrm{Var}^2/d^2$ for symmetric functions lies in $[1/16, 1]$.

Theorem 6.4 (block-symmetric functions; [*proved here*])**.** *Let $f \in \mathcal{B}_d(n)$ be invariant under all permutations inside each block of a partition of $[n]$ into b blocks. Then $\sum_i \mathrm{Inf}_i(f)^{1/2} \le 2bd^2$, and hence $\mathrm{MaxInf}(f) \ge I(f)^2/(4b^2d^4) \ge \mathrm{Var}(f)^2/(4b^2d^4)$.*

Proof. Fix a block s of size N; all $i \in s$ have the same influence, so the block contributes $N\mathrm{Inf}_i(f)^{1/2}$, which is at most $N < 2d^2$ if $N < 2d^2$. Let $N \ge 2d^2$ and fix the coordinates x' outside s. Then $f = p(\sum_{i\in s} x_i)$ with $\deg p \le d$ and $|p| \le 1$ on $\{-N, -N+2, \dots, N\}$ [40]. As in the proof of Proposition 6.2, $P(t) = p(Nt)$ is bounded by 1 on a grid of mesh $2/N$ in $[-1,1]$, so $\max_{[-1,1]} |P| \le N/(N - d^2) \le 2$ and, by Markov's inequality, $|p'| \le 2d^2/N$ on $[-N, N]$. Hence $|d_i f(x)| = |p(m+1) - p(m-1)|/2 \le 2d^2/N$ pointwise, with $m = \sum_{j\in s\setminus\{i\}} x_j$, and the block contributes at most $2d^2$. Finally $I(f) \le \mathrm{MaxInf}(f)^{1/2} \sum_i \mathrm{Inf}_i(f)^{1/2}$ and $I(f) \ge \mathrm{Var} f$. □

By Theorem 4.5, functions that become block-symmetric with $b \le \mathrm{poly}(d)$ blocks after a decision tree of depth $\mathrm{poly}(d)$ satisfy Conjecture 1.1 with polynomial loss.

6.2. **Wreath-symmetric functions.** Let $n = Mw$, with the coordinates split into M blocks of size w, and let the wreath product $S_w \wr S_M$ act by permuting coordinates inside blocks and permuting blocks. Tribes-type functions are of this kind; exact linear programs found nearly Boolean bounded polynomials of this symmetry type of moderate degree for small w, and the following theorem shows that this class nevertheless satisfies the conjecture. Since $S_w \wr S_M$ is transitive on the n coordinates, [19, Proposition 3.18] already gives $\mathrm{MaxInf} \ge \mathrm{Var}^2/(d^4 e^{2d})$; the point of Theorems B.1 and 6.5 is a bound polynomial in d and w, and then in d alone, which we have not found in print.

Theorem 6.5 (wreath-symmetric functions, uniform in w; [*proved here*])**.** *If $f \in \mathcal{B}_d(Mw)$ is invariant under $S_w \wr S_M$, then $\mathrm{MaxInf}(f) \ge \mathrm{Var}(f)^2/(2416\, d^6)$ for all w, M and d. If $d \ge 2$, $w \ge 4d^2$ and $M \ge 4\pi d^3$, then $\mathrm{MaxInf}(f) \ge \mathrm{Var}(f)^2/(728\, d^4)$.*

Theorem 6.6 (sharp law for composed wreath functions; [*proved here*])**.** *Let F be a real polynomial on $\mathbb{R}^r$ of total degree at most a with $|F| \le 1$ on $[-1,1]^r$, let $\ell_1, \dots, \ell_r$ be real polynomials of degree at most b with $|\ell_i| \le 1$ on $[-1,1]$, let $y_1, \dots, y_M$ be the block means of $x \in \{-1,1\}^{Mw}$, and let $f(x) = F\big(\frac{1}{M}\sum_l \ell_1(y_l), \dots, \frac{1}{M}\sum_l \ell_r(y_l)\big)$ (that is, $f = R(\nu(y))$ in the notation of Appendix B). Then f is invariant under $S_w \wr S_M$, $|f| \le 1$ and $\deg f \le D = ab$. If $D \ge 2$ and $w \ge 4D^2$, then for every M*

$$I(f) \le 27D^2/n, \qquad \text{hence} \qquad \mathrm{Var}(f) \le 27D^2/n \quad \text{and} \quad \mathrm{MaxInf}(f) \ge \mathrm{Var}(f)^2/(27D^2).$$

Remark 6.7 (optimality; [*proved here*]). For odd D the symmetric function $f = T_D(n^{-1}\sum_i x_i)$ is of the form of Theorem 6.6 ($r = 1$, $F = T_D$, $\ell_1(t) = t$, $a = D$, $b = 1$). Since $T_D(m) = \pm Dm + O(m^3)$ and $\mathbb{E}m^2 = 1/n$ for $m = n^{-1}\sum_i x_i$, $n\mathrm{Var} f \to D^2$ and $\mathrm{Var}(f)^2/\mathrm{MaxInf}(f) \to D^2$ as $n \to \infty$. So the constant in $\mathrm{MaxInf} \ge \mathrm{Var}^2/(CD^2)$ cannot be below $1 - o(1)$. [*numerical*]: $n\mathrm{Var} f/D^2 = 0.9920$ for $D = 3$, $n = 10^3$, and 0.9992 for $n = 10^4$ (exact computation).

Theorem 6.8 (sharp law for many huge blocks; [*proved here*])**.** *If $f \in \mathcal{B}_d(Mw)$ is invariant under $S_w \wr S_M$, $d \ge 2$, $M \ge 4\pi d^3$ and $w \ge 8d^6 \ln(2d)$, then $I(f) \le 170\, d^2/n$, hence $\mathrm{MaxInf}(f) \ge \mathrm{Var}(f)^2/(170\, d^2)$.*

The proofs of Theorems 6.5, 6.6 and 6.8, and the results they use (among them Theorems B.1 and B.8), are in Appendix B.

6.3. **Polynomials of smooth linear statistics.** Let $W \in \mathbb{R}^{r\times n}$ have rank r and columns w_j, let $Z = W([-1,1]^n) = \sum_j [-w_j, w_j]$ (a symmetric convex body) with gauge norm $\|\cdot\|_Z$, $\delta_j = \|w_j\|_Z \le 1$, $\delta = \max_j \delta_j$, and $f(x) = P(Wx)$ with P a real polynomial of degree at most d and $|P| \le 1$ on $W(Q_n)$, so that $f \in \mathcal{B}_d(n)$. Let $M = \sup_Z |P|$ and $D_v P = \langle v, \nabla P\rangle$.

Theorem 6.9 (smooth linear statistics, sharpened; [*proved here*])**.** *If $d^2 r\delta \le 1/4$, then* $\mathrm{MaxInf}(f) \ge I(f)^2/(324\, d^2 r) \ge \mathrm{Var}(f)^2/(324\, d^2 r)$, *and* $\mathrm{MaxInf}(f) \ge I(f)^2/(138\, d^2 r)$ *if $d \ge 4$.*

The proof of Theorem 6.9, which uses Theorem C.4 and Lemmas C.2, C.3 and C.5, is in Appendix C.

Proposition 6.10 (one linear form with general weights; [*proved here*])**.** *Let $f(x) = P(\sum_i w_i x_i) \in \mathcal{B}_d(n)$ with $\sum_i |w_i| = 1$, $\delta = \max_i |w_i|$ and $d^2\delta \le 1/4$. Then* $\mathrm{MaxInf}(f) \ge I(f)^2/(6.7\, d^2)$.

Proof. Here $Z = [-1,1]$, and every point of $[-1,1]$ is within δ of a value of $\sum_i w_i x_i$, so Markov's inequality gives $M \le 1/(1-d^2\delta) \le \frac{4}{3}$. With $S^{(j)} = \sum_{i\ne j} w_i x_i$, the mean value theorem gives $d_j f = w_j P'(\xi)$ with $|\xi - S^{(j)}| \le |w_j|$. If $|S^{(j)}| \le \frac{1}{2}$ then $|\xi| \le \frac{3}{4}$ and Bernstein's inequality gives $|P'(\xi)| \le \frac{4}{\sqrt{7}} dM$; always $|P'| \le d^2 M$. By Hoeffding's inequality and $\sum_i w_i^2 \le \delta \le 1/(4d^2)$, $\Pr[|S^{(j)}| > \frac{1}{2}] \le 2e^{-d^2/2}$. So $\|d_j f\|_2 \le |w_j| M d\sqrt{16/7 + 2d^2 e^{-d^2/2}} \le |w_j| M d\sqrt{16/7 + 4/e}$, and $\sum_j \|d_j f\|_2 \le \frac{4}{3}\sqrt{16/7+4/e}\, d \le 2.585\, d$, with $2.585^2 \le 6.7$. □

Let $\varrho = \sup_{t\in[-1,1]} \min_x |t - \sum_i w_i x_i|$ be the covering radius (as in Theorem C.4(b)). The same argument under the weaker hypotheses $\varrho \le 1/(2d^2)$, $\max_i |w_i| \le \frac{1}{4}$ and $\sum_i w_i^2 \le 1/(8\ln(d^2+1))$ gives the constant 17.2 (we omit the computation). Proposition 6.10 extends Proposition 6.2 (equal weights) to general weights at the scale $d^2\delta = O(1)$ of Theorem C.4. A dynamic program over the exact extremal problem ($d \le 8$, $n \le 576$) leaves slack at least 2.8 in the constant 6.7 [*numerical*].

6.4. **Functions of one linear form.** Let $f = P(\langle a, x\rangle)$ with $\deg P \le d$, $|f| \le 1$ on Q_n, and (after flipping signs and dropping zeros) $a_1 \ge \cdots \ge a_n > 0$; let $A = \sum_i a_i$, $A_i = \sum_{j>i} a_j$, and $S(a) = \{\langle a, x\rangle : x \in Q_n\} \subseteq [-A, A]$.

Lemma 6.11 ([*standard*])**.** (a) *If $S \subseteq [\alpha,\beta]$ contains α, β, has consecutive gaps at most $h \le L/(2d^2)$ with $L = \beta - \alpha$, and $|P| \le 1$ on S, then $\|P\|_{[\alpha,\beta]} \le 2$ and $\|P'\|_{[\alpha,\beta]} \le 4d^2/L$.* (b) *The maximal gap of $S(a)$ is at most $2\max(a_n, \max_i (a_i - A_i)^+)$.*

Proof. (a) Let $M = \|P\|_{[\alpha,\beta]} = |P(\xi)|$; Markov gives $\|P'\| \le 2d^2 M/L$, and some $s \in S$ has $|s - \xi| \le h/2$, so $M \le 1 + M/2$. (b) $S(a) = 2\Sigma(a) - A$ with $\Sigma(a)$ the set of subset sums. Adding the weights in increasing order $b_1 \le b_2 \le \ldots$, the subset sums $T_{k+1} = T_k \cup (T_k + b_{k+1})$ acquire no new gap if $b_{k+1} \le b_1 + \cdots + b_k$, and otherwise a single new gap $b_{k+1} - (b_1 + \cdots + b_k) = a_i - A_i$ with $i = n - k$; and $T_1 = \{0, b_1\}$. □

Call *a regular* if $a_n \le A/(2d^2)$ and $a_i - A_i \le A/(2d^2)$ for all i; then the gaps of $S(a)$ are at most $A/d^2 = L/(2d^2)$ with $L = 2A$.

Theorem 6.12 (one linear form; [*proved here*])**.** (a) *If a is regular, then $|d_i f| \le 2d^2 a_i/A$ pointwise, $\sum_i \mathrm{Inf}_i(f)^{1/2} \le 2d^2$, and* $\mathrm{MaxInf}(f) \ge I(f)^2/(4d^4)$. (b) *If a is not regular, let $m \ge 1$ be the least index such that the tail $(a_{m+1}, \ldots, a_n)$ is regular with respect to its own sum and d (the empty tail counts as regular). Then* $\mathrm{MaxInf}(f) \ge I(f)^3/(2^{11} d^5 (2d^2+m)^2 m^2)$.

Proof. (a) $d_i f(x) = (P(s+a_i) - P(s-a_i))/2$ with both arguments in $S(a) \subseteq [-A, A]$; the mean value theorem and Lemma 6.11(a) with $L = 2A$ give $|d_i f| \le a_i \cdot 4d^2/(2A)$. Sum and use $I \le \mathrm{MaxInf}^{1/2} \sum_i \mathrm{Inf}_i^{1/2}$. (b) Query $x_1, \dots, x_m$. At every leaf l, $f|_l(x') = P(c_l + \langle a', x' \rangle)$ is a regular one-form function, so by (a) $\rho_l := \sum_i \mathbb{E}[(d_i f)^2 \mid l]^{1/2} \le 2d^2 + m$ (the m queried coordinates contribute at most 1 each). Apply Lemma 5.9 to the gradient form with $q_i' = g_i^2$: then $V = I/(16d^4)$, $\bar{R}_D' \le (2d^2+m)/(4d^2)$, and $\mathrm{MaxInf}(f) \ge 16d^3 \max(\varepsilon, \tau) \ge 16d^3 \, \frac{I^3/(16d^4)^3}{16m^2(2d^2+m)^2/(16d^4)} = \frac{I^3}{256\, d^5 m^2 (2d^2+m)^2}$, which is stronger than the stated bound. □

The lacunary case $a_i = 3^{-i}$, in which every tail is irregular, is open.

Separator theorems for disjointly supported and r-sparse selectors (Theorems F.1 and F.3, Corollary F.4) are in Appendix F.

6.5. **Graph properties.** Let $M \ge 2$ and $N = \binom{M}{2}$, and index the coordinates of Q_N by the edges of K_M. A function on Q_N is a *graph property* if it is invariant under the action of S_M on edges. Since S_M is transitive on edges, [19, Proposition 3.18] gives $N\mathrm{Var} f \le d^4 e^{2d}$, so the fixed-degree form of the conjecture for graph properties is known (with $81e^6 \approx 3.3 \cdot 10^4$, or 6561 by the remark in Section 1.2, at $d = 3$); what is new below is a bound polynomial in $\mathrm{Var} f$ and d (Theorem 6.16), the limit in Theorem 6.14(ii), the constant of Theorem 6.15, and the explicit coefficient bounds of Lemmas D.1 and D.2, which do not depend on M. That the $t_\pm(H, \cdot)$ span the invariant polynomials is folklore, and the linear independence of homomorphism densities is classical [18, 34]. (Spanning: an S_M-invariant f has equal coefficients on edge sets in one S_M-orbit, and these orbits are the isomorphism classes of graphs without isolated vertices; we include a short proof of the independence.) Let $\mathcal{G}_d$ be the set of isomorphism classes of graphs H without isolated vertices and with $1 \le e(H) \le d$, let $\mathrm{aut}(H)$ be the order of the automorphism group, and $(m)_j = m(m-1)\cdots(m-j+1)$. For H with $v(H) \le M$ let $c_M(H) = (M)_{v(H)}/\mathrm{aut}(H)$ be the number of copies of H contained in K_M and $t_\pm(H, x) = c_M(H)^{-1} \sum_{H'} \prod_{e \in E(H')} x_e$, the sum over these copies. A graph property of degree at most d is $f = b_\emptyset + \sum_H b_H t_\pm(H, x)$, the sum over $H \in \mathcal{G}_d$ with $v(H) \le M$, and

$$\mathrm{Var} f = \sum_H \frac{b_H^2}{c_M(H)}, \qquad \frac{N}{c_M(H)} = \frac{\mathrm{aut}(H)}{2(M-2)_{v(H)-2}}.$$

Put $d^* = d$ for odd d and $d^* = d - 1$ for even d. For symmetric measurable $U \colon [0,1]^2 \to [-1,1]$ let $t(H, U) = \int \prod_{uv \in E(H)} U(z_u, z_v)\, dz$ be the homomorphism density, and $t(\emptyset, U) = 1$.

Lemma 6.13 (graphon transfer; [*proved here*])**.** *If $f \in \mathcal{B}_d(N)$ is a graph property then $|b_\emptyset + \sum_H b_H t(H, U)| \le 1$ for every symmetric measurable $U \colon [0,1]^2 \to [-1,1]$. Consequently $|b_{K_2}| \le d^*$, and $|b_H| \le C(d)$ for all $H \in \mathcal{G}_d$ and all M, where $C(d) < \infty$ depends only on d and $C(d) \ge 2^{d-1}$.*

Proof. Draw $z_1, \dots, z_M$ independently and uniformly from $[0,1]$ and then, independently, $x_{uv} \in \{\pm 1\}$ with $\mathbb{E}[x_{uv} \mid z] = U(z_u, z_v)$. A copy of H has distinct vertices, so $\mathbb{E} \prod_{e \in E(H')} x_e = t(H, U)$ and $\mathbb{E} f(x) = b_\emptyset + \sum_H b_H t(H, U) \in [-1, 1]$. For $U \equiv s \in [-1, 1]$ this is a polynomial $p(s)$ of degree at most d, bounded by 1 on $[-1, 1]$, whose coefficient of s is b_{K_2}; its odd part has degree at most d^* and is bounded by 1, so Bernstein's inequality at $s = 0$ gives $|b_{K_2}| \le d^*$.

The functions $t(H, \cdot)$, $H \in \mathcal{G}_d \cup \{\emptyset\}$, are linearly independent. Suppose $\sum_H c_H t(H, U) = 0$ for all U and $c_F \ne 0$ for some $F \in \mathcal{G}_d$ (if only $c_\emptyset \ne 0$ there is nothing to prove); choose such F with $V = v(F)$ maximal. For a step kernel with parts of measures $\alpha_1, \dots, \alpha_V > 0$ and values $\beta_{ij} = \beta_{ji} \in [-1, 1]$, $t(H, U) = \sum_{\phi \colon V(H) \to [V]} \prod_v \alpha_{\phi(v)} \prod_{uv \in E(H)} \beta_{\phi(u)\phi(v)}$. Multiplying the term of H by $(\sum_i \alpha_i)^{V - v(H)}$ makes the identity homogeneous of degree V in α; it holds on an open

set, hence as a polynomial identity. Place F on $[V]$. A term contributing to the monomial $\alpha_1 \cdots \alpha_V \prod_{ij \in E(F)} \beta_{ij}$ comes from an injective ϕ mapping $E(H)$ onto $E(F)$; as F has no isolated vertices, ϕ is onto $[V]$ and is an isomorphism $H \to F$. So the coefficient of this monomial is $c_F \operatorname{aut}(F) \neq 0$, a contradiction. Hence $b \mapsto \sup_U |\sum_H b_H t(H,U)|$ is a norm on $\mathbb{R}^{\mathcal{G}_d \cup \{\emptyset\}}$; let $C(d)$ be the maximum of $|b_H|$, $H \in \mathcal{G}_d$, over its (compact) unit ball. Extending b by 0 for $v(H) > M$ shows that the bound holds for every M. Finally $T_d(t(K_2,U)) = \sum_j c_j t(jK_2,U)$, with jK_2 the matching with j edges, lies in the unit ball and has top coefficient 2^{d-1}. □

Theorem 6.14 (graph properties; [*proved here*])**.** *Let $f \in \mathcal{B}_d(N)$ be a graph property on M vertices.* (i) $\operatorname{MaxInf}(f) \geq \operatorname{Var}(f)^2/K(d)$ *with* $K(d) = d^{*2} + \frac{1}{2}C(d)^2 \sum_{H \in \mathcal{G}_d,\, v(H) \geq 3} \operatorname{aut}(H)$*; a bound of this form with $K(d) = e^{O(d)}$ holds for every $f \in \mathcal{B}_d$ by the general bound quoted in the abstract, and with $K(d) = d^4 e^{2d}$ for graph properties by* [19, Proposition 3.18]*, so the content of (i) is the structure of $K(d)$, used in (ii) and in Theorem 6.15.* (ii) $\lim_{M\to\infty} \max_f N\operatorname{Var} f = d^{*2}$*, the maximum over graph properties $f \in \mathcal{B}_d(N)$.* (iii) $N\operatorname{Var} f \leq \max(2d^2, M_0^2/2)$ *with* $M_0 = \max(2d+3,\, 2 + 3C(d)^2|\mathcal{G}_d|/d^2)$.

Proof. S_M acts transitively on edges, so all $\operatorname{Inf}_e(f)$ are equal and $\operatorname{MaxInf}(f) = N^{-1}\sum_e \operatorname{Inf}_e(f) \geq \operatorname{Var}(f)/N = \operatorname{Var}(f)^2/(N\operatorname{Var} f)$. By the display above, $N\operatorname{Var} f = \sum_H b_H^2 \operatorname{aut}(H)/(2(M-2)_{v(H)-2})$, with $(M-2)_{v(H)-2} \geq 1$ since $v(H) \leq M$. The term of K_2 is $b_{K_2}^2 \leq d^{*2}$, and the others are at most $C(d)^2 \operatorname{aut}(H)/2$ by Lemma 6.13; this is (i). (ii) The terms with $v(H) \geq 3$ are $O_d(1/M)$, so the lim sup is at most d^{*2}. Conversely, $f = T_{d^*}(N^{-1}\sum_e x_e)$ is a graph property in $\mathcal{B}_d(N)$; its level-one coefficients are $T'_{d^*}(0)/N + O_d(N^{-2})$ with $|T'_{d^*}(0)| = d^*$, so $N\operatorname{Var} f \geq N^2(d^*/N - O_d(N^{-2}))^2 \to d^{*2}$. (iii) If $M \leq M_0$ then $N\operatorname{Var} f \leq N \leq M_0^2/2$. If $M > M_0$ and $3 \leq v(H) \leq 2d$, then $\operatorname{aut}(H) \leq 6(M-3)_{v(H)-3}$: for $v(H) = 3$ because $\operatorname{aut}(H) \leq 3!$, and for $v = v(H) \geq 4$ because $(M-3)_{v-3} \geq (v)_{v-3} = v!/6$ as $M - 3 \geq 2d \geq v$. So each term with $v(H) \geq 3$ is at most $3C(d)^2/(M-2)$, their sum is at most $3C(d)^2|\mathcal{G}_d|/(M-2) \leq d^2$, and $N\operatorname{Var} f \leq d^{*2} + d^2 \leq 2d^2$. □

Theorem 6.15 (degree at most 3; [*proved here, computer-assisted*])**.** *Every graph property $f \in \mathcal{B}_3(N)$ satisfies $N\operatorname{Var} f \leq 97$, hence $\operatorname{MaxInf}(f) \geq \operatorname{Var}(f)^2/97$.*

The bounds above are polynomial in $\operatorname{Var} f$ but not in d. The next one is polynomial in both; it loses one factor $\operatorname{Var} f$ against Theorem 6.14(i).

Theorem 6.16 (polynomial bound for graph properties; [*proved here*])**.** *Let $d \geq 1$ and let $f \in \mathcal{B}_d(N)$ be a graph property on $M \geq 2$ vertices. Then* $\operatorname{Var} f \leq 6.06 \cdot 10^4\, d^8/M$ *and*

$$\operatorname{MaxInf}(f) \;\geq\; \frac{\operatorname{Var}(f)^3}{1.9 \cdot 10^9\, d^{16}}.$$

In the proof (Appendix D.2), x' differs from x only on the star of a vertex u. Replace u by K clones with its star in x and K' clones with its star in x', and delete $K + K' - 1$ random other vertices. The expected value of f is then a polynomial of degree at most $2d$ in (K, K'), bounded by 1 at the lattice points, with values $f(x)$ at $(1,0)$ and $f(x')$ at $(0,1)$ (Lemma D.5). A two-variable Markov inequality (Lemma D.6) gives $|f(x) - f(x')| \leq 256e\, d^4/M$ (Theorem D.7), and a Fourier identity gives the variance bound (Lemma D.8).

Attribution. Cloning and subsampling, so that an expectation becomes a polynomial in a size parameter that is bounded at the integer points, followed by a Markov-type inequality for such polynomials, is the technique of Zhandry's small-range lemma [57, Theorem 7.3], [58, Lemma 3.2]. It is stated there for quantum query algorithms, but the proof uses only polynomiality and boundedness. Chailloux [14] used it for permutation-symmetric functions, and Ben-David, Childs, Gilyén, Kretschmer, Podder and Wang [10] for graph properties, in both cases to

bound randomized query complexity polynomially in quantum query complexity. What is new here is its use for influences of bounded polynomials, with two types of clones and a two-variable Markov inequality. The only influence bound for bounded graph properties that we found in print is $N\mathrm{Var} f \le d^4 e^{2d}$ [19, Proposition 3.18].

Remark 6.17 (improvements and extensions; [*unaudited*]). (a) In step (4) of the proof of Lemma D.6, Bernstein's inequality $|p'(\theta)| \le D \sup_{[0,1]} |p|/\sqrt{\theta(1-\theta)}$ in place of Markov's inequality in θ, together with $\int_0^1 (\theta(1-\theta))^{-1/2} d\theta = \pi$, gives $|P(1,0) - P(0,1)| \le 8\pi e D^3/M$. Hence $|f(x) - f(x')| \le 64\pi e\, d^3/M$, $\mathrm{Var} f \le 3.74 \cdot 10^4 d^6/M$ and $\mathrm{MaxInf}(f) \ge \mathrm{Var}(f)^3/(7 \cdot 10^8 d^{12})$. (b) In outline, the same proof covers k-uniform hypergraph properties ($\deg \Psi \le kd$; with (a), $\mathrm{MaxInf} \ge k! \,\mathrm{Var}^{k+1}/(C_k^k d^{6k})$, $C_k = 16\pi^2 e^2 k^6$), and directed and bipartite graph properties ($\mathrm{MaxInf} \ge c\mathrm{Var}^3/d^{12}$). (c) Zhandry's lemma [58, Lemma 3.2], applied to $\alpha \mapsto f(x^\alpha)$ for random small-range maps α of the vertex set, combined with Theorem 4.1(L1) and [19, Theorem 3.3], gives in a few lines $\mathrm{Var} f \le 13.3\, d^{8/3} N^{-1/3}$, hence $\mathrm{MaxInf}(f) \ge \mathrm{Var}(f)^4/(2400\, d^8)$. This is a better power of d and a worse power of $\mathrm{Var} f$ than Theorem 6.16; we have only a sketch of this argument. (d) [*numerical*] A linear program over all polynomials of total degree at most D bounded by 1 on the lattice points gives $M \max |P(1,0) - P(0,1)| \approx 2.4D^2$ to $2.5D^2$ for $D \le 6$ and $M \le 200$. A proof of this order would give $\mathrm{Var} f = O(d^4/M)$ and $\mathrm{MaxInf}(f) \ge c\mathrm{Var}(f)^3/d^8$. Pointwise bounds of the form of Theorem D.7 cannot give $N\mathrm{Var} f \le \mathrm{poly}(d)$, since a single star change moves $T_d(N^{-1}\sum_e x_e)$ by about $4d^2/M$.

The proofs of Theorems 6.15 and 6.16, the explicit constants behind Theorem 6.14 and two barriers are in Appendix D.

6.6. **Cyclically invariant functions.** A function on Q_n is *cyclically invariant* if it is invariant under $x \mapsto (x_2, \dots, x_n, x_1)$; the conjecture for this class is item (3) of [3, Problem 1.3]. Cyclic shifts act transitively, so [19, Proposition 3.18] gives $n\mathrm{Var} f \le d^4 e^{2d}$ in every degree (874 for $d = 2$, $3.3 \cdot 10^4$ for $d = 3$; 32 and 6561 by the remark in Section 1.2). Theorems 6.18 and 6.19 improve these constants; Theorem 6.20, which is polynomial in d for the levels $k \le 3$ in every degree, and Theorem D.11, which is polynomial in the locality, go beyond it. A bound polynomial in d for all levels is open. All influences of such an f are equal, so $\mathrm{MaxInf}(f) \ge \mathrm{Var}(f)/n = \mathrm{Var}(f)^2/(n\mathrm{Var} f)$, and it suffices to bound $n\mathrm{Var} f$. Let $W_j = \sum_{|S|=j} \hat{f}(S)^2$.

Theorem 6.18 (degree at most 2; [*proved here*])**.** *If $f \in \mathcal{B}_2(n)$ is cyclically invariant then $n\mathrm{Var} f \le 9$, hence $\mathrm{MaxInf}(f) \ge \mathrm{Var}(f)^2/9$.*

Theorem 6.19 (degree at most 3; [*proved here*])**.** *If $f \in \mathcal{B}_3(n)$ is cyclically invariant then $n\mathrm{Var} f \le 9 + 8 + 1944 = 1961$, hence $\mathrm{MaxInf}(f) \ge \mathrm{Var}(f)^2/1961$.*

Theorem 6.20 (fixed levels in every degree; [*proved here*])**.** *Let $f \in \mathcal{B}_d(n)$ be cyclically invariant and $W_k = \sum_{|S|=k} \widehat{f}(S)^2$. Then $nW_1 \le d^2$, $nW_2 \le 2d^4$ and $nW_3 \le 54d^6$ for every d; more generally $nW_k \le (k^k 2^{k-1} \tau_{d,k})^2/k!$, where $\tau_{d,k}$ is the largest modulus of the coefficient of ρ^k in a polynomial of degree at most d bounded by 1 on $[-1,1]$ (for $k = 3$ this gives the stated bound, and for $k = 2$ a sharper single-phase argument is used).*

The proofs of Theorems 6.18, 6.19 and 6.20, two barriers for the method (Propositions D.9 and D.10) and a bound for local cyclic functions (Theorem D.11) are in Appendix D.

6.7. **Torus-bounded polynomials.** Polynomials bounded on the torus $\mathbb{T}^n$ are treated in Appendix E: block-symmetric functions satisfy a degree-free bound (Theorem E.2), and for this class the conjecture reduces to the spike inequality (Conjecture E.7), which holds in each fixed degree (Theorem E.8).

7. Reformulations equivalent to the conjecture

Throughout this section $F = \sum_j y_j g_j$ is a bounded one-block form of degree k, and $\mathrm{AA}_1(c, C)$ is the statement of Section 4 that every such form satisfies $\max(\varepsilon, \tau) \geq c(V/k)^C$; by Lemma 2.4 and [43], $\mathrm{AA}_1(c, C)$ for some c, C is equivalent to Conjecture 1.1 up to polynomial losses.

7.1. Sign amplification and nearly Boolean functions.

Lemma 7.1 (bounded sign approximation; [*proved here*])**.** *Let $0 < w \leq 1$, $0 < \delta \leq \frac{1}{2}$, $s_0 = w/2$, $\mu = 2s_0^2/(1-s_0^2)$, $N = \lceil \frac{5}{3w} \ln \frac{40}{w\delta} \rceil$, $K(s) = T_N(1 + \mu - (2+\mu)s^2)$, $D = \int_0^1 K$ and $Q(t) = \int_0^t K(s)\, ds/(D+2)$. Then Q is an odd polynomial of degree $2N+1 \leq \frac{10}{3w} \ln \frac{40}{w\delta} + 3$, $|Q| \leq 1$ on $[-1,1]$, and $|Q(t) - \operatorname{sgn} t| \leq \delta$ for $w \leq |t| \leq 1$.*

Proof. $u(s) = 1 + \mu - (2+\mu)s^2$ decreases in $|s|$ with $u(s_0) = 1$ and $u(1) = -1$, so $|K| \leq 1$ on $[s_0, 1]$ and $K \geq 1$ on $[0, s_0]$. Since $u(s_0/2) = 1 + \frac{3}{4}\mu$ with $\frac{3}{4}\mu \in [\frac{3}{8}w^2, \frac{1}{2}]$ and $\operatorname{arccosh}(1+x) \geq \sqrt{x}$ on $[0,1]$, $K \geq T_N(1 + \frac{3}{4}\mu) \geq \frac{1}{2}e^N\sqrt{3\mu/4} \geq \frac{1}{2}e^{0.61Nw}$ on $[0, s_0/2]$, so $D \geq \frac{w}{8}e^{0.61Nw} - 1 \geq 4/\delta$ by the choice of N. For $t \in [0, s_0]$, $0 \leq \int_0^t K \leq D + 1$; for $t \in [s_0, 1]$, $\int_0^t K \in [D-2, D+2]$. Hence $|Q| \leq 1$, and for $t \in [w, 1]$, $Q(t) \in [\frac{D-2}{D+2}, 1]$, so $|1 - Q(t)| \leq 4/D \leq \delta$. K is even, so Q is odd. □

Sharper sign approximations (degree $O(\log(1/\delta)/w)$) are known from the quantum signal processing literature; Lemma 7.1 suffices here.

Theorem 7.2 (sign amplification; [*proved here*])**.** *Let $f \in \mathcal{B}_d$, $v = \mathrm{Var} f \in (0,1]$, $0 < w \leq v^{3/2}/2000$ and $0 < \delta \leq v^{1/2}/40$. There are θ and a polynomial P with $|P| \leq 1$ on $[-1,1]$ and $\deg P = m \leq \bar{m}(w,\delta) := \frac{20}{3w} \ln \frac{80}{w\delta} + 3$ such that, with $\sigma = \operatorname{sgn}(f - \theta)$ and $a = P \circ f$:* (i) *$a \in \mathcal{B}_{md}$ and $\Pr[|f - \theta| < w] \leq 3w/v^{1/2}$;* (ii) *$|a - \sigma| \leq \delta$ off the window $\{|f - \theta| < w\}$;* (iii) *$\mathrm{Var} a \geq 0.133v$, $\mathbb{E}(1 - a^2) \leq 2\delta + 3w/v^{1/2}$ and $\mathbb{E}(a - \sigma)^2 \leq \delta^2 + 12w/v^{1/2}$;* (iv) *$|d_i a| \leq m^2 |d_i f|$, so $\mathrm{Inf}_i(a) \leq m^4 \mathrm{Inf}_i(f)$.*

Proof. Let $q_1 = \sup\{t : \Pr[f < t] \leq v/16\}$ and $q_2 = \inf\{t : \Pr[f > t] \leq v/16\}$. Then $\Pr[f \in [q_1, q_2]] \geq 1 - v/8$, and since $|f - c| \leq 2$, $v \leq (q_2 - q_1)^2/4 + 4(v/8)$, so $\ell = q_2 - q_1 \geq (2v)^{1/2}$. Among $\lfloor \ell/(2w) \rfloor \geq \ell/(4w)$ disjoint windows $(\theta - w, \theta + w)$ in (q_1, q_2) one has mass at most $4w/\ell \leq 3w/v^{1/2}$; take its centre θ. Then $\Pr[f \geq \theta] \in [v/16, 1 - v/16]$ and $\mathrm{Var}\sigma \geq v/5$. Let $P(t) = Q((t - \theta)/2)$ with Q from Lemma 7.1 at width $w/2$ and accuracy δ; this gives $|P| \leq 1$, (ii) and the degree bound. Off the window $1 - a^2 \leq 2\delta$ and $(a - \sigma)^2 \leq \delta^2$, and on the window these are at most 1 and 4; so $\mathrm{Var}(a)^{1/2} \geq \mathrm{Var}(\sigma)^{1/2} - \|a - \sigma\|_2 \geq (0.447 - 0.082)v^{1/2}$. (iv) is Markov's inequality $\sup_{[-1,1]} |P'| \leq m^2$ with the mean value theorem. □

For a function $\beta(K)$ let NB_β be the statement: every $a \in \mathcal{B}_K$ with $\mathbb{E}(1 - a^2) \leq \beta(K)\mathrm{Var} a$ has $\mathrm{MaxInf}(a) \geq c(\mathrm{Var} a/K)^C$.

Theorem 7.3 ([*proved here*])**.** *For fixed $\gamma \in [0,1)$, $\mathrm{NB}_{K^{-\gamma}}$ implies Conjecture 1.1 with polynomial loss; conversely the conjecture implies NB_β for every β.*

Proof. The converse is trivial. Apply Theorem 7.2 with $\delta = w$, where $w = \mathrm{poly}(v/d)$ is chosen so small that $w \leq v^{3/2}/2000$ and $45\, w^{1-\gamma}(20d \ln(9/w))^\gamma \leq v^{3/2}$ (possible since $\gamma < 1$). Then $\bar{m} \leq (20/w)\ln(9/w)$, $a \in \mathcal{B}_K$ with $K = \bar{m}d$, and $\mathbb{E}(1 - a^2)/\mathrm{Var} a \leq (2w + 3w/v^{1/2})/(0.133v) \leq 45w/v^{3/2} \leq K^{-\gamma}$. So $\mathrm{NB}_{K^{-\gamma}}$ gives $\mathrm{MaxInf}(a) \geq c(0.133v/K)^C$, and (iv) gives $\mathrm{MaxInf}(f) \geq \bar{m}^{-4} c(0.133v/(\bar{m}d))^C$. □

Theorem 7.4 (the product case is complete; [*proved here*])**.** *Suppose every bounded product selector form $F = \sum_{j \in \{\pm 1\}^K} y_j \prod_{r \leq K} \frac{1 + j_r a_r}{2}$ (with the $a_r \in [-1,1]$ depending on disjoint blocks*

of variables) of degree at most D_F satisfies $\max(\varepsilon,\tau) \ge c'(V/D_F)^{C'}$ *with* $C' \ge 1 \ge c'$. *Then Conjecture 1.1 holds with polynomial loss. Conversely, the conjecture gives the hypothesis, since* $F \in \mathcal{B}_{D_F}$ *and* $\mathrm{MaxInf}(F) = \max(\varepsilon,\tau)$.

Proof. Step 1. Let $a \in \mathcal{B}_k$, $\sigma^2 = \mathrm{Var}a > 0$, $\eta = \mathbb{E}(1-a^2)$, $L = \ln(64C'k/\sigma^2) + \ln(2/c')$, $K = \lceil 8C'L/\sigma^2 \rceil$, and assume $\eta \le \sigma^2/(16C'L)$. Take K disjoint copies a_r of a; then $g_j \ge 0$, $\sum_j g_j = 1$, and exactly: $V = (1-\eta/2)^K \ge \frac{1}{2}$ (as $K\eta \le 1$); $\varepsilon = w_*^K$ with $w_* = (1+2|\mathbb{E}a| + \mathbb{E}a^2)/4$ and $1 - w_* \ge (1-|\mathbb{E}a|)/2 \ge \sigma^2/4$, so $\varepsilon \le e^{-K\sigma^2/4} \le e^{-2C'L}$; and $\mathrm{Inf}_{z_i} F = (1-\eta/2)^{K-1}\mathrm{Inf}_i(a)/2$ for z_i in a copy. As $D_F \le 2Kk$ and $e^{-2C'L} < c'(4Kk)^{-C'}$, the hypothesis gives $\mathrm{MaxInf}(a) \ge 2c'(4Kk)^{-C'}$. *Step 2.* Apply Theorem 7.2 to f with $\delta = v/(576C'L_*)$ and $w = v^{3/2}/(864C'L_*)$, where $L_* = C_3 \ln(C_4 d/(c'v))$ bounds $L(a)$, with C_3, C_4 depending on C' (the degree m depends on w, δ only logarithmically, so $L(a) \le A + B\ln L_*$ and L_* can be chosen self-consistently). Then $\mathbb{E}(1-a^2)/\mathrm{Var}a \le 1/(16C'L(a))$, Step 1 applies to $a = P \circ f$, and (iv) gives $\mathrm{MaxInf}(f) \ge m^{-4}2c'(4Kmd)^{-C'} = \mathrm{poly}(v/d)$. □

7.2. **An average-case Nisan lemma.** The *average-case Nisan lemma* $\mathrm{AN}(C_1, C_2)$ is the statement: if $\pi\colon Q_n \to \Sigma$ and $E \subseteq Q_n$ with $\mu(E) \le c_1(\varepsilon/b)^{C_1}$ are such that at every $z \notin E$ at most b pairwise disjoint blocks B satisfy $z^B \notin E$ and $\pi(z^B) \ne \pi(z)$, then π is ε-close to a decision tree of depth at most $(b/\varepsilon)^{C_2}$.

Theorem 7.5 ([*proved here*])**.** *If* $\mathrm{AN}(C_1, C_2)$ *holds for some* $C_1 < 1$, *then Conjecture 1.1 holds with polynomial loss.*

Proof. Let $v = \mathrm{Var}f$, $\varepsilon = v/2^{12}$, $w = \min(v^{3/2}/2000, [(c_1/3)v^{1/2}(\varepsilon/(2d^2))^{C_1}]^{1/(1-C_1)}) = \mathrm{poly}(v/d)$, and θ, P, a as in Theorem 7.2 with $\delta = v^{1/2}/40$; let $\pi = \sigma$ and $E = \{|f-\theta| < w\}$. At $z \notin E$, disjoint blocks $B_1, \dots, B_b$ with $z^{B_i} \notin E$ and $\sigma(z^{B_i}) \ne \sigma(z)$ change f by at least $2w$, all in the same direction; the symmetrization $Q(s)$ of f along the blocks has degree at most d, $|Q| \le 1$ on $\{0, \dots, b\}$ and $|Q(1) - Q(0)| \ge 2w$, so Lemma 6.11(a) gives $b \le b_0 := \lfloor 2d^2/w \rfloor$. Then $\mu(E) \le 3w/v^{1/2} \le c_1(\varepsilon/b_0)^{C_1}$, so there is a tree T of depth $D \le (b_0/\varepsilon)^{C_2}$ with $\Pr[T \ne \sigma] \le \varepsilon$. Now $\mathbb{E}(a-T)^2 \le 2\mathbb{E}(a-\sigma)^2 + 8\varepsilon \le 0.016v$, and $\tilde{T} = \mathbb{E}[a \mid \text{leaf of } T]$ satisfies $\mathrm{Var}\tilde{T} = \mathrm{Var}a - \mathbb{E}(a-\tilde{T})^2 \ge 0.133v - 0.016v \ge 0.1v$. Theorem 4.1(L1) gives $\mathrm{Var}\tilde{T} = \mathrm{Cov}(\tilde{T}, a) \le D\mathrm{MaxInf}(a)^{1/2}$, so $\mathrm{MaxInf}(f) \ge m^{-4}(0.1v/D)^2$. □

[*heuristic*] Fourier-tail, junta, and hypercontractive arguments do not seem able to prove AN with $C_1 < 1$: Majority appears to satisfy all averaged hypotheses of this type and needs unbounded depth when $b\mu(E) \gtrsim 1$. We have not made this precise.

A one-scale restriction transfer that is equivalent to the conjecture even for top-heavy functions (Theorem G.4), the quasi-polynomial bound it would give (Theorem G.2), and a search form of the conjecture (Theorem G.5, a proof outline) are in Appendix G.

7.3. **Degree-free formulations.** The degree enters the arguments above mainly through Lemma 2.2. Here it is replaced by a block-sensitivity parameter that is defined for every bounded function. For $x, z \in Q_n$ and $B \subseteq [n]$ let (z_B, x_{B^c}) be x with the coordinates in B overwritten by those of z.

Definition 7.6 (real block sensitivity)**.** For $g\colon Q_n \to [-1,1]$ and $x \in Q_n$ let $\mathrm{rbs}(g,x)$ be the maximum of $\sum_{i=1}^b \frac{1}{2}|g(z^{(i)}_{B_i}, x_{B_i^c}) - g(x)|$ over $b \ge 0$, pairwise disjoint $B_1, \dots, B_b \subseteq [n]$ and $z^{(1)}, \dots, z^{(b)} \in Q_n$, and let $\mathrm{rbs}(g) = \max_x \mathrm{rbs}(g,x)$.

Overwriting B by z is the same as flipping the block $\{j \in B : z_j \ne x_j\}$, so $\mathrm{rbs}(g)$ is the block sensitivity $\mathrm{BS}((1+g)/2)$ of [35, Definition 11]. Singletons give $\sum_i |d_i g(x)| \le \mathrm{rbs}(g,x)$,

hence $\mathrm{Var} g \le I(g) \le \sum_i \|d_i g\|_1 \le \mathrm{rbs}(g)$, as $|d_i g| \le 1$. Restrictions do not increase rbs, and $\mathrm{rbs}(2h-1) = \mathrm{bs}(h)$ for $h\colon Q_n \to \{0,1\}$.

Lemma 7.7 (block-sensitivity potential; [*known*]: (a) is [29, Proposition 4.3], and (b) is a variant of [29, Lemma 6.1] and [35, Claim 33] with an overwrite in place of the mean deviation; proofs included)**.** (a) *If $g \in \mathcal{B}_d(n)$ then* $\mathrm{rbs}(g) \le d^2$. (b) *Let $g\colon Q_n \to [-1,1]$, $J \subseteq [n]$, $z \in Q_n$, and $\Delta_J(g;z) = \frac{1}{2}\mathbb{E}_x|g(z_J, x_{J^c}) - g(x)|$. Then, with b uniform on $\{-1,1\}^J$ and u uniform on $\{-1,1\}^{J^c}$,*

$$\mathbb{E}_x \,\mathrm{rbs}(g,x) \;\ge\; \Delta_J(g;z) + \mathbb{E}_b \mathbb{E}_u \,\mathrm{rbs}(g|_{x_J=b}, u).$$

Proof. (a) Fix x, disjoint blocks $B_1, \dots, B_b$ and $z^{(1)}, \dots, z^{(b)}$. For $w \in Q_b$ let $y(w)$ agree with x outside $\bigcup_i B_i$ and with $\frac{1+w_i}{2}x + \frac{1-w_i}{2}z^{(i)}$ on B_i. Then $y(w) \in Q_n$, and $h(w) = g(y(w))$ has degree at most d, since each coordinate of $y(w)$ is affine in a single w_i. As $d_i h(\mathbf{1}) = \frac{1}{2}(g(x) - g(z^{(i)}_{B_i}, x_{B_i^c}))$, the sum in Definition 7.6 is $\sum_i |d_i h(\mathbf{1})| \le d^2$ by [19, Theorem 3.3] (or $4d^2$ by Lemma 2.2). (b) For $x = (b,u)$, an optimal family of blocks in J^c for $g|_{x_J=b}$ at u, together with the block J overwritten by z, is admissible for g at x. Hence $\mathrm{rbs}(g,x) \ge \frac{1}{2}|g(z_J,u) - g(b,u)| + \mathrm{rbs}(g|_{x_J=b}, u)$; take expectations. □

For a function $r(d,\eta)$ let $\mathrm{ACE}(r)$ be the statement: for every $f \in \mathcal{B}_d$ and $\eta \in (0,1]$ there are $J \subseteq [n]$ with $|J| \le r(d,\eta)$ and $b \in \{-1,1\}^J$ such that $\mathrm{Var}(f|_{x_J=b}) \le \eta$. Lovett and Zhang conjecture $\mathrm{ACE}(r)$ with $r = \mathrm{poly}(d, 1/\eta)$ [35, Conjecture 32] and relate it to the decision-tree form of the conjecture through the potential of Lemma 7.7(b) [35, Claim 33]. The following remark makes this quantitative for an arbitrary codimension r.

Remark 7.8 (ACE and the conjecture; [*unaudited*]). (a) If $\mathrm{ACE}(r)$ holds, then every $f \in \mathcal{B}_d$ with $v = \mathrm{Var} f > 0$ satisfies, with $r = r(d, v^4/256)$,

$$\max_i \|d_i f\|_1 \;\ge\; \frac{v^3}{64\, d^2 r}, \qquad \mathrm{MaxInf}(f) \;\ge\; \frac{v^6}{2^{12} d^4 r^2}.$$

(b) Conversely, if $\mathrm{MaxInf}(g) \ge c(\mathrm{Var} g/d)^C$ for all $g \in \mathcal{B}_d$, then $\mathrm{ACE}(r)$ holds with $r(d,\eta) = d^{C+1}/(c\eta^C)$. (c) The junta theorem of [16] gives $\mathrm{ACE}(r)$ with $r = 2^{O(d)}/\eta^2$, and (a) then gives a bound $2^{-O(d)}\mathrm{Var}^{O(1)}$ of the known type. ACE with $r = 2^{o(d)}\mathrm{poly}(1/\eta)$ would give $\mathrm{MaxInf}(f) \ge e^{-o(d)}\mathrm{poly}(\mathrm{Var} f)$, which is not known. Polynomial dependence on $1/\eta$ is necessary: for $\eta \le \frac{1}{4}$ and $f = m^{-1}\sum_{i \le m} x_i$ with $m = \lfloor 1/(2\eta) \rfloor$, $\mathrm{Var}(f|_{x_J=b}) \le \eta$ needs $|J| \ge m - \eta m^2 \ge 1/(4\eta) - 1$.

Proof. (a) Let $\varepsilon = v^2/4$ and $\eta = \varepsilon^2/16$. Build a decision tree as follows. At a node whose restriction g of f has $\mathrm{Var} g < \varepsilon$, stop. Otherwise take J, b from $\mathrm{ACE}(r)$ for g and η, and query the coordinates of J (they are unqueried, and $J \ne \emptyset$ since $\mathrm{Var} g > \eta$). With $m = \mathbb{E}_u g(b,u)$, $|g - m| \le 2$ gives $\mathbb{E}|g-m| \ge \mathrm{Var} g/2$, and $\mathbb{E}_u|g(b,u) - m| \le \sqrt{\eta}$, so

$$\Delta_J(g;b) \;\ge\; \tfrac{1}{2}\big(\mathbb{E}|g-m| - \sqrt{\eta}\big) \;\ge\; \tfrac{\varepsilon}{4} - \tfrac{\varepsilon}{8} \;=\; \tfrac{\varepsilon}{8}.$$

Apply Lemma 7.7(b) at every such node and sum over the tree: if N is the number of such nodes on the path of a uniform x, then $\frac{\varepsilon}{8}\mathbb{E}N \le \mathrm{rbs}(f) \le d^2$ by Lemma 7.7(a). So the revealments satisfy $\sum_i \delta_i \le r\,\mathbb{E}N \le 8d^2 r/\varepsilon$. Let $T = \mathbb{E}[f \mid \mathrm{leaf}]$. Then $\mathrm{Var} T = v - \mathbb{E}_l \mathrm{Var}(f|_l) \ge v - \varepsilon \ge v/2$ and $\mathrm{Var} T = \mathrm{Cov}(T, f)$, so Theorem 4.1(L1) gives $v/2 \le \sum_i \delta_i \|d_i f\|_1 \le (32d^2 r/v^2)\max_i \|d_i f\|_1$. The second bound follows from $\mathrm{Inf}_i(f) \ge \|d_i f\|_1^2$. (With $\varepsilon = v/2$ the same argument gives $\max_i \|d_i f\|_1 \ge v^2/(32d^2 r(d, v^2/64))$.) (b) As long as $\mathrm{Var} g > \eta$, restrict the most influential coordinate i of the current restriction g to the value β that minimizes $I(g|_{x_i=\beta})$. Since $\mathbb{E}_\beta I(g|_{x_i=\beta}) = I(g) - \mathrm{Inf}_i(g)$, each step lowers I by at least $c(\eta/d)^C$, and $0 \le I \le d$; so there are at most $d^{C+1}/(c\eta^C)$ steps.

In these arguments the degree enters only through the codimension r and the bound $\mathrm{rbs}(f) \le d^2$. The following degree-free conjecture keeps only rbs.

Conjecture 7.9 (real Nisan; [*open*])**.** *There are absolute constants $c, C > 0$ such that for every $g\colon Q_n \to [-1,1]$ and every $\varepsilon \in (0,1]$ some decision tree T of depth at most $C(\mathrm{rbs}(g)/\varepsilon)^C$ satisfies $\mathbb{E}(g-T)^2 \le \varepsilon$.*

For $g = 2h-1$ with $h\colon Q_n \to \{0,1\}$ it holds with $\varepsilon = 0$ and depth $\mathrm{rbs}(g)^3$, by Nisan's bound (Lemma 5.7); Conjecture 7.9 is its real-valued form. By Lemma 7.7(a) it implies Conjecture 1.1 (Remark 7.10). We have not found it in print. The closest statements are the degree-free bound on fractional block sensitivity in [35, Theorem 15] and the potential of [29, Lemma 6.1] and [35, Claim 33].

Remark 7.10 (real Nisan and influences; [*unaudited*]). (a) Conjecture 7.9 is equivalent, up to polynomial changes of the constants, to the degree-free influence bound bsAA: $\mathrm{MaxInf}(g) \ge c(\mathrm{Var}g/\mathrm{rbs}(g))^C$ for every $g\colon Q_n \to [-1,1]$ with $\mathrm{Var}g > 0$. Both imply Conjecture 1.1, since $\mathrm{rbs}(f) \le d^2$ for $f \in \mathcal{B}_d$. *Proof.* Let $v = \mathrm{Var}g$ and $\beta = \mathrm{rbs}(g)$. If T is as in Conjecture 7.9 with $\varepsilon = v/2$ and depth t, then $T' = \mathbb{E}[g \mid \text{leaf of } T]$ satisfies $\mathbb{E}(g-T')^2 \le \mathbb{E}(g-T)^2 \le v/2$, so $\mathrm{Var}T' \ge v/2$, and Theorem 4.1(L1) gives $v/2 \le \mathrm{Cov}(T', g) \le t \max_i \|d_i g\|_1$; hence $\mathrm{MaxInf}(g) \ge v^2/(4t^2)$. Conversely, assume bsAA and let $\theta = c(\varepsilon/(2\beta))^C$. At every node whose restriction h has $\mathrm{Var}h \ge \varepsilon/2$, query a most influential coordinate i of h; then $\mathrm{Inf}_i(h) \ge \theta$, since $\mathrm{rbs}(h) \le \beta$. As $\mathbb{E}_{x_i} I(h|_{x_i}) = I(h) - \mathrm{Inf}_i(h)$ and $I(g) \le \beta$, the expected number of queries is at most β/θ. Cut the tree at depth $t = 2\beta/(\theta\varepsilon)$ and let $T = \mathbb{E}[g \mid \text{leaf}]$. By Markov's inequality the cut leaves have total probability at most $\varepsilon/2$, and the other leaves l have $\mathrm{Var}(g|_l) < \varepsilon/2$, so $\mathbb{E}(g-T)^2 \le \varepsilon$, with depth $t = \mathrm{poly}(\beta/\varepsilon)$. (b) *Known cases.* If $g = 2h-1$ with $h\colon Q_n \to \{0,1\}$, Lemma 5.7 and Theorem 4.1(L1) give $\mathrm{MaxInf}(g) \ge \mathrm{Var}(g)/\mathrm{rbs}(g)^3$ [*standard*]. If $|g| \ge 1-\eta$ pointwise with $\eta < 1$, and $\sigma = \operatorname{sgn} g$, then a block that changes σ changes g by at least $2(1-\eta)$; so $\mathrm{bs}(\sigma) \le \mathrm{rbs}(g)/(1-\eta)$ and $\mathrm{Inf}_i(g) \ge (1-\eta)^2 \mathrm{Inf}_i(\sigma)$, and the Boolean case gives $\mathrm{MaxInf}(g) \ge (1-\eta)^5 \mathrm{Var}(\sigma)/\mathrm{rbs}(g)^3$. If $\|d_i g\|_2 \le \lambda \|d_i g\|_1$ for all i (non-spiky derivatives), then $I(g) \le \mathrm{MaxInf}(g)^{1/2} \sum_i \|d_i g\|_2 \le \lambda \mathrm{MaxInf}(g)^{1/2} \mathrm{rbs}(g)$, so $\mathrm{MaxInf}(g) \ge I(g)^2/(\lambda^2 \mathrm{rbs}(g)^2)$. For subcube selectors $\sum_l c_l \mathbf{1}_{C_l}$ with subcubes C_l and $\sum_l |c_l| \mathbf{1}_{C_l} \le 1$ we have proofs, not included here and also unaudited, of bounds of the form $\mathrm{MaxInf} \ge c\mathrm{Var}^6/\mathrm{rbs}^4$. In general bsAA is open.

7.4. **Barriers for proofs of the general statement.** This subsection uses its own notation and records two obstructions to proofs of Conjecture 1.1: the degree must enter beyond the pointwise bound of Lemma 2.2, and dual certificates must depend nonlinearly on the test function. Let $s(f) = \max_x \sum_i |d_i f(x)|$ be the real sensitivity of $f\colon Q_n \to [-1,1]$; thus $s(f) \le 4d^2$ on $\mathcal{B}_d$, and $s(f) \le \mathrm{rbs}(f)$.

The soft-max tournament. Let $\beta > 0$, $\frac{1}{1+e^{2\beta}} < \theta < \frac12$ and $c_\beta = e^{2\beta} - 1$. For $b \ge 1$ let $k_0 = \lfloor \theta b \rfloor$, $q_0 = \Pr[\mathrm{Bin}(b, \frac12) \le k_0]$ and $m = \lfloor 1/q_0 \rfloor$; as $k_0 < b/2$, $q_0 \le \frac12$, so $m \ge 2$ and $mq_0 > 1 - q_0 \ge \frac12$. Let $x \in Q_{mb}$, split into blocks $B_1, \dots, B_m$ of size b, and $y \in Q_m$; put $S_j = \sum_{i \in B_j} x_i$, $K_j = (b - S_j)/2$ (the number of -1's in B_j) and

$$p_j(x) = \frac{e^{\beta S_j}}{\sum_{l=1}^m e^{\beta S_l}}, \qquad F(x,y) = \sum_{j=1}^m p_j(x)\, y_j.$$

Let $D(x) = \int_x^{1/2} \ln \frac{1-u}{u}\, du$, the relative entropy of the Bernoulli laws with means x and $\frac12$.

Proposition 7.11 (soft-max tournament; [*proved here*])**.** $|F| \le 1$, *and:*

(i) $\sum_i |d_i F(x,y)| \le 1 + c_\beta b$ *at every point, the sum running over all $mb + m$ coordinates;*

(ii) $\mathrm{Inf}_{y_j}(F) \le 1/m$ *and* $\mathrm{Inf}_{x_i}(F) \le c_\beta^2/(2m)$, *so* $\mathrm{Var}(F)/m \le \mathrm{MaxInf}(F) \le \max(1, c_\beta^2/2)/m$, *where* $\frac{1}{2}e^{2(1/2-\theta)^2 b} \le m \le (b+1)e^{bD(k_0/b)}$;
(iii) $\mathrm{Var}F \ge c(\beta,\theta)$ *for every* b, *with* $c(\beta,\theta) > 0$ *independent of* b;
(iv) $\mathrm{rbs}(F) \ge \frac{1}{2}(m-1)(1-2me^{-2\beta b})$, *where* $me^{-2\beta b} \to 0$ *as* $b \to \infty$; *so* $\mathrm{rbs}(F) = e^{\Omega(b)}$ *and* $\deg F \ge \mathrm{rbs}(F)^{1/2} = e^{\Omega(b)}$.

Proof. $|F| \le \sum_j p_j|y_j| = 1$. Fix j, $i \in B_j$ and (x,y), let $x^\pm = x^{i\to\pm 1}$ and $p = p_j(x^-)$. Only S_j changes, by 2, so $p_j(x^+) = pe^{2\beta}/(1+pc_\beta)$ and $p_l(x^+) = p_l(x^-)/(1+pc_\beta)$ for $l \ne j$. Hence $\delta := p_j(x^+) - p_j(x^-) = c_\beta p(1-p)/(1+pc_\beta)$ and, as $\sum_l p_l = 1$, $\sum_{l\ne j}|p_l(x^+) - p_l(x^-)| = \delta$. Moreover $\delta \le c_\beta p_j(x)$: for $x_i = -1$ this reads $\delta \le c_\beta p$, and for $x_i = 1$, $\delta = c_\beta e^{-2\beta}(1-p)\,p_j(x^+)$. As $d_{x_i}F = \frac{1}{2}\sum_l (p_l(x^+) - p_l(x^-))y_l$, we get $|d_{x_i}F| \le \delta \le c_\beta p_j(x)$ and $\mathbb{E}_y(d_{x_i}F)^2 = \frac{1}{4}\sum_l (p_l(x^+) - p_l(x^-))^2 \le \frac{1}{2}\delta^2$.

(i) Summing over $i \in B_j$ and j gives $\sum_i |d_{x_i}F| \le c_\beta b \sum_j p_j = c_\beta b$, and $d_{y_j}F = p_j$. (ii) The K_j are independent $\mathrm{Bin}(b, \frac{1}{2})$ variables, so $\mathbb{E}p_j = 1/m$ and $\mathrm{Inf}_{y_j}(F) = \mathbb{E}p_j^2 \le 1/m$, and $\mathrm{Inf}_{x_i}(F) \le \frac{1}{2}c_\beta^2 \mathbb{E}p_j^2$ for $i \in B_j$. As $\mathbb{E}_y F = 0$, $\mathrm{Var}F = \mathbb{E}\sum_j p_j^2 = \sum_j \mathrm{Inf}_{y_j}(F)$. Finally $m \ge 1/q_0 - 1 \ge 1/(2q_0)$ with $q_0 \le e^{-2(1/2-\theta)^2 b}$ by Hoeffding's inequality, and $m \le 1/q_0 \le 2^b/\binom{b}{k_0} \le (b+1)e^{bD(k_0/b)}$ by the standard bound $\binom{b}{k} \ge e^{bH(k/b)}/(b+1)$, H the natural entropy, since $D = \ln 2 - H$.

(iii) $\max_j p_j = 1/W$ with $W = \sum_l e^{-2\beta(K_l - k_{\min})}$ and $k_{\min} = \min_l K_l$, so $\mathrm{Var}F \ge \mathbb{E}W^{-2} \ge \Pr[W \le \Lambda]/\Lambda^2$ for every $\Lambda > 0$. Let $\rho = \theta/(1-\theta)$; the hypothesis on θ says $1/\rho < e^{2\beta}$. Put $r' = (1/\rho + e^{2\beta})/2$ and $P(k) = \Pr[\mathrm{Bin}(b, \frac{1}{2}) = k]$. For $k \le k_0$, $P(k-1)/P(k) = k/(b-k+1) < \rho$, so $m\Pr[K_1 \le k_0 - s] \le \rho^s m q_0 \le \rho^s$ for $s \ge 0$. Since $P(k+1)/P(k) = (b-k)/(k+1)$ decreases in k, $P(k_0+u) \le P(k_0)r_0^u$ for $u \ge 0$, with $r_0 = (b-k_0)/(k_0+1) < 1/\rho + 1/(\theta b) \le r'$ once $b \ge b_0 := 2/(\theta(e^{2\beta} - 1/\rho))$; then $mP(k_0+u) \le r'^u$. Fix an integer $t \ge 0$ with $\rho^{t+1} \le \frac{1}{10}$, and let $Z = \sum_l e^{-2\beta(K_l - k_0)}\mathbf{1}[K_l \ge k_0 - t]$ and $\Gamma = \sum_{s=1}^t (\rho e^{2\beta})^s + (1 - r'e^{-2\beta})^{-1}$. For $b \ge b_0$: $\Pr[k_{\min} > k_0] = (1-q_0)^m \le e^{-mq_0} < e^{-1/2}$; $\Pr[k_{\min} < k_0 - t] \le m\Pr[K_1 \le k_0 - t - 1] \le \frac{1}{10}$; and $\mathbb{E}Z = m\sum_{k \ge k_0 - t} P(k)e^{-2\beta(k-k_0)} \le \Gamma$, so $\Pr[Z > 10\Gamma] \le \frac{1}{10}$. Outside these three events $k_0 - t \le k_{\min} \le k_0$, so $W = e^{2\beta(k_{\min}-k_0)}Z \le 10\Gamma$. As $1 - e^{-1/2} - \frac{2}{10} > 0.19$, $\mathrm{Var}F \ge 0.19/(10\Gamma)^2$ for $b \ge b_0$; for the finitely many $b < b_0$, $\mathrm{Var}F \ge \mathbb{E}(\sum_j p_j)^2/m = 1/m > 0$.

(iv) Let x be $+1$ on B_1 and -1 elsewhere, $y_1 = 1$ and $y_j = -1$ for $j \ge 2$. Then $F(x,y) = 2p_1 - 1 \ge 1 - 2(m-1)e^{-2\beta b}$, since $p_1 = 1/(1 + (m-1)e^{-2\beta b})$. Overwriting one block B_j, $j \ge 2$, by $+1$'s gives $p_1 = p_j \le \frac{1}{2}$ and $F = 2p_1 - 1 \le 0$. These $m-1$ blocks are disjoint, which gives the bound on $\mathrm{rbs}(F)$. As $D(x) \le (\frac{1}{2} - x)\ln\frac{1-x}{x}$, $D(\theta) < \frac{1}{2}\ln(1/\rho) < \beta$; since $D(k_0/b) \to D(\theta)$, (ii) gives $me^{-2\beta b} \le (b+1)e^{b(D(k_0/b) - 2\beta)} \to 0$ and $m = e^{\Omega(b)}$. The degree bound is Lemma 7.7(a). □

By Proposition 7.11, for all large b, $\mathrm{Var}F \ge c(\beta,\theta)$ and $\mathrm{MaxInf}(F) \le e^{-\kappa s(F)}$ with $\kappa = (\frac{1}{2} - \theta)^2/c_\beta$. So the *sensitivity form* of Conjecture 1.1, $\mathrm{MaxInf}(f) \ge \mathrm{poly}(\mathrm{Var}f/s(f))$ for all $f\colon Q_n \to [-1,1]$, which would imply the conjecture, is false with an exponential gap. Hence a proof of Conjecture 1.1 must use the degree beyond the bound $\sum_i |d_i f| \le O(d^2)$: an argument that uses only $|f| \le 1$, $\mathrm{Var}f$ and this pointwise bound (an induction on the dimension, a Bellman function, or a martingale or decision-tree potential in these quantities) would prove the sensitivity form. The gap is the largest possible [*proof outline*]: the i-th term of Talagrand's L^1–L^2 inequality [52] is $O(\|d_i f\|_1/\ln(1/\mathrm{MaxInf}(f)))$, so $\mathrm{MaxInf}(f) \ge \exp(-O(s(f)/\mathrm{Var}f))$. By (iv), F is consistent with Conjecture 7.9; its degree is exponential in b. The closest precedent we know is a worst-case separation: Tal [51, Theorem 10] constructs $h\colon \{0,1\}^n \to [0,1]$ with $\mathrm{bs}_\ell(h) \ge \lfloor n/\ell \rfloor$ and $s(h) \le 3\ell n^{1/\ell}$, bs_ℓ using blocks of size at most ℓ; it does not concern influences or variance.

Convex duality for the extremal problem. Fix n, d and $\tau \ge 0$, let $\langle f,g\rangle = \mathbb{E}fg$, let $\mathcal{V}$ be the space of functions $Q_n \to \mathbb{R}$ of degree at most d, $\mathcal{V}_0 = \{h \in \mathcal{V} : \mathbb{E}h = 0\}$, $P_{\le d}$ the orthogonal projection onto $\mathcal{V}$, and $\|g\|_{\mathcal{B}_d^*} = \max_{f\in\mathcal{B}_d(n)}\langle f,g\rangle$. For $h\colon Q_n \to \mathbb{R}$ let

$$\mathrm{val}_\tau(h) = \max\big\{\langle f,h\rangle :\ f \in \mathcal{B}_d(n),\ \|d_i f\|_2 \le \sqrt{\tau} \text{ for all } i\big\}.$$

For unit $h \in \mathcal{V}_0$, $\langle f,h\rangle = \langle f - \mathbb{E}f, h\rangle \le (\mathrm{Var}f)^{1/2}$, with equality for $h = (f-\mathbb{E}f)/(\mathrm{Var}f)^{1/2}$. Hence the supremum of $\mathrm{val}_\tau(h)$ over unit $h \in \mathcal{V}_0$ is $\sup\{(\mathrm{Var}f)^{1/2} : f \in \mathcal{B}_d(n),\ \mathrm{MaxInf}(f) \le \tau\}$, and Conjecture 1.1 with constants c, C says exactly that $\mathrm{val}_\tau(h) \le d^{1/2}(\tau/c)^{1/(2C)}$ for all n, d, τ and unit $h \in \mathcal{V}_0$. A *dual certificate* is a feasible point of the following dual problem.

Proposition 7.12 (second-order-cone duality; [*proved here*])**.** *For every $h\colon Q_n \to \mathbb{R}$ and $\tau \ge 0$,*

$$\mathrm{val}_\tau(h) = \min\Big\{\|h_0\|_1 + \sqrt{\tau}\sum_{i=1}^n \|u_i\|_2 :\ h + q = h_0 + \sum_{i=1}^n x_i u_i,\ q \perp \mathcal{V},\ u_i \text{ independent of } x_i\Big\},$$

and the minimum is attained. One may require $\deg u_i \le d-1$*, and, eliminating h_0 and q,*

$$\mathrm{val}_\tau(h) = \min_u \Big(\big\|P_{\le d}\big(h - \textstyle\sum_i x_i u_i\big)\big\|_{\mathcal{B}_d^*} + \sqrt{\tau}\,\textstyle\sum_i \|u_i\|_2\Big).$$

Proof. Weak duality. If f is admissible and (h_0,q,u) is feasible, then $\langle f,q\rangle = 0$, $\langle f,h_0\rangle \le \|h_0\|_1$, and $\langle f, x_i u_i\rangle = \mathbb{E}[u_i\, d_i f] \le \sqrt{\tau}\|u_i\|_2$, since u_i does not depend on x_i and $\mathbb{E}_{x_i}[x_i f] = d_i f$. *Strong duality.* For $\tau = 0$ the admissible f are the constants in $[-1,1]$, so $\mathrm{val}_0(h) = |\hat h(\emptyset)|$, and assigning each nonempty S to its smallest element writes $h - \hat h(\emptyset) = \sum_i x_i u_i$. Let $\tau > 0$. The admissible set is $\mathcal{V} \cap B_\infty \cap \bigcap_i C_i$ with $B_\infty = \{\|f\|_\infty \le 1\}$ and $C_i = \{\|d_i f\|_2 \le \sqrt{\tau}\}$, closed convex sets in $\mathbb{R}^{Q_n}$ whose relative interiors contain 0. By [45, Corollary 16.4.1] its support function val_τ is the infimal convolution of the support functions of $\mathcal{V}$, B_∞ and the C_i, and the infimum is attained. These support functions are the indicator of $\mathcal{V}^\perp$ (which gives q), $\|\cdot\|_1$ (which gives h_0), and, for C_i, $g \mapsto \sqrt{\tau}\|u\|_2$ if $g = x_i u$ with u independent of x_i and $+\infty$ otherwise, since C_i contains the functions independent of x_i and $\langle f, x_i u\rangle = \langle d_i f, u\rangle$. This is the stated dual. Replacing u_i by its part of degree at most $d-1$ lowers $\|u_i\|_2$, and the discarded part of $x_i u_i$ has degree above d, so it can be moved into q. For the last formula, $\min_{q\perp\mathcal{V}}\|g+q\|_1 = \max_{f\in\mathcal{B}_d(n)}\langle f,g\rangle = \|P_{\le d}g\|_{\mathcal{B}_d^*}$, by the same corollary for $\mathcal{V}\cap B_\infty$. □

A *linear scheme* is a family $A = (A_i)_{i\le n}$ of linear maps from $\mathcal{V}_0$ to functions on Q_n such that $A_i h$ does not depend on x_i; the maps may depend on n, d and τ, but not on h. Its cost

$$\mathrm{Cost}_A(h) = \big\|P_{\le d}\big(h - \textstyle\sum_i x_i A_i h\big)\big\|_{\mathcal{B}_d^*} + \sqrt{\tau}\,\textstyle\sum_i \|A_i h\|_2$$

is at least $\mathrm{val}_\tau(h)$ by Proposition 7.12; the ℓ^1 part (h_0,q) is still optimized for each h. Examples are $A_i h = d_i M h$ for Fourier multipliers M. For $M\chi_S = \chi_S/|S|$ $(S \ne \emptyset)$, $\sum_i x_i A_i h = h$ and, by Cauchy–Schwarz, $\mathrm{Cost}_A(h) \le (\tau n)^{1/2}(\sum_S |S|^{-1}\hat h(S)^2)^{1/2} \le (\tau n)^{1/2}$ for unit h: this certifies the trivial bound $\mathrm{Var}f \le n\mathrm{MaxInf}(f)$.

Proposition 7.13 (linear dual certificates are trivial; [*proved here*])**.** *Let $1 \le k \le d-1$, $k + 2^k \le n$ and $c_k = \sqrt{\tau}\,2^{k/2}/(k+1)$. For every linear scheme A some image h of Addr_k (on $k+2^k$ of the coordinates) under a permutation and sign changes of the coordinates has $\mathrm{Cost}_A(h) \ge \min(1, c_k)$. Hence, if $\mathrm{Cost}_A(h) \le \sqrt{\varepsilon} < 1$ for every unit $h \in \mathcal{V}_0$, then $\tau \le \varepsilon(k+1)^2 2^{-k}$; in particular $\tau \le \varepsilon d^2 2^{1-d}$ if $n \ge d - 1 + 2^{d-1}$, and $\tau \le 4\varepsilon(\log_2 n)^2/n$ if $4 \le n \le 2^d$.*

Proof. Let $\mathcal{G}$ be the group of signed permutations of the coordinates, acting on functions; it preserves $\mathcal{B}_d(n)$, $\mathcal{V}_0$, $P_{\le d}$, the L^p norms and $\|\cdot\|_{\mathcal{B}_d^*}$. If $\gamma \in \mathcal{G}$ maps x_i to $\pm x_{\pi(i)}$, let it act on tuples by $(\gamma u)_{\pi(i)} = \pm\gamma(u_i)$, with the same sign. Then $\sum_j x_j(\gamma u)_j = \gamma(\sum_i x_i u_i)$ and $(\gamma u)_{\pi(i)}$ does not

depend on $x_{\pi(i)}$, so $\mathrm{Cost}(u;h) = \|P_{\le d}(h - \sum_i x_i u_i)\|_{\mathcal{B}_d^*} + \sqrt{\tau}\sum_i \|u_i\|_2$ satisfies $\mathrm{Cost}(\gamma u;\gamma h) = \mathrm{Cost}(u;h)$, and it is convex in u. The averaged scheme $\bar{A}(h) = \mathbb{E}_\gamma\, \gamma^{-1}A(\gamma h)$ (γ uniform on $\mathcal{G}$) is linear and $\mathcal{G}$-equivariant, and $\mathrm{Cost}_{\bar{A}}(h) \le \mathbb{E}_\gamma \mathrm{Cost}_A(\gamma h)$ by Jensen's inequality. So it suffices to show $\mathrm{Cost}_{\bar{A}}(h) \ge \min(1, c_k)$ for $h = \mathrm{Addr}_k$.

The sign change of a set R of coordinates maps χ_S to $(-1)^{|S\cap R|}\chi_S$; equivariance under these maps shows that $\bar{A}_i\chi_S$ is a multiple of $\chi_{S\triangle\{i\}}$. As $\bar{A}_i\chi_S$ does not depend on x_i, $\bar{A}_i\chi_S = \beta_{i,S}\chi_{S\setminus\{i\}}$ if $i \in S$ and $\bar{A}_i\chi_S = 0$ otherwise, and equivariance under permutations gives $\beta_{i,S} = \beta_{|S|}$. Hence $h - \sum_i x_i\bar{A}_i h = Mh$ with $M\chi_S = m_{|S|}\chi_S$ and $m_l = 1 - l\beta_l$.

Let $h = \mathrm{Addr}_k$ with address bits $a_1,\dots,a_k$ and targets x_b. By Proposition 2.7, $\hat{h}(S)^2 = 4^{-k}$ for the 4^k sets $S = \{x_b\}\cup T$, $T \subseteq \{a_1,\dots,a_k\}$, and $\hat{h}(S) = 0$ otherwise. As $\pm h \in \mathcal{B}_d(n)$, the first term of the cost is at least $|\langle h, Mh\rangle| = |\mu|$ with $\mu = \mathbb{E}_T m_{|T|+1}$, T uniform. The targets alone give $\sum_b \|\bar{A}_{x_b}h\|_2 = 2^{k/2}(\mathbb{E}_T\beta_{|T|+1}^2)^{1/2} \ge 2^{k/2}\mathbb{E}_T|1 - m_{|T|+1}|/(k+1) \ge 2^{k/2}|1-\mu|/(k+1)$. So $\mathrm{Cost}_{\bar{A}}(h) \ge |\mu| + c_k|1-\mu| \ge \min(1,c_k)$. □

Proposition 7.13 covers every choice of the u_i that is linear and independent of h (for instance noise on a fixed set of coordinates, with the residual split among the u_i by a fixed linear rule), also if only the normalized $(f - \mathbb{E}f)/(\mathrm{Var}f)^{1/2}$, $f \in \mathcal{B}_d(n)$, must be certified, since these include the images of Addr_k. With $d_i Mh$ above, linear schemes thus give $\mathrm{MaxInf} \ge \mathrm{Var}/n$ and nothing better than $4\mathrm{Var}(\log_2 n)^2/n$. Choices that depend on h are much cheaper: routing each Fourier set $\{x_b\}\cup T$ with $T \ne \emptyset$ to the address bit of smallest index in T, and each $\{x_b\}$ to x_b, writes $\mathrm{Addr}_k = \sum_i x_i u_i$ with $\|u_{x_b}\|_2 = 2^{-k}$ and $\|u_{a_t}\|_2 = 2^{-t/2}$, so $\mathrm{val}_\tau(\mathrm{Addr}_k) \le \sqrt{\tau}(1 + \sum_{t\ge1} 2^{-t/2}) < 3.42\sqrt{\tau}$ for every k. Thus a dual certificate must depend nonlinearly on h; for the address function an h-dependent routing of each Fourier set to one of its coordinates suffices. Neither proposition is progress toward Conjecture 1.1; they rule out two natural strategies.

8. Toward a quasi-polynomial bound: the touching transfer

This section shows that the touching transfer, which is implied by the conjecture, would give $\mathrm{MaxInf}(f) \ge (\mathrm{Var}f/d)^{O(\log d)}$ with a polynomial loss; Section 10 proves its mass form with a quasi-polynomial prefactor, which gives Theorem 10.1. Its limits are in Appendix I, and a linear deficit transfer, which fails in the regime needed, is in Appendix H.

For a function u and $\theta > 0$ let $J_\theta(u) = \{j : \mathrm{Inf}_j(u) \ge \theta\}$ (the θ-spikes), $S_\theta(u) = \sum_{j\in J_\theta(u)} \mathrm{Inf}_j(u)$, $N_\theta(u) = I(u) - S_\theta(u)$, and let $U_\theta(u) = \sum_{S\ne\emptyset,\, S\cap J_\theta(u)=\emptyset} \hat{u}(S)^2$ be the variance not touched by θ-spikes. As before, σ is the rate-$\frac12$ restriction, ρ_t is the composition of t independent copies of σ (a rate-2^{-t} restriction), and $u_\sigma \in \mathcal{B}_d$ if $u \in \mathcal{B}_d$.

Lemma 8.1 ([*proved here*])**.** *Let $J = J_\theta(u)$.* (a) $U_\theta(u) = \mathrm{Var}\mathbb{E}[u \mid x_{J^c}]$. (b) $\mathrm{Var}u - S_\theta(u) \le U_\theta(u) \le \min(\mathrm{Var}u, N_\theta(u))$. (c) $U_\theta(u) = \mathrm{Var}u$ *if and only if* $J = \emptyset$. (d) $U_\theta(u)$ *and* $N_\theta(u)$ *are nondecreasing in* θ. (e) *At rate* p, $\mathbb{E}_\rho\mathrm{Var}(u_\rho) = \sum_S(1-(1-p)^{|S|})\hat{u}(S)^2 \in [p\mathrm{Var}u, \mathrm{Var}u]$*; in particular* $\mathbb{E}_\sigma\mathrm{Var}(u_\sigma) \ge \frac12\mathrm{Var}u$.

Proof. (a) $\mathbb{E}[u \mid x_{J^c}] = \sum_{S\cap J=\emptyset}\hat{u}(S)\chi_S$. (b) The sets meeting J carry at most $\sum_{j\in J}\mathrm{Inf}_j(u)$, and every nonempty S avoiding J contains some $j \notin J$. (c) Each $j \in J$ has $\mathrm{Inf}_j(u) \ge \theta > 0$, so some set containing j has positive weight. (d) is immediate. (e) If A is the alive set, then $\mathbb{E}\widehat{u_\rho}(T)^2 = \sum_{S\cap A = T}\hat{u}(S)^2$ for $T \subseteq A$, and $\Pr[S\cap A \ne \emptyset] = 1-(1-p)^{|S|}$. □

Definition 8.2 (touching transfer)**.** Let $C \ge 1$ (the cap exponent), $\varepsilon = \varepsilon(d) \ge 0$ and $L = L(d) \ge 1$. $\mathrm{TT}(\varepsilon, L)$ (with cap C) is the statement: for all d, all $g \in \mathcal{B}_d$ and all $0 < \theta \le \mathrm{Var}(g)^C$,

$$U_{\theta/L}(g) \;\le\; 2\,\mathbb{E}_\sigma U_\theta(g_\sigma) + \varepsilon\mathrm{Var}(g).$$

For fixed g and L let $\varepsilon_{\min}(g;L) = \sup_\theta [U_{\theta/L}(g) - 2\mathbb{E}_\sigma U_\theta(g_\sigma)]/\mathrm{Var}(g)$ over admissible θ.

Since $U_{\theta/L}(g)$ is nonincreasing in L and the right side does not depend on L, $\mathrm{TT}(\varepsilon, L)$ implies $\mathrm{TT}(\varepsilon, L')$ for $L' \geq L$, and refuting TT at L refutes it at every smaller L. The functions F of Proposition H.3 have $U_{\theta/L}(F) = 2^{-k}$ for $2^{-k} < \theta/L \leq \frac{1}{2}$, so they can violate $\mathrm{TT}(\varepsilon, L)$ only when $\varepsilon < 2^{-k}$; they do violate it at the chain's $\varepsilon = 1/(16ed)$ for small L (e.g. $k = 10$, $d = 110$, $1 \leq L < 8$; embedded in $\mathcal{B}_d$ with padding, up to $L \approx d^{0.25}$) [*numerical*] (exact evaluation), which is weaker than Corollary I.3.

The chain. The base case uses the real-sensitivity bound (Lemma 2.2) and the level-k inequality.

Lemma 8.3 (base case; [*proved here*])**.** *Let $f \in \mathcal{B}_d$, let ρ be a rate-p restriction with $p \leq 1/(2ed)$, let $\eta \in (0,1]$, $k = \lceil \log_2(d/(2\eta p)) \rceil$ and $\theta^* = (\eta p \mathrm{Var}(f)/(8d^2\, 3^k))^4$. Then $\mathbb{E}_\rho N_{\theta^*}(f_\rho) \leq \eta p \mathrm{Var}(f)$.*

Proof. Let $g \in \mathcal{B}_d$ and $\theta > 0$. For $j \notin J_\theta(g)$, Lemma 3.1(a) with $\delta = \frac{1}{3}$ gives $W^{\leq k}(d_j g) \leq 3^k \|d_j g\|_1^{3/2} \leq 3^k \theta^{1/4} \|d_j g\|_1$, since $\|d_j g\|_1 \leq \|d_j g\|_2 = \mathrm{Inf}_j(g)^{1/2} < \theta^{1/2}$. As in the proof of Theorem 3.2, $W^{\leq k}(d_j g) = \sum_{S \ni j, |S| \leq k+1} \hat{g}(S)^2$. Summing over $j \notin J_\theta(g)$, with Lemma 2.2 for the levels $\leq k+1$ and $|S| \leq d$ for the others,

$$N_\theta(g) \;\leq\; 4d^2\, 3^k \theta^{1/4} + d\, W^{>k+1}(g).$$

Take $g = f_\rho$ and $\theta = \theta^*$; the first term is then $\eta p \mathrm{Var}(f)/2$. As in the proof of Lemma G.1, $\mathbb{E} W^{>k+1}(f_\rho) = \sum_S \widehat{f}(S)^2 \Pr[\mathrm{Bin}(|S|, p) \geq k+2] \leq \mathrm{Var}(f)(edp/(k+2))^{k+2} \leq \mathrm{Var}(f) 2^{-(k+2)}$, and $d\, 2^{-(k+2)} \leq \eta p/2$ because $2^k \geq d/(2\eta p)$. ☐

(Lemma 3.1(b) gives $\theta^{1/2}$ in place of $\theta^{1/4}$, which would replace the exponent $\max(4, C)$ below by $\max(2, C)$.)

Theorem 8.4 (the chain; [*proved here*])**.** *Assume $\mathrm{TT}(\varepsilon, L)$ with cap C, where $\varepsilon = 1/(16ed)$. Let $f \in \mathcal{B}_d$, $d \geq 2$, $v = \mathrm{Var} f > 0$, $T = \lceil \log_2(2ed) \rceil$, $p = 2^{-T} \in (\frac{1}{4ed}, \frac{1}{2ed}]$, $\eta = \frac{1}{16ed}$, $k = \lceil \log_2(8ed^2/p) \rceil$ and*

$$\theta_T = \min\left\{ \left(\frac{\eta p v}{8d^2\, 3^k} \right)^4,\; \left(\frac{pv}{16ed} \right)^C \right\}.$$

Then $\mathrm{MaxInf}(f) \geq \theta_T L^{-T}$. Explicitly, $\theta_T \geq \left(v/(c_2 d^{4+3\log_2 3})\right)^{\max(4,C)}$, where $c_2 = 512e^2(64e^2)^{\log_2 3} < 6.6 \cdot 10^7$ and $4 + 3\log_2 3 < 8.76$.

Proof. Put $\theta_t = \theta_T L^{-(T-t)} \leq \theta_T$, $V_t = \mathbb{E}\mathrm{Var}(f_{\rho_t})$ and $\psi_t = \mathbb{E} U_{\theta_t}(f_{\rho_t})/V_t$. By Lemma 8.1(e), $V_t = \sum_S (1 - (1-2^{-t})^{|S|}) \widehat{f}(S)^2$ is nonincreasing in t, and $V_T \geq pv > 0$.

Step t ($1 \leq t \leq T$). Let $g = f_{\rho_{t-1}} \in \mathcal{B}_d$, so that g_σ is distributed as f_{ρ_t}. If $\theta_t \leq \mathrm{Var}(g)^C$, TT gives $U_{\theta_{t-1}}(g) \leq 2\mathbb{E}_\sigma U_{\theta_t}(g_\sigma) + \varepsilon \mathrm{Var}(g)$; otherwise $U_{\theta_{t-1}}(g) \leq \mathrm{Var}(g) < \theta_t^{1/C} \leq \theta_T^{1/C}$. Take the expectation over ρ_{t-1} and divide by $V_{t-1} \geq \max(V_t, pv)$: $\psi_{t-1} \leq 2\psi_t + \varepsilon + \theta_T^{1/C}/(pv)$.

Base. The k above is that of Lemma 8.3 for this η, since $d/(2\eta p) = 8ed^2/p$; so $\theta_T \leq \theta^*$, and by Lemma 8.1(b),(d) and Lemma 8.3, $\mathbb{E} U_{\theta_T}(f_{\rho_T}) \leq \mathbb{E} N_{\theta^*}(f_{\rho_T}) \leq \eta p v \leq \eta V_T$, i.e. $\psi_T \leq \eta$.

Iteration. Since $2^T < 4ed$ and $\theta_T^{1/C} \leq pv/(16ed)$, we get $\psi_0 \leq 2^T(\eta + \varepsilon + \theta_T^{1/C}/(pv)) \leq 4ed \cdot \frac{3}{16ed} = \frac{3}{4}$. As $V_0 = v$, this says $U_{\theta_0}(f) < \mathrm{Var} f$, so $J_{\theta_0}(f) \neq \emptyset$ by Lemma 8.1(c), and $\mathrm{MaxInf}(f) \geq \theta_0 = \theta_T L^{-T}$.

Constants. We have $\eta p \geq 1/(64e^2d^2)$ and $2^k \leq 16ed^2/p \leq 64e^2d^3$, so $3^k \leq (64e^2d^3)^{\log_2 3}$ and $\eta p v/(8d^2\, 3^k) \geq v/(c_2 d^{4+3\log_2 3}) =: b_0$. Also $pv/(16ed) \geq v/(64e^2d^2) \geq b_0$. Since $b_0 \leq 1$, $\theta_T \geq \min(b_0^4, b_0^C) \geq b_0^{\max(4,C)}$. ☐

Corollary 8.5 ([*proved here*])**.** $\mathrm{TT}(1/(16ed), d^b)$ *with cap* C *implies, for all* $f \in \mathcal{B}_d$,

$$\mathrm{MaxInf}(f) \;\geq\; \Big(\frac{\mathrm{Var} f}{c_2 d^{8.76}}\Big)^{\max(4,C)} d^{-b\log_2(4ed)} = \Big(\frac{\mathrm{Var} f}{d}\Big)^{O(C+b\log d)},$$

and $\mathrm{TT}(1/(16ed), e^{o(d/\log d)})$ *implies* $\mathrm{MaxInf}(f) \geq (\mathrm{Var} f/d)^{O(C)} e^{-o(d)}$.

Spike averaging and the sandwich.

Lemma 8.6 (spike averaging; [*proved here*])**.** *Let* $g \in \mathcal{B}_d$, $\theta > 0$, $L \geq 1$, $J' = J_{\theta/L}(g)$ *and* $h = \mathbb{E}[g \mid x_{[n]\setminus J'}] = \sum_{S\cap J'=\emptyset} \hat{g}(S)\chi_S$. *Then* $h \in \mathcal{B}_d$, $\mathrm{Var} h = U_{\theta/L}(g)$, $\mathrm{Inf}_j(h) \leq \mathrm{Inf}_j(g) < \theta/L$ *for* $j \notin J'$, *and* $\mathrm{Inf}_j(h) = 0$ *for* $j \in J'$. *If moreover* $\varepsilon > 0$, $U_{\theta/L}(g) > \varepsilon \mathrm{Var}(g)$ *and* $\theta \leq \mathrm{Var}(g)^C$, *then* $\mathrm{MaxInf}(h) < \theta/L \leq \mathrm{Var}(g)^C/L < \mathrm{Var}(h)^C/(\varepsilon^C L)$.

Proof. h is a conditional expectation of a $[-1,1]$-valued function and its Fourier support lies in that of g, so $h \in \mathcal{B}_d$; the other claims are read off the Fourier expansion. □

The map $g \mapsto h$ is the standard averaging operator; [12, §6] uses the complementary junta $\sum_{S\subseteq J} \widehat{f}(S)\chi_S$.

Theorem 8.7 (sandwich; [*proved here*])**.** *Write* $\mathrm{AA}(A,c)$ *for the statement that* $\mathrm{MaxInf}(u) \geq c(\mathrm{Var} u/d)^A$ *for all* d *and all* $u \in \mathcal{B}_d$.

(i) $\mathrm{TT}(1/(16ed), L)$ *implies* $\mathrm{MaxInf}(f) \geq \theta_T L^{-T}$ *(Theorem 8.4).*

(ii) If $\mathrm{AA}(A,c)$ *holds with* $A \leq C$, *then* $U_{\theta/L}(g) \leq \varepsilon \mathrm{Var}(g)$ *for all* $\varepsilon > 0$, $L \geq (d/\varepsilon)^A/c$, $g \in \mathcal{B}_d$ *and* $0 < \theta \leq \mathrm{Var}(g)^C$; *so* $\mathrm{TT}(\varepsilon, L)$ *holds even without its restriction term.*

(iii) Hence $\mathrm{AA}(A,c)$ *with* $A \leq C$ *implies* $\mathrm{TT}(1/(16ed), (16ed^2)^A/c)$, *which implies the quasi-polynomial bound* $\mathrm{MaxInf}(f) \geq \theta_T(c/(16ed^2)^A)^T$. *Here* $(16ed^2)^A/c \geq 1$, *since a dictator forces* $c \leq 1$.

(iv) Any g *with* $U_{\theta/L}(g) > \varepsilon \mathrm{Var} g$ *and* $\theta \leq \mathrm{Var}(g)^C$, *in particular any counterexample to* $\mathrm{TT}(\varepsilon, L)$, *yields* $h \in \mathcal{B}_d$ *with* $\mathrm{MaxInf}(h)/\mathrm{Var}(h)^C < \varepsilon^{-C}/L$. *So, for* $\varepsilon \geq d^{-O(1)}$, *refuting* $\mathrm{TT}(\varepsilon, L)$ *with* $L = d^{\omega(1)}$ *along infinitely many* d *would refute* $\mathrm{AA}(A,c)$ *for every* $A \leq C$ *and* $c > 0$. *Without a lower bound on* ε *the factor* ε^{-C} *can absorb* L.

Proof. (ii) If $U_{\theta/L}(g) > \varepsilon \mathrm{Var} g$ and h is as in Lemma 8.6, then $c(\varepsilon \mathrm{Var}(g)/d)^A < c(\mathrm{Var} h/d)^A \leq \mathrm{MaxInf}(h) < \theta/L \leq \mathrm{Var}(g)^A/L$, using $\mathrm{Var} g \leq 1$ and $A \leq C$; so $L < (d/\varepsilon)^A/c$. (iii) is (ii) with $\varepsilon = 1/(16ed)$, combined with (i). (iv) is Lemma 8.6, together with $\mathrm{Var}(h)^C \leq \mathrm{Var}(h)^A$. □

$\mathrm{AA}(A,c)$ fails for every $A < 2$ [*proved here*]: the average of M copies of the 0/1-valued AND_m on disjoint variables, scaled by $1/M$, has $\mathrm{MaxInf} = 2^{-m-1}/M^2$ and $\mathrm{Var} = 2^{-m}(1-2^{-m})/M$; let $M \to \infty$ with m fixed.

A degree-free sufficient condition. For $g\colon Q_n \to [-1,1]$, $\theta > 0$ and $L \geq 1$ let

$$\Delta(g;\theta,L) = U_{\theta/L}(g) - 2\,\mathbb{E}_\sigma U_\theta(g_\sigma),$$

so that the inequality of Definition 8.2 holds for g at θ if and only if $\Delta(g;\theta,L) \leq \varepsilon \mathrm{Var} g$. Lemmas 8.1, 8.6 and I.9 and their proofs hold for every such g, with h bounded in place of $h \in \mathcal{B}_d$; in particular $\Delta(g;\theta,L) \leq 2\mathbb{E}_\sigma V_\sigma$ by Lemma I.9. Let rbs be the real block sensitivity (Definition 7.6).

Conjecture 8.8 (block-sensitivity transfer; [*open*])**.** *There are absolute constants* $A \geq 1$ *and* $c \in (0,1]$ *such that for all* n, *all* $g\colon Q_n \to [-1,1]$ *with* $\mathrm{Var} g > 0$, *all* $L \geq 1$ *and all* $0 < \theta \leq \mathrm{Var}(g)^C$,

$$L \cdot \Delta(g;\theta,L) \;\leq\; A\,\mathrm{rbs}(g)^{1/c}.$$

Here C is the cap exponent. Equivalently, a failure of the inequality of Definition 8.2 for g at an admissible θ forces $\mathrm{rbs}(g) \geq (\varepsilon L \mathrm{Var}g/A)^c$: many disjoint sensitive blocks at one point. The chain needs only the weaker statement (W_K): for some $A \geq 1$, all such g, all $\varepsilon \in (0,1]$ and all $0 < \theta \leq \mathrm{Var}(g)^C$, the inequality of Definition 8.2 holds for every $L \geq A(\mathrm{rbs}(g)/(\varepsilon \mathrm{Var}g))^K$; as $\varepsilon \mathrm{Var}g \leq 1$, Conjecture 8.8 implies ($\mathrm{W}_{1/c}$).

Proposition 8.9 ([*proved here*])**.** *Let $f \in \mathcal{B}_d$, $d \geq 2$, $v = \mathrm{Var}f > 0$, and let T and θ_T be as in Theorem 8.4.* (a) *Under Conjecture 8.8,* TT *holds at each $g \in \mathcal{B}_d$ with $\mathrm{Var}g > 0$ for all $L \geq Ad^{2/c}/(\varepsilon \mathrm{Var}g)$, which at $\varepsilon = 1/(16ed)$ is the loss $16eAd^{1+2/c}/\mathrm{Var}g = (d/\mathrm{Var}g)^{1+2/c+o(1)}$, and*

$$\mathrm{MaxInf}(f) \;\geq\; \theta_T \Big(\frac{v}{2048 e^3 A\, d^{3+2/c}} \Big)^T = \Big(\frac{\mathrm{Var}f}{d} \Big)^{O(C + (1+2/c)\log d)}.$$

(b) *Under* (W_K), $\mathrm{MaxInf}(f) \geq \theta_T A^{-T} (v/(2048 e^3 d^5))^{KT} = (\mathrm{Var}f/d)^{O(C + K \log d)}$.

Proof. For $g \in \mathcal{B}_d$, $\mathrm{rbs}(g) \leq d^2$ by Lemma 7.7(a), so the conjecture gives $\Delta(g;\theta,L) \leq Ad^{2/c}/L \leq \varepsilon \mathrm{Var}g$ for $L \geq L_g := Ad^{2/c}/(\varepsilon \mathrm{Var}g)$ and $\theta \leq \mathrm{Var}(g)^C$; under ($\mathrm{W}_K$) the same holds with $L_g := A(d^2/(\varepsilon \mathrm{Var}g))^K$. Let ε, η, p and θ_t be as in the proof of Theorem 8.4, let $\delta = 1/(128 e^2 d^2)$, and use there the single loss $L = 2048 e^3 A d^{3+2/c}/v$ in (a) and $L = A(2048 e^3 d^5/v)^K$ in (b), so that $L \geq L_g$ whenever $\mathrm{Var}g \geq \delta v$. In step t, if $\mathrm{Var}g \geq \delta v$ the step is unchanged; if $\mathrm{Var}g < \delta v$ then $U_{\theta_{t-1}}(g) \leq \mathrm{Var}g < \delta v$. Hence $\psi_{t-1} \leq 2\psi_t + \varepsilon + \theta_T^{1/C}/(pv) + \delta/p$, and, as $2^T < 4ed$,

$$\psi_0 \;\leq\; 2^T \big(\eta + \varepsilon + \theta_T^{1/C}/(pv) \big) + (2^T - 1) 2^T \delta \;<\; \tfrac{3}{4} + 16 e^2 d^2 \delta = \tfrac{7}{8}.$$

So $J_{\theta_0}(f) \neq \emptyset$ and $\mathrm{MaxInf}(f) \geq \theta_0 = \theta_T L^{-T}$. The last forms follow from $\theta_T \geq (v/(c_2 d^{8.76}))^{\max(4,C)}$ and $T < \log_2(4ed)$. □

The following degree-free bound proves Conjecture 8.8 whenever $M_2(g) \leq \mathrm{poly}(\mathrm{rbs}(g))$.

Lemma 8.10 ([*proved here*])**.** *Let $g\colon Q_n \to [-1,1]$, $\theta > 0$, $L \geq 1$, $J' = J_{\theta/L}(g)$ and $M_2(g) = \sum_j \|d_j g\|_\infty^2$. Then*

$$L \cdot \Delta(g;\theta,L) \;\leq\; \frac{1}{2} \sum_{j \notin J'} \big(\mathrm{Inf}_j(g) + \|d_j g\|_\infty^2 \big) \;\leq\; M_2(g).$$

Proof. With the notation of Lemma I.9, $V_\sigma \leq \sum_{j \in K_\sigma} \mathrm{Inf}_j(G_\sigma) \leq \sum_{j \in K_\sigma} X_j$ with $X_j = \mathrm{Inf}_j(g_\sigma)$, since G_σ averages g_σ over coordinates other than j. As $\mathbf{1}[X_j \geq \theta] \leq X_j/\theta$ and j is alive with probability $\frac{1}{2}$, $\Delta \leq 2\mathbb{E}_\sigma V_\sigma \leq \theta^{-1} \sum_{j \notin J'} \mathbb{E}[X_j^2 \mid j \text{ alive}]$. Given that j is alive, $X_j = \mathbb{E}[\psi_j \mid x_F]$ with $\psi_j = (d_j g)^2$ and F the fixed set, which contains each $i \neq j$ independently with probability $\frac{1}{2}$; so $\mathbb{E}[X_j^2 \mid j \text{ alive}] = \sum_S 2^{-|S|} \hat{\psi}_j(S)^2 \leq \mathrm{Inf}_j(g)^2 + \frac{1}{2}(\mathbb{E}\psi_j^2 - \mathrm{Inf}_j(g)^2) \leq \frac{1}{2}\mathrm{Inf}_j(g)(\mathrm{Inf}_j(g) + \|d_j g\|_\infty^2)$. Finally $\mathrm{Inf}_j(g) < \theta/L$ for $j \notin J'$, and $\mathrm{Inf}_j(g) \leq \|d_j g\|_\infty^2$. □

For Tribes with $s \geq w$, $M_2 = w\,\mathrm{rbs} \leq \mathrm{rbs}^2$, but for the address function $M_2 \geq 2^k$ while $\mathrm{rbs} = k+1$.

Remark 8.11 (relation to real Nisan; [*unaudited*]). Conjecture 7.9 implies (W_{C_0}), with c_0, C_0 the constants of bsAA (Remark 7.10). Indeed, if the inequality of Definition 8.2 fails for g at (ε,θ,L) with $\theta \leq \mathrm{Var}(g)^C$, then h of Lemma 8.6 has $\mathrm{Var}h = U_{\theta/L}(g) > \varepsilon \mathrm{Var}g$, $\mathrm{MaxInf}(h) < \theta/L \leq 1/L$ and $\mathrm{rbs}(h) \leq \mathrm{rbs}(g)$ (h is an average of restrictions of g, and rbs is convex), so bsAA for h gives $L < c_0^{-1}(\mathrm{rbs}(g)/(\varepsilon \mathrm{Var}g))^{C_0}$. For $L\Delta$ the argument gives only $c_0^{-1/C_0} \mathrm{rbs}(g) L^{1-1/C_0}$, and $C_0 \geq 2$ is necessary (the averages of AND_m above have $\mathrm{rbs} = m/2$); whether Conjecture 7.9 implies Conjecture 8.8 is open.

Remark 8.12 (special cases and numerics; [*unaudited*]). (i) *Boolean* $g\colon Q_n \to \{-1,1\}$: $L\Delta \le L\,U_{\theta/L}(g) \le \theta D(g) \le \mathrm{rbs}(g)^3$ for $\theta \le 1$, i.e. the conjecture with $c = \frac{1}{3}$, $A = 1$: for h as in Lemma 8.6 and an optimal tree for g, Theorem 4.1(L1) gives $\mathrm{Var}h = \mathrm{Cov}(g,h) \le \sum_{i\notin J'} \delta_i \mathbb{E}|d_i h| < \theta D(g)/L$, as $\mathbb{E}|d_i h| \le \mathbb{E}|d_i g| = \mathrm{Inf}_i(g)$; then use Lemma 5.7. (ii) *AND-block compositions* $g = P(\sum_{i\le s} \mathrm{AND}_{B_i})$, with disjoint blocks of arbitrary widths and $|P| \le 1$ on $\{0,\dots,s\}$, without θ/L-spikes: $L\Delta \le 8\gamma\,\mathrm{rbs}(g) \le 16\,\mathrm{rbs}(g)$ with $\gamma = \max_k |P(k+1) - P(k)|$, so $c = 1$. (iii) *Product selectors* $F = \sum_{b\in\{-1,1\}^k} y_b \prod_r \frac{1}{2}(1 + b_r s_r(z_{B_r}))$ ($|s_r| \le 1$, disjoint blocks, $\deg F = d$) without θ/L-spikes and with $\theta \le \mathrm{Var}(F)^C$: by hypercontractivity at $p = \frac{3}{2}$, the coordinates y_b contribute at most $e^{1/2}\mathrm{Var}F$ to the bound $2L\mathbb{E}_\sigma V_\sigma$ on $L\Delta$ once $L \ge 1.7\cdot 10^4 d^5$; spikes among the z-coordinates are not covered. We have proofs of (ii) and (iii), not included here. [*numerical*] Exact computation of $\sup_{L,\theta} L\Delta$ over all 3^n restrictions, with cap 2 and with no cap, for 48 functions with $n \le 12$, and about 70,000 adversarial evaluations, give $L\Delta \le 0.4\,\mathrm{rbs}(g)$, the worst case being Tribes with $w = s = 3$; at these sizes TT fails only for $L \le 4$, so they cannot detect c. Monte Carlo for Tribes with $8 \le w \le 28$ (failing losses up to $7.7\cdot 10^3$) gives lower bounds for $\sup L\Delta$ that grow roughly like $\mathrm{rbs}(g)^{0.4}$, consistent with $c = 1$.

Appendix I shows that TT fails for losses $L \le d^{1-o(1)}$, and with cap $C \ge 2$ for $L \le d^{2C/(2C-1)-o(1)}$ (Theorem I.4, Corollaries I.3 and I.6), and that a counterexample needs the new spikes to touch a definite part of the mass (Lemma I.9).

Reduction to old-free functions. The old spikes can be removed at a polynomial cost (Lemma J.20 and Corollary J.21, [*proved here, audited*], with proofs in Appendix J). For $0 < \varepsilon \le 1$, $\theta' = \varepsilon^2\theta/(36d)$ and $L' = \varepsilon^2 L/(36d)$, and with $h = \mathbb{E}[g \mid x_{J^c}]$, $J = J_{\theta/L}(g)$, as in Lemma 8.6,

$$\Delta(g;\theta,L) \le \Delta(h;\theta',L') + \tfrac{\varepsilon}{2}\mathrm{Var}g,$$

and $\mathrm{MaxInf}(h) < \theta/L = \theta'/L'$, so h has no old spikes at (θ',L'). Hence TT at loss L' and parameter $\varepsilon/2$ for all old-free $h \in \mathcal{B}_d$ with the cap $\theta' \le \mathrm{Var}(h)^2/(36d)$ implies TT at loss $36dL'/\varepsilon^2$, which is $9216e^2d^3L'$ for $\varepsilon = 1/(16ed)$, under cap 2 for all $g \in \mathcal{B}_d$. For old-free h, $G_\sigma = h_\sigma$ in Lemma I.9, and the transfer becomes a bound on the mass of the new spikes. Section 9 studies this mass, Section 10 bounds it at a loss of θ-exponent $1/(2T)$ with a quasi-polynomial prefactor, and Section 11 studies the spike mass of a single restriction at rate $1/d$.

9. Spikes after restriction: priming, a collapse, and the remaining mass bound

Lemma I.9 reduces the touching transfer to one quantity: the variance touched by *new* spikes, that is, by coordinates whose influence crosses the threshold θ only after the rate-$\frac{1}{2}$ restriction. This section collects what is known about that quantity. Section 9.1 fixes notation and defines the spiked mass. Sections 9.2 and 9.3 give the priming tools and show that every *capture* formulation collapses: a captured spike carries little mass by a direct argument (Lemma 9.8), so capture hypotheses are either false or restate the mass bound. Section 9.4 gives the amplification theorem for size-biased laws. Section 9.5 states the remaining target, a mass bound for old-free capped functions at rate $\frac{1}{2}$, and Sections 9.6–9.7 give the proved pieces, the open statement, and the reason the available tools fall short. Section 10 proves the target with a quasi-polynomial prefactor, which gives Theorem 10.1. Section 9.8 comments on the one-shot restriction.

Two rules are followed throughout. Every hypothesis is stated in *mass form*, as a bound on a part of the spiked mass, and never relative to the spiked mass itself. Every hypothesis is compared with the mass bound SM, and we record whether it is equivalent to it or weaker. No hypothesis of capture type is presented as a reduction of the conjecture.

9.1. **Old and new spikes, spiked mass.** Fix $g \in \mathcal{B}_d$, $\theta > 0$ and $L \geq 1$, and write $I_j = \mathrm{Inf}_j(g) = \|d_j g\|_2^2$. As in Definition 8.2, $\sigma = (F, z)$ is the rate-$\frac{1}{2}$ restriction: each coordinate is fixed independently with probability $\frac{1}{2}$ (the fixed set F, with uniform values z) and is otherwise alive. The *old spikes* are $J' = J_{\theta/L}(g)$, so $I_j < \theta/L$ for $j \notin J'$. As in Lemma I.9, $G_\sigma = \mathbb{E}[g_\sigma \mid x_{A \setminus J'}]$, where A is the alive set, and the *new spikes* are $K_\sigma = J_\theta(g_\sigma) \setminus J'$: alive coordinates outside J' whose influence is multiplied by more than L under one restriction.

Definition 9.1 (spiked mass)**.** The spiked mass of j is $\mu_j = \mathbb{E}_\sigma[\mathrm{Inf}_j(G_\sigma)\mathbf{1}[j \in K_\sigma]]$. For $\varepsilon > 0$, $\mathrm{SM}(\varepsilon)$ at (g, θ, L) is the statement $\sum_j \mu_j \leq \frac{\varepsilon}{2}\mathrm{Var}(g)$.

Since G_σ averages g_σ and j is alive with probability $\frac{1}{2}$, $\mu_j \leq \frac{1}{2} I_j$. In Lemma I.9, $V_\sigma \leq \sum_{j \in K_\sigma} \mathrm{Inf}_j(G_\sigma)$, so $\mathbb{E}_\sigma V_\sigma \leq \sum_j \mu_j$ and $\Delta(g; \theta, L) \leq 2\mathbb{E}_\sigma V_\sigma \leq 2\sum_j \mu_j$. Hence:

Corollary 9.2 (SM implies TT; [*proved here, audited*])**.** *If* $\mathrm{SM}(\varepsilon)$ *holds at* (g, θ, L)*, then the inequality of Definition 8.2 holds for* g *at* θ*.*

Old-free functions and the size-biased law. By Lemma 8.6 the old spikes can be averaged out: $h = \mathbb{E}[g \mid x_{[n]\setminus J'}]$ satisfies $\mathrm{MaxInf}(h) < \theta/L$. For such h there are no old spikes, $G_\sigma = h_\sigma$, and with $\phi_j := d_j h$, $I_j = \mathrm{Inf}_j(h)$ and $Q_j := \mathrm{Inf}_j(h_\sigma) = \mathbb{E}[\phi_j^2 \mid x_F]$ (for j alive),

$$\mu_j = \mathbb{E}_\sigma\big[Q_j \mathbf{1}[j \text{ alive}, Q_j \geq \theta]\big] = \tfrac{1}{2} I_j \, \widetilde{\Pr}_j(Q_j \geq \theta).$$

Here $\widetilde{\Pr}_j$ is the *size-biased law*: x has density ϕ_j^2/I_j with respect to the uniform measure, the coordinates $v \neq j$ carry i.i.d. clocks $t_v \sim U[0,1]$ independent of x, and $F = \{v : t_v \leq \frac{1}{2}\}$ is revealed in increasing clock order. The identity holds for every event measurable with respect to (F, x_F) and the clocks, since $\mathbb{E}_z[\mathbb{E}[\phi_j^2 \mid x_F = z]\mathbf{1}_E] = I_j \widetilde{\Pr}_j(E)$. We write $\ell_j = \ln(\theta/I_j) > \ln L$. For a partial assignment ρ of the coordinates other than j let $N(\rho) = \mathbb{E}[\phi_j^2 \mid \rho]$ and let $\mathrm{RI}_u(\rho) = \mathrm{Inf}_u(\phi_{j,\rho})/N(\rho) \in [0,1]$ be the *relative influence* of u at ρ. When u is revealed at ρ, N is multiplied by $1 + c_u x_u$, with the *bias* $c_u = \mathbb{E}[d_u(\phi_j^2) \mid \rho]/N(\rho)$; by Cauchy–Schwarz $c_u^2 \leq 4\,\mathrm{RI}_u(\rho)$.

9.2. **Priming through block sensitivity.** For $g\colon Q_n \to [-1,1]$ we use the real block sensitivity $\mathrm{rbs}(g)$ of Definition 7.6. Overwriting a block is flipping a block, so $2\,\mathrm{rbs}(g) = \max_z \max \sum_i |g(z^{\oplus B_i}) - g(z)|$, the inner maximum over families of pairwise disjoint nonempty blocks B_i, where $z^{\oplus B}$ flips the coordinates in B.

Lemma 9.3 ([*known*]; this is Lemma 7.7(a))**.** $\mathrm{rbs}(g) \leq d^2$ *for* $g \in \mathcal{B}_d$*.*

Second proof. Fix z and the blocks, and put $\tilde{h}(y) = g(x)$ with $x_u = z_u y_b$ for $u \in B_b$. Then $\tilde{h} \in \mathcal{B}_d$ on Q_k and $|g(z^{\oplus B_b}) - g(z)| = 2|d_b\tilde{h}(\mathbf{1})|$, so the claim is $\sum_b |d_b\tilde{h}(\mathbf{1})| \leq d^2$, which is [19, Theorem 3.3]. (Lemma 2.2, or Markov's inequality on the two lines through $\mathbf{1}$ along the sign classes of $d_b\tilde{h}(\mathbf{1})$, gives only $4d^2$, hence $\mathrm{rbs}(g) \leq 4d^2$, without the citation.) □

Selection with relative thresholds. For $t \geq 1$ let $A_j = \{u \notin J' \cup \{j\} : \mathrm{Inf}_u(d_j g) \geq I_j/t\}$ and $K = (d-1)t + 1$.

Lemma 9.4 (weighted Helly selection; [*proved here, audited*])**.** (a) $|A_j| \leq (d-1)t$. (b) *For* $u \notin J'$, $\sum_{j\colon u \in A_j} I_j \leq (d-1)t\,\mathrm{Inf}_u(g) < (d-1)t\theta/L$. (c) *Every* $C \subseteq [n] \setminus J'$ *contains* $m > 2L\mu(C)/(K^2\theta)$ *indices whose sets* A_j *are pairwise disjoint, where* $\mu(C) = \sum_{j \in C} \mu_j$*.*

Proof. $\sum_{u \neq j} \mathrm{Inf}_u(d_j g) = \sum_{S \ni j}(|S| - 1)\hat{g}(S)^2 \leq (d-1)I_j$ gives (a), and the same identity with the roles of u and j exchanged gives (b), since $u \in A_j$ means $I_j \leq t\mathrm{Inf}_u(d_j g)$. For (c), pick greedily; a pick j removes j and every k with $A_k \cap A_j \neq \emptyset$, of total weight at most $\mu_j + \sum_{u \in A_j} \sum_{k\colon u \in A_k} \mu_k < K^2\theta/(2L)$ by (a), (b), $\mu_k \leq I_k/2$ and $I_j < \theta/L$. □

For j and $A \not\ni j$ let $\Gamma_{j,A} = \mathbb{E}[d_j g \mid x_A]$, and for $a \in \{-1,1\}^A$ let

$$q_j(A,a) = \mathbb{E}[(d_j g)^2 \mid x_A = a], \qquad \pi_j(A,a) = \mathbb{E}_{a'} \big\| d_j g(a,\cdot) - d_j g(a',\cdot) \big\|_1.$$

Lemma 9.5 (priming; [*proved here, audited*])**.** *Let $A_1, \dots, A_m$ be pairwise disjoint, $j_i \notin A_i$, and $a_i \in \{-1,1\}^{A_i}$. Then* (P) $2\,\mathrm{rbs}(g) \geq \sum_i \big(|\Gamma_{j_i,A_i}(a_i)| - I_{j_i}^{1/2}\big)$*, and* (P*) $2\,\mathrm{rbs}(g) \geq \sum_i \pi_{j_i}(A_i,a_i)$*.* (Conversion) *If $q_j(A,a) \geq 4I_j$, then $\pi_j(A,a) \geq q_j(A,a)/8$.*

Proof. Let $B_i(z) = \{u \in A_i : z_u \neq a_{i,u}\}$; these blocks are disjoint and $z^{\oplus B_i(z)} = (a_i, z_{A_i^c})$. So $2\,\mathrm{rbs}(g) \geq \mathbb{E}_z \sum_i |g(z^{\oplus B_i(z)}) - g(z)|$. Since $B_i(z)$ does not depend on z_{j_i}, averaging over z_{j_i} and the switch identity $d_j g(z^{\oplus B}) - d_j g(z) = \frac{1}{2}[\Delta_B g(z^{j\to 1}) - \Delta_B g(z^{j\to -1})]$, with $\Delta_B g(z) = g(z^{\oplus B}) - g(z)$, give $\mathbb{E}_{z_{j_i}} |\Delta_{B_i(z)} g(z)| \geq |d_{j_i} g(a_i, z_{A_i^c}) - d_{j_i} g(z)|$. Taking $\mathbb{E}_z$ gives (P*); Jensen and Cauchy–Schwarz give (P). For the conversion, write $\phi_a = d_j g(a,\cdot)$, $\bar\phi = \mathbb{E}_{a'}\phi_{a'}$, $q = q_j(A,a)$. Since $|\phi_a - \phi_{a'}| \leq 2$, $\|\phi_a - \phi_{a'}\|_1 \geq \frac{1}{2}\|\phi_a - \phi_{a'}\|_2^2$, and $\mathbb{E}_{a'}\|\phi_a - \phi_{a'}\|_2^2 = q - 2\langle \phi_a, \bar\phi\rangle + I_j \geq (\sqrt{q} - \sqrt{I_j})^2 \geq q/4$. □

Random priming. The deterministic sets A_j can be replaced by laws. For each $j \notin J'$ let ν_j be a law on pairs (A,a) with $A \subseteq [n] \setminus (J' \cup \{j\})$ and $|A| \leq s$.

Lemma 9.6 (random priming, subset form; [*proved here, audited*])**.** *Suppose that for a set N' of non-old coordinates* (P1) $\sum_{j\in N'} I_j \mathbb{E}_{\nu_j} \pi_j \geq \gamma\theta \sum_{j\in N'} \mu_j$*, and* (P2) $\sum_{j\in N'} I_j \nu_j(u \in A) \leq \lambda \mathrm{Inf}_u(g)$ *for every non-old u. Then $\sum_{j\in N'} \mu_j \leq 4(\lambda+1)(s+1)\,2\,\mathrm{rbs}(g)/(\gamma L)$.*

Proof. Draw $(A_j, a_j) \sim \nu_j$ independently, and select each $j \in N'$ independently with probability $p_j = cLI_j/\theta \leq c$, where $c = 1/(2(\lambda+1)(s+1))$. Delete every selected j whose block $\{j\} \cup A_j$ meets the block of another selected index. For fixed j and draw, a fixed u of the block lies in another selected block with probability at most $\sum_{j'} p_{j'}(\nu_{j'}(u \in A) + \mathbf{1}[j' = u]) < c(\lambda+1)$ by (P2) and $\mathrm{Inf}_u(g) < \theta/L$, so j is deleted with probability at most $\frac{1}{2}$. Hence the survivors, whose blocks are disjoint, satisfy $\mathbb{E}\sum \pi_j \geq \frac{1}{2}\sum_j p_j \mathbb{E}_{\nu_j}\pi_j \geq \frac{1}{2}cL\gamma \sum_{N'} \mu_j$, and Lemma 9.5(P*) applied to one realization gives the claim. □

The subset form is randomized capture in mass form: it bounds the mass of *any* family of spikes that is primed in the sense of (P1) with polynomial loads. By the conversion, (P1) holds when $q_j(A,a) \geq \max(4I_j, \kappa\theta)$ on the relevant event, with $\gamma = \kappa/8$. We also use an averaged form.

Lemma 9.7 (averaged priming; [*proved here, audited*])**.** *Let ν_j be sub-probability laws with $\pi_j \geq \pi_*$ ν_j-almost surely, $W = \sum_j I_j \nu_j(1)$ and $\Lambda_u = \sum_j I_j \nu_j(u \in \{j\} \cup A)$. Then $2\,\mathrm{rbs}(g) \geq \pi_* \min\{W^2/(4\sum_u \Lambda_u^2),\, W/(2\max_j I_j)\}$.*

Proof. Fix $\eta \leq 1/\max_j I_j$. Independently for each j, select j with probability $\eta I_j \nu_j(1) \leq 1$ and draw (A_j, a_j) from $\nu_j/\nu_j(1)$; the number X of selected indices has $\mathbb{E}X = \eta W$. Let Y be the number of selected j whose block $\{j\} \cup A_j$ meets the block of another selected index. By independence,

$$\mathbb{E}Y \leq \sum_u \sum_{j \neq j'} \Pr(j \text{ selected}, u \in \{j\} \cup A_j) \Pr(j' \text{ selected}, u \in \{j'\} \cup A_{j'}) \leq \sum_u (\eta \Lambda_u)^2.$$

The other $X - Y$ blocks are pairwise disjoint, so Lemma 9.5(P*) gives $2\,\mathrm{rbs}(g) \geq \pi_* \mathbb{E}(X - Y) \geq \pi_*(\eta W - \eta^2 \sum_u \Lambda_u^2)$. Take $\eta = W/(2\sum_u \Lambda_u^2)$ if this is at most $1/\max_j I_j$, and $\eta = 1/\max_j I_j$ otherwise; in the second case $\eta^2 \sum_u \Lambda_u^2 < \eta W/2$. □

9.3. **Lemma CM and the collapse of capture.** The rate-$\frac{1}{2}$ route was first organized around a capture hypothesis. Let $\beta_j = \max_a |\Gamma_{j,A_j}(a)|$ with A_j as in Lemma 9.4, and $C_\beta = \{j \notin J' : \beta_j \ge \beta\sqrt{\theta}\}$ (the *captured* spikes). Capture(β, t) at (g, θ, L) is the statement $\mu(C_\beta) \ge \frac{1}{2}\sum_j \mu_j$. An independent check showed that the captured mass is small unconditionally.

Lemma 9.8 (captured mass; [*proved here, audited*])**.** *For every $g \in \mathcal{B}_d$, $\theta > 0$, $t \ge 1$, $\beta > 0$ and $L \ge 4/\beta^2$,*

$$\mu(C_\beta) \;\le\; 2\,\mathrm{rbs}(g)\,\frac{K^2\sqrt{\theta}}{L\beta} \;\le\; \frac{2d^2K^2\sqrt{\theta}}{L\beta}.$$

Proof. If $\mu(C_\beta) > 0$, Lemma 9.4(c) gives $m > 2L\mu(C_\beta)/(K^2\theta)$ indices in C_β with disjoint A_j. Lemma 9.5(P) with a_i maximizing $|\Gamma_{j_i,A_{j_i}}|$ gives $2\,\mathrm{rbs}(g) \ge m(\beta\sqrt{\theta} - \sqrt{\theta/L}) \ge m\beta\sqrt{\theta}/2 > L\beta\mu(C_\beta)/(K^2\sqrt{\theta})$. □

Theorem 9.9 (the collapse; [*proved here, audited*])**.** *Let $\varepsilon = 1/(16ed)$, $t \ge 1$, $\beta > 0$, and $L \ge L_0 := \max\{4/\beta^2,\, 8d^2K^2\sqrt{\theta}/(\varepsilon\beta\mathrm{Var}g)\}$.*

- *(a)* Capture(β, t) *at (g, θ, L) implies* SM(ε)*, hence the inequality of Definition 8.2 at (g, θ). This implication is exactly Lemma 9.8 combined with Corollary 9.2.*
- *(b)* Capture(β, t) *is false as a hypothesis: there are g and admissible (θ, L), at every configuration the chain of Theorem 8.4 visits, at which it fails while* SM(ε) *holds (Examples 9.10 and 9.11).*
- *(c) Every repair collapses: at $L \ge L_0$ the forms in Table 3 are false, equivalent to TT at the configuration, or equivalent to SM up to a factor 2 in ε, or (the old-free, L^1 and randomized forms) imply SM.*

Proof. (a) By Lemma 9.8, $\mu(C_\beta) \le \varepsilon\mathrm{Var}(g)/4$, so Capture gives $\sum_j \mu_j \le 2\mu(C_\beta) \le \frac{\varepsilon}{2}\mathrm{Var}g$. (b) is the content of the two examples. (c) For the minimal form, Capture and $\Delta > \varepsilon\mathrm{Var}g$ cannot hold together by (a), so "Capture wherever $\Delta > \varepsilon\mathrm{Var}g$" says that $\Delta \le \varepsilon\mathrm{Var}g$. The threshold form is SM by (a) and Lemma 9.8. The absolute form gives $\sum\mu \le \frac{\varepsilon}{4}\mathrm{Var}g + \mu(C_\beta) \le \frac{\varepsilon}{2}\mathrm{Var}g$, and follows from SM$(\varepsilon/2)$. The old-free form applies Lemma 9.8 with $J' = \emptyset$, and the L^1 and randomized forms imply SM in the same way, through Lemma 9.5(P*) and Lemma 9.6 in place of Lemma 9.8. □

Example 9.10 (address function)**.** Let $g(x, y) = y_x$ with k selector bits x and 2^k targets y_a, so $d = k + 1$ and $\mathrm{Var}g = 1$. If $2^{-k} < \theta/L \le \frac{1}{2}$, the selectors are old and the targets are new. Since $d_{y_a}g = \mathbf{1}[x = a]$ depends only on old coordinates, $A_{y_a} = \emptyset$ and $\beta_{y_a} = 2^{-k} < \beta\sqrt{\theta}$ for every admissible β, so $C_\beta = \emptyset$; but $\sum_a \mu_{y_a} = \frac{1}{2}2^{-k}\sum_{r:2^{-r}\ge\theta}\binom{k}{r}2^{-r} > 0$ (r the number of alive selectors). So Capture fails, while $\sum\mu \le \frac{1}{2}(3/4)^k$ and $\Delta \le U_{\theta/L}(g) = 2^{-k}$, so TT holds once $2^{-k} \le \varepsilon$, and SM once $(3/4)^k \le \varepsilon$. Both hold in the range $k \ge 427$ below. At the chain's first step the targets are new from $k \ge 427$ at $L = 4$ [*numerical*].

Example 9.11 (a degree-3 gadget)**.** Let $g = \frac{1}{2}x_0 + \frac{1}{2}x_1\psi(y)$ with $\psi = \frac{2}{M(M-1)}\sum_{v<w} y_v y_w$ on M bits; $\mathbb{E}\psi = 0$, $d = 3$. Then $I_0 = \frac{1}{4}$ (x_0 is old), $I_1 = \frac{1}{2M(M-1)}$, and $\mathrm{Inf}_{y_v}(d_1 g) = \frac{2}{M}I_1$, so $A_1 = \emptyset$ for $t < M/2$ and $\beta_1 = |\mathbb{E}\psi|/2 = 0$: x_1 is never captured, and $\beta_{y_v} = 0$ as well. Fixing all y to $+1$ gives $\mathrm{Inf}_{x_1}(g_\sigma) = \frac{1}{4}$, so $\mu_1 > 0$ for $\theta \le \frac{1}{4}$, while a fourth-moment bound and Hoeffding's inequality give $\mu_1 \le \frac{1}{4}\exp(-(\theta/8)^{1/2}M/2)$. For $\theta \le \frac{1}{4}$, any L, t and ambient $d \ge 3$, take $M \ge 10 + \max\{(L/(2\theta))^{1/2}, 2t, 6\ln(32ed)/\sqrt{\theta}\}$: then x_1 is a new spike-capable coordinate, $C_\beta = \emptyset$ for all $\beta > 0$, and $0 < \sum\mu \le \frac{1}{128ed} \le \frac{\varepsilon}{2}\mathrm{Var}g$. Capturing x_1 needs $t \ge M/2$, which pushes L_0 above L. The gadget separates a large averaged square, $\mathbb{E}[(d_j g)^2_\sigma] \ge \theta$, from a large squared average, $\mathbb{E}[d_j g \mid x_{A_j}]^2 \ge \beta^2\theta$.

form of the capture hypothesis, at $L \geq L_0$	status
literal: Capture at every chain configuration	false (Examples 9.10, 9.11)
minimal: Capture wherever $\Delta > \varepsilon \mathrm{Var} g$	equivalent to TT at the configuration
threshold: Capture wherever $\sum \mu > \frac{\varepsilon}{2}\mathrm{Var} g$	equivalent to SM(ε)
absolute: uncaptured mass $\leq \frac{\varepsilon}{4}\mathrm{Var} g$	between SM($\varepsilon/2$) and SM(ε)
old-free literal (after Lemma 8.6)	implies SM for old-free h; not refuted
L^1-capture; randomized capture (Lemma 9.6)	imply SM in the same way

TABLE 3. Capture-type hypotheses and what they are equivalent to, or imply.

What survives is Lemma 9.8 itself: at polynomial loss, every counterexample to SM, and hence to TT, carries its spiked mass on uncaptured spikes. The rest of the section works with mass bounds directly. Whether the conjecture implies SM at polynomial loss for functions with old spikes is open (for old-free functions it does, Remark 9.18); no counterexample to SM is known.

9.4. **Amplification under the size-biased law.** Two tools control how the conditional second moment grows along the clock order. Both are stated for a multilinear $\phi\colon Q_n \to [-1,1]$ of degree at most d, with $q = \mathbb{E}\phi^2 > 0$ and $\widetilde{\Pr} = (\phi^2/q)\cdot U$. Their proofs are in Appendix J.

Theorem 9.12 (spread-direction level-1 inequality, RL1; [*proved here, audited*])**.** *Let $0 < \tau \leq 1/(2e^3d^4)$ and $\kappa = 10e/(\ln(1/\tau) - 3 - 4\ln d)$. If $c_v = \mathbb{E}_{\widetilde{\Pr}}[x_v]$ are the biases of $\widetilde{\Pr}$, then*

$$\sum_{v\colon |c_v|<\tau} c_v^2 \;\leq\; \max\{\tau,\ \kappa \ln(e/q)\}.$$

So the small biases have total square at most a constant times $\ln(e/q)$, the log-inverse *absolute* height; this factor is the source of the $\ln(e/I)$ price below.

Lemma 9.13 (load lemma; [*proved here, audited*])**.** *Consider an adaptive exploration that reveals coordinates one at a time, the choice of the next coordinate depending on the revealed values and on randomness independent of x. For a coordinate u let τ_u be the step just before u is revealed. Then the relative-influence process RI_u is a $\widetilde{\Pr}$-martingale up to τ_u, $\mathbb{E}_{\widetilde{\Pr}}[\mathrm{RI}_u(\rho_\tau)] = \mathrm{Inf}_u(\phi)/q$ for every stopping time $\tau \leq \tau_u$, and $\widetilde{\Pr}(\sup_{t\leq\tau_u}\mathrm{RI}_u \geq \beta) \leq \mathrm{Inf}_u(\phi)/(\beta q)$. Consequently, for h as in Section 9.1 and any rule that puts u into A_j only when $\mathrm{RI}_u \geq \beta$ at some time before u is revealed,*

$$\widetilde{\Pr}_j(u \in A_j) \leq \frac{\mathrm{Inf}_u(d_j h)}{\beta I_j}, \qquad \sum_j I_j \widetilde{\Pr}_j(u \in A_j) \leq \frac{d-1}{\beta}\mathrm{Inf}_u(h), \qquad \mathbb{E}_{\widetilde{\Pr}_j}|A_j| \leq \frac{d-1}{\beta}.$$

The identity holds because $\mathrm{Inf}_u(\phi_\rho) = \mathbb{E}[(d_u\phi)^2 \mid \rho]$ is a uniform-measure martingale while u is unrevealed, and the density of $\widetilde{\Pr}$ on the revealed values is $N(\rho)/q$; the last bounds use $\sum_{j\neq u}\mathrm{Inf}_u(d_j h) \leq (d-1)\mathrm{Inf}_u(h)$, as in Lemma 9.4.

Fix a threshold $\tau_1 \in (0, 1/(2e^3d^4)]$ and put $\kappa_1 = 10e/(\ln(1/\tau_1) - 3 - 4\ln d)$. Along the clock order write $q_t = \mathbb{E}[\phi^2 \mid x_{S_t}]$ with S_t the revealed set at time t, Q for the final value, c_k for the bias of the k-th revealed coordinate just before it is revealed, and

$$\mathrm{QV}^{\lg} = \sum_k c_k^2 \mathbf{1}[|c_k| \geq \tau_1], \qquad \Lambda = \ln(e/q).$$

Theorem 9.14 (amplification; [*proved here, audited*])**.** *Let the alive set have rate $p \in (0,1)$, so that each coordinate is fixed with probability $1-p$. For every $y > 0$,*

$$\widetilde{\Pr}\Big(\ln \frac{Q}{q} \;\geq\; 2y + (e-1)\ln\tfrac{1}{p} \cdot \max\{\tau_1, \kappa_1(\Lambda + y)\} + \mathrm{QV}^{\mathrm{lg}} \Big) \;\leq\; 3e^{-y}.$$

The same holds conditionally on the stopping data (B, x_B) of any exploration run before the clocks, with ϕ replaced by $\phi(x_B, \cdot)$.

Proof sketch; the full proof is Theorem J.14. Write $\ln(Q/q) = \ln Z_m + \mathrm{QV}^{\mathrm{sm}} + \mathrm{QV}^{\mathrm{lg}}$ with $Z_k = (q_{s_k}/q)\exp(-\sum_{i\leq k} c_i^2)$. One reveal multiplies q by $1 + c_k x_{v_k}$ with $\widetilde{\Pr}$-mean $1 + c_k^2$, so Z is a supermartingale and $\widetilde{\Pr}(\ln Z_m \geq y) \leq e^{-y}$. The small-bias part $\mathrm{QV}^{\mathrm{sm}}$ is controlled at every time by Theorem 9.12 applied to the restricted function, $\sum c_v(t)^2 \leq \max\{\tau_1, \kappa_1(\Lambda - \ln(q_t/q))\}$ over unrevealed v with small bias, and the clocks reveal each unrevealed coordinate at hazard rate $\ln(1/p)$; an exponential-moment bound on the hazard-weighted sum, off the event $\{\min_t \ln(q_t/q) < -y\}$ of probability $\leq e^{-y}$, gives the middle term with probability $\geq 1 - e^{-y}$. □

No regularity of ϕ is assumed. The theorem says that a large amplification needs large-bias reveals, unless Λ, the logarithm of the inverse starting height, is large.

Theorem 9.15 (staged form; [*proved here, Appendix K; audited*])**.** *Suppose the revealed set is revealed in s stages, the stage of each coordinate being chosen independently of x and of the clocks, and in clock order within each stage.* (a) *Theorem 9.14 holds with $(e-1)\ln(1/p)$ replaced by $(e-1)\,s\ln(1/p)$: each stage adds one hazard $\ln(1/p)$.* (b) *The same holds conditionally on the revealed data at any stage boundary, with one hazard per remaining stage and q, Λ taken at that boundary; here alive coordinates that are never revealed (frozen, with hazard 0) are allowed.*

Outline (the full proof is in Appendix K). Run the proof of Theorem 9.14 stage by stage: within a stage the clock order is the order of Theorem 9.14, and the stage assignment is independent of x and of the clocks, so each stage contributes its own hazard $\ln(1/p)$; for (b), condition on the data at the boundary and run the remaining stages.

9.5. **The remaining target: SM for old-free capped functions.** Let $\varepsilon = 1/(16ed)$, $T = \lceil \log_2(2ed) \rceil$ and $V = \mathrm{Var}h$.

Definition 9.16 (old-free capped; SM-OF)**.** $h \in \mathcal{B}_d$ is *old-free capped at (θ, L)* if $\mathrm{MaxInf}(h) < \theta/L$ and $\theta \leq V^2/(36d)$. For a function $L_0 = L_0(\theta, d, V)$, SM-OF(L_0) is the statement: for every h old-free capped at (θ, L) with $L \geq L_0$,

$$\sum_j \mu_j \;=\; \sum_j \tfrac{1}{2} I_j \widetilde{\Pr}_j(Q_j \geq \theta) \;\leq\; \tfrac{\varepsilon}{4} V.$$

The cap $\theta \leq V^2/(36d)$ is what the old-free reduction delivers: averaging out the old spikes as in Lemma 8.6 and adjusting parameters (Lemma J.20 and Corollary J.21; [*proved here, audited*]) turns TT for g at (θ, L) into SM-OF for h at $(\theta', L') = (\varepsilon^2\theta/(36d), \varepsilon^2 L/(36d))$, with loss $9216e^2d^3$ and cap $C = 2$. A loss that depends on θ is admissible in the chain if its θ-exponent is small.

Proposition 9.17 (the consequence; [*conditional on SM-OF; the implication is proved here, audited*])**.** *Suppose* SM-OF(L_0) *holds with $L_0 = P\,\theta^{-c}$, where $P = (d/V)^{O(1)}$ and $cT \leq \frac{1}{2}$. Then every $f \in \mathcal{B}_d$ satisfies* $\mathrm{MaxInf}(f) \geq (\mathrm{Var}f/d)^{O(\log d)}$.

Proof. By Lemma OF and Corollary 9.2 (through Lemma I.9 with $G_\sigma = h_\sigma$), TT holds at every configuration of Theorem 8.4 with the loss $L(\theta) = P'\theta^{-c}$, $P' = \mathrm{poly}(d)P$. Run the chain with this loss: $\theta_{t-1} = \theta_t / L(\theta_t)$, so $\ln(1/\theta_0) = (1+c)^T \ln(1/\theta_T) + \ln P' \cdot ((1+c)^T - 1)/c \leq 1.65\ln(1/\theta_T) + 1.30\,T\ln P'$ when $cT \leq \frac{1}{2}$. Restrictions whose variance falls below $\mathrm{poly}(V/d)$

are dropped, at the cost of an additive error in each step, by the device used in the proof of Proposition 8.9. □

So a loss θ^{-c} per level is harmless only for $c \le 1/(2T)$; a fixed power of $1/\theta$ compounds to $\theta^{-(1+c)^T}$ and destroys the bound. Below, a bound is called *θ-free* if its loss is $(d/V)^{O(1)}\theta^{-c}$ with $cT \le \frac{1}{2}$, and *one-shot* if it is a fixed power of $1/\theta$. A loss $P_d V^{-1}\theta^{-c}$ with $cT \le \frac{1}{2}$ and $\ln P_d$ polylogarithmic in d is *θ-mild*; Section 10 proves SM-OF at such a loss (Theorem 10.26).

Remark 9.18 (AA implies SM-OF vacuously). If $\mathrm{MaxInf}(u) \ge (\mathrm{Var}u/d)^A$ for all $u \in \mathcal{B}_d$, then no h is old-free capped at (θ, L) with $L \ge (36d/V^2)(d/V)^A$, since $\mathrm{MaxInf}(h) < \theta/L \le V^2/(36dL)$. So SM-OF($L_0$) holds for such L_0. SM-OF is therefore sandwiched like TT, but it is not known to be equivalent to the conjecture. At the θ-mild loss of Theorem 10.26, SM-OF is proved. A function h old-free capped at that loss has $\mathrm{MaxInf}(h) < V^3/(36dP_d)$, while $\mathrm{MaxInf}(h) > V^2/(4C^d)$ by the DFKO bound of Section 1; so such h exist only if $\ln P_d < d \ln C$, and the theorem has content only for astronomically large d.

Example 9.19 ($L \ge L_0$ is necessary; [*sketch, numerical*])**.** Let $u_t = x_t y_{t,0}\,\mathrm{AND}_w(y_{t,1}, \dots, y_{t,w})$ on disjoint blocks, $t \le N$, $W = N^{-1}\sum_t u_t$ and $h = T_D(W)$, the Chebyshev polynomial of degree $D \approx \lambda\sqrt{N2^w}$, so $d = D(w+2)$. Then h is old-free capped for $L < c_\lambda d/((w+2)^2 2^w)$ with $c_\lambda \in [0.016, 0.028]$, and under $\widetilde{\Pr}_j$ the amplification of a spike coordinate is exactly $2^{\mathrm{Bin}(w,1/2)}$ up to a factor $1 - O(\delta^2)$. For $L \approx d^{1/3}$ ($d = 10^{40}$) the spiked mass is $\sum \mu_j \approx 0.12\,V$, while $\frac{\varepsilon}{4} \approx 5.8 \cdot 10^{-43}$; the largest such L is about $\sqrt{d}/\log d$. So SM-OF fails for $L \lesssim \sqrt{d}/\log d$, and $L_0 \gtrsim \sqrt{d}/\log d$. The spikes here are captured by single gates; they survive only because the captured-mass bounds need $L \ge \mathrm{poly}(d)/V$.

Lemma 9.20 (HC; [*proved here, audited*])**.** *For every $f\colon Q_n \to [0,1]$, $\mathrm{Stab}_{1/2}(f) \le (\mathbb{E}f)^{4/3}$. In particular $1 + X \le I^{-2/3}$ for every $\phi\colon Q_n \to [-1,1]$ with $I = \mathbb{E}\phi^2$, with no regularity or degree assumption.*

Proof. $\mathrm{Stab}_{1/2}(f) = \|T_\rho f\|_2^2$ with $\rho = 1/\sqrt{2}$, and hypercontractivity gives $\|T_\rho f\|_2 \le \|f\|_{1+\rho^2} = \|f\|_{3/2}$; since $f^{3/2} \le f$, $\|f\|_{3/2} \le (\mathbb{E}f)^{2/3}$. Apply this to $f = \phi^2$. □

Proposition 9.21 (baseline with a one-shot loss; [*proved here, audited*])**.** *For every $h \in \mathcal{B}_d$ with $\mathrm{MaxInf}(h) < \theta/L$, $\sum_j \mu_j \le dV/(2\theta^{2/3}L^{1/3})$. Hence* SM-OF *holds with $L_0 = (2d/\varepsilon)^3\theta^{-2}$, and $\sum_j \frac{I_j}{2}\widetilde{\Pr}_j(E_j) \le \frac{\varepsilon}{16}V$ for any events $E_j \subseteq \{Q_j \ge \theta\}$ once $L \ge (8d/\varepsilon)^3\theta^{-2}$.*

Proof. Given that j is alive, $Q_j = \mathbb{E}[\phi_j^2 \mid x_F]$ and $\mathbb{E}Q_j^2 = \mathrm{Stab}_{1/2}(\phi_j^2) \le I_j^{4/3}$ by Lemma 9.20. By Markov, $\widetilde{\Pr}_j(Q_j \ge \theta) = \mathbb{E}[Q_j \mathbf{1}[Q_j \ge \theta]]/I_j \le \mathbb{E}Q_j^2/(\theta I_j) \le I_j^{1/3}/\theta < \theta^{-2/3}L^{-1/3}$; sum with $\sum_j I_j \le dV$. □

At the top level of the chain, $\theta \approx V^2/(36d)$, this is a polynomial loss. The whole open content of SM-OF is therefore the requirement that the loss be θ-free. Section 10 meets the θ-exponent requirement $cT \le \frac{1}{2}$ at a θ-mild loss (Theorem 10.26); at a θ-free loss, with a polynomial prefactor, SM-OF is open.

Per-spike arguments. A *per-spike argument* at (θ, L, d) is a functional $\Psi(\phi, \theta) \ge 0$ of the joint law of x, the clocks and the revealed path under $\widetilde{\Pr}_\phi$, which is *sound*, $\Psi(\phi,\theta) \ge \frac{I_\phi}{2}\widetilde{\Pr}_\phi(Q_\phi \ge \theta)$ for every $\phi\colon Q_n \to [-1,1]$ of degree at most $d-1$, together with a deduction of $\sum_j \Psi(\phi_j, \theta) \le \frac{\varepsilon}{4}V$ from the constraints (C): the ϕ_j are such functions on a common cube, $I_j < \theta/L$, $\sum_j I_j \le dV$, $V \in (0,1]$ and $\theta \le V^2/(36d)$. Constraints that couple different j, such as the loads of Lemma 9.6 or the boundedness of one common h, are not in (C). Markov and moment bounds (Lemma 9.24, Proposition 9.21) and Theorem 9.14 or Lemma 9.26 applied to one spike are per-spike arguments.

Proposition 9.22 (per-spike barrier; [*proved here, audited*])**.** *Let $t = \log_2(1/\theta)$ with $1 \le \lfloor t \rfloor \le (d-1)/2$, and $\theta \le 1/(36d)$. If $L < \theta\, 4^{\lfloor t \rfloor}$, for instance if $L \le 1/(4\theta)$, then no per-spike argument proves SM-OF at (θ, L). Hence no per-spike argument gives SM-OF with a loss $P\theta^{-c}$, $c < 1$, for θ in the window $[2^{-(d-1)/2}, (4P)^{-1/(1-c)}]$; the window is nonempty when $\ln(4P)/(1-c) \lesssim \frac{1}{2}(d-1)\ln 2$, and it contains the chain's quasi-polynomial range $\ln(1/\theta) = O(\log d \cdot \log(d/V))$.*

Proof sketch. Take $V = 1$, $w = 2\lfloor t \rfloor \le d-1$, $n_0 = \lfloor d 2^w \rfloor$ disjoint blocks B_k of size w with patterns p_k, and $\phi_k = \mathbf{1}[x_{B_k} = p_k]$, so that $I_k = 2^{-w}$ and $d - \frac{1}{2} \le \sum_k I_k \le d$. Under $\widetilde{\Pr}_k$, $x_{B_k} = p_k$ almost surely, so $Q_k = 2^{-|B_k \setminus F|}$ and $\widetilde{\Pr}_k(Q_k \ge \theta) = \Pr(\mathrm{Bin}(w, \frac{1}{2}) \le \lfloor t \rfloor) \ge \frac{1}{2}$. The cap $I_k < \theta/L$ is $L < \theta 4^{\lfloor t \rfloor}$, and $4^{\lfloor t \rfloor} > \theta^{-2}/4$. So (C) holds, while every sound Ψ has $\sum_k \Psi_k \ge \sum_k \mu_k \ge \frac{1}{4}(d - \frac{1}{2}) > \frac{\varepsilon}{4}$. Finally $P\theta^{-c} \le 1/(4\theta)$ once $\theta \le (4P)^{-1/(1-c)}$. □

Wider blocks, $w = 2\lfloor t \rfloor + s$, raise the barrier to $1 + \Theta(\sqrt{\ln d / \ln(1/\theta)})$, a correction that vanishes as $\theta \to 0$ at fixed d; with Proposition 9.21, the best per-spike exponent lies in $[1, 2]$. The family need not come from one bounded h, and it is not a counterexample to SM-OF. In an actual old-free capped h, AND-type spikes have defect 0 for the greedy and final-context laws below (every revealed block coordinate keeps relative influence $\frac{1}{2}$), so Proposition 9.25 with $D_0 = 0$ bounds their mass by $\max\{64d^2/L,\ 8d^{5/2}\beta^{-1}(V/L)^{1/2}\}$: no single h at $L \ge \mathrm{poly}(d/V)$ realizes the family. A proof of an exponent $c < 1$ must therefore use cross-spike structure of h, such as the loads in Lemma 9.6; on the uncaptured part of the mass a per-spike bound is not excluded (Section 9.7).

9.6. **The proved pieces.** Fix h old-free capped at (θ, L), write $\widetilde{\Pr}_j$, Q_j, ℓ_j as in Section 9.1, and let $\tau_1 = e^{-345} d^{-129}$, $\gamma_1 = (e-1)\ln 2$, $c_1 = (\frac{3}{4} - \gamma_1\kappa_1)/(2 + \gamma_1\kappa_1)$ and $c(\tau_1) = \gamma_1\kappa_1/(\frac{3}{4} - \gamma_1\kappa_1)$. The spiked mass splits exactly into three pieces, according to events of the enlarged probability space:

$$\sum_j \mu_j = \sum_j \tfrac{I_j}{2}\Big[\widetilde{\Pr}_j(\mathrm{sp}_j \wedge \mathrm{SJ}_j) + \widetilde{\Pr}_j(\mathrm{sp}_j \wedge \mathrm{LJ}_j \wedge \mathrm{Cap}_j) + \widetilde{\Pr}_j(\mathrm{sp}_j \wedge \mathrm{LJ}_j \wedge \neg\mathrm{Cap}_j)\Big],$$

where $\mathrm{sp}_j = \{Q_j \ge \theta\}$, $\mathrm{SJ}_j = \{\mathrm{QV}_j^{\lg} < \ell_j/4\}$ (small jumps), LJ_j is its complement, and Cap_j is a capture event defined by priming laws. SM-OF follows if each piece is at most $\frac{\varepsilon}{16}V$.

Theorem 9.23 (piece 1: small-jump spikes; [*proved here, audited*])**.** (a) $\widetilde{\Pr}_j(\mathrm{sp}_j \wedge \mathrm{SJ}_j) \le 3e^{\gamma_1\tau_1/2}(I_j/\theta)^{c_1}(e/\theta)^{c(\tau_1)c_1}$. (b) *If $L \ge L_{\mathrm{SJ}} := (24e^{\gamma_1\tau_1/2}d/\varepsilon)^{1/c_1}(e/\theta)^{c(\tau_1)}$, then $\sum_j \frac{I_j}{2}\widetilde{\Pr}_j(\mathrm{sp}_j \wedge \mathrm{SJ}_j) \le \frac{\varepsilon}{16}V$.* (c) *$c_1 \ge 0.32$ and $c(\tau_1)T \le \frac{1}{2}$ for all $d \ge 2$, so the loss is θ-free [numerical; re-run independently; the supremum is about 0.498].*

Proof. (a) is Theorem 9.14 at $p = \frac{1}{2}$ with y chosen so that $2y + \gamma_1 \max\{\tau_1, \kappa_1(\Lambda + y)\} = \frac{3}{4}\ell_j$, where $\Lambda = \ln(e/I_j) = \ell_j + \ln(e/\theta)$; on $\mathrm{sp}_j \wedge \mathrm{SJ}_j$, $\ln(Q_j/I_j) \ge \ell_j$ and $\mathrm{QV}^{\lg} < \ell_j/4$. (b) Sum (a) with $I_j < \theta/L$ and $\sum I_j \le dV$. □

Neither the cap nor any regularity is used. If the event is defined by the total quadratic variation instead, the supermartingale Z alone gives $\widetilde{\Pr}_j(\mathrm{sp}_j \wedge \{\mathrm{QV}_j < \ell_j/4\}) \le (I_j/\theta)^{3/4}$, with no θ-loss at all; so piece 1 is elementary and all the difficulty sits in the large jumps.

χ^2 *bounds.* For $\phi = \phi_j$ let $1 + X = \mathrm{Stab}_{1/2}(\phi^2)/I^2 = \mathbb{E}_{\widetilde{\Pr}}[Q/I]$.

Lemma 9.24 ([*proved here, audited*])**.** (C1) $\widetilde{\Pr}_j(\mathrm{sp}_j) \le (1 + X_j)I_j/\theta < (1 + X_j)/L$, *so $\mu_j \le I_j(1+X_j)/(2L)$, and $\sum_j I_j X_j \le K_\chi V$ implies SM-OF at $L \ge 2(d + K_\chi)/\varepsilon$.* (C2) *If $\phi = \prod_i \phi_i$ on disjoint blocks, then $Q = \prod_i Q_i$, $\widetilde{\Pr} = \bigotimes_i \widetilde{\Pr}_i$ and $1 + X = \prod_i (1 + X_i)$.* (C3) *A factor $\frac{1}{2}(1+g)$ with $g \in \mathcal{B}_k$, $\mathrm{MaxInf}(g) \le \tau$ and $e_0 = \mathbb{E}(\frac{1}{2}(1+g))^2$ has $X \le (3k^2 3^R\sqrt{\tau} + \frac{5}{32}2^{-R})/e_0^2$ for every*

level $R \geq 1$. (C4) *A product of A such factors with $e_i \geq e_0$ has $1 + X \leq e$, hence $\widetilde{\Pr}(\mathrm{sp}) < 3/L$, provided $\tau \leq \tau_{C4} := (e_0^2/(6Ak^2 3^R))^2$ with $R = \lceil \log_2(A/e_0^2) \rceil + 1$; the threshold depends on A.* (C5) *If $\mathrm{MaxInf}(\phi) \leq I^{7.2}/(10^4 d^4)$, then $X \leq 1$ and $\widetilde{\Pr}(\mathrm{sp}) < 2/L$.*

Proof. (C1) is Markov: $\widetilde{\Pr}(Q \geq \theta) \leq \mathbb{E}_{\widetilde{\Pr}}[Q]/\theta$. (C2) is independence of the blocks under the uniform measure and the clocks. (C3) applies Theorem 3.2 to g and to g^2 ($\mathrm{Inf}(g^2) \leq 4\mathrm{Inf}(g)$): the levels $1..R$ of ϕ^2 carry at most $3k^2 3^R \sqrt{\tau}$, and the levels above R are damped by 2^{-R} in $\mathrm{Stab}_{1/2}$. (C4) multiplies (C3) by (C2), and (C5) is (C3) with R optimized for ϕ itself. □

The products of (C4) are the one-shot core of the conjecture (Remark 9.34); at rate $\frac{1}{2}$ they are not spikes. CHI, the hypothesis $\sum_j I_j X_j \leq K_\chi V$, is sufficient but not necessary for a single spike: for AND_m with $m = 4\log_2(1/\theta)$, $1 + X = \theta^{-2.34}$ while $\widetilde{\Pr}(\mathrm{sp}) \leq \theta^{0.7}$.

Piece 2: captured mass with θ-free laws. A *priming law* assigns to each j a set $A_j \subseteq F \setminus \{j\}$ that is a function of x and the clocks; its *load* is λ if $\Lambda_u := \sum_j I_j \widetilde{\Pr}_j(u \in A_j \cup \{j\}) \leq (\lambda + 1)\mathrm{Inf}_u(h)$ for all u. For $D_0 > 0$ let $\mathrm{Cap}_j = \{Q_j \geq \theta,\ \ln(Q_j/q_j(A_j, x_{A_j})) \leq D_0\}$.

Proposition 9.25 (capture by a priming law; [*proved here, audited*])**.** *If $L \geq 4e^{D_0}$, then*

$$\sum_j \tfrac{I_j}{2} \widetilde{\Pr}_j(\mathrm{Cap}_j) \ \leq \ \max\Big\{ \frac{64 d^2 e^{D_0}}{L},\ 8(\lambda+1) d^{3/2} \Big(\frac{V e^{D_0}}{L} \Big)^{1/2} \Big\}.$$

The loss is θ-free exactly when λ and D_0 are.

Proof. On Cap_j, $q_j(A_j) \geq \theta e^{-D_0} \geq 4I_j$, so $\pi_j \geq \theta e^{-D_0}/8$ by the conversion. Apply Lemma 9.7 to the laws of (A_j, x_{A_j}) on Cap_j, with $W = 2\sum_j \frac{I_j}{2} \widetilde{\Pr}_j(\mathrm{Cap}_j)$, $\sum_u \Lambda_u^2 \leq (\lambda+1)^2(\theta/L) d V$, $\max I_j < \theta/L$ and $2\,\mathrm{rbs}(h) \leq 8d^2$. □

The following laws have θ-free loads, by Lemma 9.13 or by a direct computation; the loads of the greedy-closure, final-context and pivotal laws are [*proved here, audited*], and the matched-block load and mass bound are a [*proof outline; audited*].

- *Greedy closure* at threshold r: reveal, in any order, every coordinate of F whose current relative influence is at least r, until none is left; $A_j = A^\infty$ is the revealed set. Load $(d-1)/r$ at every depth, one Doob bound per coordinate.
- *Final context* at threshold β: $A_j = \{u \in F : \mathrm{RI}_u(x_{F\setminus u}) \geq \beta\}$. The event is measurable with respect to the clocks and $x_{F\setminus u}$, so $\widetilde{\Pr}_j(u \in A_j) \leq \mathrm{Inf}_u(\phi_j)/(2\beta I_j)$, and the load is at most $(d-1)/\beta$.
- *Matched block*, for gate families $\phi = \sum_i w_i \prod_{v \in B_i} \frac{1}{2}(1 + x_v)$ with M spread gates of width w and $\nu = M 2^{-w} \leq \frac{1}{10}$: $A_j = B_{i^*} \cap F$ for the first gate fully fixed to 1. Load $3(d-1)$. The defect is not bounded pointwise, but $\mathbb{E}_{\widetilde{\Pr}} Q \leq 27/M^2$ and Markov's inequality show that it exceeds D_0 only on a $\widetilde{\Pr}$-fraction $27(1 + O(\nu))e^{-D_0}$ of the spikes. (P1) holds with $\gamma = \frac{2}{17}$, so Lemma 9.6 with $\lambda = 3(d-1)$ and $s \leq d-1$ bounds the spread-gate spiked mass by $816 d^4/L$.
- *Pivotal law* at threshold β: $A_j = F \cap \{u : (d_u \phi_j)^2/\mathbb{E}_{x_u} \phi_j^2 \geq \beta\}$. Since $(d_u \phi)^2$ and $\mathbb{E}_{x_u} \phi^2$ do not depend on x_u, $\widetilde{\Pr}_j(u \in A_j) \leq \mathbb{E}[\phi_j^2 (d_u \phi_j)^2/(\beta \mathbb{E}_{x_u} \phi_j^2)]/I_j = \mathrm{Inf}_u(\phi_j)/(\beta I_j)$, so the load is $(d-1)/\beta$.

With these laws, spread gates get θ-free bounds from the matched block (through Lemma 9.6) and from the final context (through Proposition 9.25; its defect exceeds D_0 only with negligible $\widetilde{\Pr}$-probability once $\beta \leq (V/d)^b$), [*proof outline; audited*]. The analogous bounds for parity gates (relative influence $\frac{1}{2}$ in every context), AND blocks and approximate-OR cascades (greedy closure), hard-address gates (pivotal law), and products of majorities (static heavy set, relative

influence $\geq 0.39/\sqrt{d}$) are [*proved here, unaudited*]; for the cascades they hold to leading order, unbalanced fan-ins are heuristic, and these proofs are not included here. A union of laws has load at most the sum of the loads, but it is not itself a capture law, since $q_j(\cdot)$ is not monotone in the set: capture is argued law by law, by partitioning according to the first capturing law.

The θ-exponent of each law. For the same families the static laws give only one-shot losses ([*proved here, unaudited*]): with $p_0 := \varepsilon/(16d)$, a static gate-set law needs $L \geq d^4/(p_0\theta)$ (exponent 1), and a static soft-address law needs $(p_0\theta)^{-2}$ (exponent 2). Lemma 9.20 alone has exponent 2 (Proposition 9.21), and the matched-block, final-context and pivotal laws have θ-free loads (with the tags above), hence exponent 0 wherever they capture; of the families above, the capture itself is audited only for spread gates. This is a statement about each law on its families; no combination of the laws is known to have exponent at most $1/(2T)$ on every family, and the soft-addressed, semi-hard and unbalanced cases remain heuristic or open.

Activation and the light remainder. Let A^∞ be the greedy closure at threshold r, $q^\infty = \mathbb{E}[\phi_j^2 \mid x_{A^\infty}]$, and $1 + X^\infty = \mathbb{E}[Q_j^2 \mid \mathcal{G}^\infty]/(q^\infty)^2$, the χ^2 ratio of the remaining stage given the closure's σ-field $\mathcal{G}^\infty$; given $\mathcal{G}^\infty$, the unrevealed coordinates have the size-biased law of the restricted function, which is r-regular in relative influence.

Lemma 9.26 (light stage; [*proved here, audited*])**.** *For $K_0 \geq 1$, $\widetilde{\Pr}_j(Q_j \geq \theta,\ Q_j \geq K_0 q^\infty) \leq \mathbb{E}_{\widetilde{\Pr}_j}\big[(1 + X^\infty)\min\{1/K_0,\ q^\infty/\theta\}\big]$.*

Proof. $\mathbb{E}_{\widetilde{\Pr}}[Q \mid \mathcal{G}^\infty] = \mathbb{E}_U[Q^2 \mid \mathcal{G}^\infty]/q^\infty = q^\infty(1 + X^\infty)$ by the tower property; apply Markov at the level $\max\{\theta, K_0 q^\infty\}$. □

Lemma 9.27 (activation; [*proof outline; audited*])**.** *For any adaptive exploration that reads x only on F (its σ-field is contained in $\sigma(x_F, \text{clocks}, \text{coins})$; the lemma is false without this hypothesis), any v, $\beta \in (0,1]$ and $y > 0$, let $1 + X_v = \mathrm{Stab}_{1/2}((d_v\phi)^2)/\|d_v\phi\|_2^4$. Then*

$$\widetilde{\Pr}\big(\mathrm{RI}_v \text{ reaches } \beta \text{ before } v \text{ is revealed, with } N \geq Ie^{-y} \text{ until then}\big) \ \leq\ (\mathrm{RI}_v(0)/\beta)^2(1 + X_v)e^y.$$

Proof sketch. Change measure to the size-biased law of $d_v\phi$, whose density on the revealed values is the uniform-measure martingale $M = \mathbb{E}[(d_v\phi)^2 \mid \cdot]/\|d_v\phi\|_2^2$; on the event, $M \geq \beta e^{-y}/\mathrm{RI}_v(0)$, and the maximal inequality for the submartingale M^2 bounds the probability by $\mathbb{E}M_\infty^2$ times the inverse level, where M_∞ is the terminal value, and $\mathbb{E}M_\infty^2 \leq 1 + X_v$ because the revealed σ-field is contained in $\sigma(x_F, \text{clocks}, \text{coins})$. □

Compared with Lemma 9.13, activation costs an extra factor $(\mathrm{RI}_v(0)/\beta)(1 + X_v)e^y$. If every light derivative has $1 + X_v \leq 3$ (for instance by (C4) or (C5)), then with $y = \ln(32d/\varepsilon)$ and $r \leq \varepsilon^2\beta^2/(3100d^3)$, activation chains of every depth carry $\widetilde{\Pr}$-mass at most $\varepsilon/(16d)$, by summing Lemma 9.27 over the chains depth by depth [*proof outline; audited*].

Proposition 9.28 (SM-OF from a light χ^2 bound; [*conditional on CHI-L; proof outline, unaudited*])**.** *Let $r \in (0,1]$ and $K_0 \geq 64d/\varepsilon$, and let* CHI-L(r, K_0) *be the statement*

$$\sum_j I_j\, \mathbb{E}_{\widetilde{\Pr}_j}\big[(1 + X_j^\infty)\min\{1/K_0,\ q_j^\infty/\theta\}\big] \ \leq\ \tfrac{\varepsilon}{32}V.$$

Then CHI-L(r, K_0) *implies* SM-OF *at $L \geq \mathrm{poly}(d/(r\varepsilon))\, K_0/V$.*

Proof sketch. The spikes with $Q_j \geq K_0 q_j^\infty$ are bounded by Lemma 9.26 and the hypothesis. On the rest, $q_j^\infty > \theta/K_0 \geq 4I_j$ once $K_0 \leq L/4$, so the closure law satisfies (P1) of Lemma 9.6 with $\gamma \approx 1/(8K_0)$ on the subfamily where it carries half of the spiked mass; its load is $(d-1)/r$, and the size bound $s = O(d/(r\varepsilon))$ holds on the spikes that carry the mass. Lemma 9.6 gives the bound. □

CHI-L itself is false at θ-free thresholds, already for light spread gates (item (ii) of Section 9.7), so Proposition 9.28 is only a conditional reduction. No path bound is used, so neither the reordering defect of fixed-order arguments nor the $\ln(e/I)$ price of Theorem 9.14 enters. CHI-L is weaker than CHI, since $\mathbb{E}_{\widetilde{\Pr}}[(1+X^\infty)q^\infty] = I(1+X)$, and it removes the AND counterexample to CHI (AND blocks are in A^∞, so $X^\infty = 0$). In the comparison with SM, CHI-L implies SM-OF at the stated loss and is not known to follow from it. The loss is θ-free only if r and K_0 are.

9.7. **The open statement and the missing tool.** Fix $\delta = \beta = (V/d)^b$, let T_j^{gr} be the greedy closure at threshold δ and T_j^{fin} the final-context law at threshold β, and for $c \in [0,2]$ let $D_0(c) = C\ln(d/V) + c\ln(e/\theta)$ and

$$U_j^*(c) = \{Q_j \ge \theta\} \wedge \mathrm{LJ}_j \wedge \{\ln(Q_j/q_j(T_j^{\mathrm{gr}})) > D_0(c)\} \wedge \{\ln(Q_j/q_j(T_j^{\mathrm{fin}})) > D_0(c)\}.$$

Conjecture 9.29 (the graded residual hypothesis U_c^*; [*open for $c < 2$; with $c = 1/(2T)$ and a quasi-polynomial prefactor in place of $(d/V)^{C_0}$ it follows from Theorem 10.26 by the converse in Theorem 9.30, up to constants in ε*]). *There are b, C, C_0 such that for every h old-free capped at (θ, L) with $L \ge (d/V)^{C_0}\theta^{-c}$, $\sum_j \frac{I_j}{2}\widetilde{\Pr}_j(U_j^*(c)) \le \frac{\varepsilon}{16}V$.*

Theorem 9.30 (U_c^* and SM-OF; [*proved here, audited, as a conditional reduction*]). *U_c^* implies SM-OF at the loss $L_0 = \max\{L_{\mathrm{SJ}}, L_{\mathrm{cap}}(c), (d/V)^{C_0}\theta^{-c}\} = \mathrm{poly}(d/V)\,\theta^{-\max(c, c(\tau_1))}$, where $L_{\mathrm{cap}}(c) = e^{D_0(c)}\max\{4, 1024d^2/(\varepsilon V), (128d^{5/2}/(\varepsilon\beta))^2/V\} = \mathrm{poly}(d/V)(e/\theta)^c$. Conversely, SM-OF at a loss $\mathrm{poly}(d/V)\theta^{-c}$ implies U_c^* at the same loss, up to constants in ε. Moreover U_2^* holds once $(d/V)^{C_0} \ge (8d/\varepsilon)^3$.*

Proof. A large-jump spike outside $U_j^*(c)$ has defect at most $D_0(c)$ for T^{gr} or for T^{fin}, so $\{Q_j \ge \theta\}$ is covered by $\mathrm{sp}_j \wedge \mathrm{SJ}_j$, the two capture events Cap_j of Proposition 9.25 for T^{gr} and T^{fin} at $D_0 = D_0(c)$, and $U_j^*(c)$. Multiply by $I_j/2$ and sum. The first mass is at most $\frac{\varepsilon}{16}V$ by Theorem 9.23; the two captured masses are, by Proposition 9.25, since $L \ge L_{\mathrm{cap}}(c)$ gives $L \ge 4e^{D_0}$ and makes both branches of its maximum at most $\frac{\varepsilon}{16}V$; and the last is by U_c^*. The converse holds because $U_j^*(c) \subseteq \{Q_j \ge \theta\}$, and U_2^* follows from this inclusion and Proposition 9.21. □

θ-exponent c of the loss	consequence of SM-OF at $P\theta^{-c}$	status
$c = 2$	$\mathrm{MaxInf}(f) \ge \exp(-O(d^{\log_2 3}\ln(d/V)))$, weaker than $\mathrm{Var}(f)^2 e^{-O(d)}$	proved
constant $c < 1$	$\mathrm{MaxInf}(f) \ge \exp(-O_c(d^{\log_2(1+c)}\ln(d/V)))$	open
$cT \le \frac{1}{2}$	$\mathrm{MaxInf}(f) \ge (\mathrm{Var}f/d)^{O(\log d)}$	open
$cT = \frac{1}{2}$, $P = P_d/V$	$\mathrm{MaxInf}(f) \ge (\mathrm{Var}f/d)^{O(\log d)} e^{-O(\log^4 d)}$	proved

TABLE 4. The payoff of the θ-exponent ($V = \mathrm{Var}f$; $P = \mathrm{poly}(d/V)$ except in the last row, where $\ln P_d = 3.0 \cdot 10^6\, T^3$). The last row is Theorem 10.1 [*proved here; audited twice in note form, and the text audited twice*]. Every $c \ge 1$ gives only $\ln(1/\mathrm{MaxInf}) = \Omega(d\ln(d/V))$.

The rows follow from the proof of Proposition 9.17, and the last row from Lemma 10.2, Proposition 10.3 and Theorem 10.26: $\ln(1/\theta_0) = (1+c)^T\ln(1/\theta_T) + \ln P' \cdot ((1+c)^T - 1)/c$ with $(1+c)^T \le (1+c)^{3.45} d^{\log_2(1+c)}$, and $(1+c)^T \ge 2^T \ge 2ed$ when $c \ge 1$. A constant $c < 1$ is subexponential, but it beats the known bound $\mathrm{Var}(f)^2 e^{-O(d)}$ only when $\ln(1/V) = o(d^{1-\log_2(1+c)})$; the normalization $cT \le \frac{1}{2}$ is sufficient, and any $cT = O(1)$ is still quasi-polynomial. The θ-exponent budget across the pieces is a maximum, not a sum: the loss of Theorem 9.30 is the largest of the losses $P_i\theta^{-c_i}$ of the pieces, so its exponent is $\max_i c_i$, and a condition such as $c + c(\tau_1) \le 1/(2T)$ is sufficient but not necessary.

So U_c^* is SM-OF itself at the exponent $\max(c, c(\tau_1))$, restricted to the uncaptured large-jump spikes; it is not weaker. Its only content beyond SM-OF is that the small-jump and captured masses are θ-free. This is unavoidable for any formulation by uncaptured mass, since captured mass is automatically small (Lemma 9.8, Lemma 9.6, Proposition 9.25). Its content is structural: by Lemmas 9.13 and 9.27 and Proposition 9.28, a violator must amplify by more than e^{D_0} through coordinates that are large in the clock order but relatively small both online and in final context, and its closure remainder must be relatively regular, absolutely irregular at scale $(q^\infty)^2$, and stall at $q^\infty < r\,\mathrm{poly}(d/(\varepsilon\theta))$. None of the tested families (spread gates, parity gates, AND, approximate-OR cascades, products, the dense family at $L \ge L_0$) is known to have these properties; no member of U^* is known.

Question 9.31 (an adaptive scale-invariant H1)**.** Is there c with $cT \le \frac{1}{2}$ such that for every $\phi \in \mathcal{B}_{d-1}$ and every θ with $\mathbb{E}\phi^2 \le \theta^{1+c}/P$, $P = \mathrm{poly}(d/V)$, $\widetilde{\Pr}_\phi(U^*(c)) \le \varepsilon/(16d)$? Equivalently: if ϕ is relatively δ-regular online and β-regular in final context along the size-biased path, its restricted mean does not amplify by more than e^{D_0}, except with size-biased probability $\varepsilon/(16d)$, with constants independent of the height.

This local statement implies Conjecture 9.29 by summing over j, and hence SM-OF; it is stronger, since it is required for every ϕ and drops the cap on the other coordinates. The existing tools fall short of it for precise reasons. Section 10 does not answer it: it proves SM-OF with a quasi-polynomial prefactor, without a local statement, by a recursion over chain derivatives.

(i) *H1 is absolute.* At a node of height q, an amplification $\widetilde{\Pr}(Y \ge e^D q) \ge p$ of the restricted mean Y of $\psi = \phi_\rho^2$, with $pe^D \ge 2$, gives only $\mathbb{E}Y^2 \ge pe^D q^2$, i.e. low-level weight of order q^2. Theorem 3.2 then gives $\mathrm{RI}_u \ge (pe^D/2)^{5.2} q^{9.4}/(4 \cdot 10^4 d^4)$, and even a linear form of H1 would give $\mathrm{RI}_u \gtrsim pe^D q/\mathrm{poly}(d) = p\theta/\mathrm{poly}(d)$ at the root: a loss θ^{-1} at least. The hypothesis of H1 is the absolute variance, not the relative variance $\mathrm{Var}(Y)/q^2$; for AND_m the true relative influence is $\frac{1}{2}$ while $\mathrm{Var}(Y)/q \approx (3/4)^m$.

(ii) *The one-node relative statement is false.* Spread gates with $M = (\theta L)^{-1/3}$ gates have relative influence about $1/(2M)$ for every coordinate at the root and $\widetilde{\Pr}(\mathrm{sp}) \ge \frac{1}{4}$; the greedy law reveals nothing, and only the final context captures them. The matched gate has final relative influence about $\frac{1}{2}$ when one gate dominates, and about $1/(2(k^2+(k-1)^2))$ with k matched gates; missing the final context needs about $\beta^{-1/2}$ simultaneously matched gates, which has negligible $\widetilde{\Pr}$-probability for $\beta \le (V/d)^b$. So a scale-invariant H1 must be adaptive. For the same reason CHI-L fails at θ-free thresholds for light spread gates ([*proved here, unaudited*]; proof not included), and holds for them only in mass form, through Proposition 9.25.

(iii) *Theorem 9.14 has an absolute budget.* It is relative in the bias, but the small-bias budget $\kappa_1\Lambda$ is proportional to $\ln(e/q)$. In piece 1 this costs $(e/\theta)^{c(\tau_1)c_1}$, which is θ-free; in fixed-order activation arguments it puts $(I/e)^{\varepsilon'} = (\theta/e)^{\varepsilon'} e^{-\varepsilon'\ell}$ into the heavy threshold, and $e^{\varepsilon'\ell}$ is not polynomial on the relevant range $\ell \le 16(d-1)/(\tau_1^2 p_0)$, with $p_0 = \varepsilon/(16d)$ as above. Proposition 9.28 removes this factor from piece 3.

(iv) *No moment bound is θ-free.* Proposition 9.21 is the best moment bound, with loss θ^{-2}; by Proposition 9.22, no per-spike bound on the whole spiked mass has exponent below 1; and AND_m has $1 + X = (3/2)^m = I^{-0.585}$ although it is fully captured. So the missing statement is a capture statement, or a χ^2 bound after the heavy closure as in CHI-L, not a moment bound for ϕ.

The deep-chain residual. By Lemma 9.27, a light coordinate u is activated with non-negligible probability only if $1 + X_u$ is large, that is, only if the derivative is itself spiky; so the question repeats one level down. For a chain $c = (j, u_1, \dots, u_t)$ let $\psi_c = d_{u_t} \cdots d_{u_1} d_j h$, of degree at most $d - 1 - t$, with $\|\psi_c\|_2^2 \le I_j$, and let $R(t) = \sum_{|c|=t} \frac{1}{2}\|\psi_c\|_2^2 \widetilde{\Pr}_{\psi_c}$(the chain derivative reaches the height H_t), with heights $H_0 = \theta > H_1 > \cdots$ fixed by Theorem 9.14. Then $R(0) \ge \sum_j \mu_j$ and $R(d) = 0$. The one-level step (a change of measure to $\widetilde{\Pr}_{\psi_u}$ at relative influence at least $\tau_1^2/4$, and Theorem 9.14 at confidence y_t) is a [*proof outline; audited*]. With the identity $I_j \mathrm{RI}_u(0) = \mathrm{Inf}_u(\phi_j)$ it gives the mass recursion

$$R(t) \;\le\; 4e^{-y_t} W(t) + \frac{4}{\tau_1^2} R(t+1), \qquad W(t) = \sum_{|c|=t} \tfrac{1}{2}\|\psi_c\|_2^2 \le \frac{d^{t+1} V}{2},$$

which, like the arithmetic below, is [*proved here, unaudited*]. Bonami's inequality closes the recursion at any depth D: $R(D) \le 2 \cdot 9^{d-1-D}(\theta/L) W(D)/H_D$. Per unit weight one level costs $4(d-t)/\tau_1^2 \ge 4(2e^3 d^4)^2 (d-t) > 9$, since Theorem 9.14 needs $\tau_1 \le 1/(2e^3 d^4)$, while Bonami's inequality closes one degree for a factor 9. So a level of recursion always costs more than the degree it saves, the prefactor of the loss is at least 9^{d-1} for every D, and the best split is $D = 0$: plain Bonami, which gives SM-OF at $L_0 = 32 \cdot 9^{d-1} d/\varepsilon$, θ-free but exponential in d (compare Proposition I.1). Moreover, the exceptional terms sum to at most $\frac{\varepsilon}{32} V$ only if $y_t = t \ln(4d/\tau_1^2) + O(\ln(dD/\varepsilon))$ grows linearly in t, so the height loss $\ln(\theta/H_D)$ is quadratic in D. Within a loss $\mathrm{poly}(d/V)\theta^{-c}$, the recursion therefore extends the degree $O(\log(d/\theta))$ that Bonami's inequality closes by only $O((\ln(1/\theta)/\ln d)^{1/2})$ levels. With heights fixed in advance and a polynomial prefactor, the deep-chain residual is [*open*]. A non-compounding recursion would need a per-level factor below 9, a breadth bound over the activated coordinates only, or a drop of the degree by a constant factor per level, and generic residual chains lose one degree per level. Section 10 avoids the compounding differently [*proved here; audited twice in note form, and the text audited twice*]. Lightness is measured against a fixed fraction of the node's weight, which ends every chain at depth at most $S - 1 = 16T + 3$ (Corollary 10.12); each node is examined at a row of a grid of heights with one exceptional exponent for all depths, so the height loss is linear in the depth rather than quadratic, and it costs $\ln P_d = 3.0 \cdot 10^6 T^3$, a quasi-polynomial prefactor (Remark 10.27).

Lemma 9.32 (INC; [*proved here, unaudited*])**.** *For $g\colon Q_n \to \mathbb{R}$, a coordinate u and $\rho \in [0,1]$, $\mathrm{Stab}_\rho((d_u g)^2) \le \mathrm{Stab}_\rho(g^2)$, with equality for $g = x_u f$.*

Proof. Write $g_\pm$ for the restrictions to $x_u = \pm 1$, and $S_{ab} = \mathbb{E}[g_a^2(x) g_b^2(y)]$ for ρ-correlated x, y on the other coordinates. Pointwise $(d_u g)^2 = \frac{1}{4}(g_+ - g_-)^2 \le m := \frac{1}{2}(g_+^2 + g_-^2)$, and Stab_ρ is monotone on nonnegative functions, so $\mathrm{Stab}_\rho((d_u g)^2) \le \mathrm{Stab}_\rho(m) = \frac{1}{4}(S_{++} + S_{--}) + \frac{1}{2} S_{+-}$. Also $\mathrm{Stab}_\rho(g^2) = \frac{1+\rho}{4}(S_{++} + S_{--}) + \frac{1-\rho}{2} S_{+-}$, and the difference is $\frac{\rho}{4}\langle g_+^2 - g_-^2, T_\rho(g_+^2 - g_-^2)\rangle \ge 0$. □

With Lemma 9.20, $\mathrm{Stab}_{1/2}(\psi_c^2) \le \mathrm{Stab}_{1/2}(\phi_j^2) \le I_j^{4/3}$ along every chain, and $1 + X_{d_u g} \le (1 + X_g)/\mathrm{RI}_u(0)^2$. In Lemma 9.27 this cancels the factor $\mathrm{RI}_u(0)^2$, so a light coordinate is not penalized for being light. But the per-level factor becomes $(d-t)e^{y_t}/\beta^2$, and since every large-bias reveal must be caught, $\beta \le \tau_1^2/4$, so the factor is at least $16(d-t)/\tau_1^4 > 9$: charging only the increment does not stop the compounding [*proof outline; unaudited*].

Remark 9.33 (gapped address bits; [*proved here, unaudited*]). Let $w\colon Q_n \to [0,1]$ have degree at most k, and suppose that its range misses $[\tau - \eta, \tau + \eta)$, where $\frac{1}{8} \le \tau - \eta < \tau + \eta \le \frac{7}{8}$. Put $f_\tau = \mathbf{1}[w \ge \tau]$. Then

$$\mathrm{MaxInf}(w) \;\ge\; \eta^2 \mathrm{Var}(f_\tau)/D_\eta, \qquad D_\eta = C_\eta^2, \quad C_\eta = k^4(1 + 1/\eta)/(2\eta) = \mathrm{poly}(k/\eta).$$

Every value of w is at distance at least η from τ, so a minimal block that moves w across the gap has size at most $k^2/(2\eta)$, and at most $k^2(1+1/\eta)$ such blocks are disjoint; hence f_τ has certificate complexity at most C_η and a decision tree of depth at most C_η^2. The OSSS inequality gives $\mathrm{Var}(f_\tau) \le D_\eta \mathrm{MaxInf}(f_\tau)$, and a flip of f_τ forces $|d_u w| \ge \eta$, so $\mathrm{Inf}_u(f_\tau) \le \mathrm{Inf}_u(w)/\eta^2$. The range of a restriction of w lies in the range of w, so the gap is inherited, and a gapped, mildly sharp address bit of a product is captured by the greedy closure at the θ-free threshold $\beta = \eta^2/(8D_\eta)$, with load $(d-1)/\beta$ [*proof outline*]. Without a gap, the obstruction is a dense family of near-constant soft subcubes whose values fill an interval at resolution η. Whether a sharp bit of degree k can have such a family at polynomial η is [*open*].

9.8. **The one-shot restriction.**

Remark 9.34 (rate $1/d$). In this remark ρ denotes a rate-p restriction, as in Section 8 (each coordinate alive independently with probability p), $V = \mathrm{Var} f$, and $k = \lceil \log_2(d/p) \rceil$ is the level of Lemma 8.3 at $\eta = \frac{1}{2}$, so that $k + 2 \ge \log_2 d + 6$. [*Proved here, audited:*] at rate $p \le 1/(2ed)$ with $\theta_0 = (pV/(16d^2 3^k))^2 = (V/d)^{O(1)}$, the spike-forcing lemma states that $\frac{1}{p}\mathbb{E}_\rho U_{\theta_0}(f_\rho) \le \frac{V}{2}$ for every $f \in \mathcal{B}_d$; it follows from the base-case estimate of Lemma 8.3, with the exponent 2 of the remark after it. So a one-shot transfer "$\mathrm{MaxInf}(f) < \theta_0/L$ implies $\frac{1}{p}\mathbb{E}_\rho U_{\theta_0}(f_\rho) \ge \frac{3}{4}V$" is the conjecture restated, and every capture, amplification or bounded-activation hypothesis for the one-shot restriction fails for every AA-deficient f: spike forcing produces the spiked mass and Lemma 9.6 bounds its captured part. These one-shot theorems are valid implications but circular as reductions. Their structural class results survive. [*Heuristic:*] the obstruction at rate $1/d$ is the product of regular near-Boolean factors, which no uniform or size-biased priming law that we know of captures. [*Proved here, audited:*] at rate $\frac{1}{2}$ the same products are not spikes (Lemma 9.24(C4)). [*Proof outline; audited:*] Lemma 9.27, for explorations that read x only on F, kills their activation chains at every depth. This is why the remaining question of this section, with a polynomial prefactor, is posed at rate $\frac{1}{2}$; along the one-shot route of Section 11 it is the anti-spike inequality of Question 11.11.

10. A quasi-polynomial bound through activation chains

This section proves SM-OF (Definition 9.16) with a loss whose θ-exponent is $1/(2T)$ and whose prefactor is quasi-polynomial in d, and deduces a quasi-polynomial lower bound on MaxInf through a chain with a θ-dependent loss (Lemma 10.2). The route replaces the deep-chain residual of Section 9.7 by a recursion over chain derivatives. Three choices fix the exponent. A coordinate is heavy for a node if its influence is at least the fraction $M^{-1/4}$ of the node's weight, where $M = LV$ is the same at every node (Definition 10.6). Each node is examined at a whole row of heights, and the mass recursion runs over a grid of heights indexed by the depth s and the position k in the row (Definition 10.4). The exceptional term of Theorem B$^\circ$ is charged only on the event that the node's height reaches the intermediate height (Lemma 10.15); in the recursion this charge moves one step along the row, so a single exceptional exponent y, independent of the depth, serves every node. Every node is analysed in its own two-stage exploration by a restarted Theorem A (Theorem 10.14); this is legal because spikes are defined by their final height, and a sup-event costs only a factor 2 against the final height in any filtration that reads x only on F (Lemma 10.13). Section 10.8 records the status of every step and of every cited input. Throughout, $T = \lceil \log_2(2ed) \rceil \ge 4$; in Sections 10.1–10.8, $\varepsilon = 1/(16ed)$, $\gamma_1 = (e-1)\ln 2$, and $\widetilde{\Pr}_\psi$ is the size-biased law of a bounded ψ: x has density $\psi^2/\|\psi\|_2^2$ with respect to the uniform measure U, every coordinate carries an i.i.d. clock $t_v \sim U[0,1]$ independent of x, and $F = \{v : t_v \le \frac{1}{2}\}$ (clock rate $\frac{1}{2}$). For the root functions $\phi_j = d_j h$ of Section 9.1 this puts a clock also on j, which that section omits; nothing changes, since ϕ_j does not depend on x_j. In the

main text of Sections 10.1–10.8 the letter p denotes only 2^{-T} (Section 10.1), and the closure probability is p_*; the alive rate p of Theorem 9.14 and of Appendix K equals $\frac{1}{2}$ at the nodes. We write $\mathrm{bs}(h) = 2\,\mathrm{rbs}(h)$ and $\mathrm{bs}(h,y) = 2\,\mathrm{rbs}(h,y)$, with rbs as in Definition 7.6.

Theorem 10.1 (quasi-polynomial bound; [*proved here; audited twice in note form, and the text audited twice; components as tagged*])**.** *Let $d \geq 2$, $f \in \mathcal{B}_d$ and $v = \mathrm{Var} f > 0$. Then*

$$\ln \frac{1}{\mathrm{MaxInf}(f)} \;\leq\; 6.6\,\ln \frac{c_2 d^{8.76}}{v} + 1.30\,T\Big(\ln P_d + \ln \frac{1}{v} + 6.4 \ln d + 23.2\Big),$$

where $\ln P_d = 3.0 \cdot 10^6\, T^3$ and $c_2 = 512 e^2 (64e^2)^{\log_2 3} < 6.6 \cdot 10^7$ is the constant of Theorem 8.4. Expanded,

$$\ln \frac{1}{\mathrm{MaxInf}(f)} \;\leq\; 3.9 \cdot 10^6\, T^4 + (1.30\,T + 6.6) \ln \frac{1}{v} + 8.4\,T \ln d + 30.2\,T + 57.9 \ln d + 119.$$

In particular $\mathrm{MaxInf}(f) \geq (v/d)^{O(\log d)} \exp(-O(\log^4 d))$ for all $d \geq 2$.

The constant $3.0 \cdot 10^6$ is valid for every $d \geq 2$ (Proposition 10.25); the constant actually needed falls with T from $2.68 \cdot 10^6$ at $T = 4$ towards $256 K_\ell = 2.486 \cdot 10^6$, with $K_\ell = 3000 e \gamma_1$ [*numerical*]. The exponent $\log^4 d$ is what this scheme gives (Remark 10.27). The constants are large: at $v = 1$ the bound beats e^{-d} for every $d \geq 1.91 \cdot 10^{13}$ and for no $d \leq 1.90 \cdot 10^{13}$ [*numerical*].

10.1. **The chain with a θ-dependent loss.** The chain of Theorem 8.4 is stated for a loss $L = L(d)$. We need a loss that depends on θ, is quasi-polynomial in d, and is required only on restrictions with variance above a floor. Propositions 9.17 and 8.9 contain this modification in their proofs for a polynomial prefactor; we state and prove it. Lemma 10.2, Proposition 10.3 and their proofs are unchanged from the previous audited version of the section; the paragraph after Proposition 10.3 is new.

Lemma 10.2 (chain with a θ-dependent loss; [*proved here, audited*])**.** *Let $C \geq 1$, $c > 0$ with $cT \leq \frac{1}{2}$, $P^* \geq 1$ and $L^*(\theta) = P^* \theta^{-c}$. Put $\delta = 1/(128 e^2 d^2)$, $\eta = \varepsilon$, $p = 2^{-T}$ and $k = \lceil \log_2(8ed^2/p) \rceil$. Let $f \in \mathcal{B}_d$ with $v = \mathrm{Var} f > 0$, and suppose that for every $g \in \mathcal{B}_d$ with $\mathrm{Var} g \geq \delta v$ and every $0 < \theta \leq \mathrm{Var}(g)^C$,*

$$U_{\theta / L^*(\theta)}(g) \;\leq\; 2\,\mathbb{E}_\sigma U_\theta(g_\sigma) + \varepsilon \mathrm{Var}(g). \tag{TT*}$$

Put $\theta_T = \min\{(\eta p v/(8 d^2 3^k))^4, (pv/(16ed))^C\}$ and $\theta_{t-1} = \theta_t / L^*(\theta_t)$. Then $\mathrm{MaxInf}(f) \geq \theta_0$ and

$$\ln \frac{1}{\theta_0} = (1+c)^T \ln \frac{1}{\theta_T} + \ln P^* \,\frac{(1+c)^T - 1}{c} \;\leq\; 1.65 \ln \frac{1}{\theta_T} + 1.30\,T \ln P^*.$$

For $C = 2$, $\ln(1/\theta_T) \leq 4 \ln(c_2 d^{8.76}/v)$.

Proof. Let ρ_t be the rate-2^{-t} restriction, $V_t = \mathbb{E}\mathrm{Var}(f_{\rho_t}) \geq \max(V_{t+1}, 2^{-t} v)$ and $\xi_t = \mathbb{E} U_{\theta_t}(f_{\rho_t})/V_t$. In step t ($t = T, \dots, 1$) let $g = f_{\rho_{t-1}}$. If $\mathrm{Var} g < \delta v$, use $U_{\theta_{t-1}}(g) \leq \mathrm{Var} g < \delta v$. If $\mathrm{Var} g \geq \delta v$ and $\theta_t \leq \mathrm{Var}(g)^C$, (TT*) gives $U_{\theta_{t-1}}(g) \leq 2\mathbb{E}_\sigma U_{\theta_t}(g_\sigma) + \varepsilon \mathrm{Var} g$. Otherwise $U_{\theta_{t-1}}(g) \leq \mathrm{Var} g < \theta_t^{1/C} \leq \theta_T^{1/C}$, since $L^* \geq 1$ on $(0,1]$ gives $\theta_t \leq \theta_T$. Averaging, with $\mathbb{E}_{\rho_{t-1}} \mathbb{E}_\sigma = \mathbb{E}_{\rho_t}$, and dividing by $V_{t-1} \geq \max(V_t, pv)$,

$$\xi_{t-1} \leq 2\xi_t + \varepsilon + \theta_T^{1/C}/(pv) + \delta/p.$$

Since $d/(2\eta p) = 8ed^2/p$, k is the k of Lemma 8.3 and $\theta_T \leq \theta^*$, so by Lemma 8.1(b),(d) and Lemma 8.3, $\mathbb{E} U_{\theta_T}(f_{\rho_T}) \leq \mathbb{E} N_{\theta^*}(f_{\rho_T}) \leq \eta p v \leq \eta V_T$, i.e. $\xi_T \leq \eta$. Unrolling, with $2^T < 4ed$ and $\theta_T^{1/C} \leq pv/(16ed)$, the first three terms give at most $2^T \cdot 3/(16ed) < \frac{3}{4}$, and the floor adds at most $(2^T - 1) 2^T \delta < 16 e^2 d^2 \delta = \frac{1}{8}$. Hence $\xi_0 < 1$. Since ρ_0 is trivial, $V_0 = v$ and

$\xi_0 = U_{\theta_0}(f)/\mathrm{Var}f$, so $U_{\theta_0}(f) < \mathrm{Var}f$, and $\mathrm{MaxInf}(f) \ge \theta_0$ by Lemma 8.1(c). The loss enters only through $\theta_{t-1} = \theta_t/L^*(\theta_t)$ and $L^* \ge 1$, and the base case does not depend on it. The identity follows from $X_{t-1} = (1+c)X_t + \ln P^*$ for $X_t = \ln(1/\theta_t)$. Since $cT \le \frac{1}{2}$, $(1+c)^T \le e^{1/2} < 1.65$, and $((1+c)^T - 1)/c = \sum_{i<T}(1+c)^i$ is increasing in c, so it is at most $2T(e^{1/2}-1) < 1.30T$. The bound on θ_T for $C = 2$ is that of Theorem 8.4. □

In the proof, (TT*) is used only at $g = f_{\rho_{t-1}}$ and $\theta = \theta_t$, $t = 1, \dots, T$. The floor δv is what lets one loss $L^*(\theta_t)$ serve every kept restriction in step t, which the averaging over ρ_{t-1} needs; Proposition 10.3 obtains P^* independent of g from it. The proof is linear in $\ln P^*$ and never uses the size of P^*, so a quasi-polynomial P^* is allowed. It also covers any $cT = O(1)$, with constants e^{cT} and $(e^{cT}-1)/c$ (for $cT = 0.8$ these are 2.23 and $1.532\,T$); here $cT = \frac{1}{2}$.

Proposition 10.3 (SM-OF gives the chain hypothesis; [*proved here, audited*])**.** *Suppose* SM-OF(L_0) *holds with* $L_0(\theta, d, V) = P_d\, V^{-1}\theta^{-c}$, *where* $P_d \ge 1$ *depends on* d *only and* $c \le 1/8$. *Then (TT*) holds with* $C = 2$ *and* $\ln P^* \le \ln P_d + \ln(1/v) + 6.4 \ln d + 23.2$.

Proof. Let $g \in \mathcal{B}_d$, $\theta \le \mathrm{Var}(g)^2$, and suppose the inequality of Definition 8.2 fails for g at θ with loss L. By Corollary J.21 (through Lemma J.20), the averaged h is old-free capped at $(\theta', L') = (\varepsilon^2\theta/(36d), \varepsilon^2 L/(36d))$, with $\mathrm{Var}h > \varepsilon \mathrm{Var}g$ and $\Delta(h; \theta', L') > \frac{\varepsilon}{2}\mathrm{Var}h$. By Lemma I.9 and Corollary 9.2, $\Delta(h; \theta', L') \le 2\sum_j \mu_j$. So SM-OF fails at (θ', L') unless $L' < L_0(\theta', d, \mathrm{Var}h)$. Since L_0 is nonincreasing in V, TT holds for g at θ whenever

$$L \;\ge\; L_g(\theta) := (9216e^2d^3)^{1+c}\, P_d\, (\varepsilon \mathrm{Var}g)^{-1}\theta^{-c}.$$

If $\mathrm{Var}g \ge \delta v$ then $(\varepsilon\mathrm{Var}g)^{-1} \le 2048e^3d^3/v$, so $L_g(\theta) \le P^*\theta^{-c}$ with $\ln P^* = \ln P_d + \ln(1/v) + (1+c)\ln(9216e^2d^3) + \ln(2048e^3d^3) \le \ln P_d + \ln(1/v) + 6.4\ln d + 23.2$. Since $U_{\theta/L}(g)$ is nonincreasing in L, TT at L_g gives TT at $P^*\theta^{-c}$. □

So it remains to prove SM-OF(L_0) with $c = 1/(2T)$ and $\ln P_d = 3.0 \cdot 10^6\, T^3$. Fix h old-free capped at (θ, L) with $L \ge L_0$, and write $\phi_j = d_j h$, $I_j = \mathrm{Inf}_j(h)$, $V = \mathrm{Var}h$, so that $\sum_j I_j \le dV$ and $\mu_j = \frac{1}{2} I_j \widetilde{\Pr}_j(Q_j \ge \theta)$ (Section 9.1). Put

$$X = \ln\frac{1}{\theta}, \qquad M = LV.$$

Then $\mathrm{MaxInf}(h) < \theta/L$; $X \ge \ln(36d) > 0$, since $\theta \le V^2/(36d)$ and $V \le 1$; $\ln L = \ln M + \ln(1/V) \ge \ln M$; and $\ln M \ge cX + \ln P_d$, since $L \ge L_0$.

10.2. **Nodes, the grid of heights, growth and closure.** The parameters ε_1, τ_1, S and y are fixed in Section 10.7, and $\alpha = \frac{1}{4}$. Here $K = 2$ is the constant of Lemma 10.13, and

$$\beta_1^\circ = \tau_1^2/4, \qquad \beta_1 = \beta_1^\circ/K = \tau_1^2/8, \qquad \ell = \ln\frac{1}{\beta_1};$$

the extra factor $1/K$ is the price of Lemma 10.13 in Theorem 10.16. With $\varepsilon_1 = \gamma_1\kappa_1$ (κ_1 as in Theorem 9.14) put

$$r = \frac{1}{1-\varepsilon_1}, \qquad b_0 = 1 + \ln 2 + 2y + \varepsilon_1(1+y), \qquad b_1 = rb_0 + y + \ell, \qquad b_2 = rb_0 + \ln 2.$$

Definition 10.4 (the grid of heights)**.** For $0 \le s, k \le S-1$ define $Z_{s,k}$ and $z_{s,k}$ by $Z_{0,0} = X$, $z_{s,k} = r(Z_{s,k} + b_0)$, and

$$Z_{s,k} = \max\{z_{s-1,k} + y + \ell,\; z_{s,k-1} + \ln 2\} \qquad ((s,k) \ne (0,0)), \tag{2}$$

where the first branch is omitted if $s = 0$ and the second if $k = 0$. The *heights* are $H_{s,k} = e^{-Z_{s,k}}$ and the *intermediate heights* are $h_{s,k} = e^{-z_{s,k}}$. The *bottom row* is $Z_s^* = Z_{s,S-1}$, $H_s^* = H_{s,S-1}$, $h_s^* = h_{s,S-1}$; put $D_s^* = \ln(H_s^*/h_s^*)$ and $B_s^* = Z_{s+1}^* - Z_s^*$.

The grid is deterministic: it depends on θ, d and T only. Row s lists the heights at which the nodes of depth s are examined. The step $k \to k+1$ along a row is the step from the height H to the threshold $h/2$ of the refined exceptional term (Lemma 10.15); the step $s \to s+1$ is the step from a parent to its children (Theorem 10.16).

Lemma 10.5 (the grid; [*proved here, audited*])**.** *Suppose $\gamma_1\tau_1 < 1$ and $y + \ell \geq \ln 2$. For all $0 \leq s, k \leq S-1$:*

(G0) $0 < h_{s,k} < H_{s,k} \leq \theta$, *and*

$$\ln \frac{H_{s,k}}{2h_{s,k}} \;=\; 1 + 2y + \varepsilon_1\Big(\ln \frac{e}{h_{s,k}} + y\Big) \;>\; 2y + \gamma_1 \max\Big\{\tau_1,\; \kappa_1\Big(\ln \frac{e}{h_{s,k}} + y\Big)\Big\}. \tag{3}$$

(G1) $H_{s,k+1} \leq h_{s,k}/2$ *if* $k < S-1$.

(G2) $H_{s+1,k} \leq \beta_1 h_{s,k} e^{-y}$ *if* $s < S-1$.

(G3) $Z_{s,k}$ *is strictly increasing in* s *and in* k, *and* $Z_{s,k} \geq X$. *Hence* $H^*_s \leq H_{s,k} \leq \theta$ *and* $h^*_s \leq h_{s,k}$.

(G4) $Z_{s,k} = r^{s+k}X + r^k\sigma_s + \omega_k$, *where* $\sigma_s = b_1(r^s-1)/(r-1)$ *and* $\omega_k = b_2(r^k-1)/(r-1)$.

Consequently, for $0 \leq s \leq S-1$ (and $s \leq S-2$ for B^*_s),

$$B^*_s = \varepsilon_1 r^{s+S}X + r^{s+S-1}b_1, \qquad Z^*_0 - X = (r^{S-1}-1)X + \omega_{S-1},$$
$$D^*_s = \varepsilon_1 r^{s+S}X + (r-1)(r^{S-1}\sigma_s + \omega_{S-1}) + rb_0,$$
$$\ln \frac{\theta}{h^*_s} = (r^{s+S}-1)X + r(r^{S-1}\sigma_s + \omega_{S-1}) + rb_0,$$

and $\sigma_s \leq s r^{s-1} b_1$, $\omega_k \leq k r^{k-1} b_2$.

Proof. (G3) Since $r \geq 1$ and $b_0 > 0$, $r(Z + b_0) > Z$ for $Z \geq 0$. So each branch of (2) exceeds the entry it is computed from, and induction from $Z_{0,0} = X > 0$ gives the claims; $z_{s,k}$ is increasing in $Z_{s,k}$.

(G0) Write $Z = Z_{s,k}$ and $z = z_{s,k}$. Then $z > Z$ gives $h_{s,k} < H_{s,k}$, and (G3) gives $H_{s,k} \leq e^{-X} = \theta$. From $(1-\varepsilon_1)z = Z + b_0$ we get $z - Z = b_0 + \varepsilon_1 z = 1 + \ln 2 + 2y + \varepsilon_1(1 + z + y)$, and $1 + z = \ln(e/h_{s,k})$; this is the identity in (3). With $a = \ln(e/h_{s,k}) + y > 0$ and $\gamma_1\kappa_1 = \varepsilon_1$, $\gamma_1 \max\{\tau_1, \kappa_1 a\} \leq \gamma_1\tau_1 + \varepsilon_1 a < 1 + \varepsilon_1 a$, which is the inequality.

(G1) and (G2) are the two branches of (2), since $e^{-\ln 2} = \frac{1}{2}$ and $e^{-\ell} = \beta_1$.

(G4) By induction on $s + k$, using $r\sigma_{s-1} + b_1 = \sigma_s$ and $r\omega_{k-1} + b_2 = \omega_k$. For $k = 0$ only the first branch is present, and $r(r^{s-1}X + \sigma_{s-1} + b_0) + y + \ell = r^sX + \sigma_s$. For $s = 0$ only the second is present, and $r(r^{k-1}X + \omega_{k-1} + b_0) + \ln 2 = r^kX + \omega_k$. For $s, k \geq 1$ the second branch is $r(r^{s+k-1}X + r^{k-1}\sigma_s + \omega_{k-1} + b_0) + \ln 2 = r^{s+k}X + r^k\sigma_s + \omega_k$, the first is $r^{s+k}X + r^{k+1}\sigma_{s-1} + r\omega_k + b_1$, and the second minus the first is

$$r^k(\sigma_s - r\sigma_{s-1}) - b_1 - (r-1)\omega_k = (b_1 - b_2)(r^k - 1) = (y + \ell - \ln 2)(r^k - 1) \;\geq\; 0.$$

So the maximum is the second branch, which is the closed form.

The consequences: $\sigma_{s+1} - \sigma_s = r^s b_1$, $\sigma_0 = \omega_0 = 0$ and $r - 1 = \varepsilon_1 r$ give B^*_s and $Z^*_0 - X$; $D^*_s = z_{s,S-1} - Z^*_s = (r-1)Z^*_s + rb_0$ and $\ln(\theta/h^*_s) = z_{s,S-1} - X = rZ^*_s + rb_0 - X$; and $(r^n - 1)/(r-1) = \sum_{i<n} r^i \leq n r^{n-1}$. □

Definition 10.6 (nodes)**.** A *chain* at level s is a sequence $c = (j, u_1, \ldots, u_s)$ of distinct coordinates. Its *node function* is $\psi_c = d_{u_s} \cdots d_{u_1} d_j h$, with $w_c = \|\psi_c\|_2^2$, *law* $\widetilde{\Pr}_c = \widetilde{\Pr}_{\psi_c}$ (standard rate-$\frac{1}{2}$ clocks on all coordinates) and final height $Q^c = \mathbb{E}[\psi_c^2 \mid x_F]$. Its *amplification* is $A_c = H^*_s / w_c$, and its *heavy set* is $\mathcal{H}_c = \{v : \mathrm{Inf}_v(\psi_c) \geq M^{-\alpha} w_c\}$. Put

$$p_* = \frac{\varepsilon}{64Sd}\Big(\frac{\beta_1}{d}\Big)^{S-1}$$

and $\mathcal{C}_0 = \{(j) : I_j > 0\}$; a node $c \in \mathcal{C}_s$ is *closed* if $A_c \geq p_*^{-3}(H_s^*)^{-2}$ and *open* otherwise, and $\mathcal{C}_{s+1} = \{c \cdot v : c \in \mathcal{C}_s \text{ open},\ v \notin \mathcal{H}_c \cup c,\ \mathrm{Inf}_v(\psi_c) > 0\}$.

The heavy threshold $M^{-\alpha}$ does not depend on the node, and the closure probability p_* does not depend on the level.

Lemma 10.7 (node facts; [*proved here, audited; (a), (b) and (d) as in the previous version*])**.** (a) $|\psi_c| \leq 1$, ψ_c *does not depend on* x_u *for* $u \in c$, *and* $\deg \psi_c \leq d-1-s$. (b) $w_{c\cdot v} = \mathrm{Inf}_v(\psi_c) = \rho_v^c\, w_c$ *with* $\rho_v^c := \mathrm{Inf}_v(\psi_c)/w_c$, *and* $\sum_v \rho_v^c \leq d$. (c) $|\mathcal{H}_c| \leq dM^\alpha$ *(not used below),* $\mathcal{H}_c \cap c = \emptyset$, *and* $\mathcal{H}_c$ *is a function of the chain* c *alone (given* h *and* L*).* (d) $\sum_{c\in\mathcal{C}_s} w_c \leq d^{s+1}V$. (e) *At the root,* $\{Q^{(j)} \geq H_{0,0}\} = \{Q_j \geq \theta\}$, *so* $\mu_j = \frac{1}{2} I_j \widetilde{\Pr}_{(j)}(Q^{(j)} \geq H_{0,0})$.

Proof. (a) d_u is the half-difference, so $|d_u g| \leq 1$ if $|g| \leq 1$. (b) $\|d_v\psi\|_2^2 = \mathrm{Inf}_v(\psi)$ and $\sum_v \mathrm{Inf}_v(\psi) \leq \deg(\psi)\|\psi\|_2^2$. (c) $\sum_v \mathrm{Inf}_v(\psi_c) \leq dw_c$ gives $|\mathcal{H}_c| M^{-\alpha} w_c \leq dw_c$; $\mathrm{Inf}_u(\psi_c) = 0$ for $u \in c$, while $M^{-\alpha}w_c > 0$; and M and w_c are determined by h, L and c. (d) Iterate (b) from $\sum_j I_j \leq dV$. (e) $H_{0,0} = e^{-X} = \theta$ and $\psi_{(j)} = \phi_j$; the rest is the identity after Definition 9.1, since h is old-free. □

Lemma 10.8 (closure; [*proved here, audited*])**.** *For* $\psi\colon Q_n \to [-1,1]$ *with* $w = \|\psi\|_2^2 > 0$, $Q = \mathbb{E}[\psi^2 \mid x_F]$ *and* $H > 0$, $\widetilde{\Pr}_\psi(Q \geq H) \leq w^{1/3}/H$.

Proof. Q is $\sigma(\text{clocks}, x_F)$-measurable, so $\widetilde{\Pr}_\psi(Q \geq H) = \mathbb{E}[Q\mathbf{1}[Q \geq H]]/w \leq \mathbb{E}[Q^2]/(Hw)$, and $\mathbb{E}[Q^2] = \mathrm{Stab}_{1/2}(\psi^2) \leq w^{4/3}$ by Lemma 9.20. □

Lemma 10.9 (closed nodes; [*proved here, audited*])**.** *If* $c \in \mathcal{C}_s$ *is closed, then* $\widetilde{\Pr}_c(Q^c \geq H_{s,k}) \leq p_*$ *for every* $k \leq S-1$.

Proof. Closed means $w_c = H_s^*/A_c \leq p_*^3 (H_s^*)^3$. By Lemma 10.8 and (G3), $\widetilde{\Pr}_c(Q^c \geq H_{s,k}) \leq w_c^{1/3}/H_{s,k} \leq p_* H_s^*/H_{s,k} \leq p_*$. □

Lemma 10.10 (growth; [*proved here, audited*])**.** *For* $s \leq S-2$, $c \in \mathcal{C}_s$ *and* $c \cdot v \in \mathcal{C}_{s+1}$, $\ln A_{c\cdot v} > \ln A_c + \alpha \ln M - B_s^*$.

Proof. $v \notin \mathcal{H}_c$ gives $0 < w_{c\cdot v} = \mathrm{Inf}_v(\psi_c) < M^{-\alpha} w_c$, so

$$A_{c\cdot v} = H_{s+1}^*/w_{c\cdot v} > (H_{s+1}^*/H_s^*) M^\alpha (H_s^*/w_c) = e^{-B_s^*} M^\alpha A_c.$$

□

Corollary 10.11 (amplification floor; [*proved here, audited*])**.** *Put* $\lambda_s = \ln L + X - Z_s^* + s\alpha \ln M$. *Then* $\ln A_c > \lambda_s$ *for every* $s \leq S-1$ *and every* $c \in \mathcal{C}_s$. *If*

$$(G'') \qquad \alpha \ln M \ \geq\ B_s^* \qquad \text{for every } s \leq S-2,$$

then $\lambda_s \geq \lambda_0 = \ln L - (Z_0^* - X)$ *for every* $s \leq S-1$.

Proof. At the root, $0 < I_j \leq \mathrm{MaxInf}(h) < \theta/L$ and $w_{(j)} = I_j$, so $A_{(j)} = H_0^*/I_j > LH_0^*/\theta$, and $\ln(H_0^*/\theta) = X - Z_0^*$; this is $\ln A_{(j)} > \lambda_0$. The step is Lemma 10.10, since $\lambda_{s+1} - \lambda_s = \alpha \ln M - B_s^*$. Under (G″) these increments are nonnegative for $s \leq S-2$. □

The *root loss* $Z_0^* - X = (r^{S-1} - 1)X + \omega_{S-1}$ is the price of measuring amplification against the bottom row; it enters (CV″) and (MW″) below.

Corollary 10.12 (depth; [*proved here, audited; the inequality it assumes is part of Proposition 10.25*])**.** *If*

$$(CL'') \qquad (1 + (S-1)\alpha) \ln M \ \geq\ 3Z_{S-1}^* - X + 3 \ln \frac{1}{p_*},$$

then every node of $\mathcal{C}_{S-1}$ *is closed and* $\mathcal{C}_S = \emptyset$.

Proof. By $\ln L \geq \ln M$ and (CL″), $\lambda_{S-1} = \ln L + X - Z^*_{S-1} + (S-1)\alpha \ln M \geq 3\ln(1/p_*) + 2Z^*_{S-1} = \ln(p_*^{-3}(H^*_{S-1})^{-2})$. For $c \in \mathcal{C}_{S-1}$, $\ln A_c > \lambda_{S-1}$ by Corollary 10.11, so c is closed, and closed nodes have no children. □

This uses only the first claim of Corollary 10.11, not (G″). Every chain ends at depth at most $S - 1 = 16T + 3$.

10.3. **Filtration robustness.** This subsection is unchanged from the audited version of the section.

Lemma 10.13 (FR; [*proved here, audited*])**.** *Let $\psi\colon Q_n \to [-1,1]$ with $w = \|\psi\|_2^2 > 0$ and $Q = \mathbb{E}[\psi^2 \mid x_F]$. Let $\mathbb{F} = (\mathbb{F}_t)_{t=0,1,\dots,N}$, $N < \infty$, be a filtration in discrete time (in Theorem 10.16, t is the reveal index of an exploration) contained in $\sigma(\text{clocks}, x_F, \text{coins})$, the coins independent of x (it reads x only on F, adaptively or not), with terminal σ-field $\mathbb{F}_N$ containing $\sigma(F, x_F)$; put $q_t = \mathbb{E}_U[\psi^2 \mid \mathbb{F}_t]$. For every $\mathbb{F}$-stopping time τ and $K > 1$, $\widetilde{\Pr}_\psi(\exists t \geq \tau : q_t < q_\tau/K \mid \mathbb{F}_\tau) \leq 1/K$. Hence*

$$\widetilde{\Pr}_\psi\big(\max_t q_t \geq H\big) \;\leq\; \tfrac{K}{K-1}\,\widetilde{\Pr}_\psi(Q \geq H/K) \qquad (H > 0).$$

Proof. The density of $\widetilde{\Pr}_\psi$ on $\mathbb{F}_t$ is q_t/w. The terminal value of q is Q: since x is independent of the clocks and coins under U, $\mathbb{E}_U[\psi^2 \mid \text{clocks}, x_F, \text{coins}] = Q$, and Q is $\sigma(F, x_F)$-measurable, hence measurable for the terminal σ-field. Optional stopping at $\tau' = \min\{t \geq \tau : q_t < q_\tau/K\}$ ($\tau' = \infty$ if there is none) gives the first bound. For the display take $\tau = \min\{t : q_t \geq H\}$; on $\{\tau < \infty\}$ we have $q_\tau \geq H$, and then $Q \geq H/K$ with conditional probability at least $1 - 1/K$ by the same density argument at the final time. □

The hypothesis that $\mathbb{F}$ reads x only on F is the one of Lemma 9.27 and cannot be dropped. Exact enumeration ($n = 6$, an optimal adaptive order) gives worst ratio 0.525 in the display at $K = 2$ [*numerical*].

10.4. **Theorem B° at a node and Theorem DNC.** Fix an open $c \in \mathcal{C}_s$ and numbers $0 < h < H \leq 1$ and $y > 0$. The exploration $\mathcal{G}^c$ of c has two stages and no pre-revealed set: stage 1 reveals $\mathcal{H}_c \cap F$ with its values at once ($\mathcal{H}_c$ is fixed by the index c); stage 2 reveals $F \setminus \mathcal{H}_c$ in increasing clock order, at reveal indices $k = 1, \dots, m$; the coordinates of $\mathcal{H}_c \setminus F$ are frozen (alive, never revealed, hazard 0). Since stage 1 is revealed at once, this is not a staged exploration in the sense of Theorem 9.15; Theorem 10.14 treats it from the stage boundary on, through Lemma K.2 and Proposition K.3. Let v_k be the k-th stage-2 revealed coordinate and S_k the revealed set after k stage-2 reveals, $q^{(k)} = \mathbb{E}_U[\psi_c^2 \mid x_{S_k}]$, c_k the bias of the k-th stage-2 reveal just before it, and call it *large* if $|c_k| \geq \tau_1$. Let $\mathcal{G}^c_k$ be the σ-field of the clocks and values revealed up to and including the k-th stage-2 value (stage 1 only for $k = 0$), and for $k > m$ the terminal σ-field; we drop the superscript c when the node is fixed. Let $\mathcal{G}^{c,+}_{k-1}$ be the σ-field of the clocks and values revealed before the k-th stage-2 value, together with the identity and clock of the k-th revealed coordinate. Define

$$\mathrm{cap}_c(h) = \{q^{(0)} \geq h\}, \qquad \zeta = \min\{k \geq 0 : q^{(k)} \geq h\} \quad (\zeta = \infty \text{ if no } k \leq m \text{ qualifies}),$$

$$E^c_v = \{q^{(0)} < h,\ \zeta \leq m,\ \exists k > \zeta :\ v_k = v,\ |c_k| \geq \tau_1,\ |c_j| < \tau_1\ (\zeta < j < k),\ q^{(k-1)} \geq he^{-y}\}.$$

The exploration, and with it v_k, S_k, $q^{(k)}$, c_k and the σ-fields, depends on the chain c alone, since $\mathcal{H}_c$ does (Lemma 10.7(c)); ζ, $\mathrm{cap}_c(h)$ and E^c_v depend also on (h, y). The search for ζ starts at the stage boundary, so one hazard ($\leq \ln 2$) is paid after it. The events E^c_v are pairwise disjoint, since v is the first large reveal *strictly* after ζ, and $E^c_v \cap \{v_k = v\} \in \mathcal{G}^{c,+}_{k-1}$. Only $v \notin \mathcal{H}_c$ occur. Counting the crossing reveal itself ($k \geq \zeta$) would make E^c_v depend on x_v; exact enumeration shows that this breaks the change of measure by 43–48% [*numerical*].

Theorem 10.14 (Theorem B°, node form; [*proved here, Appendix K; audited*])**.** *Let $\psi\colon Q_n \to [-1,1]$ be multilinear of degree at most d with $w = \|\psi\|_2^2 > 0$, let $\mathcal{H}$ be a fixed set, and let $\mathcal{G}$, $q^{(k)}$, ζ, cap and E_v be as above with $(\psi,\mathcal{H})$ in place of $(\psi_c,\mathcal{H}_c)$. Let $\tau_1 \in (0, 1/(2e^3d^4)]$, κ_1 as in Theorem 9.14, $0 < h < H \le 1$ and $y > 0$ with $\ln(H/(2h)) \ge 2y + \gamma_1 \max\{\tau_1, \kappa_1(\ln(e/h)+y)\}$. Write $\widetilde{\Pr}$ for $\widetilde{\Pr}_\psi$ and $\widetilde{\Pr}_v$ for $\widetilde{\Pr}_{d_v\psi}$, with the same clocks, and $Q = \mathbb{E}[\psi^2 \mid x_F]$.*

(a) $\widetilde{\Pr}\big(Q \ge H,\ q^{(0)} < h,\ \text{no } E_v \text{ occurs}\big) \le 3e^{-y}$.

(b) $E_v \cap \{v_k = v\} \in \mathcal{G}^+_{k-1}$, *and on this event* $\rho_v(S_{k-1}) := \mathbb{E}_U[(d_v\psi)^2 \mid x_{S_{k-1}}]/q^{(k-1)} \ge c_k^2/4 \ge \beta_1^\circ$ *and* $\mathbb{E}_U[(d_v\psi)^2 \mid x_{S_{k-1}}] \ge \beta_1^\circ h e^{-y}$.

(c) $\widetilde{\Pr}(E_v) \le (\rho_v/\beta_1^\circ)\,\widetilde{\Pr}_v(E_v)$, *where* $\rho_v = \mathrm{Inf}_v(\psi)/w$.

Proof outline; the full proof is in Appendix K. Strong Markov at a stopping index. For a $(\mathcal{G}_k)$-stopping index τ, on $\{\tau = k\}$ and given $\mathcal{G}_\tau$: $x_{S_k^c}$ has law $(\psi_{S_k}^2/q^{(k)})\cdot U$, where $\psi_{S_k} = \psi(x_{S_k},\cdot)$; the clocks of the unrevealed stage-2 coordinates are i.i.d. $U(s_k,1]$, where s_k is the clock of v_k and $s_0 = 0$, and independent of x; and the frozen coordinates are never revealed. This holds because the stage-2 order is a function of the clocks and x is independent of the clocks.

The proof of Theorem 9.14, run from the state at ζ. Its inputs are supplied by the strong Markov property at s_ζ; after rescaling time the remaining hazard is at most $\ln 2$. Three exceptional events each have conditional probability at most e^{-y} given $\mathcal{G}_\zeta$:

- E_1 (Ville): $(q^{(j)}/q^{(\zeta)})\exp(-\sum_{\zeta<i\le j} c_i^2)$ is a supermartingale.
- E_2 (optional stopping): the height falls below $q^{(\zeta)}e^{-y}$.
- E_3 (compensator): $\mathbb{E}[\exp(\mathrm{QV}^{\mathrm{sm}}_{>\zeta} - (e-1)A_{>\zeta}) \mid \mathcal{G}_\zeta] \le 1$. Here $\mathrm{QV}^{\mathrm{sm}}_{>\zeta}$ and $\mathrm{QV}^{\mathrm{lg}}_{>\zeta}$ are the sums of c_i^2 over the small and the large reveals $i > \zeta$, and $A_{>\zeta}$ is the hazard-weighted small-bias sum after ζ (the quantity A in the proof of Proposition K.3). Off E_2, the bound of RL1 on the small-bias sum, which includes the frozen coordinates, gives $A_{>\zeta} \le \ln 2 \cdot \max\{\tau_1, \kappa_1(\ln(e/h)+y)\}$.

Off $E_1 \cup E_2 \cup E_3$, $\ln(Q/q^{(\zeta)}) < 2y + \gamma_1 \max\{\tau_1, \kappa_1(\ln(e/h)+y)\} + \mathrm{QV}^{\mathrm{lg}}_{>\zeta}$.

Dichotomy. On $\{Q \ge H, q^{(0)} < h\}$ we have $1 \le \zeta \le m$ and $q^{(\zeta)} \le 2q^{(\zeta-1)} < 2h$, so $\ln(Q/q^{(\zeta)}) > \ln(H/(2h))$. Off the three events, some large reveal occurs strictly after ζ. The first one, k, has $q^{(k-1)} \ge he^{-y}$: if $k-1 = \zeta$ this is $q^{(\zeta)} \ge h$, and otherwise it holds because we are off E_2. So E_{v_k} occurs, which gives (a) with $3e^{-y}$; the lower excursion E_2 is counted once.

Change of measure. On $\mathcal{G}^+_{k-1}$, $d\widetilde{\Pr}/d\widetilde{\Pr}_v = \rho_v/\rho_v(S_{k-1})$, and $\rho_v(S_{k-1}) \ge c_k^2/4 \ge \beta_1^\circ$ on $E_v \cap \{v_k = v\}$. Summing over k gives (c), since v is revealed at most once. Only times strictly before v's value are used. □

Exact enumeration ($n = 6$, all coin configurations) gives worst ratios 0.69, 0.46 and 0.44 for E_1, E_2, E_3 and 0.004 for the dichotomy, each a supremum over every look-ahead-free stopping index; the change-of-measure identity holds to $1.8\cdot 10^{-14}$ for the four value-dependent stopping rules on four families that were tested [*numerical*]. An independent enumeration ($n = 5, 6$) reproduces the identity to $2.6\cdot 10^{-13}$, with worst ratio 0.925 in (c) [*numerical*].

The proof of part (a) gives more than its statement: the exceptional term is paid only on the event that the height reaches h.

Lemma 10.15 (the refined exceptional term; [*proved here, audited*])**.** *Under the hypotheses of Theorem 10.14,*

$$\widetilde{\Pr}\big(Q \ge H,\ q^{(0)} < h,\ \text{no } E_v \text{ occurs}\big) \ \le\ 3e^{-y}\,\widetilde{\Pr}(\zeta < \infty) \ \le\ 6e^{-y}\,\widetilde{\Pr}(Q \ge h/2).$$

Proof. We use Steps 2 and 3 of the proof of Theorem 10.14 in Appendix K without change, and write $\mathcal{X}$ for the exceptional event that Step 2 calls X. Step 2 fixes, for each k, the event

X_k of Proposition K.3 for the start state $\mathcal{G}_k$, with $\widetilde{\Pr}(X_k \mid \mathcal{G}_k) \le 3e^{-y}$ on $\{k \le m\}$, and puts $\mathcal{X} = \bigcup_k (\{\zeta = k\} \cap X_k)$. Step 3 shows $\{Q \ge H,\ q^{(0)} < h\} \cap \{\text{no } E_v \text{ occurs}\} \subseteq \mathcal{X}$. Since $\{\zeta = k\} \in \mathcal{G}_k$ and $\{\zeta = k\} \subseteq \{k \le m\}$, the display of Step 2 with its weights kept reads

$$\widetilde{\Pr}(\mathcal{X}) = \sum_k \mathbb{E}\big[\mathbf{1}(\zeta = k)\, \widetilde{\Pr}(X_k \mid \mathcal{G}_k)\big] \le 3e^{-y} \sum_k \widetilde{\Pr}(\zeta = k) = 3e^{-y}\, \widetilde{\Pr}(\zeta < \infty);$$

Appendix K bounds the last sum by 1. This is the first inequality. For the second, $\{\zeta < \infty\} = \{\max_{0 \le k \le m} q^{(k)} \ge h\}$, where m is random. Apply Lemma 10.13 with $K = 2$ to ψ under $\widetilde{\Pr}$ and to the filtration $(\mathcal{G}_t)_{t=0,1,\dots,n}$ of the exploration, where $\mathcal{G}_t$ is the terminal σ-field for $t > m$. It reads x only on F, and its terminal σ-field contains $\sigma(F, x_F)$ because no coordinate of F is frozen (Remark K.4). Its $q_t = \mathbb{E}_U[\psi^2 \mid \mathcal{G}_t]$ is $q^{(t)}$ for $t \le m$, because the revealed set is a function of the clocks and x is independent of the clocks under $U \otimes$ clocks. Hence $\widetilde{\Pr}(\zeta < \infty) \le \widetilde{\Pr}(\max_t q_t \ge h) \le 2\, \widetilde{\Pr}(Q \ge h/2)$. □

This is the same use of Lemma 10.13, on the same filtration, as in the proof of Theorem 10.16 below, applied to ψ itself instead of $d_v\psi$.

Theorem 10.16 (deep-node capture on the grid, DNC; [*proved here, audited*])**.** *Let $c \in \mathcal{C}_s$ be open, and let $0 < h < H \le 1$ and $y > 0$ satisfy $\ln(H/(2h)) \ge 2y + \gamma_1 \max\{\tau_1, \kappa_1(\ln(e/h) + y)\}$. Then*

$$\widetilde{\Pr}_c(Q^c \ge H) \ \le\ \widetilde{\Pr}_c(\mathrm{cap}_c(h)) + 6e^{-y}\, \widetilde{\Pr}_c(Q^c \ge h/2) + \beta_1^{-1} \sum_{v:\, c\cdot v \in \mathcal{C}_{s+1}} \rho_v^c\, \widetilde{\Pr}_{c\cdot v}\big(Q^{c\cdot v} \ge \beta_1 h e^{-y}\big).$$

The sum runs over the $v \notin \mathcal{H}_c \cup c$ with $\rho_v^c > 0$, for which the law $\widetilde{\Pr}_{c\cdot v}$ is defined.

Proof. Theorem 10.14 applies at $(\psi_c, \mathcal{H}_c, h, H, y)$: ψ_c is multilinear of degree at most d with $w_c > 0$ (Lemma 10.7(a)), $\mathcal{H}_c$ is a fixed set (Lemma 10.7(c)), $\tau_1 \le 1/(2e^3 d^4)$ (Lemma 10.24), and the condition on (h, H, y) is assumed. Now

$$\{Q^c \ge H\} \subseteq \mathrm{cap}_c(h) \cup \big(\{Q^c \ge H,\ q^{(0)} < h\} \cap \{\text{no } E_v^c \text{ occurs}\}\big) \cup \textstyle\bigcup_v E_v^c,$$

and Lemma 10.15 bounds the probability of the middle event by $6e^{-y}\widetilde{\Pr}_c(Q^c \ge h/2)$. Only $v \notin \mathcal{H}_c$ occur, and if $\rho_v^c = 0$ (in particular if $v \in c$) then $d_v\psi_c = 0$ and $E_v^c = \emptyset$ by part (b); so only v with $c\cdot v \in \mathcal{C}_{s+1}$ remain. By part (b) and $d_v\psi_c = \psi_{c\cdot v}$, on E_v^c the child's height in $\mathcal{G}^c$ just before v's value is $\mathbb{E}_U[\psi_{c\cdot v}^2 \mid x_{S_{k-1}}] \ge \beta_1^\circ h e^{-y} = K\beta_1 h e^{-y}$. So $E_v^c \subseteq \{\max_k \mathbb{E}_U[\psi_{c\cdot v}^2 \mid \mathcal{G}_k^c] \ge K\beta_1 h e^{-y}\}$. The exploration $\mathcal{G}^c$ reads x only on F and its terminal σ-field contains $\sigma(F, x_F)$ (Remark K.4), so Lemma 10.13 under $\widetilde{\Pr}_{c\cdot v}$ bounds this event's probability by $2\widetilde{\Pr}_{c\cdot v}(Q^{c\cdot v} \ge \beta_1 h e^{-y})$. With part (c), each child costs $2\rho_v^c/\beta_1^\circ = \rho_v^c/\beta_1$. □

The activation event is not implied by a final-height spike of the child; it is dominated by one, under the child's own law, at a cost of 2 in probability and 2 in height. The parent's exploration is used only to locate the activation. In $\mathcal{G}^c$ every stage-2 coordinate lies outside $\mathcal{H}_c$, so no hypothesis on static capture in an inherited filtration is needed. The exploration $\mathcal{G}^c$ does not depend on (H, h, y), so Theorem 10.16 is applied at one node once for every position k of its row, at $(H, h, y) = (H_{s,k}, h_{s,k}, y)$; each application is a separate use of Theorem 10.14, with its own ζ and its own events E_v^c, and (3) supplies its hypothesis.

10.5. **The mass recursion on the grid and the ledger.** Put $\chi = 6e^{-y}$. For $0 \le s, k \le S-1$ let

$$R_{s,k} = \sum_{c \in \mathcal{C}_s} \beta_1^{-s} \tfrac{1}{2} w_c \, \widetilde{\Pr}_c(Q^c \ge H_{s,k}), \qquad \mathrm{CL}_{s,k} = \sum_{c \in \mathcal{C}_s \text{ closed}} \beta_1^{-s} \tfrac{1}{2} w_c \, \widetilde{\Pr}_c(Q^c \ge H_{s,k}),$$

$$\mathrm{CAPT}_{s,k} = \sum_{c \in \mathcal{C}_s \text{ open}} \beta_1^{-s} \tfrac{1}{2} w_c \, \widetilde{\Pr}_c(\mathrm{cap}_c(h_{s,k})), \qquad N_s = \tfrac{1}{2} \beta_1^{-s} d^{s+1} V,$$

and set $R_{s,S} = N_s$ and $R_{S,k} = 0$.

Proposition 10.17 (mass recursion on the grid; [*proved here, audited*])**.** *Suppose every node of $\mathcal{C}_{S-1}$ is closed. Then $R_{0,0} = \sum_j \mu_j$, and for all $0 \le s, k \le S-1$,*

$$R_{s,k} \;\le\; \mathrm{CL}_{s,k} + \mathrm{CAPT}_{s,k} + \chi R_{s,k+1} + R_{s+1,k}.$$

Proof. $R_{0,0} = \sum_j \mu_j$ is Lemma 10.7(e), with $\mu_j = 0$ when $I_j = 0$. Closed nodes contribute $\mathrm{CL}_{s,k}$ to $R_{s,k}$, and open nodes occur only at levels $s \le S-2$. For an open $c \in \mathcal{C}_s$ apply Theorem 10.16 at $(H, h, y) = (H_{s,k}, h_{s,k}, y)$, which is admissible by (G0). In its middle term, $\widetilde{\Pr}_c(Q^c \ge h_{s,k}/2) \le \widetilde{\Pr}_c(Q^c \ge H_{s,k+1})$ for $k < S-1$ by (G1); for $k = S-1$ we bound this probability by 1. In its last term, $\widetilde{\Pr}_{c\cdot v}(Q^{c\cdot v} \ge \beta_1 h_{s,k} e^{-y}) \le \widetilde{\Pr}_{c\cdot v}(Q^{c\cdot v} \ge H_{s+1,k})$ by (G2). Multiply by $\beta_1^{-s}\frac{1}{2} w_c$ and sum over the open $c \in \mathcal{C}_s$. The first terms give $\mathrm{CAPT}_{s,k}$. The middle terms give at most $\chi R_{s,k+1}$ if $k < S-1$, since adding the closed nodes only enlarges the sum, and at most $\chi \sum_{c \in \mathcal{C}_s} \beta_1^{-s}\frac{1}{2} w_c \le \chi N_s = \chi R_{s,S}$ if $k = S-1$, by Lemma 10.7(d). In the last terms, $\beta_1^{-s}\beta_1^{-1}\frac{1}{2} w_c \rho_v^c = \beta_1^{-(s+1)}\frac{1}{2} w_{c\cdot v}$ by Lemma 10.7(b), and each $c' \in \mathcal{C}_{s+1}$ is $c \cdot v$ for exactly one pair (c, v), so they sum to $R_{s+1,k}$. At $s = S-1$ every node is closed, so $R_{S-1,k} = \mathrm{CL}_{S-1,k}$. □

Corollary 10.18 (unrolling; [*proved here, audited*])**.** *Under the hypothesis of Proposition 10.17,*

$$\sum_j \mu_j \;\le\; \sum_{s=0}^{S-1} \sum_{k=0}^{S-1} \binom{s+k}{k} \chi^k \big(\mathrm{CL}_{s,k} + \mathrm{CAPT}_{s,k}\big) + \sum_{s=0}^{S-1} \binom{s+S-1}{S-1} \chi^S N_s.$$

Proof. Write $E_{s,k} = \mathrm{CL}_{s,k} + \mathrm{CAPT}_{s,k}$, and for $0 \le s, k \le S-1$ let

$$U_{s,k} = \sum_{s'=s}^{S-1} \sum_{k'=k}^{S-1} \binom{s'-s+k'-k}{k'-k} \chi^{k'-k} E_{s',k'} + \sum_{s'=s}^{S-1} \binom{s'-s+S-1-k}{S-1-k} \chi^{S-k} N_{s'},$$

with $U_{s,S} = N_s$ and $U_{S,k} = 0$. We claim $U_{s,k} = E_{s,k} + \chi U_{s,k+1} + U_{s+1,k}$ for $0 \le s, k \le S-1$. Write $a = s' - s$. The coefficient of $E_{s',k'}$ on the right, with $b = k' - k$, is $\mathbf{1}(a = b = 0) + \mathbf{1}(b \ge 1)\binom{a+b-1}{b-1}\chi^b + \mathbf{1}(a \ge 1)\binom{a+b-1}{b}\chi^b$, which is $\binom{a+b}{b}\chi^b$ by Pascal's rule (with $\binom{n}{-1} = \binom{n}{n+1} = 0$). The coefficient of $N_{s'}$ on the right, with $n = S-1-k$, is $\chi \cdot \mathbf{1}(a = 0)$ from $\chi U_{s,S} = \chi N_s$ if $n = 0$, or $\binom{a+n-1}{n-1}\chi^{n+1}$ from $\chi U_{s,k+1}$ if $n \ge 1$, plus $\mathbf{1}(a \ge 1)\binom{a-1+n}{n}\chi^{n+1}$ from $U_{s+1,k}$; in both cases the total is $\binom{a+n}{n}\chi^{n+1}$. All terms are nonnegative, so Proposition 10.17 and induction on (s, k) in decreasing order of $s + k$, starting from $R_{s,S} = U_{s,S}$ and $R_{S,k} = U_{S,k}$, give $R_{s,k} \le U_{s,k}$. At $(s, k) = (0, 0)$ this is the claim. □

Corollary 10.19 (ledger; [*proved here, audited*])**.** *Suppose that every node of $\mathcal{C}_{S-1}$ is closed, that $\mathrm{CAPT}_{s,k} \le 3\varepsilon V/(32S)$ for all $0 \le s, k \le S-1$, that $S \ge 68$ and $\chi \le 1/(4S)$, and that*

$$\textit{(TERM)} \qquad S(y - \ln 6) \;\ge\; (S-1)(\ell + \ln d + \ln 4) + \ln(1024 e S) + 2 \ln d.$$

Then $\sum_j \mu_j \le \frac{\varepsilon V}{96} + \frac{\varepsilon V}{8} + \frac{\varepsilon V}{128} = \frac{55}{384}\varepsilon V < \frac{\varepsilon}{4} V$, which is SM-OF *at (θ, L).*

Proof. For $s \le S-1$, $\sum_{k=0}^{S-1}\binom{s+k}{k}\chi^k \le (1-\chi)^{-(s+1)} \le (1-\chi)^{-S} \le \exp(S\chi/(1-\chi)) \le \exp(\frac{1}{4}\cdot\frac{272}{271}) < \frac{4}{3}$. By Lemma 10.9 and Lemma 10.7(d), $\mathrm{CL}_{s,k} \le \frac{1}{2}\beta_1^{-s}p_*d^{s+1}V = \frac{\varepsilon V}{128S}(\beta_1/d)^{S-1-s} \le \frac{\varepsilon V}{128S}$. So the CL part of Corollary 10.18 is at most $S\cdot\frac{4}{3}\cdot\frac{\varepsilon V}{128S} = \frac{\varepsilon V}{96}$, and the CAPT part is at most $S\cdot\frac{4}{3}\cdot\frac{3\varepsilon V}{32S} = \frac{\varepsilon V}{8}$. In the terminal part, $\binom{s+S-1}{S-1} \le 2^{s+S-1} \le 4^{S-1}$, $\chi^S = 6^S e^{-Sy}$ and $N_s \le \frac{1}{2}\beta_1^{-(S-1)}d^SV$, so it is at most $S\,4^{S-1}6^Se^{-Sy}\,\frac{1}{2}\beta_1^{-(S-1)}d^SV$. Taking logarithms, this is at most $V/(2048ed) = \varepsilon V/128$ exactly when (TERM) holds. Finally $\frac{1}{96}+\frac{1}{8}+\frac{1}{128} = \frac{55}{384} < \frac{1}{4}$. □

The terminal part is the price of truncating each row at $k = S-1$; since χ is small, it is controlled by the single condition (TERM) on y, which does not grow with the depth beyond $\ell + \ln d + O(1)$.

10.6. **Priming for chain families.** For a function g, a set A and $a \in \{-1,1\}^A$ let $q^g(A,a) = \mathbb{E}[g^2 \mid x_A = a]$ and $\pi^g(A,a) = \mathbb{E}_z|g(z^{A\leftarrow a}) - g(z)|$, where $z^{A\leftarrow a}$ overwrites z on A and z is uniform. Lemmas 10.20, 10.21 and 10.22 are unchanged from the previous audited version of the section; they concern chains and blocks only, not the heights or the heavy sets.

Lemma 10.20 (conversion; [*proved here, audited*])**.** *If $|g| \le 1$ and $w = \|g\|_2^2$, then $\pi^g(A,a) \ge q - \sqrt{qw}$ with $q = q^g(A,a)$. In particular $q \ge 4w$ implies $\pi^g(A,a) \ge q/2$. The bound is pointwise in (A,a), so it holds for random blocks.*

Proof. With $X = g(z^{A\leftarrow a})$ and $Y = g(z)$, $|X-Y| \ge |X|(|X|-|Y|)$ because $|X| \le 1$, so $\pi \ge \mathbb{E}X^2 - \mathbb{E}|XY| \ge q - \sqrt{qw}$. □

Lemma 10.21 (priming for chain derivatives; [*proved here, audited*])**.** *Let $A_1,\dots,A_m$ be pairwise disjoint, $c_1,\dots,c_m$ chains with $c_i \cap A_i = \emptyset$ (the chains may repeat, meet other blocks, or have different lengths), and $a_i \in \{-1,1\}^{A_i}$. Then $\sum_i \pi^{\psi_{c_i}}(A_i,a_i) \le \mathbb{E}_y\mathrm{bs}(h,y) \le \mathrm{bs}(h) = 2\,\mathrm{rbs}(h) \le 2d^2$.*

Proof. For a chain $c = (j,u_1,\dots,u_s)$ write $d_c = d_{u_s}\cdots d_{u_1}d_j$, so that $\psi_c = d_ch$. Fix i, put $c = c_i$ and $G(z) = h(z^{A_i\leftarrow a_i}) - h(z)$. Since $A_i \cap c = \emptyset$, d_c commutes with the overwrite on A_i, so $\psi_c(z^{A_i\leftarrow a_i}) - \psi_c(z) = d_cG(z)$, and $|d_cG(z)| \le \mathbb{E}_\sigma|G(z^{c\leftarrow\sigma})|$ with σ uniform on $\{-1,1\}^c$ and $y = z^{c\leftarrow\sigma}$ uniform. So $\pi^{\psi_{c_i}}(A_i,a_i) \le \mathbb{E}_y|h(y^{A_i\leftarrow a_i}) - h(y)|$, and for fixed y the disjoint blocks are admissible in $\mathrm{bs}(h,y)$. The identity $\mathrm{bs} = 2\,\mathrm{rbs}$ is the normalisation of Section 9.2, and $\mathrm{rbs}(h) \le d^2$ is Lemma 9.3, that is, Lemma 7.7(a) [*known*]. □

Without the citation, the proof of Lemma 9.3 gives $\mathrm{rbs}(h) \le 4d^2$, hence $\mathrm{bs}(h) \le 8d^2$. This multiplies the first bound on $W_{s,k}$ in the proof of Corollary 10.23 by 4 and the second by 2, which adds $\ln 4$ to the intercept of (MW″) and $2\ln 4$ to that of (LD″). By Proposition 10.25, $\ln P_d$ exceeds every intercept by more than $2.7\cdot 10^5\,T^3 > 2\ln 4$, so the proposition and its constants are unchanged [*proved here, audited*].

Lemma 10.22 (averaged priming for chain families; [*proved here, audited*])**.** *Let $\mathcal{C}$ be a finite family of chains with weights $M_c > 0$. For each $c \in \mathcal{C}$ let ν_c be a sub-probability law on pairs (A,a) with $A \cap c = \emptyset$ and $\pi^{\psi_c}(A,a) \ge \pi_*$ ν_c-almost surely. With $W = \sum_c M_c\nu_c(1)$ and $\Lambda_u = \sum_c M_c\nu_c(u \in A)$,*

$$\mathrm{bs}(h) \;\ge\; \pi_*\min\Big\{\frac{W^2}{4\sum_u\Lambda_u^2},\ \frac{W}{2\max_c M_c}\Big\}.$$

Proof. Select each c independently with probability $\eta M_c\nu_c(1)$ and draw $(A_c,a_c) \sim \nu_c/\nu_c(1)$. The expected number of selected c whose block meets another selected block is at most $\eta^2\sum_u\Lambda_u^2$, and the remaining ones satisfy Lemma 10.21. So $\mathrm{bs}(h) \ge \pi_*(\eta W - \eta^2\sum_u\Lambda_u^2)$; optimise $\eta \le 1/\max_c M_c$. This is Lemma 9.7 with weights M_c, in which only the sets A need to be disjoint. □

Corollary 10.23 (captured mass at (s,k); [*proved here, audited*])**.** *Let $s \le S-1$. Suppose $\ln A_c > \lambda_0$ for every $c \in \mathcal{C}_s$, where $\lambda_0 = \ln L - (Z_0^* - X)$, and $\lambda_0 \ge D_s^* + \ln 4$. Then $\mathrm{CAPT}_{s,k} \le 3\varepsilon V/(32S)$ for every $k \le S-1$, provided*

$$(MW'')\quad \ln M \ge (Z_0^* - X) + D_s^* + s\ell + \ln\frac{512Sd^2}{3\varepsilon} \qquad \text{and} \qquad (LD'')\quad M^{1-2\alpha} \ge \Big(\frac{64Sd^{s+2.5}}{3\varepsilon\beta_1^s}\Big)^2 \frac{\theta}{h_s^*}.$$

Proof. Fix k and apply Lemma 10.22 to the open $c \in \mathcal{C}_s$, with $M_c = w_c > 0$ and ν_c the $\widetilde{\Pr}_c$-law of $(\mathcal{H}_c \cap F, x_{\mathcal{H}_c \cap F})$ restricted to $\mathrm{cap}_c(h_{s,k})$. This is a sub-probability law, and its blocks avoid c by Lemma 10.7(c). Then $W = W_{s,k} := \sum_{c \text{ open}} w_c \widetilde{\Pr}_c(\mathrm{cap}_c(h_{s,k}))$ and $\mathrm{CAPT}_{s,k} = \frac{1}{2}\beta_1^{-s} W_{s,k}$.

Gain. $w_c = H_s^*/A_c < H_s^* e^{-\lambda_0} \le H_s^* e^{-D_s^*}/4 = h_s^*/4 \le h_{s,k}/4$, by (G3). On $\mathrm{cap}_c(h_{s,k})$ the block height is $q^{\psi_c}(\mathcal{H}_c \cap F, x_{\mathcal{H}_c \cap F}) = q^{(0)} \ge h_{s,k} \ge 4w_c$, so Lemma 10.20 gives $\pi^{\psi_c} \ge q^{(0)}/2 \ge h_s^*/2$. We use the weaker $\pi_* = h_s^*/8$ of Lemma 9.5, the same for every k. Also $\max_c M_c \le H_s^* e^{-\lambda_0}$.

Loads. Under $\widetilde{\Pr}_c$, $u \in F$ with probability $\frac{1}{2}$ independently of x, so $\Lambda_u \le \frac{1}{2}\sum_{c:\, u \in \mathcal{H}_c} w_c$. Now $u \in \mathcal{H}_c$ means $w_c \le M^\alpha \mathrm{Inf}_u(\psi_c)$, and $\sum_{c \in \mathcal{C}_s} \mathrm{Inf}_u(\psi_c) \le d^{s+1}\mathrm{Inf}_u(h)$, because a Fourier set $J \ni u$ contains at most d^{s+1} chains of length $s+1$ avoiding u. Hence $\Lambda_u \le \frac{1}{2}M^\alpha d^{s+1}\mathrm{Inf}_u(h)$, and with $\sum_u \mathrm{Inf}_u(h) \le dV$ and $\mathrm{MaxInf}(h) < \theta/L$,

$$\sum_u \Lambda_u^2 \;\le\; \tfrac{1}{4}M^{2\alpha}d^{2s+2}\mathrm{MaxInf}(h)\sum_u \mathrm{Inf}_u(h) \;<\; \tfrac{1}{4}M^{2\alpha}d^{2s+3}\,\theta V/L.$$

Conclusion. With $\mathrm{bs}(h) \le 2d^2$ (Lemma 10.21), Lemma 10.22 gives

$$W_{s,k} \;\le\; \frac{2\,\mathrm{bs}(h)\max_c M_c}{\pi_*} \;\le\; 32d^2 e^{D_s^* - \lambda_0} \qquad \text{or} \qquad W_{s,k}^2 \;\le\; \frac{4\,\mathrm{bs}(h)\sum_u \Lambda_u^2}{\pi_*} \;<\; \frac{16M^{2\alpha}d^{2s+5}\theta V}{Lh_s^*}.$$

In the first case, $\frac{1}{2}\beta_1^{-s}W_{s,k} \le 3\varepsilon V/(32S)$ follows from $\lambda_0 \ge \ln(512Sd^2/(3\varepsilon)) + D_s^* + s\ell + \ln(1/V)$, which is (MW″), since $\ln L = \ln M + \ln(1/V)$. In the second case, squaring shows that $\frac{1}{2}\beta_1^{-s} \cdot 4M^\alpha d^{s+2.5}(\theta V/(Lh_s^*))^{1/2} \le 3\varepsilon V/(32S)$ is equivalent to $(64Sd^{s+2.5}/(3\varepsilon\beta_1^s))^2 M^{2\alpha}\theta/h_s^* \le LV = M$, that is, to (LD″). □

The load factor is $M^{2\alpha}$, and M is the quantity being bounded below, so in (LD″) it moves to the left-hand side as $M^{1-2\alpha} = M^{1/2}$ and costs a factor 2 in $\ln M$, but no θ-exponent.

10.7. **Parameters and the final arithmetic.** Take

$$\alpha = \frac{1}{4}, \qquad c = \frac{1}{2T}, \qquad S = 16T + 4, \qquad \varepsilon_1 = \frac{1}{150T^2}, \qquad \ln\frac{1}{\tau_1} = 3 + 4\ln d + \frac{10e\gamma_1}{\varepsilon_1},$$
$$y = \ell + \ln d + \ln 24 + t_S, \qquad t_S = \frac{\ln(1024eS) + 2\ln d}{S}, \qquad \ln P_d = 3.0 \cdot 10^6\, T^3,$$

so that $\ell = 2\ln(1/\tau_1) + \ln 8 = K_\ell T^2 + 6 + \ln 8 + 8\ln d$ with $K_\ell = 3000e\gamma_1 = 9712.602\ldots$

Lemma 10.24 (the parameters; [*proved here, audited*])**.** (P1) *$\gamma_1\kappa_1 = \varepsilon_1$ and $\tau_1 \le 1/(2e^3d^4)$; also $\gamma_1\tau_1 < 1$ and $y + \ell \ge \ln 2$.* (P2) *(TERM) of Corollary 10.19 holds.* (P3) *$S \ge 68$ and $\chi = 6e^{-y} \le 1/(4S)$.*

Proof. (P1) $\kappa_1 = 10e/(\ln(1/\tau_1) - 3 - 4\ln d) = \varepsilon_1/\gamma_1$, and $\ln(1/\tau_1) \ge 3 + \ln 2 + 4\ln d = \ln(2e^3d^4)$ because $10e\gamma_1/\varepsilon_1 > \ln 2$. Then $\gamma_1\tau_1 < \gamma_1/(2e^3) < 1$, and $y + \ell \ge \ell > \ln 2$. (P2) $S(y - \ln 6) = S(\ell + \ln d + \ln 4) + St_S = S(\ell + \ln d + \ln 4) + \ln(1024eS) + 2\ln d$, which exceeds the right side of (TERM) by $\ell + \ln d + \ln 4 > 0$. (P3) $S = 16T + 4 \ge 68$, and $y \ge \ell > K_\ell T^2$ with $24S = 24(16T + 4) < e^{K_\ell T^2}$, so $4S\chi < 1$. □

Proposition 10.25 (the arithmetic; [*proved here for every $d \geq 2$, Appendix L; audited*])**.** *If $\ln M \geq cX + \ln P_d$ with $\ln P_d = 3.0 \cdot 10^6\, T^3$, then the following hold.*

- *(G″) of Corollary 10.11: $\alpha \ln M \geq B_s^*$ for every $s \leq S-2$.*
- *(CV″) $\ln M \geq (Z_0^* - X) + D_s^* + \ln 4$ for every $s \leq S-1$.*
- *(MW″) and (LD″) of Corollary 10.23, for every $s \leq S-1$.*
- *(CL″) of Corollary 10.12.*

These hold for every $d \geq 2$.

Proof. Exact forms. By Lemma 10.5, with $\ln(1/\varepsilon) = \ln(16e) + \ln d$, $\ln(1/p_*) = \ln(64Sd/\varepsilon) + (S-1)(\ell + \ln d)$ and $1 + (S-1)\alpha = 4T + \frac{7}{4}$, each condition reads $\ln M \geq mX + \Gamma$ with m and Γ free of θ:

- (G″): $m = 4\varepsilon_1 r^{s+S}$, $\Gamma = 4r^{s+S-1}b_1$;
- (CV″): $m = r^{S-1} - 1 + \varepsilon_1 r^{s+S}$, $\Gamma = \omega_{S-1} + (r-1)(r^{S-1}\sigma_s + \omega_{S-1}) + rb_0 + \ln 4$;
- (MW″): the same m, and $\Gamma = \omega_{S-1} + (r-1)(r^{S-1}\sigma_s + \omega_{S-1}) + rb_0 + s\ell + \ln(512Sd^2/(3\varepsilon))$;
- (LD″): $m = 2(r^{s+S} - 1)$, $\Gamma = 2\big[2\ln(64Sd^{s+2.5}/(3\varepsilon)) + 2s\ell + r(r^{S-1}\sigma_s + \omega_{S-1}) + rb_0\big]$;
- (CL″): $m = (3r^{2S-2} - 1)/(4T + \frac{7}{4})$, $\Gamma = 3\big(r^{S-1}\sigma_{S-1} + \omega_{S-1} + \ln(1/p_*)\big)/(4T + \frac{7}{4})$.

In the first four, m and Γ are nondecreasing in s, because $r > 1$, σ_s increases and $d \geq 2$; so it suffices to take $s = S-1$, and $s = S-2$ for (G″). Two of these requirements are redundant and are kept only for clarity: (MW″) implies (CV″), since $s\ell + \ln(512Sd^2/(3\varepsilon)) \geq \ln 4$; and capture is needed only for $s \leq S-2$, since every node of $\mathcal{C}_{S-1}$ is closed. Since $X > 0$, a condition holds for every admissible θ once $m \leq c$ and $\Gamma \leq \ln P_d$; the conditions are separate lower bounds on M, so $\ln P_d$ must dominate their maximum.

Slopes. Write $u = 1/T \leq \frac{1}{4}$. Since $\varepsilon_1 = u^2/150$, Appendix L (Lemma L.3) gives $r^n \leq 1 + 0.2317939\, u$ for $n \leq 2S-1$ and $r^{S-1} \leq 1 + 0.1133175\, u$. Hence:

- (LD″): $Tm \leq 2 \cdot 0.2317939 = 0.4635878 < \frac{1}{2} = Tc$;
- (CL″): $m \leq (2 + 3 \cdot 0.2317939\, u)/(4T + \frac{7}{4})$, which is at most c because $6 \cdot 0.2317939 = 1.3907634 \leq \frac{7}{4}$;
- (MW″) and (CV″): $Tm \leq 0.1133175 + (1 + 0.2317939/4)/600 \leq 0.1150808$;
- (G″): $Tm \leq 4(1 + 0.2317939/4)/600 \leq 0.0070530$.

Intercepts. Appendix L proves, for every $T \geq 4$ and every d with $2ed \leq 2^T$,

$$\Gamma_{\mathrm{LD}} \leq 16K_\ell T^3\Big(16 + \frac{6.196076}{T}\Big) \leq 2.727147 \cdot 10^6\, T^3, \qquad \Gamma_{\mathrm{CV}} \leq \Gamma_{\mathrm{MW}} \leq 5.125043 \cdot 10^5\, T^3,$$

$$\Gamma_{\mathrm{G}} \leq 4.112130 \cdot 10^4\, T^3, \qquad \Gamma_{\mathrm{CL}} \leq 2.242490 \cdot 10^5\, T^3.$$

Each bound is T^3 times a polynomial in u with nonnegative coefficients, evaluated at $u = \frac{1}{4}$ (for Γ_{LD}, after the chord bound of Lemma L.4) [*computer-assisted: a finite evaluation, printed in Appendix L*]. The main contribution to Γ_{LD} is $2r(r^{S-1}\sigma_{S-1} + \omega_{S-1}) + 4(S-1)\ell \approx 2 \cdot 6S\ell + 4S\ell = 16S\ell$, which gives $256K_\ell T^3$. So every slope is below c and every intercept is below $3.0 \cdot 10^6\, T^3$, for every $d \geq 2$. Exact evaluation at the worst d ($\ln d = T\ln 2 - \ln(2e)$) for $T = 4, \ldots, 40$ and $T \in \{50, 64, 100, 128, 200, 256, 500, 1000, 2000\}$ agrees: the needed constant falls from $2.679 \cdot 10^6$ at $T = 4$ to $2.487 \cdot 10^6$ at $T = 2000$ [*numerical*]. □

Theorem 10.26 (SM-OF at a quasi-polynomial, θ-mild loss; [*proved here, audited; components as tagged*])**.** SM-OF(L_0) *holds with $L_0 = P_d\, V^{-1}\theta^{-1/(2T)}$ and $\ln P_d = 3.0 \cdot 10^6\, T^3$.*

The θ-exponent $1/(2T)$ is that of a θ-free loss, but the prefactor P_d is quasi-polynomial in d, whereas a θ-free loss in the sense of Section 9.5 has prefactor $(d/V)^{O(1)}$. SM-OF at a θ-free loss in that sense remains open [*open*].

d	T	S	worst slope·T	(LD″) slope·T	needed $\ln P_d/T^3$
2	4	68	-0.0105 (CL″)	-0.0370	$2.679\cdot10^6$
10^2	10	164	-0.0052 (CL″)	-0.0592	$2.562\cdot10^6$
10^8	30	484	-0.0019 (CL″)	-0.0687	$2.511\cdot10^6$
10^{32}	109	1748	-0.0005 (CL″)	-0.0721	$2.493\cdot10^6$
10^{128}	428	6852	-0.0001 (CL″)	-0.0730	$2.488\cdot10^6$

TABLE 5. Exact evaluation of the constraints of Proposition 10.25 at sample degrees (slope relative to c, i.e. $(m-c)T$; the needed constant is the largest intercept, always that of (LD″)) [*numerical;* `supplement/qpg_table.py`].

Proof. Let h be old-free capped at (θ, L) with $L \ge L_0$, and let X and M be as at the end of Section 10.1; then $\ln M \ge cX + \ln P_d$. Proposition 10.25 gives (G″), (CV″), (MW″), (LD″) and (CL″), and Lemma 10.24 gives (P1)–(P3), hence the hypotheses of Lemma 10.5. By Corollary 10.11 and (G″), $\ln A_c > \lambda_s \ge \lambda_0$ for every $c \in \mathcal{C}_s$, $s \le S-1$. By Corollary 10.12 and (CL″), every node of $\mathcal{C}_{S-1}$ is closed and $\mathcal{C}_S = \emptyset$. Since $\ln L \ge \ln M$, (CV″) gives $\lambda_0 \ge D_s^* + \ln 4$, so Corollary 10.23 with (MW″) and (LD″) gives $\mathrm{CAPT}_{s,k} \le 3\varepsilon V/(32S)$ for all $s, k \le S-1$. Proposition 10.17 (Theorem 10.16 at every open node and every position of its row), Corollary 10.18 and Corollary 10.19, with (P2) and (P3), give $\sum_j \mu_j < \frac{\varepsilon}{4}V$. □

Proof of Theorem 10.1. By Theorem 10.26 and Proposition 10.3 (with $c = 1/(2T) \le 1/8$ and $P_d \ge 1$), (TT*) holds with $C = 2$ and $\ln P^* \le \ln P_d + \ln(1/v) + 6.4\ln d + 23.2$. Lemma 10.2 with $cT = \frac{1}{2}$ gives the bound, with $6.6 = 1.65\cdot 4$. For the expansion, $1.30\cdot 3.0\cdot 10^6 = 3.9\cdot 10^6$, $1.30\cdot 6.4 \le 8.4$, $1.30\cdot 23.2 \le 30.2$, $6.6\cdot 8.76 \le 57.9$ and $6.6\ln c_2 \le 6.6\ln(6.6\cdot 10^7) < 119$. Finally $T \le \log_2(4ed)$, so the first term is $O(\log^4 d)$ and the others are $O(\log d\cdot\ln(d/v))$. □

Remark 10.27 (why $\log^4 d$, and what would give $\log^3 d$; [*heuristic; not a lower bound for SM-OF*]). In this scheme four constraints force $\ln P_d = \Theta(T^3)$. (a) *Depth.* The slope of (CL″) is about $2/(1+(S-1)\alpha)$ and must be at most c, so $S = \Theta(T)$. (b) *Drift, in two places.* Both come from the term $\varepsilon_1\ln(e/h)$ in (3), where $\ln(1/h) \ge X$. The factor θ/h_s^* of (LD″) gives the slope $2(r^{2S-1}-1) \approx 4S\varepsilon_1$, and the root loss $Z_0^* - X$ gives (CV″) and (MW″) the slope $r^{S-1}-1 \approx S\varepsilon_1$; each must be at most c, so $\varepsilon_1 = O(1/T^2)$. (c) *Large-jump threshold.* Theorem 9.14 then forces $\ell \ge 20e\gamma_1/\varepsilon_1 = \Theta(T^2)$. (d) *Heights.* Along the bottom row, $Z_{S-1}^* - X \approx (S-1)(b_1+b_2) \approx 6S\ell$, and the (LD″) intercept is about $16S\ell = \Theta(T^3)$. The chain of Lemma 10.2 multiplies by T, which gives $T^4 = \Theta(\log^4 d)$. The previous version of this scheme measured lightness relative to the node's amplification, which forced $S = \Theta(T\log T)$, and used a level-dependent exceptional exponent $y_s \approx s\ell$, which made the heights fall like $S^2\ell$; the heavy threshold $M^{-1/4}$ and the grid remove both.

To reach $\log^3 d$ one needs $\varepsilon_1 = \Theta(1/T)$: then $\ell = \Theta(T)$ and every intercept is $O(S\ell) = O(T^2)$. This needs both θ-slopes of (b) removed. A Theorem 9.14 whose drift term is relative to the node would remove both. A capture bound for chain families that does not go through $\mathrm{MaxInf}(h) < \theta/L$, and so has no factor θ/h_s^*, would remove only the first: the root loss $Z_0^* - X$ would still force $\varepsilon_1 = O(1/T^2)$. Both statements are open [*open*], and neither is used in this section. The root loss is paid because a node of depth s is examined at the bottom-row height H_s^*, while the root mass μ_j needs only $H_{0,0} = \theta$. On the staircase $\{k \le s\}$ of the grid, with a node of depth s examined at the diagonal height $H_{s,s}$, the root loss would vanish, and a capture bound measured against $H_{s,s}$ rather than through $\mathrm{MaxInf}(h) < \theta/L$ (also open) would alone give $\ln P_d = O(T^2)$, hence $\log^3 d$.

10.8. **Status.** The cited inputs carry these tags:

- Theorem 8.4, Corollary 8.5, Lemma 8.3 and Lemma I.9: [*proved here*]; audited and re-derived independently.
- Lemma J.20 and Corollary J.21: [*proved here, audited*], re-audited and re-derived independently.
- Theorem 9.14 (full proof: Theorem J.14) with RL1, Lemmas 9.5, 9.7 and 9.20, and Corollary 9.2: [*proved here, audited*]; Theorem 9.14 and RL1 were also re-audited and re-derived independently.
- Theorem 10.14 with its proof in Appendix K, including Steps 2 and 3, which Lemma 10.15 reads without change, and Remark K.4: [*proved here, Appendix K; audited*].
- Theorem 9.15, for context only (the node explorations use Proposition K.3 directly): [*proved here, Appendix K; audited*].
- Lemma 9.3, that is, Lemma 7.7(a): [*known*]; Lemma 10.21 records the citation-free alternative.

Lemma 9.27 is used in no proof of this section. Proposition 9.17, whose statement assumes a polynomial prefactor, is replaced by Lemma 10.2 and Proposition 10.3. Theorem 9.23 and Lemma 9.6 are not needed: the small-jump piece is the exceptional term of Theorem 10.16, whose θ-cost sits in (3), and every captured piece is bounded by Lemma 10.22.

Reused without change of statement or proof from the previous audited version of this section: Lemma 10.2, Proposition 10.3, Lemma 10.8 (its general part), Lemma 10.13, Theorem 10.14 (except its substitution clause, which now names only $(\psi, \mathcal{H})$, since (h, H, y) are free parameters of the exploration paragraph), Lemmas 10.20, 10.21 and 10.22, and parts (a), (b), (d) of Lemma 10.7 [*proved here, audited*]. Their hypotheses hold here as noted where they are used: the node functions are bounded multilinear functions of degree at most d; every heavy set is a fixed set, given h and L; every exploration reads x only on F and freezes no coordinate of F; and the blocks of every captured piece avoid their chains. New or modified, [*proved here, audited*]: Definitions 10.4 and 10.6, Lemma 10.5, Lemma 10.7(c) and the form of (e), Lemmas 10.9 and 10.10, Corollaries 10.11 and 10.12, Lemma 10.15, Theorem 10.16, Remark K.5 (items 2 and 3, rewritten for the grid), Proposition 10.17, Corollaries 10.18, 10.19 and 10.23, Lemma 10.24, Proposition 10.25 with Appendix L, Theorem 10.26 and Theorem 10.1. The chain needs two conditions, both met here: the loss in step t is the same function of θ_t for every kept restriction, which the floor δv in (TT*) supplies; and SM-OF is used uniformly in θ, which Proposition 10.25 provides because every slope is at most c. The boundary case $L = L_0$ is not needed: $\mathrm{Var}h > \varepsilon \mathrm{Var}g$ gives strict slack in Proposition 10.3. Lean covers the constant-loss chain and Lemma I.9 only.

Audit status. The mathematics of every step listed above as new or modified was audited twice, independently and with agreeing verdicts, in the form of the working notes from which this text is transcribed. This text was then audited twice, independently: once against the notes, as a transcription, and once as a referee reads it; each of these two audits recomputed every displayed number of Appendix L. Both confirmed the text with fixes, which are made in it, in Appendix K and in the passages of other sections that quote this section. All four audits found Theorem 10.1 proved, with no mathematical error. The slope margin of (CL$''$) is rigorous but thin: $(c - m)/c = 0.021$ at $T = 4$, and about $0.12/T$ for large T [*numerical; Table 5*]. Any change of S, α, ε_1 or the grid must re-check (CL$''$) first.

10.9. **Classical simulation of quantum query algorithms.** The motivation of Conjecture 1.1 in [2] is the classical simulation of quantum query algorithms on most inputs. If a quantum algorithm makes t queries to $X \in \{0,1\}^n$, its acceptance probability $p(X)$ is a real multilinear polynomial of degree at most $2t$ with values in $[0,1]$ [9]; with $x_i = 1 - 2X_i$ and

$f = 2p - 1$ this gives $f \in \mathcal{B}_{2t}(n)$. In this subsection t is the number of quantum queries and $T = \lceil \log_2(2ed) \rceil$ keeps its meaning from Theorem 10.1, so $T = \lceil \log_2(4et) \rceil$ for $d = 2t$. Here ε is the accuracy, δ the failure fraction, p the acceptance probability and τ a stopping time; the $\varepsilon = 1/(16ed)$, $\delta = 1/(128e^2d^2)$ and $p = 2^{-T}$ of Sections 10.1–10.8 do not appear. The reduction from influences to classical query algorithms is the greedy algorithm of [2, §3.1, proof of Theorem 22 of the arXiv version]: while the restricted function has variance above a threshold, query a variable whose influence exceeds the guaranteed lower bound (we take one of maximal influence). We need it with explicit parameters; [24, Theorem 3.4] states a quantitative version of the same reduction.

Lemma 10.28 (greedy reduction; [*known* [2, §3.1]*; this quantitative form proved here, audited*])**.** *Let $d \geq 1$, $f \in \mathcal{B}_d(n)$, $v_0 > 0$ and $\gamma > 0$, and suppose that* $\mathrm{MaxInf}(h) \geq \gamma$ *for every restriction h of f (any set of coordinates fixed to any values) with* $\mathrm{Var} h > v_0$*. Let $\delta \in (0, 1]$ and let N be an integer with $N \geq 2I(f)/(\delta\gamma)$; since $I(f) \leq d\mathrm{Var} f \leq d$, $N = \lceil 2d/(\delta\gamma) \rceil$ will do. Then there is a deterministic adaptive classical algorithm that makes at most N queries to $x \in \{-1, 1\}^n$ and outputs a number $A(x)$ such that, for x uniform on $\{-1, 1\}^n$,*

$$\Pr_x \big[|A(x) - f(x)| > a \big] \;\leq\; \frac{\delta}{2} + \frac{v_0}{a^2} \qquad \text{for every } a > 0.$$

Proof. Write f_ρ for the restriction of f by a partial assignment ρ. The algorithm keeps the assignment ρ_k of the coordinates read so far, with $\rho_0 = \emptyset$. At step k it stops if $\mathrm{Var} f_{\rho_k} \leq v_0$ or $k = N$, and outputs $A(x) = \mathbb{E} f_{\rho_k}$, the mean over the unread coordinates. Otherwise it reads a coordinate i_k maximizing $\mathrm{Inf}_i(f_{\rho_k})$, with ties broken by index, and extends ρ_k by the value x_{i_k} to ρ_{k+1}; since f_{ρ_k} is not constant, i_k is unread. The choice of i_k depends only on the values already read, so the algorithm is a decision tree of depth at most N, determined by f.

For $g \in \mathcal{B}_d$, a coordinate i and $b \in \{-1, 1\}$, let $g_{i,b}$ be the restriction $x_i = b$. For $S \not\ni i$, $\widehat{g_{i,b}}(S) = \widehat{g}(S) + b\,\widehat{g}(S \cup \{i\})$, so for $j \neq i$ and b uniform

$$\mathbb{E}_b \mathrm{Inf}_j(g_{i,b}) = \sum_{S \ni j,\, S \not\ni i} \big(\widehat{g}(S)^2 + \widehat{g}(S \cup \{i\})^2 \big) = \mathrm{Inf}_j(g), \qquad \text{hence} \qquad \mathbb{E}_b I(g_{i,b}) = I(g) - \mathrm{Inf}_i(g).$$

Let x be uniform, let τ be the first k with $\mathrm{Var} f_{\rho_k} \leq v_0$ (it exists, since $\mathrm{Var} f_\rho = 0$ once every coordinate is read), and put $\Phi_k = I(f_{\rho_{\min(k,\tau,N)}})$. If $k < \min(\tau, N)$, the bit x_{i_k} is unread, hence uniform and independent of ρ_k, and the identity gives $\mathbb{E}[\Phi_{k+1} \mid \rho_k] = \Phi_k - \mathrm{MaxInf}(f_{\rho_k})$; if $k \geq \min(\tau, N)$ then $\Phi_{k+1} = \Phi_k$. Summing over $k < N$,

$$\mathbb{E} \sum_{k < \min(\tau, N)} \mathrm{MaxInf}(f_{\rho_k}) \;=\; \Phi_0 - \mathbb{E}\Phi_N \;\leq\; I(f).$$

For $k < \tau$ we have $\mathrm{Var} f_{\rho_k} > v_0$, so each term is at least γ by hypothesis, and $\gamma\,\mathbb{E}\min(\tau, N) \leq I(f)$. By Markov's inequality, $\Pr[\tau > N] \leq \Pr[\min(\tau, N) \geq N] \leq I(f)/(\gamma N) \leq \delta/2$.

On the event $\tau \leq N$ the output is $\mathbb{E} f_{\rho_\tau}$ and $\mathrm{Var} f_{\rho_\tau} \leq v_0$. For a leaf ρ of the tree, the event $\{\rho_\tau = \rho\}$ is the event that x agrees with ρ on the coordinates fixed by ρ, so given it the unread coordinates are uniform, and Chebyshev's inequality gives $\Pr[|f(x) - \mathbb{E} f_\rho| > a \mid \rho_\tau = \rho] \leq \mathrm{Var}(f_\rho)/a^2 \leq v_0/a^2$. Averaging over the leaves and adding $\Pr[\tau > N]$ gives the claim. □

Corollary 10.29 (classical simulation of quantum query algorithms; [*proved here from Theorem 10.1, audited*])**.** *Let $t \geq 1$, $0 < \varepsilon < \frac{1}{2}$ and $0 < \delta \leq 1$, and put $T = \lceil \log_2(4et) \rceil$. Let $p(X)$ be the acceptance probability of a quantum algorithm that makes t queries to $X \in \{0, 1\}^n$. There is a deterministic classical algorithm that makes at most N queries to X and outputs a number within ε of $p(X)$ for all but a δ fraction of the inputs $X \in \{0, 1\}^n$, where*

$$\ln N \;\leq\; \ln \frac{8t}{\delta} + 6.6 \ln \frac{c_2 (2t)^{8.76}}{2\varepsilon^2 \delta} + 1.30\, T \Big(\ln P_{2t} + \ln \frac{1}{2\varepsilon^2 \delta} + 6.4 \ln(2t) + 23.2 \Big),$$

with $\ln P_{2t} = 3.0 \cdot 10^6\, T^3$ *and* $c_2 < 6.6 \cdot 10^7$ *as in Theorem 10.1. In particular*

$$\ln N \;\le\; 3.91 \cdot 10^6\, T^4 + (1.30\, T + 6.6) \ln \frac{1}{\varepsilon^2 \delta} + \ln \frac{1}{\delta}.$$

Hence, for $t \ge 3$,

$$N \;\le\; \frac{1}{\delta} \Big(\frac{1}{\varepsilon^2 \delta} \Big)^{1.88 \ln t + 12.4} e^{O(\log^4 t)} \;=\; \exp\Big(O\big(\log^4 t + \log t \cdot \log \tfrac{1}{\varepsilon\delta} \big) \Big),$$

and in particular $N \le \exp(O(L^4))$ *with* $L = \log(t/(\varepsilon\delta))$.

Proof. By [9], p is a real multilinear polynomial in X of degree at most $2t$ with values in $[0,1]$ on $\{0,1\}^n$. With $x_i = 1 - 2X_i$, the function $f(x) = 2p(X) - 1$ lies in $\mathcal{B}_d(n)$ for $d = 2t \ge 2$, and so does every restriction of f. Put $v_0 = 2\varepsilon^2\delta$ and let $R(v)$ be the right-hand side of Theorem 10.1 for this d; it is decreasing in v. If h is a restriction of f with $\mathrm{Var} h > v_0$, Theorem 10.1 applied to h gives $\ln(1/\mathrm{MaxInf}(h)) \le R(\mathrm{Var} h) \le R(v_0)$. So the hypothesis of Lemma 10.28 holds with $\gamma = e^{-R(v_0)}$, and the lemma with $N = \lceil 2d/(\delta\gamma) \rceil = \lceil 4t/(\delta\gamma) \rceil$ and $a = 2\varepsilon$ gives an algorithm with $\Pr_x[|A(x) - f(x)| > 2\varepsilon] \le \delta/2 + v_0/(4\varepsilon^2) = \delta$. The simulation outputs $(1 + A(x))/2$, which is within ε of $p(X)$ whenever $|A(x) - f(x)| \le 2\varepsilon$; querying x_i is querying X_i, and X is uniform when x is.

Since $4t/(\delta\gamma) \ge 1$, $N \le 8t/(\delta\gamma)$, so $\ln N \le \ln(8t/\delta) + R(v_0)$, which is the first bound. For the second, the expanded form of Theorem 10.1 gives $R(v_0) \le 3.9 \cdot 10^6\, T^4 + (1.30\, T + 6.6) \ln(1/v_0) + 8.4\, T \ln d + 30.2\, T + 57.9 \ln d + 119$, and $\ln(1/v_0) \le \ln(1/(\varepsilon^2\delta))$. Since $2^T \ge 2ed$, $\ln d \le T \ln 2 - \ln(2e) < 0.7\, T - 1.69$, so

$$\begin{aligned} \ln(8t) + 8.4\, T \ln d + 30.2\, T + 57.9 \ln d + 119 \;&=\; \ln 4 + (8.4\, T + 58.9) \ln d + 30.2\, T + 119 \\ &<\; 5.9\, T^2 + 58\, T + 21, \end{aligned}$$

which is less than $10^4\, T^4$ because $T \ge 4$. Theorem 10.1 holds with the constant $3.0 \cdot 10^6$ for every $d \ge 2$, so both bounds hold for every $t \ge 1$. Finally, $T \le \log_2(4et) + 1$ gives $1.30\, T + 6.6 \le 1.30 \log_2 t + 12.4 \le 1.88 \ln t + 12.4$, and $T = O(\log t)$ for $t \ge 3$. For the last form, $\log t \le L$ and $\log t \cdot \log(1/(\varepsilon\delta)) \le L^2$, and $L^2 \le L^4$ since $L > \log 6 > 1$. □

Remark 10.30 (what the reduction uses). (a) Lemma 10.28 uses the influence bound at a single scale, the stopping threshold $v_0 = 2\varepsilon^2\delta$, but for every restriction of f with variance above v_0 that the tree can reach, not only for f. Theorem 10.1 holds on all of $\mathcal{B}_d$, which is closed under restrictions, and its bound is monotone in v, so its value at v_0 serves at every node. No bound below the stopping scale is used, and a bound for f alone, or for a random restriction of f, would not suffice.

(b) The statement concerns query complexity only. The tree is determined by p, and following it requires the variances and influences of restrictions of p; as in [2], no bound on the running time is claimed. The tree is adaptive.

(c) The ε and δ dependence enters only through the threshold $v_0 = 2\varepsilon^2\delta$ (Chebyshev at the leaves) and the factor $1/\delta$ (Markov for the depth). Since $\ln(1/\mathrm{MaxInf})$ in Theorem 10.1 grows like $(1.30\, T + 6.6) \ln(1/v)$, the overhead is $(\varepsilon^2\delta)^{-O(\log t)}$ rather than polynomial in $1/(\varepsilon\delta)$.

Remark 10.31 (comparison with known simulations). Under Conjecture 1.1 the same reduction gives $\mathrm{poly}(t, 1/\varepsilon, 1/\delta)$ queries [2, Theorem 22 of the arXiv version]; this folklore simulation statement remains open, and Corollary 10.29 does not give it. Appendix M proposes a route to it: the quantum OSSS inequality of Conjecture M.3, even with its constant $4t$ replaced by any polynomial in t, would give Conjecture 1.1 for the biases $2p - 1$ of quantum query algorithms, and hence the polynomial simulation (Proposition M.12). For general t-query algorithms the known bounds were exponential in t. The DFKO inequality gives $\mathrm{MaxInf}(f) > \mathrm{Var}(f)^2/(4C^d)$

for $f \in \mathcal{B}_d$ with an unspecified absolute constant C (Section 1), and Lemma 10.28 then gives $N = \lceil 4t\,C^{2t}/(\varepsilon^4\delta^3)\rceil = 2^{O(t)}/(\varepsilon^4\delta^3)$ adaptive queries; [2, Corollary 25 of the arXiv version] gives $2^{O(t)}/(\varepsilon^4\delta^4)$ nonadaptive queries through the junta theorem [2, Theorem 9 of the arXiv version]. To our knowledge, Corollary 10.29 is the first simulation of general t-query algorithms with an overhead subexponential in t, and its bound is quasi-polynomial in $t/(\varepsilon\delta)$.

The improvement is asymptotic only. At $t = 1$ the leading term of the bound is $1.0 \cdot 10^9$. The leading term $3.91 \cdot 10^6\, T^4$ is below $t \ln 2$ only for $t \geq 3.25 \cdot 10^{13}$, and below $10t$ only for $t \geq 1.46 \cdot 10^{12}$. Since C is not explicit, the crossover with the DFKO route is not determined; it is at about $3.3 \cdot 10^{13}$ if $C = \sqrt{2}$ and about $1.5 \cdot 10^{12}$ if $C = e^5$, and it also requires $\log(1/(\varepsilon\delta))$ to be small compared with $t/\log t$.

Polynomial simulations are known for restricted classes. By [4, Theorem 1.3], the functions $2p - 1$ with p the acceptance probability of a t-query algorithm are exactly the functions $x \mapsto F((x, \mathbf{1}))$ with F a form of degree $2t$ and completely bounded norm at most 1. For completely bounded block-multilinear forms of degree d, [7, Theorem 1.3] gives $\mathrm{MaxInf} \geq \mathrm{Var}^2/(e(d+1)^4)$, and [7, Corollary 1.7] gives a simulation with $O(d^5\varepsilon^{-8}\delta^{-5})$ queries, in its notation, for algorithms that query each block of variables once, a class containing k-fold Forrelation. Further, [23, Corollary 1.7] gives $\mathrm{MaxInf} \geq \mathrm{Var}^2$ for t-query algorithms whose output is homogeneous of degree $2t$, and [24, Corollary 1.5] gives $O(t^2/(\varepsilon^2\delta))$ nonadaptive queries for block-query algorithms. General algorithms reduce to block form only on an exponentially small set of inputs [7, §4].

Bounds also depend on the number of rounds. For t-query algorithms with r rounds of adaptivity, [13, Theorem 1] gives $2^{O(r^2)}(t\log(1/\delta)/\varepsilon)^{O(r)}$ queries, with an announced improvement to $t^{O(r)}$, and [33] gives $\mathrm{poly}(t)$ for nonadaptive algorithms (Corollary M.17 gives $\lceil 4t^2/(\varepsilon^4\delta^3)\rceil$) and $(rt/(\varepsilon\sqrt{\delta}))^{O(4^r)}$ for r rounds [33, Theorem 1.8]. For fixed ε, δ, and up to unspecified constants, the bound of [13, Theorem 1] is smaller than that of Corollary 10.29 for r up to about $\log^2 t$ (up to about $\log^3 t$ with the announced $t^{O(r)}$), and larger for more rounds, in particular for fully adaptive algorithms.

Bhattacharya proved the conjecture for random restrictions [12, Theorems 6.3–6.4]: at rate $\log d/(Cd)$ the restricted function is, with high probability, close to a junta on influential coordinates. This controls a random restriction, not every restriction along a query path, so it does not give the hypothesis of Lemma 10.28.

11. Random priming at rate $1/d$, a one-shot conditional theorem, and the regular core

Section 9 studies the new spikes of one rate-$\frac{1}{2}$ restriction inside the chain of Theorem 8.4, and Section 10 bounds their mass with a quasi-polynomial prefactor, which gives Theorem 10.1. This section follows the other route: a single restriction at rate $p = 1/d$, at one threshold $\theta_0 = (\mathrm{Var} f/d)^{O(1)}$, so that every loss is paid once. Spike forcing (Lemma 11.1) shows that at this threshold a definite part of the variance is always carried by spikes of the restricted function, and random priming (Lemma 9.6) turns every family of primable spikes into large real block sensitivity; together they give a polynomial bound whenever the forced spikes can be primed. Theorem 11.5 carries this out in three forms (capture, amplification and bounded delayed activation) with explicit polynomial exponents, and Corollary 11.7 gives classes for which the hypotheses hold structurally. As reductions of the conjecture the hypotheses are circular (Remark 9.34 and Remark 11.6). Sections 11.4 and 11.5 locate what is left, the *regular core*: at a heavy spike whose derivative is relatively regular every admissible exploration is empty, priming on small fixed sets gives nothing (Proposition 11.13), and products of nearly Boolean functions of lower degree are hard instances whenever the conjecture fails at that degree (Remark 11.14). The gap becomes the anti-spike inequality of Question 11.11, which implies the

conjecture (Proposition 11.12) and follows from the large-jump bound (LJ) of Definition J.16. Nothing in this section proves Conjecture 1.1.

Tools. The section uses the following results of Section 9 and Appendix J, all [*proved here, audited*] except Lemma 9.3 ([*known*]), with their proofs given there. For $g\colon Q_n \to \mathbb{R}$, a coordinate j, a set $A \not\ni j$ and $a \in \{-1,1\}^A$, $q_j(A,a) = \mathbb{E}[(d_j g)^2 \mid x_A = a]$ and $\pi_j(A,a) = \mathbb{E}_{a'}\|d_j g(a,\cdot) - d_j g(a',\cdot)\|_1$ are as in Section 9.2.

- *Priming and conversion* (Lemmas 9.5 and 9.3). If $A_1,\dots,A_m$ are pairwise disjoint, $j_i \notin A_i$ and $a_i \in \{-1,1\}^{A_i}$, then $\sum_i \pi_{j_i}(A_i,a_i) \le 2\,\mathrm{rbs}(g)$, where rbs is the real block sensitivity of Definition 7.6. For $g \in \mathcal{B}_d$, $\mathrm{rbs}(g) \le d^2$ is known (Lemma 9.3); the elementary bound $\mathrm{rbs}(g) \le 4d^2$ noted in its second proof needs no citation and gives the form $2\,\mathrm{rbs}(g) \le 8d^2$ used below. If $q_j(A,a) \ge 4\mathrm{Inf}_j(g)$, then $\pi_j(A,a) \ge q_j(A,a)/8$.
- *Random priming, subset form* (Lemma 9.6). Let $\mathrm{MaxInf}(g) < \theta/L$, let ν_j be laws of pairs (A,a) with $j \notin A$ and $|A| \le s$, and let $\mu_j \ge 0$. If, for a set N' of coordinates, (P1) $\sum_{j\in N'} \mathrm{Inf}_j(g)\mathbb{E}_{\nu_j}\pi_j \ge \gamma\theta \sum_{j\in N'}\mu_j$ and (P2) $\sum_{j\in N'}\mathrm{Inf}_j(g)\nu_j(u \in A) \le \lambda \mathrm{Inf}_u(g)$ for every u, then $\sum_{j\in N'}\mu_j \le 4(\lambda+1)(s+1)\,2\,\mathrm{rbs}(g)/(\gamma L)$. The numbers μ_j enter the proof only through (P1), so they may be the spike masses of any restriction, and the sets A may contain coordinates of any kind.
- *The Load Lemma* (Lemmas 9.13 and J.10, Corollaries J.11 and J.12). Let ϕ have $\mathbb{E}\phi^2 > 0$, and consider its size-biased law and an adaptive revelation strategy whose coins are independent of x. The relative influence RI_u is a martingale until u is revealed, and at the revelation time $\widetilde{\mathbb{E}}[\mathrm{RI}_u(\rho_u)] = \mathrm{Inf}_u(\phi)/\mathbb{E}\phi^2$. Hence RI_u reaches β before u is revealed with probability at most $\mathrm{Inf}_u(\phi)/(\beta\mathbb{E}\phi^2)$, the expected number of reveals at relative influence at least β is at most $\deg\phi/\beta$, and a reveal at bias $|c_u| \ge \tau$ has probability at most $4\mathrm{Inf}_u(\phi)/(\tau^2\mathbb{E}\phi^2)$. For $h \in \mathcal{B}_d$ the weighted loads satisfy $\sum_j I_j \widetilde{\Pr}_j(\cdot) \le ((d-1)/\beta)\mathrm{Inf}_u(h)$, and $(4(d-1)/\tau^2)\mathrm{Inf}_u(h)$ for bias thresholds.
- *RL1* (Theorem J.1, proved in Appendix J.1, and Corollary J.3). Let $\phi\colon Q_n \to [-1,1]$ have degree at most k and $M = \mathbb{E}\phi^2 > 0$. For a unit vector a with $\max_i |a_i| \le \eta$ and $\varepsilon_* = e^3k^4\eta^2 \le \frac12$, the size-biased mean $\mu = \widetilde{\mathbb{E}}\langle a,y\rangle$ satisfies $\mu^2 \le 10e\ln(e/M)/\ln(1/\varepsilon_*)$. In the form used below: if $0 < \tau \le 1/(2e^3k^4)$, the biases c_u of the size-biased law satisfy

$$\sum_{u:\,|c_u|<\tau} c_u^2 \;\le\; \max\{\tau,\ \kappa\ln(e/M)\}, \qquad \kappa = \frac{10e}{\ln(1/\tau) - 3 - 4\ln k}.$$

11.1. **Setting and spike forcing.** Fix $d \ge 2$ and $f \in \mathcal{B}_d$ with $V = \mathrm{Var} f > 0$, and let $p = 1/d$. Write $\phi_j = d_j f$ and $I_j = \mathrm{Inf}_j(f) = \mathbb{E}\phi_j^2$, so that $\sum_j I_j \le dV$. Every coordinate u carries a clock $t_u \sim U[0,1]$; the clocks are i.i.d. and independent of x. The fixed set is $F = \{u : t_u \le 1-p\}$ and the alive set is $A = F^c$, so each coordinate is alive independently with probability p, and $\sigma = (F, x_F)$ is the rate-p restriction, with restricted function f_σ. Put $Q_j = \mathbb{E}[\phi_j^2 \mid x_F]$, which equals $\mathrm{Inf}_j(f_\sigma)$ when $j \in A$. As in Section 9.1, $\widetilde{\Pr}_j$ is the size-biased law of ϕ_j: x has density ϕ_j^2/I_j with respect to the uniform measure, and the clocks are independent of x; $\widetilde{\mathbb{E}}_j$ is its expectation. Since ϕ_j does not depend on x_j,

$$\mathbb{E}\big[Q_j \mathbf{1}_E\big] = I_j\,\widetilde{\Pr}_j(E) \tag{4}$$

for every event E determined by $(t_u, x_u\mathbf{1}[u \in F])_{u\ne j}$ and by coins independent of x, the left side being taken under the uniform measure. The quantities $N(\rho)$, $\mathrm{RI}_u(\rho)$ and c_u are as in Section 9.1: revealing u at ρ multiplies the height N by $1 + c_u x_u$, the $\widetilde{\Pr}_j$-law of x_u given ρ is $(1 \pm c_u)/2$, and $c_u^2 \le 4\,\mathrm{RI}_u(\rho)$.

Parameters. Throughout,

$$k_0 = \lceil \log_2(8d) \rceil + 1, \qquad \theta_0 = \Big(\frac{pV}{16d^2\, 3^{k_0}}\Big)^2, \qquad p_0 = \frac{1}{4d},$$
$$\tau = \theta_0^{20}, \qquad \beta = \frac{\tau^2}{4}, \qquad \lambda = \frac{2(d-1)}{\beta}, \qquad s = \Big\lfloor \frac{64d(d-1)}{\beta} \Big\rfloor. \tag{5}$$

The threshold θ_0 has the form of the one in Remark 9.34, with the level k_0 adapted to $p = 1/d$, and p_0 is not the p_0 of Section 9.6. Since $k_0 \geq 5$, $2^{k_0+2} \geq 64d$ and $3^{k_0} \leq 243\, d^{\log_2 3}$, we have $3888d^3 \leq 16d^3 3^{k_0} \leq 3888 d^{4.585} \leq d^{16.59}$, and therefore

$$\Big(\frac{V}{d}\Big)^{34} \leq \theta_0 \leq \frac{d^{-6}}{3888^2}, \qquad \ln\frac{1}{\tau} \geq 120 \ln d + 330, \qquad (\lambda+1)(s+1) \leq \frac{128 d^3}{\beta^2} = 2048\, d^3 \theta_0^{-80}. \tag{6}$$

Put $\kappa = 10e/(\ln(1/\tau) - 3 - 4\ln(d-1))$ and $\kappa_d = 2\kappa \ln d$. The denominator of κ is at least $116 \ln d + 327$, so $\kappa_d \leq 20e/116 < \frac{1}{2}$; and $\tau \leq 1/(2e^3(d-1)^4)$, so the displayed form of RL1 applies with $k = d-1$ to every restriction of every ϕ_j.

Spikes. A coordinate j is a *spike* of σ if $j \in A$ and $Q_j \geq \theta_0$. Write $\mathrm{sp}_j = \{Q_j \geq \theta_0\}$. Since $\mathbf{1}[j \in A]$ is independent of everything else, (4) gives the spike mass

$$\mu_j = \mathbb{E}_\sigma\big[Q_j \mathbf{1}[j \in A,\ Q_j \geq \theta_0]\big] = pI_j\, \widetilde{\Pr}_j(\mathrm{sp}_j).$$

The *heavy spikes* are the coordinates in $\mathcal{N} = \{j : \widetilde{\Pr}_j(\mathrm{sp}_j) \geq p_0\}$. If $\mathrm{MaxInf}(f) < \theta_0/L$ with $L \geq 1$, every spike is new in the sense of Section 9.1: its influence is multiplied by more than L.

Lemma 11.1 (spike forcing at rate $1/d$; [*proved here, audited*])**.** (a) $\mathbb{E}_\sigma U_{\theta_0}(f_\sigma) < pV/2$. (b) $\sum_j \mu_j > pV/2$. (c) $\sum_{j\in\mathcal{N}} \mu_j > pV/4$.

Proof. (a) Fix σ, let $J = J_{\theta_0}(f_\sigma)$ and $u' = \mathbb{E}[f_\sigma \mid x_{J^c}]$. Then $u' \in \mathcal{B}_d$, $\mathrm{MaxInf}(u') < \theta_0$ and $U_{\theta_0}(f_\sigma) = \mathrm{Var} u'$ (Lemma 8.1(a)). Corollary 3.3 with $k = k_0$ gives $W^{[1,k_0+1]}(u') < 4d^2 3^{k_0} \theta_0^{1/2} = pV/4$. The Fourier support of u' lies in that of f_σ, so $W^{\geq k_0+2}(u') \leq W^{\geq k_0+2}(f_\sigma)$, and by the formula in Lemma 8.1(e), $\mathbb{E}_\sigma W^{\geq k_0+2}(f_\sigma) = \sum_S \widehat{f}(S)^2 \Pr[|S \cap A| \geq k_0+2]$. Since $\Pr[\mathrm{Bin}(m,p) \geq r] \leq (mp)^r/r!$ and $mp \leq dp = 1$ for $m \leq d$, this is at most $V/(k_0+2)! \leq V 2^{-k_0-2} \leq pV/64$. Hence $\mathbb{E}_\sigma U_{\theta_0}(f_\sigma) < pV/4 + pV/64$. (b) By Lemma 8.1(b), $\mathrm{Var} f_\sigma - U_{\theta_0}(f_\sigma) \leq S_{\theta_0}(f_\sigma) = \sum_{j\in A} Q_j \mathbf{1}[Q_j \geq \theta_0]$, whose expectation is $\sum_j \mu_j$, and $\mathbb{E}_\sigma \mathrm{Var} f_\sigma \geq pV$ by Lemma 8.1(e). Now use (a). (c) For $j \notin \mathcal{N}$, $\mu_j < pp_0 I_j$, so these coordinates carry less than $pp_0 dV = pV/4$. □

Remark 9.34 states spike forcing for $p \leq 1/(2ed)$, through the base case of Lemma 8.3; at $p = 1/d$ the high-degree tail is controlled by $dp = 1$ and the level k_0, as above. If $\mathrm{MaxInf}(f) < \theta_0/L$, the heavy spikes carry more than a quarter of the forced mass and all of them are new; the rest of the section asks when they can be primed.

11.2. **Admissible explorations and the small-bias tail.** For each j let $H_j = \{u \neq j : \mathrm{Inf}_u(\phi_j) \geq \beta I_j\}$. Since $\sum_u \mathrm{Inf}_u(\phi_j) \leq (d-1) I_j$, $|H_j| \leq (d-1)/\beta$.

Definition 11.2 (admissible exploration)**.** An *admissible exploration* for j is a sequential procedure, run under $\widetilde{\Pr}_j$, that queries coordinates $u \neq j$ one at a time; the next query, and the decision to stop, depend only on the observations so far and on coins independent of x and of the clocks. A query of u observes $\mathbf{1}[u \in F]$ and, if $u \in F$, the value x_u, and nothing else about the clocks: the exploration is *membership-blind*. An unqueried coordinate may be queried only if $u \in H_j$ or u is *β-large*, that is, $\mathrm{RI}_u \geq \beta$ at the time of the query, computed from the values revealed so far. The output is the queried set B and the data $z_B = (B \cap F, x_{B\cap F})$, with height $q_j(z_B) = \mathbb{E}[\phi_j^2 \mid x_{B\cap F}]$. The *truncation* B^{tr} stops after s queries. The *relative-influence exploration* B_0 queries the smallest unqueried u with $\mathrm{RI}_u \geq \beta$, as long as one exists.

When B_0 stops, every unqueried coordinate has $\mathrm{RI}_u < \beta$, hence bias $|c_u| < 2\sqrt{\beta} = \tau$. Queried alive coordinates are not revealed, so the restricted function need not be β-regular in them; it is β-regular when no query hits an alive coordinate. A bias trigger at level τ, which queries u when $|c_u| \geq \tau$, is admissible, since $|c_u| \geq \tau$ forces $\mathrm{RI}_u \geq \tau^2/4 = \beta$.

Lemma 11.3 (loads and sizes; [*proved here, audited*])**.** *For every family of admissible explorations B_j and every u:* (a) $\widetilde{\Pr}_j(u \in B_j) \leq 2\mathrm{Inf}_u(\phi_j)/(\beta I_j)$ *for $u \neq j$, hence* $\widetilde{\mathbb{E}}_j|B_j| \leq 2(d-1)/\beta$*;* (b) $\sum_j I_j \widetilde{\Pr}_j(u \in B_j) \leq \lambda \mathrm{Inf}_u(f)$*;* (c) $\widetilde{\Pr}_j(B_j^{\mathrm{tr}} \neq B_j) \leq p_0/8$*, and* (a) *and* (b) *hold for* B_j^{tr}*.*

Proof. The memberships $\mathbf{1}[u \in F]$ are coins independent of x, so an admissible exploration is an adaptive revelation strategy in the sense of Lemma 9.13, in which a queried alive coordinate is never revealed. (a) If $u \in H_j$, then $1 \leq \mathrm{Inf}_u(\phi_j)/(\beta I_j)$. Otherwise u is queried only if RI_u reaches β before u is revealed, which has probability at most $\mathrm{Inf}_u(\phi_j)/(\beta I_j)$ by the Load Lemma. Sum over u, using $\sum_u \mathrm{Inf}_u(\phi_j) \leq (d-1)I_j$. (b) Multiply (a) by I_j and sum over $j \neq u$, using $\sum_{j \neq u} \mathrm{Inf}_u(d_j f) = \sum_{S \ni u}(|S|-1)\widehat{f}(S)^2 \leq (d-1)\mathrm{Inf}_u(f)$. (c) $B_j^{\mathrm{tr}} \neq B_j$ only if $|B_j| \geq s+1$, so Markov's inequality and (a) give $\widetilde{\Pr}_j(B_j^{\mathrm{tr}} \neq B_j) \leq 2(d-1)/(\beta(s+1)) \leq 1/(32d) = p_0/8$; and $B_j^{\mathrm{tr}} \subseteq B_j$. □

The second stage. Condition on the data $\mathcal{D}$ of an admissible exploration for j: the queried set, the memberships of the queried coordinates, and the revealed values $z = x_{B \cap F}$. The exploration reads x only on $B \cap F$ and its other inputs are independent of x, so the conditional law of the unrevealed values is the size-biased law of $\phi_j(z, \cdot)$, of height $q = q_j(z_B)$; and the clocks of the unqueried coordinates are still i.i.d. $U[0,1]$ and independent of x and $\mathcal{D}$. Reveal each unqueried $u \in F$ at time t_u. Given the past at time t, each unrevealed unqueried coordinate is revealed at hazard rate $1/(1-t)$, and $\int_0^{1-p} dt/(1-t) = \ln d$; the queried alive coordinates are never revealed. The final height is Q_j, and $\ln(Q_j/q) = \sum_k \ln(1 + c_k x_{v_k})$, where v_k is the k-th revealed coordinate and c_k its bias just before. A reveal is *large* if $|c_k| \geq \tau$, and N^{lg} denotes the number of large reveals; each contributes at most $\ln 2$. By the Load Lemma, $\widetilde{\mathbb{E}}_j[N^{\mathrm{lg}} \mid \mathcal{D}] \leq 4(d-1)/\tau^2 = (d-1)/\beta$, which is polynomial in d/V; but each large reveal may double the height, so this alone gives only $q \geq \theta_0 \exp(-\mathrm{poly}(d/V))$ on the spike. The small reveals are controlled by RL1.

Lemma 11.4 (the small-bias tail at rate $1/d$; [*proved here, audited*])**.** *Fix the data $\mathcal{D}$ of an admissible exploration for j, with height $q > 0$, and let $\delta, y > 0$. With conditional $\widetilde{\Pr}_j$-probability at least $1 - \delta - e^{-y}$,*

$$\ln \frac{Q_j}{q} \;\leq\; N^{\mathrm{lg}} \ln 2 + 2\tau \ln d + \kappa_d \ln \frac{e}{\delta q} + 2y.$$

Proof. Let N_t be the height at time t and σ_δ the first time at which $N_t < \delta q$. On the revealed values the density of the conditional law with respect to the uniform measure is N_t/q, so $\widetilde{\Pr}_j(\sigma_\delta \leq 1 - p \mid \mathcal{D}) \leq \delta$, as in step (ii) of the proof of Theorem J.14; stop all processes at σ_δ. The large reveals contribute at most $N^{\mathrm{lg}} \ln 2$. The small reveals contribute $W = W^c + M$, where

$$W^c = \int_0^{1-p} \sum_u h(c_u(t)) \frac{dt}{1-t}, \qquad h(c) = \sum_{m \geq 1} \frac{c^{2m}}{(2m-1)2m} = \mathbb{E} \ln(1 + c x_u) \quad (x_u \sim (1 \pm c)/2),$$

the sum running over the unrevealed unqueried u with $|c_u(t)| < \tau$, is the compensator, and M is a martingale. Since $|\ln(1 \pm c)| \leq 2|c|$ and $\mathbb{E} \ln^2(1 + c x_u) \leq c^2$ for $|c| \leq \frac{1}{2}$, the jumps of M are at most $2\tau \leq \frac{1}{2}$ in absolute value and its predictable quadratic variation is at most $v = \int_0^{1-p} \sum_u c_u(t)^2 \, dt/(1-t)$, with the same range of summation. Before σ_δ, RL1 with $k = d-1$, applied to the restricted function of height $N_t \geq \delta q$, gives $\sum_{|c_u(t)| < \tau} c_u(t)^2 \leq \tau + \kappa \ln(e/(\delta q))$,

the sum over all unrevealed coordinates. As $h(c) \le (\ln 2)c^2$, this gives $W^c \le (\ln 2)v \le v \le v_0 := \ln d\,(\tau + \kappa \ln(e/(\delta q)))$. Freedman's inequality [20], in its standard form for martingales with jumps at most $\frac{1}{2}$, gives $\widetilde{\Pr}(M \ge z,\ \langle M\rangle \le v_0) \le \exp(-z^2/(2v_0+z))$, and with $z = v_0 + 2y$ the exponent is $(v_0+2y)^2/(3v_0+2y) \ge y$. So, with probability at least $1-\delta-e^{-y}$, $W \le 2v_0 + 2y = 2\tau \ln d + \kappa_d \ln(e/(\delta q)) + 2y$. □

11.3. **The conditional theorem and the covered classes.** For each j fix an admissible exploration B_j, for instance B_0, and write $q_j = q_j(z_{B_j})$ and N_j^{lg} for the number of large reveals of the second stage.

Theorem 11.5 (one-shot conditional theorem; [*proved here, audited*])**.** *Let $f \in \mathcal{B}_d$, $d \ge 2$, $V = \mathrm{Var} f > 0$, $m \in (0,1]$ and $C_0 \ge 0$.*

(i) Capture. *If $\widetilde{\Pr}_j(q_j \ge m \mid \mathrm{sp}_j) \ge \frac{5}{8}$ for every $j \in \mathcal{N}$, then $\mathrm{MaxInf}(f) \ge \theta_0/L_{\mathrm{cap}}$, where $L_{\mathrm{cap}} = 2^{22} d^5 \theta_0^{-79}/(mV)$; in particular $\mathrm{MaxInf}(f) \ge (V/d)^{2747} m$.*

(ii) Amplification. *If $\widetilde{\mathbb{E}}_j[\ln(Q_j/q_j) \mid \mathrm{sp}_j] \le C_0 \ln(d/V)$ for every $j \in \mathcal{N}$, then $\mathrm{MaxInf}(f) \ge (V/d)^{2825+8C_0}$.*

(iii) Bounded delayed activation. *If $\widetilde{\Pr}_j(N_j^{\mathrm{lg}} \le C_0 \log_2(d/V) \mid \mathrm{sp}_j) \ge \frac{3}{4}$ for every $j \in \mathcal{N}$, then $\mathrm{MaxInf}(f) \ge (V/d)^{2855+2C_0}$.*

Parts (i) *and* (ii) *do not use RL1.*

Proof. (i) Suppose that $\mathrm{MaxInf}(f) < \theta_0/L_{\mathrm{cap}}$; then $I_j < m/4$ for every j. Let ν_j be the $\widetilde{\Pr}_j$-law of $(B_j^{\mathrm{tr}} \cap F, x_{B_j^{\mathrm{tr}} \cap F})$, a law of pairs (A,a) with $|A| \le s$. For $j \in \mathcal{N}$, the event $\mathrm{sp}_j \cap \{q_j \ge m\} \cap \{B_j^{\mathrm{tr}} = B_j\}$ has probability at least $(\frac{5}{8} - \frac{1}{8})\widetilde{\Pr}_j(\mathrm{sp}_j)$, by Lemma 11.3(c) and $\widetilde{\Pr}_j(\mathrm{sp}_j) \ge p_0$, and on it the conversion gives $\pi_j \ge q_j/8 \ge m/8$. Hence $I_j \mathbb{E}_{\nu_j} \pi_j \ge \frac{m}{16} I_j \widetilde{\Pr}_j(\mathrm{sp}_j) = m\mu_j/(16p)$, which is (P1) on $N' = \mathcal{N}$ with $\theta = \theta_0$ and $\gamma = m/(16p\theta_0)$. (P2) holds with λ by Lemma 11.3(b),(c). Random priming with $L = L_{\mathrm{cap}}$, $2\,\mathrm{rbs}(f) \le 8d^2$ and (6) gives

$$\sum_{j \in \mathcal{N}} \mu_j \ \le\ \frac{4 \cdot 2048 d^3 \theta_0^{-80} \cdot 8d^2 \cdot 16p\theta_0}{mL_{\mathrm{cap}}} \ =\ \frac{2^{20} p d^5 \theta_0^{-79}}{mL_{\mathrm{cap}}} \ =\ \frac{pV}{4},$$

contradicting Lemma 11.1(c). Finally $\theta_0/L_{\mathrm{cap}} = 2^{-22} d^{-5} \theta_0^{80} mV \ge (V/d)^{2747} m$ by (6), since $2^{-22} d^{-5} V \ge (V/d)^{27}$.

(ii) It suffices to verify the hypothesis of (i) with $m = (V/d)^{78+8C_0}$. Fix $j \in \mathcal{N}$, let $\delta' = p_0/8$ and $Y = \ln(Q_j/q_j) + \ln(1/\delta')$. Given the exploration data and the clocks, the density of the second-stage values with respect to the uniform measure is Q_j/q_j, so $\widetilde{\mathbb{E}}_j[q_j/Q_j \mid \mathcal{D}, \mathrm{clocks}] \le 1$ and $\widetilde{\Pr}_j(Y < -y) \le \delta' e^{-y}$ for $y \ge 0$. Hence $\widetilde{\mathbb{E}}_j[Y_- \mid \mathrm{sp}_j] \le \delta'/p_0 = \frac{1}{8}$ and $\widetilde{\mathbb{E}}_j[Y_+ \mid \mathrm{sp}_j] \le C_0 \ln(d/V) + \ln(1/\delta') + \frac{1}{8}$. By Markov's inequality, with conditional probability at least $\frac{7}{8}$, $\ln(Q_j/q_j) \le 8C_0 \ln(d/V) + 7\ln(32d) + 1$; on sp_j this gives $q_j \ge \theta_0 e^{-1} (32d)^{-7} (V/d)^{8C_0} \ge (V/d)^{78+8C_0}$, since $(32d)^7 \le d^{42}$ and $e \le d^2$.

(iii) It suffices to verify the hypothesis of (i) with $m = (V/d)^{108+2C_0}$. Fix $j \in \mathcal{N}$ and apply Lemma 11.4 with $\delta = e^{-y} = p_0/16$. Its exceptional event has conditional probability at most $(p_0/8)/p_0 = \frac{1}{8}$ given sp_j, so with conditional probability at least $\frac{5}{8}$ the spike occurs, $N_j^{\mathrm{lg}} \ln 2 \le C_0 \ln(d/V)$, and

$$\ln \frac{1}{q_j} \ \le\ \ln \frac{1}{\theta_0} + C_0 \ln \frac{d}{V} + 2\tau \ln d + \kappa_d \Big(1 + \ln \frac{1}{\delta} + \ln \frac{1}{q_j}\Big) + 2y.$$

With $\kappa_d < \frac{1}{2}$, $2\tau \ln d \le 1$, $\ln(1/\theta_0) \le 34 \ln(d/V)$ and $\ln(1/\delta) = y = \ln(64d) \le 7\ln(d/V)$, this gives $\ln(1/q_j) \le (107.4 + 2C_0)\ln(d/V)$. □

Remark 11.6 (what the theorem gives; [*proved here, audited*]). (a) For a single f, each hypothesis implies the bound; the converse is not claimed.

(b) Call f *deficient* if $\mathrm{MaxInf}(f) < \theta_0/L_{\mathrm{cap}}$. Quantified over the deficient f, each hypothesis is equivalent to the conclusion; quantified over all f, it is at least as strong. Indeed, if f is deficient, the proof of (i), applied to the set $N' \subseteq \mathcal{N}$ of heavy spikes at which the capture hypothesis holds, gives $\sum_{j\in N'} \mu_j \le pV/4$ (and $\le pV/8$ if $\mathrm{MaxInf}(f) < \theta_0/(2L_{\mathrm{cap}})$), while $\sum_{j\in\mathcal{N}} \mu_j > pV/4$. So the hypothesis fails at some heavy spike of every deficient f, and the statement that it holds for every deficient f is equivalent to the statement that no deficient f exists, that is, to the conclusion for all f. The same holds for (ii) and (iii), with $m = (V/d)^{78+8C_0}$ and $m = (V/d)^{108+2C_0}$ in L_{cap}. This is the circularity of Remark 9.34: the one-shot theorems are valid implications, but not reductions of the conjecture.

(c) The hypotheses carry information only at spikes where explorations find something. At a heavy spike j with $\mathrm{Inf}_u(\phi_j) < \beta I_j$ for every u, no coordinate may be queried, every admissible exploration is empty, and $q_j = I_j$; each hypothesis then forces $I_j \ge \mathrm{poly}(V/d)$, that is, it asserts that j itself is polynomially influential. What survives is the localization of the difficulty to heavy spikes with relatively regular derivatives (Sections 11.4 and 11.5), and the structural classes of Corollary 11.7.

Corollary 11.7 (covered classes; [*proved here, audited*])**.** *Let $f \in \mathcal{B}_d$, $d \ge 2$ and $V = \mathrm{Var} f > 0$.*

(a) Heavy juntas. *If, for every $j \in \mathcal{N}$, the derivative ϕ_j depends only on x_{H_j}, then* $\mathrm{MaxInf}(f) \ge (V/d)^{2825}$.

(b) The class $\mathcal{G}$. *Suppose that, for every $j \in \mathcal{N}$, there are a set $R \not\ni j$ and, for each $r \in \{-1,1\}^R$, disjoint sets D_r and Z_r, disjoint from $R\cup\{j\}$, with $\phi_j(r,\cdot) = \psi_r(x_{D_r})\chi_r(x_{Z_r})$, such that $R\cup\bigcup_r D_r \subseteq H_j$ and, for every r and every restriction of χ_r to part of Z_r, each unassigned coordinate of Z_r has bias less than τ in absolute value under the size-biased law of the restricted χ_r. Then* $\mathrm{MaxInf}(f) \ge (V/d)^{2855}$.

Proof. (a) Query H_j. This is admissible, $|H_j| \le (d-1)/\beta \le s$, and $q_j = \mathbb{E}[\phi_j^2 \mid x_{H_j\cap F}] = Q_j$ identically, so Theorem 11.5(ii) applies with $C_0 = 0$. (b) Query R, and then D_r for every r that agrees with the revealed values on $R \cap F$; all queries lie in H_j. At every later time the conditional law of the unrevealed values is a mixture, over the compatible r, of the size-biased laws of the restrictions of $\phi_j(r,\cdot)$, and each component is the product of a law on D_r, the size-biased law of a restriction of χ_r on Z_r, and the uniform law elsewhere. A coordinate revealed in the second stage lies in no queried set, so in each component its bias is 0 or less than τ in absolute value, and its bias under the mixture is a convex combination of these. Hence $N_j^{\mathrm{lg}} = 0$, and Theorem 11.5(iii) applies with $C_0 = 0$. □

Examples. For a product $\phi = \prod_i g_i$ on disjoint blocks, $\mathrm{Inf}_u(\phi)/\mathbb{E}\phi^2 = \mathrm{Inf}_u(g_i)/\mathbb{E}g_i^2$ for u in the i-th block. A parity gate $\frac{1}{2}(1 + x_u x_v)$ has ratio $\frac{1}{2}$ and $\mathbf{1}[\mathrm{Maj}_3 = 1]$ has ratio $\frac{1}{4}$, so derivatives that are products of such gates, for instance $\mathrm{AND} \circ \mathrm{Maj}_3$, are heavy juntas. A spread gate $\phi = \frac{\alpha}{M}\sum_{i\le M} \mathbf{1}[x_{D_i} = b^{(i)}]$, with disjoint blocks D_i of width w, $\varpi = 2^{-w}$ and $(M-1)\varpi \le 1$, has $\mathbb{E}\phi^2 = \frac{\alpha^2}{M}\varpi(1 + (M-1)\varpi)$ and $\mathrm{Inf}_u(\phi) = \alpha^2\varpi/(2M^2)$ for u in a block, so its ratio is at least $1/(4M)$, and it is a heavy junta when $M \le \theta_0^{-40}$. [*Proof outline, unaudited:*] if $\mathrm{MaxInf}(f) < \theta_0/L$ with $L \ge 480d$, a fourth-moment bound on the number of matched blocks gives $\widetilde{\mathrm{Pr}}_j(\mathrm{sp}_j) \le 30\alpha^2/(\theta_0 M^2) + 60/L$, so a heavy spike of this form has $M \le 16\sqrt{d/\theta_0} \le \theta_0^{-2}$, and the size condition is automatic. Members of $\mathcal{G}$ include exclusive selectors $\phi_j(r,\cdot) = \psi_r(x_{D_r})$ with R and all D_r in H_j, and, [*proved here, unaudited*], products $\psi(x_D)\prod_{u\in Z}(1+a_u x_u)/(1+|a_u|)$ with $D \subseteq H_j$, Z disjoint from $D\cup\{j\}$ and $|a_u| < \tau/2$, since the bias of x_u is $2a_u/(1+a_u^2)$ in every restriction.

11.4. **Why bias triggers fail, and the limits of the rate-$\frac{1}{2}$ chain.** A bias trigger at level τ is admissible, and Lemma 11.4 controls the biases below τ. Two families show that bias-triggered exploration nevertheless fails, at rate $1/d$ and at rate $\frac{1}{2}$, while the relative-influence trigger captures both. In each, the coordinates that make the spike have small or zero bias until other coordinates are revealed, although their relative influence is large from the start.

Proposition 11.8 (address functions; [*proved here, audited*])**.** *Let $k, w \geq 1$ with $k + w \leq d - 1$. Take selectors $y \in \{-1,1\}^k$, disjoint blocks D_b of w coordinates for $b \in \{-1,1\}^k$, $\phi = \prod_{u \in D_y} \frac{1}{2}(1 + x_u)$ and $f = x_j \phi$, so that $f \in \mathcal{B}_d$ and $\phi_j = \phi$.*

(a) $\mathbb{E}\phi^2 = 2^{-w}$. Under $\widetilde{\Pr}_j$ each selector has bias 0 and relative influence $\frac{1}{2}(1 - 2^{-w})$, and each block coordinate has bias 2^{-k} and relative influence 2^{-k-1}.

(b) At rate $1/d$, $\widetilde{\Pr}_j(Q_j = 1) \geq (1 - 1/d)^{k+w} \geq 1/e$.

(c) If $\tau > 2^{-k}$, the bias-triggered exploration at level τ is empty, so $q_j = 2^{-w}$, and on $\{Q_j = 1\}$ the second stage multiplies the height by 2^w.

(d) Order the coordinates with the selectors first. If $\beta \leq \frac{1}{2}(1 - 2^{-w})$, the relative-influence exploration B_0 queries every selector before any other coordinate; on the event that all selectors are fixed it then queries exactly D_y, and $q_j(z_{B_0}) = Q_j$.

Proof. $\phi \in \{0,1\}$, and under $\widetilde{\Pr}_j$ the selector value y is uniform, $x_{D_y} = \mathbf{1}$, and the other blocks are uniform. For a selector y_i, $(d_{y_i}\phi)^2 = \frac{1}{4}(\mathbf{1}[x_{D_{y^{i\to 1}}} = \mathbf{1}] - \mathbf{1}[x_{D_{y^{i\to -1}}} = \mathbf{1}])^2$ has mean $\frac{1}{2}2^{-w}(1 - 2^{-w})$; a coordinate $u \in D_b$ has $\mathrm{Inf}_u(\phi) = 2^{-k}2^{-w}/2$, and under $\widetilde{\Pr}_j$ it equals 1 when its block is selected, which has probability 2^{-k}, and is uniform otherwise. This gives (a). For (b), $Q_j = 1$ once all selectors and all of D_y are fixed. (c) follows from (a). For (d), revealing selectors leaves an address function with fewer selectors, in which the remaining selectors still have relative influence $\frac{1}{2}(1 - 2^{-w}) \geq \beta$, so B_0 takes them first; once all selectors are revealed, $\phi = \mathbf{1}[x_{D_y} = \mathbf{1}]$, whose coordinates have relative influence $\frac{1}{2}$ in every restriction, while all other coordinates have relative influence 0. □

The address function is not claimed to be covered by Corollary 11.7: when some selectors are alive, several blocks stay active. At rate $\frac{1}{2}$ the same mechanism refutes the delayed-activation hypothesis that the chain of Section 9 would need. For $L \geq 1$, $\tau' \in (0,1]$ and $p_* \in (0,1]$, let $\mathrm{NA}(L, \tau', p_*)$ be the statement: for every θ, every $h \in \mathcal{B}_d$ that is old-free capped at (θ, L) (Definition 9.16) and every j with $\widetilde{\Pr}_j(Q_j \geq \theta) \geq p_*$, where Q_j and $\widetilde{\Pr}_j$ refer to the rate-$\frac{1}{2}$ restriction, after the bias-triggered exploration at level τ' the second stage reveals a coordinate of bias at least τ' in absolute value with $\widetilde{\Pr}_j$-probability at most $p_*/6$.

Proposition 11.9 (parity gates; [*proved here, audited*])**.** *Let $L \geq 2$ and $16(1 + \log_2 L) \leq m \leq (d-1)/2$. Let $\phi_m = \prod_{i \leq m} \frac{1}{2}(1 + x_{u_i} x_{v_i})$, let $h^* \in \mathcal{B}_d$ depend only on coordinates other than j, u_i, v_i, with $\mathrm{MaxInf}(h^*) < 2^{1-m}$ and $\mathrm{Var}(h^*)^2 \geq 288 d L 2^{-m}$, and put $h = \frac{1}{2}(x_j \phi_m + h^*)$ and $\theta = L2^{-m-1}$. At rate $\frac{1}{2}$:*

(a) $h \in \mathcal{B}_d$ is old-free capped at (θ, L), and $\widetilde{\Pr}_j(Q_j \geq \theta) \geq 0.89$.

(b) The bias of a gate coordinate is 0 until its partner is revealed and ± 1 afterwards. So the bias-triggered exploration is empty, every spike path reveals a coordinate of bias ± 1, and $\mathrm{NA}(L, \tau', p_*)$ fails for all $\tau' \leq 1$ and $p_* \leq 0.89$.

(c) Every unrevealed gate coordinate has relative influence $\frac{1}{2}$ at every stage, so a relative-influence exploration at any $\beta \leq \frac{1}{2}$ queries exactly the gate coordinates, and $q_j(z_B) = Q_j$.

Consequently, if $\mathrm{NA}(L, \tau', p_*)$ held for some $\tau' \leq 1$ and $p_* \leq 0.89$, then every $h^* \in \mathcal{B}_d$ would satisfy $\mathrm{MaxInf}(h^*) \geq \mu_1 := \min\{(2L)^{-16}, \mathrm{Var}(h^*)^2/(288dL)\}$ whenever $\mu_1 \geq 2^{-(d-3)/2}$.

Proof. Here $I_j = \frac{1}{4}\mathbb{E}\phi_m^2 = 2^{-m}/4$, a gate coordinate has influence $2^{-m}/8$, and the other coordinates have influence $\mathrm{Inf}(h^*)/4 < \theta/L$, so $\mathrm{MaxInf}(h) < \theta/L = 2^{-m-1}$; and $\mathrm{Var}h = \frac{1}{4}(2^{-m} + \mathrm{Var}h^*)$, so $\theta \le \mathrm{Var}(h)^2/(36d)$. Under $\widetilde{\Pr}_j$ every gate is satisfied and $Q_j = \frac{1}{4}2^{N-m}$, where $N \sim \mathrm{Bin}(m, \frac{1}{4})$ counts the gates with both coordinates fixed; the spike is $\{N \ge 1+\log_2 L\}$, and since $m \ge 32$ the Chernoff bound gives $\Pr[N < m/16] \le e^{-9m/128} \le e^{-2.25} < 0.11$. This proves (a). For (b), $x_{u_i} = x_{v_i}$ under $\widetilde{\Pr}_j$, and x_{v_i} is uniform until it is revealed. For (c), $(d_{u_i}\phi_m)^2 = \frac{1}{4}\prod_{i' \ne i}\frac{1}{2}(1 + x_{u_{i'}}x_{v_{i'}})$, and the conditional mean of $\frac{1}{2}(1 + x_{u_i}x_{v_i})$ is $\frac{1}{2}$ while u_i is unrevealed. For the consequence, if $\mathrm{MaxInf}(h^*) < \mu_1$, take $m = \lceil \log_2(1/\mu_1) \rceil$: then $16(1+\log_2 L) \le m \le (d-1)/2$, $\mathrm{MaxInf}(h^*) < \mu_1 < 2^{1-m}$ and $\mathrm{Var}(h^*)^2 \ge 288dL\mu_1 \ge 288dL2^{-m}$, so (b) contradicts $\mathrm{NA}(L, \tau', p_*)$. □

So bias-triggered no-activation at the polynomial losses of the chain is at least as hard as a polynomial bound for every h^* in this range, and the obstruction is local: one block of parity gates attached to an arbitrary function. Both families are captured by the relative-influence trigger, which is why Definition 11.2 triggers on relative influence.

Remark 11.10 (the rate-$\frac{1}{2}$ chain and spread gates). (a) [*Proved here.*] The chain of Theorem 8.4, with a uniform loss or with a loss $P\theta^{-c}$ per step, gives no polynomial bound. For a uniform loss L: since TT fails for $L \le d^{1-o(1)}$ (Theorem I.4 and Corollary I.6), each of its $T = \lceil \log_2(2ed) \rceil$ steps divides the threshold by at least $d^{1-o(1)}$, so the bound it yields is at most $\theta_T d^{-(1-o(1))T} = d^{-\Omega(\log d)}$. A loss $P\theta^{-c}$ per step, with $P \ge 1$, multiplies $\log(1/\theta)$ by at least $1 + c$ at each step, which compounds to $(1+c)^T = \Theta(d^{\log_2(1+c)})$, as in the proof of Proposition 9.17; since the chain starts from $\theta_T \le (V/d)^2$, the bound is no better than $\exp(-\Omega(d^{\log_2(1+c)}\log(d/V)))$, which beats $\mathrm{Var}(f)^2 e^{-O(d)}$ only when $c < 1$ and $\log(1/V) = o(d^{1-\log_2(1+c)})$. Theorem 10.1 is this chain at $c = 1/(2T)$, where $(1+c)^T \le e^{1/2}$, with a quasi-polynomial prefactor (Lemma 10.2); it gives $(\mathrm{Var}f/d)^{O(\log d)}\exp(-O(\log^4 d))$. The one-shot route pays its loss once.

(b) [*Heuristic.*] For spread gates $M^{-1}\sum_{i \le M}\prod_{u \in D_i}\frac{1}{2}(1 + x_u)$ of width w with $M2^{-w} \le 1$, after k reveals in the active block the next bias is about $2^k/(M + 2^k)$ and the next relative influence about $\frac{1}{2} \cdot 2^k/(M + 2^k)$. The bias trigger fires after about $\log_2(\tau M)$ reveals, and the relative-influence trigger after about $\log_2(\tau^2 M)$. The defect $\ln(Q/q(z_B))$, minus the small-jump price, is $\ln(\tau M) + O(1)$, respectively $\ln(\tau^2 M) + O(1)$, when $\tau M^2 \lesssim 2^{w/2}$, and it saturates to $O(1)$ for $M \gtrsim 2^{w/2}$. At $M \approx 2^{w/4}\tau^{-1/2}$, with $w = d - 1$ and $\tau = d^{-O(1)}$, the bias-trigger defect is $\frac{w}{4}\ln 2 - \frac{1}{2}\ln(1/\tau) - O(1) = \Theta(d)$. [*Numerical:*] for the bias trigger, the mean of $\ln(Q/q(z_B))$ is 1.91, 3.14, 3.94, 5.44, against $\ln(\tau M/(1-\tau)) = 1.96, 3.35, 3.99, 5.37$, for $(M, w, \tau) = (64, 24, 0.1)$, $(256, 32, 0.1)$, $(1024, 40, 0.05)$, $(4096, 48, 0.05)$. In the chain these gates are handled by mass accounting, through the matched-block law of Section 9.6; at rate $1/d$ they are heavy juntas when $M \le \theta_0^{-40}$ (Corollary 11.7).

11.5. **The regular core and the remaining gap.** By Remark 11.6(c) the difficulty sits at heavy spikes whose derivatives are relatively regular. A condition that is not circular must therefore hold uniformly over all regular bounded functions, not only over the derivatives of one given f.

Question 11.11 (the anti-spike inequality $\mathrm{ASR}(C', C'')$; [*open*])**.** Is it true that for all $d \ge 2$, all $0 < \theta \le (16d^3)^{-2}$ and every $\psi\colon Q_n \to [-1, 1]$ of degree at most d with $q = \mathbb{E}\psi^2 \le (\theta/d)^{C''}$ and $\mathrm{Inf}_u(\psi) \le (\theta/d)^{C'}q$ for every u,

$$\widetilde{\Pr}_\psi\big(\mathbb{E}[\psi^2 \mid x_F] \ge \theta\big) \;\le\; \frac{1}{32d},$$

where $\widetilde{\Pr}_\psi = (\psi^2/q)\cdot U$ and F contains each coordinate independently with probability $1-1/d$, independently of x?

Proposition 11.12 (ASR implies the conjecture; [*proved here; the reduction is audited, the constants are not*])**.** *If* $\mathrm{ASR}(C',C'')$ *holds, then every* $f\in\mathcal{B}_d$ *with* $V=\mathrm{Var}f>0$ *satisfies*

$$\mathrm{MaxInf}(f)\ \ge\ \frac{V}{212992\,d^5}\Big(\frac{\theta_0}{2d}\Big)^{2C'+C''}\ \ge\ \Big(\frac{V}{d}\Big)^{23+72C'+36C''}.$$

Proof. Let $\beta'=(\theta_0/(2d))^{C'}$, $m=(\theta_0/(2d))^{C''}$, $s'=\lceil 64d^2/\beta'\rceil$ and $L=212992\,d^5\theta_0/(\beta'^2mV)$, and suppose that $\mathrm{MaxInf}(f)<\theta_0/L$; then $I_j<m/4$ for all j. For $j\in\mathcal{N}$, run the relative-influence exploration at threshold β' on the full cube under $\widetilde{\Pr}_j$, independently of the clocks: it reveals x_u for every queried u, fixed or alive, and it is truncated after s' queries. Revealing only fixed coordinates would not suffice, since alive coordinates could then never be revealed and could keep relative influence about $\frac12$. Let (B,x_B) be the output, $q_B=\mathbb{E}[\phi_j^2\mid x_B]$, E_0 the event that the exploration stops before the truncation, and $E_1=\{q_B\ge m\}$. By the Load Lemma, $\widetilde{\Pr}_j(E_0^c)\le(d-1)/(\beta'(s'+1))\le p_0/16$. On E_0 the restriction $\phi_j(x_B,\cdot)$ is β'-regular, and given (B,x_B) the set $F\setminus B$ is a fresh subset of B^c of rate $1-p$. Put $Q'=\mathbb{E}[\phi_j^2\mid x_B,x_{F\setminus B}]$, so that $\mathbb{E}[Q'\mid F,x_F]=Q_j$ under the uniform measure, and $S_j=\mathbb{E}[Q_j\mathbf{1}[\mathrm{sp}_j]]=I_j\widetilde{\Pr}_j(\mathrm{sp}_j)$. Then

$$S_j=\mathbb{E}[Q'\mathbf{1}[\mathrm{sp}_j]]\le\mathbb{E}[Q'\mathbf{1}_{E_0^c}]+\mathbb{E}[Q'\mathbf{1}_{E_0\cap E_1}]+\mathbb{E}[Q'\mathbf{1}[Q'\ge\theta_0/2]\mathbf{1}_{E_0\setminus E_1}]+\tfrac12 S_j,$$

because $\mathbb{E}[Q'\mathbf{1}[Q'<\theta_0/2]\mathbf{1}[\mathrm{sp}_j]]\le\frac{\theta_0}{2}\Pr[\mathrm{sp}_j]\le\frac12 S_j$. The first two terms are $I_j\widetilde{\Pr}_j(E_0^c)\le I_jp_0/16$ and $I_j\widetilde{\Pr}_j(E_0\cap E_1)$. On $E_0\setminus E_1$ the function $\phi_j(x_B,\cdot)$ has degree at most $d-1$, height below $m=((\theta_0/2)/d)^{C''}$ and relative influences at most $\beta'=((\theta_0/2)/d)^{C'}$, so ASR at $\theta=\theta_0/2\le(16d^3)^{-2}$ bounds the third term by $I_j/(32d)=I_jp_0/8$. As $S_j\ge p_0I_j$ for $j\in\mathcal{N}$, it follows that $\widetilde{\Pr}_j(E_0\cap E_1)\ge\frac{5}{16}\widetilde{\Pr}_j(\mathrm{sp}_j)$. On $E_0\cap E_1$ the conversion gives $\pi_j(B,x_B)\ge m/8$, so (P1) holds on $\mathcal{N}$ with $\gamma=5m/(128p\theta_0)$, and (P2) holds with $\lambda'=(d-1)/\beta'$ by the weighted load bound. Random priming, whose priming sets may contain alive coordinates, with $\lambda'+1\le d/\beta'$, $s'+1\le 65d^2/\beta'$ and $2\,\mathrm{rbs}(f)\le 8d^2$, gives $\sum_{j\in\mathcal{N}}\mu_j\le 53248\,p\theta_0d^5/(\beta'^2mL)=pV/4$, contradicting Lemma 11.1(c). The last inequality of the statement uses $212992\le 2^{18}\le d^{18}$ and $\theta_0/(2d)\ge(V/d)^{36}$. □

From (LJ) to ASR. Let $C'\ge256$ and $C''\ge16$. By Corollary J.17(b), applied with the empty exploration, (LJ) at (θ,C',C'') gives $\widetilde{\Pr}(Q\ge\theta)\le p_1/8=1/(512ed^2)\le1/(32d)$ for every $\theta^{2C'}/4$-regular bounded function of degree at most d and height at most $(\theta/d)^{C''}$, under the rate-$1/d$ clocks. Since $(\theta/d)^{2C'+1}\le\theta^{2C'}/4$, (LJ) for all $\theta\le(16d^3)^{-2}$ implies $\mathrm{ASR}(2C'+1,C'')$, and hence the conjecture with exponent $95+144C'+36C''$. The mechanism is Corollary J.15: outside probability $p_1/16$, a spike needs large-bias quadratic variation at least $\Lambda/4=\frac14\ln(e/q)$, measured at the bias level $\theta^{C'/2}$, while a regular function starts with all biases at most $2\sqrt{\beta}=\theta^{C'}$. The large biases must therefore be created along the clock order (*delayed activations*), and (LJ) bounds the probability of a spike with at least $\Lambda/4$ large reveals by $p_1/16$.

Status of ASR and (LJ). Both quantify over all regular bounded functions of small height. Applied only to the restricted derivatives of one fixed f, they fail whenever f violates the bound, by the proof of Proposition 11.12; this is the sense in which (LJ) is called circular as a reduction after Definition J.16 and in Remark 9.34. As universal statements they are sufficient for the conjecture, at least as hard as it, and not known to follow from it; some of their hard instances are described in Remark 11.14.

Proposition 11.13 (fixed-set priming of regular derivatives; [*proved here, audited*])**.** *Let* $\phi\colon Q_n\to[-1,1]$ *with* $I=\mathbb{E}\phi^2>0$ *and* $\mathrm{Inf}_u(\phi)\le\beta I$ *for every* u*, and let* A *be a set of* s

coordinates that is fixed, or chosen independently of x. Write $\pi(A,a)=\mathbb{E}_{a'}\|\phi(a,\cdot)-\phi(a',\cdot)\|_1$. Then (a) $\mathbb{E}_a\pi(A,a)\le\sqrt{2s\beta I}$ *for uniform a, and* (b) $\mathbb{E}_{a\sim\widetilde{\Pr}_A}\pi(A,a)\le\sqrt{2s\beta I}+\sqrt{8\ln 2\cdot s\beta}$, *where $\widetilde{\Pr}_A$ is the law of x_A under the size-biased law of ϕ. Consequently, if every ν_j is uniform or size-biased on a set of s coordinates chosen independently of x, then (P1) with $\theta=\theta_0$ fails on every family of heavy spikes with β-regular derivatives once $\beta<(\gamma\theta_0 pp_0)^2/(16s)$, since there $\mu_j\ge pp_0 I_j$.*

Proof. (a) By Cauchy–Schwarz, $(\mathbb{E}_{a,a'}\|\phi(a,\cdot)-\phi(a',\cdot)\|_1)^2\le\mathbb{E}_{a,a'}\|\phi(a,\cdot)-\phi(a',\cdot)\|_2^2=2\sum_{S\cap A\neq\emptyset}\hat\phi(S)^2\le 2\sum_{u\in A}\mathrm{Inf}_u(\phi)\le 2s\beta I$. (b) Since $0\le\pi\le 2$, the two expectations differ by at most $2\,\mathrm{TV}(\widetilde{\Pr}_A,U_A)\le\sqrt{2\,\mathrm{KL}(\widetilde{\Pr}_A\|U_A)}$. By the chain rule in a fixed order of A, $\mathrm{KL}=\sum_k\widetilde{\mathbb{E}}[h(c_k)]$, where c_k is the bias of the k-th coordinate given the earlier ones and h is as in the proof of Lemma 11.4; with $h(c)\le(\ln 2)c^2$, $c^2\le 4\,\mathrm{RI}$ and the Load Lemma, $\mathrm{KL}\le 4\ln 2\sum_{u\in A}\mathrm{Inf}_u(\phi)/I\le 4\ln 2\cdot s\beta$. For the consequence, $\mathbb{E}_{\nu_j}\pi_j\le(\sqrt{2}+\sqrt{8\ln 2})\sqrt{s\beta}<4\sqrt{s\beta}<\gamma\theta_0 pp_0$, so $\sum_j I_j\mathbb{E}_{\nu_j}\pi_j<\gamma\theta_0\sum_j\mu_j$ on such a family. □

For laws given by decision trees of depth s, passing to the fixed union of the nodes gives bounds of the same form, with s replaced by 2^s-1 (by $(3^s-1)/2$ if the tree also reads membership coins) and the constant 2 in the uniform term, [*proved here, audited*]. Indeed, let W be the set of coordinates at the nodes of the tree, so that $|W|$ is at most this number, and let (A,x_A) be its output, with $A\subseteq W$. For uniform $b\in\{-1,1\}^{W\setminus A}$ and $z\in\{-1,1\}^W$, inserting $\phi(z,\cdot)$ between $\phi(a,b,\cdot)$ and $\phi(a',b,\cdot)$ gives $\pi(A,a)\le\mathbb{E}_b\pi(W,(a,b))+\mathbb{E}_z\pi(W,z)$. The path of the tree depends only on x_A and on coins independent of x, so the kernel K that replaces $x_{W\setminus A}$ by a fresh uniform b maps the uniform law U_W on $\{-1,1\}^W$ to itself. Hence $\mathbb{E}\pi(A,x_A)\le 2\mathbb{E}_z\pi(W,z)$ for uniform x; under the size-biased law the first term changes by at most $2\,\mathrm{TV}(K\widetilde{\Pr}_W,KU_W)\le 2\,\mathrm{TV}(\widetilde{\Pr}_W,U_W)$, by data processing, and the proofs of (a) and (b) for the fixed set W finish the argument. For $s\le\frac12\log_2(1/\beta)$ (for $s\le\frac12\log_3(1/\beta)$ with membership coins) $|W|\beta<\beta^{1/2}$, so (P1) fails as above once $6\beta^{1/4}<\gamma\theta_0 pp_0$. General adaptive laws are not covered, and admissible explorations on a β-regular derivative are empty anyway. Since the truncation of the explorations is of order $1/\beta$, only genuinely small priming sets are excluded: uniform and size-biased priming on small fixed sets, or on shallow trees, gives no transfer at regular spikes.

Remark 11.14 (products of nearly Boolean factors; [*proof outline*]). Let $a_1,\dots,a_A\in\mathcal{B}_k$ live on disjoint blocks, be nearly Boolean to precision $o(1/A)$, satisfy $\Pr[a_i\approx 1]\le\frac12$ (replace a_i by $-a_i$ if necessary) and $\mathrm{MaxInf}(a_i)\le\tau'$, and put $\phi=\prod_i\frac12(1+a_i)$, of degree at most Ak. The relative influence of a coordinate u of the i-th block is $\mathrm{Inf}_u(a_i)/(4\mathbb{E}(\frac12(1+a_i))^2)$, at most about $\tau'/\mathrm{Var}(a_i)$ (and $\tau'/2$ for balanced Boolean factors), and $\mathbb{E}\phi^2\approx\prod_i\Pr[a_i\approx 1]\le 2^{-A}$. By Lemma 9.24(C2) the size-biased law is a product, concentrated on $\{a_i\approx 1\text{ for all }i\}$. At rate $1/d$ with $d\approx Ak+1$, each factor loses only $O(k/d)=O(1/A)$ in log-height in expectation, since a Boolean factor g of degree k has $\mathbb{E}\mathrm{Var}(g\mid x_F)\le pk\mathrm{Var}g$; so $\mathbb{E}[\phi^2\mid x_F]$ stays of order 1 with $\widetilde{\Pr}_\phi$-probability bounded below. With $A\approx C''\log_2(d/\theta)$ and $\tau'\le(\theta/d)^{C'}\min_i\mathrm{Var}(a_i)$, ϕ violates $\mathrm{ASR}(C',C'')$, and so does the derivative of $h=\frac12(x_j\phi+h^*)$ as in Proposition 11.9. Such factors exist only if the conjecture fails for nearly Boolean functions of degree $k\approx d/A$, in the sense that $\mathrm{MaxInf}(a_i)\le(\theta/d)^{C'}\mathrm{Var}(a_i)$; and by Theorem 7.3, nearly Boolean counterexamples exist whenever the conjecture fails, up to polynomial losses. So some hard instances of ASR are products of counterexamples of lower degree, and ASR at degree d implies the conjecture for nearly Boolean functions of degree about $d/(C''\log_2(d/\theta))$. The construction is exact for Boolean factors. [*Numerical:*] for Maj_k factors with $k\in\{3,5,7,9\}$ and $A\in\{4,8,16\}$, $\widetilde{\mathbb{E}}\ln(1/Q)$

lies between 0.17 and 0.33 and $\widetilde{\Pr}(Q \geq \frac{1}{4}) \geq 0.987$; for nearly Boolean factors the precision $o(1/A)$ is needed, since with 10% smoothing at $k = 3$, $\widetilde{\mathbb{E}} \ln(1/Q)$ grows as 0.58, 0.86, 1.41 for $A = 4, 8, 16$.

[*Heuristic:*] if the conjecture held at degree d/A with exponent $C(d/A)$, the relative influences of such products would be about $\beta' \approx (V/d)^{C(d/A)}$, and the loss $(d/\beta')^2$ of Proposition 11.12 would give $C(d) \approx 2C(d/A) + O(1)$ if every uncapturable regular spike had this structure. Then $\log(1/\mathrm{MaxInf}) = O(d^{\log_A 2} \log(d/V))$, which beats $\mathrm{Var}(f)^2 e^{-O(d)}$ for every constant $A \geq 3$ when $\log(1/V) = o(d^{1-\log_A 2})$, but is not polynomial; a loss linear in $1/\beta'$ would give $C(d) = C(d/A) + O(1)$, a quasi-polynomial bound.

The remaining gap. Along the one-shot route the implications are

$$\text{(LJ) for all } \theta \implies \mathrm{ASR}(2C' + 1, C'') \implies \text{Conjecture 1.1},$$

and the tools of this section and of Appendix J handle everything else: structured heavy spikes are captured (Corollary 11.7), small biases cost only the fraction $\kappa_d < \frac{1}{2}$ of the log-height (Lemma 11.4), and a spike without about $\Lambda/4$ large reveals is rare (Corollary J.15). What remains is to show that a regular bounded function of degree at most d and small height cannot, jointly with a spike, produce many delayed activations, that is, biases that are small when the restriction starts and become large along the clock order. This mechanism is invisible to bias-triggered and relative-influence explorations, which stop at once on regular functions, and to priming on small fixed sets (Proposition 11.13); and products of nearly Boolean factors realize it whenever the conjecture fails at a lower degree (Remark 11.14). The residual is therefore the regular core: regular derivatives of small height that are spiked with non-negligible probability. Its known hard instances are products of nearly Boolean counterexamples of lower degree (Remark 11.14), and the nearly Boolean case of precision $K^{-\gamma}$, for fixed $\gamma < 1$, is equivalent to the conjecture up to polynomial losses (Theorem 7.3). ASR and (LJ) imply the conjecture and are not known to follow from it. A proof of either needs an inequality for regular bounded functions with polynomial dependence on the degree; the standard tools for regular functions, invariance principles and hypercontractive estimates, lose factors exponential in the degree.

12. What remains open

Conjecture 1.1 is not proved in this paper [*open*]. This section states the remaining gap and lists the open problems, grouped by what a solution would give. The general bound proved here is the quasi-polynomial bound of Theorem 10.1; its constants are astronomically large, and at $\mathrm{Var} f = 1$ it beats e^{-d} for every $d \geq 1.91 \cdot 10^{13}$ and for no $d \leq 1.90 \cdot 10^{13}$ (first at $d = 1.903 \cdot 10^{13}$).

The gap. By Lemma 2.4 and Theorem 4.15, the conjecture is equivalent, up to polynomial losses, to the existence, for every bounded one-block form of degree k and variance V, of a z-decision tree of depth $\mathrm{poly}(k/V)$ whose leaves carry, on average, a y-variable of influence $\mathrm{poly}(V/k)$. By Theorem 4.14 it suffices to answer Question 4.16 positively: to find such a tree whose leaves carrying half of the variance have root sums $R_l \leq \mathrm{poly}(k/V)$. Since $R \leq e^k$ for every bounded family, this amounts to halving the root sum $O(k)$ times by shallow trees, which must in general be adaptive (the adaptive address function, Section 4.4). The variant that allows discarding mass $V/4$ leaf by leaf is equivalent to the conjecture (Theorem 4.17); whether Question 4.16 is strictly stronger, i.e. whether the conjecture implies a positive answer to it (without discarding indices inside the kept leaves), is open. Section 7.4 rules out arguments that use the degree only through $\sum_i |d_i f| \leq O(d^2)$ and dual certificates that depend linearly on the test function.

Statements that would imply the conjecture.

(i) A positive answer to Question 4.16.

(ii) The average-case Nisan lemma $\mathrm{AN}(C_1, C_2)$ with exponent $C_1 < 1$ (Theorem 7.5), or an approximate-certificate condition implying it. It is an integral, dense-domain analogue of [35, Conjecture 37]. That conjecture is false (Proposition H.4): its symmetric form holds for complementary sets (Remark H.5) and fails when a gap region of constant measure separates X from Y, whereas AN allows only a small exceptional set. AN may be stronger than the conjecture, and it is open even for $C_1 \geq 1$.

(iii) The degree-free Conjecture 7.9, equivalently $\mathrm{MaxInf}(g) \geq \mathrm{poly}(\mathrm{Var}g/\mathrm{rbs}(g))$ for every $g\colon Q_n \to [-1,1]$ (Remark 7.10), which is known for Boolean, nearly Boolean and non-spiky g, and which is false with the real sensitivity in place of rbs (Proposition 7.11); and ACE with codimension $\mathrm{poly}(d, 1/\eta)$ [35, Conjecture 32]. Already codimension $2^{o(d)}\mathrm{poly}(1/\eta)$ would give $\mathrm{MaxInf}(f) \geq e^{-o(d)}\mathrm{poly}(\mathrm{Var}f)$ (Remark 7.8).

(iv) Any of the statements that are equivalent to the conjecture up to polynomial losses (and hence not easier): the leaf-wise tree statement $\mathrm{TR}_{\mathrm{leaf}}$ and the halving lemma with heavy-y exit (Theorem 4.17); $\mathrm{NB}_{K^{-\gamma}}$ for any fixed $\gamma < 1$ (Theorem 7.3); the product-selector case (Theorem 7.4); the search form (Theorem G.5, a proof outline); and the one-scale transfer OS, even restricted to top-heavy functions (Theorem G.4). In the very nearly Boolean regime $\beta(K)K \lesssim 1$, which the amplification of Theorem 7.3 does not reach, NB_β is open and is not known to imply the conjecture.

(v) The anti-spike inequality $\mathrm{ASR}(C', C'')$ of Question 11.11 (Proposition 11.12), or the large-jump bound (LJ) of Definition J.16 for all θ, which for $C' \geq 256$ and $C'' \geq 16$ implies $\mathrm{ASR}(2C'+1, C'')$ (Corollary J.17 and Section 11.5). Both quantify over all regular bounded functions of small height; they are not known to follow from the conjecture, and some of their hard instances are products of nearly Boolean counterexamples of lower degree (Remark 11.14). The hypotheses of Theorem 11.5 are not candidates: quantified over the functions below its threshold, each is equivalent to the conclusion, and quantified over all f it is at least as strong (Remark 11.6).

The quasi-polynomial bound, and statements that would improve it. Theorem 10.1 gives $\mathrm{MaxInf}(f) \geq (\mathrm{Var}f/d)^{O(\log d)}\exp(-O(\log^4 d))$ [*proved here; audited twice in note form, and the text audited twice*]. The statements below would give $(\mathrm{Var}f/d)^{O(\log d)}$, without the second factor. The touching transfer TT of Definition 8.2 is implied by the conjecture (Theorem 8.7), and $\mathrm{TT}(1/(16ed), \mathrm{poly}(d))$ gives $\mathrm{MaxInf}(f) \geq (\mathrm{Var}f/d)^{O(\log d)}$ (Theorem 8.4, Corollary 8.5). It fails for $L = O(1)$ already for Tribes (Corollary I.3), for $L \leq d^{1-o(1)}$, and with cap $C \geq 2$ for $L \leq d^{2C/(2C-1)-o(1)}$; concretely $\mathrm{TT}(0.96, d^{1.2})$ with cap 2 is false (Theorem I.4, Corollary I.6). Under the conjecture with exponent $A \leq C$ the threshold exponent $a^*(\varepsilon)$ lies in $[2C/(2C-1), A]$ for every fixed $\varepsilon \in (0,1)$, and below $2A$ at the chain's $\varepsilon = 1/(16ed)$ (Remark I.8), and no Boolean inner function beats the amplification of AND in log-average (Proposition I.7). Whether TT holds with some polynomial loss is open; by Lemma I.9 (with h as there) a counterexample needs the new spikes to touch, in expectation, more than $\frac{1}{2}\sum_{|S|\geq 2}(1-2^{1-|S|})\hat{h}(S)^2 + \frac{\varepsilon}{2}\mathrm{Var}g$, a threshold between a quarter and a half of $W^{\geq 2}(h)$ (plus $\frac{\varepsilon}{2}\mathrm{Var}g$). The degree-free Conjecture 8.8 would give $\mathrm{MaxInf}(f) \geq (\mathrm{Var}f/d)^{O(C+(1+2/c)\log d)}$ (Proposition 8.9); it holds where $M_2(g) \leq \mathrm{poly}(\mathrm{rbs}(g))$ (Lemma 8.10), and its weak form follows from Conjecture 7.9 (Remark 8.11). The linear deficit transfer of Definition H.1 would also give a quasi-polynomial bound (Remark H.2, a proof outline), but it is false in the regime that argument needs, already for Boolean functions with $\mathrm{MaxInf} = \frac{1}{2}$ (Proposition H.3). The quasi-polynomial bound of Theorem G.2, conditional on OS, is a reformulation, since OS is equivalent to the conjecture. Section 9 treats the mass form of the touching transfer that its methods address: by Lemma I.9 it suffices that the new spikes of every old-free capped function carry small mass after one rate-$\frac{1}{2}$ restriction (SM-OF). SM-OF implies TT and so gives the same bound, but only at a θ-free loss, of θ-exponent at

most $1/(2T)$ and with a polynomial prefactor. With a quasi-polynomial prefactor it is proved (Theorem 10.26), and this gives Theorem 10.1. At the loss $\mathrm{poly}(d)\,\theta^{-2}$ it holds unconditionally (Proposition 9.21) and gives nothing new, and Table 4 lists the payoff of the intermediate exponents. That losses above a polynomial threshold are needed rests on a numerical sketch (Example 9.19: $L_0 \gtrsim \sqrt{d}/\log d$). Every capture-type strengthening is false, equivalent to TT, or implies SM, because captured spikes carry little mass automatically (Lemma 9.8, Theorem 9.9). Small-jump spikes and primable spikes are handled there. What remains for a polynomial prefactor is Conjecture 9.29, the mass of large-jump spikes that neither the greedy-closure nor the final-context law captures, which is SM-OF itself on that part of the mass, and its local form, a scale-invariant version of the top-heaviness inequality under size-biased laws (Question 9.31). The deep activation chains, on which the fixed-height recursion of Section 9.7 compounds, no longer block a quasi-polynomial bound: measuring lightness against a fixed fraction of the node's weight, on a grid of heights, ends every chain at depth at most $S-1 = 16T+3$ (Corollary 10.12), at the price of the prefactor $\ln P_d = 3.0 \cdot 10^6\, T^3$ (Remark 10.27). With a uniform loss or a loss $P\theta^{-c}$ per step, the chain of Theorem 8.4 gives no polynomial bound (Remark 11.10): a uniform loss is at least $d^{1-o(1)}$ at each of its $\Theta(\log d)$ steps, and a loss $P\theta^{-c}$ compounds; the one-shot route of Section 11 pays its loss once. Theorem 10.1 is this chain at $c = 1/(2T)$ with a quasi-polynomial prefactor.

Statements that would extend the special cases.

- *Graph properties:* $N\mathrm{Var} f \le \mathrm{poly}(d)$, which would improve Theorem 6.16 to $\mathrm{MaxInf}(f) \ge \mathrm{Var}(f)^2/\mathrm{poly}(d)$, with the optimal power of $\mathrm{Var} f$. Theorem 6.16 gives only $N\mathrm{Var} f \le \mathrm{poly}(d)\, M$, and pointwise star bounds cannot give more (Remark 6.17(d)), so a bound for the typical star change is needed. Lemma D.2 makes the constant of Lemma 6.13 explicit, $2^{d-1} \le C(d) \le (4d)^{2d}$, and the three- and four-vertex terms are $\mathrm{poly}(d)/M$ (Corollary D.3); by Remark D.4 the bound can be reached neither by level-by-level coefficient bounds nor by a single-scale perturbative blow-down.
- *Cyclically invariant functions:* a bound polynomial in d, item (3) of [3, Problem 1.3]. Summing Theorem 6.20 over all levels gives only $e^{O(d \log d)}$, and Propositions D.9 and D.10 show that uniform or single-resonance symbol bounds cannot give less than $e^{\Theta(d)}$ at level d.
- *Wreath-symmetric functions:* the d^2 law $\mathrm{MaxInf} \ge \mathrm{Var}^2/(Cd^2)$ in general, in particular for blocks of size $w < 4d^2$. The general bound is $\mathrm{Var}^2/(2416d^6)$ (Theorem 6.5), and the sharp order d^2 is proved only for composed functions and for many huge blocks (Theorems 6.6 and 6.8).
- *Smooth linear statistics:* removing the hypothesis $d^2 r\delta \le 1/4$ in Theorem C.4; this would follow from a bulk bound $\sup_{(1-\eta)Z} |P| = O(1)$ at the Rakhmanov scale $d\delta \le c$ for generic W (open).
- *One linear form:* the lacunary case (Theorem 6.12 covers regular weights, and regular tails after m leading weights with a bound polynomial in m).
- *Subcube classes:* an analogue of the subcube cover lemma (Lemma 5.1) for superlevel sets of spiky low-degree derivatives.
- *Torus-bounded polynomials:* the spike inequality (Conjecture E.7), $\sup_g \theta(g) < 1$ over analytic multilinear g that are not multiples of monomials, which would give the conjecture for this class (Appendix E), and the torus selection conjecture (Conjecture E.5). Products of one-variable factors give exactly $\theta_1 = 0.302992$, and no g with larger θ was found numerically. In each fixed degree the spike inequality holds uniformly in the number of variables, but with a proven gap $\eta_d \ge c\, 2^{-3d/2}/d$, exponentially small in the degree (Theorem E.8), which gives no gain for Conjecture 1.1.

- *Biases of quantum query algorithms:* the quantum OSSS inequality $\mathrm{Var}(f)^2 \le 4t\sum_k \delta_k \mathrm{Inf}_k(f)$ for the bias f of every t-query algorithm, with δ_k the query magnitudes of [11] averaged over the input (Conjecture M.3, Appendix M). It is proved for one query, for two queries up to the constant, and for nonadaptive algorithms. With any constant polynomial in t it would give Conjecture 1.1 for these biases, and hence a classical simulation of quantum query algorithms on most inputs with polynomially many queries (Proposition M.12).

Appendix A. Remarks on the literature

This appendix collects remarks on results in the literature that are not used in our proofs. They concern the random-restriction results of [12].

- *Random restrictions.* Bhattacharya [12, Theorem 6.4] proved: if $f\colon Q_n \to [0,1]$ has degree $d \ge 2$ and $\mathrm{Var} f \ge 1/d$, and ρ is a random restriction with survival probability $\log d/(C_1 d)$, then with probability at least $\mathrm{Var}(f)\log(d)/(50C_1 d)$ the restriction f_ρ has a coordinate of influence at least $\mathrm{Var}(f)^2/d^{C_2}$. He also states a tail theorem [12, Theorem 6.2]: with $\theta = \mathrm{Var}(f)^2/(d^C C^k)$ and $J = \{j : \mathrm{Inf}_j f \ge \theta\}$, if $\sum_{S\not\subseteq J}\widehat{f}(S)^2 \ge \mu$ then $\sum_{|S|>k}\widehat{f}(S)^2 \ge \mu/(d^C C^k)$. *Caveat.* As printed this statement is false when $J \neq \emptyset$: for $k=2$, $d=3$ and $f = \frac{1}{2} + \frac{\varepsilon}{2}x_1 + \frac{1-\varepsilon-\eta}{2}x_2 + \frac{\eta}{2}x_3x_4x_5$ (range $[0,1]$, $\mathrm{Var} f \approx 1/4$) one may choose $\varepsilon,\eta > 0$ with $\varepsilon^2/4 < \theta$ and $\eta^2(3^C C^2 - 1) < \varepsilon^2$, so that $J = \{2\}$, $\mu = (\varepsilon^2+\eta^2)/4$ and $\sum_{|S|>2}\widehat{f}(S)^2 = \eta^2/4 < \mu/(3^C C^2)$ (for larger d pad with a negligible monomial of degree d). The proof in [12] is consistent with the threshold $\theta = \mu^2/(d^C C^k)$; moreover, in our reading the published argument for the underlying [12, Theorem 6.1] yields an error bound of order between $\mu^5/(C^k d^C)$ and $\mu^4/(C^k d^C)$ rather than $\mu/(C^k d^C)$. We do not use [12, Theorems 6.1–6.2]; the results of Section 3 are self-contained. We have not checked whether these issues affect [12, Theorem 6.4].

Remark A.1 (relation to the tail theorem of [12]). A bound of the same shape can be derived from the tail theorem [12, Theorem 6.2] (case $J = \emptyset$): apply it to $g = (T_\rho f + 1)/2$, which satisfies $\mathrm{Var} g \ge \rho^{2j} W^j(f)/4$ and $\sum_{|S|>k}\widehat{g}(S)^2 \le \rho^{2k+2}/4$, with $k = 2j$ and $\rho^{2(j+1)}$ proportional to $W^j(f)/(d^C C^{2j})$. Taking the printed statement at face value this gives $\mathrm{MaxInf}(f) \ge W^j(f)^4/(16 d^{3C} C^{6j})$, with an unspecified constant C; with a tail bound $\mu^4/(C^k d^C)$ (the better end of the range in the item above) the same computation, now with $k = 5j-1$ and $\rho^{2j} = W^j(f)^4/(128 d^C C^{5j-1})$, gives $W^j(f)^{10}/(4^8 d^{3C} C^{15j})$, and with $\mu^5/(C^k d^C)$ (and $k = 6j-1$) it gives $W^j(f)^{12}/(2^{20} d^{3C} C^{18j})$ (for $C \ge 1$). Corollary 3.3 supersedes both: it is explicit, uses the cumulative weight $W^{[1,j]}$ instead of a single level, has exponent 2, and is independent of [12].

Appendix B. Wreath-symmetric functions: proofs and constants

This appendix contains the proofs of the wreath-symmetric bounds stated in Section 6 and the results they use. The constant chains are given in full.

Theorem B.1 (wreath-symmetric functions; [*proved here*]; the Bernstein inequality in the Fisher metric is classical)**.** *If $f \in \mathcal{B}_d(Mw)$ is invariant under $S_w \wr S_M$, then* $\mathrm{MaxInf}(f) \ge \mathrm{Var}(f)^2/(40\, d^2 w(w+1))$.

Proof. Structure. By symmetrization inside blocks [40] and the fundamental theorem on symmetric polynomials across blocks, $f(x) = R(\nu(x))$, where $\nu(x) = M^{-1}\sum_l \delta_{s_l(x)}$ is the empirical law of the block sums $s_l \in \{0,\dots,w\}$ and R is a real polynomial of degree at most d on $\mathbb{R}^{w+1}$

(each power sum $\sum_l s_l^j$ is linear in ν, and each monomial in the power sums has at most d factors). Moreover $|R| \le 1$ on the grid $G_M = \{\nu \in \Delta_w : M\nu \in \mathbb{Z}^{w+1}\}$. Let $S = \sup_{\Delta_w} |R|$.

Bernstein on the sphere. Put $\tilde{P}(u) = R(u_0^2, \dots, u_w^2)$, of degree at most $2d$; $u \mapsto (u_s^2)_s$ maps the unit sphere onto Δ_w, so $\sup_{S^w} |\tilde{P}| = S$. On a great circle $t \mapsto u\cos t + e \sin t$, $\tilde{P}$ is a trigonometric polynomial of degree at most $2d$ bounded by S, so its derivative is at most $2dS$ by Bernstein's inequality; as geodesic distance is at most $\frac{\pi}{2}$ times chordal distance, $|\tilde{P}(u) - \tilde{P}(u')| \le \pi d S |u - u'|$ on S^w.

Discretization. Let $|R(\nu^*)| = S$ and round $M\nu^*$ to integers with sum M, each within distance 1, giving $\nu' \in G_M$; then $|\sqrt{\nu'_s} - \sqrt{\nu^*_s}| \le M^{-1/2}$, so $S \le 1 + \pi d S \sqrt{(w+1)/M}$ and $S \le 2$ if $M \ge 4\pi^2 d^2 (w+1)$.

Variance. Assume $M \ge 4\pi^2 d^2 (w+1)$. Re-randomizing block l moves one atom of ν from $s = s_l$ to s'; with $k_t = M\nu(t)$, $|\sqrt{\nu} - \sqrt{\nu'}|^2 \le M^{-1}(1/k_s + 1/(k_{s'} + 1))$. The block sums are i.i.d. with law $\beta = \mathrm{Bin}(w, \frac{1}{2})$ having $w+1$ positive atoms, and $\mathbb{E}[1/(1 + \mathrm{Bin}(N,p))] \le 1/((N+1)p)$ gives $\mathbb{E}[1/k_{s_l}] \le (w+1)/M$ and likewise for s'. Since $S \le 2$, $|f(x) - f(x^{(l)})| \le 2\pi d |\sqrt{\nu} - \sqrt{\nu'}|$, and the Efron–Stein inequality gives $\mathrm{Var} f \le \frac{1}{2} \sum_l \mathbb{E}(f - f^{(l)})^2 \le 4\pi^2 d^2 (w+1)/M$.

Influences. The group is transitive on coordinates, so $\mathrm{MaxInf}(f) = I(f)/(Mw) \ge \mathrm{Var}(f)/(Mw)$. If $M \ge 4\pi^2 d^2 (w+1)$ then $1/M \ge \mathrm{Var}(f)/(4\pi^2 d^2 (w+1))$; otherwise $1/M > 1/(4\pi^2 d^2 (w+1)) \ge \mathrm{Var}(f)/(4\pi^2 d^2 (w+1))$. In both cases $\mathrm{MaxInf}(f) \ge \mathrm{Var}(f)^2/(4\pi^2 d^2 w(w+1))$, and $4\pi^2 < 40$. □

For balanced Tribes ($M \approx 2^w \ln 2$) the variance bound in the proof shows that a bounded polynomial of this symmetry type with constant variance has degree at least $c\, 2^{w/2}/\sqrt{w}$. The same proof gives the following.

Theorem B.2 (block-pattern symmetry; [*proved here*])**.** *Let $f \in \mathcal{B}_d(Mw)$ be invariant under permutations of the M blocks (with a fixed identification of the blocks and no assumption inside them), and suppose that f depends only on the block states $\phi(x_{B_1}), \dots, \phi(x_{B_M})$ for a fixed map $\phi\colon \{\pm 1\}^w \to [m]$ whose m values all have positive probability. Then $\mathrm{Var}(f) \le 4\pi^2 d^2 m/M$ whenever $M \ge 4\pi^2 d^2 m$, and $\mathrm{MaxInf}(f) \ge \mathrm{Var}(f)^2/(40 d^2 wm)$.*

Proof. Only the structure step and the last step of the proof of Theorem B.1 change. *Full patterns.* Let first ϕ be the identity. Averaging a monomial over the block permutations gives an orbit sum $\sum_{(l_1,\dots,l_k)\ \text{injective}} \prod_{r \le k} h_r(x_{B_{l_r}})$ with $k \le d$ nonempty block parts; by Möbius inversion over set partitions of $[k]$, such a sum is a linear combination of products over the parts β of a partition of $\sum_l \prod_{r \in \beta} h_r(x_{B_l})$, each factor being M times a linear functional of the empirical law of the block states. So $f = R(\nu)$ with $\deg R \le d$. *Coarse states.* For general ϕ, $f = \mathbb{E}[f \mid \phi(x_{B_1}), \dots, \phi(x_{B_M})]$, since f depends only on the coarse states. Given the coarse states, the blocks are independent and uniform on the fibres of ϕ, so the conditional expectation of a product of k linear functionals of the fine empirical law is, after grouping equal block indices, a polynomial of degree at most k in linear functionals of the coarse empirical law. Hence $f = R(\nu)$ with ν the empirical law of the coarse states and $\deg R \le d$; every empirical law of M coarse states occurs, and the sphere, discretization and variance steps apply with $w+1$ replaced by m. *Influences.* All blocks carry the same total influence, so $\mathrm{MaxInf}(f) \ge I(f)/(Mw) \ge \mathrm{Var}(f)/(Mw)$, and the case analysis in the last step of the proof of Theorem B.1 gives $\mathrm{MaxInf}(f) \ge \mathrm{Var}(f)^2/(4\pi^2 d^2 wm)$. □

With full block states $m = 2^w$, so this bound is exponential in w.

A bound uniform in w. Besides Theorem B.1 for small blocks we use the following facts for a real polynomial p of degree at most d: Bernstein's inequality at the centre of an interval of half-length h, $|p'| \le d \sup|p|/h$; for $|t| \le b < 1$, $|p''(t)| \le 2d(d-1) \sup_{[-1,1]} |p|/(1 - b^2)$ (Bernstein's

inequality twice, on $[-a,a]$ with $a^2=(1+b^2)/2$); Markov's inequality $\sup_{[0,1]}|p'|\le 2d^2\sup_{[0,1]}|p|$; and V. Markov's inequality $\sup_{[-1,1]}|p''|\le \frac{1}{3}d^2(d^2-1)\sup_{[-1,1]}|p|$ [36].

Lemma B.3 (ℓ^1 Bernstein inequality on the cube; [*standard*])**.** *If P is a polynomial on $\mathbb{R}^M$ of total degree at most d and $z\in[-1,1]^M$, then $\sum_l\sqrt{1-z_l^2}\,|\partial_l P(z)|\le d\sup_{[-1,1]^M}|P|$.*

Proof. Write $z_l=\cos\theta_l$. $F(\theta)=P(\cos\theta_1,\dots,\cos\theta_M)$ is a trigonometric polynomial with frequencies $k\in\mathbb{Z}^M$, $\sum_l|k_l|\le d$. For $h\in\{\pm1\}^M$ the function $\varepsilon\mapsto F(\theta+\varepsilon h)$ has integer frequencies $\langle k,h\rangle\in[-d,d]$, so by Bernstein's inequality its derivative at $\varepsilon=0$, which is $-\sum_l h_l\sin\theta_l\,\partial_l P(z)$, has modulus at most $d\sup|P|$. Take $h_l=-\mathrm{sgn}(\sin\theta_l\,\partial_l P(z))$. □

For the next two statements let $R(\mu)=Q(m_1(\mu),\dots,m_d(\mu))$, $m_j(\mu)=\int t^j\,d\mu(t)$, where every monomial of Q is a product $m_{j_1}\cdots m_{j_k}$ with $\sum_r j_r\le d$; R is defined on probability measures μ on $[-1,1]$ and $\varepsilon\mapsto R((1-\varepsilon)\mu+\varepsilon\mu')$ is a polynomial of degree at most d. Let $S_R=\sup_\mu|R(\mu)|$ and let $\psi_\mu(t)=\sum_j(\partial_j Q)(m(\mu))\,t^j$ be the first variation, so that $\frac{d}{d\varepsilon}R(\mu+\varepsilon\eta)|_{\varepsilon=0}=\int\psi_\mu\,d\eta$ for signed η. *Carathéodory reduction:* $R(\mu)$ depends only on $m(\mu)\in\mathrm{conv}\{(t,\dots,t^d):t\in[-1,1]\}$, a compact set, so $S_R<\infty$, and by Carathéodory's theorem every value of R is attained at a measure with at most $d+1$ atoms. For $z\in[-1,1]^N$ and $\nu(z)=N^{-1}\sum_l\delta_{z_l}$, the polynomial $P_N(z)=R(\nu(z))$ has total degree at most d and $\partial_l P_N(z)=N^{-1}\psi'_{\nu(z)}(z_l)$.

Proposition B.4 (Wasserstein–Bernstein; [*proved here*])**.** (a) *For every probability measure μ on $[-1,1]$, $\int\sqrt{1-t^2}\,|\psi'_\mu(t)|\,d\mu(t)\le d\,S_R$, and this is sharp.* (b) *The pointwise form $|\psi'_\mu(t)|\le Cd\,S_R/\sqrt{1-t^2}$ is false for every constant C and $d>C$.*

Proof. (a) For empirical measures $\mu=\nu(z)$, $z\in[-1,1]^N$ (any N), this is Lemma B.3 applied to P_N, since $\sup|P_N|\le S_R$; general μ follow by weak approximation, both sides being continuous in μ. Equality holds for $R=T_d(m_1)$ and $\mu=\delta_t$, $t=\cos(\pi/(2d))$, where $\sqrt{1-t^2}\,|T'_d(t)|=d$. (b) Take $R=T_d(m_1)$, so $S_R=1$ and $\psi'_\mu\equiv T'_d(m_1(\mu))$, and $\mu=(1-\alpha)\delta_1+\alpha\delta_0$: then $\psi'_\mu(0)=T'_d(1-\alpha)\to d^2$ as $\alpha\to0$, while the bound at $t=0$ is Cd. □

The example in (b) is realized by wreath configurations ($M-1$ blocks of $+1$'s and one block with mean 0, w even, $M\ge d^2$), where $|\partial_l P|$ is of order d^2/M at an interior coordinate (about $0.70\,d^2/M$ at $M=d^2$); and $R=T_a(\int T_2\,d\mu)$ with $\mu=(1-\frac{1}{M})\delta_0+\frac{1}{M}\delta_{1/2}$ gives $|\psi'_\mu(\frac12)|\approx 2a^2=d^2/2$ even in the bulk. So a bound of order d^2 must average over typical μ, which (a) does.

Lemma B.5 (threshold; [*proved here*])**.** *Let $|R(\nu)|\le 2$ for every empirical measure ν of M points of $[-1,1]$, and let A be such that $|\partial_\theta[\psi_\mu(\cos\theta)]|\le A\,S_R$ for all μ and θ. If $M\ge 2\pi A$, then $S_R\le 4$. The constant $A=2d^3$ is admissible, so $S_R\le4$ whenever $M\ge 4\pi d^3$.*

Proof. For $A=2d^3$: $\varepsilon\mapsto R((1-\varepsilon)\mu+\varepsilon\delta_t)$ on $[0,1]$ has degree at most d, modulus at most S_R, and derivative $\psi_\mu(t)-\int\psi_\mu\,d\mu$ at 0, so Markov's inequality on $[0,1]$ gives $|\psi_\mu-c|\le 2d^2S_R$ on $[-1,1]$; Bernstein's inequality for the trigonometric polynomial $\theta\mapsto\psi_\mu(\cos\theta)$ of degree at most d gives $A=2d^3$. By the Carathéodory reduction it suffices to bound $|R(\mu)|$ for μ with finitely many atoms. Write the atoms as $\cos\theta$, $\theta\in[0,\pi]$, and split $[0,\pi]$ into $K=\lceil\pi/h\rceil$ intervals of length $\pi/K\le h$ with centres $c_1<\cdots<c_K$. (1) Move each atom linearly in θ to the centre of its interval (displacement at most $h/2$). Along the path all measures μ_τ are probability measures, and $|\frac{d}{d\tau}R(\mu_\tau)|\le\frac{h}{2}\int\sqrt{1-t^2}\,|\psi'_{\mu_\tau}|\,d\mu_\tau\le\frac{h}{2}dS_R$ by Proposition B.4(a); so the result $\mu'=\sum_i p_i\delta_{\cos c_i}$ has $|R(\mu)-R(\mu')|\le hdS_R/2$. (2) Round the cumulative sums $F_i=p_1+\cdots+p_i$ to nearest multiples F'_i of $1/M$, with $F'_K=1$. Rounding is monotone, so $p'_i=F'_i-F'_{i-1}\ge0$, and $\nu=\sum_i p'_i\delta_{\cos c_i}$ is an empirical measure of M points, so $|R(\nu)|\le2$. Summation by parts gives $\nu-\mu'=\sum_{i<K}e_i(\delta_{\cos c_i}-\delta_{\cos c_{i+1}})$ with $|e_i|\le1/(2M)$, so on the segment $[\mu',\nu]$ the derivative

of R is at most $\sum_i |e_i|\,|\psi(\cos c_i) - \psi(\cos c_{i+1})| \le (K-1)\frac{1}{2M}\frac{\pi}{K} A S_R \le \pi A S_R/(2M)$. (3) With $h = 1/(2d)$ and $M \ge 2\pi A$, $|R(\mu)| \le 2 + S_R/4 + S_R/4$, so $S_R \le 2 + S_R/2$ and $S_R \le 4$. □

Grid measures. Let $Y_w = \{-1, -1+\frac{2}{w}, \dots, 1\}$, let Δ_w be the set of probability measures on Y_w, $S_\Delta = \sup_{\Delta_w} |R|$, $\bar\psi_\mu = \psi_\mu - \int \psi_\mu\, d\mu$ and $K_\mu(a,b) = D^2R(\mu)[\delta_a - \mu, \delta_b - \mu]$. For $\mu \in \Delta_w$ and $a, b \in Y_w$ the function $g(\alpha,\beta) = R(\mu + \alpha(\delta_a - \mu) + \beta(\delta_b - \mu))$ on the triangle $\{\alpha, \beta \ge 0,\ \alpha+\beta \le 1\}$ has degree at most d and $|g| \le S_\Delta$; Markov's inequality on $[0,1]$, Markov's inequality on $[0, \frac12]$ twice, and V. Markov's inequality on $[0,1]$ give

$$|\bar\psi_\mu(a)| \le 2d^2 S_\Delta, \qquad |K_\mu(a,b)| \le 16 d^4 S_\Delta, \qquad |K_\mu(a,a)| \le \tfrac43 d^4 S_\Delta. \tag{G}$$

We also use the inequality of Szegő and van der Corput–Schaake [*known*] [50, 53]: a real trigonometric polynomial T of degree at most k with $|T| \le S$ satisfies $T'^2 + k^2 T^2 \le k^2 S^2$; equivalently, a real polynomial p of degree at most k with $|p| \le 1$ on $[-1,1]$ satisfies $(1-t^2)p'(t)^2 + k^2 p(t)^2 \le k^2$ there.

Lemma B.6 (L^2 first-variation bound; [*proved here*])**.** *Let R be as above, let ν be a probability measure on $[-1,1]$ and $S_\nu = \sup\{|R(\mu)| : \mu \ll \nu$ a probability measure$\}$. Then*

$$\mathrm{Var}_\nu(\psi_\nu) = \int \bar\psi_\nu^2\, d\nu \le d^2\,(S_\nu^2 - R(\nu)^2).$$

This is sharp: equality holds for $R = T_d(m_1)$ with d odd and $\nu = \frac12(\delta_{-1} + \delta_1)$.

Proof. Let $h \in L^2(\nu)$ with $\int h\, d\nu = 0$ and $\int h^2\, d\nu = 1$, and let $u_\theta = \cos\theta + h\sin\theta$. Then $\int u_\theta^2\, d\nu = 1$, so $\mu_\theta = u_\theta^2 \nu$ is a probability measure with $\mu_\theta \ll \nu$. Each $m_j(\mu_\theta) = \cos^2\theta\, m_j(\nu) + \sin 2\theta \int t^j h\, d\nu + \sin^2\theta \int t^j h^2\, d\nu$ is affine in $(\cos 2\theta, \sin 2\theta)$, and every monomial of Q has at most d factors, so $R(\mu_\theta) = H(2\theta)$ with H a real trigonometric polynomial of degree at most d and $|H| \le S_\nu$ on $\mathbb{R}$. Since $\frac{d}{d\theta}\mu_\theta|_{\theta=0} = 2h\nu$ has mass 0, $2H'(0) = 2\int \psi_\nu h\, d\nu = 2\int \bar\psi_\nu h\, d\nu$, and the Szegő inequality gives $|\int \bar\psi_\nu h\, d\nu| = |H'(0)| \le d\sqrt{S_\nu^2 - R(\nu)^2}$. Take $h = \bar\psi_\nu / \|\bar\psi_\nu\|_{L^2(\nu)}$ (the claim is trivial if $\bar\psi_\nu = 0$ ν-a.e.). In the equality case $\psi_\nu(t) = T_d'(0)\,t = \pm d\,t$, so $\mathrm{Var}_\nu(\psi_\nu) = d^2$, while $S_\nu = 1$ and $R(\nu) = T_d(0) = 0$. □

[*numerical*]: adversarial searches over random (non-composition) R of weighted degree at most 5 and ν with 2 to 4 atoms, with S_ν computed numerically, found no ratio above 1. The lemma is a bound in the reweighting direction, while the sharp law needs the transport direction $\int (1-t^2)\psi_\nu'^2\, d\nu$ (Lemma B.10 below, for compositions): for $R = T_d(m_1)$ and ν the empirical measure of the block means, $\mathrm{Var}_\nu(\psi_\nu)$ is about d^2/w, while the proof of Theorem B.8 below uses only $S_\nu \le S_\Delta \le 2$, and this is where it loses a factor w.

Lemma B.7 (threshold on the grid; [*proved here*])**.** *Let $w, d \ge 1$ and $M > 2d^3$, and let $|R(\nu)| \le 1$ for every empirical measure ν of M points of Y_w. Then $S_\Delta \le 1/(1 - 2d^3/M)$; in particular $S_\Delta \le 2$ for $M \ge 4d^3$.*

Proof. Let $\mu \in \Delta_w$, $y_i = -1 + 2i/w$ and $F_i = \mu(\{y_0, \dots, y_i\})$, and let F_i' be the multiple of $1/M$ nearest to F_i (with a fixed tie rule, so $F_w' = 1$). Rounding is monotone, so $\nu = \sum_i (F_i' - F_{i-1}')\delta_{y_i}$ ($F_{-1}' = 0$) is an empirical measure of M points of Y_w, and $|R(\nu)| \le 1$. Summation by parts gives $\nu - \mu = \sum_{i<w} e_i(\delta_{y_i} - \delta_{y_{i+1}})$ with $|e_i| \le 1/(2M)$, so for $\mu_\tau = \mu + \tau(\nu - \mu) \in \Delta_w$, $|\frac{d}{d\tau} R(\mu_\tau)| \le \frac{1}{2M}\sum_{i<w} |\psi_{\mu_\tau}(y_{i+1}) - \psi_{\mu_\tau}(y_i)|$. The differences $\psi(y_{i+1}) - \psi(y_i) = \frac{2}{w}\psi'(\xi_i)$, with $\xi_i \in (y_i, y_{i+1})$ increasing, change sign at most $d-1$ times, since ψ' has degree at most $d-1$. So the sequence $(\psi(y_i))_i$ has at most d monotone runs, and its total variation is at most $d(\max_i \psi(y_i) - \min_i \psi(y_i)) \le 4d^3 S_\Delta$ by (G). Hence $|R(\mu)| \le 1 + 2d^3 S_\Delta/M$ for every $\mu \in \Delta_w$; as $S_\Delta < \infty$ (Δ_w is compact), $S_\Delta \le 1/(1 - 2d^3/M)$. □

For grid measures and any w, this replaces the threshold $4\pi^2 d^2(w+1)$ of the discretization step in the proof of Theorem B.1 by $4d^3$. It says nothing about measures off the grid, where Lemma B.5 ($w \geq 4d^2$) is unchanged, and it does not lower the order d^3.

Theorem B.8 (block-level d^2 law for any block size; [*proved here*])**.** *If $f \in \mathcal{B}_d(Mw)$ is invariant under $S_w \wr S_M$, $d \geq 2$, $w \geq 1$ and $M \geq 100d^3$, then* $\mathrm{Var}(f) \leq 15d^2/M$, *hence* $\mathrm{MaxInf}(f) \geq \mathrm{Var}(f)^2/(15d^2 w)$.

Proof. By the structure step of the proof of Theorem B.1, written in the block means $y_l \in Y_w$, $f = R(\nu(y))$ with R as above and $|R| \leq 1$ on empirical measures of M points of Y_w; so Lemma B.7 gives $S := S_\Delta \leq 2$. The y_l are i.i.d. Let y' be an independent copy of y_1, $\eta = \delta_{y'} - \delta_{y_1}$ and $\nu' = \nu + \eta/M$ (block 1 re-randomized). The Efron–Stein inequality over blocks and exchangeability give $\mathrm{Var} f \leq \frac{M}{2}\mathbb{E}(R(\nu') - R(\nu))^2$. (a) $G(\varepsilon) = R(\nu + \varepsilon\eta)$ has $\nu + \varepsilon\eta \in \Delta_w$ for $\varepsilon \in [0, 1/M]$, $G'(0) = \bar\psi_\nu(y') - \bar\psi_\nu(y_1)$ and $|G''| = |K(y', y') - 2K(y', y_1) + K(y_1, y_1)| \leq (\frac{8}{3} + 32)d^4 S \leq 34.7 d^4 S$ by (G), all at $\nu + \varepsilon\eta$; so $|R(\nu') - R(\nu) - G'(0)/M| \leq 17.4 d^4 S/M^2$. (b) By exchangeability and Lemma B.6 (with $S_\nu \leq S$, since $\mu \ll \nu$ implies $\mu \in \Delta_w$), $\mathbb{E}\bar\psi_\nu(y_1)^2 = \mathbb{E}\frac{1}{M}\sum_l \bar\psi_\nu(y_l)^2 = \mathbb{E}\mathrm{Var}_\nu(\psi_\nu) \leq d^2 S^2$. (c) For $t \in Y_w$, $\frac{d}{d\varepsilon}\bar\psi_{\nu+\varepsilon\eta}(t) = K(y', t) - K(y_1, t) - (\bar\psi(y') - \bar\psi(y_1))$, all at $\nu + \varepsilon\eta$, so by (G) $|\bar\psi_{\nu'}(t) - \bar\psi_\nu(t)| \leq (32d^4 + 4d^2)S/M \leq 33 d^4 S/M$. Since y' is an atom of ν' and (ν', y') has the law of (ν, y_1), (b) gives $\mathbb{E}\bar\psi_{\nu'}(y')^2 \leq d^2 S^2$, hence $\mathbb{E}\bar\psi_\nu(y')^2 \leq 2d^2 S^2 + 2(33 d^4 S/M)^2$. (d) So $\mathbb{E}G'(0)^2 \leq 2\mathbb{E}\bar\psi_\nu(y_1)^2 + 2\mathbb{E}\bar\psi_\nu(y')^2 \leq 6d^2 S^2 + 4356 d^8 S^2/M^2$. With $(a+b)^2 \leq 1.1a^2 + 11b^2$ and $d^8/M^2 \leq d^2/10^4$,

$$\begin{aligned}\mathrm{Var} f &\leq \frac{M}{2}\Big[\frac{1.1\,\mathbb{E}G'(0)^2}{M^2} + \frac{11\,(17.4)^2 d^8 S^2}{M^4}\Big] \leq \frac{S^2 d^2}{2M}\big[1.1\,(6 + 0.4356) + 11 \cdot 0.0303\big] \\ &\leq \frac{3.706\, S^2 d^2}{M} \leq \frac{14.9\, d^2}{M}.\end{aligned}$$

Finally, by transitivity $\mathrm{MaxInf} = I/n \geq \mathrm{Var}/n$, and $\mathrm{Var} \leq 15d^2 w/n$ gives $\mathrm{MaxInf} \geq \mathrm{Var}^2/(15d^2 w)$. □

For $M \geq 100d^3$, Lemma B.7 gives $S \leq 1/(1 - 0.02)$, and the same computation gives $\mathrm{Var} f \leq 3.9\, d^2/M$.

Corollary B.9 (small blocks; [*proved here*])**.** *If $f \in \mathcal{B}_d(Mw)$ is invariant under $S_w \wr S_M$ and $w < 4d^2$, then* $\mathrm{MaxInf}(f) \geq \mathrm{Var}(f)^2/(400d^5)$ *for every M.*

Proof. For $d = 1$, $\mathrm{MaxInf} \geq \mathrm{Var}^2$ (the reductions in the proof of Theorem 6.5). Let $d \geq 2$. If $M \geq 100d^3$, Theorem B.8 gives $\mathrm{MaxInf} \geq \mathrm{Var}^2/(15d^2 w) \geq \mathrm{Var}^2/(60d^4)$. If $M < 100d^3$, then $n = Mw < 400d^5$, and $\mathrm{MaxInf} \geq \mathrm{Var}/n \geq \mathrm{Var}^2/(400d^5)$ since $\mathrm{Var} \leq 1$. □

Theorem B.1 gives $640d^6$ here. The d^2 law for $w < 4d^2$ is [*open*]: Theorem B.8 loses a factor w against it, in the step $S_\nu \leq S_\Delta$ (see the remark after Lemma B.6).

Proof of Theorem 6.5. Reductions. Let $n = Mw$. The group is transitive, so $\mathrm{MaxInf}(f) = I(f)/n \geq \mathrm{Var}(f)/n$; hence $I \leq B/n$ implies $\mathrm{MaxInf} \geq \mathrm{Var}^2/B$, and $\mathrm{Var} \leq B'/w$ implies $\mathrm{MaxInf} \geq \mathrm{Var}/(Mw) \geq \mathrm{Var}^2/(B'M)$. For $d = 1$, $f = a + b\sum_i x_i$ with $|a| + n|b| \leq 1$, so $\mathrm{Var} f = nb^2 \leq 1/n$ and $\mathrm{MaxInf} = b^2 = \mathrm{Var}/n \geq \mathrm{Var}^2$. Let $d \geq 2$. If $w < 4d^2$, Corollary B.9 gives $\mathrm{MaxInf} \geq \mathrm{Var}^2/(400d^5) \geq \mathrm{Var}^2/(200d^6)$ (Case A). Let $w \geq 4d^2$, so $w \geq 16$.

Structure and tools. By the structure step of the proof of Theorem B.1, written in the block means $y_l = 2s_l/w - 1$, $f = P(y_1, \ldots, y_M)$ with P symmetric of total degree at most d and $|P| \leq 1$ on the grid $\{-1, -1 + \frac{2}{w}, \ldots, 1\}^M$; by the fundamental theorem on symmetric polynomials $P(z) = R(\nu(z))$ with R as above. (T1) Every point of $[-1, 1]^M$ is within ℓ^∞-distance $1/w \leq 1/(4d^2)$ of the grid, so Lemma C.1 (as in the boundedness step of the proof of Theorem C.4)

gives $\sup_{[-1,1]^M}|P|\le 2$. (T2) By Hoeffding's inequality $\Pr[|y_l|>\frac12]\le 2e^{-w/8}\le 2e^{-d^2/2}$, and $d^4e^{-d^2/2}\le 16e^{-2}<2.166$. (T3) For i in block l let $\varphi(t)=P(y$ with y_l replaced by $t)$ and let $t_\pm$ be the values of y_l for $x_i=\pm1$, so $t_+-t_-=2/w$ and $y_l\in\{t_-,t_+\}$. Taylor's formula at y_l gives $d_if=\frac12(\varphi(t_+)-\varphi(t_-))=w^{-1}\partial_lP(y)+r_i$ with $|r_i|\le w^{-2}\sup_{[t_-,t_+]}|\varphi''|$; the main term is common to the w coordinates of the block, so

$$I(f)\le\frac{2}{w}\mathbb{E}|\nabla P(y)|^2+2\sum_i\mathbb{E}r_i^2,$$

and $[t_-,t_+]\subseteq[-\frac58,\frac58]$ if $|y_l|\le\frac12$.

Case B: $M<4d^3$ (few large blocks; Euclidean gradient). We show $I\le(32d^2+147M)/w$; this part holds for every M. If $\|y\|_\infty\le\frac12$, then $y+tv\in[-1,1]^M$ for every Euclidean unit vector v and $|t|\le\frac12$, so Bernstein's inequality at the centre and (T1) give $|D_vP(y)|\le 4d$, i.e. $|\nabla P(y)|_2\le 4d$; always $|\nabla P|_2\le\sup_{\|v\|_\infty\le1}D_vP\le 4d^2$ by Lemma C.1. With (T2) and a union bound, $\mathbb{E}|\nabla P|^2\le 16d^2+16d^4\cdot 2Me^{-d^2/2}\le 16d^2+69.3M$. For the remainder, if $|y_l|\le\frac12$ then $|\varphi''|\le 2d(d-1)\cdot 2/(1-\frac{25}{64})\le 6.57d^2$ on $[t_-,t_+]$, and always $|\varphi''|\le\frac23d^2(d^2-1)\le 2d^4$ (V. Markov); with $w^2\ge 16d^4$, $\sum_i\mathbb{E}r_i^2\le\frac{Mw}{w^4}(43.2d^4+8d^8e^{-w/8})\le\frac{M}{w}(2.7+0.5\cdot 2.166)\le 3.8\frac{M}{w}$. Hence $I\le(32d^2+146.2M)/w$, so $\mathrm{Var}\le I\le(32d^2+147M)/w$ and $\mathrm{MaxInf}\ge\mathrm{Var}^2/(M(32d^2+147M))$. For $M<4d^3$ and $d\ge2$, $M(32d^2+147M)<4d^3(32d^2+588d^3)=128d^5+2352d^6\le 2416d^6$.

Case C: $M\ge4d^3$ (grid measures only). By Lemma B.7, $S:=S_\Delta\le 2$. For $\nu\in\Delta_w$, (G) gives $|\bar\psi_\nu|\le 2d^2S$ on Y_w; every point of $[-1,1]$ is within $1/w$ of Y_w and $|\bar\psi'_\nu|\le d^2\sup_{[-1,1]}|\bar\psi_\nu|$ (Markov), so $\sup_{[-1,1]}|\bar\psi_\nu|\le 2d^2S/(1-d^2/w)\le 4d^2S$. Hence, for every $\nu\in\Delta_w$, $|\psi'_\nu(t)|\le 4d^3S/\sqrt{1-t^2}\le 8d^3S/\sqrt3$ for $|t|\le\frac12$ (Bernstein) and $|\psi'_\nu|\le 4d^4S$ on $[-1,1]$ (Markov). For i in block l, with $t_\pm$ as in (T3) and $\nu^\pm$ the empirical measures of the block means with $y_l=t_\pm$, we have $\nu^+-\nu^-=\eta:=(\delta_{t_+}-\delta_{t_-})/M$, and $\nu^-+\tau\eta\in\Delta_w$ for $\tau\in[0,1]$, so

$$|d_if|=\tfrac12\Big|\int_0^1\int\psi_{\nu^-+\tau\eta}\,d\eta\,d\tau\Big|\le\frac{1}{2M}\cdot\frac{2}{w}\sup_{\tau\in[0,1],\,t\in[t_-,t_+]}|\psi'_{\nu^-+\tau\eta}(t)|.$$

So $|d_if|\le 8d^3S/(\sqrt3\,n)$ if $[t_-,t_+]\subseteq[-\frac12,\frac12]$, and $|d_if|\le 4d^4S/n$ always. The first case fails only if $|y_l|>\frac12-\frac2w\ge\frac38$, which has probability at most $2e^{-9w/128}\le 2e^{-9d^2/32}$ (Hoeffding, $w\ge16$), and $d^2e^{-9d^2/32}\le 4e^{-9/8}\le 1.299$ for $d\ge2$. Since $I=n\mathrm{Inf}_i$ by transitivity,

$$I\le\frac{S^2}{n}\Big[\frac{64}{3}d^6+32d^8e^{-9d^2/32}\Big]\le\frac{(21.34+41.57)\,S^2d^6}{n}\le\frac{62.91\,S^2d^6}{n}\le\frac{252\,d^6}{n},$$

so $\mathrm{MaxInf}\ge\mathrm{Var}^2/(252d^6)$.

Case C′: $M\ge4\pi d^3$. By (T1), $|R(\nu)|=|P(z)|\le 2$ for every empirical measure of M points, so Lemma B.5 gives $S_R\le 4$. As in its proof, $|\psi_\mu-c|\le 8d^2$ on $[-1,1]$, so Bernstein's and Markov's inequalities give $|\psi'_\mu(t)|\le 8d^3/\sqrt{1-t^2}$ and $|\psi'_\mu|\le 8d^4$; since $\partial_lP(z)=M^{-1}\psi'_{\nu(z)}(z_l)$,

$$|\partial_lP(z)|\le 16d^3/(\sqrt3M)\ \text{ if }|z_l|\le\tfrac12,\qquad |\partial_lP(z)|\le 8d^4/M\ \text{ always.}$$

Main term. Let $L=\{l:|y_l|\le\frac12\}$. Lemma B.3 with (T1) gives $\sum_{l\in L}|\partial_lP(y)|\le 2d/(\sqrt3/2)=4d/\sqrt3$, so $\sum_{l\in L}(\partial_lP)^2\le\max_L|\partial_lP|\sum_L|\partial_lP|\le 64d^4/(3M)$; with (T2), $\mathbb{E}\sum_{l\notin L}(\partial_lP)^2\le 2Me^{-w/8}\cdot 64d^8/M^2$, so $\mathbb{E}|\nabla P(y)|^2\le(\frac{64}{3}d^4+128d^8e^{-d^2/2})/M\le 298.5d^4/M$. *Remainder.* For $\mu=\nu(z)$, $\partial_l^2P(z)=M^{-1}\psi''_\mu(z_l)+M^{-2}\partial_s\partial_t\tilde K_\mu(z_l,z_l)$, where $\tilde K_\mu(s,t)=D^2R(\mu)[\delta_s-\mu,\delta_t-\mu]$. The function $g(a,b)=R(\mu+a(\delta_s-\mu)+b(\delta_t-\mu))$ has degree at most d on $[0,\frac12]^2$, $|g|\le4$ and $\partial_a\partial_bg(0,0)=\tilde K_\mu(s,t)$, so Markov's inequality on $[0,\frac12]$ twice gives $|\tilde K_\mu|\le 64d^4$; as $\tilde K_\mu$ has degree at most d in each variable, $|\partial_s\partial_t\tilde K_\mu|\le 64d^8$, and $\le 64d^6\cdot\frac{64}{39}\le 105.1d^6$ for $|s|,|t|\le\frac58$ (Bernstein

twice). Also $|\psi_\mu''| \le \frac{8}{3}d^6$ (V. Markov) and $|\psi_\mu''| \le 2d(d-1)8d^2 \cdot \frac{64}{39} \le 26.26d^4$ on $[-\frac{5}{8},\frac{5}{8}]$. With $M \ge 4\pi d^3$ and $d \ge 2$ this gives $|\partial_l^2 P(z)| \le 30.44d^4/M$ if $|z_l| \le \frac{5}{8}$ and $|\partial_l^2 P(z)| \le 5.22d^6/M$ always. By (T3) and (T2), and then $w^2 \ge 16d^4$,

$$\sum_i \mathbb{E} r_i^2 \le \frac{n\,[(30.44d^4)^2 + (5.22d^6)^2\, 2e^{-w/8}]}{M^2 w^4} \le \frac{57.92d^4 + 3.41d^8 e^{-d^2/2}}{n} \le \frac{65.3\,d^4}{n}.$$

Hence $I \le (2\cdot 298.5 + 2\cdot 65.3)d^4/n = 727.6\,d^4/n \le 728d^4/n$, and $\mathrm{MaxInf} \ge \mathrm{Var}^2/(728d^4)$. Cases A, B and C give the first statement, with the constant $\max(200, 2416, 252)d^6 = 2416d^6$, and Case C′ gives the second. □

Without Lemma B.5, the Fisher–Bernstein discretization of the proof of Theorem B.1, applied on the simplices of at most $d+1$ atoms given by the Carathéodory reduction, gives $S_R \le 4$ only for $M \ge 4\pi^2 d^2(d+1)$ (and then 728 improves to 700). The exponent 6 comes from Case B with $d^2 \ll M < 4d^3$ (the term $147M$) and from Case C (the pointwise bound on the first variation); Case A is now of order d^5. A pointwise bound $A = O(d^2)$ in Lemma B.5 would move the threshold of Case C′ to $O(d^2)$, and Proposition B.4(b) shows that A must be at least of order d^2. Local maximization of $\mathrm{Var} f$ by successive exact linear programs over wreath-invariant bounded f ($2 \le d \le 4$, $M \le 4$, $w \le 512$, not all pairs) found no f with $n\mathrm{Var} f > d^2$; these local maxima are lower bounds on the supremum, not upper bounds. The largest values found are attained by fully symmetric f, and for odd d the symmetric polynomial $T_d(n^{-1}\sum_i x_i)$ has $n\mathrm{Var} f \to d^2$ and $\mathrm{MaxInf} \cdot d^2/\mathrm{Var}(f)^2 \to 1$ as $n \to \infty$ (for $d = 3$ and $n = 1024$, $n\mathrm{Var} f = 0.9922\,d^2$) [*numerical*]. So the truth appears to be of order d^2; by Example 6.3 this order cannot be improved for the class, since symmetric functions are wreath-symmetric. The d^2 law is proved below for composed wreath functions (Theorem 6.6, any M) and for all wreath-symmetric f with many huge blocks (Theorem 6.8); in the remaining ranges the best bounds are those of Theorem 6.5.

The sharp d^2 law in two regimes. We use the Szegő inequality stated before Lemma B.6. Let

$$R(\mu) = F(L_1(\mu), \dots, L_r(\mu)), \qquad L_i(\mu) = \int \ell_i\, d\mu,$$

where F is a real polynomial on $\mathbb{R}^r$ of total degree at most a with $|F| \le 1$ on $[-1,1]^r$, and the ℓ_i are real polynomials of degree at most b with $|\ell_i| \le 1$ on $[-1,1]$. Then R is of the form above with d replaced by $D = ab$, $S_R \le 1$, and $\psi_\mu'(t) = \sum_i \partial_i F(L(\mu))\,\ell_i'(t)$.

Lemma B.10 (L^2 Wasserstein–Bernstein inequality for compositions; [*proved here*])**.** *For every probability measure μ on $[-1,1]$, $\int (1-t^2)\,\psi_\mu'(t)^2\, d\mu(t) \le a^2b^2\,(1-R(\mu)^2)$. This is sharp: equality holds for $r = 1$, $F = T_a$, $\ell_1 = T_b$ and $\mu = \delta_t$, for every $t \in [-1,1]$.*

Proof. Let $L = L(\mu) \in [-1,1]^r$ and $c_i = \sqrt{1-L_i^2}\,|\partial_i F(L)| \ge 0$. (i) Write $L_i = \cos\theta_i$ and choose $h \in \{\pm 1\}^r$ as in the proof of Lemma B.3. Then $\varepsilon \mapsto F(\cos(\theta_1 + \varepsilon h_1), \dots, \cos(\theta_r + \varepsilon h_r))$ is a real trigonometric polynomial of degree at most a, bounded by 1 because the curve stays in the cube, and its derivative at $\varepsilon = 0$ has modulus $\sum_i c_i$; the Szegő inequality at $\varepsilon = 0$ gives $\sum_i c_i \le a\sqrt{1-F(L)^2}$. (ii) If $|L_i| = 1$, then $\ell_i = \pm 1$ μ-a.e.; these are extreme values of ℓ_i on $[-1,1]$, so $\ell_i' = 0$ at such points of $(-1,1)$, and $\sqrt{1-t^2}\,\ell_i'(t) = 0$ μ-a.e. For $|L_i| < 1$ put $x_i(t) = \sqrt{1-t^2}\,\ell_i'(t)/\sqrt{1-L_i^2}$. Then μ-a.e. $\sqrt{1-t^2}\,|\psi_\mu'(t)| \le \sum_{|L_i|<1} c_i |x_i(t)|$, and the weighted Cauchy–Schwarz inequality gives $(1-t^2)\psi_\mu'(t)^2 \le (\sum_i c_i)(\sum_{|L_i|<1} c_i x_i(t)^2)$. (iii) The Szegő inequality for ℓ_i and Jensen's inequality give $\int (1-t^2)\ell_i'^2\, d\mu \le b^2(1 - \int \ell_i^2\, d\mu) \le b^2(1-L_i^2)$, that is, $\int x_i^2\, d\mu \le b^2$. Integrating (ii) and using (i), $\int (1-t^2)\psi_\mu'^2\, d\mu \le b^2(\sum_i c_i)^2 \le a^2b^2(1-F(L)^2)$. In the equality case, with $t = \cos\theta$, $R(\mu) = \cos(ab\theta)$ and $(1-t^2)\psi_\mu'(t)^2 = a^2b^2\sin^2(ab\theta)$. □

The pointwise analogue $(1-t^2)\psi'_\mu(t)^2 \le Ca^2b^2$ is false, by Proposition B.4(b) with $a=d$ and $b=1$; the lemma averages over μ.

Proof of Theorem 6.6. We use (T2), (T3), the bound $I \le \frac{2}{w}\mathbb{E}|\nabla P|^2 + 2\sum_i \mathbb{E}r_i^2$ and the reduction ($I \le B/n$ implies MaxInf $\ge \mathrm{Var}^2/B$) from the proof of Theorem 6.5. Here $f = P(y)$ with $P(z) = F(\Lambda(z))$, $\Lambda(z) = \frac{1}{M}\sum_k \ell(z_k) \in [-1,1]^r$, so $|P| \le 1$ on all of $[-1,1]^M$ and (T1) is not needed. Moreover $\partial_l P(z) = M^{-1}\psi'_{\nu(z)}(z_l)$, so $M|\nabla P(z)|^2 = \int \psi'^2_{\nu(z)}\, d\nu(z)$, and $\partial_l^2 P = M^{-1}\langle \nabla F(\Lambda), \ell''(z_l)\rangle + M^{-2}D^2F(\Lambda)[\ell'(z_l),\ell'(z_l)]$. Lemma C.1 on $[-1,1]^r$, applied to F and to D_vF, gives $\|\nabla F\|_1 \le 2a^2$ and $|D^2F[v,v]| \le 4a^2(a-1)^2\|v\|_\infty^2$.

Main term. By Lemma B.10, $\int(1-t^2)\psi'^2\, d\nu \le D^2$, so the atoms y_l with $|y_l| \le \frac12$ contribute at most $\frac43 D^2$ to $\int \psi'^2\, d\nu$. Always $|\psi'| \le \|\nabla F\|_1 \max_i \sup|\ell_i'| \le 2a^2b^2 = 2D^2$ (Markov). By (T2) and $xe^{-x/2} \le 4e^{-2}$ for $x = D^2 \ge 4$, $\mathbb{E}M|\nabla P|^2 \le \frac43 D^2 + 8D^4e^{-w/8} \le 5.67D^2$, so $\frac{2}{w}\mathbb{E}|\nabla P|^2 \le 11.4D^2/n$.

Remainder, $M \ge D$. If $|y_l| \le \frac12$, then on $[t_-,t_+] \subseteq [-\frac58,\frac58]$ Bernstein's inequality (twice, resp. once) gives $|\ell_i''| \le 3.282\, b(b-1)$ and $|\ell_i'| \le 1.281\, b$, so $|\partial_l^2 P| \le 6.564\,[a^2b(b-1)/M + a^2(a-1)^2b^2/M^2] \le 6.564(\frac12 D^3 + D^3)/M \le 9.85D^3/M$, using $a^4b^2/M^2 = D^2a^2/M^2 \le D^3/M$. Otherwise Markov's and V. Markov's inequalities give $|\partial_l^2 P| \le \frac23 a^2b^4/M + 4a^4b^4/M^2 \le 4.67D^4/M$. By (T3), (T2) and $w^2 \ge 16D^4$, $2\sum_i \mathbb{E}r_i^2 \le [194.1D^6 + 87.3D^8e^{-w/8}]/(nw^2) \le [12.13D^2 + 5.46D^4e^{-D^2/2}]/n \le 15.1D^2/n$.

Remainder, $M < D$. The function φ of (T3) has degree at most D and $|\varphi| \le 1$ on $[-1,1]$, so $|\varphi''| \le 3.282D^2$ on $[-\frac58,\frac58]$ and $|\varphi''| \le D^4/3$ everywhere; hence $2\sum_i \mathbb{E}r_i^2 \le \frac{2M^2}{D^2w^2}[10.8D^4 + \frac29 D^8e^{-w/8}]\frac{D^2}{n} \le 1.41D^2/n$.

So $I \le (11.4 + 15.1)D^2/n \le 27D^2/n$. □

The law is in $D = ab$, which is only an upper bound on $\deg f$. The constant 27 is not optimized: keeping the exact D-dependence of the tail terms gives about $12.7D^2$ (tighter bookkeeping, not included). [*numerical*]: exact computation of $nI(f)$ from the distribution of the block sums, for compositions with $M \le 3$, $w \in \{4D^2, 8D^2\}$ and $D \le 6$, gave $nI(f)/D^2 \le 0.964$.

Proof of Theorem 6.8. We use the notation, (T1)–(T3) and Case C′ of the proof of Theorem 6.5 ($w \ge 4d^2$ holds). Lemma B.5 gives $S_R \le 4$, so $|\psi_\mu - \int \psi_\mu\, d\mu| \le 8d^2$ for every μ; hence $|\psi'_\mu(t)| \le 8d^3/\sqrt{1-t^2}$ (Bernstein), $|\psi'_\mu| \le 8d^4$ (Markov) and, with $\tau = d^{-3} \le \frac18$, $|\psi''_\mu| \le 16d^3(d-1)/(1-d^{-6}) \le 16d^4$ on $[-\tau,\tau]$ (Bernstein twice). Let $c = \sqrt{1-\tau^2} \ge \sqrt{63/64}$, $W = \{l : |y_l| \le \tau\}$, $N = |W|$ and $\psi = \psi_{\nu(y)}$.

(a) *Window.* Lemma B.3 and (T1) give $\sum_{l\in W}|\psi'(y_l)| = M\sum_{l\in W}|\partial_l P(y)| \le 2dM/c$. On $[-\tau,\tau]$, ψ' oscillates by at most $2\tau\cdot 16d^4 = 32d$, so $\max_W|\psi'| \le 2dM/(cN) + 32d$, and $\frac1M\sum_{l\in W}\psi'(y_l)^2 \le \frac1M \max_W|\psi'|\sum_W|\psi'| \le 4d^2M/(c^2N) + 64d^2/c$.

(b) *Off the window.* By Hoeffding's inequality and $\tau^2 w \ge 8\ln(2d)$, $p_\tau := \Pr[|y_l| > \tau] \le 2e^{-\tau^2 w/2} \le 1/(8d^4)$. For $\tau < |t| \le \frac12$, $\psi'(t)^2 \le \frac43\, 64d^6 \le 85.4d^6$; for $|t| > \frac12$ (probability at most $2e^{-w/8}$), $\psi'(t)^2 \le 64d^8$.

(c) The y_l are independent, so N is binomial with parameters M and $q \ge 1 - p_\tau \ge \frac{127}{128}$. Since $\mathbf{1}_{N\ge1}/N \le 2/(N+1)$ and $\mathbb{E}[1/(N+1)] \le 1/((M+1)q)$, $\mathbb{E}[\mathbf{1}_{N\ge1}/N] \le 2.02/M$. Summing (a) and (b),

$$\mathbb{E}M|\nabla P|^2 \le \frac{8.08d^2}{c^2} + \frac{64d^2}{c} + 85.4d^6p_\tau + 128d^8e^{-w/8} \le (8.21 + 64.6 + 10.7 + 0.01)\, d^2 \le 83.6\, d^2,$$

so $\frac{2}{w}\mathbb{E}|\nabla P|^2 \le 167.2\, d^2/n$.

(d) *Remainder.* By the bounds on $\partial_l^2 P$ in Case C′, $|r_i| \le 30.5d^4/(Mw^2)$ if $|y_l| \le \frac{1}{2}$ and $|r_i| \le 5.22d^6/(Mw^2)$ otherwise, so $2\sum_i \mathbb{E}r_i^2 \le [1861d^8 + 109d^{12}e^{-w/8}]/(nw^2) \le 0.96/n$, because $w^2 \ge 64(\ln 4)^2 d^{12} > 122.9\, d^{12}$.

Hence $I \le 167.2\, d^2/n + 0.96/n \le 170\, d^2/n$. □

Remark B.11 (a conditional threshold in w, and its barrier; the implication is [*proved here*], its hypothesis is [*open*]). Let $K(d)$ be the supremum, over all R of weighted degree at most d with $S_R \le 4$ and over all probability measures μ on $[-1,1]$, of $\sup_{|t|\le 5/8} |\psi_\mu''(t)|$; Bernstein's inequality twice gives $K(d) \le 26.3d^4$. The proof of Theorem 6.8 uses K in two places: the window step needs $\tau K = O(d)$ and $\tau^2 w \ge 8\ln(2d)$, and the remainder needs $w^2 \ge C(K + 8.4d^3)^2/d^2$, since $|\partial_l^2 P| \le (K + 8.4d^3)/M$ in the bulk when $M \ge 4\pi d^3$. Hence $K(d) = O(d^k)$ gives the d^2 law for $M \ge 4\pi d^3$ and $w \ge Cd^{2k-2}\ln(2d)$: $k = 4$ is Theorem 6.8, and $k = 2$ would give w of order $d^2 \ln d$. This route cannot go below $k = 2$: for $R(\mu) = T_a(\int T_b\, d\mu)$, $d = ab$, any even b, and $\mu = \delta_0$, one has $T_b(0) = \pm 1$ and $|\psi_\mu''(0)| = |T_a'(T_b(0))\, T_b''(0)| = a^2b^2 = d^2$, so $K(d) \ge d^2$ (and $K(d) \ge 4d^2$, using $4R$). This is a barrier for this route only, not a lower bound for the true threshold in w. [*open*]: the averaged L^2 bound of Lemma B.10 for polynomials in the moments that are bounded only on the moment body $\mathrm{conv}\{(t,\dots,t^d)\}$ (step (i) of its proof fails, since the curves used there leave the moment body); the threshold $M \ge 4\pi d^3$, which needs a pointwise bound $A = O(d^2)$ in Lemma B.5; and Case A ($w < 4d^2$).

Appendix C. Smooth linear statistics: proofs

This appendix contains the proof of Theorem 6.9 and the results it uses, among them the unsharpened Theorem C.4. The setting and notation are those of Section 6.

Lemma C.1 (Markov on a symmetric convex body; [*known*]: the first bound with d^2 in place of $2d^2$ is [46, Theorem 2], see also [30] (balls), [56] (convex bodies) and [8, Theorem 2(c)]; proof included)**.** *Let $K \subset \mathbb{R}^r$ be compact, convex and symmetric with nonempty interior, and $M_K = \sup_K |P|$. For $u \in K$ and $v \in \mathbb{R}^r$, $|D_v P(u)| \le 2d^2\|v\|_K M_K$, and $|D_{v_1}\cdots D_{v_k}P(u)| \le \prod_{i=0}^{k-1} 2(d-i)^2 \prod_i \|v_i\|_K\, M_K$.*

Proof. Let $\|v\|_K = 1$. For $|s| + |t| \le 1$, $su + tv \in K$. The polynomials $p_\pm(\tau) = P((1-\tau)u \pm \tau v)$, $\tau \in [0,1]$, have degree at most d and are bounded by M_K, so Markov's inequality on $[0,1]$ gives $|p_\pm'(0)| \le 2d^2 M_K$; and $D_v P(u) = (p_+'(0) - p_-'(0))/2$. Iterate, using $\deg D_{v_1} P \le d-1$. □

Lemma C.2 (leverage; [*proved here*])**.** *$\sum_j \delta_j \le r$; hence $\sum_j \delta_j^2 \le r\delta$.*

Proof. For $t \in [0,\infty)^n$ let $V(t) = \mathrm{vol}_r \sum_j t_j[-w_j, w_j]$. For a convex body K, $\mathrm{vol}(K + [-sv, sv]) = \mathrm{vol}K + 2s|v|\mathrm{vol}_{r-1}(\pi_v K)$ (Fubini over the fibres in direction v), so V is affine in each t_j, with $\partial_j V(t) = 2|w_j|\mathrm{vol}_{r-1}(\pi_j Z(t))$, and V is homogeneous of degree r, so $\sum_j \partial_j V(\mathbf{1}) = r\,\mathrm{vol}Z$ by Euler's identity. If $p + [a,b]w_j \subseteq Z$, then by symmetry and convexity $\frac{b-a}{2} w_j \in Z$, so every fibre of Z in direction w_j has length at most $2|w_j|/\delta_j$ and $\mathrm{vol}Z \le 2|w_j|\mathrm{vol}_{r-1}(\pi_j Z)/\delta_j$, i.e. $\delta_j \le \partial_j V(\mathbf{1})/\mathrm{vol}Z$. Sum over j. □

Lemma C.3 (vertex rounding; [*proved here*])**.** *For every $u \in Z$ there is $x \in Q_n$ with $\|u - Wx\|_Z \le r\delta$.*

Proof. Let y' be a vertex of the polytope $\{y \in [-1,1]^n : Wy = u\}$ and $J = \{j : |y_j'| < 1\}$. If $|J| > r$, some $c \ne 0$ supported on J has $Wc = 0$ and $y' \pm \varepsilon c$ stays in the polytope, contradicting extremality; so $|J| \le r$. Put $x_j = y_j'$ for $j \notin J$ and $x_j = \operatorname{sgn} y_j'$ ($\operatorname{sgn} 0 = 1$) for $j \in J$; then $u - Wx = \sum_{j\in J}(y_j' - x_j)w_j$ with $|y_j' - x_j| \le 1$. □

Theorem C.4 (smooth linear statistics; [*proved here*])**.** (a) *If* $d^2r\delta \le 1/4$, *then* $M \le 2$, $\sum_j \|d_j f\|_2 \le 6d^2\sqrt{r}$, *and* $\mathrm{MaxInf}(f) \ge I(f)^2/(36d^4r) \ge \mathrm{Var}(f)^2/(36d^4r)$. (b) *Let* $\varrho = \sup_{u\in Z}\min_{x\in Q_n}\|u - Wx\|_Z$ $(\le r\delta)$. *If* $d^2\varrho \le 1/4$, *then* $M \le 2$, $|d_j f| \le 4d^2\delta_j$ *pointwise,* $\sum_j \|d_j f\|_2 \le 4d^2 r$, *and* $\mathrm{MaxInf}(f) \ge I(f)^2/(16d^4r^2)$.

Proof. Boundedness on Z. For $u \in Z$ take x with $\|u - Wx\|_Z \le \varrho$ and $v = u - Wx$; the segment from Wx to u lies in Z, so Lemma C.1 gives $|P(u)| \le 1 + 2d^2\varrho M$, hence $M \le 1/(1 - 2d^2\varrho) \le 2$; under (a), $\varrho \le r\delta$ by Lemma C.3.

The discrete derivative. Let $\phi(t) = P(Wx - x_j w_j + t w_j)$, a polynomial of degree at most d whose arguments lie in Z for $|t| \le 1$. Then $d_j f(x) = (\phi(1) - \phi(-1))/2$. For (b), the mean value theorem and Lemma C.1 give $|d_j f(x)| \le 2d^2\delta_j M \le 4d^2\delta_j$; sum with Lemma C.2 and use $I \le \mathrm{MaxInf}^{1/2}\sum_j \|d_j f\|_2$. For (a), let $\ell_j(x) = \phi'(x_j) = \langle w_j, \nabla P(Wx)\rangle$. Since $x_j = \pm 1$, $d_j f(x) - \phi'(x_j) = \frac{1}{2}\int_{-1}^{1}(\phi'(t) - \phi'(x_j))\,dt$, so $|d_j f - \ell_j| \le \sup_{|t|\le 1}|\phi''| \le 4d^2(d-1)^2\delta_j^2 M$ by Lemma C.1 with $k = 2$. The ℓ_j lie in the span of the r functions $(\partial_a P)(Wx)$, and pointwise $\sum_j |\ell_j(x)| = \sup_{c\in[-1,1]^n}\langle Wc, \nabla P(Wx)\rangle = \sup_{v\in Z} D_v P(Wx) \le 2d^2 M$; so Proposition 4.12 gives $\sum_j \|\ell_j\|_2 \le 2d^2 M\sqrt{r}$. With Lemma C.2, $\sum_j \|d_j f\|_2 \le 2d^2 M\sqrt{r} + 4d^2(d-1)^2 M\delta r \le 4d^2\sqrt{r} + 2(d-1)^2 \le 6d^2\sqrt{r}$, using $M \le 2$ and $8d^2\delta r \le 2$. Conclude with $I(f) \le \mathrm{MaxInf}(f)^{1/2}\sum_j \|d_j f\|_2$. □

The scale $d^2\delta = O(1)$ is sharp up to the factor r: for $r = 1$ and equal weights, $f = T_d(n^{-1}\sum_i x_i)$ with $n = 4d^2$ satisfies the hypothesis with equality, and beyond this scale the supremum over Z can grow like $e^{cd^2\delta}$ (Coppersmith–Rivlin). We have not found an AA-type statement for $P(Wx)$ in print; the theorem is new when W has many distinct columns; for W with B distinct columns the conjecture already follows from Theorem 6.4. Via Theorem 4.5, it extends to functions that become such polynomials (with leaf-dependent P_l, W_l, $r_l \le r$) after a decision tree of depth t: e.g. under (b) at every leaf, $\mathrm{MaxInf}(f) \ge \min(\mathrm{Var}^2/(4t^2), \mathrm{Var}^4/(589824\,d^{13}r^4t^2))$. The factor d^4 in (a) comes from squaring a Markov bound on all of Z; Bernstein's inequality in the bulk and a tail bound reduce it to d^2, as follows.

The factor d^2. We now remove one factor d^2 from Theorem C.4(a). The proof imports from the proof of Theorem C.4: the boundedness step ($M \le 2$ under $d^2r\delta \le 1/4$, via Lemmas C.1 and C.3); the discrete-derivative step, $|d_j f - \ell_j| \le \sup_{|t|\le 1}|\phi''(t)|$ with $\phi(t) = P(Wx - x_j w_j + t w_j)$ and $\ell_j = \langle w_j, \nabla P(Wx)\rangle$; Lemma C.2; and the kernel-diagonal step of the proof of Proposition 4.12: if h_j lie in a D-dimensional subspace of $L^2(Q_n)$ with kernel diagonal Λ, then $\sum_j \|h_j\|_2 \le \mathbb{E}[(\sum_j |h_j|)\Lambda^{1/2}] \le \sqrt{D}\,\|\sum_j |h_j|\|_2$ by Cauchy–Schwarz and $\mathbb{E}\Lambda = D$. We also use the one-sided Hoeffding inequality $\Pr[\sum_j a_j x_j > t] \le e^{-t^2/(2\sum_j a_j^2)}$.

Lemma C.5 (tail; [*proved here*])**.** $\Pr[\|Wx\|_Z > \frac{1}{2}] \le 9^r \exp(-9/(128\delta))$.

Proof. $h_Z(\theta) = \sup_{u\in Z}\langle\theta, u\rangle = \sum_j |\langle\theta, w_j\rangle|$ is the dual norm of $\|\cdot\|_Z$ (a norm, as W has rank r). Let N be a $\frac{1}{4}$-net of $\{h_Z = 1\}$ in h_Z with $|N| \le 9^r$. Given u, choose θ with $h_Z(\theta) = 1$, $\langle\theta, u\rangle = \|u\|_Z$ and $\theta' \in N$ with $h_Z(\theta - \theta') \le \frac{1}{4}$; then $\|u\|_Z \le \langle\theta', u\rangle + \frac{1}{4}\|u\|_Z$, so $\|Wx\|_Z > \frac{1}{2}$ forces $\langle\theta', Wx\rangle > \frac{3}{8}$ for some $\theta' \in N$. Here $\langle\theta', Wx\rangle = \sum_j x_j\langle\theta', w_j\rangle$ with $|\langle\theta', w_j\rangle| \le h_Z(\theta')\delta_j \le \delta$, so $\sum_j\langle\theta', w_j\rangle^2 \le \delta\sum_j |\langle\theta', w_j\rangle| = \delta$. Apply the one-sided Hoeffding inequality with $t = \frac{3}{8}$ and a union bound over N. □

Proof of Theorem 6.9. By the boundedness step, $M \le 2$. For $d \le 3$, Theorem C.4(a) gives $36d^4r \le 324d^2r$; let $d \ge 4$, so $\delta \le 1/64$. Let $G = \{\|Wx\|_Z \le \frac{1}{2}\}$ and $p = \Pr[G^c]$. Since $9/(128\delta) \ge 9d^2r/32$, Lemma C.5 gives $p \le (9e^{-9d^2/32})^r \le 9e^{-9d^2/32}$, and $d^4 p \le 9d^4 e^{-9d^2/32} \le$

25.6 for $d \geq 4$. *Main terms.* Pointwise $\sum_j |\ell_j(x)| = h_Z(\nabla P(Wx)) = \sup_{v\in Z} D_v P(Wx)$. On G, $Wx + tv \in Z$ for $v \in Z$ and $|t| \leq \frac{1}{2}$, so Bernstein's inequality at the centre gives $\sum_j |\ell_j| \leq 2dM$; always $\sum_j |\ell_j| \leq 2d^2 M$ by Lemma C.1. The ℓ_j lie in the span of the r functions $(\partial_a P)(Wx)$, so the kernel-diagonal step gives $\sum_j \|\ell_j\|_2 \leq \sqrt{r}\, M\sqrt{4d^2 + 4d^4 p}$. *Remainders.* On G the points $Wx + (t - x_j)w_j$, $|t| \leq 1$, have gauge at most $\frac{1}{2} + 2\delta \leq \frac{5}{8}$, so for each such point the chord centred at it in direction w_j/δ_j with gauge half-length $\frac{1}{4}$ lies in Z (its gauge is at most $\frac{7}{8}$), and Bernstein's inequality twice at the centre gives $|\phi''| \leq 32d(d-1)M\delta_j^2 \leq 37d^2 M\delta_j^2$; always $|\phi''| \leq 4d^2(d-1)^2 M\delta_j^2$ by Lemma C.1. By the discrete-derivative step and Lemma C.2, $\sum_j \|d_j f - \ell_j\|_2 \leq \delta r M d^2\sqrt{1369 + 16d^4 p} \leq \frac{M}{4}\sqrt{1369 + 16d^4 p}$. Hence $\sum_j \|d_j f\|_2 \leq 2\sqrt{r}\sqrt{4d^2 + 102.4} + \frac{1}{2}\sqrt{1778.6} \leq 11.73\, d\sqrt{r}$ for $d \geq 4$, and $I(f) \leq \mathrm{MaxInf}(f)^{1/2} \sum_j \|d_j f\|_2$ with $11.73^2 \leq 138$ concludes. □

Appendix D. Graph properties and cyclically invariant functions

This appendix contains the proofs of the results on graph properties and cyclically invariant functions stated in Section 6, with their explicit constants and computer-assisted parts, including the polynomial bound for graph properties (Appendix D.2). It also records barriers for the method and a bound for local cyclic functions.

D.1. **Graph properties.**

Proof of Theorem 6.15. $\mathcal{G}_3$ consists of $K_2, P_3, 2K_2, K_3, P_4, K_{1,3}, P_3 + K_2, 3K_2$, with automorphism groups of orders $2, 2, 8, 6, 2, 6, 4, 48$. By Lemma 6.13, $|b_{K_2}| \leq 3$. For each other $H \in \mathcal{G}_3$ we use a linear programming dual certificate: rational step kernels U_k (the constants $s \in \frac{1}{100}\mathbb{Z} \cap [-1, 1]$ and random two-, three- and four-block kernels with weights and entries in $\frac{1}{64}\mathbb{Z}$) and rational multipliers y_k with $\sum_k y_k t(H', U_k) = [H' = H]$ for all $H' \in \mathcal{G}_3 \cup \{\emptyset\}$, verified in exact rational arithmetic by computer (for each H the certificate uses at most 18 kernels; the kernels and the multipliers y_k are in the file `g3_certificate.json` of the supplementary directory `supplement/`, and the stand-alone checker `verify_g3.py` there verifies all identities in exact rational arithmetic and evaluates the variance bound, $N\mathrm{Var} f \leq 96.2655$ at $M = 8$). Then $b_H = \sum_k y_k \sum_{H'} b_{H'} t(H', U_k)$, so Lemma 6.13 gives $|b_H| \leq \sum_k |y_k| =: B_H$, with $B_{P_3}, B_{2K_2} \leq 4.7408$, $B_{K_3} = 1$, $B_{P_4} \leq 15.4952$, $B_{K_{1,3}} \leq 8.8836$, $B_{P_3+K_2} \leq 42.4344$ and $B_{3K_2} \leq 22.6220$. For $M \geq 8$ every $H \in \mathcal{G}_3$ has $v(H) \leq 6 \leq M$, and

$$N\mathrm{Var} f \leq 9 + \sum_{H\in\mathcal{G}_3\setminus\{K_2\}} B_H^2 \frac{\mathrm{aut}(H)}{2(M-2)_{v(H)-2}},$$

which is nonincreasing in M and equals $96.27\ldots$ at $M = 8$. For $M \leq 7$, $N\mathrm{Var} f \leq N \leq 21$. The second statement follows as in Theorem 6.14(i). □

An independent certificate (constants $s \in \frac{1}{20}\mathbb{Z}$ and two- and three-block kernels with entries in $\{0, \pm\frac{1}{2}, \pm 1\}$) gives $B_{P_3} = B_{2K_2} = 128/27$, $B_{P_4} \leq 15.862$, $B_{K_{1,3}} \leq 9.437$, $B_{P_3+K_2} \leq 43.955$, $B_{3K_2} \leq 23.282$, and the weaker constant 102.

Coefficient bounds level by level. For a kernel $V\colon [0,1]^2 \to [-1,1]$ and $s \in [-1,1]$, the function $s \mapsto b_\emptyset + \sum_H b_H t(H, sV) = b_\emptyset + \sum_k s^k \sum_{e(H)=k} b_H t(H, V)$ is a polynomial of degree at most d, bounded by 1 on $[-1,1]$ by Lemma 6.13. Let $m_k(d) = |[s^k]T_{d'}(s)|$, where $d' \in \{d, d-1\}$ has the parity of k; thus $m_2 = d'^2/2$, $m_3 = d'(d'^2 - 1)/6$, $m_4 = d'^2(d'^2 - 4)/24$, $m_k(d) \leq d^k/k!$ by the arithmetic–geometric mean inequality, and $m_d(d) = 2^{d-1}$. Let K_k° be the set of $\beta \in \mathbb{R}^{\{H:\, e(H)=k\}}$

with $|\sum_{e(H)=k}\beta_H t(H,V)|\le 1$ for all symmetric measurable $V\colon[0,1]^2\to[-1,1]$, and $\nu_H=\max\{|\beta_H|:\beta\in K^\circ_{e(H)}\}$.

Lemma D.1 (rays; [*proved here*]). *K°_k is compact and does not depend on d, $1\le\nu_H<\infty$, and every graph property $f\in\mathcal{B}_d(N)$ satisfies $|b_H|\le\nu_H\,m_{e(H)}(d)$ for every H. Moreover $\nu_{K_3}=\nu_{C_4}=1$, $\nu_{P_3}\le 4$ and $\nu_{2K_2}\le 5$.*

Proof. If $|p|\le 1$ on $[-1,1]$ and $\deg p\le d$, the coefficient of s^k lies in the parity-k part $(p(s)\pm p(-s))/2$, which is bounded by 1 and has degree at most d'; by V. A. Markov's coefficient inequality [36] it is at most $m_k(d)$ in absolute value. Applied to the polynomial above, this says $(b_H)_{e(H)=k}/m_k(d)\in K^\circ_k$. Compactness follows from the linear independence of the densities (proof of Lemma 6.13), and $\nu_H\ge 1$ because $|t(H,V)|\le 1$. For a balanced $g\colon[0,1]\to\{\pm1\}$, the kernel $g(z)g(z')$ has $t(H,\cdot)=1$ if all degrees of H are even and 0 otherwise, and K_3 and C_4 are the only graphs without isolated vertices with 3, resp. 4, edges and all degrees even; so $\nu_{K_3}=\nu_{C_4}=1$. The kernel $(g(z)+g(z'))/2$ has $t(P_3,\cdot)=\frac14$ and $t(2K_2,\cdot)=0$, so $|\beta_{P_3}|\le 4$, and the constant kernel 1 gives $|\beta_{P_3}+\beta_{2K_2}|\le 1$, hence $|\beta_{2K_2}|\le 1+|\beta_{P_3}|\le 5$. □

Lemma D.2 (explicit extraction; [*proved here*]). *Let $V\ge\max v(H)$ over the graphs in a finite family $\mathcal{F}$ without isolated vertices, and let b satisfy $|b_\emptyset+\sum_{H\in\mathcal{F}}b_H t(H,U)|\le 1$ for all U. Then $|b_H|\,\mathrm{aut}(H)\,(V-v(H))!\le 2^V V^V$ for every $H\in\mathcal{F}$. In particular $C(d)\le(4d)^{2d}$ (take $V=2d$), and $\nu_H\le(4e(H))^{2e(H)}/(\mathrm{aut}(H)(2e(H)-v(H))!)$ (take $V=2e(H)$).*

Proof. For step kernels with parts of measures $\alpha_1,\dots,\alpha_V$ and values β_{ij}, let $P(\alpha,\beta)=b_\emptyset(\sum\alpha)^V+\sum_H b_H t(H,U_{\alpha,\beta})(\sum\alpha)^{V-v(H)}$, homogeneous of degree V in α and bounded by 1 on the closed simplex times $[-1,1]^{\binom V2}$ (with the diagonal values β_{ii} fixed, say 0; they do not enter the coefficient). For a homogeneous polynomial of degree V, inclusion–exclusion gives $[\alpha_1\cdots\alpha_V]P=\sum_{S\subseteq[V]}(-1)^{V-|S|}|S|^V P(\mathbf{1}_S/|S|,\cdot)$, which is bounded by $2^V V^V$. Only injective maps $V(H)\to[V]$ contribute to this coefficient, so it is multilinear in the off-diagonal β_{ij}, and each of its coefficients is bounded by its supremum over $\{\pm1\}^{\binom V2}$. The coefficient of $\prod_{ij\in E(F_0)}\beta_{ij}$, for a labelled copy F_0 of H, is $b_H\,\mathrm{aut}(H)(V-v(H))!$. For $C(d)$ apply this to the unit ball of Lemma 6.13, and for ν_H to $K^\circ_{e(H)}$. □

Through Theorem 6.14(iii), Lemma D.2 gives the explicit $K(d)\le(4d)^{12d}$. As a variance bound this is far weaker than $N\mathrm{Var}f\le d^4e^{2d}$ from [19, Proposition 3.18], for every d; what it adds is an explicit bound on every coefficient b_H that does not depend on M.

Corollary D.3 (three- and four-vertex terms; [*proved here, computer-assisted*]). *For every graph property $f\in\mathcal{B}_d(N)$ on $M\ge 4$ vertices,*

$$\sum_{3\le v(H)\le 4}\frac{b_H^2 N}{c_M(H)}\;\le\;\frac{4d^4+d^6/12}{M-2}+\frac{c_4(d)}{(M-2)(M-3)},$$

where $c_4(d)\le 12d^4+14d^6+d^8+1.8\cdot10^{15}d^{10}+3.3\cdot10^{16}d^{12}$.

Proof. By Lemma D.1 and $m_k\le d^k/k!$, $|b_{P_3}|\le 2d^2$ and $|b_{K_3}|\le d^3/6$, which gives the first term. The graphs with four vertices are $2K_2,P_4,K_{1,3},C_4$, the paw, the diamond and K_4, with $(e(H),\mathrm{aut}(H))=(2,8),(3,2),(3,6),(4,8),(4,2),(5,4),(6,24)$, and $c_4(d)=\sum_{v(H)=4}\nu_H^2 m_{e(H)}(d)^2\mathrm{aut}(H)/2$. Exact rational dual certificates over 3001 rational step kernels with at most three blocks and entries in $\{0,\pm\frac12,\pm1\}$ give $\nu_{2K_2}<3.41$, $\nu_{P_4}<15.27$, $\nu_{K_{1,3}}<9.39$ and $\nu_{\mathrm{paw}}<21.45$ (computer-assisted; the certificates are in the file `mlg4_nu_certificates.json` of the supplementary directory `supplement/`, and the

stand-alone checker `verify_mlg4.py` there rebuilds the list of all graphs with $k \le 4$ edges from the ten connected graphs with at most four edges and verifies every identity in exact rational arithmetic), and $\nu_{C_4} = 1$; these give the terms $12d^4 + 14d^6 + d^8$. For the diamond and K_4 we use Lemma D.2. □

With floating-point LP values ($\nu_{\text{diamond}} \approx 27.93$, $\nu_{K_4} \approx 28.72$, without certificates) the last two coefficients become about 0.11 and 0.02 [*numerical*]. So the three- and four-vertex part of the multi-level bound is $\text{poly}(d)/M$, and $|b_H| = O_H(d^{e(H)})$ for every fixed H. A bound polynomial in d for all H is open. Even such a bound would not by itself give a polynomial bound on $N\text{Var}f$, because $\mathcal{G}_d$ contains $2^{\Theta(d)}$ dense graphs with $O(\sqrt{d})$ vertices (compare Remark D.4(a)). The bottleneck is dense H.

Remark D.4 (two barriers; [*proved here*] for (a) and for the asymptotic estimates in (b); the conclusion about Neumann-series inversion is an interpretation of these estimates). (a) *Level-by-level bounds cannot prove* $K(d) \le 2^{o(\sqrt{d})}$. Let H be K_v with $e(K_v) = d$ (for other d, K_v plus a vertex joined to $d - \binom{v}{2}$ of its vertices), so $v \approx \sqrt{2d}$. Since $|t(H,V)| \le 1$, $b = 2^{d-1}e_H = m_d(d)e_H$ satisfies every level constraint $(b_{H'})_{e(H')=k} \in m_k(d)K_k^\circ$, and its term in $N\text{Var}f$ is $4^{d-1}\text{aut}(H)/(2(M-2)_{v(H)-2}) \ge 4^{d-1}/(2M^{v(H)-2})$, which is at most $\text{poly}(d)$ only for $M \ge 2^{\Omega(\sqrt{d})}$; together with $N\text{Var}f \le N$, the relaxation allows $N\text{Var}f = 2^{\Theta(\sqrt{d})}$ (numerically $\log_2(\cdot)/\sqrt{d} \approx 2.1$ to 3.1 for $d \le 210$). Cross-level constraints are needed. (b) *The perturbative blow-down fails.* Map $[M]$ to $[m]$ by a uniformly random σ and put $x_{uv} = y_{\sigma(u)\sigma(v)}$ for $\sigma(u) \ne \sigma(v)$ (and a $[-1,1]$-valued within-class variable otherwise). Averaging over σ gives a graph property of y in $\mathcal{B}_d$ with coefficients $B_Q = \sum_H b_H \pi_{H\to Q}$, so $\sum_Q B_Q^2 w_m(Q) \le \binom{m}{2}$, $w_m(Q) = \text{aut}(Q)/(2(m-2)_{v(Q)-2})$. For $H = Q \sqcup K_{2,t}$, merging the two vertices on the 2-side doubles all $2t$ edges of $K_{2,t}$, which cancel, so $\pi_{H\to Q} \ge (1 - O(v^2/m))/m$, and the corresponding entry of the correction operator in the weighted norm is $\pi_{H\to Q}(w_m(Q)/w_m(H))^{1/2} \asymp c\, m^{t/2}/\sqrt{\text{aut}(K_{2,t})}$ with $c = 1$ for $t \ne 2$; for $C_4 = K_{2,2}$ both diagonals collapse, $c = 2$, and the entry is about $m/\sqrt{2}$. So the correction operator has weighted ℓ^2 norm at least of order $m^{t/2}/\sqrt{2\,t!}$ for every $t \le (d - e(Q))/2$, and no single-scale Neumann-series inversion in this norm gives $\text{poly}(d)$. Within-class variables do not help here: merging u, v removes $2|N(u) \cap N(v)| + [uv \in E]$ edges, so when $e(H) - e(Q)$ is even (as for $Q \sqcup K_{2,t}$) every one-merge reduction merges a non-adjacent pair and uses no within-class variable, and the reductions with at least two merges have total weight $O(v^4/m^2)$; thus $\pi_{H\to Q} = (c + O(v^4/m))/m$ with an integer $c \ge 1$, so $\pi_{H\to Q} \asymp c/m$ for $m \ge Cv^4$. (For odd $e(H) - e(Q)$ adjacent merges occur, e.g. $K_2 \sqcup K_3 \to K_2$, and within-class variables matter.) By Lemma D.1 each correction is $O_H(d^{e(H)}/m)$, so the blow-down works for any fixed finite family of H; uniformity over dense H with $e(H) \approx d$ is the open part.

D.2. **Proof of Theorem 6.16.** We use the notation of Section 6.5. Each copy of H in K_M is the image of exactly $\text{aut}(H)$ injective maps $V(H) \to [M]$, so for $v = v(H) \le M$

$$t_\pm(H,x) = \frac{1}{(M)_v} \sum_{\phi\colon V(H)\hookrightarrow[M]} \prod_{ab\in E(H)} x_{\phi(a)\phi(b)}.$$

This formula defines $t_\pm(H,G)$, and through $f = b_\emptyset + \sum_H b_H t_\pm(H,\cdot)$ also $f(G)$, for every symmetric array $G = (G_{pq})_{p\ne q}$ with entries in $[-1,1]$ indexed by an arbitrary M-element set. If the entries are ± 1, then $f(G) = f(x)$ for the graph x obtained by transporting G to $[M]$ along any bijection, because f is S_M-invariant; in particular $|f(G)| \le 1$. An injective ϕ maps distinct edges of H to distinct pairs, so $t_\pm(H,G)$ is multilinear in the entries of G. Hence, if some entries are independent uniform signs, independent of the other entries, then $\mathbb{E}f(G)$ is the value of f at G with these entries replaced by 0.

Fix $x \in Q_N$, a vertex u, $W = [M] \setminus \{u\}$ and $y' \in \{-1,1\}^W$. Let $y = (x_{uw})_{w \in W}$ be the star of u in x, and let x' be x with this star replaced by y'. Let $T^* = \{(K,K') \in \mathbb{Z}^2 : K, K' \geq 0,\ 1 \leq K + K' \leq M\}$. For $(K,K') \in T^*$ let $R \subseteq W$ be a uniformly random set of size $M - K - K' \in [0, M-1]$, and let $G(K,K')$ be the array on the M-element set $\{a_1, \dots, a_K\} \cup \{a'_1, \dots, a'_{K'}\} \cup R$ with entries $x_{ww'}$ between $w, w' \in R$, y_w between a_i and $w \in R$, y'_w between a'_j and $w \in R$, and independent uniform signs, independent of R, between any two clones $a_i, a_{i'}, a'_j, a'_{j'}$. Put $\Psi(K,K') = \mathbb{E} f(G(K,K'))$.

Lemma D.5 (two-type clone family; [*proved here*])**.** (i) $|\Psi| \leq 1$ *on* T^*. (ii) $\Psi(1,0) = f(x)$ *and* $\Psi(0,1) = f(x')$. (iii) *There is a real polynomial* P *in* (K,K') *of total degree at most* $2d$ *with* $P = \Psi$ *on* T^*.

Proof. (i) Every realization of $G(K,K')$ is a ± 1-labelled complete graph on M vertices. (ii) If $K + K' = 1$ then $R = W$, there are no clone–clone pairs, and $G(1,0)$, $G(0,1)$ are x, x' with u renamed. (iii) It suffices to treat $\mathbb{E} t_{\pm}(H, G(K,K'))$ for $H \in \mathcal{G}_d$ with $v = v(H) \leq M$. Classify the injective maps ϕ from $V(H)$ to the vertex set of $G(K,K')$ by $A = \phi^{-1}\{a_1, \dots, a_K\}$, $A' = \phi^{-1}\{a'_1, \dots, a'_{K'}\}$ and $\psi = \phi|_C$, an injection $C \to R$, where $C = V(H) \setminus (A \cup A')$. If an edge of H has both ends in $A \cup A'$, the product contains a clone–clone sign exactly once and has expectation 0. Otherwise the product equals

$$F_{A,A'}(\psi) = \prod_{ab \in E(H[C])} x_{\psi(a)\psi(b)} \prod_{\substack{ab \in E(H) \\ a \in A,\, b \in C}} y_{\psi(b)} \prod_{\substack{ab \in E(H) \\ a \in A',\, b \in C}} y'_{\psi(b)},$$

which does not depend on $\phi|_{A \cup A'}$, and there are $(K)_{|A|}(K')_{|A'|}$ choices of $\phi|_{A \cup A'}$. For a fixed injection $\psi \colon C \to W$, $\Pr[\psi(C) \subseteq R] = (M - K - K')_{|C|}/(M-1)_{|C|}$, which is 0 when $|C| > M - K - K'$. Hence

$$\mathbb{E} t_{\pm}(H, G(K,K')) = \frac{1}{(M)_v} \sum_{A,A'} \frac{(K)_{|A|}(K')_{|A'|}(M - K - K')_{|C|}}{(M-1)_{|C|}} \sum_{\psi \colon C \hookrightarrow W} F_{A,A'}(\psi),$$

the outer sum over disjoint $A, A' \subseteq V(H)$ such that no edge of H lies in $A \cup A'$. If $|C| = M$, which is possible only when $v(H) = M \leq 2d$, then $(M-1)_{|C|} = 0$, but the inner sum is empty since $|W| = M - 1$; such terms are read as 0. The inner sums do not depend on (K,K'), and $(K)_{|A|}(K')_{|A'|}(M-K-K')_{|C|}$ is a polynomial of total degree $|A| + |A'| + |C| = v(H) \leq 2e(H) \leq 2d$. □

Subsampling W keeps the normalization $(M)_v$ fixed, which is what makes Ψ a polynomial. The value $P(0,0)$ is not a value of the family and can exceed 1 (it reaches 1.5 for $M = 6$, $d = 2$ [*numerical*]); the proof below uses it only through the bound of step (3).

We use three facts about a real polynomial p of degree at most D on an interval J of length ℓ: Markov's inequality $\sup_J |p'| \leq (2D^2/\ell) \sup_J |p|$; the rounding bound of Lemma 6.11(a), that $\sup_J |p| \leq 2B$ if $|p| \leq B$ on a set of points of J that contains both endpoints and has gaps at most $\ell/(2D^2)$; and Chebyshev's extrapolation bound $|p(s)| \leq T_D(1 + 2\delta/\ell) \sup_J |p|$ for s at distance δ from J, with $T_D(1+\eta) = \cosh(D \operatorname{arccosh}(1+\eta)) \leq e^{D\sqrt{2\eta}}$.

Lemma D.6 (a two-variable Markov inequality; [*proved here*])**.** *Let* $D \geq 1$ *be an integer,* $M \geq 10D^4$*, and let* P *be a real polynomial in* (K,K') *of total degree at most* D *with* $|P| \leq 1$ *on* T^**. Then* $|P(1,0) - P(0,1)| \leq 16e\, D^4/M$.

Proof. Let $Q(s,\theta) = P(s\theta, s(1-\theta))$. A monomial $K^i K'^j$ becomes $s^{i+j}\theta^i(1-\theta)^j$, so Q has degree at most D in s and at most D in θ. Let $a = 2D^2$. (1) *Lines.* For an integer $s \in [a, M]$ and $k = 0, \dots, s$, $Q(s, k/s) = P(k, s-k)$ with $(k, s-k) \in T^*$. These points of $[0,1]$ contain 0 and 1

and have gaps $1/s \le 1/(2D^2)$, so $|Q(s,\theta)| \le 2$ for $\theta \in [0,1]$. (2) *Columns.* For fixed $\theta \in [0,1]$, $s \mapsto Q(s,\theta)$ is bounded by 2 at the integers of $[a,M]$, an interval of length $M-a \ge 2D^2$. So $|Q(s,\theta)| \le 4$ for $s \in [a,M]$. (3) *Corner.* For $s \in [0,a]$ the distance to $[a,M]$ is at most a, so $|Q(s,\theta)| \le 4T_D(1+\eta)$ with $\eta = 2a/(M-a)$. Since $M \ge 10D^4 \ge 8D^4 + 2D^2 = a(4D^2+1)$, we have $2D^2\eta \le 1$, i.e. $D\sqrt{2\eta} \le 1$, and so $|Q| \le 4e$ on $[0,M]\times[0,1]$. Thus $|P| \le 4e$ on the closed triangle $\{K, K' \ge 0,\ K + K' \le M\}$, including the corner $(0,0)$, where no lattice bound is available. (4) *Markov in θ, then in s.* Fix $\theta \in [0,1]$ and let $h(s) = \partial_\theta Q(s,\theta)$, a polynomial in s of degree at most D. As a polynomial identity $Q(0,\theta) = P(0,0)$ for all θ, so $h(0) = 0$. Markov's inequality in θ on $[0,1]$ and step (3) give $|h(s)| \le 8eD^2$ for $s \in [0,M]$; Markov's inequality in s on $[0,M]$ gives $|h'| \le 16eD^4/M$ there. Hence $|h(1)| = |h(1)-h(0)| \le 16eD^4/M$. (5) $P(1,0) - P(0,1) = Q(1,1) - Q(1,0) = \int_0^1 \partial_\theta Q(1,\theta)\,d\theta$, and step (4) holds for every θ. □

Theorem D.7 (a star change moves a graph property little; [*proved here*])**.** *Let $d \ge 1$, $M \ge 160d^4$, and let $f \in \mathcal{B}_d(N)$ be a graph property on M vertices. If $x, x' \in Q_N$ differ only on the star of one vertex, then $|f(x) - f(x')| \le 256e\,d^4/M < 696\,d^4/M$.*

Proof. Apply Lemma D.6 to the polynomial P of Lemma D.5, with $D = 2d$; then $10D^4 = 160d^4$ and $16eD^4 = 256e\,d^4$. □

Lemma D.8 (variance through stars; [*proved here*])**.** *Let $f\colon Q_N \to \mathbb{R}$, let $B_u = \{uw : w \ne u\}$ be the star of u, and let ω be the largest change of f when the coordinates in one star are changed. Then $2\mathrm{Var} f \le \sum_{u\in[M]} \mathbb{E}\mathrm{Var}_{B_u} f \le M\omega^2/4$, where Var_{B_u} is the variance over the coordinates in B_u with the others fixed.*

Proof. For a set J of coordinates, $\mathbb{E}\mathrm{Var}_J f = \sum_{S\cap J\ne\emptyset} \widehat{f}(S)^2$. An edge set S meets B_u if and only if S covers u, so $\sum_u \mathbb{E}\mathrm{Var}_{B_u} f = \sum_S |V_S|\widehat{f}(S)^2$, with V_S the set of vertices covered by S, and $|V_S| \ge 2$ for $S \ne \emptyset$. This is an identity, so the overlaps $B_u \cap B_w = \{uw\}$ do no harm. With the other coordinates fixed, f takes values in an interval of length at most ω on each star, so $\mathrm{Var}_{B_u} f \le \omega^2/4$. □

Proof of Theorem 6.16. Let $C = 8192e^2 < 6.06\cdot 10^4$. If $M \ge 160d^4$, Theorem D.7 and Lemma D.8 give $\mathrm{Var} f \le \frac{M}{8}(256e\,d^4/M)^2 = Cd^8/M$. If $M < 160d^4$, then $\mathrm{Var} f \le 1 < 160d^4/M \le Cd^8/M$. Since S_M is transitive on edges, all influences of f are equal, and $\mathrm{MaxInf}(f) = I(f)/N \ge \mathrm{Var}(f)/N$. If $\mathrm{Var} f > 0$, then $M \le Cd^8/\mathrm{Var} f$ and $N < M^2/2$, so $\mathrm{MaxInf}(f) > 2\mathrm{Var}(f)^3/(C^2d^{16})$, and $C^2/2 < 1.84\cdot 10^9$. □

D.3. **Cyclically invariant functions.**

Proof of Theorem 6.18. Write $f(x) = c + a\sum_i x_i + \frac{1}{2}x^{\mathsf{T}}Qx$ with Q a symmetric circulant matrix with zero diagonal, and let $\lambda_m = \sum_k Q_{0k}\cos(2\pi mk/n)$, $m \in \mathbb{Z}_n$, be its eigenvalues (indices mod n). Then $\sum_m \lambda_m = \mathrm{tr}\, Q = 0$, $W_1 = na^2$ and $W_2 = \frac{1}{2}\mathrm{tr}\, Q^2 = \frac{1}{2}\sum_m \lambda_m^2$. The multilinear extension of f is bounded by 1 on $[-1,1]^n$. On the diagonal $x = (s,\dots,s)$ the odd part of f is nas, so $|na| \le 1$. For $x_i = \cos(\vartheta + 2\pi mi/n)$ with ϑ uniform on $[0,2\pi]$, $\mathbb{E}x_i = 0$ and $\mathbb{E}x_ix_j = \frac{1}{2}\cos(2\pi m(i-j)/n)$, so $\mathbb{E}f = c + \frac{n}{4}\lambda_m \in [-1,1]$, i.e. $\lambda_m \in [-\frac{4}{n}(1+c), \frac{4}{n}(1-c)]$. A mean-zero random variable with values in $[-A,B]$ has second moment at most AB, so $\frac{1}{n}\sum_m \lambda_m^2 \le \frac{16}{n^2}(1-c^2)$ and $nW_2 \le 8$. Hence $n\mathrm{Var} f = (na)^2 + nW_2 \le 9$. □

For $k \ge 1$ let $A_k(s) = \hat{f}(\{s_1,\dots,s_k\})$ for $s \in \mathbb{Z}_n^k$ with distinct entries, $A_k(s) = 0$ otherwise, and $\sigma_k(\xi) = \sum_{s\in\mathbb{Z}_n^k} A_k(s)e^{-2\pi i\xi\cdot s/n}$. Cyclic invariance gives $A_k(s + r\mathbf{1}) = A_k(s)$, so σ_k vanishes off $H_k = \{\xi \in \mathbb{Z}_n^k : \sum_t \xi_t = 0\}$, a set of n^{k-1} elements, and Parseval's identity gives

$$nW_k = \frac{1}{k!}\,\mathbb{E}_{\xi\in H_k}|\sigma_k(\xi)|^2.$$

Let $\tau_{d,k}$ be the largest modulus of the coefficient of ρ^k in a real polynomial of degree at most d bounded by 1 on $[-1,1]$. By V. Markov's classical theorem the extremal polynomial is T_d or T_{d-1}, so $\tau_{d,1} \le d$, $\tau_{d,2} \le d^2/2$, $\tau_{d,3} \le d(d^2-1)/6 \le d^3/6$ and $\tau_{d,d} = 2^{d-1}$.

Proof of Theorem 6.19. Write $f = c + a\sum_i x_i + f_2 + f_3$ with f_j homogeneous of degree j and $f_2 = \frac{1}{2}x^{\mathsf T}Qx$ as before. The diagonal $p(s) = f(s,\dots,s)$ is a cubic bounded by 1 on $[-1,1]$ with $p'(0) = na$, so $|na| \le \tau_{3,1} = 3$ and $nW_1 = (na)^2 \le 9$. In the single-phase process of the proof of Theorem 6.18 every monomial of odd degree has odd ϑ-charge and averages to 0, so $\mathbb{E}f = c + \frac{n}{4}\lambda_m$ as before; hence $nW_2 \le 8$ and $|\lambda_m| \le 4(1+|c|)/n$. For level three fix $\xi \in H_3$ and let $x_i = \frac{1}{3}\sum_{l=1}^{3}\cos(\vartheta_l + 2\pi\xi_l i/n)$, with ϑ uniform on $\{\vartheta_1+\vartheta_2+\vartheta_3 = \psi\}$. The average of $e^{iq\cdot\vartheta}$ over this torus is $e^{im\psi}$ if $q = m(1,1,1)$ and 0 otherwise. So the linear terms vanish, the quadratic terms leave only same-phase pairs, giving $Z = \frac{n}{36}\sum_l \lambda_{\xi_l}$, independent of ψ, and the cubic terms leave only $q = \pm(1,1,1)$, giving $\frac{1}{108}\mathrm{Re}(e^{i\psi}\overline{\sigma_3(\xi)}) = R\cos(\psi-\phi)$ with $R = |\sigma_3(\xi)|/108$. Thus $c + Z + R\cos(\psi-\phi) \in [-1,1]$ for every ψ, and comparing $\psi = \phi$ with $\psi = \phi+\pi$ gives $R \le 1$. So $|\sigma_3| \le 108$ on H_3 and $nW_3 \le 108^2/6 = 1944$. □

Proof of Theorem 6.20. Let f_k be the homogeneous part of degree k, fix $\xi \in H_k$ and let $y_i(\vartheta) = \frac{1}{k}\sum_{l=1}^{k}\cos(\vartheta_l + 2\pi\xi_l i/n)$, with ϑ uniform on $\{\sum_l \vartheta_l = \psi\}$. For $\rho \in [-1,1]$, $q_\psi(\rho) = \mathbb{E}_\vartheta f(\rho y(\vartheta))$ is a polynomial of degree at most d bounded by 1, and since $f_j(\rho y) = \rho^j f_j(y)$ its coefficient of ρ^k is $\mathbb{E}_\vartheta f_k(y(\vartheta))$, of modulus at most $\tau_{d,k}$. Only the ϑ-characters $m(1,\dots,1)$ with $|m| \le 1$ survive the average. The terms with $m = \pm 1$ take one factor from each phase with equal signs; summing over the $k!$ assignments and using the symmetry of A_k gives $k^{-k}2^{1-k}\mathrm{Re}(e^{i\psi}\overline{\sigma_k(\xi)})$. The term $m = 0$, which occurs only for even k, gives a quantity $Z_k(\xi)$ independent of ψ and in general nonzero. So the coefficient equals $Z_k + R_k\cos(\psi-\phi)$ with $R_k = k^{-k}2^{1-k}|\sigma_k(\xi)|$, and comparing $\psi = \phi$ with $\psi = \phi+\pi$ gives $R_k \le \tau_{d,k}$. Parseval's identity gives the bound on nW_k. For $k = 1$ the diagonal gives $|na| \le \tau_{d,1} \le d$. For $k = 2$ the single-phase input $\rho\cos(\vartheta + 2\pi mi/n)$ has ρ^2-coefficient exactly $\frac{n}{4}\lambda_m$, so $|\lambda_m| \le 4\tau_{d,2}/n \le 2d^2/n$ and $nW_2 = \frac{n}{2}\sum_m \lambda_m^2 \le 2d^4$. For $k = 3$, $(108\tau_{d,3})^2/6 \le 54d^6$. □

For every transitive group $G \le S_n$, every degree d and every G-invariant $f \in \mathcal{B}_d$, the level-two bound $nW_2 \le 6.7d^4$ also holds [*proof outline*]: if $\frac{1}{2}x^{\mathsf T}Bx$ is the level-two part, the spectral projector of an extreme eigenvalue of B commutes with G, hence has constant diagonal and rescales to a correlation matrix ρ; sign rounding of a Gaussian vector with covariance ρ (the Grothendieck identity and Schur's product theorem) gives a bounded even polynomial whose coefficient of r^2 (for the input scaled by $r \in [-1,1]$) has modulus at least $(\frac{2}{\pi} - \frac{1}{2})n\|B\|$, Markov's coefficient inequality [36] bounds it by $d^2/2$, and $nW_2 \le n^2\|B\|^2/2$. The resulting degree-two total $n\mathrm{Var}f \le 108.2$ is not an improvement over the baseline 32 from the proof of [19, Proposition 3.18] (Section 1.2), so we omit it.

Summing Theorem 6.20 over $k \le d$ gives only $n\mathrm{Var}f \le e^{O(d\log d)}$, worse than [19, Proposition 3.18]. The next two statements show that the uniform symbol bounds used here cannot give a polynomial bound by themselves.

Proposition D.9 (uniform symbol bounds; [*proved here*])**.** *For $d \ge 2$, $f = T_d(n^{-1}\sum_i x_i) \in \mathcal{B}_d(n)$ is cyclically invariant, $\frac{1}{d!}\sup_{H_d}|\sigma_d|^2 \ge \frac{1}{d!}|\sigma_d(0)|^2 = d!\,4^{d-1}\big((n)_d/n^d\big)^2 \to d!\,4^{d-1}$, and $nW_k(f) = O_d(n^{1-k}) \to 0$ for every $k \ge 2$. So any argument that bounds nW_d by $\frac{1}{d!}C_d^2$, where $C_d \ge \sup_{H_d}|\sigma_d|$ holds on the whole class, cannot give a bound on nW_d below $d!\,4^{d-1}(1-o(1))$, although $nW_d(f) \to 0$ for this f.*

Proof. The level-d coefficients come only from the top term $2^{d-1}(n^{-1}\sum_i x_i)^d$, so $\hat f(S) = 2^{d-1}d!/n^d$ for $|S| = d$ and $\sigma_d(0) = (n)_d 2^{d-1}d!/n^d$. For $|S| = k$, $\hat f(S) = O_d(n^{-k})$, so $W_k = O_d(n^{-k})$. □

The identity for $\sigma_k(0)$ is exact only at $k = d$: for $k < d$ the multilinearization of lower powers changes it at finite n, and $\sigma_k(0) = 0$ for $k \not\equiv d \pmod 2$. The large values of σ_d sit near $\xi = 0$, a set of measure about n^{1-d} in H_d; bounds depending on ξ and direct L^2 arguments are not excluded.

Proposition D.10 (single resonances; [*proved here*] except the last sentence)**.** *Let $F\colon \mathbb{T}^r \to [-1,1]$ and let $q_1,\dots,q_k$ be characters of $\mathbb{T}^r$ such that the $2k$ characters $\pm q_t$ are distinct and nonzero. Then $|\prod_t \hat F(q_t)| \le (2k)^{-k/2}$. Consequently, a test input $x_i = F(\vartheta + i\alpha)$ that sees σ_k only through the $k!$ orderings of one such tuple and its conjugate, with the ρ^k-coefficient extracted by Markov's inequality, gives at best $|\sigma_k(\zeta)| \le \tau_{d,k}/(2|\prod_t \hat F(q_t)|)$, hence a level-$k$ bound of at least $\tau_{d,k}^2(2k)^k/(4\,k!)$; at $k = d$, where $\tau_{d,d} = 2^{d-1}$, this is at least $(8e)^d/\mathrm{poly}(d)$. [heuristic]: sign inputs $F = \mathrm{sgn}(\sum_l \cos\vartheta_l)$ attain the amplitude bound up to a factor c^k (central limit heuristic; Monte Carlo for $k \le 8$).*

Proof. By Parseval, $\sum_t 2|\hat F(q_t)|^2 \le \mathbb{E}F^2 \le 1$, since $|\hat F(-q)| = |\hat F(q)|$ and the $\pm q_t$ are distinct; the AM–GM inequality gives $\prod_t |\hat F(q_t)|^2 \le (2k)^{-k}$. The rest is the computation in the proof of Theorem 6.20 with the weight $\prod_t \hat F(q_t)$ in place of $(2k)^{-k}$. □

The scope is exactly this: Proposition D.10 excludes bounds on σ_k at a point (uniform in ξ or not) obtained from one non-aliased resonant tuple per test input with Markov extraction. It does not exclude aliased or multi-tuple inputs (the diagonal input at $\xi = 0$ is one), L^2-averaged families of tests with controlled cross terms, extraction that uses more structure of q_ψ, non-quasi-periodic inputs, or arguments that are not based on test processes (induction on the degree, hypercontractivity).

Theorem D.11 (local cyclic functions; [*proved here*])**.** *Let $l \ge 1$ and let $f\colon Q_n \to [-1,1]$ be multilinear and cyclically invariant, with every monomial contained in an arc of $l+1$ consecutive coordinates (so $\deg f \le l+1$). If $n \ge 2l+2$ then $n\mathrm{Var} f \le (2l+1)(4l+1)^2\gamma^2$ with $\gamma = (n+2l+1)/(n-2l-1)$. For all n, $n\mathrm{Var} f \le 50(l+1)^3$, hence $\mathrm{MaxInf}(f) \ge \mathrm{Var}(f)^2/(50(l+1)^3)$.*

Proof. Let $n \ge 2l+2$. Each orbit of monomials has n elements and a unique representative $S \subseteq \{0,\dots,l\}$ with $0 \in S$, so $f(x) = \frac{1}{n}\sum_{i\in\mathbb{Z}_n} h(x_i,\dots,x_{i+l})$ with $h(y) = c + n\sum_S \hat f(S) y^S$, the sum over these representatives. Cyclic words of length n correspond to closed walks of length n in the de Bruijn graph of order l (vertices $\{\pm1\}^l$, edges the windows in $\{\pm1\}^{l+1}$), and $f(x)$ is $\frac{1}{n}$ times the sum $h(\cdot)$ of h over the edges of the walk. *Repetition bound.* Let C be a closed walk of length $q < n$ through the vertex $\mathbf{1}$, let L_0 be the loop at $\mathbf{1}$, $a = \lfloor n/q \rfloor$ and $c' = n - aq < q$. The walk $C^a L_0^{c'}$ is a cyclic word of length n and $h(L_0) = f(\mathbf{1}) \in [-1,1]$, so $a|h(C)| \le n + c'$ and $|h(C)| \le q(n+q)/(n-q)$. *Potential.* For a vertex v let P_v be the path of length l from $\mathbf{1}$ to v, R_v the path of length l from v to $\mathbf{1}$, and $u(v) = h(P_v)$. For an edge e from v to v' put $g(e) = h(e) + u(v) - u(v')$. Then $g(e) = h(P_v e R_{v'}) - h(P_{v'}R_{v'})$ is a difference of the sums over two closed walks through $\mathbf{1}$ of lengths $2l+1$ and $2l$, so $\|g\|_\infty \le (4l+1)\gamma$ by the repetition bound. The potential terms telescope on closed walks, so $f(x) = \frac{1}{n}\sum_i G_i$ with $G_i = g(x_i,\dots,x_{i+l})$. Under uniform x the G_i are stationary, and G_i, G_j are independent when the windows are disjoint, so $n\mathrm{Var} f = \sum_{|k|\le l} \mathrm{Cov}(G_0, G_k) \le (2l+1)\|g\|_\infty^2$. For $n \ge 10(2l+1)$, $\gamma \le \frac{11}{9}$ and $(2l+1)(4l+1)^2\gamma^2 \le 32(l+1)^3 \cdot \frac{121}{81} < 50(l+1)^3$; for smaller n, $n\mathrm{Var} f \le n < 50(l+1)^3$. □

For comparison, iterated linear-programming ascent over l-local cyclic functions (each step an exact LP over all 2^n points) finds $n\mathrm{Var} f = 1.25,\ 3.00,\ 4.27$ for $l = 1,2,3$ at $n = 12$ and

$n = 16$. These are lower bounds for the maximum (1.25 is the exact maximum for $l = 1$ and even n: with s, t the means of x_i and $x_i x_{i+1}$, the achievable (s, t) span the triangle with vertices $(\pm 1, 1), (0, -1)$, and maximizing $A^2 + B^2$ subject to $|c + As + Bt| \le 1$ there gives $\frac{5}{4}$), and they suggest growth roughly linear in l [*numerical*]; averaging T_{l+1} of the normalized window mean over all rotations gives only $n\mathrm{Var} f \le 1$. Since $\deg f \le l + 1$, [19, Proposition 3.18] gives only $e^{O(l)}$.

Appendix E. Torus-bounded polynomials

This appendix treats polynomials that are bounded on the torus $\mathbb{T}^n$. It contains degree-free bounds for block-symmetric functions and product selectors, and the reduction of the conjecture for this class to the spike inequality. Call a multilinear polynomial $f = \sum_S \widehat{f}(S) z^S$ (complex coefficients) of degree at most d *torus-bounded* if $\sup_{\mathbb{T}^n} |f| \le 1$, equivalently (maximum principle) $|f| \le 1$ on the closed polydisc. Fourier coefficients, $\mathrm{Var} f = \sum_{S \ne \emptyset} |\widehat{f}(S)|^2$ and $\mathrm{Inf}_i f = \sum_{S \ni i} |\widehat{f}(S)|^2$ are those of the restriction to $Q_n \subset \mathbb{T}^n$, so torus-bounded f with real coefficients lie in $\mathcal{B}_d(n)$, and AA for this class is a special case of the conjecture. A one-block form $F(y, z) = \sum_j y_j g_j(z)$ is torus-bounded iff $\sum_j |g_j(z)| \le 1$ on $\mathbb{T}^n$. For such families write $V = \sum_j \|g_j\|_2^2$ and $\varepsilon = \max_j \|g_j\|_2^2$. *AA for torus-bounded f remains open*; this subsection reduces it to a single-function inequality (Conjecture E.7). Except where tagged otherwise, everything below is [*proved here*].

Theorem E.1 (one-block reduction; [*proved here*])**.** *If f is torus-bounded of degree d, then $\sum_j |\partial_j f(z)| \le \frac{\pi}{2} d$ for every $z \in \mathbb{T}^n$. Hence $G(y, z) = \frac{2}{\pi d} \sum_j y_j \partial_j f(z)$ is torus-bounded, with $\mathrm{Var} G = 4I(f)/(\pi^2 d^2)$, $\mathrm{Inf}_{y_j} G = 4\mathrm{Inf}_j(f)/(\pi^2 d^2)$ and $\mathrm{Inf}_{z_i} G \le 4(d-1)\mathrm{Inf}_i(f)/(\pi^2 d^2)$; so AA for torus-bounded one-block forms implies AA for all torus-bounded f with a loss polynomial in d.*

Proof. Fix z and let $c_j = z_j \partial_j f(z)$. For $\theta \in \{-1, 0, 1\}^n$, $\phi(t) = f(z_1 e^{i\theta_1 t}, \dots, z_n e^{i\theta_n t})$ is a trigonometric polynomial of degree at most d bounded by 1, so $|\sum_j \theta_j c_j| = |\phi'(0)| \le d$ by Bernstein's inequality. For $\alpha \in [0, 2\pi)$ take $\theta_j = \mathrm{sgn}\,\mathrm{Re}(e^{-i\alpha} c_j)$: then $\sum_j |\mathrm{Re}(e^{-i\alpha} c_j)| \le d$, and averaging over α with $\mathbb{E}_\alpha |\mathrm{Re}(e^{-i\alpha} c)| = \frac{2}{\pi}|c|$ gives the bound. The identities for G are read off from the Fourier expansion. □

On the cube the corresponding reduction (Lemma 2.4) loses d^4.

Theorem E.2 (block-symmetric torus-bounded functions; [*proved here*])**.** *Let $[n] = B_1 \cup \cdots \cup B_b$ and let f be torus-bounded and invariant under permutations inside each block. Then $\mathrm{MaxInf}(f) \ge (\mathrm{Var} f / b)^2$, with no dependence on d. If $b = 1$ and $d < n$, then $\mathrm{Var} f \le (1 - |\widehat{f}(\emptyset)|^2)/n$.*

Proof. Write $f = \sum_m p_m \prod_s e_{m_s}(z_{B_s}) / \binom{n_s}{m_s}$ with $n_s = |B_s|$ and e_μ the elementary symmetric polynomials. The block-diagonal restriction $p(w) = f(w_1 1_{B_1}, \dots, w_b 1_{B_b}) = \sum_m p_m w^m$ is bounded by 1 on $\mathbb{T}^b$, so $\sum_m |p_m|^2 \le 1$. With $q_m = |p_m|^2 / \prod_s \binom{n_s}{m_s}$, $\mathrm{Var} f = \sum_{m \ne 0} q_m$ and $\mathrm{Inf}_i(f) = n_s^{-1} \sum_m m_s q_m$ for $i \in B_s$. Since $\binom{N}{\mu} \ge N/\mu$ for $1 \le \mu \le N$, $q_m \le |p_m|^2 m_s / n_s$ when $m_s \ge 1$, so by Cauchy–Schwarz $\beta_s := \sum_{m_s \ge 1} q_m \le (\sum_m q_m m_s / n_s)^{1/2} (\sum_m |p_m|^2)^{1/2} \le \mathrm{Inf}_i(f)^{1/2}$; and $\mathrm{Var} f \le \sum_s \beta_s$. For $b = 1$ and $d < n$, $q_m \le |p_m|^2 / n$ for $1 \le m \le d$. □

The bound is attained by $n^{-1} \sum_i z_i$. On the cube symmetric functions only satisfy $\mathrm{MaxInf} \ge \mathrm{Var}^2/(16 d^2)$ (Proposition 6.2), and the Chebyshev examples $T_d(n^{-1} \sum_i x_i)$ show that a factor depending on d is needed there: the diagonal circle $\{w\mathbf{1}\} \subset \mathbb{T}^n$ sees every level, while the cube diagonal sees two points.

A converse [*proved here*], via the Grace–Walsh–Szegő coincidence theorem [49, 54] [*known*]. Conversely, every polynomial p with $|p| \le 1$ on $\mathbb{T}^b$ and $\deg_{w_s} p \le n_s$ is the block-diagonal restriction of the block-symmetric multilinear f obtained by polarizing blockwise, and $\sup_{\mathbb{T}^n} |f| = \sup_{\mathbb{T}^b} |p|$: for fixed other variables f is symmetric and multi-affine in z_{B_s}, and the closed disc is a convex circular region, so $f(z) = f(\zeta 1_{B_s}, \text{rest})$ for some ζ in the closed disc; iterate over the blocks and use the maximum principle. The argument of Theorem E.2, applied to $F = \sum_j y_j g_j$ with the constant parts $\sum_j |\hat{g}_j(\emptyset)|^2 \le \sqrt{\varepsilon}$ and the z-part at most $\mathrm{Inf}_{z_i}(F)^{1/2}$, shows that one-block forms with symmetric selectors g_j satisfy $\max(\varepsilon, \max_i \mathrm{Inf}_{z_i} F) \ge V^2/4$ (this is a two-sided bound, not a bound on ε alone).

Theorem E.3 (address families on the torus; [*proved here*])**.** *Let $h = \sum_{T \subseteq [k]} c_T a^T$ be multilinear and $g_b(a) = h(b \odot a)$, $b \in \{\pm 1\}^k$, with $\sum_b |g_b| \le 1$ on $\mathbb{T}^k$ (or only $\mathbb{E}_a \sum_b |g_b(a)| \le 1$). Then $V = 2^k \|c\|_2^2 \le (\pi^2/16)^k$. On the cube the address function has $V = 1$ for every k.*

Proof. Pair lemma: for $H_0, H_1 \in \mathbb{C}^m$, $\mathbb{E}_{u \in \mathbb{T}} |H_0 + uH_1| \ge \frac{2\sqrt{2}}{\pi}(|H_0|^2 + |H_1|^2)^{1/2}$. Indeed $|H_0 + uH_1|^2 = \alpha + 2\beta\cos(\vartheta + \varphi)$ with $\alpha = |H_0|^2 + |H_1|^2$ and $\beta = |\langle H_0, H_1\rangle| \le \alpha/2$; $\mathbb{E}\sqrt{\alpha + 2\beta\cos\vartheta}$ is non-increasing in β (Chebyshev's association inequality) and equals $\frac{2\sqrt{2}}{\pi}\sqrt{\alpha}$ at $\beta = \alpha/2$. Let $H^{(j)} = (\sum_{T' \subseteq [k-j]} c_{T' \cup U} a^{T'})_{U \subseteq \{k-j+1, \dots, k\}}$, vector-valued. Splitting $H^{(j)} = H_0 + b_{k-j} a_{k-j} H_1$ with $|(H_0, H_1)| = |H^{(j+1)}|$, both signs of b_{k-j} give the same expectation over $u = a_{k-j}$, so the pair lemma gives $\mathbb{E}_a \sum_b |h(b \odot a)| \ge (4\sqrt{2}/\pi)^j \, \mathbb{E} \sum_{b_{\le k-j}} |H^{(j)}(b_{\le k-j} \odot a_{\le k-j})|$ by induction; at $j = k$ the right side is $(4\sqrt{2}/\pi)^k \|c\|_2$. □

For M-ary equivariant families $g_\omega(a) = h(\omega \odot a)$, $\omega \in \mu_M^k$, the same peeling gives $V \le (\pi^2/(8M))^k$ and, as $\varepsilon = V/M^k$, $\varepsilon \ge V^{1 + \log M / \log(8M/\pi^2)}$, with exponent at most 2.435 (attained at $M = 2$).

Proposition E.4 (soft bits; [*proved here*] for (i) and (ii); the consequence is [*heuristic*])**.** (i) *For analytic a on $\mathbb{T}^n$, $\mathrm{Var}(a) = 2\|\mathrm{Im}\, a - \mathrm{Im}\, \hat{a}(\emptyset)\|_2^2 \le 2 \sup |\mathrm{Im}\, a|^2$. Hence for a product family $g_b = \prod_l (1 + b_l a_l)/(2 s_l)$ of soft bits a_l on disjoint blocks, with $s = \sup(|1 + a| + |1 - a|)/2$, one has $V = \prod_l \varphi(a_l)$ with $\varphi(a) = \mathbb{E}(1 + |a|^2)/(2s^2) \le 1 - \min(\mathrm{Var} a, 1)/6$.* (ii) *If P is analytic multilinear of degree at most k with $\mathrm{Re}\, P \ge 0$ on $\mathbb{T}^n$ and $P(0) = 1$, then $\sup_{\mathbb{T}^n} |P^{=m}| \le 2$ for every level $m \ge 1$, so $\mathrm{Var} P \le 4k$.*

Proof. (i) $\mathrm{Im}\, a - \mathrm{Im}\, \hat{a}(\emptyset) = \frac{1}{2i} \sum_{S \ne \emptyset} (\hat{a}(S) z^S - \overline{\hat{a}(S)} z^{-S})$, and the frequencies S, $-S$ are distinct. For the family, $\sum_b |g_b| \le 1$ by construction, $\mathbb{E}|a|^2 \le s^2$, and $s^2 \ge 1 + \sup |\mathrm{Im}\, a|^2 \ge 1 + \mathrm{Var}(a)/2$ (Minkowski), so $\varphi \le \frac{1}{2} + 1/(2 + \mathrm{Var} a) \le 1 - \min(\mathrm{Var} a, 1)/6$. (ii) For fixed $z \in \mathbb{T}^n$, $w \mapsto P(wz) = \sum_m w^m P^{=m}(z)$ has nonnegative real part on the disc and value 1 at 0, so Carathéodory's coefficient inequality gives $|P^{=m}(z)| \le 2$. □

So every soft bit with $\mathrm{Var} a \ge v$ costs a factor at least $1 - \min(v, 1)/6$ in V (on the cube a biased Boolean bit costs nothing); combining (i) and (ii) suggests a root-sum loss $V^{-O(\sqrt{k})}$ for products of soft bits [*heuristic*].

Conjecture E.5 (torus selection conjecture; [*open*])**.** *If $g_1, \dots, g_N$ are analytic multilinear of degree at most $k - 1$ with $\sum_j |g_j| \le 1$ on $\mathbb{T}^n$, then $\varepsilon \ge (V/k)^{O(1)}$.*

With Theorem E.1, Conjecture E.5 with exponent C gives $\mathrm{MaxInf}(f) \ge c(\mathrm{Var} f/d)^{3C}$ for torus-bounded f. It is false on the cube (address function) and it needs multilinearity: in one variable the Fejér family $g_\omega(w) = N^{-1} \omega^{-M} w^M K_M(\omega^{-1} w)$, $\omega \in \mu_N$, $N = 2M + 1$ (K_M the Fejér kernel; $\sum_\omega |g_\omega| \equiv 1$ by quadrature) has $V_1 = \|K_M\|_2^2/N \ge \frac{1}{3}$ and $\varepsilon_1 = V_1/N$, and its

tensor powers in m variables (degree $2Mm$, i.e. $k = 2Mm+1$) violate $\varepsilon \geq (V/k)^C$ as $m \to \infty$ once $N > 3^{C-1}$. A sufficient (not equivalent) spike form is $\mathbb{E}_z \max_j |g_j(z)| \leq \mathrm{poly}(k)\,\varepsilon^c$, since $V \leq \mathbb{E}_z \max_j |g_j|$. *Small cases* [*proved here*]: constant selectors ($k=1$) give $\varepsilon \geq V^2$, tight; affine selectors ($k=2$) give $\varepsilon \geq V^2/2$ by Weissler's inequality $\|g(\cdot/\sqrt{2})\|_2 \leq \|g\|_1$ [55], which in general gives $\varepsilon \geq 2^{1-k}V^2$; analytic selectors of any degree D in one variable give $\varepsilon \geq V^2/(D+1)$ (the kernel-diagonal argument of Proposition 4.12; the span of $1, w, \dots, w^D$ has constant kernel diagonal $D+1$), sharp up to a constant by the Fejér family; selectors depending on a common set of m variables give $\varepsilon \geq V^2/\sum_{i<k}\binom{m}{i}$; and equivariant families give $\varepsilon \geq V^{2.435}$ (above).

Theorem E.6 (product selectors; [*proved here*])**.** *Let* $\theta_1 = 2\log(\pi/(2\sqrt{2}))/\log 2 = 2\ln(\pi/2)/\ln 2 - 1 = 0.302992\ldots$. *If every* $g_j = c_j \prod_{i\in S_j}(\alpha_{ji} + \beta_{ji} z_i)$ *is a product of affine factors in distinct variables (supports arbitrary and overlapping, any* N *and degree) and* $\sum_j |g_j| \leq 1$ *on* $\mathbb{T}^n$*, then* $\varepsilon \geq V^{2/(1-\theta_1)} = V^{2.869\ldots}$.

Proof. For $x \in [0,1]$ let $\ell(x) = 1/\mathbb{E}_\vartheta\sqrt{1+x\cos\vartheta}$ and $\gamma(x) = \sqrt{1+x}$; for $\alpha+\beta z$ with $x = 2|\alpha||\beta|/(|\alpha|^2+|\beta|^2)$, $\|\cdot\|_2/\|\cdot\|_1 = \ell(x)$ and $\|\cdot\|_\infty/\|\cdot\|_2 = \gamma(x)$. The function $F(x) = \frac{\theta_1}{2}\log(1+x) + \log \mathbb{E}\sqrt{1+x\cos\vartheta}$ is concave with $F(0) = F(1) = 0$, so $\ell \leq \gamma^{\theta_1}$ on $[0,1]$. Norms of products in distinct variables multiply, so $\|g_j\|_2/\|g_j\|_1 \leq (\|g_j\|_\infty/\|g_j\|_2)^{\theta_1} \leq \|g_j\|_2^{-\theta_1}$, as $\|g_j\|_\infty \leq 1$. Hence $V = \sum_j \|g_j\|_2^{1+\theta_1}\|g_j\|_2^{1-\theta_1} \leq \varepsilon^{(1-\theta_1)/2}\sum_j \|g_j\|_1 \leq \varepsilon^{(1-\theta_1)/2}$. □

By Theorem E.1, torus-bounded f all of whose derivatives $\partial_j f$ are products of affine factors in distinct variables satisfy $\mathrm{MaxInf}(f) \geq (4/\pi^2)^{1.87}\mathrm{Var}(f)^{2.87}/d^{3.74}$ (exponents rounded up, which is safe as $4/\pi^2, \mathrm{Var} f \leq 1 \leq d$). *The exponent must exceed* 2.1289 [*proved here*]. For $M=4$ the product family $\sigma_4^{-k}\prod_l(1+\omega_l z_l)$, $\omega \in \mu_4^k$, with $\sigma_4 = \sup_u \sum_{\omega\in\mu_4}|1+\omega u| = 4(\cos\frac{\pi}{8} + \sin\frac{\pi}{8})$, has $V = (2+\sqrt{2})^{-k}$ and $\varepsilon = V/4^k$, so $\varepsilon = V^{1+\ln 4/\ln(2+\sqrt{2})} = V^{2.12895\ldots}$ with $V \to 0$; hence Conjecture E.5 fails for every exponent $C < 2.12895$, even in the form $\varepsilon \geq c(V/k)^C$. The optimal exponent lies in $[2.1289, 2.4348]$ for equivariant families and in $[2.1289, 2.8695]$ for product families.

Gluing on the torus. The vector OSSS inequality (L2) of Theorem 4.1 holds on every product probability space $\Omega = \prod_i \Omega_i$, for every decision tree of depth at most t with measurable adaptive queries, in the form $|\mathrm{Cov}(f,h)| \leq \|f - \mathbb{E}f\|_2\,(t\sum_i \delta_i \mathbb{E}\|h - h^{(i)}\|^2)^{1/2}$, where $h^{(i)}$ is h with coordinate i re-randomized [*proved here*]. *Proof*: let x, y be independent, let $i_1, i_2, \dots$ be the coordinates queried on input x, and let $u^{[j]}$ be x with $x_{i_1}, \dots, x_{i_j}$ replaced by $y_{i_1}, \dots, y_{i_j}$. Given the path of x, the unqueried coordinates of x are still product-distributed and y is independent, so $u^{[s]}$ (s the number of queries) has law μ and is independent of the path, and $\mathbb{E}\langle f(x), h(u^{[s]})\rangle = \langle \mathbb{E}f, \mathbb{E}h\rangle$; given the first $j-1$ steps (which fix $i_j = i$), $u^{[j-1]}$ has law μ and $u^{[j]}$ is $u^{[j-1]}$ with coordinate i re-randomized. So $\mathrm{Cov}(f,h) = \mathbb{E}\langle f(x) - \mathbb{E}f, \sum_{j\leq t}(h(u^{[j-1]}) - h(u^{[j]}))\rangle$, and Cauchy–Schwarz twice gives the claim; finiteness is not used. Consequently [*proved here*]: if D is a phase-revealing tree of depth $t \geq 1$ on $\mathbb{T}^n$ (a decision tree on $\mathbb{T}^n$ that at each node queries one coordinate z_i, reveals its value, and chooses the next query measurably; a leaf is the sequence of revealed values, and δ_i is the probability that z_i is queried) with leaf masses $m_j(l) = \mathbb{E}[|g_j|^2 \mid l]$, $\bar{R}_D = \mathbb{E}_l \sum_j m_j^{1/2}$ and τ_D the largest $\mathrm{Inf}_{z_i} F$ over queried i, then $\max(\varepsilon, \tau_D) \geq V^3/(32t^2\bar{R}_D^2)$. Indeed, with $q = (|g_j|^2)_j$ and the leaf-determined $f = (m_j)_j$, $\kappa = \mathbb{E}_l\sum_j m_j^2$: $\mathrm{Cov}(f,q) \geq \kappa - \varepsilon V$, $\|f - \mathbb{E}f\|_2^2 \leq \kappa$, $\mathbb{E}\|q - q^{(i)}\|^2 \leq 4\mathbb{E}\sum_j |g_j - g_j^{(i)}|^2 = 8\mathrm{Inf}_{z_i}F$ (as $|g_j| \leq 1$ and $\mathbb{E}|z_i - z_i'|^2 = 2$), and $\sum_i \delta_i \leq t$; so $\varepsilon < \kappa/(2V)$ forces $\kappa < 32t^2\tau_D$, while Hölder gives $\kappa \geq V^3/\bar{R}_D^2$. If all selectors depend on a common set of t variables, revealing them gives $\bar{R}_D \leq 1$ and $\max(\varepsilon, \tau) \geq V^3/(32t^2)$, polynomial in t (the dimension bound gives only 2^{-t}).

The spike inequality. For analytic multilinear g that is not a multiple of a monomial let $a(g) = \log(\|g\|_2/\|g\|_1)$, $b(g) = \log(\|g\|_\infty/\|g\|_2)$ and $\theta(g) = a(g)/b(g) \in [0,1]$ (Hölder gives $a \le b$; multiples of monomials, where θ is $0/0$, are excluded). Let $\Psi_\theta(g) = e^{a-\theta b}$; the inequality $\mathrm{S}(\theta, C_k)$ is $\|g\|_2^{1+\theta} \le C_k \|g\|_1 \|g\|_\infty^\theta$ for all g of degree at most k, i.e. $\Psi_\theta \le C_k$.

Conjecture E.7 (spike inequality; [*open*])**.** $\sup_g \theta(g) < 1$*, the supremum over all analytic multilinear* g *(any number of variables) that are not multiples of monomials.*

[*proved here*]: (a) *Tensorization.* $\|\cdot\|_1, \|\cdot\|_2, \|\cdot\|_\infty$ are multiplicative over products in disjoint variables, so a, b are additive and $\theta(g_1 g_2)$ lies between $\theta(g_1)$ and $\theta(g_2)$; hence $\mathrm{S}(\theta, 1)$ holds iff $\theta(h) \le \theta$ for every indecomposable h. Since $\Psi_\theta(g^{\otimes m}) = \Psi_\theta(g)^m$ in degree km, $\mathrm{S}(\theta, e^{o(k)})$ implies $\mathrm{S}(\theta, 1)$ for fixed θ. So *Conjecture E.7 is equivalent to* $\mathrm{S}(\theta, \mathrm{poly}(k))$ *for some* $\theta < 1$, the form in which it is used. (b) For one-variable factors $\alpha + \beta z$, θ increases in $x = 2|\alpha||\beta|/(|\alpha|^2 + |\beta|^2)$ and equals $\theta_1 = 0.302992$ at $|\alpha| = |\beta|$ (proof of Theorem E.6); so $\mathrm{S}(\theta_1, 1)$ holds on products of one-variable factors in distinct variables, and Conjecture E.7 concerns non-product atoms (an affine form in several variables is itself such an atom). Products of $(1+z)/2$ give $\Psi_\theta = 2^{k(\theta_1-\theta)/2}$, so $\mathrm{S}(\theta, C_k)$ with $\theta < \theta_1$ needs $\log C_k \ge k(\theta_1 - \theta)(\ln 2)/2$. (c) *Reduction.* $\mathrm{S}(\theta, C_k)$ gives, exactly as in the proof of Theorem E.6, $\varepsilon \ge (V/C_k)^{2/(1-\theta)}$, hence by Theorem E.1 $\mathrm{MaxInf}(f) \ge \frac{\pi^2 d^2}{4}\big(4\mathrm{Var} f/(\pi^2 d^2 C_d)\big)^{2/(1-\theta)}$ for torus-bounded f; so Conjecture E.7 implies polynomial AA for torus-bounded f. (d) *Weak forms.* Hölder and Weissler give $\mathrm{S}(\theta, 2^{k(1-\theta)/2})$ for every θ, which reproduces only the bound $\varepsilon \ge 2^{-k} V^2$. A weak form with parameters (θ_d, C_d) gives a bound $e^{-o(d)}$ (better than 2^{-d}) exactly when $(\log C_d + \log(d/\mathrm{Var} f))/(1-\theta_d) = o(d)$; requiring separately $\log C_k/(1-\theta_k) = o(k)$ and $\log(1/\mathrm{Var})/(1-\theta) = o(d)$ is not enough ($C_k = 1$, $1 - \theta_k = 2\ln k/k$, $\mathrm{Var} \ge \frac{1}{2}$ gives only $e^{-(2+o(1))d}$). Fixed $\theta < 1$ with $C_k = C^{\sqrt{k}}$ is equivalent to $\mathrm{S}(\theta, 1)$ by (a), so the genuinely intermediate weak forms have $\theta_k \to 1$. [*numerical*]: searches over general g in at most 6 variables and symmetric polarized g in at most 16 variables found no g with $\theta(g)$ above θ_1 beyond the grid error (± 0.002); the best points found were products of balanced affine factors. The search is local, with few restarts for 5 and 6 variables, so it does not show that products are the only maximizers.

Entropy criterion and the flat regime [*proved here*]. For analytic multilinear g let $F = |g|^2$, $q = \mathbb{E}F$, $\nu = F\,dm/q$ and $\mathrm{KL}(\nu) = \mathrm{Ent}(F)/q$. Jensen's inequality for $\mathbb{E}_\nu F^{-1/2}$ gives $\log(\|g\|_2/\|g\|_1) \le \frac{1}{2}\mathrm{KL}(\nu)$, so $\theta(g) \le \theta_{\mathrm{ent}}(g) := \mathrm{KL}(\nu)/\log(\|g\|_\infty^2/\|g\|_2^2)$. This step uses no analyticity (it holds for every measurable g), so analyticity can enter only through bounds on $\mathrm{KL}(\nu)$. Since KL and $\log(\|g\|_\infty^2/\|g\|_2^2)$ are additive over products in disjoint variables, θ_{ent} obeys the same mediant rule as θ; but $\theta_{\mathrm{ent}}((1+z)/2) = (1-\ln 2)/\ln 2 = 0.442695\ldots > \theta_1$, so this route gives at best the exponent $2/(1-0.4427) \approx 3.59$ in Conjecture E.5. Weissler's inequality $\|g\|_1 \ge \|g(\cdot/\sqrt{2})\|_2$ and Jensen's inequality for $2^{-|S|}$ give $\log(\|g\|_2/\|g\|_1) \le \frac{\ln 2}{2}\bar{d}$, with $\bar{d} = \sum_S |S||\hat{g}(S)|^2/\|g\|_2^2$ the mean level; hence $\theta(g) < 1 - \eta$ unless g is in the *flat regime* $\|g\|_2 \ge 2^{-\bar{d}/(2(1-\eta))}\|g\|_\infty$. So the bound $\theta(g) \le 1 - \eta_d$ for all g of degree at most d is equivalent to the same bound in the flat regime; it would give $\varepsilon \ge V^{2/\eta_d}$ and $\mathrm{MaxInf}(f) \ge (c\mathrm{Var} f/d^2)^{O(1/\eta_d)}$ for torus-bounded f (for instance $(\mathrm{Var} f/d)^{O(\sqrt{d})}$ if $\eta_d = c/\sqrt{d}$), and a subexponential bound by this route needs $\eta_d = \omega(\log(d/\mathrm{Var} f)/d)$. Two-norm bounds $\log(\|g\|_2/\|g\|_1) \le c\kappa$ alone (κ the mean level or the number of variables) give only $\varepsilon \ge e^{-2c\kappa} V^2$, exponential in the degree. [*numerical*]: local searches in at most 4 variables found no g with $\theta_{\mathrm{ent}}(g) > 0.442695$, the best points being products of balanced factors, possibly in substituted variables z^S.

Conjecture E.7 in each fixed degree. Here g is an analytic polynomial on $\mathbb{T}^n$ of total degree at most d, multilinear or not, in any number n of variables, with $\|g\|_\infty = 1$ and $q = \|g\|_2^2 < 1$; $\theta(g)$ is defined as above. Let Θ_d be the supremum of $\theta(g)$ over all such g that are not multiples

of monomials, and let

$$c_4(d) = \inf_p \frac{\mathbb{E}|p| - \mathbb{E}|p|^2}{1 - \mathbb{E}|p|^2},$$

the infimum over one-variable analytic polynomials p of degree at most d with $\sup_{\mathbb{T}} |p| = 1$ and $|p| \not\equiv 1$.

Theorem E.8 (the spike inequality in fixed degree, uniformly in n; [*proved here*])**.** (i) $c_4(d) \geq 1/(6\pi d)$ *for every* $d \geq 1$*, and* $c_4(1) = 4/\pi - 1$. (ii) *For all* $\lambda, \tau \in (0,1)$ *with* $\lambda^2 + \tau < 1$ *and every* g *as above,*

$$\theta(g) \leq \max\Big(\frac{d \ln 2}{\ln(1/(\lambda^2+\tau))},\ 1 - c_4(d)\,\lambda(1-\lambda^2)\tau\Big).$$

(iii) *Hence* $\Theta_d \leq 1 - \eta_d$ *with* $\eta_d \geq c\, c_4(d)\, 2^{-3d/2} \geq c'\, 2^{-3d/2}/d$ *for absolute constants* $c, c' > 0$*;* $\Theta_1 \leq 0.96922$ (*all affine forms, any* n)*, and* $\Theta_2 \leq 0.99878$.

Proof. Circle lemma. If $\deg p \leq d$ and $M = \sup_{\mathbb{T}} |p| \leq 1$, then $\mathbb{E}|p|(1-|p|) \geq c_4(d)\, M\, \mathbb{E}(1-|p|^2)$. Indeed, write $p = M\tilde{p}$, $\alpha = \mathbb{E}|\tilde{p}|$ and $\beta = \mathbb{E}|\tilde{p}|^2 \geq \alpha^2$; the ratio $\rho(M) = (\alpha - M\beta)/(1 - M^2\beta)$ satisfies $\mathbb{E}|p|(1-|p|) = M\rho(M)\mathbb{E}(1-|p|^2)$ and $\rho'(M) = \beta(-1 + 2M\alpha - M^2\beta)/(1-M^2\beta)^2 \leq -\beta(1-M\alpha)^2/(1-M^2\beta)^2 \leq 0$, so $\rho(M) \geq \rho(1) \geq c_4(d)$ (if $|\tilde{p}| \equiv 1$, then $\rho = 1/(1+M) \geq \frac{1}{2} > c_4(1) \geq c_4(d)$).

(ii) Let $L = \{|g| < \lambda\}$. If $\Pr[L] > 1-\tau$, then $q < \lambda^2 + \tau$, and Weissler's inequality [55] in degree d, $\|g\|_2 \leq 2^{d/2}\|g\|_1$, gives $a(g) \leq \frac{d}{2}\ln 2$ while $b(g) = \frac{1}{2}\ln(1/q)$, so $\theta(g) < d\ln 2/\ln(1/(\lambda^2+\tau))$. Let $\Pr[L] \leq 1 - \tau$. (1) For $s \in \{0,1\}^n$ and $\xi \in \mathbb{T}^n$, $p_{\xi,s}(w) = g(\xi_1 w^{s_1}, \dots, \xi_n w^{s_n})$ is a one-variable analytic polynomial of degree at most d, and $(\xi, w) \mapsto (\xi_i w^{s_i})_i$ preserves Haar measure; so $\mathbb{E}|g|(1-|g|) = \mathbb{E}_\xi \mathbb{E}_w |p_{\xi,s}|(1-|p_{\xi,s}|)$. (2) Apply the circle lemma on the s-circles $\{(\xi_i w^{s_i})_i : w \in \mathbb{T}\}$ on which $\sup|g| \geq \lambda$, and drop the others (the integrand is nonnegative). With Z_s the union of the s-circles on which $\sup|g| < \lambda$ ($Z_s \subseteq L$), $\mathbb{E}|g|(1-|g|) \geq c_4\lambda\,\mathbb{E}[(1-|g|^2)\mathbf{1}_{Z_s^c}] \geq c_4\lambda[(1-\lambda^2)\Pr[L\setminus Z_s] + \mathbb{E}(1-|g|^2)\mathbf{1}_{L^c}]$. (3) Let A_s be the average along s-circles, i.e. the Fourier projection onto $\{\alpha \in \mathbb{Z}^n : \langle \alpha, s\rangle = 0\}$. Since $A_s\mathbf{1}_L = 1$ on Z_s, $\Pr[Z_s] \leq \|A_s\mathbf{1}_L\|_2^2 = \sum_{\langle\alpha,s\rangle=0} |\hat{\mathbf{1}}_L(\alpha)|^2$, and for $\alpha \neq 0$ and uniform s, $\Pr_s[\langle\alpha, s\rangle = 0] \leq \frac{1}{2}$ (flip one s_i with $\alpha_i \neq 0$). So $\mathbb{E}_s \Pr[Z_s] \leq \Pr[L]^2 + \frac{1}{2}(\Pr[L] - \Pr[L]^2)$, that is, $\mathbb{E}_s \Pr[L \setminus Z_s] \geq \frac{1}{2}\Pr[L](1 - \Pr[L])$, independently of n. (4) Averaging (2) over s and using $1 - \Pr[L] \geq \tau$, $(1-\lambda^2)\tau/2 \leq 1$ and $1 - q \leq \Pr[L] + \mathbb{E}(1-|g|^2)\mathbf{1}_{L^c}$, we get $\|g\|_1 - q = \mathbb{E}|g|(1-|g|) \geq c_5(1-q)$ with $c_5 = c_4(d)\lambda(1-\lambda^2)\tau/2$. Then $\|g\|_1/q \geq 1 + c_5(1/q - 1) \geq q^{-c_5}$ (as $\ln(1 + c(u-1)) \geq c\ln u$ for $c \leq 1 \leq u$), so $b - a = \ln(\|g\|_1/q) \geq 2c_5 b$ and $\theta(g) \leq 1 - 2c_5$.

(i) *Lower bound.* Let $u = |p|$, $\sup u = 1$; by Bernstein's inequality u is d-Lipschitz in the angle. If $\min u \geq m$, then $u(1-u) \geq m(1-u) \geq \frac{m}{2}(1-u^2)$, so the ratio is at least $m/2$. Otherwise each of the two arcs from a minimum point to a maximum point crosses every level $v \in [m, 1)$, and the area formula gives $\mathbb{E}u(1-u) \geq \frac{1}{\pi d}(\frac{1}{6} - \frac{m^2}{2} + \frac{m^3}{3})$; also $1 - \mathbb{E}u^2 \leq 1 - 1/(3\pi d)$, since $u \geq 1 - d|\phi - \phi^*|$ near a maximum point ϕ^*, so the ratio is at least $(\frac{1}{6} - \frac{m^2}{2} + \frac{m^3}{3})/(\pi d - \frac{1}{3})$. Taking $m = 1/(3\pi d)$ gives $c_4(d) \geq 1/(6\pi d)$. *Degree* 1. For $p = \alpha + \beta w$ with $\alpha, \beta \geq 0$, $\alpha + \beta = 1$ and $k = 2\sqrt{\alpha\beta}$, $\mathbb{E}|p| = \frac{2}{\pi}\mathbf{E}(k)$ (complete elliptic integral of the second kind) and $1 - \mathbb{E}|p|^2 = k^2/2$; the series $\frac{2}{\pi}\mathbf{E}(k) = 1 - \frac{k^2}{4} - \sum_{j\geq 2} e_j k^{2j}$ with $e_j > 0$ shows that the ratio $\frac{1}{2} - 2\sum_{j\geq 2} e_j k^{2j-2}$ decreases in k, so its minimum is at $k = 1$, where it equals $4/\pi - 1$.

(iii) Optimizing (ii) over (λ, τ) with $c_4(1) = 4/\pi - 1$ gives $0.969217\ldots \leq 0.96922$ for $d = 1$, and with $c_4(2) \geq 0.02776$ (the maximum over m in the proof of (i)) gives 0.99878 for $d = 2$. For general d take $\lambda^2 = x/3$ and $\tau = 2x/3$ with $x = 2^{-d/(1-\eta)}$; this gives $\eta_d \geq \frac{2}{3\sqrt{3}} c_4(d) 2^{-3d/2}(1 - o(1))$. ∎

Remarks on Theorem E.8. (a) [*conditional*]: the explicit p with $|p| = |1+r\cos\phi|$, $r = 1.180544$, gives $c_4(2) \le 0.1886352$, and global numerical searches find no smaller value; if $c_4(2) \ge 0.1886$ (not certified), then $\Theta_2 \le 0.9918$. (b) *No AA gain* [*proved here*]: case $\Pr[L] > 1-\tau$ gives $\theta < 1$ only if $\lambda^2 + \tau < 2^{-d}$, and then $\lambda(1-\lambda^2)\tau < 0.385 \cdot 2^{-3d/2}$, so every parameter choice in this argument gives $\eta_d < 0.385\, c_4(d) 2^{-3d/2}$, exponentially small; a bound $e^{-o(d)}$ through $\theta \le 1-\eta_d$ needs $\eta_d = \omega(\log(d/\mathrm{Var} f)/d)$, (item (d) above with $C_d = 1$), so Theorem E.8 gives no AA gain. The bottleneck is Weissler's factor $2^{d/2}$ in the spiky case. (c) *Scope.* For fixed d the form with a constant is trivial (Weissler gives $\mathrm{S}(0, 2^{d/2})$), so the content of Theorem E.8 is the constant-1 form $\theta(g) \le 1 - \eta_d$. Some decay of η_d is forced for non-multilinear g: the Dirichlet kernels $D = \sum_{k=0}^{d} w^k$ have $\|D\|_\infty = d+1$, $\|D\|_2 = \sqrt{d+1}$ and $\|D\|_1 \sim \frac{4}{\pi^2} \ln d$ (Lebesgue constants), so $\theta(D) = 1 - \ln \|D\|_1 / \ln \sqrt{d+1} \to 1$. For multilinear g, Conjecture E.7 (degree-free) remains [*open*], and the bound $\Theta_1 \le 0.96922$ is far from $\theta_1 = 0.302992$, the value at $(1+z)/2$.

Appendix F. Separator theorems

This appendix contains separator theorems for disjointly supported and sparse selectors. They are only loosely connected to the main line of the paper.

F.1. **Disjointly supported selectors.** The following separator theorem extends a two-polynomial lemma of [32, Lemma 3.5] to any number of polynomials, with a depth independent of their number.

Theorem F.1 (disjoint-support separator; [*proved here*])**.** *Let $g_1, \dots, g_m : Q_n \to \mathbb{R}$ be multilinear of degree at most k with $g_j g_l \equiv 0$ for $j \ne l$. There is a decision tree of depth at most $(2k-1) \cdot 32k^3 \le 64k^4$ such that on every leaf at most one g_j is not identically zero.*

Proof. Work on the current subcube S; stop if at most one g_l is nonzero on S. Otherwise every nonzero g_l has degree at least 1 on S. Let $Q = \sum_l g_l^2$, $z^* \in S$ a maximizer of Q, j^* the unique index with $g_{j^*}(z^*) \ne 0$, and $s = g_{j^*}^2 - \sum_{l \ne j^*} g_l^2$ (degree at most $2k$); disjointness gives $|s| = Q \le s(z^*)$ on S. Let $H = (g_l)_{l \ne j^*}$, $D_H = \deg H$, and $M_1, \dots, M_b$ a maximal family of pairwise disjoint monomials of degree D_H of the g_l, $l \ne j^*$. Restricting H to the subcube that varies M_i (other coordinates as in z^*) keeps the top coefficient $\widehat{H}(M_i) \ne 0$, so some $E_i \subseteq M_i$ has $H(z^{*E_i}) \ne 0$, hence $g_{j^*}(z^{*E_i}) = 0$ and $s(z^{*E_i}) < 0$. The function $r(w) = s(z^* \text{ with } E_i \text{ flipped iff } w_i = -1)/s(z^*)$ on Q_b has degree at most $2k$, $|r| \le 1$, $r(\mathbf{1}) = 1$ and $r(\mathbf{1}^{\oplus i}) < 0$, so $|d_i r(\mathbf{1})| > \frac{1}{2}$ for all i and Lemma 2.2 gives $b < 32k^2$. Query $U = M_1 \cup \dots \cup M_b$ ($|U| < 32k^3$). By maximality every top monomial of every g_l, $l \ne j^*$, of degree D_H meets U, so afterwards these g_l have degree at most $D_H - 1$. The potential $\Phi = D_1 + D_2$ (the two largest degrees among the nonzero restricted g_l, with multiplicity) drops by at least 1 in each stage, is at most $2k$ initially and at least 2 while the procedure runs; so there are at most $2k-1$ stages. □

Corollary F.2 ([*proved here*])**.** *If $F = \sum_j y_j g_j$ with $|g_j| \le 1$, $\deg g_j \le k$ and pairwise disjoint supports, then $\max(\varepsilon, \tau) \ge V^2/(16t^2) \ge V^2/(2^{16}k^8)$.*

Proof. On each leaf of the tree of Theorem F.1 at most one c_{jl} is nonzero, so $\kappa_D = \sum_l p_l V_l^2 \ge V^2$. By Theorem 4.2, $\varepsilon \ge V/2$ or $\tau \ge V^2/(16t^2)$. (If $t = 0$, at most one g_j is nonzero and $\varepsilon = V$.) □

For two polynomials this is the separator of [32]; the content for the conjecture is the independence of the depth from m. The level-set analogue (approximately disjoint supports) is the L^2-sharp selector case, for which the average-case Nisan lemma of Section 7 is the missing ingredient.

F.2. **Sparse supports.** For multilinear $g_1, \dots, g_m$ on Q_n let $A(z) = \{l : g_l(z) \neq 0\}$. On a subcube S write $g_l|_S$ for the restriction (a multilinear polynomial in the free coordinates) and $\deg 0 = 0$.

Theorem F.3 (r-sparse separator; [*proved here*])**.** *Let $g_1, \dots, g_m \colon Q_n \to \mathbb{R}$ be multilinear of degree at most k with $|A(z)| \leq r$ for every z. There is a decision tree of depth at most $8r^2(r+1)k^4 \leq 16r^3k^4$, independent of m and n, such that on every leaf at most r of the g_l are not identically zero.*

Proof. Let $N(S)$ be the number of l with $g_l|_S \not\equiv 0$; stop if $N(S) \leq r$. Otherwise let $r' = \max_{z \in S} |A(z)| \in [1, r]$, choose z^* maximizing $\prod_{a \in A(z)} |g_a(z)|$ over $z \in S$ with $|A(z)| = r'$, and put $A^* = A(z^*)$, $\pi = \prod_{a \in A^*} g_a$. Let $D_H = \max_{l \notin A^*} \deg g_l|_S$, let $M_1, \dots, M_b$ be a maximal family of pairwise disjoint sets of size D_H such that each M_i is a monomial of some $g_l|_S$ with $l \notin A^*$, query $U = M_1 \cup \dots \cup M_b$, and recurse on every child. Write z^{*E} for z^* with the coordinates in E flipped.

(C0) As $N(S) > r \geq |A^*|$, some $l \notin A^*$ has $g_l|_S \not\equiv 0$; it vanishes at z^*, so it is not constant and $D_H \geq 1$. (C1) If $\pi(z) \neq 0$ then $A^* \subseteq A(z)$ and $|A(z)| \leq r'$, so $A(z) = A^*$ and z was admissible; hence $\pi(z^*) \neq 0$ and $|\pi| \leq |\pi(z^*)|$ on S. (C2) For each i there is a nonempty $E_i \subseteq M_i$ with $\pi(z^{*E_i}) = 0$: take $l \notin A^*$ having M_i as a monomial and fix the coordinates outside M_i as in z^*; the resulting polynomial on M_i has nonzero top coefficient, since no monomial of $g_l|_S$ strictly contains M_i, so it is nonzero at some z^{*E_i}, and $E_i \neq \emptyset$ because $g_l(z^*) = 0$. At z^{*E_i} the index $l \notin A^*$ is active and at most r' indices are active, so some $a \in A^*$ is not. (C3) The function $\rho(w) = \pi(z^{*E(w)})/\pi(z^*)$, $E(w) = \bigcup_{i : w_i = -1} E_i$, is multilinear on Q_b of degree at most $r'k$ (substitute $z_j = z_j^* w_i$ for $j \in E_i$), $|\rho| \leq 1$ by (C1), $\rho(\mathbf{1}) = 1$, and $\rho(\mathbf{1}^{\oplus i}) = 0$ by (C2). So $|d_i\rho(\mathbf{1})| = \frac{1}{2}$ for every i, and Lemma 2.2 gives $b/2 \leq 4(r'k)^2$. (C4) Hence $|U| \leq bD_H \leq 8r^2k^3$. For every answer on U, each $l \notin A^*$ with $\deg g_l|_S = D_H$ drops in degree (each of its top monomials meets U, by maximality), and no degree increases. (C5) Let $\Phi(S)$ be the sum of the $r+1$ largest entries of the multiset $\{\deg g_l|_S\}_l$. Then Φ drops by at least 1 from S to each child. Otherwise a set T' of $r+1$ indices attaining Φ at the child attains $\Phi(S)$ and avoids $L = \{l \notin A^* : \deg g_l|_S = D_H\} \neq \emptyset$; since $|T'| > |A^*|$, some $l_1 \in T' \setminus A^*$ has degree less than D_H, and exchanging l_1 for an element of L would increase the sum. (C6) If $N(S) > r$ then $\Phi(S) \geq 1$, because nonzero constants are active on all of S, so there are at most r' of them. As $\Phi \leq (r+1)k$ at the root, every path has at most $(r+1)k$ stages of at most $8r^2k^3$ queries each. □

For $r = 1$ the potential is $D_1 + D_2 \geq 2$ and $\pi = g_{j^*}$ has degree k, so the depth in Theorem F.1 improves to $(2k-1)8k^3 < 16k^4$.

Corollary F.4 ([*proved here*])**.** (a) *If $F = \sum_j y_j g_j$ is a bounded one-block form with $\deg g_j \leq k$ and at most r of the g_j nonzero at every point, then $\max(\varepsilon, \tau) \geq V^2/(16rt^2) \geq V^2/(2^{12}r^7k^8)$, where $t \leq 16r^3k^4$ is the depth in Theorem F.3.* (b) (Sparse sensitivity.) *Let $f \colon Q_n \to \mathbb{R}$ be multilinear of degree d with $s(f) = \max_z \#\{i : f(z) \neq f(z^{\oplus i})\} \leq r$. Then f is computed exactly by a decision tree of depth $D(f) \leq 8r^2(r+1)(d-1)^4 + r \leq 17r^3d^4$. If moreover $|f| \leq 1$, then $\mathrm{MaxInf}(f) \geq (\mathrm{Var} f/D(f))^2 \geq \mathrm{Var}(f)^2/(289r^6d^8)$.*

Proof. (a) Each leaf has at most r nonzero c_{jl}, so $V_l^2 \leq r\sum_j c_{jl}^4$ and $\kappa_D \geq \sum_l p_l V_l^2/r \geq V^2/r$. By Theorem 4.2, $\kappa_D - \varepsilon V \leq 8t^2\tau$; if $\varepsilon \leq V/(2r)$ this gives $\tau \geq V^2/(16rt^2)$, and otherwise $\varepsilon > V/(2r) \geq V^2/(16rt^2)$ (if $t = 0$, at most r selectors are nonzero and $\varepsilon \geq V/r$). (b) Apply Theorem F.3 to $g_i = d_i f$, of degree at most $d-1$, with $d_i f(z) \neq 0$ if and only if i is sensitive at z. On a leaf, $(d_i f)|_{\mathrm{leaf}} = d_i(f|_{\mathrm{leaf}})$ for free i, and at most r of these are nonzero, so $f|_{\mathrm{leaf}}$

depends on at most r free coordinates; query them. For $|f| \le 1$, Theorem 4.1(L1) with $h = f$ gives $\mathrm{Var} f \le \sum_i \delta_i \mathbb{E}|d_i f| \le D(f)\mathrm{MaxInf}(f)^{1/2}$. □

Exact zeros are essential: $f = n^{-1}\sum_i x_i$ has degree 1 and $s(f) = n$, and the level-set analogue fails for Majority. For Boolean f, part (b) is weaker than what follows from Huang's sensitivity theorem [25] and [40]; its content is for real-valued f.

Appendix G. A multi-scale transfer and a search form

This appendix contains the one-scale restriction transfer, the spike gadget and the search form mentioned in Section 7. By Theorem G.4 the transfer is equivalent to the conjecture, so the conditional bound of Theorem G.2 is a reformulation, not a weakening. The iterated-restriction algorithm (query the most influential variable, restrict, repeat) is that of [2], and restriction lemmas of a related kind, at rate $\log d/(Cd)$, are [12, Lemmas 5.3–5.4, Theorems 6.3–6.4]; the multi-scale statements below are reformulations of the conjecture in this spirit. Let σ be a random restriction in which each coordinate is alive with probability $\frac{1}{2}$ and fixed to ± 1 with probability $\frac{1}{4}$ each. The *one-scale transfer* $\mathrm{OS}(C, a, L)$ is the statement: for all $d \ge 2$, $g \in \mathcal{B}_d$ and $\delta, q \in (0, 1]$, if $\Pr_\sigma[\mathrm{Var}(g_\sigma) > 0,\ \mathrm{MaxInf}(g_\sigma) \ge \delta \mathrm{Var}(g_\sigma)^C] \ge q$ then $\mathrm{MaxInf}(g) \ge \delta q^a \mathrm{Var}(g)^C / L(d)$.

Lemma G.1 (base case; [*proved here*])**.** *Let $f \in \mathcal{B}_d$, $v = \mathrm{Var} f > 0$, $p \le 1/(2ed)$ and $k = \lceil \log_2(8/(p^2 v)) \rceil$. With probability at least $pv/4$ over a rate-p restriction ρ, $\mathrm{MaxInf}(f_\rho) \ge (pv/(16 d^2 3^k))^2$.*

Proof. $\mathbb{E}\mathrm{Var}(f_\rho) = \sum_S (1 - (1-p)^{|S|}) \widehat{f}(S)^2 \ge pv$ and $\mathrm{Var} f_\rho \le 1$, so $\Pr[\mathrm{Var} f_\rho \ge pv/2] \ge pv/2$. Also $\mathbb{E} W^{>k+1}(f_\rho) = \sum_{S \ne \emptyset} \widehat{f}(S)^2 \Pr[\mathrm{Bin}(|S|, p) > k+1] \le v(edp/(k+2))^{k+2} \le v 2^{-(k+2)}$, so by Markov $\Pr[W^{>k+1}(f_\rho) > pv/4] \le 2^{-k}/p \le pv/8$. On the intersection $W^{[1,k+1]}(f_\rho) \ge pv/4$, and Corollary 3.3 applies to $f_\rho \in \mathcal{B}_d$. □

Theorem G.2 ([*proved here*]; a reformulation by Theorem G.4)**.** *If $\mathrm{OS}(C, a, L)$ holds with $C \ge 2$, then for $f \in \mathcal{B}_d$, $d \ge 2$, $v = \mathrm{Var} f$, $T = \lceil \log_2(2ed) \rceil$, $p = 2^{-T}$, k as in Lemma G.1 and $\delta_T = (pv/(16 d^2 3^k))^2$,*

$$\mathrm{MaxInf}(f) \ \ge\ v^C\, \delta_T\, (pv/4)^{aT}\, 2^{-aT(T+1)/2}\, L(d)^{-T}.$$

For $L(d) = d^b$ this is $\exp(-K(1+a+b+C)\log(2d)\log(2d/v))$ with K absolute, i.e. $(v/d)^{O(\log d)}$; for $L(d) = e^{o(d/\log d)}$ it is $e^{-o(d)}$ whenever $\log(1/v) = o(d/\log d)$.

Proof. Let ρ_t be the composition of t independent rate-$\frac{1}{2}$ restrictions (a rate-2^{-t} restriction) and $E_\delta(g) = \{\mathrm{Var} g > 0,\ \mathrm{MaxInf}(g) \ge \delta \mathrm{Var}(g)^C\}$. By Lemma G.1, $\Pr[E_{\delta_T}(f_{\rho_T})] \ge q_T := pv/4$. If $\Pr[E_{\delta_t}(f_{\rho_t})] \ge q_t$, then with probability at least $q_t/2$ over ρ_{t-1} we have $\Pr_\sigma[E_{\delta_t}((f_{\rho_{t-1}})_\sigma)] \ge q_t/2$, and OS for $f_{\rho_{t-1}} \in \mathcal{B}_d$ gives $E_{\delta_{t-1}}(f_{\rho_{t-1}})$ with $\delta_{t-1} = \delta_t (q_t/2)^a / L(d) \le 1$ and $q_{t-1} = q_t/2$. At $t = 0$ this says $\mathrm{MaxInf}(f) \ge \delta_0 v^C$, and $\prod_{t=1}^T q_T 2^{t-T-1} = q_T^T 2^{-T(T+1)/2}$. □

Proposition G.3 ([*proved here*])**.** (i) *Conjecture 1.1 with constants (c_{AA}, C_{AA}) implies $\mathrm{OS}(C, 0, d^{C_{AA}}/c_{AA})$ for every $C \ge C_{AA}$.* (ii) *For $C \ge 2$, OS holds with $L = e^{O(d)}$ (Proposition 2.5).* (iii) *For $C \ge 2$ and $L(d) \ge (24d^3)^2$, OS holds automatically for every g with $W^{[1,k+1]}(g) \ge \mathrm{Var}(g)/2$, $k = \lceil \log_3 d \rceil$.* (iv) *For $C < 2$, $\mathrm{OS}(C, a, L)$ is false for every $L = 2^{o(d)}$; and without the restriction $\delta \le 1$ it is false for every $L = 2^{o(d)}$ and every C.*

Proof. (i) $\mathrm{MaxInf}(g) \ge c_{AA}(\mathrm{Var} g/d)^{C_{AA}} \ge c_{AA} d^{-C_{AA}} \mathrm{Var}(g)^C \ge \delta q^0 \mathrm{Var}(g)^C / L$. (ii) Proposition 2.5. (iii) Corollary 3.3 gives $\mathrm{MaxInf}(g) \ge \mathrm{Var}(g)^2/(8d^2 3^k)^2 \ge \mathrm{Var}(g)^C/(24d^3)^2$. (iv) Let $g_m = M^{-1}\sum_{c=1}^M \mathrm{AND}_m(x^{(c)})$ on disjoint blocks, $\mathrm{AND}_m = \prod_{l \le m}(1 + x_l)/2$, with $M \ge \lceil 1/\pi_1 \rceil$,

$\pi_1 = 2m4^{-m}$. With probability at least $1 - 1/e$ some copy keeps exactly one alive variable with all its fixed variables equal to $+1$, so $\mathrm{MaxInf}((g_m)_\sigma) \geq 1/(4M^2) = 2^{m-1}\mathrm{MaxInf}(g_m)$; and $\mathrm{Var}((g_m)_\sigma) \leq 4\mathrm{Var}(g_m)$ with probability at least $\frac{3}{4}$. Hence with probability at least 0.38, $D_C((g_m)_\sigma) \geq 2^{m-1-2C} D_C(g_m)$, where $D_C(g) = \mathrm{MaxInf}(g)/\mathrm{Var}(g)^C$. For $C < 2$, $D_C(g_m) \leq 2^{(C-1)(m+1)} M^{C-2} \to 0$ as $M \to \infty$, so the hypothesis holds with $\delta = 2^{m-1-2C} D_C(g_m) < 1$ and $q = 0.38$, and OS would force $L(m) \geq 2^{m-1-2C} 0.38^a$. Without the cap the same computation applies for every C. □

By (iii), only top-heavy g matter, and one might hope that OS for them is easier than the conjecture. It is not: gluing a spike average onto a variance donor on fresh variables makes the hypothesis of OS hold at full strength, while the spikes contribute almost nothing to MaxInf, so the conclusion must be paid by the donor.

Theorem G.4 (spike gadget; [*proved here*])**.** *Fix $C \geq 2$, $a \geq 0$ and a nondecreasing $L \geq 1$; let $q = 1 - 1/e - 1/4 > 0.38$ and $\kappa(D) = q^a/(4^{C+1}L(D))$. Assume $\mathrm{OS}(C, a, L)$ only for top-heavy g, i.e. for $g \in \mathcal{B}_D$ with $W^{[1,k+1]}(g) < \mathrm{Var}(g)/2$, $k = \lceil \log_3 D \rceil$. Let $h \in \mathcal{B}_d$ with $v = \mathrm{Var}h > 0$, and let $D \geq d$ satisfy $m := \lceil \log_2(1/\kappa(D)) \rceil + 1 \leq D$. Then, with $k = \lceil \log_3 D \rceil$,*

$$\mathrm{MaxInf}(h) \;\geq\; \min\left\{ \left(\frac{v}{16 d^2 3^k} \right)^2,\; \frac{\kappa(D)^5 v^C}{256} \right\}.$$

Consequently OS*, even restricted to top-heavy functions, is equivalent to Conjecture 1.1 up to polynomial losses:* $\mathrm{OS}(C, a, Kd^b)$ *gives* $\mathrm{MaxInf}(h) \geq c\,\mathrm{Var}(h)^C d^{-\max(6,5b)}$ *with $c = c(C, a, K, b) > 0$, and* OS *with a loss L satisfying $L(d) \leq q^a 2^{d-2C-4}$ gives $\mathrm{MaxInf}(h)/\mathrm{Var}(h)^C \geq \min\{d^{-6}/2304,\; q^{5a}4^{-5C-9}L(d)^{-5}\}$, which is $\mathrm{poly}(1/(dL(d)))$; for larger L Proposition 2.5 gives a bound of the same form.*

Proof. Case 1: $W^{[1,k+1]}(h) \geq v/4$. Corollary 3.3 gives the first term.

Case 2: $W^{[1,k+1]}(h) < v/4$. Let $s = M^{-1} \sum_{c=1}^{M} \mathrm{AND}_m(x^{(c)})$ on M fresh disjoint blocks of m variables, $\mathrm{AND}_m = \prod_{l \leq m}(1 + x_l)/2$, with $\pi_1 = 2m4^{-m}$ and $M = \max(\lceil 1/\pi_1 \rceil, \lceil 1/(4v^{C/2}) \rceil)$, and let $g = (h + s)/2 \in \mathcal{B}_D$. Since $\kappa \leq 4^{-3}$ we have $m \geq 7$.

(i) *g is top-heavy.* $\mathrm{Var}s \leq 2^{-m}/M \leq 4 \cdot 2^{-m} v^{C/2} < v/2$, since $C \geq 2$ and $v \leq 1$. As h and s depend on disjoint variables, $W^{[1,k+1]}(g) \leq (v/4 + \mathrm{Var}s)/4 < (v + \mathrm{Var}s)/8 = \mathrm{Var}(g)/2$.

(ii) *The hypothesis of* OS *holds for g* with $\delta = 1/(16M^2(4\mathrm{Var}g)^C)$, which is at most 1 because $4\mathrm{Var}g \geq v$ and $16M^2 v^C \geq 1$. A given copy of AND_m keeps exactly one alive variable with all its fixed variables equal to $+1$ with probability π_1, so some copy does with probability at least $1 - (1 - \pi_1)^M \geq 1 - 1/e$; that variable has influence $(1/(4M))^2 = 1/(16M^2)$ in g_σ. Since $\mathbb{E}_\sigma \mathrm{Var}(g_\sigma) \leq \mathrm{Var}g$, Markov's inequality gives $\mathrm{Var}(g_\sigma) \leq 4\mathrm{Var}g$ with probability at least $\frac{3}{4}$. With probability at least q both hold, and then $\mathrm{Var}(g_\sigma) > 0$ and $\mathrm{MaxInf}(g_\sigma) \geq 1/(16M^2) = \delta(4\mathrm{Var}g)^C \geq \delta \mathrm{Var}(g_\sigma)^C$.

(iii) OS for g gives $\mathrm{MaxInf}(g) \geq \delta q^a \mathrm{Var}(g)^C/L(D) = \kappa(D)/(4M^2)$. By disjointness $\mathrm{MaxInf}(g) = \frac{1}{4}\max(\mathrm{MaxInf}h, 2^{-m-1}/M^2)$, and $2^{-m-1} \leq \kappa/4$ by the choice of m; hence $\mathrm{MaxInf}(h) \geq \kappa/M^2$.

(iv) $M \leq \max(2/\pi_1, 1/(2v^{C/2}))$, so $M^2 \leq 4/(\pi_1^2 v^C)$; and $2^m \leq 4/\kappa$ gives $\pi_1 \geq m\kappa^2/8 \geq \kappa^2/8$. So $\mathrm{MaxInf}(h) \geq \kappa \pi_1^2 v^C/4 \geq \kappa^5 v^C/256$.

For the consequences take $D = \max(d, d_0)$ with $d_0 = d_0(C, a, K, b)$ such that $m \leq D$, and use $3^k \leq 3D$ and $(v/(48d^3))^2 \geq v^C/(48^2 d^6)$; the converse implication is Proposition G.3(i). □

So Theorem G.2 is a reformulation of the conjecture, not a weakening, and Theorem G.4 gives the polynomial bound directly. The same gadget defeats every variant in which a statement about $(\mathrm{MaxInf}, \mathrm{Var})$ at scale $\frac{1}{2}$ implies one at scale 1 under a cap $\delta \leq \mathrm{poly}(\mathrm{Var}/d)$, with any

probability $q<1$ (take $M\gg 1/\pi_1$ and Markov at level $1/(1-q)$), as well as averaged versions and low-fraction mass versions: if the conclusion contains $\mathrm{MaxInf}(g)\ge\theta/L$ and a negligible spike part can satisfy the hypothesis, the statement is equivalent to the conjecture.

The search form below says that the transfer problem is a query-complexity problem. Fix $A\subseteq[n]$, let $y\in\{\pm1\}^{A^c}$ be uniform, $r\in(0,1]$, and $\varphi_j(y)=\mathrm{Inf}_j^{(r)}(f_{A,y})$ for $j\in A$ (noisy influence $\|T_{\sqrt{r}}d_j\cdot\|_2^2$ of the restriction). $\mathrm{JA}(Q,\varepsilon)$ at (A,θ,r) is the statement: a randomized decision tree on y with expected number of queries at most Q outputs $\tilde\varphi(y)\in[0,1]^A$ with $\mathbb{E}\sum_j|\tilde\varphi_j-\varphi_j(y)|\mathbf{1}[\max(\tilde\varphi_j,\varphi_j(y))\ge\theta]\le\varepsilon\theta$.

Theorem G.5 (search form; [*proof outline*])**.** *The following are equivalent, with polynomial changes of constants:* (a) *Conjecture 1.1;* (b) *there is K such that for all $f\in\mathcal{B}_d$, A, $\theta,\varepsilon\in(0,1]$ and $k\ge1$, $\mathrm{JA}(Q,\varepsilon)$ holds at $(A,\theta,1-1/k)$ with $Q\le(dk/(\theta\varepsilon))^K$.*

Proof outline. (b)⇒(a). (1) Since $\sum_j|\varphi_j(y)-\varphi_j(y^{\oplus l})|\le 4d\sqrt{\psi_l(y_{-l})}$ with $\mathbb{E}\psi_l=\mathrm{Inf}_l(f)$, for θ' uniform on $[\theta,2\theta]$ the junta map $J_{\theta'}(y)=\{j:\varphi_j(y)\ge\theta'\}$ satisfies $\mathbb{E}_{\theta'}\Pr_y[J_{\theta'}(y)\ne J_{\theta'}(y^{\oplus l})]\le 4d\sqrt{\mathrm{Inf}_l f}/\theta$. (2) A set-valued OSSS inequality in covariance form, proved by the hybrid argument of Theorem 4.1 using only the triangle inequality of the discrete metric: for a tree T computing G and any J, $\Pr[G=J]-\sum_S\Pr[G=S]\Pr[J=S]\le\sum_l\delta_l(T)\Pr_z[J(z)\ne J(z^{(l)})]$, where $z^{(l)}$ re-randomizes coordinate l. Applied to $G_{\theta'}=\{j:\tilde\varphi_j\ge\theta'\}$ and $J_{\theta'}$, with $q=\Pr_y[J_{2\theta}\ne\emptyset]$ and $\tau=\mathrm{MaxInf}(f)$, it gives $q-\varepsilon-\tau/\theta\le 2dQ\sqrt{\tau}/\theta$, so $\tau\ge\min(\theta q/4,(\theta q/(4dQ))^2)$ when $\varepsilon\le q/4$. (3) Lemma G.1 at $p=2^{-\lceil\log_2(2ed)\rceil}$ provides A with $q\ge pv/4$ for $\theta=\mathrm{poly}(v/d)$ (noisy influences at $r=1-1/(k'+2)$ are at least e^{-1} times the low-degree influences). *(a)⇒(b).* Under the conjecture the greedy algorithm of [2] (query the most influential variable while the restricted variance is at least ε_1) gives a tree T of depth $D=d^{C+1}/(c\varepsilon_1^{C+1})$ with $\|f-T\|_2^2\le 2\varepsilon_1$. For noisy influences with $r=1-1/k$, $\sum_j|\varphi_j-\varphi_j^T|\le 2k\|f_{A,y}-T_{A,y}\|_2$ (the seminorm $\|T_{\sqrt{r}}d_j\cdot\|$ and $\max_s s\,r^{s-1}\le k$), and each φ_j^T is estimated from $R=(6eD/(\theta^2\varepsilon))^2$ samples by following at most $2D(D+1)$ paths; a Chernoff bound for the small $\iota_j=\mathrm{Inf}_j(T_{A,y})$ and the choice $\varepsilon_1=(\varepsilon\theta^2/(24k))^2$ give total error at most $\varepsilon\theta$ with $Q\le 2D(D+1)R$. □

Appendix H. Refuted routes

This appendix records two routes that fail. The deficit transfer, a route to a quasi-polynomial bound measured in influence mass, is false in the regime that the argument needs. The L^∞ Talagrand inequality of [35, Conjecture 37], which would imply Conjecture 1.1, is false, also in symmetric form (Appendix H.2).

H.1. The deficit transfer.

For a function u and $\theta>0$ let $S_\theta(u)=\sum_j\mathrm{Inf}_j(u)\mathbf{1}[\mathrm{Inf}_j(u)\ge\theta]$ be the spike mass and $N_\theta(u)=I(u)-S_\theta(u)$ the non-spike mass; recall that $2\mathbb{E}_\sigma I(g_\sigma)=I(g)$ for the rate-$\frac12$ restriction σ.

Definition H.1 (deficit transfer)**.** $\mathrm{DT}(K,\varepsilon,L)$ (with cap C) is the statement: for all $g\in\mathcal{B}_d$ and $0<\theta\le\mathrm{Var}(g)^C$,

$$N_{\theta/L}(g)\ \le\ 2K\,\mathbb{E}_\sigma N_\theta(g_\sigma)+\varepsilon\,I(g).$$

DT says that non-spike mass at scale 1 is controlled by non-spike mass at scale $\frac12$. Unlike OS it is linear in g, and the spike gadget does not hit it: for $g=(h+s)/2$ with h spike-free the right side is about $K\,I(g)$, which is at least the left side.

Remark H.2 (consequences of DT; [*proof outline*]). Since DT is linear in g, it can be averaged along the chain of scales of Theorem G.2. With $\varphi_t=1-\mathbb{E}S_{\theta_t}(f_{\rho_t})/\mathbb{E}I(f_{\rho_t})$ and $\theta_{t-1}=\theta_t/L$ one gets $\varphi_{t-1}\le K\varphi_t+\varepsilon+e_t$, where $e_t\le d\,\theta_t^{1/C}2^t/I(f)$ accounts for restrictions with $\mathrm{Var}^C<\theta_t$, and

at the bottom scale $T = O(\log d)$ a base-case estimate in the spirit of Lemma G.1 (not included here) gives $\varphi_T \le 1/(4K^T)$ at a threshold $\theta_T = \mathrm{poly}(v/d)$. Thus $\mathrm{DT}(O(1), 1/\mathrm{poly}(d), \mathrm{poly}(d))$ would give $\varphi_0 \le \frac{3}{4}$ and $\mathrm{MaxInf}(f) \ge \theta_T L^{-T} = (\mathrm{Var} f/d)^{O(\log d)}$, and a loss $L = e^{o(d/\log d)}$ would give $e^{-o(d)}$ when $\log(1/\mathrm{Var} f) = o(d/\log d)$.

The next proposition shows that DT fails in exactly this regime, already for Boolean functions that satisfy the conjecture.

Proposition H.3 (parity-target address functions; [*proved here*])**.** *Let $k \ge 4$, let $y \in Q_k$ be address bits, let B_a ($a \in Q_k$) be 2^k disjoint blocks of w target bits, and let $F(y,x) = \chi_{B_y}(x) = \prod_{i \in B_y} x_i$. Then F is Boolean of degree $k+w$, $\mathrm{Var} F = 1$ and $\mathrm{MaxInf}(F) = \frac{1}{2}$. For $w = k^2$, $1 \le K \le 2^{k^2-k-1}$, $\varepsilon < \frac{1}{2} - \frac{1}{k}$, $r_\theta = \lceil k/2 + (k \ln(2K)/2)^{1/2} \rceil$, $\theta = 2^{-r_\theta}$ and $1 \le L < 2^{k-r_\theta}$, the inequality of Definition H.1 fails for $g = F$ (with any cap C). In particular, in degree $d = k^2 + k$, $\mathrm{DT}(K, \varepsilon, L)$ is false for every constant K, every $\varepsilon < \frac{1}{2} - o(1)$ and every $L \le 2^{c_K \sqrt{d}}$. Since DT quantifies over all $g \in \mathcal{B}_d$, one may also take F with $k = \Theta(\log^2 d)$ address bits (degree $k^2 + k \le d$); then $\theta = 2^{-r_\theta} = d^{-\Theta(\log d)}$ lies in the range of thresholds used in Remark H.2, and $L < 2^{k-r_\theta} = d^{\Theta(\log d)}$ covers every polynomial loss.*

Proof. Each address bit has influence $\frac{1}{2}$ (flipping it switches between two independent uniform parities), and each target bit has influence 2^{-k}, so $I(F) = w + k/2$. As $\theta/L > 2^{-k}$, $N_{\theta/L}(F) \ge w$. Under σ let r be the number of alive address bits. An alive target bit in B_a has influence 2^{-r} in F_σ if the fixed address bits agree with a and 0 otherwise; it is a non-spike only if $r > r_\theta$. Hence the target bits contribute

$$2^k \cdot w \cdot \tfrac{1}{2} \cdot \mathbb{E}\big[2^{-r} 2^{-(k-r)} \mathbf{1}[r > r_\theta]\big] = \tfrac{w}{2} P', \qquad P' = \Pr[\mathrm{Bin}(k, \tfrac{1}{2}) > r_\theta] \le \frac{1}{2K}$$

to $\mathbb{E}_\sigma N_\theta(F_\sigma)$, by Hoeffding's inequality. An alive address bit has influence $\frac{1}{2} \ge \theta$ in F_σ unless some block is entirely fixed, which has probability at most $2^k 2^{-w}$; so the address bits contribute at most $\frac{k}{2} 2^{k-w}$. Therefore

$$2K\, \mathbb{E}_\sigma N_\theta(F_\sigma) + \varepsilon I(F) \le \frac{w}{2} + Kk 2^{k-w} + \Big(\frac{1}{2} - \frac{1}{k}\Big)\Big(w + \frac{k}{2}\Big) < w \le N_{\theta/L}(F),$$

using $w = k^2$, $(\frac{1}{2} - \frac{1}{k})(k^2 + \frac{k}{2}) < k^2/2 - 3k/4$ and $Kk 2^{k-k^2} \le k/2$. Finally $\theta \le 1 = \mathrm{Var}(F)^C$, and $k - r_\theta \ge k/2 - (k \ln(2K)/2)^{1/2} - 1$ with $k \ge \sqrt{d} - 1$. □

The chain of Remark H.2 concludes that $S_{\theta_0}(f) \ge I(f)/4$ with $\theta_0 = d^{-O(\log d)}$; for F this is false, since $S_\theta(F) = k/2$ for $2^{-k} < \theta \le \frac{1}{2}$ while $I(F) = k^2 + k/2$. So no transfer whose output is that a constant fraction of the influence mass lies on quasi-polynomially large spikes can hold, and influence mass is the wrong currency. In particular DT is false although F satisfies the conjecture. A transfer measured in variance is not refuted by these examples; the current candidate is the touching transfer of Section 8, whose limits are in Appendix I.

H.2. **Talagrand's inequality for L^∞.** For $x \in \{0,1\}^n$ and $K \subseteq [0,1]^n$ Lovett and Zhang [35, Section 4] put $d_\infty(x, K) = \min_{y \in K} \|x - y\|_\infty$, with $\mathrm{conv}(Y)$ the convex hull of Y in $[0,1]^n$, and [35, Definition 36]: "Let $X \subset \{0,1\}^n$. The i-th influence of X is the probability that a random element in X moves outside X when the i-th bit is flipped: $\mathrm{Inf}_i[X] = \Pr_{x \in X}[x \oplus e_i \notin X]$. The maximal influence of X is $\mathrm{Inf}_\infty[X] = \max_i \mathrm{Inf}_i[X]$" (the source has $\mathrm{Inf}_\infty[x]$, a typo). Their conjecture reads:

> **Conjecture 37** (Talagrand for L^∞)**.** Let $X, Y \subset \{0,1\}^n$. Assume that $\mathrm{Inf}_\infty[X], \mathrm{Inf}_\infty[Y] \le \tau$. Then there exists $x \in X$ such that $d_\infty(x, \mathrm{conv}(Y)) \le \mathrm{poly}(\tau)$.

By [35, Claim 38] it implies Conjecture 1.1. The statement has no condition on the sizes of X and Y or on n, and we read $\mathrm{poly}(\tau)$ as $C\tau^c$ with absolute constants $C, c > 0$, which is what Claim 38 needs. Both examples below are built on Tribes: k disjoint blocks (tribes) $T_1, \dots, T_k$ of width $w \geq 2$, $n = kw$ and $p = 2^{-w}$. A tribe T is *full* at x if $x_T = 1^w$ and *empty* if $x_T = 0^w$, and $T(i)$ is the tribe containing i.

Proposition H.4 (Conjecture 37 of [35] is false; [*proved here*])**.** *Let $k = \lceil 2^w \ln 2 \rceil$.* (a) *Let X be the set of x with a full tribe and $Y = \{0,1\}^n \setminus X$. Then $d_\infty(x, \operatorname{conv} Y) \geq 1/w$ for every $x \in X$, $\mathrm{Inf}_\infty[Y] = p/(1-p)$, $\mathrm{Inf}_\infty[X] = p(1-p)^{k-1}/(1-(1-p)^k)$, and $\mu(X), \mu(Y) \to \frac{1}{2}$ as $w \to \infty$.* (b) *Let X' be the set of x with a full tribe and no empty tribe, and $Y' = \{1 - x : x \in X'\}$ the set of x with an empty tribe and no full tribe. Then $d_\infty(x, \operatorname{conv} Y') \geq 1/w$ for every $x \in X'$ and $d_\infty(y, \operatorname{conv} X') \geq 1/w$ for every $y \in Y'$, $\mathrm{Inf}_\infty[X'] = \mathrm{Inf}_\infty[Y'] = p(1-p)^{k-1}/((1-p)^k - (1-2p)^k)$, and $\mu(X') = \mu(Y') \to \frac{1}{4}$.*

In (a), $\tau = \max(\mathrm{Inf}_\infty[X], \mathrm{Inf}_\infty[Y]) = (1 + o(1))2^{-w}$, and in (b) $\tau = (2 + o(1))2^{-w}$. So $1/w = (1 + o(1))/\log_2(1/\tau)$, which exceeds $C\tau^c$ for all fixed $C, c > 0$ once w is large. Hence Conjecture 37 of [35] is false, and by (b) so is its symmetric form, in which the conclusion is that some point of X is close to $\operatorname{conv} Y$ or some point of Y is close to $\operatorname{conv} X$.

Proof. (a) Let $x \in X$ and let T be full at x. Every $y \in Y$ has a 0 in T, so $\sum_{i \in T}(1 - y_i) \geq 1$ on Y and hence on $\operatorname{conv} Y$. For $z \in \operatorname{conv} Y$ this gives $\sum_{i \in T}(x_i - z_i) \geq 1$, so $\|x - z\|_\infty \geq 1/w$. Under the uniform measure on Y the tribes are independent and uniform on $\{0,1\}^w \setminus \{1^w\}$, and $y \oplus e_i \notin Y$ if and only if $y_{T(i)} = 1^w \oplus e_i$, so $\mathrm{Inf}_i[Y] = 1/(2^w - 1)$. Further, $x \oplus e_i \notin X$ for $x \in X$ if and only if $T(i)$ is the only full tribe of x, an event of probability $p(1-p)^{k-1}$ under the uniform measure on $\{0,1\}^n$, while $\mu(X) = 1 - (1-p)^k$. As $w \to \infty$, $(1-p)^k \to \frac{1}{2}$, which gives the remaining claims. (b) As in (a), every $x \in X'$ has a full tribe T and every $y \in Y'$ has a 0 in T, so $d_\infty(x, \operatorname{conv} Y') \geq 1/w$. The map $x \mapsto 1 - x$ exchanges X' and Y' and preserves influences and d_∞, which gives the claims for Y'. Flipping i moves $x \in X'$ out of X' exactly when $x_i = 1$ and either $T(i)$ is the only full tribe, which has probability $p(1-2p)^{k-1}$ under the uniform measure, or $x_{T(i)} = e_i$ while another tribe is full and none is empty, which has probability $p((1-p)^{k-1} - (1-2p)^{k-1})$. The sum is $p(1-p)^{k-1}$, and $\mu(X') = (1-p)^k - (1-2p)^k \to \frac{1}{2} - \frac{1}{4}$. □

We have found no erratum or later discussion of [35, Conjecture 37].

Remark H.5 (readings that survive; [*proved here*] for (ii), [*open*] for (iii) and (iv))*.* (i) Adding dummy variables changes neither the influences nor the distances, since $\operatorname{conv}(Y \times \{0,1\}^m) = \operatorname{conv}(Y) \times [0,1]^m$; so neither a factor $\mathrm{polylog}(n)$ nor an additive term $(\log n)^{-c}$ in the bound rescues the statement. In (a) the order of X and Y matters: averaging the k points obtained from $y \in Y$ by filling one tribe gives $d_\infty(y, \operatorname{conv} X) \leq 1/k = O(\tau)$. (ii) *Complementary sets.* If $Y = \{0,1\}^n \setminus X$ with $X, Y \neq \emptyset$, and $d_\infty(x, \operatorname{conv} Y) \geq \eta$ for all $x \in X$ and $d_\infty(y, \operatorname{conv} X) \geq \eta$ for all $y \in Y$, then $\max(\mathrm{Inf}_\infty[X], \mathrm{Inf}_\infty[Y]) \geq \eta^3$. So the symmetric form holds for complementary sets with $\mathrm{poly}(\tau) = \tau^{1/3}$, while (a) shows that the one-sided form fails even for them. *Proof.* Let $h = \mathbf{1}_X$. If $B_1, \dots, B_b$ are disjoint blocks with $x^{B_j} \in Y$ for some $x \in X$, the average of the x^{B_j} lies in $\operatorname{conv} Y$ at distance $1/b$ from x, so $b \leq 1/\eta$; the same holds at points of Y. So $\mathrm{bs}(h) \leq 1/\eta$, and $D(h) \leq \eta^{-3}$ by Lemma 5.7. For $F = 2h - 1$, Theorem 4.1(L1) applied to an optimal tree for F gives $\mathrm{Var} F \leq D(h) \max_i \mathrm{Inf}_i(F)$, where $\mathrm{Var} F = 4\mu(X)\mu(Y)$ and $\mathrm{Inf}_i(F) = 2\mu(X)\mathrm{Inf}_i[X] = 2\mu(Y)\mathrm{Inf}_i[Y]$. Hence $\mathrm{Inf}_\infty[X] \geq 2\mu(Y)/D(h)$ and $\mathrm{Inf}_\infty[Y] \geq 2\mu(X)/D(h)$, and $\max(\mu(X), \mu(Y)) \geq \frac{1}{2}$. (iii) *The KKL scale.* The bound $d_\infty(x, \operatorname{conv} Y) \leq C/\log(1/\tau)$ for some $x \in X$ is not refuted, and Tribes shows that it would be sharp up to the constant. Through [35, Claim 38] it would give only $\tau^{-1} \leq \exp(\mathrm{poly}(d, 1/\varepsilon, \log n))$, which does not imply the conjecture. (iv) *Level sets of one polynomial.* [35, Claim 38] applies Conjecture 37 only to $X = \{f \leq t - \alpha\}$ and $Y = \{f \geq t + \alpha\}$ for one f of degree d, with $t = \mathbb{E} f$, $\alpha \in [\varepsilon, 2\varepsilon]$ and $|X|, |Y| \geq \varepsilon 2^n$; there

the distance bound $d_\infty \geq p$ holds at every point of X and of Y, so the symmetric form would suffice. The pair of (a) is of this kind, but only with a small gap [*unaudited*]: $f = 1 - 2N/k$, with N the number of full tribes, lies in $\mathcal{B}_w$, $\mathbb{E}f = 1-2p$, and $X \subseteq \{f \leq t-\alpha\}$, $Y \subseteq \{f \geq t+\alpha\}$ for $0 < \alpha \leq \min(2p, 2/k - 2p)$. Conversely, suppose $X \subseteq \{f \geq t+\alpha\}$ and $Y \subseteq \{f \leq t-\alpha\}$ (or the reverse) for some $f \in \mathcal{B}_d$. At a point $y \in Y$ at which every tribe has exactly one 0, flipping that 0 moves y into X, so $|d_i f(y)| \geq \alpha$ for k coordinates, and Lemma 2.2 gives $k\alpha \leq 4d^2$, i.e. $d^2\tau/\alpha \geq d^2 p/\alpha \geq (\ln 2)/4$; as $\alpha \leq 1$, also $d/\alpha \geq \max(d, k/(4d)) \geq \sqrt{k}/2 = \Omega(\sqrt{n/\log n})$. So Tribes does not refute a level-set version with a bound polynomial in $d^2\tau/\alpha$; such a version suffices for [35, Claim 38] and is essentially the level-set form of Conjecture 1.1. Proposition H.4 refutes the degree-free statement, not this degree-dependent route.

Appendix I. Limits of the touching transfer

This appendix contains the negative results on the touching transfer of Section 8. They show that the loss L must be at least $d^{1-o(1)}$.

Proposition I.1 (spike-free functions; [*proved here*])**.** *If $g \in \mathcal{B}_d$ and $J_{\theta/L}(g) = \emptyset$, then $2\mathbb{E}_\sigma U_\theta(g_\sigma) \geq \mathrm{Var}(g)(1 - d\,9^{d-1}/L)$. So $\mathrm{TT}(\varepsilon, L)$ holds for every such g once $L \geq d\,9^{d-1}/\varepsilon$, with no cap.*

Proof. By Lemma 8.1(b),(e), $2\mathbb{E}U_\theta(g_\sigma) \geq \mathrm{Var}g - 2\mathbb{E}S_\theta(g_\sigma)$. For an alive j the derivative of g_σ in direction j is $(d_j g)_\sigma$, so $X_j := \mathrm{Inf}_j(g_\sigma) = \mathbb{E}_{x_A}[(d_j g)_\sigma^2]$; by Jensen's inequality, uniformity of the fixed values and Bonami's inequality ($\deg d_j g \leq d-1$), $\mathbb{E}[X_j^2 \mid j \text{ alive}] \leq \mathbb{E}(d_j g)^4 \leq 9^{d-1}\mathrm{Inf}_j(g)^2$. Hence $\mathbb{E}[X_j \mathbf{1}[X_j \geq \theta] \mid j \text{ alive}] \leq 9^{d-1}\mathrm{Inf}_j(g)^2/\theta < 9^{d-1}\mathrm{Inf}_j(g)/L$, and summing, $\mathbb{E}S_\theta(g_\sigma) \leq 9^{d-1}I(g)/(2L) \leq d\,9^{d-1}\mathrm{Var}(g)/(2L)$. □

[*proof outline*] moments of order $2 + 2\delta$ with $\delta \to 0$ improve 9^{d-1} to $e^{(2+o(1))d}$ provided $\log(1/\varepsilon) = o(\delta d)$. No argument reaches $L = d^a$ with $a < 2C/(2C-1)$ for spike-free g: the families below have no θ/L-spikes and violate TT there.

Compositions of disjoint ANDs. Take s disjoint blocks of width w, let $A_i \in \{0,1\}$ indicate that all bits of block i equal 1, and put $N = \sum_i A_i$, $\lambda = s2^{-w}$, $\pi_0 = (1-2^{-w})^s$ and $\pi_0' = (1-2^{-w})^{s-1}$. For a polynomial P of degree q with $|P| \leq 1$ on $\{0, \dots, s\}$ let $g = P(N)$ and $\Delta P(t) = P(t+1) - P(t)$. As $A_i^2 = A_i$, N^j is a combination of products of the A_i of distinct blocks, so $g \in \mathcal{B}_{qw}$; every bit has influence $\iota_0 = \frac{2^{-w}}{2}\mathbb{E}\Delta P(N')^2$ with $N' \sim \mathrm{Bin}(s-1, 2^{-w})$. Under σ, block i is dead if some bit is fixed to -1 (put $\phi_i = 0$), fully fixed if all its bits are fixed to 1 ($\phi_i = 1$), and otherwise alive with $r_i \geq 1$ alive bits ($\phi_i = 2^{-r_i}$); it is alive with r alive bits with probability $\kappa_r = \binom{w}{r}2^{-r}4^{-(w-r)}$. Then $g_\sigma = P(N_\sigma)$, and an alive bit of block i has influence $\iota_i = \frac{\phi_i}{2}\mathbb{E}\Delta P(N_{-i})^2 \geq \frac{\phi_i}{2}\Delta P(0)^2\Pi^{(-i)}$, where N_{-i} counts the other blocks and $\Pi^{(-i)} = \Pr[N_{-i} = 0] = \prod_{i' \neq i}(1-\phi_{i'})$ is independent of ϕ_i, with $\mathbb{E}\Pi^{(-i)} = \pi_0'$. Let W be the set of alive blocks with $\iota_i < \theta$, N_W their count and $N_{\mathrm{rest}} = N_\sigma - N_W$. Then $U_\theta(g_\sigma) = \mathrm{Var}Q(N_W)$ with $Q(t) = \mathbb{E}P(t + N_{\mathrm{rest}})$, $|Q| \leq 1$, and $\Pi_0 := \Pr[N_\sigma = 0] = \Pi_W \Pr[N_{\mathrm{rest}} = 0]$ with $\Pi_W = \Pr[N_W = 0]$. For $\theta = (1+\xi)L\iota_0$ with $\xi > 0$ there are no θ/L-spikes, so $U_{\theta/L}(g) = \mathrm{Var}g$, and if $\theta \leq \mathrm{Var}(g)^C$ then $\varepsilon_{\min}(g; L) \geq 1 - 2\mathbb{E}_\sigma U_\theta(g_\sigma)/\mathrm{Var}g$.

Proposition I.2 (Tribes; [*proved here*])**.** *Let $P(t) = 1 - 2 \cdot \mathbf{1}[t \geq 1]$, so that $g = 1 - 2\,\mathrm{OR}_i A_i$ is Tribes of degree $d = sw$, $\iota_0 = 2^{1-w}\pi_0'$, $\mathrm{Var}g = 4\pi_0(1-\pi_0)$ and $\iota_i = 2\phi_i\Pi^{(-i)}$. For every $\theta > 0$,*

$$\mathbb{E}_\sigma U_\theta(g_\sigma) \leq 4s \sum_{r=1}^{w} \kappa_r (1 - 2^{-r}) \min(2^{-r}\pi_0', \theta/2).$$

Proof. Let Π_S, Π_W be the products of $1-\phi_i$ over the alive blocks outside and inside W, so $\Pi_0 = \Pi_S\Pi_W$ ($U=0$ if some block is fully fixed). By Lemma 8.1(a), $\mathbb{E}[g_\sigma \mid x_W] = -1 + 2\Pi_S Z_W$ with $Z_W \in \{0,1\}$ of mean Π_W, so $U_\theta(g_\sigma) = 4\Pi_S^2\Pi_W(1-\Pi_W) \le 4\Pi_0(1-\Pi_W) \le 4\sum_{i\in W}\phi_i\Pi_0$. For $i \in W$, $4\phi_i\Pi_0 = 4\phi_i(1-\phi_i)\Pi^{(-i)} = 2(1-\phi_i)\iota_i \le 2(1-\phi_i)\min(\iota_i,\theta)$. Take expectations; by Jensen's inequality (min is concave) and independence, $\mathbb{E}\min(2^{1-r}\Pi^{(-i)},\theta) \le \min(2^{1-r}\pi_0',\theta)$. □

Corollary I.3 (Tribes threshold; [*proved here*])**.** *In Proposition I.2 let $\theta = (1+\xi)L\iota_0$ with $\xi > 0$ and $\theta \le \mathrm{Var}(g)^C$, and suppose $(1+\xi)L \le 2^{\alpha w}$ with $\alpha < \frac12$ and $(\frac12-\alpha)w \ge 1$. Then*

$$\varepsilon_{\min}(g;L) \;\ge\; 1 - \frac{2\lambda\pi_0'}{\pi_0(1-\pi_0)}\Big(e^{-2((1/2-\alpha)w-1)^2/w} + B_\alpha\Big),$$

with $B_\alpha = e^{-2(1/2-\alpha)^2 w}$ for $\alpha \in [\frac13,\frac12)$ and $B_\alpha = 0.945^w$ for $\alpha < \frac13$. With $s = \mathrm{round}(2^w \ln 2)$: for every fixed $a < \frac12$ and every cap exponent, $\varepsilon_{\min}(g;d^a) \ge 1 - d^{-\Omega((1/2-a)^2)}$; and for every fixed L, $\varepsilon_{\min} \to 1$ as $w\to\infty$, so $\mathrm{TT}(\varepsilon, O(1))$ is false for every $\varepsilon < 1$.

Proof. Put $x = 1-\alpha$ and split the sum of Proposition I.2 at $r_c = \lfloor xw \rfloor$. For $r \ge r_c$ use $s\kappa_r 2^{-r} = \lambda\Pr[\mathrm{Bin}(w,\frac12) = r]$ and Hoeffding's inequality, whose deviation $r_c - \frac{w}{2} \ge (\frac12-\alpha)w - 1$ is nonnegative. For $r < r_c$ use $\theta/2 \le 2^{(\alpha-1)w}\pi_0'$ and $s\kappa_r 2^{(\alpha-1)w} = \lambda\Pr[\mathrm{Bin}(w,\frac12) = r]\,2^{r-xw}$; then $\sum_{r\le xw}\Pr[\mathrm{Bin}(w,\frac12)=r]2^{r-xw} \le \min_{t\in[0,1]} 2^{-w}2^{(t-1)xw}(1+2^{1-t})^w$, which is $e^{-wD(x\|1/2)} \le e^{-2(x-1/2)^2 w}$ for $x \le \frac23$ (take $2^{1-t} = x/(1-x)$) and at most 0.945^w for $x > \frac23$ (take $t=0$). Divide $2\mathbb{E}U$ by $\mathrm{Var}g$. For this s, $\mathrm{Var}g = 1 - O(4^{-w})$, so the cap holds for every fixed C, and $\log_2 d = w + \log_2(\lambda w)$, so $L = d^a$ corresponds to $\alpha = a(1 + O(\log w/w))$. □

At small size, $\varepsilon_{\min} = 0.578$ at $L = 2$ under the cap Var^2 (Tribes with $w=2$, $s=5$, all 3^{10} restrictions) [*numerical*]. For $a \ge \frac12$ the behaviour of Tribes is known only numerically: Monte Carlo gives $\varepsilon_{\min} < 0$ at $a = \frac12$ for $8 \le w \le 28$ [*numerical*].

Theorem I.4 (compositions; [*proved here*])**.** *Assume $\Delta P(0) \ne 0$. Let $\theta = (1+\xi)L\iota_0^+$ with $\xi > 0$ and $\iota_0^+ \ge \iota_0$, and assume $\theta \le \mathrm{Var}(g)^C$ and $\lambda(1+(3/4)^w) < \ln 4$. Put $\theta' = 8\theta/\Delta P(0)^2$ and $r_1 = \lfloor \log_2(1/\theta') \rfloor + 1$. Then, for every L satisfying these conditions,*

$$\varepsilon_{\min}(g;L) \;\ge\; 1 - \frac{8\lambda}{\mathrm{Var}g}\Big(\Pr\big[\mathrm{Bin}(w,\tfrac12) \ge r_1\big] + \pi_4\Big), \qquad \pi_4 = \lambda 2^{-w} + \frac{\frac94\lambda(3/4)^w}{(\ln 4 - \lambda(1+(3/4)^w))^2}.$$

Proof. As $|Q| \le 1$, $U_\theta(g_\sigma) = \mathrm{Var}Q(N_W) \le \mathbb{E}(Q(N_W)-Q(0))^2 \le 4\Pr[N_W \ge 1] \le 4\sum_{i\in W}\phi_i$. If $i \in W$ and $\Pi^{(-i)} \ge \frac14$ then $\phi_i < \theta'$, since otherwise $\iota_i \ge \theta'\Delta P(0)^2/8 = \theta$. Hence, by independence and $\mathbb{E}\phi_i\mathbf{1}[r_i \ge 1] \le 2^{-w}$, $\mathbb{E}[\phi_i\mathbf{1}[i\in W]] \le \mathbb{E}[\phi_i\mathbf{1}[\phi_i < \theta']] + 2^{-w}\Pr[\Pi^{(-i)} < \frac14]$, and $\mathbb{E}[\phi_i\mathbf{1}[\phi_i<\theta']] = \sum_{r\ge r_1}\binom{w}{r}4^{-w} = 2^{-w}\Pr[\mathrm{Bin}(w,\frac12)\ge r_1]$. The event $\Pi^{(-i)} < \frac14$ requires that another block is fully fixed, which has probability at most $s4^{-w} = \lambda 2^{-w}$, or that $Z = \sum_{i'}(\tilde\phi_{i'} + \tilde\phi_{i'}^2) > \ln 4$, where $\tilde\phi = \phi\mathbf{1}[r \ge 1] \le \frac12$ and $-\ln(1-y) \le y + y^2$ on $[0,\frac12]$. Since $\mathbb{E}\tilde\phi \le 2^{-w}$, $\mathbb{E}\tilde\phi^2 = 4^{-w}((3/2)^w - 1)$ and $(\tilde\phi + \tilde\phi^2)^2 \le \frac94\tilde\phi^2$, we have $\mathbb{E}Z \le \lambda(1+(3/4)^w)$ and $\mathrm{Var}Z \le \frac94\lambda(3/4)^w$, and Chebyshev's inequality bounds $\Pr[Z > \ln 4]$. Summing over the s blocks gives $\mathbb{E}U_\theta(g_\sigma) \le 4\lambda(\Pr[\mathrm{Bin}(w,\frac12)\ge r_1] + \pi_4)$. □

Proposition I.5 (near-OR compositions, first moment; [*proved here*])**.** *Suppose $P(0) \le 1$ and $P(t) \in [-1, -1+4\delta']$ for $t \ge 1$, with $0 \le \delta' \le \frac18$, and put $\theta'' = 2\theta/\Delta P(0)^2$. Then*

$$\mathbb{E}_\sigma U_\theta(g_\sigma) \le 5s\sum_{r=1}^{w}\kappa_r(1-2^{-r})\min(2^{-r}\pi_0', \theta'') + 100\lambda\delta'^2.$$

Proof. Let $Z = Q(N_W) + 1 \geq 0$. On $\{N_W = 0\}$, $Z = z_0 := Q(0) + 1 \leq 2\Pr[N_{\mathrm{rest}} = 0] + 4\delta'$, and on $\{N_W \geq 1\}$, $Z \in [0, 4\delta']$. Conditioning on $\mathbf{1}[N_W \geq 1]$ gives $U_\theta(g_\sigma) = \mathrm{Var} Z \leq (1 - \Pi_W)[\Pi_W(z_0^2 + 16\delta'^2) + 4\delta'^2]$. By $(a+b)^2 \leq \frac{5}{4}a^2 + 5b^2$, $z_0^2 \leq 5\Pr[N_{\mathrm{rest}} = 0] + 80\delta'^2$, and as $\Pi_W \Pr[N_{\mathrm{rest}} = 0] = \Pi_0$, $U_\theta(g_\sigma) \leq 5\Pi_0(1 - \Pi_W) + 100\delta'^2(1 - \Pi_W)$. Now $1 - \Pi_W \leq \sum_{i \in W} \phi_i$, $\Pi_0 \phi_i = (1 - \phi_i)\phi_i \Pi^{(-i)}$, and $\phi_i \Pi^{(-i)} < \theta''$ for $i \in W$ by the lower bound on ι_i. Finish as in Proposition I.2, using $\mathbb{E}(1 - \Pi_W) \leq \sum_i \mathbb{E}\phi_i \mathbf{1}[r_i \geq 1] \leq \lambda$. □

For Tribes, $\delta' = 0$ and the exact formula for U give the constant 4 of Proposition I.2.

Corollary I.6 (approximate OR; [*proved here*])**.** *Let $c > 0$ be fixed, $s \geq 2$, $x(t) = (s + 1 - 2t)/(s - 1)$, $q = \lceil c\sqrt{s} \rceil$, $R(t) = T_q(x(t))/T_q(x(0))$, $\delta = 1/T_q(x(0))$ and $P = (2R - 1)/(1 + 2\delta)$, so that $d = qw$.*

- *(a) P satisfies the hypothesis of Proposition I.5 with $\delta' = \delta/(1 + 2\delta)$; moreover $\delta \leq 1/\cosh(2c)$, $|\Delta P(0)| \geq 2 - 6\delta'$, $\mathrm{Var} g \geq \pi_0(1 - \pi_0)(2 - 6\delta')^2$ and $\iota_0 \leq \iota_0^+ := \frac{2^{-w}}{2}(4\pi_0' + 16\delta'^2)$.*
- *(b)* Balanced. *With $s = \mathrm{round}(2^w \ln 2)$, $\log_2 d = w/2 + O(\log w)$, and for every fixed $a < 1$ and every cap exponent, $\varepsilon_{\min}(g; d^a) \geq 1 - d^{-\Omega((1-a)^2)}$. So $\mathrm{TT}(\varepsilon, L)$ fails for $L \leq d^{1-o(1)}$ and $\varepsilon < 1 - o(1)$.*
- *(c)* Sparse. *With cap $C \geq 2$ and $s = \mathrm{round}(2^{w(1-1/(2C))})$, so that $\lambda \approx 2^{-w/(2C)}$ and $\mathrm{Var} g \approx 4\lambda$, $\log_2 d = (\frac{1}{2} - \frac{1}{4C})w + O(\log w)$, and $\mathrm{TT}(1 - o(1), d^a)$ fails for every $a < 2C/(2C - 1)$ ($a < \frac{4}{3}$ for $C = 2$). The same computation for sparse Tribes, via Corollary I.3, gives only $a < C/(2C - 1)$.*
- *(d) For $C = 2$, $w = 1000$, $c = 3$ (so $s = 2^{750}$, $q = 3 \cdot 2^{375}$, $\log_2 d = 386.55\ldots$) and $L = d^{1.2}$, Theorem I.4 gives $r_1 = 535$ and $\varepsilon_{\min} \geq 0.9700637$, by an exact evaluation with integer binomial tails [proved here, computer-assisted]. So $\mathrm{TT}(0.97, d^{1.2})$ with cap 2 is false, by a thin margin: the bound is a step function of the exponent and drops to 0.9649 by $a = 1.20025$. The robust statement is that $\mathrm{TT}(0.96, d^{1.2})$ with cap 2 is false (margin 0.01).*

Proof. Since x maps $\{1, \ldots, s\}$ into $[-1, 1]$ and $x(0) > 1$, $R(0) = 1$ and $|R(t)| \leq \delta$ for $t \geq 1$, so $|P| \leq 1$, $P(0) = 1/(1 + 2\delta)$ and $P(t) \in [-1, -1 + 4\delta']$ for $t \geq 1$. Write $x(0) = 1 + 2/(s - 1) = \cosh \tau_0$; then $\tanh(\tau_0/2) = s^{-1/2}$, so $\tau_0 \geq 2\tanh(\tau_0/2) = 2/\sqrt{s}$, $q\tau_0 \geq 2c$ and $T_q(x(0)) = \cosh(q\tau_0) \geq \cosh(2c)$. Hence $|\Delta P(0)| \geq 2 - 6\delta'$, $\mathrm{Var} g \geq \pi_0(1 - \pi_0)(2 - 6\delta')^2$ (Cauchy–Schwarz against $\mathbf{1}[N = 0]$) and $\iota_0 \leq \iota_0^+$, as $|\Delta P(t)| \leq 4\delta'$ for $t \geq 1$. In Theorem I.4, $\log_2(1/\theta') = w - \log_2 L - O(1)$ and $\pi_4 = O(\lambda(3/4)^w)$. (b) Here $\lambda = \ln 2 + o(1)$, $\mathrm{Var} g$ is bounded below and $\log_2 L = \frac{a}{2}w + O(\log w)$, so $r_1 \geq (1 - \frac{a}{2})w - O(\log w)$ and Hoeffding's inequality bounds the tail by $e^{-\Omega((1-a)^2 w)}$; the cap holds since $\theta = O(L2^{-w})$. (c) Here $8\lambda/\mathrm{Var} g = O(1)$, and $\log_2 L = a(\frac{1}{2} - \frac{1}{4C})w + O(\log w) < \frac{w}{2} - \Omega(w)$ for $a < 2C/(2C - 1)$, so again $r_1 - \frac{w}{2} = \Omega(w)$ and the tail and π_4 tend to 0; the cap $\theta \approx 2L2^{-w} \leq (4\lambda(1 - 3\delta')^2)^C \approx 4^C 2^{-w/2}(1 - 3\delta')^{2C}$ holds for large w. (d) is a direct evaluation of the bound. □

Proposition I.5 gives $1 - O(\delta'^2) - d^{-\Omega((1-a)^2)}$ in (b). It beats Theorem I.4 numerically only for Tribes and for approximate OR at $a \geq 0.9$ (0.969 against 0.958 at $w = 1000$, $a = 0.9$). Elsewhere it is worse, because its constant is 5 rather than 4 and it has a floor $1 - O(\delta'^2) \approx 0.9966$: for the sparse family at $w = 1000$ it gives 0.961 against 0.970 at $a = 1.2$ and 0.568 against 0.630 at $a = 1.25$; at $w = 400$, $a = 1.2$, 0.309 against 0.400; and for approximate OR at $w = 400$, $a = 0.8$, 0.9920 against 0.9936 (cap 2; [*numerical*]). The sparse rows with $w \leq 200$ fail the cap near $a = \frac{4}{3}$, so only $w \geq 400$ is usable there.

Other inner functions. Let $f = \mathbf{1}_A \colon \{-1, 1\}^n \to \{0, 1\}$ with $p = \mathbb{E}f = 2^{-\ell} > 0$, let $\mu_\sigma = \mathbb{E}f_\sigma$ and let $Y_\sigma = \log_2(\mu_\sigma/p)$ be the amplification under σ. The firing-mass measure is $\nu(\sigma) =$

$\Pr[\sigma]\mu_\sigma/p$; equivalently, draw x uniform in A and an alive set D (each coordinate independently with probability $\frac{1}{2}$) and fix x on $F = [n] \setminus D$.

Proposition I.7 (no inner function beats AND on average; [*proved here*])**.** (a) $\mathbb{E}_\nu 2^Y = \mathrm{Stab}_{1/2}(f)/p^2 \le p^{-2/3}$. (b) $\mathbb{E}_\nu Y \le \ell/2$, *with equality if and only if A is a subcube; for* AND_ℓ, $Y \sim \mathrm{Bin}(\ell, \frac{1}{2})$ *under* ν. (c) *For every* $\beta > \frac{1}{2}$, $\nu(Y \ge \beta\ell) \le \min\{(1+p^{2\beta-1})/2,\ p^{\beta-2/3}\}$. *So no Boolean inner function, of any degree, amplifies its firing mass by more than AND's* $p^{-1/2}$ *in geometric mean, and for every fixed* $\beta > \frac{1}{2}$ *at most a fraction* $(1+p^{2\beta-1})/2 \to \frac{1}{2}$ *(as* $p \to 0$*) of the firing mass is amplified beyond* $p^{-\beta}$.

Proof. Couple $\sigma = (D, x_F)$ with $\sigma^* = (F, x_D)$, built from the same x; both are ν-distributed, since D and F have the same law. (a) $\mathbb{E}_\nu 2^Y = \mathbb{E}\mu_\sigma^2/p^2$, and $\mathbb{E}\mu_\sigma^2 = \sum_S 2^{-|S|}\widehat{f}(S)^2 \le \|f\|_{3/2}^2 = p^{4/3}$ by hypercontractivity. (b) Fix the partition and let X be uniform on A. Given X_F, X_D is uniform on a fibre of size $\mu_\sigma 2^{|D|}$, so $H(X_D \mid X_F) = |D| + \mathbb{E}\log_2 \mu_\sigma$, and likewise $H(X_F \mid X_D) = |F| + \mathbb{E}\log_2 \mu_{\sigma^*}$. Then $n - \ell = H(X) = H(X_F) + H(X_D \mid X_F) \ge H(X_F \mid X_D) + H(X_D \mid X_F)$, and averaging over partitions gives $2\mathbb{E}_\nu Y \le \ell$. Equality forces X_F and X_D to be independent for every partition, so X has independent coordinates and A is a subcube. For AND_ℓ, Y is the number of fixed relevant bits. (c) Fix the partition and put $\tau = p^{1-\beta}$. The fibres partition A, so at most $(p/\tau)2^{|F|}$ values of x_F have $\mu_\sigma \ge \tau$, and at most $(p/\tau)2^{|D|}$ values of x_D have $\mu_{\sigma^*} \ge \tau$. Hence $\Pr_{x\in A}[Y_\sigma \ge \beta\ell,\ Y_{\sigma^*} \ge \beta\ell] \le (p/\tau)^2 2^n/(p2^n) = p^{2\beta-1}$; if $\nu(Y \ge \beta\ell) = 1-\zeta$, this probability is at least $1 - 2\zeta$. The second bound is Markov's inequality applied to (a). □

In a composition $g = P(\sum_i f(\text{block } i))$ one has $d_j g = \Delta P(N_{-i})\, d_j f_i$, so an alive bit j of block i is a θ-spike only if $\mathrm{Inf}_j(f_\sigma)$ exceeds $L \cdot \mathrm{MaxInf}(f)$ times a pivotality ratio. Moreover $\mathrm{Inf}_j(f_\sigma) \le \mu_\sigma/2$, and if $\deg f = k$ then $2^{-k-1} \le \mathrm{Inf}_j(f) \le p/2$ for every relevant j, because the nonzero function $d_j f$ of degree $\le k-1$ is nonzero with probability at least 2^{1-k}; so L times the ratio is below $2^{Y+k-\ell}$ [*proved here*]. Hence inner functions of minimal degree ($k = \ell$, e.g. AND) are covered by Proposition I.7; inner functions with $k > \ell$ and $\mathrm{MaxInf}(f) \ll p$ are not, since the proposition bounds the amplification of μ_σ while TT spikes are measured against $\mathrm{MaxInf}(f)$. Threshold inner functions Thr_k^m with $m \le 80$ give amplification $p^{-0.40}$ to $p^{-0.46}$, below AND's $p^{-1/2}$ [*numerical*]. The factor $\frac{1}{2}$ in (c) is exactly TT's factor 2: a composition in which half of each block's firing mass stays untouched has $2\mathbb{E}U \approx \mathrm{Var}g$, i.e. $\varepsilon_{\min} \approx 0$ [*heuristic*], so (c) alone does not prove TT for compositions.

Remark I.8 (where the threshold lies). Throughout, $C \ge 2$: $\mathrm{AA}(A, c)$ needs $A \ge 2$, so the hypothesis $A \le C$ forces $C \ge 2$, and then $2C/(2C-1) \le \frac{4}{3} < 2 \le A$. Let $a^*(\varepsilon) = \inf\{a : \mathrm{TT}(\varepsilon, Kd^a)$ holds for some constant $K\}$. [*proved here*] For every fixed $\varepsilon \in (0,1)$, $a^*(\varepsilon) \ge 2C/(2C-1)$ (Corollary I.6(c)); under $\mathrm{AA}(A, c)$ with $A \le C$ also $a^*(\varepsilon) \le A$ (Theorem 8.7(ii)), and at the chain's $\varepsilon = 1/(16ed)$ the upper bound becomes $2A$, because $(d/\varepsilon)^A = (16ed^2)^A$. Spike-free g satisfy TT for $L \ge d\,9^{d-1}/\varepsilon$ unconditionally (Proposition I.1). Without the conjecture, TT is proved only at a θ-dependent loss with a quasi-polynomial prefactor (Proposition 10.3, through Theorem 10.26), and $\mathrm{TT}(1/(16ed), d^b)$ for a constant b would give $(\mathrm{Var}f/d)^{O(\log d)}$ (Corollary 8.5). [*heuristic*] (i) In compositions $P(N)$ a block becomes a spike only through $\phi_i = 2^{-r_i}$ times a pivotality ratio, and the ϕ-mass over r is exactly $\mathrm{Bin}(w, \frac{1}{2})$ (proved); so the amplification is at most $2^{w/2+O(\sqrt{w})}$ if the pivotality ratio is $O(1)$, which is unproved for general P. (ii) By Coppersmith–Rivlin and Markov (classical), if $|P| \le 1$ on $\{0, \dots, s\}$ and $q \le \sqrt{s}$, then $|\Delta P| = O(q^2/s)$, so a jump $\gamma = |\Delta P(0)|$ needs $q = \Omega(\sqrt{\gamma s})$. (iii) $\mathrm{Var}g \le 4\lambda$ and $\iota_0 \ge \frac{2^{-w}}{2}\gamma^2\pi_0'$ (proved). Optimizing $\log L/\log d$ under (i)–(iii), with the model $\mathrm{Var}g \approx \gamma^2\lambda$, gives $a \le 2C/(2C-1) + o(1)$ for $C \ge 2$, so AND-block compositions should not go beyond

Corollary I.6(c). The unproved ingredients are the pivotality ratio, a lower bound on U for the non-spike mass (Theorem I.4 and Proposition I.5 bound U only from above), and the model $\mathrm{Var} g \approx \gamma^2\lambda$. (iv) For other inner functions, Proposition I.7 shows that hypercontractivity caps only the mean amplification of the firing mass, at $p^{-2/3}$, while for every fixed $\beta > \frac{1}{2}$ at most a fraction tending to $\frac{1}{2}$ of it is amplified beyond $p^{-\beta}$, AND being the equality case in log-average; inner f with $\deg f > \log_2(1/p)$ and $\mathrm{MaxInf}(f) \ll p$ are not covered.

So any proof of TT with polynomial loss must use $L \geq d^{2C/(2C-1)-o(1)}$. The next lemma localizes what a counterexample must do.

Lemma I.9 (new spikes; [*proved here*])**.** *Let $g \in \mathcal{B}_d$, $\theta > 0$, $L \geq 1$, and J', h as in Lemma 8.6. For σ with alive set A let $G_\sigma = \mathbb{E}[g_\sigma \mid x_{A\setminus J'}]$, $K_\sigma = J_\theta(g_\sigma) \setminus J'$ (the new spikes) and $V_\sigma = \sum_{T\subseteq A\setminus J', T\cap K_\sigma\neq\emptyset} \widehat{G_\sigma}(T)^2$. Then*

$$2\mathbb{E}_\sigma U_\theta(g_\sigma) \;\geq\; 2\mathbb{E}_\sigma \mathrm{Var}(G_\sigma) - 2\mathbb{E}_\sigma V_\sigma \;\geq\; \sum_{S\neq\emptyset} \hat{h}(S)^2(2 - 2^{1-|S|}) - 2\mathbb{E}_\sigma V_\sigma.$$

Hence the inequality of Definition 8.2 holds for g at this θ whenever $\mathbb{E}_\sigma V_\sigma \leq \frac{1}{2}\sum_{|S|\geq 2} \hat{h}(S)^2(1 - 2^{1-|S|}) + \frac{\varepsilon}{2}\mathrm{Var}(g)$.

Proof. Fix σ, $u = g_\sigma$, $K = K_\sigma$. The sum defining $U_\theta(u)$ contains every $T \neq \emptyset$ with $T \subseteq A \setminus (J_\theta(u) \cup J') = A \setminus (J' \cup K)$, so by the tower property $U_\theta(u) \geq \mathrm{Var}(\mathbb{E}[u \mid x_{A\setminus(J'\cup K)}]) = \mathrm{Var}(\mathbb{E}[G_\sigma \mid x_{A\setminus(J'\cup K)}]) = \mathrm{Var}(G_\sigma) - V_\sigma$, as G_σ depends only on $x_{A\setminus J'}$. Condition on A and on the fixed values outside J': the fixed values on $J' \setminus A$ are uniform, and averaging G_σ over them gives the restriction h_σ of h; by convexity of the variance, $\mathbb{E}_\sigma \mathrm{Var}(G_\sigma) \geq \mathbb{E}_\sigma \mathrm{Var}(h_\sigma) = \sum_{S\neq\emptyset} \hat{h}(S)^2 \Pr[S \cap A \neq \emptyset] = \sum_{S\neq\emptyset} \hat{h}(S)^2(1 - 2^{-|S|})$. For the last claim use $U_{\theta/L}(g) = \mathrm{Var} h = \sum_{S\neq\emptyset} \hat{h}(S)^2$ and note that $1 - 2^{1-|S|}$ vanishes for $|S| = 1$. □

Thus TT fails only if the new spikes, coordinates whose influence grows by a factor larger than L under one restriction, touch inside G_σ more than $\frac{1}{2}\sum_{|S|\geq 2}(1-2^{1-|S|})\hat{h}(S)^2 + \frac{\varepsilon}{2}\mathrm{Var} g$; the first term lies between $\frac{1}{4}$ and $\frac{1}{2}$ of the variance $W^{\geq 2}(h)$ of h above level 1, which is what the narrowing in the AND-block compositions touches.

Appendix J. Size-biased laws: RL1, the Load Lemma and Theorem A

Throughout this appendix logarithms are natural, the cube $\{-1,1\}^n$ carries the uniform measure U, and $\mathbb{E}$ and $\|\cdot\|_2$ refer to U unless stated otherwise. Every function on the cube is identified with its multilinear polynomial $\psi(x) = \sum_S \widehat{\psi}(S)x^S$, which also defines ψ on $\mathbb{R}^n$. For a coordinate u we write $D_u\psi(x) = (\psi(x^{u\to 1}) - \psi(x^{u\to -1}))/2$ and $\mathrm{Inf}_u(\psi) = \mathbb{E}[(D_u\psi)^2] = \sum_{S\ni u} \widehat{\psi}(S)^2$; D_u is the derivative written d_u in the main text. As there, L denotes the loss and T the length of the chain, while T_r is the noise operator. For φ with $M := \mathbb{E}[\varphi^2] > 0$ the *size-biased law* is the probability measure

$$\widetilde{\Pr} := (\varphi^2/M) \cdot U,$$

and $\widetilde{\mathbb{E}}$ is its expectation. For a partial assignment ρ, φ_ρ is the restriction of φ, a function of the unassigned coordinates.

J.1. **RL1: a level-1 inequality for size-biased laws in spread directions.** *Setting.* Let $n, k \geq 1$ and let $\varphi\colon \{-1,1\}^n \to [-1,1]$ have degree at most k. Put $M := \mathbb{E}[\varphi^2] > 0$ (so $M \leq 1$) and $\Lambda_M := \ln(1/M) \geq 0$. Let $a \in \mathbb{R}^n$ satisfy $\sum_i a_i^2 = 1$ and $\max_i |a_i| \leq \eta$, and put

$$\mu := \frac{\mathbb{E}[\varphi(y)^2 \langle a, y\rangle]}{M} = \widetilde{\mathbb{E}}\, \langle a, y\rangle.$$

Theorem J.1 (RL1; [*proved here, audited*])**.** *Put $\varepsilon_* := e^3 k^4 \eta^2$. If $\varepsilon_* \leq 1/2$, then*

$$\mu^2 \leq 10e\, \frac{\ln(e/M)}{\ln(1/\varepsilon_*)} \qquad (10e = 27.18\ldots).$$

Corollary J.2 (RL1 in terms of η; [*proved here, audited*])**.** *(a) If $\eta \leq e^{-3}k^{-4}$, then $\mu^2 \leq 10e\, \ln(e/M)/\ln(1/\eta)$.*

(b) For every $\eta < 1$, $\mu^2 \leq \max\{10e,\, 6 + 8\ln k\}\, \ln(e/M)/\ln(1/\eta)$.

The chain interface is the following consequence. It applies with $k = d-1$ to every partial restriction of a derivative $D_j g$ with $g \in \mathcal{B}_d$ (such restrictions are bounded by 1), and with $k = d$ to every restriction of a function in $\mathcal{B}_d$.

Corollary J.3 (chain interface; [*proved here, audited*])**.** *Let $k \geq 1$ and $0 < \tau \leq 1/(2e^3k^4)$. Let $\varphi\colon \{-1,1\}^n \to [-1,1]$ have degree at most k and $M = \mathbb{E}[\varphi^2] > 0$. Put $c_u := \widetilde{\mathbb{E}}\, y_u$ and $c^{\mathrm{sm}} := (c_u \mathbf{1}[|c_u| < \tau])_u$. Then*

$$\|c^{\mathrm{sm}}\|_2^2 \leq \max\{\tau,\ \kappa \ln(e/M)\}, \qquad \kappa := \frac{10e}{\ln(1/\tau) - 3 - 4\ln k}.$$

No lower bound on M and no condition on the direction of c is needed.

Multilinear calculus. For multilinear ψ let $\partial_i \psi := \sum_{S \ni i} \widehat{\psi}(S) x^{S\setminus\{i\}}$ be the formal partial derivative; on the cube $\partial_i \psi = D_i \psi$. For $v \in \mathbb{R}^n$ put $D_v := \sum_i v_i \partial_i$. Since the ∂_i commute and $\partial_i^2 = 0$, we have $D_v^j = j! \sum_{|R|=j} v^R \partial_R$. Hence $D_v^j \psi$ is multilinear of degree at most $\deg \psi - j$, and $D_{rv}^j = r^j D_v^j$ for $r \in \mathbb{R}$.

(T1) *Taylor.* For $x, v \in \mathbb{R}^n$, $t \mapsto \psi(x+tv)$ is a polynomial of degree at most $\deg \psi$, and $(d/dt)^j \psi(x+tv) = (D_v^j \psi)(x+tv)$. Indeed, $\frac{d}{dt}\psi(x+tv) = \sum_i v_i (\partial\psi/\partial x_i)(x+tv)$; for a multilinear ψ the analytic partial $\partial\psi/\partial x_i$ equals $\partial_i \psi$, and $D_v \psi$ is again multilinear, so we may iterate.

(T2) *Solid cube.* For $x \in [-1,1]^n$ let Y have independent ± 1 coordinates with $\mathbb{E}Y_i = x_i$. Then $\mathbb{E}Y^S = x^S$, so $\varphi(x) = \mathbb{E}\varphi(Y)$ and $|\varphi(x)| \leq 1$ on $[-1,1]^n$.

(T3) *Noise inversion.* Put $(T_r\psi)(y) := \psi(ry)$, so that $\widehat{T_r\psi}(S) = r^{|S|}\widehat{\psi}(S)$. If $\deg\psi \leq m$ and $0 < r \leq 1$, then $\|\psi\|_2^2 = \sum_S \widehat{\psi}(S)^2 \leq r^{-2m} \sum_S r^{2|S|}\widehat{\psi}(S)^2$, so $\|\psi\|_2 \leq r^{-m}\|T_r\psi\|_2 \leq r^{-m} \max_{y\in\{-1,1\}^n} |\psi(ry)|$.

Lemma J.4 (the tilt is a directional derivative; [*proved here, audited*])**.** *For each i, $\mathbb{E}[\varphi^2 y_i] = 2\mathbb{E}[\varphi\, \partial_i \varphi]$. Hence $M\mu = 2\mathbb{E}[\varphi D_a \varphi]$ and*

$$\mu^2 \leq 4\|D_a\varphi\|_2^2 / M. \tag{7}$$

Proof. Write $\varphi = A + y_i B$, where $A := \mathbb{E}_{y_i}\varphi$ and $B := \partial_i \varphi$ do not depend on y_i. Since $y_i^2 = 1$, we have $\varphi^2 y_i = (A^2 + B^2) y_i + 2AB$, so $\mathbb{E}[\varphi^2 y_i] = 2\mathbb{E}[AB]$. Also $\mathbb{E}[\varphi\, \partial_i \varphi] = \mathbb{E}[AB] + \mathbb{E}[y_i B^2] = \mathbb{E}[AB]$. Summing with weights a_i gives $M\mu = 2\mathbb{E}[\varphi D_a \varphi] \leq 2\|\varphi\|_2 \|D_a\varphi\|_2 = 2\sqrt{M}\|D_a\varphi\|_2$. □

Hermite expansion along a. Let $G \sim N(0, I_n)$. For multilinear ψ, χ we have $\mathbb{E}[\psi(G)\chi(G)] = \sum_S \widehat{\psi}(S)\widehat{\chi}(S) = \mathbb{E}[\psi\chi]$, because the monomials G^S are orthonormal. Put $s := \langle a, G\rangle \sim N(0,1)$ and $Z := G - sa$. The pair (s, Z) is jointly Gaussian with $\mathrm{Cov}(s, Z_i) = a_i - a_i = 0$, so s and Z are independent, and $G = Z + sa$. For fixed z, $h_z(t) := \varphi(z + ta)$ is a polynomial of degree at most k, and $h_z^{(j)}(t) = (D_a^j\varphi)(z + ta)$ by (T1).

Let He_j be the probabilists' Hermite polynomials, so that $\mathbb{E}[He_i He_j] = j!\,\delta_{ij}$ under $N(0,1)$ and $He_j' = jHe_{j-1}$. For a polynomial h, j Gaussian integrations by parts (using $He_j\gamma_{\mathrm{N}} = (-1)^j\gamma_{\mathrm{N}}^{(j)}$ for the standard normal density γ_{N}) give $\mathbb{E}[h(s)He_j(s)] = \mathbb{E}[h^{(j)}(s)] =: b_j$. Hence $h = \sum_{j\le \deg h}(b_j/j!)He_j$, $\mathbb{E}h(s)^2 = \sum_j b_j^2/j!$ and $\mathbb{E}h'(s)^2 = \sum_{j\ge1} jb_j^2/j!$. Define, for $0 \le j \le k$,

$$b_j(z) := \mathbb{E}_s[(D_a^j\varphi)(z + sa)], \qquad W_j := \mathbb{E}_Z[b_j(Z)^2]/j!.$$

Lemma J.5 ([*proved here, audited*])**.** *$M = \sum_{j=0}^k W_j$, $\|D_a\varphi\|_2^2 = \sum_{j=1}^k jW_j$, and $0 \le W_j \le \|D_a^j\varphi\|_2^2/j!$. Consequently*

$$\mu^2 \le 4\,\frac{\sum_{j=1}^k jW_j}{\sum_{j=0}^k W_j}. \tag{8}$$

Proof. Condition on Z and apply the Hermite facts to h_Z, noting $h_Z'(s) = (D_a\varphi)(G)$. This gives $\mathbb{E}[\varphi(G)^2] = \mathbb{E}_Z\sum_j b_j(Z)^2/j!$ and $\mathbb{E}[(D_a\varphi)(G)^2] = \mathbb{E}_Z\sum_j jb_j(Z)^2/j!$. Since φ and $D_a\varphi$ are multilinear, the left sides are M and $\|D_a\varphi\|_2^2$. Finally $b_j(Z) = \mathbb{E}[(D_a^j\varphi)(G) \mid Z]$, so by Jensen $\mathbb{E}b_j(Z)^2 \le \mathbb{E}[(D_a^j\varphi)(G)^2] = \|D_a^j\varphi\|_2^2$, because $D_a^j\varphi$ is multilinear. Inequality (8) follows from (7). □

Bernstein-type bounds.

Lemma J.6 (derivatives at the centre; [*proved here, audited*])**.** *Let p be a real polynomial of degree at most k with $|p| \le 1$ on $[-h, h]$, $h > 0$. Then $|p^{(j)}(0)| \le e\sqrt{j}\,(k/h)^j$ for $1 \le j \le k$.*

Proof. By scaling we may take $h = 1$. Let $J(w) = (w + w^{-1})/2$. The function $F(w) := p(J(w))w^{-k}$ contains only the powers $w^{-2k}, \dots, w^0$, so it is analytic on $\{|w| > 1\} \cup \{\infty\}$. On $|w| = 1$ we have $J(w) \in [-1, 1]$, so $|F| \le 1$ there, and by the maximum principle $|p(J(w))| \le |w|^k$ for $|w| \ge 1$. The map J sends $\{1 \le |w| \le R\}$ onto the closed region bounded by the ellipse with foci ±1 and semi-axes $(R + R^{-1})/2 \ge (R - R^{-1})/2 =: r$. This region contains the disc $|z| \le r$, so $|p| \le R^k$ there (the Bernstein–Walsh bound; see, e.g., Borwein–Erdélyi, *Polynomials and Polynomial Inequalities*, §5.1). Cauchy's estimate gives $|p^{(j)}(0)| \le j!R^k/r^j$. Take $R = e^b$ with $b = j/k$, so that $r = \sinh b \ge b$; then $|p^{(j)}(0)| \le j!\,e^j(k/j)^j$. Finally $j! \le e\,j^{j+1/2}e^{-j}$ for all $j \ge 1$: this is an equality at $j = 1$, and for $j \ge 2$ Robbins' bound $j! \le \sqrt{2\pi}\,j^{j+1/2}e^{-j}e^{1/(12j)}$ applies with $\sqrt{2\pi}e^{1/24} < 2.62 < e$. Hence $|p^{(j)}(0)| \le e\sqrt{j}\,k^j$. □

Lemma J.7 (directional Bernstein inequality on the cube; [*proved here, audited*])**.** *For $1 \le j \le k$ and every $r \in (0, 1)$,*

$$\|D_a^j\varphi\|_2 \le e\sqrt{j}\,r^{-(k-j)}\Big(\frac{k\eta}{1-r}\Big)^j. \tag{9}$$

Consequently $\|D_a^j\varphi\|_2 \le e\sqrt{j}\,(ek^2\eta/j)^j$ for $1 \le j \le k$.

Proof. Fix $y \in \{-1, 1\}^n$ and put $h := (1 - r)/(r\eta)$. For real $|t| \le h$ the point $x(t) := r(y + ta)$ satisfies $|x_i| \le r(1 + |t|\eta) \le 1$, so $|\varphi(x(t))| \le 1$ by (T2). The polynomial $p(t) := \varphi(ry + t\,ra)$ has degree at most k, and $p^{(j)}(0) = r^j(D_a^j\varphi)(ry)$ by (T1). Lemma J.6 gives $r^j|(D_a^j\varphi)(ry)| \le$

$e\sqrt{j}\,(kr\eta/(1-r))^j$, that is, $\max_y |(D_a^j\varphi)(ry)| \le e\sqrt{j}\,(k\eta/(1-r))^j$. Now apply (T3) to $\psi = D_a^j\varphi$, which has degree at most $m = k-j$; this gives (9).

For $j < k$ take $r = 1-j/k$. Then $r^{-(k-j)} = \exp(-(k-j)\ln(1-j/k)) \le \exp((k-j)\frac{j/k}{1-j/k}) = e^j$, using $-\ln(1-x) \le x/(1-x)$, and $k\eta/(1-r) = k^2\eta/j$. For $j = k$ the factor $r^{-(k-j)}$ equals 1, and letting $r \to 0^+$ gives $\|D_a^k\varphi\|_2 \le e\sqrt{k}\,(k\eta)^k \le e\sqrt{k}\,(ek^2\eta/k)^k$. □

Proof of Theorem J.1. By Lemmas J.5 and J.7, and $j! \ge (j/e)^j$, for $1 \le j \le k$:

$$W_j \le \frac{e^2 j\,(ek^2\eta/j)^{2j}}{j!} \le e^2 j\,\varepsilon_*^j\, j^{-3j}, \qquad \text{so} \qquad jW_j \le e^2\varepsilon_*^j,$$

since $j^2 j^{-3j} \le 1$ for $j \ge 1$. Fix an integer $J \ge 0$ (the value $J = 0$ is allowed) and split $\sum_{j\ge1} jW_j$ at J. The part $1 \le j \le J$ is at most $J\sum_{j\le J} W_j \le JM$; it is empty if $J = 0$. The tail is at most $\sum_{j>J} e^2\varepsilon_*^j \le e^2\varepsilon_*^{J+1}/(1-\varepsilon_*) \le 2e^2\varepsilon_*^{J+1}$, and it is empty if $J \ge k$. With (8) and $\Lambda_* := \ln(1/\varepsilon_*) \ge \ln 2$,

$$\mu^2 \le 4\big(J + 2e^2\exp(\Lambda_M - (J+1)\Lambda_*)\big) \qquad \text{for every integer } J \ge 0. \tag{10}$$

Case A: $\Lambda_M \le \Lambda_*/5$. Take $J = 0$. With $u = 4\Lambda_*/5$ and $ue^{-u} \le 1/e$, $\mu^2 \le 8e^2e^{-4\Lambda_*/5} \le 8e^2/(e\cdot\frac45\Lambda_*) = 10e/\Lambda_* \le 10e(1+\Lambda_M)/\Lambda_*$.

Case B: $\Lambda_M > \Lambda_*/5$. Take $J := \lceil(\Lambda_M + \ln(2e^2))/\Lambda_*\rceil - 1 \ge 0$. Then $(J+1)\Lambda_* \ge \Lambda_M + \ln(2e^2)$, so the exponential term in (10) is at most $1/(2e^2)$ and $\mu^2 \le 4(J+1)$. Also $J + 1 < (\Lambda_M + \ln(2e^2))/\Lambda_* + 1 < (\Lambda_M + 2.70 + 5\Lambda_M)/\Lambda_*$, using $1 < 5\Lambda_M/\Lambda_*$. Hence $\mu^2 < 4(6\Lambda_M + 2.70)/\Lambda_* \le 24(1+\Lambda_M)/\Lambda_*$.

In both cases $\mu^2 \le 10e(1+\Lambda_M)/\Lambda_* = 10e\ln(e/M)/\ln(1/\varepsilon_*)$. □

Proof of Corollary J.2. (a) If $\eta \le e^{-3}k^{-4}$, then $\varepsilon_* = (e^3k^4\eta)\eta \le \eta \le e^{-3} < 1/2$ and $\ln(1/\varepsilon_*) \ge \ln(1/\eta)$.

(b) It remains to treat $e^{-3}k^{-4} < \eta < 1$, where $0 < \ln(1/\eta) < 3 + 4\ln k$. We use the level-1 bound $\mu^2 \le 2\Lambda_M$. For every $\lambda \in \mathbb{R}$, Jensen's inequality gives

$$\ln \mathbb{E}e^{\lambda\langle a,y\rangle} \ge \ln \widetilde{\mathbb{E}}\big[e^{\lambda\langle a,y\rangle} M/\varphi^2\big] \ge \lambda\mu - \widetilde{\mathbb{E}}\ln(\varphi^2/M) \ge \lambda\mu - \Lambda_M,$$

because $\varphi^2 \le 1$. Since $\ln \mathbb{E}e^{\lambda\langle a,y\rangle} = \sum_i \ln\cosh(\lambda a_i) \le \lambda^2/2$, taking $\lambda = \mu$ gives $\mu^2 \le 2\Lambda_M \le 2\ln(e/M) \le (6 + 8\ln k)\ln(e/M)/\ln(1/\eta)$. □

Proof of Corollary J.3. If $c^{\mathrm{sm}} = 0$ there is nothing to prove. Otherwise put $a := c^{\mathrm{sm}}/\|c^{\mathrm{sm}}\|_2$, a unit vector. Then $\widetilde{\mathbb{E}}\langle a, y\rangle = \sum_u a_u c_u = \|c^{\mathrm{sm}}\|_2$, and $\eta := \max_u |a_u| < \tau/\|c^{\mathrm{sm}}\|_2$. If $\|c^{\mathrm{sm}}\|_2^2 < \tau$ we are done. Otherwise $\eta < \tau/\sqrt{\tau} = \sqrt{\tau}$, so $\varepsilon_* = e^3k^4\eta^2 < e^3k^4\tau \le 1/2$, and Theorem J.1 gives $\|c^{\mathrm{sm}}\|_2^2 \le 10e\ln(e/M)/\ln(1/\varepsilon_*) \le 10e\ln(e/M)/\ln(1/(e^3k^4\tau)) = \kappa\ln(e/M)$. The denominator is $\ln(1/\tau) - 3 - 4\ln k \ge \ln 2 > 0$. □

Remark J.8 (constants for a fixed-order route). This remark belongs to a fixed-order small-jump argument, in which the bound of Corollary J.3 is paid once at each of the T levels of the chain of Theorem 8.4. It is not used in Section 9.6: there Theorem 9.23 needs only $c(\tau_1)T \le \frac12$, which its threshold $\tau_1 = e^{-345}d^{-129}$ achieves. With $k = d-1$, the fixed-order argument needs $\kappa \le 1/(8(\log_2 d + 3.45))$, that is, $\ln(1/\tau) \ge 3 + 4\ln(d-1) + 80e(\log_2 d + 3.45)$. Since $80e < 217.5$ and $\log_2 d < 1.4427\ln d$, the choice $\tau := e^{-754}d^{-318}$ suffices: $217.5\cdot 1.4427 + 4 < 318$ and $217.5\cdot 3.45 + 3 < 754$. This τ also satisfies $\tau \le 1/(2e^3(d-1)^4)$. In the generic form $\|c^{\mathrm{sm}}\|^2 \le 2^\beta C/\ln(1/\tau)^\beta$ with $C = 10e$ and $\beta = 1$, one would get $\kappa = 20e/\ln(1/\tau)$; the case split in the proof of Corollary J.3 saves this factor 2.

Remark J.9 (sharpness). [*numerical*] Over all $\eta < 1$, the constant in Corollary J.2(b) must grow with k: at $k = 65$ and $\eta = 1/(2k)$ the ratio $\mu^2 \ln(1/\eta)/\ln(e/M)$ is 3.18, and it grows like $c \ln k$. Exact computations for symmetric φ in the regime $\varepsilon_* \le 1/2$ suggest that the truth is close to $\mu^2 \lesssim 0.5 \ln(e/M)/\ln(1/(k\eta))$; the factor k^2 in Lemma J.7, which comes from the noise inversion (T3), is probably not sharp.

J.2. **The Load Lemma for adaptive revelation strategies.** *Setting.* Let $\varphi\colon \{-1,1\}^n \to \mathbb{R}$ be arbitrary (no degree or boundedness assumption), $M = \mathbb{E}[\varphi^2] > 0$, and $\widetilde{\Pr} = (\varphi^2/M) \cdot U$. An *adaptive revelation strategy* $\mathcal{S}$ is a randomized decision tree. It uses coins ω independent of x (under U and under $\widetilde{\Pr}$). At each step it either stops or queries an unrevealed coordinate, as a function of the values revealed so far and of ω. It may never query some coordinates. For a coordinate u, let ρ_u be the partial assignment revealed just before u is queried, or the final assignment if u is never queried; thus $u \notin \operatorname{dom} \rho_u$. Write $N(\rho) := \mathbb{E}[\varphi_\rho^2]$, where the expectation is over the unassigned coordinates, and note $\mathrm{Inf}_u(\varphi_\rho) = \mathbb{E}[(D_u\varphi)_\rho^2]$ for $u \notin \operatorname{dom}\rho$. Since $(D_u\psi)^2 \le (\psi(x^{u\to 1})^2 + \psi(x^{u\to -1})^2)/2$, we have $\mathrm{Inf}_u(\varphi_\rho) \le N(\rho)$. In particular $\mathrm{Inf}_u(\varphi_\rho) = 0$ when $N(\rho) = 0$, and the ratio below is defined $\widetilde{\Pr}$-almost surely.

Lemma J.10 (Load Lemma; [*proved here, audited*])**.** *For every adaptive revelation strategy and every coordinate u,*

$$\widetilde{\mathbb{E}}\Big[\frac{\mathrm{Inf}_u(\varphi_{\rho_u})}{N(\rho_u)}\Big] = \frac{\mathrm{Inf}_u(\varphi)}{M},$$

where the expectation is over $x \sim \widetilde{\Pr}$ and the coins.

Proof. Fix a partial assignment ρ with $u \notin \operatorname{dom}\rho$. On inputs consistent with ρ, whether the run reveals exactly $\operatorname{dom}\rho$ and then queries u (or stops without querying u) depends only on the answers in $\operatorname{dom}\rho$, which ρ fixes, and on ω. Hence $\{\rho_u = \rho\} = \{x_{\operatorname{dom}\rho} = \rho\} \cap E_\rho$, where E_ρ is an event of the coins alone. Since $\widetilde{\Pr}(x_{\operatorname{dom}\rho} = \rho) = 2^{-|\rho|}N(\rho)/M$, this gives

$$\widetilde{\Pr}(\rho_u = \rho) = \Pr(E_\rho)\, 2^{-|\rho|} N(\rho)/M = \Pr_U(\rho_u = \rho)\, N(\rho)/M.$$

Summing over ρ with $N(\rho) > 0$, the left side of the lemma equals $\sum_\rho \Pr_U(\rho_u = \rho)\mathrm{Inf}_u(\varphi_\rho)/M = \mathbb{E}_U[\mathrm{Inf}_u(\varphi_{\rho_u})]/M$.

Let $\mathcal{F}_u$ be the information available just before u is queried, namely ω and the values revealed so far. Under U, conditionally on $\mathcal{F}_u$ with $\rho_u = \rho$, the unrevealed coordinates are uniform, because the conditioning event is $\{x_{\operatorname{dom}\rho} = \rho\}$ intersected with an event of the coins. Since $D_u(\varphi_\rho) = (D_u\varphi)_\rho$ for $u \notin \operatorname{dom}\rho$, it follows that $\mathrm{Inf}_u(\varphi_{\rho_u}) = \mathbb{E}_U[(D_u\varphi)^2 \mid \mathcal{F}_u]$. The tower property then gives $\mathbb{E}_U[\mathrm{Inf}_u(\varphi_{\rho_u})] = \mathbb{E}[(D_u\varphi)^2] = \mathrm{Inf}_u(\varphi)$. □

The proof uses neither the degree nor boundedness of φ, and it allows any adaptive order.

Corollary J.11 (L1: large jumps are rare in proportion to relative influence; [*proved here, audited*])**.** *Call u a β-*large jump *of $(\varphi, \mathcal{S})$ at x if $\mathrm{Inf}_u(\varphi_{\rho_u}) \ge \beta N(\rho_u)$. Then*

$$\widetilde{\Pr}(u \text{ is a } \beta\text{-large jump}) \le \frac{\mathrm{Inf}_u(\varphi)}{\beta M}, \qquad \widetilde{\mathbb{E}}\big[\#\{\beta\text{-large jumps}\}\big] \le \frac{I(\varphi)}{\beta M} \le \frac{\deg\varphi}{\beta},$$

where $I(\varphi) = \sum_u \mathrm{Inf}_u(\varphi)$. Moreover the conditional bias $b_u := \widetilde{\mathbb{E}}[x_u \mid \mathcal{F}_u]$ satisfies

$$b_u = \frac{2\mathbb{E}[AB]}{\mathbb{E}[A^2 + B^2]}, \qquad b_u^2 \le \frac{4\mathrm{Inf}_u(\varphi_{\rho_u})}{N(\rho_u)}, \qquad \widetilde{\Pr}(|b_u| \ge \beta) \le \frac{4\mathrm{Inf}_u(\varphi)}{\beta^2 M},$$

where $\varphi_{\rho_u} = A + x_u B$ with $B = D_u(\varphi_{\rho_u})$ and A not depending on x_u.

Proof. The first bound is Markov's inequality applied to Lemma J.10. Summing over u gives the second, and $I(\varphi) = \sum_S |S|\widehat{\varphi}(S)^2 \le \deg(\varphi)M$. For the bias: by the proof of Lemma J.10, conditionally on $\mathcal{F}_u$ with $\rho_u = \rho$, the $\widetilde{\Pr}$-law of the unrevealed coordinates has density proportional to φ_ρ^2. Since $(A + x_u B)^2 x_u = (A^2 + B^2)x_u + 2AB$, this gives $b_u = 2\mathbb{E}[AB]/\mathbb{E}[A^2 + B^2]$, and $\mathbb{E}[A^2 + B^2] = N(\rho)$. By Cauchy–Schwarz and $\mathbb{E}[A^2] \le N(\rho)$, $b_u^2 \le 4\mathbb{E}[A^2]\mathbb{E}[B^2]/N(\rho)^2 \le 4\mathbb{E}[B^2]/N(\rho) = 4\mathrm{Inf}_u(\varphi_\rho)/N(\rho)$. Hence $\{|b_u| \ge \beta\}$ is contained in the event that u is a $(\beta^2/4)$-large jump, and the first bound applies. □

Corollary J.12 (L2: weighted load, any adaptive orders; [*proved here, audited*])**.** *Let h have degree at most d. For each j with $I_j := \mathrm{Inf}_j(h) > 0$ put $\varphi_j := D_j h$ and $\widetilde{\Pr}_j := (\varphi_j^2/I_j)\cdot U$, and let $\mathcal{S}_j$ be any adaptive revelation strategy (one for each j; the sets of large jumps may depend on the point). Then for every coordinate u,*

$$\sum_{j:\, I_j>0} I_j \,\widetilde{\Pr}_j\big(u \text{ is a } \beta\text{-large jump of } (\varphi_j, \mathcal{S}_j)\big) \le \frac{1}{\beta}\sum_{j\neq u} \mathrm{Inf}_u(D_j h) \le \frac{d-1}{\beta}\mathrm{Inf}_u(h).$$

With a bias threshold τ in place of β-large jumps, the load is at most $(4(d-1)/\tau^2)\mathrm{Inf}_u(h)$.

Proof. Corollary J.11 applied to φ_j (with $M = I_j$) bounds the j-th term by $\mathrm{Inf}_u(D_j h)/\beta$. For $j = u$ the term vanishes, because $D_u D_u h = 0$. For $j \neq u$, $\mathrm{Inf}_u(D_j h) = \sum_{S\ni u,j} \widehat{h}(S)^2$, so $\sum_{j\neq u}\mathrm{Inf}_u(D_j h) = \sum_{S\ni u}(|S|-1)\widehat{h}(S)^2 \le (d-1)\mathrm{Inf}_u(h)$. The bias version follows in the same way from the last bound of Corollary J.11. □

Remark J.13. [*numerical*] The identity of Lemma J.10 was checked on random φ with $n = 6$ and degrees 3–4, under deterministic greedy and randomized adaptive rules with random stopping; the maximal error is $4.6\cdot 10^{-15}$. The bound $b_u^2 \le 4\mathrm{Inf}_u(\varphi_\rho)/N(\rho)$ is attained in these tests. If an exploration queries only coordinates whose current bias is at least τ in absolute value, then Corollary J.11 gives $\widetilde{\Pr}(u \text{ is queried}) \le 4\mathrm{Inf}_u(\varphi)/(\tau^2 M)$ and $\widetilde{\mathbb{E}}|\{\text{queried coordinates}\}| \le 4\deg(\varphi)/\tau^2$ for every rate, with no maximal inequality. For the one-shot restriction of Remark 9.34, Corollary J.12 shows that point-dependent large-bias sets have polynomial load under every adaptive revelation order as soon as the bias threshold is at least $1/\mathrm{poly}(d)$.

J.3. **Theorem A: the small-bias tail cannot produce a spike.** *Hypotheses.*

(H1) $\varphi\colon \{-1,1\}^n \to [-1,1]$ has degree at most d, $d \ge 2$.

(H2) *Oblivious exploration.* (B, z_B), with $z_B := x_B$, is the stopping data of an exploration run on the full cube. The exploration is a randomized decision tree that reads coordinates of $x \sim \widetilde{\Pr}_{\mathrm{full}} := (\varphi^2/\mathbb{E}\varphi^2)\cdot U$ and uses coins independent of everything else. Every coordinate v carries a clock $t_v \sim U[0,1]$; the clocks are i.i.d. and independent of x and of the coins. For $v \notin B$, v is *alive* if $t_v > 1-p$ and *fixed* otherwise. The coordinates of B are revealed at time 0, and those of $F \setminus B = \{v \notin B : t_v \le 1-p\}$ are revealed in increasing clock order.

Consequence of (H2). Fix $\zeta_B := (B = b,\ z_b = \zeta)$ with positive probability. The event $\{B = b,\ x_b = \zeta\}$ is $\{x_b = \zeta\}$ intersected with an event of the coins. Hence, given ζ_B, the law of x_{b^c} is $\widetilde{\Pr} := (\varphi_b^2/q)\cdot U$, with $\varphi_b := \varphi(\zeta,\cdot)$ and $q := \mathbb{E}\varphi_b^2 > 0$, and the clocks $(t_v)_{v\notin b}$ remain i.i.d. $U[0,1]$ and independent of x. From now on everything is conditional on ζ_B, and we write φ for φ_b.

Notation.

- $S_t := \{v \notin b : t_v \le t\}$ and $t_* := 1-p$. The σ-field $\mathcal{G}_t$ is generated by ζ_B and $\{(v, t_v, x_v) : t_v \le t\}$; for an unrevealed v it records only that $t_v > t$.

- $q_t := \mathbb{E}_U[\varphi^2 \mid x_{S_t}]$, $Y_t := \ln(q_t/q)$ and $\Lambda := \ln(e/q)$. The restricted norm at the end is $Q := q_{t_*} = \mathbb{E}[\varphi_B^2 \mid z_B, z_{F\setminus B}]$.
- $c_v(t) := \widetilde{\mathbb{E}}[x_v \mid x_{S_t}]$ for $v \notin S_t$. Under $\widetilde{\Pr}$, $q_t > 0$ almost surely.
- The hazard is $H(s,u) := \ln((1-s)/(1-u))$ for $0 \le s \le u < 1$; note $H(0,t_*) = \ln(1/p)$.
- The reveal times $s_1 < s_2 < \cdots < s_m$ are the clock values in $(0,t_*]$ (almost surely distinct), with $s_0 := 0$. The k-th revealed coordinate is v_k, and $c_k := c_{v_k}(s_{k-1})$ is its bias just before it is revealed.
- For a threshold $\tau_1 \in (0, 1/(2e^3d^4)]$ put $\mathrm{QV}^{\mathrm{sm}} := \sum_k c_k^2 \mathbf{1}(|c_k| < \tau_1)$, $\mathrm{QV}^{\mathrm{lg}} := \sum_k c_k^2 \mathbf{1}(|c_k| \ge \tau_1)$ and $N^{\mathrm{lg}} := \#\{k : |c_k| \ge \tau_1\}$. Then $\mathrm{QV}^{\mathrm{lg}} \le N^{\mathrm{lg}}$.
- $\kappa_1 := 10e/(\ln(1/\tau_1) - 3 - 4\ln d)$, which is positive because $\ln(1/\tau_1) \ge 3 + \ln 2 + 4\ln d$.

Facts.

(F0) *One step.* Let $S \subseteq b^c$, $v \notin S$, $q_S := \mathbb{E}_U[\varphi^2 \mid x_S]$ and $c := \widetilde{\mathbb{E}}[x_v \mid x_S]$. Then $\mathbb{E}_U[\varphi^2 \mid x_S, x_v] = q_S(1 + c x_v)$ and $\widetilde{\Pr}(x_v = \pm 1 \mid x_S) = (1 \pm c)/2$.

Indeed, $\mathbb{E}_U[\varphi^2 \mid x_S, x_v]$ is a function of $x_v \in \{\pm 1\}$, hence affine in x_v. Its mean is q_S, and its slope is $\mathbb{E}_U[\varphi^2 x_v \mid x_S] = q_S c$ by the definition of c. Then $\widetilde{\Pr}(x_v = 1 \mid x_S) = \mathbb{E}_U[\varphi^2 \mathbf{1}(x_v = 1) \mid x_S]/q_S = (1+c)/2$.

(F1) *Drift.* $\widetilde{\mathbb{E}}[q_{S\cup\{v\}}/q_S \mid x_S] = ((1+c)^2 + (1-c)^2)/2 = 1 + c^2$.

(F2) *Clocks do not bias values.* Conditionally on $\mathcal{G}_{s_{k-1}}$ and on (s_k, v_k), the value x_{v_k} has $\widetilde{\Pr}$-law $(1 \pm c_k)/2$. The reason is that, given $\mathcal{G}_{s_{k-1}}$, the unrevealed clocks are independent of the unrevealed values, and (s_k, v_k) is a function of these clocks; then apply (F0).

(F3) *Hazard.* Given $\mathcal{G}_s$, the unrevealed clocks are i.i.d. uniform on $(s,1]$. So each of them survives to time u with probability $e^{-H(s,u)}$, and in hazard time $\lambda_v := H(s,t_v)$ they are i.i.d. exponential with mean 1.

(F4) *RL1 at every time.* For every t, the restriction φ_{S_t} is bounded by 1, has degree at most d, norm $\mathbb{E}\varphi_{S_t}^2 = q_t$, and size-biased biases $c_v(t)$. Since $\tau_1 \le 1/(2e^3d^4)$, Corollary J.3 with $k = d$ gives

$$\sum_{v\notin S_t,\ |c_v(t)|<\tau_1} c_v(t)^2 \le \max\{\tau_1,\ \kappa_1 \ln(e/q_t)\} = \max\{\tau_1,\ \kappa_1(\Lambda - Y_t)\}.$$

The sum runs over all unrevealed coordinates, alive ones included; it is used only as an upper bound.

Theorem J.14 (Theorem A; [*proved here, audited*])**.** *Assume (H1) and (H2), and condition on ζ_B. For every $p \in (0,1)$, every $\tau_1 \in (0, 1/(2e^3d^4)]$ and every $y > 0$,*

$$\widetilde{\Pr}\Big(Y_{t_*} \ge 2y + (e-1)\ln(1/p)\max\{\tau_1,\ \kappa_1(\Lambda + y)\} + \mathrm{QV}^{\mathrm{lg}}\Big) \le 3e^{-y}.$$

No regularity of φ_B is assumed, and no "no cascade" hypothesis: large-bias reveals are carried by $\mathrm{QV}^{\mathrm{lg}}$.

Proof. Write $Y_{t_*} = \ln Z_m + \mathrm{QV}^{\mathrm{sm}} + \mathrm{QV}^{\mathrm{lg}}$, where $Z_k := (q_{s_k}/q)\exp(-\sum_{i\le k} c_i^2)$; here we use $q_{t_*} = q_{s_m}$.

(i) Z is a supermartingale. By (F1) and (F2), $\mathbb{E}[Z_k/Z_{k-1} \mid \mathcal{G}_{s_{k-1}}, s_k, v_k] = (1 + c_k^2)e^{-c_k^2} \le 1$. Freezing Z after the last reveal, (Z_k) is a nonnegative $\widetilde{\Pr}$-supermartingale with $Z_0 = 1$. By Ville's inequality the event $E_1 := \{\max_k Z_k \ge e^y\}$ has $\widetilde{\Pr}(E_1) \le e^{-y}$.

(ii) Lower excursions are unlikely. Restricted to $\mathcal{G}_{s_k}$, the density of $\widetilde{\Pr} \otimes$ clocks with respect to $U \otimes$ clocks is q_{s_k}/q, because the clocks are independent of x. Let $k^* := \min\{k \le m : Y_{s_k} \le -y\}$, a bounded stopping index ($k^* = \infty$ if there is no such k), and $E_2 := \{k^* \le m\}$. Since

$\{k^* = k\} \in \mathcal{G}_{s_k}$ and $q_{s_k}/q \le e^{-y}$ on it,

$$\widetilde{\Pr}(E_2) = \sum_k \mathbb{E}_U\big[(q_{s_k}/q)\mathbf{1}(k^* = k)\big] \le e^{-y}.$$

Since q_t changes only at reveal times and $Y_0 = 0$, off E_2 we have $Y_t > -y$ for all $t \le t_*$.

(iii) Exponential compensator. Let σ_k be the sum of $c_v(s_{k-1})^2$ over the coordinates v that are unrevealed after $k-1$ reveals and have $|c_v(s_{k-1})| < \tau_1$. Then σ_k is $\mathcal{G}_{s_{k-1}}$-measurable and constant on $[s_{k-1}, s_k)$. Put $A := \sum_{k=1}^{m+1} \sigma_k H(s_{k-1}, s_k \wedge t_*)$ with $s_{m+1} := 1$, which includes the last waiting period up to t_*.

One step. Fix $\mathcal{G}_s$, where s is 0 or a reveal time, with R unrevealed coordinates. Let $a_v := c_v^2\mathbf{1}(|c_v| < \tau_1) \in [0,1]$, $\alpha := \sum_v (e^{a_v} - 1)$ and $\lambda_{t_*} := H(s, t_*)$. By (F3), the first reveal after s is v at hazard time λ with density $e^{-R\lambda}$, for each v. Therefore

$$\mathbb{E}\Big[e^{a_{v^*} - \alpha\lambda^*}\mathbf{1}(\lambda^* \le \lambda_{t_*}) + e^{-\alpha\lambda_{t_*}}\mathbf{1}(\lambda^* > \lambda_{t_*})\Big] = \sum_v e^{a_v}\frac{1 - e^{-(\alpha+R)\lambda_{t_*}}}{\alpha + R} + e^{-(\alpha+R)\lambda_{t_*}} = 1,$$

because $\sum_v e^{a_v} = \alpha + R$; if $R = 0$ the left side is 1 trivially. Since $e^a - 1 \le (e-1)a$ for $a \in [0,1]$, we have $\alpha \le (e-1)\sigma$ with $\sigma := \sum_v a_v$. So the same expectation with $(e-1)\sigma$ in place of α is at most 1.

The marks of the revealed coordinates are $a_{v_k}(s_{k-1}) = c_k^2\mathbf{1}(|c_k| < \tau_1)$, and the process has at most $|b^c| + 1$ steps. Iterating the one-step bound with the tower property therefore gives $\mathbb{E}\exp(\mathrm{QV}^{\mathrm{sm}} - (e-1)A) \le 1$. For a constant a (depending only on ζ_B), let $E_3 := \{\mathrm{QV}^{\mathrm{sm}} \ge (e-1)a + y,\ A \le a\}$. On E_3 we have $\mathrm{QV}^{\mathrm{sm}} - (e-1)A \ge y$, so Markov's inequality gives $\widetilde{\Pr}(E_3) \le e^{-y}$.

Choice of a. Off E_2, $\Lambda - Y_t < \Lambda + y$ for all $t \le t_*$, so (F4) gives $\sigma_k \le \max\{\tau_1, \kappa_1(\Lambda + y)\}$ for every k. By additivity of H, $\sum_k H(s_{k-1}, s_k \wedge t_*) = H(0, t_*) = \ln(1/p)$. Hence, off E_2,

$$A \le \ln(1/p)\max\{\tau_1,\ \kappa_1(\Lambda + y)\} =: a.$$

(iv) Conclusion. Off $E_1 \cup E_2 \cup E_3$ we have $\ln Z_m < y$, $A \le a$ and hence $\mathrm{QV}^{\mathrm{sm}} < (e-1)a + y$, so $Y_{t_*} < 2y + (e-1)a + \mathrm{QV}^{\mathrm{lg}}$. Therefore the event in the theorem is contained in $E_1 \cup E_2 \cup E_3$, which has probability at most $3e^{-y}$. □

One-shot constants. These constants concern the one-shot restriction of Section 9.8; in this subsection only, τ_1 denotes $\theta^{C'/2}$ rather than the threshold of Theorem 9.23. Let $p = 1/d$, $\theta \in (0, (16d^3)^{-2}]$, $C' \ge 4$ and $\tau_1 := \theta^{C'/2}$. Then:

- $\ln(1/\theta) \ge 2\ln 16 + 6\ln d \ge 5.54 + 6\ln d$, so $\ln d \le \ln(1/\theta)/6$.
- $(C'/4)\ln(1/\theta) \ge \ln(1/\theta) \ge 3 + 4\ln d$, so $\ln(1/\tau_1) - 3 - 4\ln d \ge (C'/4)\ln(1/\theta)$ and $\kappa_1 \le 40e/(C'\ln(1/\theta))$.
- $\tau_1 \le \theta^2 \le (16d^3)^{-4} \le 1/(2e^3d^4)$.
- $(e-1)\ln d \cdot \kappa_1 \le (e-1)\,40e/(6C') < 31.2/C'$ and $(e-1)\ln d \cdot \tau_1 \le 0.01$.

Using $\max\{a,b\} \le a + b$, Theorem J.14 gives

$$\widetilde{\Pr}\Big(Y_{t_*} \ge \tfrac{32}{C'}(\Lambda + y) + 2y + 0.01 + \mathrm{QV}^{\mathrm{lg}}\Big) \le 3e^{-y}. \tag{11}$$

Corollaries. In the rest of this subsection assume the one-shot constants with $C' \ge 256$, together with $C'' \ge 16$, $q \le (\theta/d)^{C''}$ and $p_1 := 1/(64ed^2)$ (not the $p_0 = \varepsilon/(16d)$ of Section 9.6). The *spike* is the event $\{Q \ge \theta\}$. Note that $\Lambda \ge 1 + C''\ln(d/\theta) \ge 1 + 16(5.54 + 7\ln d) \ge 89$.

Corollary J.15 (A1: no spike without large-bias quadratic variation; [*proved here, audited*])**.**

$$\widetilde{\Pr}\big(Q \ge \theta \text{ and } \mathrm{QV}^{\mathrm{lg}} < \Lambda/4\big) \le 3(q/e)^{1/8} \le p_1/16.$$

Proof. We have $\ln(1/\theta) \le \ln(1/q)/C'' \le \Lambda/C''$. On $\{Q \ge \theta\}$, therefore, $Y_{t_*} = \ln(Q/q) \ge \ln\theta + \Lambda - 1 \ge (1 - 1/C'')\Lambda - 1 \ge 0.9375\Lambda - 1$. Take $y = \Lambda/8$ in (11). The threshold there is at most $(32/C')(9/8)\Lambda + \Lambda/4 + 0.01 + \mathrm{QV}^{\lg} \le 0.3907\Lambda + 0.01 + \mathrm{QV}^{\lg}$, and on $\{\mathrm{QV}^{\lg} < \Lambda/4\}$ it is less than $0.6407\Lambda + 0.01$. Since $\Lambda \ge 89$, $0.9375\Lambda - 1 > 0.6407\Lambda + 0.01$. So the event lies in the exceptional set of (11), whose probability is at most $3e^{-\Lambda/8} = 3(q/e)^{1/8}$. Finally $3(q/e)^{1/8} \le 3(\theta/d)^{C''/8} \le 3(\theta/d)^2 \le 3(256d^7)^{-2} \le 1/(1024ed^2) = p_1/16$. □

Definition J.16 (β-regularity and the large-jump bound (LJ))**.** We call φ_B *β-regular* if its relative influences satisfy $\mathrm{Inf}_u(\varphi_B) \le \beta\,\mathbb{E}[\varphi_B^2]$ for every $u \notin b$, and we fix

$$\beta := \theta^{2C'}/4.$$

The *large-jump bound* (LJ) at (θ, C', C'') is the statement: for every β-regular φ_B with $q \le (\theta/d)^{C''}$,

$$\widetilde{\Pr}\big(Q \ge \theta \text{ and } N^{\lg} \ge \Lambda/4\big) \le p_1/16.$$

(LJ) is [*open*]. With this value, $2\sqrt{\beta} = \theta^{C'}$ is the bias level at which a relative-influence exploration stops. (LJ) is a hypothesis about the one-shot restriction; as a reduction of the conjecture it is circular, by Remark 9.34. The results below are stated for this value of β only; the variant $\beta = (\theta/d)^{C'}$ is not treated.

Corollary J.17 (A2: spikes need large-bias reveals)**.** *(a)* [*proved here, audited*] $\widetilde{\Pr}\big(Q \ge \theta \text{ and } \mathrm{QV}^{\lg} \le \Lambda/2\big) \le 3(q/e)^{1/8}$.

(b) [*conditional on (LJ)*] *Assume (LJ) at (θ, C', C''). Then every β-regular φ_B with $q \le (\theta/d)^{C''}$ satisfies $\widetilde{\Pr}(Q \ge \theta) \le p_1/8$. With $\theta = \theta_0/2$, for the threshold θ_0 of Remark 9.34, this is a bound on regular spikes for the one-shot restriction, and it inherits the circularity of that remark.*

Proof. (a) As in the proof of Corollary J.15, off the exceptional set of (11) with $y = \Lambda/8$, the event $\{Q \ge \theta\}$ forces $\mathrm{QV}^{\lg} > 0.9375\Lambda - 1 - 0.3907\Lambda - 0.01 \ge 0.5468\Lambda - 1.01$. This exceeds $\Lambda/2$, because $\Lambda \ge 89 > 21.6$.

(b) Since $\mathrm{QV}^{\lg} \le N^{\lg}$, Corollary J.15 and (LJ) give $\widetilde{\Pr}(Q \ge \theta) \le \widetilde{\Pr}(Q \ge \theta,\ \mathrm{QV}^{\lg} < \Lambda/4) + \widetilde{\Pr}(Q \ge \theta,\ N^{\lg} \ge \Lambda/4) \le p_1/16 + p_1/16$. □

Remark J.18 (a triggered form is strictly stronger). Suppose that, for some $C > 0$, $\widetilde{\mathbb{E}}[\mathrm{QV}^{\lg} \mid Q \ge \theta] \le C\ln(d/\theta) + C\Lambda/\ln(1/\tau_1)$, and that $C'' \ge 8C + 16$ and $C' \ge 8C$. Markov's inequality gives $\widetilde{\Pr}(\mathrm{QV}^{\lg} \ge \Lambda/2 \mid Q \ge \theta) \le 2C/C'' + 4C/(C'\ln(1/\theta)) < 0.35$. With Corollary J.17(a) this gives $\widetilde{\Pr}(Q \ge \theta) \le 6(q/e)^{1/8}$, which is far smaller than $p_1/8$ when q is small. The converse fails. The bound $p_1/8$ allows $\widetilde{\Pr}(Q \ge \theta)$ of order $p_1/100$, while by Corollary J.17(a), $\mathrm{QV}^{\lg} > \Lambda/2$ on the spike outside a set of mass $3(q/e)^{1/8} \ll p_1/100$. Then $\widetilde{\mathbb{E}}[\mathrm{QV}^{\lg} \mid Q \ge \theta] \gtrsim \Lambda/2$, which can be far above $C\ln(d/\theta)$. So the triggered form is not equivalent to the regular-spike bound. This is harmless downstream: heavy spikes have $\widetilde{\Pr}(Q \ge \theta) \ge p_1$, and either form excludes them.

Remark J.19 (scope). (1) The factor $\ln(1/p)$ in Theorem J.14 is the hazard integral $\int_0^{1-p} du/(1-u)$. In the one-shot constants it is absorbed by $\kappa_1 = O(1/(C'\ln(1/\theta)))$, because $\ln d \le \ln(1/\theta)/6$. (2) The term $\mathrm{QV}^{\lg}$ cannot be dropped: without it the statement is false. Consider regular spread gates, the average over $M = 1/(2\beta)$ blocks of ANDs of width w. There the chosen block cascades, and [*numerical*] the exact bias after j matches in that block is $2^j/(M - 1 + 2^j)$. Hence $N^{\lg} \approx w - \log_2(M\tau_1)$ and $\mathrm{QV}^{\lg} \approx w - \log_2 M - O(1)$, consistent with (11), since $Y_{t_*} \approx (w - \log_2 M)\ln 2 \le \mathrm{QV}^{\lg}$. (3) The threshold $\tau_1 = \theta^{C'/2}$ lies far above the stopping level $2\sqrt{\beta} = \theta^{C'}$ of a relative-influence exploration, so tail biases of size between these two levels count as small.

J.4. **The old-free reduction.** For $g \in \mathcal{B}_d$, $\theta > 0$ and $L \geq 1$ let $J = J_{\theta/L}(g)$, $h = \mathbb{E}[g \mid x_{J^c}]$ as in Lemma 8.6, and $\Delta(g;\theta,L) = U_{\theta/L}(g) - 2\mathbb{E}_\sigma U_\theta(g_\sigma)$. Let G_σ and K_σ be as in Lemma I.9, and write $I^{\mathrm{new}}(g) = \sum_{j\notin J} \mathrm{Inf}_j(g) \leq d\mathrm{Var}g$.

Lemma J.20 (Lemma OF; [*proved here, audited*])**.** *For every $\theta' > 0$ and $\eta > 0$, with $L' = L\theta'/\theta$,*

$$\Delta(g;\theta,L) \;\leq\; \Delta(h;\theta',L') + 2\eta\,\mathcal{G} + (1+1/\eta)\,(\theta'/\theta)\,I^{\mathrm{new}}(g),$$

where $\mathcal{G} = \mathbb{E}_\sigma[\mathbb{E}_{y'}\mathrm{Var}(G_\sigma) - \mathrm{Var}(h_\sigma)] = \sum_{S\cap J\neq\emptyset} \hat{g}(S)^2(2^{-|S\cap J|} - 2^{-|S|}) \leq \mathrm{Var}g - \mathrm{Var}h$ and y' denotes the values of the fixed old coordinates. Since $\theta'/L' = \theta/L > \mathrm{MaxInf}(h)$, the function h has no old spikes at (θ', L').

Proof. For $S \subseteq A \setminus J$ write $\widehat{G_\sigma}(S) = \widehat{h_\sigma}(S) + r_S(y')$ with $\mathbb{E}_{y'} r_S = 0$; then $\mathbb{E}_{y'} G_\sigma = h_\sigma$ and $\sum_{S\neq\emptyset} \mathbb{E}_{y'} r_S^2 = \mathbb{E}_{y'}\mathrm{Var}(G_\sigma) - \mathrm{Var}(h_\sigma)$. As in the proof of Lemma I.9, $U_\theta(g_\sigma) \geq \sum_{\emptyset\neq S\subseteq A\setminus J,\, S\cap K_\sigma=\emptyset} \widehat{G_\sigma}(S)^2$. Drop the S that meet the spikes $J_{\theta'}(h_\sigma)$, which do not depend on y'. For the remaining S, with $t_S = \mathbf{1}[S \cap K_\sigma \neq \emptyset]$ and $c_S = \widehat{h_\sigma}(S)$ (which does not depend on y'), $\mathbb{E}_{y'}(c_S + r_S)^2 = c_S^2 + \mathbb{E}_{y'} r_S^2$ and $t_S(c_S + r_S)^2 \leq t_S\big((1+1/\eta)c_S^2 + (1+\eta)r_S^2\big)$ give $\mathbb{E}_{y'}[(1-t_S)(c_S+r_S)^2] \geq c_S^2 - (1+1/\eta)c_S^2\mathbb{E}_{y'}t_S - \eta\mathbb{E}_{y'}r_S^2$; summed over S, the first terms give $U_{\theta'}(h_\sigma)$. A set S with $t_S = 1$ that avoids $J_{\theta'}(h_\sigma)$ contains a spike of g_σ whose h_σ-influence is below θ', so $\sum_S c_S^2 t_S \leq \theta'|K_\sigma|$, and $\Pr[j \in K_\sigma] \leq \mathrm{Inf}_j(g)/(2\theta)$. Summing, averaging over σ, and using $U_{\theta/L}(g) = \mathrm{Var}h$ (Lemma 8.6) and $\Delta(h;\theta',L') = \mathrm{Var}h - 2\mathbb{E}_\sigma U_{\theta'}(h_\sigma)$ gives the claim. □

Corollary J.21 (Corollary OF; [*proved here, audited*])**.** *Let $0 < \varepsilon \leq 1$, $\eta = \varepsilon/8$ and $\theta' = \varepsilon^2\theta/(36d)$, so that $L' = \varepsilon^2 L/(36d)$. Then $\Delta(g;\theta,L) \leq \Delta(h;\theta',L') + \frac{\varepsilon}{2}\mathrm{Var}g$. In particular, if the inequality of Definition 8.2 fails for g at some $\theta \leq \mathrm{Var}(g)^2$, then $\mathrm{Var}h > \varepsilon\mathrm{Var}g$, $\mathrm{MaxInf}(h) < \theta'/L'$, $\Delta(h;\theta',L') > \frac{\varepsilon}{2}\mathrm{Var}h$ and $\theta' < \mathrm{Var}(h)^2/(36d)$. Hence the touching transfer at loss L' and parameter $\varepsilon/2$ for all old-free $h \in \mathcal{B}_d$ with the cap $\theta' \leq \mathrm{Var}(h)^2/(36d)$ implies the touching transfer at loss $36dL'/\varepsilon^2 = 9216e^2d^3L'$ (for $\varepsilon = 1/(16ed)$) under cap 2 for all $g \in \mathcal{B}_d$.*

Proof. The error terms are $2\eta\mathcal{G} \leq \frac{\varepsilon}{4}\mathrm{Var}g$ and $(1+8/\varepsilon)\frac{\varepsilon^2}{36d}d\mathrm{Var}g \leq \frac{\varepsilon}{4}\mathrm{Var}g$. If the inequality fails for g, then $\mathrm{Var}h \geq U_{\theta/L}(g) \geq \Delta(g;\theta,L) > \varepsilon\mathrm{Var}g$, and the remaining claims follow from $\theta'/L' = \theta/L$ and $\theta' = \varepsilon^2\theta/(36d) \leq \varepsilon^2\mathrm{Var}(g)^2/(36d) < \mathrm{Var}(h)^2/(36d)$. □

Appendix K. Proofs of the two-stage theorems

This appendix proves Theorem 9.15, the staged form of Theorem 9.14, and Theorem 10.14, Theorem B° in the node form of Section 10.4. Both proofs rerun the proof of Theorem J.14 in another reveal order, started from the state reached at a stage boundary or at a stopping index inside a stage. Proposition K.3 does this once for all such starts. Theorem 9.15 is deduced in Section K.2 and Theorem 10.14 in Section K.3. We use the hazard $H(s,u) = \ln((1-s)/(1-u))$, the facts (F0) and (F1) and the notation of Appendix J.3. Throughout, $d \geq 2$, $\tau_1 \in (0, 1/(2e^3d^4)]$ and $\kappa_1 = 10e/(\ln(1/\tau_1) - 3 - 4\ln d)$. For a function φ and a set D of coordinates, $\varphi_D := \varphi(x_D, \cdot)$ denotes the restriction to the values x_D. In Section K.3 the letter H also denotes a target height; the hazard always appears with two arguments.

K.1. Staged explorations and a restarted Theorem A.

Definition K.1 (staged start state)**.** Let $p \in (0,1)$, $t_* = 1-p$, let $\mathcal{G}_\circ$ be a σ-field and $r \geq 1$. A *staged start state at $\mathcal{G}_\circ$ with r stages* consists of the following $\mathcal{G}_\circ$-measurable data: a set R of unrevealed coordinates; a multilinear φ on $\{-1,1\}^R$ with $|\varphi| \leq 1$, degree at most d and $q := \mathbb{E}_U\varphi^2 > 0$; an ordered partition $R = \mathcal{P}_0 \sqcup \mathcal{P}_1 \sqcup \cdots \sqcup \mathcal{P}_r$, whose parts may be empty; and a start clock $u_\circ \in [0, t_*]$. Conditionally on $\mathcal{G}_\circ$, the following must hold.

(S1) The values x_R have law $(\varphi^2/q)\cdot U$.
(S2) The clocks t_v, $v \in \mathcal{P}_1 \cup \cdots \cup \mathcal{P}_r$, are independent of x_R and of each other. For $v \in \mathcal{P}_1$ the clock t_v is uniform on $(u_\circ, 1]$, and for $v \in \mathcal{P}_j$ with $j \geq 2$ it is uniform on $[0,1]$.
(S3) The exploration runs the stages $j = 1, \dots, r$ in this order. Stage j reveals the coordinates $v \in \mathcal{P}_j$ with $t_v \leq t_*$, together with their values, one at a time in increasing clock order. The coordinates of $\mathcal{P}_0$ are never revealed; they are called *frozen*.

Here $\widetilde{\Pr}$ denotes the underlying law of x, the clocks and the coins; in the applications it is a size-biased law with independent clocks.

The reveals are numbered $k = 1, \dots, m$ in the order in which they occur. Thus $m \leq |R \setminus \mathcal{P}_0|$ is random, and v_k, j_k and $u_k := t_{v_k}$ denote the coordinate, the stage and the clock of the k-th reveal. Put $D_k = \{v_1, \dots, v_k\}$, $q_k := \mathbb{E}_U[\varphi^2 \mid x_{D_k}]$ (so $q_0 = q$), $Q := q_m$, and let $c_k := \mathbb{E}_U[\varphi^2 x_{v_k} \mid x_{D_{k-1}}]/q_{k-1}$ be the bias of the k-th reveal just before it. Define $\mathrm{QV}^{\mathrm{sm}}$, $\mathrm{QV}^{\mathrm{lg}}$ and N^{lg} along this sequence as in Appendix J.3.

Time is the pair (stage, clock), ordered lexicographically. An *epoch* is the start $(1, u_\circ)$, a reveal (j_k, u_k), or the end (j, t_*) of a stage $j < r$. The end of stage j is also the start of stage $j+1$. The σ-field $\mathcal{F}_\varepsilon$ at an epoch ε is generated by $\mathcal{G}_\circ$, the stage of ε, and the coordinate, clock and value of each reveal up to and including ε. Write $\mathcal{G}_k$ for $\mathcal{F}_\varepsilon$ at the k-th reveal, with $\mathcal{G}_0 := \mathcal{G}_\circ$; for $k > m$, $\mathcal{G}_k$ is the terminal σ-field, which also records that the exploration has ended. A *stopping index* is a random $\tau \in \{0, 1, \dots\} \cup \{\infty\}$ with $\{\tau = k\} \in \mathcal{G}_k$ for every k. The information in $\mathcal{F}_\varepsilon$ about the clocks of unrevealed coordinates is implicit. If ε lies in stage j at clock u, then the unrevealed coordinates of $\mathcal{P}_j$ have clocks $> u$, those of the earlier stages have clocks $> t_*$, and nothing is known about the later stages.

Lemma K.2 (strong Markov property of staged explorations; [*proved here, audited*])**.** *Let $\mathcal{G}_\circ$ be a staged start state with r stages, and let ε be an epoch. Write j for its stage and u for its clock, where the end of stage $j-1$ counts as stage j with clock $u = 0$. Let D be the set revealed by ε. Then $\mathcal{F}_\varepsilon$ is a staged start state with $r - j + 1$ stages and the following data: the unrevealed set $R \setminus D$; the function φ_D and height $\mathbb{E}_U[\varphi^2 \mid x_D]$; the parts $\mathcal{P}_j \setminus D, \mathcal{P}_{j+1}, \dots, \mathcal{P}_r$; the start clock u; and the frozen part*

$$\mathcal{P}_0 \cup \big((\mathcal{P}_1 \cup \cdots \cup \mathcal{P}_{j-1}) \setminus D\big).$$

Its exploration is the remaining part of the original exploration. In particular this holds at the k-th reveal on $\{k \leq m\}$, given $\mathcal{G}_k$. If τ is a stopping index, then on $\{\tau = k \leq m\}$ the conditional law given $\mathcal{G}_\tau$ coincides with the conditional law given $\mathcal{G}_k$.

Proof. Condition on $\mathcal{G}_\circ$ throughout. Let $\mathcal{I}_\varepsilon$ be the clock part of $\mathcal{F}_\varepsilon$: the stage of ε and the coordinates and clocks of the reveals up to ε. Then $\mathcal{F}_\varepsilon = \sigma(\mathcal{G}_\circ, \mathcal{I}_\varepsilon, x_D)$, and D is $\mathcal{I}_\varepsilon$-measurable. Since the reveal order is determined by the partition and the clocks, $\mathcal{I}_\varepsilon$ is a function of the clocks, which are independent of x_R by (S2).

Values. Given $(\mathcal{I}_\varepsilon, x_D)$, the values $x_{R\setminus D}$ have the conditional law of $(\varphi^2/q)\cdot U$ given x_D, with the set D held fixed. By Bayes' rule this law is $(\varphi_D^2/\mathbb{E}_U[\varphi^2 \mid x_D])\cdot U$. The function φ_D is multilinear, bounded by 1 and of degree at most d, and its norm $\mathbb{E}_U[\varphi^2 \mid x_D]$ is positive almost surely, because $\mathbb{E}_U[\varphi^2 \mid x_D]/q$ is the density of $\widetilde{\Pr}$ with respect to U on $\sigma(x_D)$.

Clocks. Fix a value of $\mathcal{I}_\varepsilon$. The event that $\mathcal{I}_\varepsilon$ takes it is, apart from the values of the revealed clocks themselves, an intersection of conditions on single clocks: $t_w > t_*$ for the unrevealed w of the stages before j; $t_w > u$ for the unrevealed $w \in \mathcal{P}_j$; and no condition on the clocks of $\mathcal{P}_{j+1}, \dots, \mathcal{P}_r$. The clocks are independent, so conditionally on this event they remain independent. The unrevealed clocks of $\mathcal{P}_j$ are uniform on $(u, 1]$; for $j = 1$ this uses that they were uniform on $(u_\circ, 1]$ and that $u \geq u_\circ$. Those of the later stages are uniform on $[0,1]$. The unrevealed coordinates of the earlier stages have clocks $> t_*$, so they are never revealed. This is

the order-statistics fact that, given the smallest clocks and their owners, the other i.i.d. uniform clocks are i.i.d. uniform above the last revealed one. Since the clocks are independent of x_R, conditioning in addition on x_D does not change their law, and they remain independent of $x_{R\setminus D}$. This proves (S1) and (S2) for $\mathcal{F}_\varepsilon$. After ε the exploration reveals, in clock order, the coordinates of $\mathcal{P}_j \setminus D$ with clocks in $(u, t_*]$, and then runs the stages $j+1, \dots, r$; this is (S3).

Stopping indices. If $A \in \mathcal{G}_\tau$, then $A \cap \{\tau = k\} \in \mathcal{G}_k$. Hence $\mathbb{E}[Y \mid \mathcal{G}_\tau] = \sum_k \mathbf{1}(\tau = k)\mathbb{E}[Y \mid \mathcal{G}_k]$ almost surely for every bounded Y. □

Proposition K.3 (Theorem A from a staged start state; [*proved here, audited*])**.** *Let $\mathcal{G}_\circ$ be a staged start state with r stages, start clock $u_\circ$ and height q. Put $\Lambda_\circ := \ln(e/q)$ and*

$$\ell_\circ := H(u_\circ, t_*) + (r-1)\ln(1/p) \;\le\; r\ln(1/p).$$

For every $y > 0$ there is an event X with $\widetilde{\Pr}(X \mid \mathcal{G}_\circ) \le 3e^{-y}$ off which:

(a) $q_k > q\,e^{-y}$ *for all* $0 \le k \le m$;

(b) $\ln(Q/q) < 2y + (e-1)\,\ell_\circ \max\{\tau_1,\ \kappa_1(\Lambda_\circ + y)\} + \mathrm{QV}^{\mathrm{lg}}$.

Proof. Write Pr for $\widetilde{\Pr}(\cdot \mid \mathcal{G}_\circ)$, the joint law of x_R and the clocks. The Pr-law of the clocks of $\mathcal{P}_0$ plays no role, since these coordinates are never revealed. Put $Z_k := (q_k/q)\exp(-\sum_{i\le k} c_i^2)$ for $k \le m$ and $Z_k := Z_m$ for $k > m$. Since $\ln(Q/q) = \sum_{k\le m}\ln(q_k/q_{k-1})$ and $\sum_{k\le m} c_k^2 = \mathrm{QV}^{\mathrm{sm}} + \mathrm{QV}^{\mathrm{lg}}$,

$$\ln(Q/q) = \ln Z_m + \mathrm{QV}^{\mathrm{sm}} + \mathrm{QV}^{\mathrm{lg}}.$$

(i) Z is a supermartingale. Fix $k \ge 1$. By Lemma K.2 at the $(k-1)$-th reveal, the following holds given $\mathcal{G}_{k-1}$. The unrevealed values have law $(\varphi^2_{D_{k-1}}/q_{k-1})\cdot U$ and are independent of the clocks. Whether a k-th reveal occurs, and if so its stage, clock and coordinate, are functions of the clocks. Hence, given $\mathcal{G}_{k-1}$ and these data, on $\{k \le m\}$ the value x_{v_k} has law $(1 \pm c_k)/2$ and $q_k/q_{k-1} = 1 + c_k x_{v_k}$, by (F0) applied to $\varphi_{D_{k-1}}$. By (F1) the conditional mean of Z_k/Z_{k-1} is $(1+c_k^2)e^{-c_k^2} \le 1$, and on $\{k > m\}$ the ratio is 1. So $(Z_k)_{k\ge 0}$ is a nonnegative $(\mathcal{G}_k)$-supermartingale with $Z_0 = 1$. It is constant after $m \le |R|$ steps. By Ville's inequality, $E_1 := \{\max_k Z_k \ge e^y\}$ has $\Pr(E_1) \le e^{-y}$.

(ii) Lower excursions. Let $\Pr^0$ be the law under which, given $\mathcal{G}_\circ$, x_R is uniform and the clocks are independent of x_R with their Pr-law. Then $d\Pr/d\Pr^0 = \varphi^2/q$. Every $\mathcal{G}_k$ is generated by $\mathcal{G}_\circ$, clock information and x_{D_k}, with D_k determined by the clock information. Under $\Pr^0$, x_R is uniform and independent of the clocks, so $\mathbb{E}_{\Pr^0}[\varphi^2 \mid \mathcal{G}_k] = \mathbb{E}_U[\varphi^2 \mid x_{D_k}] = q_k$. Thus the density of Pr on $\mathcal{G}_k$ is q_k/q. Let $k^* := \min\{k \le m : q_k \le qe^{-y}\}$, with $k^* = \infty$ if there is no such k, and put $E_2 := \{k^* \le m\}$. Since $\{k^* = k\} \in \mathcal{G}_k$ and $q_k \le qe^{-y}$ on it,

$$\Pr(E_2) = \sum_k \mathbb{E}_{\Pr^0}\big[(q_k/q)\mathbf{1}(k^* = k)\big] \le e^{-y}.$$

Off E_2, (a) holds. The height changes only at reveals, so off E_2 the height at every epoch exceeds qe^{-y}.

(iii) Exponential compensator. The epochs cut the run into waiting periods. A period starts at an epoch ε in stage j at clock u. It ends at the next reveal of stage j, if that reveal has clock $\le t_*$, and otherwise at the end (j, t_*) of stage j. So there are exactly $m + r$ periods, at most $|R| + r$. For the period starting at ε, let R_ε be its at-risk set, the unrevealed coordinates of $\mathcal{P}_j$. For $w \in R_\varepsilon$ let $c_w(\varepsilon)$ be the bias of w under $(\varphi^2_D/\mathbb{E}_U[\varphi^2 \mid x_D])\cdot U$, where D is the set revealed by ε. Put

$$a_w := c_w(\varepsilon)^2\mathbf{1}(|c_w(\varepsilon)| < \tau_1) \in [0,1], \qquad \sigma_\varepsilon := \sum_{w\in R_\varepsilon} a_w, \qquad \alpha_\varepsilon := \sum_{w\in R_\varepsilon}(e^{a_w} - 1),$$

all $\mathcal{F}_\varepsilon$-measurable. Let $\lambda_* := H(u, t_*)$, and let $\lambda_\varepsilon := H(u, u')$ be the hazard length of the period, where u' is its end clock. The *mark* of the period is a_w if it ends with the reveal of w, and 0 if it ends at the end of the stage.

By Lemma K.2, given $\mathcal{F}_\varepsilon$ the clocks of R_ε are i.i.d. uniform on $(u, 1]$. So their hazard times $\lambda_w := H(u, t_w)$ are i.i.d. exponential with mean 1. With $n_\varepsilon := |R_\varepsilon| \geq 1$, the first of them is w with hazard time in $d\lambda$ with probability $e^{-n_\varepsilon \lambda} d\lambda$, for each w. Hence, since $\sum_w e^{a_w} = \alpha_\varepsilon + n_\varepsilon$,

$$\mathbb{E}\big[e^{\text{mark}-\alpha_\varepsilon\lambda_\varepsilon} \;\big|\; \mathcal{F}_\varepsilon\big] = \sum_{w\in R_\varepsilon} e^{a_w} \int_0^{\lambda_*} e^{-(\alpha_\varepsilon+n_\varepsilon)\lambda}\, d\lambda + e^{-(\alpha_\varepsilon+n_\varepsilon)\lambda_*} = 1.$$

If $R_\varepsilon = \emptyset$, the period ends at the end of the stage with mark 0 and $\alpha_\varepsilon = 0$, and the left side is 1 trivially.

The period that ends with the k-th reveal starts at an epoch whose revealed set is D_{k-1}, because no value is revealed at the end of a stage. So its mark is $c_k^2 \mathbf{1}(|c_k| < \tau_1)$, and the sum of all marks is $\mathrm{QV}^{\mathrm{sm}}$. The factor $e^{\text{mark}-\alpha_\varepsilon\lambda_\varepsilon}$ of a period is measurable at the next epoch. There are at most $|R| + r$ periods, so iterating the one-step identity with the tower property gives

$$\mathbb{E}\exp\Big(\mathrm{QV}^{\mathrm{sm}} - \sum_\varepsilon \alpha_\varepsilon\lambda_\varepsilon\Big) = 1.$$

Since $e^a - 1 \leq (e-1)a$ for $a \in [0, 1]$, we have $\alpha_\varepsilon \leq (e-1)\sigma_\varepsilon$. Hence $\mathbb{E}\exp(\mathrm{QV}^{\mathrm{sm}} - (e-1)A) \leq 1$ with $A := \sum_\varepsilon \sigma_\varepsilon\lambda_\varepsilon$. For a $\mathcal{G}_\circ$-measurable a, let $E_3 := \{\mathrm{QV}^{\mathrm{sm}} \geq (e-1)a + y,\ A \leq a\}$. On E_3 we have $\mathrm{QV}^{\mathrm{sm}} - (e-1)A \geq y$, so Markov's inequality gives $\Pr(E_3) \leq e^{-y}$.

Choice of a. At an epoch ε with revealed set D, the restriction φ_D is bounded by 1 and has degree at most d. Its norm is $q_\varepsilon := \mathbb{E}_U[\varphi^2 \mid x_D] > 0$, and its size-biased biases are the $c_w(\varepsilon)$. Corollary J.3 with $k = d$ therefore gives

$$\sum_{w\in R\setminus D,\ |c_w(\varepsilon)|<\tau_1} c_w(\varepsilon)^2 \;\leq\; \max\{\tau_1,\ \kappa_1 \ln(e/q_\varepsilon)\}.$$

This is (F4) at ε. Its sum runs over all unrevealed coordinates: the at-risk set R_ε, the coordinates of the later stages and the frozen ones. The sum σ_ε is over the subset R_ε, so off E_2, where $\ln(e/q_\varepsilon) < \Lambda_\circ + y$,

$$\sigma_\varepsilon \leq \max\{\tau_1,\ \kappa_1(\Lambda_\circ + y)\} \qquad \text{for every period.}$$

By additivity of H, the hazard lengths of the periods of stage 1 add up to $H(u_\circ, t_*)$, and those of every later stage to $H(0, t_*) = \ln(1/p)$. Hence off E_2,

$$A \leq \ell_\circ \max\{\tau_1,\ \kappa_1(\Lambda_\circ + y)\} =: a,$$

and a is $\mathcal{G}_\circ$-measurable. Also $H(u_\circ, t_*) \leq H(0, t_*)$, so $\ell_\circ \leq r \ln(1/p)$.

(iv) Conclusion. Let $X := E_1 \cup E_2 \cup E_3$, so $\Pr(X) \leq 3e^{-y}$. Off X we have $\ln Z_m < y$ and $A \leq a$, hence $\mathrm{QV}^{\mathrm{sm}} < (e-1)a + y$. So $\ln(Q/q) < 2y + (e-1)a + \mathrm{QV}^{\mathrm{lg}}$, which is (b); (a) holds off E_2. □

For $r = 1$, $u_\circ = 0$ and $\mathcal{P}_0 = \emptyset$, Proposition K.3(b) is Theorem J.14, and the proof above is its proof. The only changes are the at-risk sets and the bookkeeping of the hazard over several stages.

K.2. **The staged form of Theorem A.** We restate Theorem 9.15.

Theorem 9.15 (staged form)**.** *Suppose the revealed set is revealed in s stages, the stage of each coordinate being chosen independently of x and of the clocks, and in clock order within each stage.* (a) *Theorem 9.14 holds with $(e-1)\ln(1/p)$ replaced by $(e-1)\,s\ln(1/p)$: each stage adds one hazard $\ln(1/p)$.* (b) *The same holds conditionally on the revealed data at any stage boundary,*

with one hazard per remaining stage and q, Λ taken at that boundary; here alive coordinates that are never revealed (frozen, with hazard 0) are allowed.

Hypotheses made precise. Assume (H1) and (H2) of Appendix J.3, except for the order in which $F \setminus B$ is revealed, which is replaced by the following.

(H2s) *Staged oblivious exploration.* Given the stopping data $\zeta_B = (B = b,\ x_b = \zeta)$, an ordered partition $b^c = \mathcal{P}_0 \sqcup \mathcal{P}_1 \sqcup \cdots \sqcup \mathcal{P}_s$ ($s \geq 1$; parts may be empty) is a function of ζ_B and of coins that are independent of x and of the clocks. In particular it does not depend on x_{b^c} or on the clocks. Stage $i = 1, \dots, s$ reveals $F \cap \mathcal{P}_i$ in increasing clock order, and the stages run in the order $1, \dots, s$. The coordinates of $\mathcal{P}_0$ are never revealed.

In the statement of Theorem 9.15(a), $\mathcal{P}_0 = \emptyset$. Then the final revealed set is F, as in Theorem 9.14, so $Q = \mathbb{E}[\varphi_B^2 \mid z_B, z_{F\setminus B}]$ is the same random variable; only the reveal path, and with it $\mathrm{QV}^{\mathrm{lg}}$, changes. The partition, and s, may depend on the stopping data ζ_B and on the coins. As in Appendix J.3, write φ for φ_b, $q = \mathbb{E}\varphi_b^2$ and $\Lambda = \ln(e/q)$. Write $\mathrm{QV}^{\mathrm{lg}}$ for the large-bias quadratic variation of the staged reveal sequence. Let $\mathcal{G}^{(i-1)}$ be the σ-field at the stage boundary before stage i, generated by ζ_B, the coins, and the coordinates, clocks and values revealed in stages $1, \dots, i-1$. Then $q^{(i-1)}$ denotes the height there, $\Lambda^{(i-1)} = \ln(e/q^{(i-1)})$, and $\mathrm{QV}^{\mathrm{lg}}_{\geq i}$ is the large-bias quadratic variation of the reveals of stages $i, \dots, s$. The two parts of the theorem then read:

$$\widetilde{\Pr}\Big(\ln \frac{Q}{q} \geq 2y + (e-1)\, s \ln \tfrac{1}{p} \max\{\tau_1,\ \kappa_1(\Lambda + y)\} + \mathrm{QV}^{\mathrm{lg}} \ \Big|\ \zeta_B \Big) \leq 3e^{-y}, \tag{12}$$

$$\widetilde{\Pr}\Big(\ln \frac{Q}{q^{(i-1)}} \geq 2y + (e-1)(s-i+1) \ln \tfrac{1}{p} \max\{\tau_1,\ \kappa_1(\Lambda^{(i-1)} + y)\} + \mathrm{QV}^{\mathrm{lg}}_{\geq i} \ \Big|\ \mathcal{G}^{(i-1)} \Big) \leq 3e^{-y} \tag{13}$$

almost surely, for every $p \in (0,1)$, $\tau_1 \in (0, 1/(2e^3 d^4)]$, $y > 0$ and $i \in \{1, \dots, s\}$. Here $\mathcal{P}_0$ is allowed in both parts; when $\mathcal{P}_0 \neq \emptyset$, Q denotes the final height of the staged exploration, the conditional second moment of φ_B given everything it reveals, and the coordinates of $\mathcal{P}_0$ may lie in F, since they are never revealed. The unconditional form of (12) for $B = \emptyset$ is Theorem 9.14 with s hazards.

Proof of Theorem 9.15. (a) Condition on ζ_B and on the coins. By the consequence of (H2) in Appendix J.3, given ζ_B the values x_{b^c} have law $(\varphi_b^2/q) \cdot U$, and the clocks $(t_v)_{v \notin b}$ are i.i.d. $U[0,1]$ and independent of x. The event $\{B = b,\ x_b = \zeta\}$ is $\{x_b = \zeta\}$ intersected with an event of the coins. The coins are independent of x and of the clocks. So the same holds given ζ_B and all the coins. The partition is a function of ζ_B and the coins. Hence $\mathcal{G}_\circ := \sigma(\zeta_B, \text{coins})$ is a staged start state with $r = s$ stages $\mathcal{P}_1, \dots, \mathcal{P}_s$, frozen part $\mathcal{P}_0$, start clock $u_\circ = 0$, function φ_b and height q; its exploration is the staged exploration (H2s). Proposition K.3(b) with $\ell_\circ = H(0, t_*) + (s-1)\ln(1/p) = s \ln(1/p)$ gives (12) conditionally on ζ_B and the coins. Integrating over the coins gives (12). Note that s may depend on the coins, since each conditional bound has its own s.

(b) Fix i; it may depend on ζ_B and the coins. The stage boundary before stage i is the epoch "end of stage $i-1$" of this start state (the start if $i = 1$), and its σ-field is $\mathcal{G}^{(i-1)}$. By Lemma K.2, $\mathcal{G}^{(i-1)}$ is a staged start state with $r = s - i + 1$ stages $\mathcal{P}_i, \dots, \mathcal{P}_s$, start clock 0 and height $q^{(i-1)}$. Its frozen part is $\mathcal{P}_0 \cup \{v \in \mathcal{P}_1 \cup \cdots \cup \mathcal{P}_{i-1} : t_v > t_*\}$. The frozen coordinates are alive and never revealed, and they lie in no at-risk set. Proposition K.3 allows them, because its RL1 bound (F4) sums over all unrevealed coordinates; this is the point of allowing a part $\mathcal{P}_0$ in (H2s). The reveals of the restarted exploration are those of stages $i, \dots, s$, so their large-bias quadratic variation is $\mathrm{QV}^{\mathrm{lg}}_{\geq i}$. Proposition K.3(b) with $\ell_\circ = (s-i+1)\ln(1/p)$ gives (13). $\square$

Remark K.4 (what the hypotheses exclude, and reading x only on F). The proof uses the stage assignment only through Lemma K.2. That lemma needs the reveal order to be a function of the clocks and of $\mathcal{G}_\circ$, with the clocks independent of the unrevealed values. A stage assignment that depends on x_{b^c}, or on the clocks, breaks this, and the theorem does not cover point-dependent sets. Every σ-field of a staged exploration is contained in $\sigma(\zeta_B, \text{coins}, \text{clocks}, x_{F\setminus B}) \subseteq \sigma(x_F, \text{clocks}, \text{coins})$, because $B \subseteq F$. So a staged exploration reads x only on F. This is the hypothesis of Lemma 9.27, which is false without it, so that lemma applies to the σ-fields of every staged exploration. Lemma 10.13 needs in addition that the terminal σ-field contain $\sigma(F, x_F)$; this holds when $\mathcal{P}_0 \cap F = \emptyset$, as in the node explorations of Section 10.4, to which it therefore applies.

K.3. **Theorem B°.** We recall the definitions of Section 10.4, written for $(\psi, \mathcal{H})$ in place of $(\psi_c, \mathcal{H}_c)$; the numbers $0 < h < H \le 1$ and $y > 0$ are free parameters there and here. Under $\widetilde{\Pr} = \widetilde{\Pr}_\psi$, x has density ψ^2/w with respect to U, the clocks t_v are i.i.d. $U[0,1]$ and independent of x, and $F = \{v : t_v \le \frac{1}{2}\}$. The exploration $\mathcal{G}$ has two stages.

- Stage 1 reveals $\mathcal{H} \cap F$, with its values, at once.
- Stage 2 reveals $F \setminus \mathcal{H}$ in increasing clock order, at reveal indices $k = 1, \dots, m$.
- The coordinates of $\mathcal{H} \setminus F$ are frozen: alive, never revealed, hazard 0.

Let $S_k = (\mathcal{H} \cap F) \cup \{v_1, \dots, v_k\}$ be the revealed set after k stage-2 reveals, $q^{(k)} = \mathbb{E}_U[\psi^2 \mid x_{S_k}]$, and let $c_k = \mathbb{E}_U[\psi^2 x_{v_k} \mid x_{S_{k-1}}]/q^{(k-1)}$ be the bias of the k-th stage-2 reveal just before it. The reveal is *large* if $|c_k| \ge \tau_1$. Let $\mathcal{G}_k$ be the σ-field of the clocks and values revealed up to and including the k-th stage-2 value; for $k > m$ it is the terminal σ-field. Let $\mathcal{G}^+_{k-1}$ be the σ-field of the clocks and values revealed before the k-th stage-2 value, together with the identity and clock of the k-th revealed coordinate (or the information that there is none). Then

$$\text{cap} = \{q^{(0)} \ge h\}, \qquad \zeta = \min\{k \ge 0 : q^{(k)} \ge h\},$$

$$E_v = \{q^{(0)} < h,\ \zeta \le m,\ \exists k > \zeta :\ v_k = v,\ |c_k| \ge \tau_1,\ |c_j| < \tau_1\ (\zeta < j < k),\ q^{(k-1)} \ge he^{-y}\},$$

with $\zeta = \infty$ if no $k \le m$ qualifies. Also $\beta_1^\circ = \tau_1^2/4$ (Section 10.2), $\gamma_1 = (e-1)\ln 2$, and $d_v\psi = (\psi(x^{v\to 1}) - \psi(x^{v\to -1}))/2$, so that $\psi = \mathbb{E}_{x_v}\psi + x_v\, d_v\psi$ and $\|d_v\psi\|_2^2 = \text{Inf}_v(\psi)$. We restate Theorem 10.14.

Theorem 10.14 (Theorem B°, node form)**.** *Let $\psi\colon Q_n \to [-1,1]$ be multilinear of degree at most d with $w = \|\psi\|_2^2 > 0$, let $\mathcal{H}$ be a fixed set, and let $\mathcal{G}$, $q^{(k)}$, ζ, cap and E_v be as above with $(\psi, \mathcal{H})$ in place of $(\psi_c, \mathcal{H}_c)$. Let $\tau_1 \in (0, 1/(2e^3d^4)]$, κ_1 as in Theorem 9.14, $0 < h < H \le 1$ and $y > 0$ with $\ln(H/(2h)) \ge 2y + \gamma_1 \max\{\tau_1, \kappa_1(\ln(e/h) + y)\}$. Write $\widetilde{\Pr}$ for $\widetilde{\Pr}_\psi$ and $\widetilde{\Pr}_v$ for $\widetilde{\Pr}_{d_v\psi}$, with the same clocks, and $Q = \mathbb{E}[\psi^2 \mid x_F]$.*

(a) $\widetilde{\Pr}\big(Q \ge H,\ q^{(0)} < h,\ \text{no } E_v \text{ occurs}\big) \le 3e^{-y}$.

(b) $E_v \cap \{v_k = v\} \in \mathcal{G}^+_{k-1}$, *and on this event* $\rho_v(S_{k-1}) := \mathbb{E}_U[(d_v\psi)^2 \mid x_{S_{k-1}}]/q^{(k-1)} \ge c_k^2/4 \ge \beta_1^\circ$ *and* $\mathbb{E}_U[(d_v\psi)^2 \mid x_{S_{k-1}}] \ge \beta_1^\circ he^{-y}$.

(c) $\widetilde{\Pr}(E_v) \le (\rho_v/\beta_1^\circ)\,\widetilde{\Pr}_v(E_v)$, *where* $\rho_v = \text{Inf}_v(\psi)/w$.

Proof of Theorem 10.14. Step 1: the stage boundary is a staged start state. Here $Q = \mathbb{E}_U[\psi^2 \mid x_F] = q^{(m)}$, because $S_m = F$. The set $S_0 = \mathcal{H} \cap F$ is a function of the clocks of $\mathcal{H}$, and $\mathcal{G}_0$ is generated by S_0, the clocks of S_0 and x_{S_0}. Under $\widetilde{\Pr}$, x is independent of the clocks. Hence, given $\mathcal{G}_0$, the values $x_{S_0^c}$ have law $(\psi^2_{S_0}/q^{(0)}) \cdot U$ by Bayes' rule, and $q^{(0)} > 0$ almost surely. The clocks of $\mathcal{H}^c$ are independent of x and of the clocks of $\mathcal{H}$, so given $\mathcal{G}_0$ they are i.i.d. $U[0,1]$ and independent of x. The coordinates of $\mathcal{H} \setminus F$ are never revealed. So $\mathcal{G}_0$ is a staged start state (Definition K.1) with the following data: $p = \frac{1}{2}$, $t_* = \frac{1}{2}$; one stage $\mathcal{P}_1 = \mathcal{H}^c$; frozen part

$\mathcal{P}_0 = \mathcal{H} \setminus F$; start clock 0; function ψ_{S_0}, of degree at most d; and height $q^{(0)}$. Its reveals, heights, biases and σ-fields are the stage-2 quantities v_k, $q^{(k)}$, c_k and $\mathcal{G}_k$. That stage 1 is revealed at once rather than in clock order does not matter, because only the σ-fields from the stage boundary on are used.

Step 2: the proof of Theorem A, run from the state at ζ. The index ζ is a $(\mathcal{G}_k)$-stopping index: $\{\zeta = k\} = \{q^{(j)} < h \text{ for } j < k,\ q^{(k)} \ge h,\ k \le m\} \in \mathcal{G}_k$. Since ζ depends on the stage-2 clocks and values, Theorem 9.14 is not quoted at ζ; instead, Proposition K.3 reruns its proof from $\mathcal{G}_\zeta$, with the inputs supplied by Lemma K.2. Fix k. By Lemma K.2 applied to the start state $\mathcal{G}_0$, on $\{k \le m\}$ the σ-field $\mathcal{G}_k$ is a staged start state with the following data: one stage, consisting of the unrevealed coordinates of $\mathcal{H}^c$; start clock $u_k = t_{v_k} \le \frac{1}{2}$ ($u_0 := 0$); height $q^{(k)}$; and frozen part $\mathcal{H} \setminus F$. Its reveals are the stage-2 reveals $j > k$, and its biases are the c_j. Let X_k be the event of Proposition K.3 for this start state, so $\widetilde{\Pr}(X_k \mid \mathcal{G}_k) \le 3e^{-y}$ on $\{k \le m\}$, and put $X := \bigcup_k (\{\zeta = k\} \cap X_k)$. Since $\{\zeta = k\} \in \mathcal{G}_k$,

$$\widetilde{\Pr}(X) = \sum_k \mathbb{E}\big[\mathbf{1}(\zeta = k)\, \widetilde{\Pr}(X_k \mid \mathcal{G}_k)\big] \le 3e^{-y}.$$

On $\{\zeta = k\}$ we have $q^{(k)} \ge h$, hence $\ln(e/q^{(k)}) \le \ln(e/h)$. The hazard satisfies $\ell_\circ = H(u_k, \frac{1}{2}) = \ln(2(1-u_k)) \le \ln 2$. So off X, on $\{\zeta \le m\}$, Proposition K.3 gives

$$\ln \frac{Q}{q^{(\zeta)}} < 2y + \gamma_1 \max\{\tau_1,\ \kappa_1(\ln(e/h) + y)\} + \mathrm{QV}^{\mathrm{lg}}_{>\zeta}, \tag{14}$$

$$q^{(j)} > q^{(\zeta)} e^{-y} \ge h e^{-y} \qquad (\zeta \le j \le m), \tag{15}$$

where $\mathrm{QV}^{\mathrm{lg}}_{>\zeta} = \sum_{j>\zeta} c_j^2 \mathbf{1}(|c_j| \ge \tau_1)$. Only the hazard after ζ, at most $\ln 2$, is paid. The RL1 bound in the proposition includes the frozen coordinates $\mathcal{H} \setminus F$.

Step 3: the dichotomy and (a). Let $G := \{Q \ge H,\ q^{(0)} < h\}$. On G, $q^{(m)} = Q \ge H > h$, so $\zeta \le m$, and $\zeta \ge 1$ because $q^{(0)} < h$. By minimality $q^{(\zeta-1)} < h$. By (F0), $q^{(\zeta)} = q^{(\zeta-1)}(1 + c_\zeta x_{v_\zeta}) \le 2q^{(\zeta-1)} < 2h$. Hence, by the hypothesis on (h, H, y),

$$\ln \frac{Q}{q^{(\zeta)}} > \ln \frac{H}{2h} \ge 2y + \gamma_1 \max\{\tau_1,\ \kappa_1(\ln(e/h) + y)\}.$$

Off X, (14) then forces $\mathrm{QV}^{\mathrm{lg}}_{>\zeta} > 0$, so some stage-2 reveal strictly after ζ is large. Let k be the first one. Then $|c_j| < \tau_1$ for $\zeta < j < k$. Also $q^{(k-1)} \ge h e^{-y}$: if $k - 1 = \zeta$ this is $q^{(\zeta)} \ge h$, and if $k - 1 > \zeta$ it is (15). So E_{v_k} occurs. Therefore $G \cap \{\text{no } E_v \text{ occurs}\} \subseteq X$, and (a) follows from $\widetilde{\Pr}(X) \le 3e^{-y}$. The three exceptional events of the restarted run are each counted once. In particular the lower excursion is counted once, and no stage-1 term arises, since the search for ζ starts at the stage boundary.

Step 4: (b). Fix v and k. Since v is revealed at most once, on $\{v_k = v\}$ the index in the definition of E_v must be this k, so

$$E_v \cap \{v_k = v\} = \{q^{(0)} < h\} \cap \{\zeta \le k-1\} \cap \bigcap_{\zeta < j < k} \{|c_j| < \tau_1\} \\ \cap \{k \le m,\ v_k = v,\ |c_k| \ge \tau_1\} \cap \{q^{(k-1)} \ge h e^{-y}\}.$$

We check that each set is in $\mathcal{G}^+_{k-1}$.

- $q^{(0)}$ and $q^{(k-1)}$ are functions of x_{S_0}, $x_{S_{k-1}}$ and of these sets, so they are $\mathcal{G}_{k-1}$-measurable.
- $\{\zeta \le k-1\} \in \mathcal{G}_{k-1}$, because ζ is a stopping index.
- For $j < k$, c_j is a function of v_j and $x_{S_{j-1}}$, so it is $\mathcal{G}_{k-1}$-measurable.

- $\{k \le m\}$, v_k and c_k are $\mathcal{G}^+_{k-1}$-measurable, since c_k is a function of v_k and $x_{S_{k-1}}$.

The value x_v does not enter. Hence $E_v \cap \{v_k = v\} \in \mathcal{G}^+_{k-1}$. This uses that E_v looks only at reveals strictly after ζ. The crossing reveal $k = \zeta$ is decided by $q^{(\zeta)} \ge h$, which depends on x_{v_ζ}. The same description shows that the events E_v are pairwise disjoint: on E_v, v is the first large reveal strictly after ζ.

For the bounds, let $S = S_{k-1}$. On $\{v_k = v\}$ we have $v \notin S$. Write $\psi_S = A + x_v B$, where $B = (d_v\psi)_S$ and A do not depend on x_v. Averaging over the unrevealed coordinates under U,

$$q^{(k-1)} = \mathbb{E}_U[A^2 + B^2 \mid x_S], \quad \mathbb{E}_U[\psi^2 x_v \mid x_S] = 2\mathbb{E}_U[AB \mid x_S], \quad \mathbb{E}_U[(d_v\psi)^2 \mid x_S] = \mathbb{E}_U[B^2 \mid x_S].$$

By Cauchy–Schwarz, with $\rho := \rho_v(S) = \mathbb{E}_U[B^2 \mid x_S]/q^{(k-1)}$,

$$c_k^2 = \frac{4\,\mathbb{E}_U[AB \mid x_S]^2}{(q^{(k-1)})^2} \le \frac{4\,\mathbb{E}_U[A^2 \mid x_S]\,\mathbb{E}_U[B^2 \mid x_S]}{(q^{(k-1)})^2} = 4\rho(1-\rho) \le 4\rho.$$

This is the bound $c_k^2 \le 4\rho_v(S_{k-1})$ behind Corollary J.11. On $E_v \cap \{v_k = v\}$ we have $|c_k| \ge \tau_1$, so $\rho_v(S_{k-1}) \ge c_k^2/4 \ge \tau_1^2/4 = \beta_1^\circ$. With $q^{(k-1)} \ge he^{-y}$ this gives $\mathbb{E}_U[(d_v\psi)^2 \mid x_{S_{k-1}}] = \rho_v(S_{k-1})\, q^{(k-1)} \ge \beta_1^\circ he^{-y}$.

Step 5: (c), change of measure. If $\mathrm{Inf}_v(\psi) = 0$ then $d_v\psi = 0$, so $\rho_v(S_{k-1}) = 0$ and $E_v = \emptyset$ by (b); assume $\mathrm{Inf}_v(\psi) > 0$. Let $\Pr^0 := U \otimes$ clocks. Then $\widetilde{\Pr}$ and $\widetilde{\Pr}_v$ have densities ψ^2/w and $(d_v\psi)^2/\mathrm{Inf}_v(\psi)$ with respect to $\Pr^0$, and they have the same clocks. The σ-field $\mathcal{G}^+_{k-1}$ is generated by clock information and $x_{S_{k-1}}$, with S_{k-1} determined by the clock information. Under $\Pr^0$, x is uniform and independent of the clocks, so on $\mathcal{G}^+_{k-1}$

$$\frac{d\widetilde{\Pr}}{d\Pr^0} = \frac{q^{(k-1)}}{w}, \qquad \frac{d\widetilde{\Pr}_v}{d\Pr^0} = \frac{\mathbb{E}_U[(d_v\psi)^2 \mid x_{S_{k-1}}]}{\mathrm{Inf}_v(\psi)}.$$

On the event $E_v \cap \{v_k = v\} \in \mathcal{G}^+_{k-1}$, the second density is positive by (b), and the ratio of the two is

$$\frac{q^{(k-1)}}{w} \cdot \frac{\mathrm{Inf}_v(\psi)}{\mathbb{E}_U[(d_v\psi)^2 \mid x_{S_{k-1}}]} = \frac{\rho_v}{\rho_v(S_{k-1})} \le \frac{\rho_v}{\beta_1^\circ}.$$

Hence $\widetilde{\Pr}(E_v \cap \{v_k = v\}) = \mathbb{E}_{\widetilde{\Pr}_v}\big[(\rho_v/\rho_v(S_{k-1}))\mathbf{1}(E_v \cap \{v_k = v\})\big] \le (\rho_v/\beta_1^\circ)\,\widetilde{\Pr}_v(E_v \cap \{v_k = v\})$. The events $\{v_k = v\}$, $k \ge 1$, are disjoint and cover E_v, since v is revealed at most once. Summing over k gives (c). Only σ-fields strictly before the value x_v are used. □

Remark K.5 (constants and conventions).

- *Exceptional mass.* The exceptional mass in (a) is $3e^{-y}$. These are the three events E_1, E_2, E_3 of the run restarted at ζ. The start height $q^{(\zeta)} \ge h$ is known at ζ, so no separate lower excursion before ζ is needed; a priming bound, as for a stage-1 loss, would cost a fourth e^{-y}.
- *The constants β_1° and β_1.* Theorem 10.14 is stated and proved with $\beta_1^\circ = \tau_1^2/4$, the constant that the bound $c_k^2 \le 4\rho_v(S_{k-1})$ of Corollary J.11 gives for a large reveal. The grid of Section 10.2 and the recursion of Proposition 10.17 use $\beta_1 = \beta_1^\circ/2 = \tau_1^2/8$. The factor 2 is the factor $K/(K-1)$ of Lemma 10.13 at $K = 2$, and it enters only in the proof of Theorem 10.16, as follows. By (b), on E_v^c the child's height just before v's value is at least $\beta_1^\circ he^{-y} = 2\beta_1 he^{-y}$. The node exploration reads x only on F, and its terminal σ-field contains $\sigma(F, x_F)$ (Remark K.4), so Lemma 10.13 bounds the probability of the corresponding sup-event by $2\widetilde{\Pr}_{c\cdot v}(Q^{c\cdot v} \ge \beta_1 he^{-y})$. Combined with (c), each child then costs $2\rho_v^c/\beta_1^\circ = \rho_v^c/\beta_1$. In the recursion $(h, H) = (h_{s,k}, H_{s,k})$, and (G2) of Lemma 10.5

gives $\beta_1 h_{s,k} e^{-y} \geq H_{s+1,k}$. Lemma 10.13 at $K = 2$, applied to ψ itself, gives the second inequality of Lemma 10.15; that factor 2 goes into $\chi = 6e^{-y}$, not into β_1.

- *Hypothesis.* The hypothesis $\ln(H/(2h)) \geq 2y + \gamma_1 \max\{\tau_1, \kappa_1(\ln(e/h) + y)\}$ is used only in Step 3. In Section 10.4 it is a hypothesis of Theorem 10.16; at the grid heights $(H_{s,k}, h_{s,k}, y)$, where Proposition 10.17 applies that theorem, it is supplied by (3), display (G0) of Lemma 10.5.

Appendix L. The arithmetic of Proposition 10.25

This appendix proves Proposition 10.25 with every inequality written out. Each step is either an elementary inequality, proved where it is used, or a finite evaluation of explicit expressions at $T = 4$. The evaluations are tagged and their values are printed; the scripts `qpg_arith_bounds.py` and `qpg_appendix_consts.py` of the supplementary directory `supplement/` recompute every printed number except those of the elementary bounds (E4) of Section L.3, and check the inequalities between the numbers they recompute; for (E4), which is proved where it is stated, they check the three resulting inequalities at sample values of T. A printed value used as an upper bound is rounded up in its last digit. No step maximises a function of T numerically: every intercept bound is T^3 times a polynomial in $u = 1/T$ with nonnegative coefficients, which is largest at $u = \frac{1}{4}$, and the one linear bound in u comes from Lemma L.4.

L.1. **Setting and notation.** Throughout, $d \geq 2$, $T = \lceil \log_2(2ed) \rceil$ and $\varepsilon = 1/(16ed)$, as in the chain (Theorem 8.4). The notation is that of Sections 10.2 and 10.7:

$$\alpha = \frac{1}{4}, \qquad c = \frac{1}{2T}, \qquad S = 16T + 4, \qquad \varepsilon_1 = \frac{1}{150T^2}, \qquad \gamma_1 = (e-1)\ln 2,$$

$$K_\ell = 3000 e \gamma_1, \qquad \ell = \ln \frac{1}{\beta_1} = K_\ell T^2 + 6 + \ln 8 + 8 \ln d,$$

$$y = \ell + \ln d + \ln 24 + t_S, \qquad t_S = \frac{\ln(1024 e S) + 2 \ln d}{S},$$

$$r = \frac{1}{1-\varepsilon_1}, \qquad b_0 = 1 + \ln 2 + 2y + \varepsilon_1(1+y), \qquad b_1 = r b_0 + y + \ell, \qquad b_2 = r b_0 + \ln 2,$$

$$\sigma_s = b_1 \frac{r^s - 1}{r - 1}, \qquad \omega_k = b_2 \frac{r^k - 1}{r - 1}, \qquad \ln \frac{1}{p_*} = \ln \frac{64 S d}{\varepsilon} + (S-1)(\ell + \ln d).$$

Here $\ell = 2\ln(1/\tau_1) + \ln 8$ with $\ln(1/\tau_1) = 3 + 4\ln d + 10 e \gamma_1/\varepsilon_1$ and $20 e \gamma_1/\varepsilon_1 = K_\ell T^2$. We also write

$$u = \frac{1}{T} \in \left(0, \frac{1}{4}\right], \qquad \Lambda = K_\ell T^2, \qquad P_0 = \frac{r b_0}{\Lambda}, \qquad P_1 = \frac{b_1}{\Lambda}, \qquad P_2 = \frac{b_2}{\Lambda},$$

so that $\varepsilon_1 = u^2/150$, $\varepsilon_1 \Lambda = K_\ell/150 = 20 e \gamma_1$, $16T\Lambda = 16 K_\ell T^3$, $(S-1)/(16T) = 1 + \frac{3}{16}u$ and $1 + (S-1)\alpha = 4T + \frac{7}{4}$. Since $2ed \leq 2^T$ and $d \geq 2$, we have $T \geq 4$ and

$$0 < \ln 2 \leq \ln d \leq (T-1)\ln 2 - 1 < T \ln 2. \tag{16}$$

Everything below holds for every real $T \geq 4$ and every d satisfying (16); integrality of T is not used. In every quantity bounded below, $\ln d$ occurs only with nonnegative coefficients, and each occurrence is bounded through (16); so no separate monotonicity argument in d is needed.

L.2. **The five conditions as linear conditions in X.** Recall that $X = \ln(1/\theta) > 0$ and that, by Lemma 10.5(G4), $Z_{s,k} = r^{s+k} X + r^k \sigma_s + \omega_k$ and $z_{s,k} = r(Z_{s,k} + b_0)$; on the bottom row, $Z^*_s = Z_{s,S-1}$ and $\ln(1/h^*_s) = z_{s,S-1}$.

quantity	value	quantity	value
$K_\ell = 3000e\gamma_1$	9712.602050	$16K_\ell$	155401.6328
$256K_\ell$	$2.486426 \cdot 10^6$	$20e\gamma_1 = \varepsilon_1\Lambda$	64.750680
$e_1 = 9\ln 2/K_\ell$	$\le 6.422918 \cdot 10^{-4}$	$r_0 = 2400/2399$	1.000416840
$\rho_r = r_0/600$	≤ 0.001667362	e_b	0.002995276
$\xi_c = r_0(32 + \frac{7}{4})/150$	≤ 0.2250938	ρ_2	0.2317939
$\xi_K = r_0(16 + \frac{3}{4})/150$	≤ 0.1117133	ρ_K	0.1133175
$a_0 = \ln(1024e) - 2 - 2\ln 2$	4.545177	$\tau(4)$	0.210439
S at $T = 4$	68	ε_1 at $T = 4$	$1/2400$
Λ at $T = 4$	155401.6328	$3\ln 2 - 1$	1.079442

TABLE 6. Constants used below; e_b, ρ_2, ρ_K and τ are defined in Lemma L.3. All but the last two rows are independent of T [*computer-assisted: finite evaluation, values printed*].

Lemma L.1 (exact forms; [*proved here, audited*])**.** *For $0 \le s \le S-1$,*

$$Z_0^* - X = (r^{S-1}-1)X + \omega_{S-1}, \qquad D_s^* = \varepsilon_1 r^{s+S}X + (r-1)(r^{S-1}\sigma_s + \omega_{S-1}) + rb_0,$$
$$\ln\frac{\theta}{h_s^*} = (r^{s+S}-1)X + r(r^{S-1}\sigma_s + \omega_{S-1}) + rb_0,$$

and $B_s^* = \varepsilon_1 r^{s+S}X + r^{s+S-1}b_1$ for $s \le S-2$. Consequently the conditions of Proposition 10.25 read $\ln M \ge mX + \Gamma$, with m and Γ free of θ:

- *(G″), $s \le S-2$: $m = 4\varepsilon_1 r^{s+S}$, $\Gamma = 4r^{s+S-1}b_1$;*
- *(CV″): $m = r^{S-1} - 1 + \varepsilon_1 r^{s+S}$, $\Gamma = \omega_{S-1} + (r-1)(r^{S-1}\sigma_s + \omega_{S-1}) + rb_0 + \ln 4$;*
- *(MW″): the same m, and $\Gamma = \omega_{S-1} + (r-1)(r^{S-1}\sigma_s + \omega_{S-1}) + rb_0 + s\ell + \ln(512Sd^2/(3\varepsilon))$;*
- *(LD″): $m = 2(r^{s+S}-1)$, $\Gamma = 2\big[2\ln(64Sd^{s+2.5}/(3\varepsilon)) + 2s\ell + r(r^{S-1}\sigma_s + \omega_{S-1}) + rb_0\big]$;*
- *(CL″): $m = (3r^{2S-2}-1)/(4T+\frac{7}{4})$, $\Gamma = 3\big(r^{S-1}\sigma_{S-1} + \omega_{S-1} + \ln(1/p_*)\big)/(4T+\frac{7}{4})$.*

Proof. Since $\sigma_0 = 0$, $Z_0^* = r^{S-1}X + \omega_{S-1}$. Next, $z_{s,S-1} - Z_{s,S-1} = (r-1)Z_{s,S-1} + rb_0$, and $(r-1)r^{s+S-1} = \varepsilon_1 r^{s+S}$ because $1 - 1/r = \varepsilon_1$; this gives $D_s^* = \ln(H_s^*/h_s^*)$, and $\ln(\theta/h_s^*) = z_{s,S-1} - X$. Since $\sigma_{s+1} - \sigma_s = b_1 r^s$, we get $B_s^* = Z_{s+1}^* - Z_s^* = (r^{s+S} - r^{s+S-1})X + r^{S-1}b_1 r^s$. Now insert these into (G″), multiplied by $1/\alpha = 4$; into (CV″) and (MW″); into the logarithm of (LD″), $\frac{1}{2}\ln M \ge 2\ln(64Sd^{s+2.5}/(3\varepsilon)) + 2s\ell + \ln(\theta/h_s^*)$, multiplied by 2; and, with $Z_{S-1}^* = r^{2S-2}X + r^{S-1}\sigma_{S-1} + \omega_{S-1}$, into (CL″), divided by $1 + (S-1)\alpha = 4T + \frac{7}{4}$. □

Since $X > 0$, the hypothesis $\ln M \ge cX + \ln P_d$ of Proposition 10.25 implies a condition $\ln M \ge mX + \Gamma$ as soon as

$$m \le c \qquad \text{and} \qquad \Gamma \le \ln P_d = 3.0 \cdot 10^6\, T^3. \tag{17}$$

Lemma L.5 and Propositions L.6–L.9 verify (17) for every condition, and Theorem L.10 collects them.

Lemma L.2 (monotonicity in s and geometric sums; [*proved here, audited*])**.** (a) *r^s, σ_s, $s\ell$ and $\ln(64Sd^{s+2.5}/(3\varepsilon))$ are nondecreasing in s. Hence in (G″), (CV″), (MW″) and (LD″) the slope and the intercept are nondecreasing in s, and in (G″) all exponents of r are at most $2S-2$.* (b) *For $n \ge 1$, $(r^n - 1)/(r-1) = \sum_{j<n} r^j \le nr^{n-1}$. Hence $\sigma_{S-1} \le (S-1)r^{S-2}b_1$ and*

$\omega_{S-1} \le (S-1)r^{S-2}b_2$, *and therefore*

$$r^{S-1}\sigma_{S-1} + \omega_{S-1} \le (S-1)\big(r^{2S-3}b_1 + r^{S-2}b_2\big),$$
$$r\big(r^{S-1}\sigma_{S-1} + \omega_{S-1}\big) \le (S-1)\big(r^{2S-2}b_1 + r^{S-1}b_2\big).$$

Proof. (a) $r > 1$, $b_1 > 0$, $\ell > 0$ and $d \ge 2$. (b) Each of the n terms r^j, $j < n$, is at most r^{n-1}. □

L.3. **Elementary bounds.**

Lemma L.3 (elementary bounds; [*proved here, audited; the constants are finite evaluations, printed*])**.** *For every $T \ge 4$ and every d satisfying (16):*

(E1) $\ell \le \ell + \ln d \le \Lambda(1+e_1u) - 3 - 6\ln 2$ *and* $y \le \Lambda(1+e_1u)$, *where* $e_1 = 9\ln 2/K_\ell \le 6.422918 \cdot 10^{-4}$.

(E2) $\varepsilon_1 \le 1/2400$ *and* $r \le 1 + \rho_r u$ *with* $\rho_r = 0.001667362$. *For* $0 \le n \le 2S-1$, $r^n \le 1 + \rho_2 u$ *with* $\rho_2 = 0.2317939$; *for* $0 \le n \le S-1$, $r^n \le 1 + \rho_K u$ *with* $\rho_K = 0.1133175$.

(E3) $b_0 \le \Lambda(2 + e_b u)$ *with* $e_b = 0.002995276$. *Consequently, coefficientwise in u,*

$$P_0 \le \bar P_0(u) := (1+\rho_r u)(2+e_b u) = 2 + 0.00633\,u + 4.994209382 \cdot 10^{-6}u^2,$$
$$P_1 \le \bar P_1(u) := \bar P_0(u) + 2(1+e_1u) = 4 + 0.0076145836\,u + 4.994209382 \cdot 10^{-6}u^2,$$
$$P_2 \le \bar P_2(u) := \bar P_0(u) + 1.784144 \cdot 10^{-5}u = 2 + 0.00634784144\,u + 4.994209382 \cdot 10^{-6}u^2.$$

(E4) $R'' := 2\ln\big(64Sd^{(S-1)+2.5}/(3\varepsilon)\big) \le 26.7\,T^2$, $\ln\big(512Sd^2/(3\varepsilon)\big) \le 4T^2$ *and* $\ln(64Sd/\varepsilon) \le 4T$. *Hence* $\ln(1/p_*) \le (S-1)\Lambda(1+e_1u) + 4T$.

Proof. (E1) $\ell + \ln d - \Lambda = 6 + \ln 8 + 9\ln d \le 6 + 3\ln 2 + 9\big((T-1)\ln 2 - 1\big) = 9T\ln 2 - 3 - 6\ln 2$ by (16), and $9T\ln 2 = \Lambda \cdot 9\ln 2/(K_\ell T) = \Lambda e_1 u$. Next, $y - (\ell + \ln d) = \ln 24 + t_S$. By (16), $t_S \le \tau(T) := (a_0 + \ln(16T+4) + 2T\ln 2)/(16T+4)$ with $a_0 = \ln(1024e) - 2 - 2\ln 2 = 4.545177$. Then $\tau'(T)(16T+4)^2 = 16 + 8\ln 2 - 16a_0 - 16\ln(16T+4)$, which decreases in T and equals $-118.6898 < 0$ at $T = 4$; so $t_S \le \tau(4) = 0.210439 \le 0.22$. Since $-3 - 6\ln 2 + \ln 24 + 0.22 = -3.7608 < 0$, $y \le \Lambda(1+e_1u)$.

(E2) $\varepsilon_1 = u^2/150 \le 1/2400$, so $r - 1 = \varepsilon_1/(1-\varepsilon_1) \le r_0\varepsilon_1$ with $r_0 = 2400/2399$, and $r_0\varepsilon_1 = r_0u^2/150 \le (r_0/600)\,u$. Since $\ln r = -\ln(1-\varepsilon_1) \le \varepsilon_1/(1-\varepsilon_1) \le r_0\varepsilon_1$, for $n \le 2S-1 = 32T+7$ we have $\xi := n\ln r \le r_0(32u+7u^2)/150 \le \xi_c u \le \xi_0 := \xi_c/4 \le 0.0562735$. By Taylor's formula, $e^\xi \le 1 + \xi + \frac12\xi^2 e^\xi \le 1 + \xi\big(1 + \frac12\xi_0 e^{\xi_0}\big)$, so $r^n \le 1 + \xi_c\big(1 + \frac12\xi_0 e^{\xi_0}\big)u \le 1 + \rho_2 u$. For $n \le S-1 = 16T+3$, in the same way $\xi \le r_0(16u+3u^2)/150 \le \xi_K u \le \xi_{K,0} := \xi_K/4 \le 0.0279284$, and $\xi_K\big(1 + \frac12\xi_{K,0}e^{\xi_{K,0}}\big) \le \rho_K$.

(E3) $b_0 = 1 + \ln 2 + \varepsilon_1 + (2+\varepsilon_1)y$. By (E1), $\varepsilon_1\Lambda = 20e\gamma_1$ and $1 + \ln 2 + \varepsilon_1 \le 1 + \ln 2 + 1/2400 \le 1.7$, we get

$$b_0 \le 2\Lambda(1+e_1u) + 1.7 + 20e\gamma_1(1+e_1u) \le 2\Lambda(1+e_1u) + 66.461078,$$

where $66.461078 \ge 1.7 + 20e\gamma_1(1 + e_1/4)$. Since $1/\Lambda = u^2/K_\ell \le u/(4K_\ell)$, this is at most $\Lambda\big(2 + (2e_1 + 66.461078/(4K_\ell))u\big)$, and $2e_1 + 66.461078/(4K_\ell) \le 0.0012845836 + 0.001710692 \le e_b$. The bound on P_0 follows from (E2). For P_1, $y + \ell \le 2\Lambda(1+e_1u)$ by (E1). For P_2, $\ln 2/\Lambda \le (\ln 2/(4K_\ell))u \le 1.784144 \cdot 10^{-5}u$. The products are expanded exactly.

(E4) By $1/\varepsilon = 16ed$, $R'' = 2\ln(1024eS/3) + 2(S+2.5)\ln d$. By (16), $2(S+2.5)\ln d \le 2(16T+6.5)T\ln 2 \le 22.19T^2 + 9.02T$, and $2\ln(1024e/3) \le 13.67$; so $26.7T^2 - R'' \ge \varphi(T) := 4.51T^2 - 9.02T - 13.67 - 2\ln(16T+4)$. Here $\varphi(4) = 13.97 > 0$ and $\varphi'(T) = 9.02(T-1) - 32/(16T+4) > 0$, so $\varphi > 0$. Next, $\ln(512Sd^2/(3\varepsilon)) = \ln(8192eS/3) + 3\ln d \le 8.92 + \ln(16T+4) + 2.08T$, which is at most $21.47 \le 4T^2$ at $T = 4$, and its derivative $16/(16T+4) + 2.08$ is at most $8T$, the derivative of $4T^2$. Finally, $\ln(64Sd/\varepsilon) = \ln(1024eS) + 2\ln d \le a_0 + \ln(16T+4) + 2T\ln 2$, which equals $14.309863 \le 16$ at $T = 4$ and has derivative $16/(16T+4) + 2\ln 2 < 4$. The bound on $\ln(1/p_*)$ follows with (E1). □

Lemma L.4 (chord bound; [*proved here, audited*])**.** *Let $f(u)=\sum_{k=0}^{n} f_k u^k$ with all $f_k \ge 0$, and let $u_0 > 0$. Then $f(u) \le f(u_0)$ and $f(u) \le f_0 + (u/u_0)\big(f(u_0)-f_0\big)$ for $0 \le u \le u_0$. For $u_0 = \frac14$, the terms of degree $k \ge 2$ contribute at most $u\sum_{k\ge2} f_k 4^{1-k}$.*

Proof. $u^k \le u_0^k$ and, for $k \ge 1$, $u^k = u\cdot u^{k-1} \le u\,u_0^{k-1}$; sum against $f_k \ge 0$. □

This is the chord lemma of the audited version of this appendix, restated so that the appendix is self-contained.

L.4. **Slopes.**

Lemma L.5 (slopes; [*proved here, audited; constants evaluated, printed*])**.** *For every $s \le S-1$ ($s \le S-2$ for (G″)):*

- *(LD″): $Tm \le 2\rho_2 = 0.4635878$;*
- *(CL″): $m \le (2+3\rho_2 u)/(4T+\frac74)$, and $c-m \ge (\frac74 - 6\rho_2)/\big(2T(4T+\frac74)\big)$ with $\frac74 - 6\rho_2 = 0.3592366 > 0$; at $T=4$, $Tm \le 0.4898807$;*
- *(MW″) and (CV″): $Tm \le \rho_K + (1+\rho_2/4)/600 \le 0.1150808$;*
- *(G″): $Tm \le 4(1+\rho_2/4)/600 \le 0.0070530$.*

In particular $m \le c = 1/(2T)$ for all five conditions.

Proof. All exponents of r below are at most $2S-1$, so (E2) applies. (LD″): $m = 2(r^{s+S}-1) \le 2\rho_2 u$. (CL″): $3r^{2S-2} - 1 \le 2 + 3\rho_2 u$, and $\frac{1}{2T} - \frac{2+3\rho_2 u}{4T+7/4} = \frac{4T+7/4-4T-6\rho_2}{2T(4T+7/4)}$, since $2T\cdot 3\rho_2 u = 6\rho_2$; at $T=4$, $T(2+3\rho_2/4)/17.75 \le 0.4898807$. (MW″), (CV″): $m = r^{S-1} - 1 + \varepsilon_1 r^{s+S} \le \rho_K u + (u^2/150)(1+\rho_2 u)$, so $Tm \le \rho_K + (u/150)(1+\rho_2 u) \le \rho_K + (1+\rho_2/4)/600$. (G″): $m = 4\varepsilon_1 r^{s+S} \le (4u^2/150)(1+\rho_2 u)$, so $Tm \le 4(1+\rho_2/4)/600$. □

L.5. **Intercepts.** By Lemma L.2(a) it suffices to bound each intercept at $s = S-1$; for (G″) the bound below holds for every $s \le S-2$ directly. Below, $\bar P_0$, $\bar P_1$, $\bar P_2$ are the polynomials of (E3), and $\bar P_{12} := \bar P_1 + \bar P_2 = 6 + 0.01396242504\,u + 9.988418764\cdot 10^{-6} u^2$.

Proposition L.6 ((LD″); [*proved here, audited; constants evaluated at $u=\frac14$, printed*])**.** *For every $s \le S-1$, $\Gamma_{\mathrm{LD}} \le 16K_\ell T^3\,(16 + 6.196076\,u) \le 2.727147\cdot 10^6\,T^3$. As $T\to\infty$ the bound tends to $256K_\ell T^3 = 2.486426\cdot 10^6\,T^3$.*

Proof. At $s = S-1$, $\Gamma_{\mathrm{LD}} = 2\big[R'' + 2(S-1)\ell + r(r^{S-1}\sigma_{S-1} + \omega_{S-1}) + rb_0\big]$. By (E4), $R'' \le 26.7T^2$; by (E1), $\ell \le \Lambda(1+e_1 u)$; by Lemma L.2(b) and (E2), $r(r^{S-1}\sigma_{S-1}+\omega_{S-1}) \le (S-1)(1+\rho_2 u)\Lambda\bar P_{12}(u)$; and $rb_0 \le \Lambda \bar P_0(u)$. Divide by $16T\Lambda = 16K_\ell T^3$ and use $(S-1)/(16T) = 1 + \frac{3}{16}u$, $1/(16T) = u/16$ and $T^2/(16T\Lambda) = u/(16K_\ell)$:

$$\frac{\Gamma_{\mathrm{LD}}}{16K_\ell T^3} \le f_{\mathrm{LD}}(u) := 4\Big(1+\frac{3u}{16}\Big)(1+e_1 u) + 2\Big(1+\frac{3u}{16}\Big)(1+\rho_2 u)\bar P_{12}(u) + \frac{u}{8}\bar P_0(u) + 2\cdot 0.0001718129\,u,$$

where $0.0001718129 \ge 26.7/(16K_\ell)$. The polynomial f_{LD} has the nonnegative coefficients 16, 6.062364443, 0.53453794, 0.0012226523, $8.682204526\cdot 10^{-7}$ (of $u^0,\dots,u^4$), and $f_{\mathrm{LD}}(0) = 4+2\cdot 6 = 16$. Its four terms at $u=\frac14$ are 4.188172399, 13.298211071, 0.062549463 and $8.591\cdot 10^{-5}$, so $f_{\mathrm{LD}}(\frac14) \le 17.549019$ [*finite evaluation, printed*]. By Lemma L.4, $f_{\mathrm{LD}}(u) \le 16 + 4\big(f_{\mathrm{LD}}(\frac14) - 16\big)u \le 16 + 6.196076\,u$; the terms of degree at least 2 contribute at most $0.133711\,u$. Finally $16K_\ell \cdot 17.549019 \le 2.727147\cdot 10^6$ and $16K_\ell\cdot 16 = 256K_\ell$. □

Proposition L.7 ((MW″) and (CV″); [*proved here, audited; constants evaluated at $u=\frac14$, printed*])**.** *For every $s \le S-1$, $\Gamma_{\mathrm{CV}} \le \Gamma_{\mathrm{MW}} \le 16K_\ell T^3 f_{\mathrm{MW}}(\frac14) \le 5.125043\cdot 10^5\,T^3$, with f_{MW} defined in the proof.*

Proof. For each s, $\Gamma_{\mathrm{CV}} \le \Gamma_{\mathrm{MW}}$, since $s\ell \ge 0$ and $512Sd^2/(3\varepsilon) \ge 4$. At $s = S-1$, by Lemma L.2(b) and (E2), $\omega_{S-1} + (r-1)(r^{S-1}\sigma_{S-1} + \omega_{S-1}) \le (S-1)(1+\rho_2 u)\big[b_2 + r_0\varepsilon_1(b_1+b_2)\big]$. With $rb_0 \le \Lambda \bar{P}_0$, $(S-1)\ell \le (S-1)\Lambda(1+e_1 u)$ by (E1), and $\ln(512Sd^2/(3\varepsilon)) \le 4T^2$ by (E4), dividing by $16K_\ell T^3$ gives

$$\frac{\Gamma_{\mathrm{MW}}}{16K_\ell T^3} \le f_{\mathrm{MW}}(u) := \Big(1+\frac{3u}{16}\Big)(1+\rho_2 u)\Big[\bar{P}_2(u) + \frac{r_0 u^2}{150}\bar{P}_{12}(u)\Big] + \frac{u}{16}\bar{P}_0(u) + \Big(1+\frac{3u}{16}\Big)(1+e_1 u) + 2.573976\cdot 10^{-5}u,$$

where $2.573976\cdot 10^{-5} \ge 4/(16K_\ell)$. The coefficients of f_{MW} are 3, 1.158103673, 0.1301220462, 0.01715016076, 0.001778507913, $4.075124714\cdot 10^{-6}$, $2.89527454\cdot 10^{-9}$, all nonnegative. Its four terms at $u = \frac{1}{4}$ are 2.219609202, 0.031274731, 1.047043100 and $6.435\cdot 10^{-6}$, so $f_{\mathrm{MW}}(\frac{1}{4}) \le 3.2979335$ and $16K_\ell \cdot 3.2979335 \le 5.125043\cdot 10^5$ [*finite evaluation, printed*]. Lemma L.4 gives $f_{\mathrm{MW}}(u) \le f_{\mathrm{MW}}(\frac{1}{4})$. □

Proposition L.8 ((G″); [*proved here, audited; constants evaluated at* $u = \frac{1}{4}$*, printed*])**.** *For every* $s \le S-2$, $\Gamma_{\mathrm{G}} \le 4.112130\cdot 10^4\, T^3$.

Proof. $s + S - 1 \le 2S-3$, so by (E2) and (E3), $\Gamma_{\mathrm{G}} = 4r^{s+S-1}b_1 \le 4(1+\rho_2 u)\Lambda \bar{P}_1(u) = T^3 g_{\mathrm{G}}(u)$ with $g_{\mathrm{G}}(u) := 4K_\ell u(1+\rho_2 u)\bar{P}_1(u)$. Its coefficients are 0, 155401.6328, 36316.98021, 68.7655426, 0.04497429198, and $g_{\mathrm{G}}(\frac{1}{4}) = 41121.2941 \le 4.112130\cdot 10^4$ [*finite evaluation, printed*]. □

Proposition L.9 ((CL″); [*proved here, audited; constants evaluated at* $u = \frac{1}{4}$*, printed*])**.** $\Gamma_{\mathrm{CL}} \le 2.242490\cdot 10^5\, T^3$.

Proof. By Lemma L.2(b) and (E2), $r^{S-1}\sigma_{S-1} + \omega_{S-1} \le (S-1)(1+\rho_2 u)\Lambda \bar{P}_{12}(u)$, and by (E4), $\ln(1/p_*) \le (S-1)\Lambda(1+e_1 u) + 4T$. With $4T + \frac{7}{4} \ge 4T$ and $(S-1)\Lambda = 16K_\ell T^3(1+\frac{3}{16}u)$,

$$\frac{\Gamma_{\mathrm{CL}}}{T^3} \le f_{\mathrm{CL}}(u) := 12K_\ell u\Big(1+\frac{3u}{16}\Big)\Big[(1+\rho_2 u)\bar{P}_{12}(u) + (1+e_1 u)\Big] + 3u^3.$$

Its coefficients are 0, 815858.5722, 316770.8573, 31093.37894, 71.21443133, 0.05059607848. At $u = \frac{1}{4}$: $(1+\frac{3}{64})(1+\frac{\rho_2}{4})\bar{P}_{12}(\frac{1}{4}) = 6.649105535$, $(1+\frac{3}{64})(1+\frac{e_1}{4}) = 1.047043100$ and $12K_\ell/4 = 29137.80615$, so $f_{\mathrm{CL}}(\frac{1}{4}) = 224248.934 \le 2.242490\cdot 10^5$ [*finite evaluation, printed*]. □

L.6. **The final comparison.**

Theorem L.10 ([*proved here, audited; the constants are finite evaluations, printed*])**.** *For every real* $T \ge 4$ *and every* d *satisfying* (16)*, each of the conditions* (G″) ($s \le S-2$)*,* (CV″)*,* (MW″)*,* (LD″) ($s \le S-1$) *and* (CL″) *satisfies* (17)*: its slope is at most* c*, and its intercept is at most* $2.727147\cdot 10^6\, T^3 < 3.0\cdot 10^6\, T^3 = \ln P_d$*. Hence Proposition 10.25 holds for every* $d \ge 2$.

Proof. The slopes are Lemma L.5. The intercepts are Propositions L.6–L.9, with $\max\{2.727147\cdot 10^6,\ 5.125043\cdot 10^5,\ 4.112130\cdot 10^4,\ 2.242490\cdot 10^5\} = 2.727147\cdot 10^6$. By Lemma L.1 and (17), the hypothesis $\ln M \ge cX + \ln P_d$ then implies every condition, for every θ. □

Remark L.11 (comparison with the proof of Proposition 10.25). The bounds stated in the proof of Proposition 10.25 are those proved here. The largest intercept is always that of (LD″), and its bound exceeds the exact value at $T = 4$ by less than 1.8%. The main term $256K_\ell T^3$ of Γ_{LD} comes from $2r(r^{S-1}\sigma_{S-1} + \omega_{S-1}) \approx 2\cdot 6S\Lambda$ and $4(S-1)\ell \approx 4S\Lambda$, with $S \approx 16T$. The tightest slope is that of (CL″), whose margin $c - m$ is of order $1/T^2$; every other slope is below c by a margin of order $1/T$.

condition	$T \cdot m$, bound	$T \cdot m$, exact	Γ/T^3, bound	Γ/T^3, exact
(LD″)	0.4635878	0.462996	$2.727147 \cdot 10^6$	$2.679188 \cdot 10^6$
(MW″)	0.1150808	0.115027	$5.125043 \cdot 10^5$	$4.981268 \cdot 10^5$
(CV″)	0.1150808	0.115027	$5.125043 \cdot 10^5$	$3.354230 \cdot 10^5$
(CL″)	0.4898807	0.489533	$2.242490 \cdot 10^5$	$1.980183 \cdot 10^5$
(G″)	0.0070530	0.007050	$4.112130 \cdot 10^4$	$4.108287 \cdot 10^4$

TABLE 7. The bounds of this appendix at $T = 4$, against the exact values from the closed forms of Lemma L.1 at $\ln d = 3\ln 2 - 1$, maximised over s (the values at $d = 2$ agree to four digits). The targets are $T \cdot c = 0.5$ for the slopes and $3.0 \cdot 10^6$ for Γ/T^3 [*computer-assisted: finite evaluation, values printed; the exact values are not used in the proofs*].

APPENDIX M. A QUANTUM OSSS CONJECTURE

For the bias $f = 2p - 1$ of a randomized decision tree of depth t with revealments r_i (p the acceptance probability; the revealments are written δ_i in Theorem 4.1, while in this appendix δ_k denotes averaged query magnitudes), inequality (L1) of Theorem 4.1 [42] with $h = f$, averaged over the random choice of the tree, gives $\mathrm{Var} f \le \sum_i r_i \mathbb{E}|d_i f|$; by Cauchy–Schwarz and $\sum_i r_i \le t$,

$$\mathrm{Var}(f)^2 \le t \sum_i r_i \mathrm{Inf}_i(f). \tag{18}$$

This appendix formulates the analogue of (18) for quantum query algorithms, with the revealments replaced by the query magnitudes of [11] averaged over the input, and proves it in several cases. A version with a constant polynomial in t would imply Conjecture 1.1 for acceptance probabilities of quantum query algorithms, and with it the classical simulation of quantum query algorithms on most inputs (Proposition M.12); that case of the conjecture is open [2, 13], so we do not expect a general proof to be easy. We prove the inequality with the sharp constant for one query (Proposition M.5), with an absolute constant for two queries (Proposition M.6), with a linear constant for a nonadaptive classical round followed by one quantum query (Proposition M.7), and with the conjectured constant $4t$ for algorithms whose averaged magnitude profile μ_k is constant (e.g. if the query distribution does not depend on the input), which includes nonadaptive algorithms (Corollary M.15); for nonadaptive algorithms the factor t can be dropped altogether (Proposition M.16), which gives their classical simulation with $\lceil 4t^2/(\varepsilon^4\delta^3)\rceil$ queries (Corollary M.17); Remark M.18 discusses algorithms with r rounds of adaptivity. For general t we outline an $e^{O(t)}$ bound (Remark M.8). Inner products are conjugate-linear in the first argument. The symbols δ_k, μ_k, Λ, κ, Γ, T and $K(t)$ defined in this appendix are local to it.

M.1. **The model.** Let $n, w \ge 1$ and $\mathcal{H} = \mathbb{C}^{n+1} \otimes \mathbb{C}^w$, a query register with basis $e_0, \dots, e_n$ and a workspace. Let $P_i = |e_i\rangle\langle e_i| \otimes I_w$ for $0 \le i \le n$. For $x \in Q_n$ put $x_0 = 1$ and $O_x = \sum_{i=0}^n x_i P_i$; the index 0 is the *null query*.

Definition M.1 ([*standard*])**.** A *t-query algorithm* consists of a unit vector $\psi_0 \in \mathcal{H}$, unitaries $U_1, \dots, U_t$ on $\mathcal{H}$ and a projection Π. Its states are $\psi_s(x) = U_s O_x \psi_{s-1}(x)$ for $1 \le s \le t$, and its *bias* is

$$f(x) = \langle \psi_t(x), (2\Pi - I)\psi_t(x)\rangle = 2\Pr[\text{accept}] - 1.$$

The *query magnitude* of $k \in [n]$ at x is $m_k(x) = \sum_{s=1}^t \|P_k \psi_{s-1}(x)\|^2$, and $\delta_k = \mathbb{E}_x m_k(x)$. We write $\mu_k(x) = \mathbb{E}_{x_k} m_k(x)$, which does not depend on x_k; thus $\mathbb{E}\mu_k = \delta_k$.

Remark M.2 (conventions; [*standard*]). (a) $\sum_{k=1}^{n} m_k(x) = t - \sum_{s=1}^{t} \|P_0\psi_{s-1}(x)\|^2 \le t$, so $\sum_k \delta_k \le t$. Every coordinate of $\psi_t(x)$ is a polynomial of degree at most t in x, so $f \in \mathcal{B}_{2t}(n)$.

(b) With $\Phi(x) = O_x\psi_{t-1}(x)$ and $M = U_t^\dagger(2\Pi - I)U_t$, a Hermitian unitary, $f = \langle \Phi, M\Phi\rangle$. Conversely, every Hermitian M with $\|M\| \le 1$ occurs after doubling the workspace, because $\begin{pmatrix} M & (I-M^2)^{1/2} \\ (I-M^2)^{1/2} & -M \end{pmatrix}$ is a Hermitian unitary; the magnitudes do not change. In the proofs, M is therefore an arbitrary Hermitian contraction.

(c) The XOR oracle $|i,b\rangle \mapsto |i, b \oplus X_i\rangle$ becomes, after a Hadamard gate on b, the phase x_i on the branch $b = 1$ and the identity on the branch $b = 0$; routing the branch $b = 0$ to the null index (and storing i in the workspace) gives the present model, with magnitudes at most the usual magnitudes of [11]. Controlled phase oracles are handled in the same way, and intermediate measurements can be deferred. Since δ_k appears on the right-hand side of every inequality below, these inequalities hold a fortiori with the usual magnitudes.

(d) $\langle O_x\psi, O_{x^{\oplus k}}\psi\rangle = 1 - 2\|P_k\psi\|^2$ for a unit vector ψ, so a query that determines x_k with certainty, made on a state that does not depend on x_k, has magnitude exactly $\frac{1}{2}$ on k. A randomized decision tree with revealments r_i is therefore simulated with $\delta_i = r_i/2$, and (18) reads $\mathrm{Var}(f)^2 \le 2t\sum_i \delta_i \mathrm{Inf}_i(f)$.

(e) For $\psi\colon Q_n \to \mathcal{H}$ we use the vector-valued notation of Section 2 ($\mathcal{H}$ viewed as a real Hilbert space with inner product $\mathrm{Re}\langle\cdot,\cdot\rangle$): $d_k\psi = \frac{1}{2}(\psi(x^{k\to 1}) - \psi(x^{k\to -1}))$ and $\mathrm{Inf}_k(\psi) = \mathbb{E}\|d_k\psi\|^2$. For fixed k we write $\psi = \psi^+ + x_k\psi^-$ with $\psi^\pm$ independent of x_k; thus $\psi^- = d_k\psi$.

(f) If $\delta_k = 0$ then $m_k \equiv 0$, the algorithm never puts amplitude on k, and f does not depend on x_k. Removing such coordinates, we assume $\delta_k > 0$ for all k whenever we divide by δ_k.

M.2. **The conjecture.** We say that *QOSSS holds with constant* $K(t)$ if $\mathrm{Var}(f)^2 \le K(t)\sum_{k=1}^{n}\delta_k\mathrm{Inf}_k(f)$ for the bias f of every t-query algorithm, for all n and w.

Conjecture M.3 (QOSSS; [*open*])**.** *For every $t \ge 1$ and the bias f of every t-query algorithm,*

$$\mathrm{Var}(f)^2 \le 4t\sum_{k=1}^{n}\delta_k\mathrm{Inf}_k(f).$$

By Remark M.2(d), classical randomized decision trees satisfy the conjecture with 2 in place of 4. The variance must be squared: for $n = 1$, the one-query algorithm with $\psi_0 = \sqrt{1-\varepsilon}\,e_0 + \sqrt{\varepsilon}\,e_1$ and $M = |e_0\rangle\langle e_1| + |e_1\rangle\langle e_0|$ has $f = 2\sqrt{\varepsilon(1-\varepsilon)}\,x_1$ and $\delta_1 = \varepsilon$, so $\mathrm{Var}f = 4\varepsilon(1-\varepsilon)$ while $\delta_1\mathrm{Inf}_1(f) = 4\varepsilon^2(1-\varepsilon)$. Interference with the null branch produces a bias of size $\sqrt{\varepsilon}$ from magnitude ε, which no decision tree can do. The constant $4t$ cannot be lowered:

Remark M.4 (sharpness; [*proved here, audited*]). Let T be odd, $0 < \varepsilon < 1$, $\bar{x} = n^{-1}\sum_{k=1}^{n} x_k$ and $u = n^{-1/2}\sum_{k=1}^{n} e_k$ (with $w = 1$ before dilation). Let $R = P_0 + 2|u\rangle\langle u| - (I - P_0)$, the reflection about u on $\mathrm{span}\{e_1,\dots,e_n\}$, which fixes e_0. Consider the T-query algorithm with $\psi_0 = \sqrt{1-\varepsilon}\,e_0 + \sqrt{\varepsilon}\,u$, $U_1 = \cdots = U_{T-1} = R$, and $M = |e_0\rangle\langle u| + |u\rangle\langle e_0|$. On $\mathrm{span}\{u, O_x u\}$ the operator RO_x is a rotation by $\arccos\bar{x}$, and $Ru = u$, so

$$f(x) = 2\sqrt{\varepsilon(1-\varepsilon)}\,\mathrm{Re}\langle u, (RO_x)^T u\rangle = 2\sqrt{\varepsilon(1-\varepsilon)}\,T_T(\bar{x}),$$

a multiple of the Chebyshev family of Example 6.3. The algorithm and the uniform measure are invariant under permutations of $[n]$, and $\sum_{k\ge 1}\|P_k(RO_x)^s u\|^2 = 1$, so $\delta_k = \varepsilon T/n$. Expanding $T_T(\bar{x})$ in monomials, the coefficient of a set S with $|S| = j$ is $O(n^{-j})$ and the coefficient of $\{k\}$ is $2\sqrt{\varepsilon(1-\varepsilon)}\,T_T'(0)/n + O(n^{-2})$, where $|T_T'(0)| = T$. Hence $\mathrm{Var}f = 4\varepsilon(1-\varepsilon)T^2/n + O(n^{-2})$ and $\mathrm{Inf}_k(f) = 4\varepsilon(1-\varepsilon)T^2/n^2 + O(n^{-3})$, so

$$\frac{\mathrm{Var}(f)^2}{T\sum_k \delta_k\mathrm{Inf}_k(f)} \longrightarrow 4(1-\varepsilon) \qquad (n\to\infty).$$

The ratio therefore comes arbitrarily close to 4 for every odd T. For $T=1$ this is the example above, spread over n coordinates. For every T, Grover search [22] with $T-1$ iterations followed by one query reading the coordinate found (magnitude $\frac12$) has $f=T_{2T-1}(\bar x)$ and $\delta_k=(T-\frac12)/n$; the same computation gives the limit $4-2/T$. An exact symmetric reduction (orthonormal Krawtchouk polynomials, $n=10^6$, $\varepsilon=10^{-3}$) gives $\mathrm{Var}(f)^2/\sum_k\delta_k\mathrm{Inf}_k(f)=3.996$, 11.988, 19.980, 35.961, 67.912 for $T=1,3,5,9,17$, against $4T$ [*numerical*]. In particular the factor t in Conjecture M.3 is necessary.

M.3. **One query.**

Proposition M.5 (one query; [*proved here, audited*])**.** *Let f be the bias of a one-query algorithm. Then*

$$\sum_k\frac{\mathrm{Inf}_k(f)}{\delta_k}\le 4,\qquad \textit{hence}\qquad \mathrm{Var}(f)^2\le I(f)^2\le 4\sum_k\delta_k\mathrm{Inf}_k(f).$$

The constant 4 is best possible, and $\mathrm{Var}(f)^2<4\sum_k\delta_k\mathrm{Inf}_k(f)$ for every nonconstant f; so 4 is a supremum and is not attained. The first and the last inequality of the display can be equalities: $\psi_0=(e_1+e_2)/\sqrt2$ and $M=|e_1\rangle\langle e_2|+|e_2\rangle\langle e_1|$ give $f=x_1x_2$ and $\delta_1=\delta_2=\frac12$.

Proof. Let $M=U_1^\dagger(2\Pi-I)U_1$, $v_i=P_i\psi_0$ and $a_i=\|v_i\|^2$, so that $\sum_{i\ge0}a_i=1$ and $\delta_k=a_k$ for $k\ge1$. Then $O_x\psi_0=\sum_{i=0}^n x_iv_i$ and, with $c_{ij}=\mathrm{Re}\langle v_i,Mv_j\rangle=c_{ji}$,

$$f(x)=\sum_{i,j=0}^n x_ix_j\langle v_i,Mv_j\rangle=\sum_i c_{ii}+2\sum_{k\ge1}c_{0k}x_k+2\sum_{1\le j<k}c_{jk}x_jx_k.$$

Hence $\widehat f(\{k\})=2c_{0k}$, $\widehat f(\{j,k\})=2c_{jk}$ and $\mathrm{Inf}_k(f)=4\sum_{j\ge0,\,j\ne k}c_{kj}^2$. Write $v_i=\sqrt{a_i}\,u_i$ for $a_i>0$; the u_i are orthonormal because the ranges of the P_i are orthogonal, and $c_{kj}=\sqrt{a_ka_j}\,\mathrm{Re}\langle u_k,Mu_j\rangle$. By Bessel's inequality,

$$\sum_k\frac{\mathrm{Inf}_k(f)}{a_k}=4\sum_k\sum_{j\ne k}a_j\big(\mathrm{Re}\langle u_k,Mu_j\rangle\big)^2\le4\sum_j a_j\sum_{k\ne j}|\langle u_k,Mu_j\rangle|^2\le4\sum_j a_j\|Mu_j\|^2\le4,$$

all sums running over indices with positive a. Then $\mathrm{Var}f\le I(f)=\sum_k(\mathrm{Inf}_k/a_k)^{1/2}(a_k\mathrm{Inf}_k)^{1/2}\le 2(\sum_k\delta_k\mathrm{Inf}_k(f))^{1/2}$ by Cauchy–Schwarz.

Remark M.4 with $T=1$ gives ratios $4(1-\varepsilon)-o(1)$, so 4 is best possible. Let f be nonconstant. If some $\widehat f(\{j,k\})\ne0$, then $\mathrm{Var}f=\sum_{S\ne\emptyset}\widehat f(S)^2<I(f)$, so $\mathrm{Var}(f)^2<I(f)^2\le4\sum_k\delta_k\mathrm{Inf}_k(f)$. Otherwise f has degree 1 and $\mathrm{Inf}_k(f)=4c_{0k}^2$; then $a_0>0$, and $a_0<1$ (if $a_0=1$ then f is constant). Bessel's inequality applied to the single vector Mu_0 gives $\sum_k\mathrm{Inf}_k/a_k=4a_0\sum_{k\ge1}(\mathrm{Re}\langle u_k,Mu_0\rangle)^2\le4a_0$, hence $\mathrm{Var}(f)^2=I(f)^2\le4a_0\sum_k\delta_k\mathrm{Inf}_k(f)<4\sum_k\delta_k\mathrm{Inf}_k(f)$, since $\sum_k\delta_k\mathrm{Inf}_k(f)>0$. In both cases $\mathrm{Var}(f)^2<4\sum_k\delta_k\mathrm{Inf}_k(f)$. □

M.4. **Two queries.**

Proposition M.6 (two queries; [*proved here, audited*])**.** *Let f be the bias of a two-query algorithm and $K_2=8(1620\sqrt3+80)\approx23{,}087$. Then*

$$\sum_k\frac{\mathrm{Inf}_k(f)}{\delta_k}\le K_2,\qquad \textit{hence}\qquad \mathrm{Var}(f)^2\le K_2\sum_k\delta_k\mathrm{Inf}_k(f).$$

Proof. Let $U=U_1$, $M=U_2^\dagger(2\Pi-I)U_2$, $v_j=P_j\psi_0$, $a_j=\|v_j\|^2$, $\varphi=O_x\psi_0$, $\chi=U\varphi=\psi_1$ and $\Phi=O_x\chi$, so that $f=\langle\Phi,M\Phi\rangle$. Let $w_{ij}=P_iUv_j$ and $b_i=\sum_j\|w_{ij}\|^2$. Since $P_k\chi=\sum_j x_jw_{kj}$ and $\mathbb{E}x_ix_j=[i=j]$ (with $x_0=1$), $\delta_k=a_k+b_k$.

Step 1 (derivative formula). Fix $k \geq 1$ and let $O'_k = O_x - x_k P_k$ and $\varphi'_k = O'_k \psi_0 = \varphi - x_k v_k$, both independent of x_k. Since $x_k^2 = 1$,

$$\Phi = (x_k P_k + O'_k) U (x_k v_k + \varphi'_k) = \Phi^0_k + x_k \Phi^1_k, \qquad \Phi^0_k = w_{kk} + O'_k U \varphi'_k, \quad \Phi^1_k = P_k U \varphi'_k + O'_k U v_k,$$

with Φ^0_k, Φ^1_k independent of x_k. Hence $d_k f = 2\mathrm{Re}\langle M\Phi^0_k, \Phi^1_k\rangle = 2\mathrm{Re}A_k + 2\mathrm{Re}B_k$, where

$$A_k = \langle P_k M \Phi^0_k, P_k U \varphi'_k\rangle, \qquad B_k = \langle P_k U^\dagger O'_k M \Phi^0_k, v_k\rangle$$

(O'_k is Hermitian and $v_k = P_k v_k$), and $\mathrm{Inf}_k(f) \leq 8\mathbb{E}|A_k|^2 + 8\mathbb{E}|B_k|^2$.

Step 2 (Cauchy–Schwarz). $|B_k|^2 \leq a_k h_k$ with $h_k = \|P_k U^\dagger O'_k M \Phi^0_k\|^2$, and $|A_k|^2 \leq G_k \Delta_k$ with $G_k = \|P_k M \Phi^0_k\|^2$ and $\Delta_k = \|P_k U \varphi'_k\|^2 = \|\sum_{j\neq k} x_j w_{kj}\|^2$.

Step 3 (decoupling). If $g \geq 0$ is a polynomial of degree at most D, then $\|g\|_2 \leq 3^D \mathbb{E} g$: Bonami's inequality $\|g\|_4 \leq 3^{D/2}\|g\|_2$ (Section 2, [41, Chapter 9]) and Hölder's inequality $\|g\|_2 \leq \|g\|_1^{1/3}\|g\|_4^{2/3}$ give $\|g\|_2 \leq 3^D \|g\|_1$. Every coordinate of Φ^0_k is a polynomial of degree at most 2, so $\deg G_k \leq 4$ and $\|G_k\|_2 \leq 81 \mathbb{E} G_k$. Next, $\Delta_k = \sum_{i,j\neq k} R_{ij} x_i x_j$ with $R = (\mathrm{Re}\langle w_{ki}, w_{kj}\rangle)_{i,j\neq k}$ positive semidefinite; distinct pairs $\{i,j\}$ give distinct characters (with $x_0 = 1$), so $\mathbb{E}\Delta_k = \mathrm{tr} R \leq b_k$ and $\mathrm{Var}\Delta_k = 2\sum_{i\neq j} R_{ij}^2 \leq 2\sum_{i\neq j} R_{ii}R_{jj} \leq 2(\mathrm{tr}R)^2$, whence $\|\Delta_k\|_2 \leq \sqrt{3}\, b_k$. Therefore

$$\mathbb{E}|A_k|^2 \leq \mathbb{E}[G_k\Delta_k] \leq \|G_k\|_2 \|\Delta_k\|_2 \leq 81\sqrt{3}\, b_k \mathbb{E} G_k, \qquad \mathbb{E}|B_k|^2 \leq a_k \mathbb{E} h_k.$$

Step 4 (pointwise sums over k). We use $\sum_{k\geq 1}\|P_k y\|^2 \leq \|y\|^2$ and that U, M, O_x, O'_k are contractions. Since $P_k U \varphi'_k = P_k \chi - x_k w_{kk}$ and $\|w_{kk}\| \leq \|v_k\|$,

$$\|\Phi^1_k\| \leq \|P_k \chi\| + \|w_{kk}\| + \sqrt{a_k}, \qquad \sum_k \|\Phi^1_k\|^2 \leq 3(1+1+1) = 9.$$

Since $\Phi^0_k = \Phi - x_k \Phi^1_k$, $G_k \leq 2\|P_k M\Phi\|^2 + 2\|\Phi^1_k\|^2$ and $\sum_k G_k \leq 2 + 18 = 20$. Since $O'_k M \Phi^0_k = O_x M\Phi - x_k O_x M \Phi^1_k - x_k P_k M \Phi + P_k M \Phi^1_k$,

$$h_k \leq 4\big(\|P_k U^\dagger O_x M\Phi\|^2 + \|P_k M\Phi\|^2 + 2\|\Phi^1_k\|^2\big), \qquad \sum_k h_k \leq 4(1+1+18) = 80.$$

Step 5. By Steps 1–3, $\mathrm{Inf}_k(f) \leq 8\cdot 81\sqrt{3}\, b_k \mathbb{E} G_k + 8 a_k \mathbb{E} h_k \leq \delta_k(8\cdot 81\sqrt{3}\,\mathbb{E}G_k + 8\mathbb{E}h_k)$. Dividing by δ_k and summing with Step 4 gives $\sum_k \mathrm{Inf}_k(f)/\delta_k \leq 8(81\sqrt{3}\cdot 20 + 80) = K_2$. The second claim follows from $\mathrm{Var} f \leq I(f)$ and Cauchy–Schwarz as in Proposition M.5. □

The constant is far from sharp: on hill-climbed instances ($n \leq 6$, $w \leq 3$) the largest observed values were 8.6 for $\sum_k \mathrm{Inf}_k/\delta_k$ and 3.93 for $\mathrm{Var}(f)^2/\sum_k \delta_k \mathrm{Inf}_k(f)$ [*numerical*]. The proof gets n-independence from Step 3: a sensitivity weight G_k that is a nonnegative polynomial of low degree cannot concentrate on the inputs where the magnitude Δ_k is large. The intermediate quantity $\sum_k \mathrm{Inf}_k/\delta_k$ is not polynomially bounded in t, however: for the address function with k address bits (Section 2.4), read classically with $t = k+1$ queries, it equals $k + 2^{k+1}$. Bounds through it are therefore exponential in t (Remark M.8), although that function satisfies Conjecture M.3 by Remark M.2(d).

M.5. **A classical round followed by one quantum query.**

Proposition M.7 ([*proved here, audited*])**.** *Let $S \subseteq [n]$ with $|S| = m$ and $t = m+1$. Consider an algorithm that first reads the coordinates x_i, $i \in S$, one per query and in a fixed order: the read of x_i is a query made on a state with amplitude only on the indices 0 and i of the query register (e.g. $(e_0 + e_i)/\sqrt{2}$ tensored with a workspace vector), it determines x_i with certainty, and its outcome is copied into the workspace. The algorithm then makes one more query; that is, after the reads it runs a one-query algorithm A_a that depends on $a = x_S$. Then its bias satisfies*

$$\mathrm{Var}(f)^2 \leq 16t \sum_k \delta_k \mathrm{Inf}_k(f).$$

Proof. Write $z = x_{S^c}$ and $X = \sum_k \delta_k \mathrm{Inf}_k(f)$. By Remark M.2(d), $\delta_i \geq \frac{1}{2}$ for $i \in S$. Since $a = x_S$ is known after the reads, a query of A_a to $i \in S$ can be replaced by the fixed phase a_i (applied by a unitary, with the branch routed to the null index); this changes neither f nor the magnitudes outside S. So $f(x) = g_{x_S}(z)$, where g_a is the bias of a one-query algorithm on z with magnitudes $\delta_k(A_a)$, and $\delta_k = \mathbb{E}_a \delta_k(A_a)$ for $k \notin S$, because the reads put no amplitude on $k \notin S$. Let $\bar{g} = \mathbb{E}_a g_a$. It is the bias of a one-query algorithm: with $\psi_0^{(a)}$ and M_a the initial state and measurement operator of A_a, prepare $2^{-m/2}\sum_a |a\rangle \otimes \psi_0^{(a)}$ in an extra register, query, and measure $\sum_a |a\rangle\langle a| \otimes M_a$. Its magnitudes are $\mathbb{E}_a \delta_k(A_a) = \delta_k$.

(i) $\mathrm{Var} f = \mathbb{E}_{a,z}(g_a(z) - \bar{g}(z))^2 + \mathrm{Var}\bar{g}$, since the cross term vanishes ($\mathbb{E}_a(g_a - \bar{g}) = 0$).

(ii) For fixed z, the Poincaré inequality on $\{-1,1\}^S$ gives $\mathbb{E}_a(g_a(z) - \bar{g}(z))^2 \leq \sum_{i\in S} \mathbb{E}_a(d_i g_\cdot(z))^2$, and $d_i g_{x_S}(z) = d_i f(x)$ for $i \in S$. Hence the first term in (i) is at most $\sum_{i\in S} \mathrm{Inf}_i(f) \leq 2\sum_{i\in S} \delta_i \mathrm{Inf}_i(f) \leq 2X$.

(iii) For $k \notin S$, Jensen's inequality gives $\mathrm{Inf}_k(\bar{g}) = \mathbb{E}_z(\mathbb{E}_a d_k g_a(z))^2 \leq \mathbb{E}_{a,z}(d_k g_a(z))^2 = \mathrm{Inf}_k(f)$.

(iv) Proposition M.5 applied to $\bar{g}$ and (iii) give $\mathrm{Var}(\bar{g})^2 \leq 4\sum_{k\notin S} \delta_k \mathrm{Inf}_k(\bar{g}) \leq 4X$.

(v) Thus $\mathrm{Var} f \leq 2X + 2\sqrt{X}$. Since $\mathrm{Inf}_k(f) \leq 1$ and $\sum_k \delta_k \leq t$, we have $X \leq t$, so $X \leq \sqrt{tX}$ and $\mathrm{Var} f \leq 2\sqrt{tX} + 2\sqrt{X} \leq 4\sqrt{tX}$. □

This class contains the address function, so here the conclusion holds even though $\sum_k \mathrm{Inf}_k/\delta_k$ is exponential in t. The variance that the quantum magnitudes cannot pay for (the dependence of g_a on a) is charged linearly to the influences of the classically read coordinates, which is affordable because $X \leq \sqrt{tX}$. The argument does not extend to a quantum first round. The analogue of $\bar{g}$ is then the bias with the query register measured after the first query, and there are two-query examples in which this measurement increases an influence from 0 to a positive value, so step (iii) has no analogue. The extension to an adaptive classical round is open.

M.6. **General t: an $e^{O(t)}$ bound.**

Remark M.8 ([*proof outline*]). For every t-query bias, $\sum_k \mathrm{Inf}_k(f)/\delta_k \leq K_{\mathrm{out}}(t) := 12t\,4^{t-1}3^{6t}(1+t^2+4^{t-1}t^2) = e^{O(t)}$, hence QOSSS holds with constant $K_{\mathrm{out}}(t)$, independently of n and w. The proof follows that of Proposition M.6; we list the steps.

(i) *Expansion.* Fix k and write $O_x = x_k P_k + O'_k$ in each of the t oracles. Then $\Phi = O_x\psi_{t-1} = \sum_{A\subseteq[t]} x_k^{|A|}\Phi_A$, where Φ_A uses P_k at the queries in A and O'_k at the others; all Φ_A are independent of x_k, and $d_k f = 2\mathrm{Re}\langle M\Phi_k^0, \Phi_k^1\rangle$ with $\Phi_k^0 = \sum_{|A| \text{ even}} \Phi_A$ and $\Phi_k^1 = \sum_{|A| \text{ odd}} \Phi_A$.

(ii) *Pairing.* For odd A with $r = \min A$ and $s = \max A$, $\Phi_A = T'_s P_k C P_k \psi'_{r-1}$, where C is a contraction, ψ'_{r-1} is the state before query r computed with the oracles O'_k, and T'_s is the part of the circuit after query s, also with the oracles O'_k. Hence $|\langle M\Phi_k^0, \Phi_A\rangle|^2 \leq G_{k,s}\Delta_{k,r}$ with $G_{k,s} = \|P_k T_s'^\dagger M\Phi_k^0\|^2$ and $\Delta_{k,r} = \|P_k\psi'_{r-1}\|^2$. With Cauchy–Schwarz over the 2^{t-1} odd sets, $\mathrm{Inf}_k(f) \leq 4\cdot 2^{t-1}\sum_{A \text{ odd}} \mathbb{E}[G_{k,\max A}\Delta_{k,\min A}]$.

(iii) *Decoupling.* $\deg G_{k,s} \leq 2(2t-s)$ and $\deg \Delta_{k,r} \leq 2(r-1)$, so as in Step 3 of Proposition M.6, $\mathbb{E}[G_{k,s}\Delta_{k,r}] \leq 3^{6t}\mathbb{E}G_{k,s}\,\mathbb{E}\Delta_{k,r}$.

(iv) *Hybrid.* $\psi'_{r-1} - \psi_{r-1} = -\sum_{q<r}(U_{r-1}O'_k\cdots U_q)\,x_k P_k\psi_{q-1}$, so $\Delta_{k,r} \leq r\sum_{q\leq r}\|P_k\psi_{q-1}(x)\|^2$ and $\mathbb{E}\Delta_{k,r} \leq t\delta_k$.

(v) *Sums over k.* Write $G_{k,s} \leq 3(\|P_k T_s^\dagger M\Phi\|^2 + \|P_k(T_s'^\dagger - T_s^\dagger)M\Phi\|^2 + \|\Phi_k^1\|^2)$, where T_s uses the true oracles. The first terms sum to at most 1. Expanding $T_s'^\dagger - T_s^\dagger$ by the hybrid argument with the true oracles applied first, each term is a contraction applied to $P_k y_q$ with a unit vector y_q independent of k; there are at most t terms, so the second terms sum to at most t^2. Likewise $\|\Phi_k^1\| \leq \sum_{A \text{ odd}}\sum_{q\leq\min A}\|P_k\psi_{q-1}\|$, so $\sum_k\|\Phi_k^1\|^2 \leq 4^{t-1}t^2$. Hence $\sum_k G_{k,s} \leq 3(1+t^2+4^{t-1}t^2)$ pointwise.

(vi) *Total.* $\sum_k \mathrm{Inf}_k/\delta_k \le 4\cdot 2^{t-1}\cdot 3^{6t}\cdot t\cdot 2^{t-1}\cdot 3(1+t^2+4^{t-1}t^2) = K_{\mathrm{out}}(t)$.

The pointwise identities and inequalities in (i), (ii), (iv) were checked numerically for $t \le 4$ [*numerical*], and an independent audit found every step correct; the bookkeeping has not been written out in full. The consequences are only of the strength of [16]: $\mathrm{MaxInf}(f) \ge \mathrm{Var}(f)^2 e^{-O(t)}$. What the bound adds is that the weighted form holds with Var^2 rather than Var^4 and without dependence on n. By the address function, no bound on $\sum_k \mathrm{Inf}_k/\delta_k$ can be polynomial in t.

M.7. **Two lemmas on the final state.** The following lemmas control the derivatives of the final state by the query magnitudes, in L^2 over the input.

Lemma M.9 ([*proved here, audited*])**.** *For every $k \in [n]$ and $0 \le s \le t$,*

$$\mathrm{Inf}_k(\psi_s) = \sum_{r=0}^{s-1} \big(\mathbb{E}\|P_k\psi_r^+\|^2 - \mathbb{E}\|P_k\psi_r^-\|^2\big) \le \sum_{r=0}^{s-1} \mathbb{E}\|P_k\psi_r\|^2.$$

In particular $\mathrm{Inf}_k(\psi_t) \le \delta_k$. *Moreover* $|d_k f| \le 2\|d_k\psi_t\|$ *pointwise, so* $\mathrm{Inf}_k(f) \le 4\delta_k$.

Proof. Let $O_x^{(k)} = P_k + \sum_{i\ne k} x_i P_i$, which is unitary, independent of x_k and commutes with P_k; then $O_x = O_x^{(k)}(I - P_k + x_k P_k)$. With $\psi_r = \psi_r^+ + x_k\psi_r^-$ and $x_k^2 = 1$,

$$O_x\psi_r = O_x^{(k)}\big[(I-P_k)\psi_r^+ + P_k\psi_r^-\big] + x_k O_x^{(k)}\big[(I-P_k)\psi_r^- + P_k\psi_r^+\big],$$

so $\psi_{r+1}^- = U_{r+1}O_x^{(k)}[(I-P_k)\psi_r^- + P_k\psi_r^+]$. The two vectors in the bracket are orthogonal, so $\|\psi_{r+1}^-\|^2 = \|\psi_r^-\|^2 - \|P_k\psi_r^-\|^2 + \|P_k\psi_r^+\|^2$ pointwise. Since $\psi_0^- = 0$, summing gives the identity, because $\mathrm{Inf}_k(\psi_s) = \mathbb{E}\|\psi_s^-\|^2$. The inequality follows from $\mathbb{E}\|P_k\psi_r\|^2 = \mathbb{E}\|P_k\psi_r^+\|^2 + \mathbb{E}\|P_k\psi_r^-\|^2$ (the cross term is odd in x_k). Finally, with $\psi_\pm = \psi_t(x^{k\to\pm 1})$ and $Z = 2\Pi - I$ Hermitian and unitary, $f(x^{k\to 1}) - f(x^{k\to -1}) = \mathrm{Re}\langle \psi_+ - \psi_-, Z(\psi_+ + \psi_-)\rangle$, which is at most $2\|\psi_+ - \psi_-\| = 4\|d_k\psi_t\|$ in absolute value. □

The hybrid argument of [11] gives $|d_k f(x)| \le 2\sum_s \|P_k\psi_{s-1}(x)\|$ pointwise, hence only $\mathrm{Inf}_k(f) \le 4t\delta_k$; compare [13, Proposition A.2]. In Lemma M.9 the amplitude toggled by a query is orthogonal to the odd part that survives, so on average over x_k the odd parts do not add up coherently. The difference $\delta_k - \mathrm{Inf}_k(\psi_t) = 2\sum_r \mathbb{E}\|P_k\psi_r^-\|^2$ is the magnitude spent on uncomputation. All relations between magnitudes and influences that we know of in the literature (and Lemma M.9) bound influences by magnitudes. Conjecture M.3 goes the other way, as do Propositions M.5, M.6, M.7 and M.16 and Corollary M.15.

Lemma M.10 (directional derivatives; [*proved here, audited*])**.** *For every k and $0 \le s < t$, pointwise in x,*

$$d_k\psi_{s+1} = U_{s+1}O_x\big[(I-2P_k)\,d_k\psi_s + x_k P_k\psi_s\big].$$

Consequently, if $c_k\colon Q_n \to \mathbb{R}$ ($k \in [n]$) do not depend on x_k, then

$$\mathbb{E}\Big\|\sum_k c_k\, d_k\psi_t\Big\|^2 \le t\sum_k \mathbb{E}\big[c_k^2\mu_k\big];$$

in particular $\mathbb{E}\|\sum_k c_k\, d_k\psi_t\|^2 \le t\sum_k c_k^2\delta_k$ *for constants c_k.*

Proof. In the proof of Lemma M.9, $O_x^{(k)} = O_x(I - P_k + x_kP_k)$, so $d_k\psi_{s+1} = U_{s+1}O_x[(I-P_k)\psi_s^- + x_kP_k\psi_s^+]$, and $x_k\psi_s^+ = x_k\psi_s - \psi_s^-$ gives the recursion. Let $Y_s = \sum_k c_k d_k\psi_s$. The recursion gives $Y_{s+1} = U_{s+1}O_x[Y_s + Z_s]$ with

$$Z_s = \sum_k c_k P_k\big(x_k\psi_s - 2d_k\psi_s\big) = \sum_k c_k P_k\big(x_k\psi_s^+ - \psi_s^-\big),$$

where $\psi_s^{\pm}$ refer to the coordinate k of the summand. The summands lie in the orthogonal ranges of the P_k, and averaging over x_k removes the cross term, so $\mathbb{E}\|Z_s\|^2 = \sum_k \mathbb{E}[c_k^2\|P_k\psi_s\|^2] = \sum_k \mathbb{E}[c_k^2\mathbb{E}_{x_k}\|P_k\psi_s\|^2]$. Since $Y_0 = 0$, Minkowski's inequality in $L^2(\mathcal{H})$ and Cauchy–Schwarz give $\|Y_t\|_{L^2} \le \sum_{s<t}\|Z_s\|_{L^2} \le \sqrt{t}\,(\sum_{s<t}\mathbb{E}\|Z_s\|^2)^{1/2} = \sqrt{t}\,(\sum_k \mathbb{E}[c_k^2\mu_k])^{1/2}$. □

For $c = e_k$, Lemma M.10 is weaker than Lemma M.9 by the factor t. For general c the factor t is necessary, by the following sharp level-1 bound and Remark M.4. Equivalently, the Gram matrix $\Gamma_{jk} = \mathrm{Re}\,\mathbb{E}\langle d_j\psi_t, d_k\psi_t\rangle$ satisfies $\Gamma \preceq t\,\mathrm{diag}(\delta)$, while $\Gamma_{kk} \le \delta_k$ by Lemma M.9.

Theorem M.11 (sharp level-1 bound; [*proved here, audited*])**.** *For the bias f of every t-query algorithm, $\sum_k \widehat{f}(\{k\})^2/\delta_k \le 4t$. The constant is sharp: the family of Remark M.4 gives $4(1-\varepsilon)t - o(1)$ for odd t.*

Proof. Let $Z = 2\Pi - I$. Since $f = \langle\psi_t^+, Z\psi_t^+\rangle + \langle\psi_t^-, Z\psi_t^-\rangle + 2x_k\mathrm{Re}\langle Z\psi_t^+, \psi_t^-\rangle$, we get $\widehat{f}(\{k\}) = \mathbb{E}d_k f = 2\mathrm{Re}\,\mathbb{E}\langle Z\psi_t^+, \psi_t^-\rangle = 2\mathrm{Re}\,\mathbb{E}\langle Z\psi_t, d_k\psi_t\rangle$, because the term $x_k\langle Z\psi_t^-, \psi_t^-\rangle$ has mean 0. For constants c_k, Lemma M.10 gives $\sum_k c_k\widehat{f}(\{k\}) = 2\mathrm{Re}\,\mathbb{E}\langle Z\psi_t, \sum_k c_k d_k\psi_t\rangle \le 2(t\sum_k c_k^2\delta_k)^{1/2}$. Taking $c_k = \widehat{f}(\{k\})/\delta_k$ gives the bound. In Remark M.4, $\sum_k \widehat{f}(\{k\})^2/\delta_k = 4(1-\varepsilon)T + O(1/n)$. □

M.8. **Reductions.**

Proposition M.12 ([*proved here, audited*])**.** *Let f be the bias of a t-query algorithm and $v = \mathrm{Var}f$.*

(a) If QOSSS holds with constant $K(t)$, then $\mathrm{MaxInf}(f) \ge v^2/(tK(t))$ and $\max_k \delta_k \ge v^2/(4tK(t))$.

(b) $\mathrm{MaxInf}(f)^2/4 \le \sum_k \delta_k \mathrm{Inf}_k(f) \le t\mathrm{MaxInf}(f)$. Hence the weak form *"$\sum_k \delta_k\mathrm{Inf}_k(f) \ge (v/t)^C/C$ for all t-query biases" is equivalent to Conjecture 1.1 for all t-query biases (with $d = 2t$), up to changing the constant C.*

(c) If Conjecture 1.1 holds for all biases of quantum query algorithms, then for every t-query algorithm and $\varepsilon, \delta > 0$ there is a deterministic classical algorithm making $\mathrm{poly}(t, 1/\varepsilon, 1/\delta)$ queries that approximates f within ε on all but a δ fraction of inputs.

(d) By Theorem 10.1 with $d = 2t$ and (b), the weak form holds with a quasi-polynomial loss: $\sum_k \delta_k\mathrm{Inf}_k(f) \ge \frac{1}{4}\mathrm{MaxInf}(f)^2 \ge (v/2t)^{O(\log 2t)}\exp(-O(\log^4(2t)))$.

In particular QOSSS with a polynomial constant implies Conjecture 1.1 for biases of quantum query algorithms, which implies polynomial classical simulation on most inputs.

Proof. (a) $v^2 \le K(t)\sum_k \delta_k\mathrm{Inf}_k(f) \le K(t)(\sum_k \delta_k)\mathrm{MaxInf}(f) \le tK(t)\mathrm{MaxInf}(f)$. Also $\sum_k \delta_k\mathrm{Inf}_k(f) \le \max_k \delta_k \cdot I(f)$, and $I(f) \le 4\sum_k \mathrm{Inf}_k(\psi_t) \le 4t$ by Lemma M.9. (b) The upper bound uses $\sum_k \delta_k \le t$. By Lemma M.9, $\delta_k \ge \mathrm{Inf}_k(f)/4$, so $\sum_k \delta_k\mathrm{Inf}_k(f) \ge \frac{1}{4}\sum_k \mathrm{Inf}_k(f)^2 \ge \frac{1}{4}\mathrm{MaxInf}(f)^2$. (c) This is the greedy argument of [2, §3.1] in the form of Lemma 10.28 and Corollary 10.29, which applies the influence bound only to restrictions of f. A restriction of a t-query bias is again a t-query bias: fixing $x_i = b$ turns x_iP_i into the fixed operator bP_i, which is absorbed into the adjacent unitary after routing the index i to the null index. (d) Immediate from Theorem 10.1 and (b). □

By (a), QOSSS with a polynomial constant also gives the variance form of [13, Conjecture 1]: their parameter $\min(\Pr[p \ge \frac{2}{3}], \Pr[p \le \frac{1}{3}])$ is at most $9\mathrm{Var}f$, and their query weights dominate δ_k. By [4, Theorem 1.3], the biases of t-query algorithms are exactly the functions $x \mapsto T(\tilde{x}, \dots, \tilde{x})$, $\tilde{x} = (x, 1)$, with T a $2t$-linear form of cb-norm at most 1. The case of Conjecture 1.1 in (b) is therefore the conjecture for diagonals of completely bounded forms. The bounds of [7, 24] for cb-bounded block-multilinear forms concern independent blocks and do not apply to diagonals (see the caveat in Section 1.2). Up to the exponent of Var, the weighted inequality and the influence bound for biases are the same problem.

M.9. **The global quantity Λ and a transference principle.** For a t-query bias f let $\delta(S) = \sum_{k\in S}\delta_k$ and

$$\Lambda(f) = \sum_{S\neq\emptyset} \frac{\widehat{f}(S)^2}{\delta(S)}.$$

Proposition M.13 ([*proved here, audited*])**.** *(a)* $\Lambda(f) = \sup_F \mathrm{Cov}(f,F)^2/\sum_k \delta_k \mathrm{Inf}_k(F)$*, the supremum over nonconstant* $F\colon Q_n\to\mathbb{R}$*.*

(b) $\mathrm{Var}(f)^2/\sum_k \delta_k \mathrm{Inf}_k(f) \le \Lambda(f) \le \sum_k \mathrm{Inf}_k(f)/\delta_k$*. Hence* $\Lambda\le 4$ *for* $t=1$*,* $\Lambda\le K_2$ *for* $t=2$*, and a bound* $\Lambda\le Kt$ *implies QOSSS with constant* Kt*.*

(c) If f *is the bias of a randomized decision tree of depth* t*, simulated as in Remark M.2(d), then* $\Lambda(f)\le 2t$*.*

Proof. (a) $\mathrm{Cov}(f,F) = \sum_{S\neq\emptyset}\widehat{f}(S)\widehat{F}(S)$ and $\sum_k \delta_k\mathrm{Inf}_k(F) = \sum_S \widehat{F}(S)^2\delta(S)$. Apply Cauchy–Schwarz, with equality for $\widehat{F}(S) = \widehat{f}(S)/\delta(S)$. (b) Take $F=f$ in (a) for the lower bound. For the upper bound, $1/\delta(S)\le |S|^{-1}\sum_{k\in S}1/\delta_k$, so $\Lambda(f)\le \sum_k \delta_k^{-1}\sum_{S\ni k}\widehat{f}(S)^2$. Then apply Propositions M.5 and M.6. (c) With revealments $r_i = 2\delta_i$, (L1) of Theorem 4.1, averaged over the trees, gives $|\mathrm{Cov}(f,F)|\le \sum_i r_i\mathbb{E}|d_iF| \le (t\sum_i r_i\mathrm{Inf}_i(F))^{1/2}$, so $\Lambda(f)\le 2t$ by (a). □

By (a), the bound $\Lambda(f)\le Ct$ is an OSSS-type covariance inequality, $\mathrm{Cov}(f,F)^2 \le Ct\sum_k\delta_k\mathrm{Inf}_k(F)$ for all F. It is the exact analogue of (c), and it is the natural global strengthening of Conjecture M.3. Theorem M.11 is its restriction to F of degree 1. For a magnitude profile $\mu=(\mu_k)_k$ (Definition M.1) let

$$\kappa(\mu) = \sup_h \frac{\sum_{S\neq\emptyset}\hat{h}(S)^2/\delta(S)}{\sum_k \mathbb{E}[(d_kh)^2/\mu_k]},$$

over nonconstant $h\colon Q_n\to\mathbb{R}$, with $0/0=0$ and $a/0=\infty$ for $a>0$. Then $\kappa(\mu)$ depends only on the functions μ_k (recall $\delta_k=\mathbb{E}\mu_k$).

Theorem M.14 (transference; [*proved here, audited*])**.** *For the bias* f *of every* t*-query algorithm,* $\Lambda(f)\le 4t\,\kappa(\mu)$*.*

Proof. Let F be nonconstant and $\tilde{F}=F-\mathbb{E}F$. Write $\tilde{F}=\sum_k x_kc_k$ with c_k independent of x_k. Such decompositions exist: for instance $c_k=\sum_{S:\min S=k}\widehat{F}(S)x^{S\setminus\{k\}}$. Since $\mathbb{E}[fx_kc_k]=\mathbb{E}[c_kd_kf]$ and, as in the proof of Theorem M.11, $\mathbb{E}[c_kd_kf] = 2\mathrm{Re}\,\mathbb{E}[c_k\langle Z\psi_t, d_k\psi_t\rangle]$, Lemma M.10 gives

$$\mathrm{Cov}(f,F) = 2\mathrm{Re}\,\mathbb{E}\Big\langle Z\psi_t, \sum_k c_kd_k\psi_t\Big\rangle, \qquad \mathrm{Cov}(f,F)^2 \le 4t\sum_k\mathbb{E}[\mu_kc_k^2].$$

Assume first that all $\mu_k>0$. Minimizing the positive definite quadratic form $\sum_k\mathbb{E}[\mu_kc_k^2]$ over the affine space of decompositions, and using finite-dimensional duality (the adjoint of $(c_k)\mapsto\sum_k x_kc_k$ is $h\mapsto(d_kh)_k$, because $\mathbb{E}[hx_kc_k]=\mathbb{E}[c_kd_kh]$), gives

$$\min_c\sum_k\mathbb{E}[\mu_kc_k^2] = \sup_h\frac{\mathbb{E}[h\tilde{F}]^2}{\sum_k\mathbb{E}[(d_kh)^2/\mu_k]} \le \kappa(\mu)\sup_h\frac{\mathbb{E}[h\tilde{F}]^2}{\sum_{S\neq\emptyset}\hat{h}(S)^2/\delta(S)} = \kappa(\mu)\sum_k\delta_k\mathrm{Inf}_k(F),$$

the last equality by Cauchy–Schwarz as in Proposition M.13(a). Hence $\mathrm{Cov}(f,F)^2\le 4t\kappa(\mu)\sum_k\delta_k\mathrm{Inf}_k(F)$, and Proposition M.13(a) gives the claim. In general, the bound $\mathrm{Cov}(f,F)^2\le 4t\sum_k\mathbb{E}[(\mu_k+\eta)c_k^2]$ holds for every $\eta>0$, so the argument gives $\Lambda(f)\le 4t\kappa_\eta$, where κ_η is defined like $\kappa(\mu)$ with $\mu_k+\eta$ in the denominator and the same $\delta(S)$ in the numerator. As $\eta\downarrow 0$, κ_η decreases to $\kappa(\mu)$, by compactness of the set of mean-zero h with $\sum_S\hat{h}(S)^2/\delta(S)=1$ and lower semicontinuity of $h\mapsto\sum_k\mathbb{E}[(d_kh)^2/\mu_k]$. □

The profile μ absorbs everything else about the algorithm: the measurement, interference and uncomputation enter only through Lemma M.10.

Corollary M.15 ([*proved here, audited*])**.** *If every μ_k is constant, then $\mathrm{Cov}(f,F)^2 \le 4t\sum_k \delta_k \mathrm{Inf}_k(F)$ for all F; in particular $\Lambda(f) \le 4t$ and Conjecture M.3 holds for f. This applies to all one-query algorithms, and to nonadaptive algorithms, where the t queries are made in parallel and simulated sequentially by swapping query registers, so that $m_k(x)$ does not depend on x.*

Proof. Here $\mu_k \equiv \delta_k$. In the proof of Theorem M.14 take $c_k = \sum_{S\ni k} \omega_k(S)\widehat{F}(S)x^{S\setminus\{k\}}$ with $\omega_k(S) = \delta_k^{-1}/\sum_{j\in S}\delta_j^{-1}$. Then $\sum_k x_k c_k = \tilde{F}$ and

$$\sum_k \delta_k \mathbb{E}c_k^2 = \sum_{S\neq\emptyset} \widehat{F}(S)^2\Big(\sum_{j\in S}\delta_j^{-1}\Big)^{-1} \le \sum_S \widehat{F}(S)^2\delta(S) = \sum_k \delta_k \mathrm{Inf}_k(F).$$

In the sequential simulation of a nonadaptive algorithm, the unitaries between queries only swap the active query register with stored ones, and the oracles only multiply amplitudes by signs. So the distribution of the active query register in the computational basis does not depend on x. □

The same decomposition gives $\mathrm{Cov}(f,F)^2 \le 4t\sum_{S\neq\emptyset}\widehat{F}(S)^2(\sum_{j\in S}\delta_j^{-1})^{-1}$. By the duality of Proposition M.13(a), with these weights in place of $\delta(S)$, this is $\sum_k \mathrm{Inf}_k(f)/\delta_k \le 4t$ when every μ_k is constant; for $t=1$ it is the first inequality of Proposition M.5. For nonadaptive algorithms the factor t is unnecessary:

Proposition M.16 (nonadaptive algorithms; [*proved here, audited*])**.** *Let f be the bias of a nonadaptive t-query algorithm. Then for every $F\colon Q_n \to \mathbb{R}$,*

$$\mathrm{Cov}(f,F)^2 \le 4\sum_k \delta_k \mathrm{Inf}_k(F).$$

Hence $\Lambda(f) \le 4$, $\mathrm{Var}(f)^2 \le 4\sum_k \delta_k \mathrm{Inf}_k(f)$ and $\mathrm{MaxInf}(f) \ge \mathrm{Var}(f)^2/(4t)$, for every t. The constant 4 in the first two inequalities is sharp for every t.

Proof. The algorithm has query registers $Q_1,\dots,Q_t$, each a copy of $\mathbb{C}^{n+1}$, and a workspace. It prepares $\psi_0 = \sum_\tau |\tau\rangle\otimes w_\tau$, applies O_x to every query register, and measures. Here $\tau = (\tau_1,\dots,\tau_t)$ ranges over $\{0,\dots,n\}^t$, $|\tau\rangle = e_{\tau_1}\otimes\cdots\otimes e_{\tau_t}$, and the workspace vectors w_τ satisfy $\sum_\tau \|w_\tau\|^2 = 1$. Put $p_\tau = \|w_\tau\|^2$, let $N_k(\tau)$ be the number of s with $\tau_s = k$, and let $A(\tau) = \{k\in[n] : N_k(\tau) \text{ is odd}\}$. In the sequential simulation of Corollary M.15 the s-th query is made on Q_s, so $\delta_k = \sum_\tau p_\tau N_k(\tau)$. Since $x_0 = 1$ and $x_k^2 = 1$, the state after the queries is

$$\Psi(x) = \sum_\tau x^{A(\tau)}|\tau\rangle\otimes w_\tau,$$

and $f = \langle\Psi, M\Psi\rangle$ for a Hermitian contraction M (Remark M.2(b)). For $k\in[n]$, $d_k\Psi = \sum_{\tau\colon k\in A(\tau)} x^{A(\tau)\setminus\{k\}}|\tau\rangle\otimes w_\tau$. Let $c_k\colon Q_n\to\mathbb{R}$ be independent of x_k. Since $x^{A\setminus\{k\}} = x_k x^A$ for $k\in A$ and the vectors $|\tau\rangle$ are orthonormal, pointwise

$$\Big\|\sum_k c_k d_k\Psi\Big\|^2 = \sum_\tau p_\tau\Big(x^{A(\tau)}\sum_{k\in A(\tau)} x_k c_k\Big)^2 = \sum_\tau p_\tau\Big(\sum_{k\in A(\tau)} x_k c_k\Big)^2,$$

and hence

$$\mathbb{E}\Big\|\sum_k c_k d_k\Psi\Big\|^2 = \sum_\tau p_\tau\,\mathbb{E}\Big(\sum_{k\in A(\tau)} x_k c_k\Big)^2.$$

Now fix F and let $\tilde{F} = F - \mathbb{E}F$. Choose weights $\omega_k(S) \geq 0$ with $\sum_{k\in S}\omega_k(S) = 1$ for every nonempty S, and put $c_k = \sum_{S\ni k}\omega_k(S)\widehat{F}(S)x^{S\setminus\{k\}}$, which is independent of x_k. Then $\sum_k x_k c_k = \tilde{F}$, and for every $A \subseteq [n]$

$$\sum_{k\in A} x_k c_k = \sum_{S\neq\emptyset} \omega_S(A\cap S)\,\widehat{F}(S)x^S, \qquad \omega_S(B) = \sum_{k\in B}\omega_k(S) \in [0,1].$$

By Parseval, $\mathbb{E}(\sum_{k\in A} x_k c_k)^2 = \sum_S \omega_S(A\cap S)^2\widehat{F}(S)^2 \leq \sum_{S:\,S\cap A\neq\emptyset}\widehat{F}(S)^2$. Let τ be distributed according to p; then $\Pr[k\in A(\tau)] \leq \Pr[N_k(\tau)\geq 1] \leq \mathbb{E}N_k = \delta_k$, and so

$$\begin{aligned}\mathbb{E}\Big\|\sum_k c_k d_k \Psi\Big\|^2 &\leq \sum_{S\neq\emptyset}\widehat{F}(S)^2 \Pr_\tau[A(\tau)\cap S\neq\emptyset] \leq \sum_S \widehat{F}(S)^2 \sum_{k\in S}\Pr_\tau[k\in A(\tau)]\\ &\leq \sum_S \widehat{F}(S)^2\delta(S) = \sum_k \delta_k \mathrm{Inf}_k(F).\end{aligned}$$

As in the proof of Theorem M.14, with Ψ and M in place of ψ_t and Z,

$$|\mathrm{Cov}(f,F)| = \Big|\sum_k \mathbb{E}[c_k d_k f]\Big| = \Big|2\mathrm{Re}\,\mathbb{E}\Big\langle M\Psi, \sum_k c_k d_k\Psi\Big\rangle\Big| \leq 2\Big(\mathbb{E}\Big\|\sum_k c_k d_k\Psi\Big\|^2\Big)^{1/2},$$

which gives the inequality. The consequences follow from Proposition M.13(a), from $F = f$, and from $\sum_k \delta_k \leq t$. For sharpness, take the one-query example with $n = 1$ in Section M.2 and pad it with $t-1$ null queries (registers $Q_2,\dots,Q_t$ in the state e_0). The result is nonadaptive, has $\delta_1 = \varepsilon$, and has $\mathrm{Var}(f)^2/(\delta_1\mathrm{Inf}_1(f)) = 4(1-\varepsilon)$. □

Proposition M.16 was also checked numerically [*numerical*]. The proof gives the slightly stronger bound $\sum_{S\neq\emptyset}\widehat{f}(S)^2/\Pr_\tau[A(\tau)\cap S\neq\emptyset] \leq 4$. So for $t\geq 2$ the constant $4t$ of Corollary M.15 is not sharp for nonadaptive algorithms; the examples of Remark M.4 have nonconstant profiles. With Lemma 10.28, Proposition M.16 gives an explicit classical simulation.

Corollary M.17 (classical simulation of nonadaptive algorithms; [*proved here, audited*])**.** *Let $t\geq 1$, $0<\varepsilon<\frac{1}{2}$ and $0<\delta\leq 1$. Let $p(X)$ be the acceptance probability of a nonadaptive quantum algorithm that makes t queries to $X\in\{0,1\}^n$. There is a deterministic adaptive classical algorithm that makes at most*

$$N = \left\lceil \frac{4t^2}{\varepsilon^4\delta^3}\right\rceil \;\leq\; \frac{4t^2}{\varepsilon^4\delta^3}+1$$

queries to X and outputs a number within ε of $p(X)$ for all but a δ fraction of the inputs $X\in\{0,1\}^n$. Equivalently (replace ε by $\varepsilon/2$), $\lceil 64t^2/(\varepsilon^4\delta^3)\rceil$ queries approximate the bias $2p-1$ within ε, for $0<\varepsilon<1$.

Proof. By Remark M.2(c), applied to each query register, $f = 2p-1$ (with $x_i = 1-2X_i$) is the bias of a nonadaptive t-query algorithm as in the proof of Proposition M.16, with final state $\Psi(x) = \sum_\tau x^{A(\tau)}|\tau\rangle\otimes w_\tau$ and a Hermitian contraction M; and $f\in\mathcal{B}_{2t}(n)$ by Remark M.2(a).

Restrictions. Let h be the restriction of f by $x_i = b_i$ for $i\in R\subseteq[n]$. Then h is again the bias of a nonadaptive algorithm with t queries: a query to a fixed coordinate is a known phase, which is absorbed into the unitaries. Explicitly, $h = \langle\Psi_h, M\Psi_h\rangle$ with $\Psi_h(x) = \sum_\tau b^{A(\tau)\cap R}x^{A(\tau)\setminus R}|\tau\rangle\otimes w_\tau$. Let $\rho(\tau)$ replace every entry of τ that lies in R by the null index 0, so that $A(\rho(\tau)) = A(\tau)\setminus R$, and let V be the isometry $|\tau\rangle\otimes w\mapsto|\rho(\tau)\rangle\otimes|\tau\rangle\otimes w$, which records τ in an enlarged workspace. Then

$$V\Psi_h(x) = \sum_\tau x^{A(\rho(\tau))}|\rho(\tau)\rangle\otimes\big(|\tau\rangle\otimes b^{A(\tau)\cap R}w_\tau\big)$$

is the final state of the nonadaptive algorithm with the unit initial vector $\sum_\tau |\rho(\tau)\rangle \otimes |\tau\rangle \otimes b^{A(\tau)\cap R} w_\tau$, and $h = \langle V\Psi_h, (VMV^\dagger)V\Psi_h\rangle$, where $VMV^\dagger$ is a Hermitian contraction. This algorithm puts no amplitude on the indices in R, so its bias does not depend on x_R, and it equals h.

By Proposition M.16, $\mathrm{MaxInf}(h) \ge \mathrm{Var}(h)^2/(4t)$ for every restriction h of f. Put $v_0 = 2\varepsilon^2\delta$. Every restriction h with $\mathrm{Var}h > v_0$ has $\mathrm{MaxInf}(h) \ge v_0^2/(4t) = \varepsilon^4\delta^2/t =: \gamma$. Lemma 10.28 with $d = 2t$, $N = \lceil 2d/(\delta\gamma)\rceil = \lceil 4t^2/(\varepsilon^4\delta^3)\rceil$ and $a = 2\varepsilon$ gives an algorithm with $\Pr_x[|A(x) - f(x)| > 2\varepsilon] \le \delta/2 + v_0/(4\varepsilon^2) = \delta$. As in the proof of Corollary 10.29, the output $(1 + A(x))/2$ is within ε of $p(X)$ whenever $|A(x) - f(x)| \le 2\varepsilon$. □

Liu and Mutreja [33] proved that nonadaptive algorithms can be simulated on most inputs with $\mathrm{poly}(t, 1/\varepsilon, 1/\delta)$ classical queries; Corollary M.17 recovers this with the explicit bound $\lceil 4t^2/(\varepsilon^4\delta^3)\rceil$.

Remark M.18 (r rounds of adaptivity). An *r-round* algorithm makes its t queries in r rounds of $t_1, \dots, t_r$ parallel queries ($t_1 + \cdots + t_r = t$), each round as in the proof of Proposition M.16, with arbitrary unitaries between the rounds; its magnitudes are those of the sequential simulation of Corollary M.15. Nonadaptive algorithms are the case $r = 1$, and general t-query algorithms the case $r = t$. *Open problem* [*open*]: is $\Lambda(f) \le C(r)\, t$ for the bias f of every r-round algorithm, with $C(r)$ independent of n, w and the t_j? For $r = 1$, Proposition M.16 gives $\Lambda \le 4$.

(a) *The first round is free* [*proved here, audited*]. Let $\delta_k^{(1)}$ be the magnitudes of the first round, $\delta^{(1)}(S) = \sum_{k\in S}\delta_k^{(1)}$, and $\mu'_k = \mathbb{E}_{x_k}\sum_{s=t_1}^{t-1}\|P_k\psi_s\|^2$ the profile of the later queries in the sequential simulation. Then $\sum_k \widehat{f}(\{k\})^2/\delta_k \le 4(t - t_1 + 1)$, and $\Lambda(f) \le 4(t - t_1 + 1)$ if every μ'_k is constant. In particular, one parallel round followed by $r - 1$ single queries gives $\sum_k \widehat{f}(\{k\})^2/\delta_k \le 4r$. *Proof.* The first t_1 queries form a nonadaptive round, so $\psi_{t_1} = W\Psi$ with W a fixed unitary and Ψ as in the proof of Proposition M.16. Given F, take c_k as there; that proof gives $\mathbb{E}\|\sum_k c_k d_k \psi_{t_1}\|^2 \le \sum_S \widehat{F}(S)^2\delta^{(1)}(S)$. For the remaining $t - t_1$ queries, the recursion of Lemma M.10, Minkowski's inequality and Cauchy–Schwarz, as in its proof, add at most $(t-t_1)^{1/2}(\sum_k \mathbb{E}[c_k^2\mu'_k])^{1/2}$ to the L^2 norm of $\sum_k c_k d_k\psi_s$. As in the proof of Theorem M.14,

$$\mathrm{Cov}(f,F)^2 \le 4\Big(\Big(\sum_S \widehat{F}(S)^2\delta^{(1)}(S)\Big)^{1/2} + \Big((t-t_1)\sum_k \mathbb{E}[c_k^2\mu'_k]\Big)^{1/2}\Big)^2.$$

Write A and B for the two sums in the display and $m = t - t_1$, so that $\mathrm{Cov}(f,F)^2 \le 4(\sqrt{A} + \sqrt{mB})^2 \le 4(1+m)(A+B)$. For $F = \sum_k a_k x_k$ we have $c_k = a_k$ and $\mathbb{E}\mu'_k = \delta_k - \delta_k^{(1)}$, so $A + B = \sum_k a_k^2\delta_k$ and $\mathrm{Cov}(f,F)^2 \le 4(t - t_1 + 1)\sum_k a_k^2\delta_k$; take $a_k = \widehat{f}(\{k\})/\delta_k$. If every μ'_k is constant, then $B = \sum_S \widehat{F}(S)^2 \sum_{k\in S}\omega_k(S)^2(\delta_k - \delta_k^{(1)}) \le \sum_S \widehat{F}(S)^2(\delta(S) - \delta^{(1)}(S))$, so $\mathrm{Cov}(f,F)^2 \le 4(t - t_1 + 1)\sum_k \delta_k \mathrm{Inf}_k(F)$, and Proposition M.13(a) gives the bound on Λ. For $t_1 = t$ this is Proposition M.16; for $t_1 = 1$ it gives Theorem M.11 and Corollary M.15.

(b) *Obstruction* [*proved here, audited*]. The proof of Proposition M.6 bounds the larger quantity $\sum_k \mathrm{Inf}_k(f)/\delta_k \ge \Lambda(f)$ (Proposition M.13(b)), by decoupling, with hypercontractivity, the pointwise magnitude of the later query from a sensitivity weight. This route cannot give $C(r)\, t$, already for $r = 2$. The classical address function that reads t_1 address bits in one round and the addressed bit in a second round, simulated as in Remark M.2(d), has $\mathrm{Inf}_k = \delta_k = \frac{1}{2}$ for the address bits and $\mathrm{Inf}_k = 2\delta_k = 2^{-t_1}$ for the 2^{t_1} data bits, so $\sum_k \mathrm{Inf}_k(f)/\delta_k = 2^{t_1+1} + t_1$. For a data bit, the pointwise magnitude of the second-round query and $(d_k f)^2$ are multiples of the indicator of the event that the address points to it, which has probability 2^{-t_1}; so the mean of their product is 2^{t_1} times the product of their means, and decoupling them loses the factor 2^{t_1}. Yet $\Lambda(f) \le 2t$ by Proposition M.13(c), and $\Lambda(f) \le 3.1$ for $t_1 \le 4$ (Section M.10).

(c) *Evidence* [*numerical*]. Exact computations with hill-climbing over two- and three-round algorithms with one or two queries per round ($n \le 5$) never gave $\Lambda > 4r$; the largest values found were 3.999 for $r = 2$ and 6.84 for $r = 3$.

(d) *Consequence* [*proof outline*]. A restriction of an r-round bias is an r-round bias with the same t_j, by the argument in the proof of Corollary M.17; for the rounds $j \ge 2$, the known signs and the routing are absorbed into the unitaries before and after round j, as in Proposition M.12(c). So $\Lambda \le C(r)\,t$ would give $\mathrm{MaxInf}(h) \ge \mathrm{Var}(h)^2/(C(r)t^2)$ for every restriction h, and the proof of Corollary M.17 with $\gamma = v_0^2/(C(r)t^2)$ would simulate r-round algorithms with $\lceil C(r)t^3/(\varepsilon^4\delta^3)\rceil$ queries. This compares with $2^{O(r^2)}(t\log(1/\delta)/\varepsilon)^{O(r)}$ queries in [13, Theorem 1] and $(rt/(\varepsilon\sqrt{\delta}))^{O(4^r)}$ in [33, Theorem 1.8]: the exponent of t would not grow with r, although the dependence on δ would be worse than in [13]. Since $r \le t$, $C(r) = \mathrm{poly}(r)$ would give the polynomial simulation of all quantum query algorithms, which is open (cf. Proposition M.12); already $C(r) = 2^{O(r)}$ would give $\mathrm{poly}(t, 1/\varepsilon, 1/\delta)$ queries for $r = O(\log t)$.

By Theorem M.14, the spectral form of Conjecture M.3 reduces to an inequality about the profile alone:

Conjecture M.19 (Conjecture K; [*open*])**.** *There is an absolute constant C such that $\kappa(\mu) \le C$ for the magnitude profile μ of every quantum query algorithm.*

Conjecture M.19 implies $\Lambda(f) \le 4Ct$, hence QOSSS with constant $4Ct$, which is optimal in its dependence on t by Remark M.4. By Proposition M.12 it is at least as hard as Conjecture 1.1 for biases. It is a weighted Poincaré inequality for one vector of nonnegative functions μ_k not depending on x_k, with $\sum_k \mathbb{E}\mu_k \le t$, and non-constant profiles are the only obstruction (Corollary M.15). Single characters $h = x^S$ give ratio at most 1, since $\mathbb{E}[1/\mu_k] \ge 1/\delta_k$ by Jensen's inequality. *Evidence* [*numerical*]: $\kappa(\mu)$ is the top eigenvalue of a generalized eigenproblem, computed exactly for each profile. Hill-climbing over random circuits gave $\kappa \le 1.13$, 1.16, 1.13 at $(t, n) = (2, 3..5)$, $(3, 3..5)$, $(4, 4)$ and $\kappa \le 1.08$, 1.06 at $(2, 6)$, $(3, 6)$. The profile of the classical address function has $\kappa = 1$. The extremal families of Remark M.4 have $\kappa = 1.003$–1.013 for $n = 6, \dots, 10$ (decreasing in n), so Theorem M.14 is asymptotically tight on them. On random circuits the ratio $\Lambda/(4t\kappa)$ depends on the ensemble: the largest values observed were 0.58 at $t = 1$ (0.89 for circuits started near the null state), 0.08 (0.21) at $t = 2$ and 0.04 (0.08) at $t = 3$.

M.10. **Evidence and false variants.** *Searches* [*numerical*]. Independent adversarial searches covered hill-climbing over general circuits ($t \le 4$, small n and w, including near-null starts) and structured families: address functions, phase kickback, symmetric Grover-type and quantum signal processing algorithms, block-symmetric algorithms, and Forrelation [1]. None of them found a violation of Conjecture M.3. The largest value of $R(f) = \mathrm{Var}(f)^2/(t\sum_k \delta_k \mathrm{Inf}_k(f))$ found is 4, approached only in limits (small off-null mass, $n \to \infty$), as in Remark M.4; R shows no growth in n or t. Symmetric optima are local maxima of R among all algorithms with $w = 2$. For one-query Forrelation $R = 1$, and for k-fold Forrelation with the standard $\lceil k/2\rceil$-query algorithm $R = 1/t^2$ (exact computations). The quantity Λ stays bounded on all examples: 2.67 to 3.10 for address functions with at most four address bits, 2 for phase kickback, 1 for Forrelation, and $4T - 2$ for Grover with a check query.

False variants. The following natural strengthenings or substitutes fail.

(i) *Final-state-only forms* [*numerical*]. Replace δ_k by the final-state quantity $W_k = \mathrm{Inf}_k(\psi_t) \le \delta_k$ (Lemma M.9); such a form would depend only on ψ_t and M. It is false at $t = 2$. In a two-query family on m variables, the first query writes the signal and some garbage, a swap parks the signal on the null index, and the second query uncomputes the garbage. This gives $f = 2\eta\sqrt{1-\eta^2}\,\bar{x}$, $\delta_k \approx 2\eta^2/m$ and $W_k \approx 3\eta^2/m^2$. For $\eta = 0.2$ and $m =$

$2, 4, \dots, 12$, $\mathrm{Var}(f)^2/\sum_k W_k\mathrm{Inf}_k(f)$ takes the values $3.8, 6.1, 8.6, 11.2, 13.7, 16.3$ (growing like $4(1-\eta^2)m/3$), while $\Lambda \le 2.6$. Garbage that has been uncomputed is invisible in the final state, but its magnitude is paid in δ, so the conjecture must depend on the query history.

(ii) *Operator-valued forms* [*proved here, audited*]. The analogue for the state itself fails at $t = 1$. For $\psi_0 = n^{-1/2}\sum_k e_k$, the vector-valued function $F = O_x\psi_0$ has $\mathrm{Var}(F) = 1$, $\mathrm{Inf}_k(F) = 1/n$ and $\delta_k = 1/n$, so $\mathrm{Var}(F)^2/\sum_k \delta_k\mathrm{Inf}_k(F) = n$. The same happens for $|\psi_t\rangle\langle\psi_t|$ and for the corresponding operator-valued covariance inequality. The scalar measurement must come before the variance.

(iii) *Measured-query forms* [*proved here, audited*]. Measuring the query register (with an XOR oracle) produces a randomized decision tree whose revealments are the measured magnitudes, and (18) applies to that tree. But such trees cannot approximate f in L^2 with comparable revealments. For $f = 2\sqrt{\varepsilon(1-\varepsilon)}\,x_1$ as above ($\delta_1 = \varepsilon$), a randomized tree computing f_D with $|f_D| \le 1$ and $\mathbb{E}(f - f_D)^2 \le \mathrm{Var}(f)/4$ has revealment $r_1 \ge |\mathbb{E}[x_1 f_D]| \ge |\widehat{f}(\{1\})|/2 = \sqrt{\varepsilon(1-\varepsilon)}$. In the norm $\|g\|_\Lambda^2 = \sum_{S\neq\emptyset}\hat{g}(S)^2/\delta(S)$, dual to $(\sum_k \delta_k\mathrm{Inf}_k(\cdot))^{1/2}$ by Proposition M.13(a), the existence of a tree with $r \le C\delta$ and $\|f - f_D\|_\Lambda = O(\sqrt{t})$ is equivalent to $\Lambda(f) = O(t)$. One direction takes f_D constant; the other uses the triangle inequality and the proof of Proposition M.13(c).

Appendix N. Numerical evidence

This appendix summarizes numerical searches and exact computations carried out as evidence. Results tagged [*numerical*] are not proofs. Each function in the table was evaluated exactly (exhaustively over the cube or by an exact linear program) unless marked Monte Carlo. The outcomes are extremes over the searched populations: a reported minimum is an upper bound on the true infimum over the class, a reported worst case of an adversarial search is a lower bound on the true supremum, and LP optima obtained by local or iterated ascent are lower bounds. They are evidence, not proofs, and small sizes cannot decide asymptotics. Table 8 summarizes the searches.

search	families	outcome
extremal functions (about 1800 functions)	symmetric optima up to $d = 31$, Chebyshev compositions, block-symmetric, small-n LP populations, address averages	$\min d^2\mathrm{MaxInf}/\mathrm{Var}^2 \approx 1$ (odd Chebyshev, $\mathrm{Var} \to 0$); flat in d at fixed Var; $16/9$ without level-1 weight
restriction amplification	same families, $p \in \{\frac{1}{2}, \frac{1}{d}\}$	$\Pr[K \ge 4] \le 0.11$, $\Pr[K \ge 8] \le 0.019$, $\Pr[K \ge 16] \le 3 \cdot 10^{-4}$ for the gain K of MaxInf
shared-variable soft addresses (145 instances)	Jackson-threshold, Chebyshev and soft-Tribes bits, $n \le 256$ (Monte Carlo for $n \ge 36$)	$\min \max(\varepsilon, \tau)\deg^2/V^2 = 46.9$; fitted $\tau \sim \deg^{-\alpha}$, $\alpha \in [1.69, 2.18]$
product soft addresses, $n = 12$	disjoint, shared, mixtures, biased	$\max(\varepsilon, \tau) \ge 0.022$ throughout
AN, $n \le 10$	Majority, Tribes, $\mathrm{Maj}_3 \circ \mathrm{Maj}_3$, address, mod-3, random; adversarial E with $\mu(E) \le 0.12$	E never lowers b in structured families; worst $D_{0.05}/b \approx 2$

Table 8. Summary of the numerical searches.

The Tribes-symmetric warning and its resolution. Exact $S_w \wr S_M$-invariant linear programs (the optimum may be symmetrized without loss) for bounded degree-d approximations p of balanced Tribes gave the following L^2-optimal values.

h	n	d	$\beta = \mathbb{E}(1-p^2)/\mathrm{Var}p$	$\mathrm{MaxInf}(p)$	$\mathrm{MaxInf}\, d^2/\mathrm{Var}^2$
$\mathrm{Tribes}_{3,5}$	15	5 / 6 / 8	0.33 / 0.078 / 0.021	0.109 / 0.142 / 0.145	4.9 / 6.0 / 9.7
$\mathrm{Tribes}_{4,11}$	44	8 / 10 / 12	0.23 / 0.055 / 0.020	0.054 / 0.065 / 0.065	5.3 / 7.2 / 9.8
$\mathrm{Tribes}_{5,22}$	110	8 / 10 / 12	1.04 / 0.33 / 0.088	0.012 / 0.023 / 0.031	3.7 / 4.1 / 5.3

The degree at which p becomes nearly Boolean ($\beta \approx 0.09$) was 6, about 9.3, and 12 for $w = 3, 4, 5$, consistent with a linear law $d \approx 3w$ while $\mathrm{MaxInf}(p) \approx 2^{-w}$; if that law persisted it would refute NB_β and the conjecture. Theorem B.1 applies exactly to this symmetry class, and the variance step of its proof excludes it asymptotically: $\mathrm{Var}p \geq c$ forces $M \leq 40d^2(w+1)/c$, i.e. $d \geq \sqrt{cM/(40(w+1))} \approx 2^{w/2}/\sqrt{w}$ for $M \approx 2^w \ln 2$, so the linear law must fail for large w (at $w \leq 5$ the bound is not yet restrictive). In the computed range every optimum has $\mathrm{MaxInf} \geq 1.19\mathrm{Var}^2/d^2$.

Remark N.1 (on the factor t). For f and h both computed by D, whether the factor t in the L^2 form can be replaced by an absolute constant is a question of [42, §5]; by [42, Figure 1] it cannot be replaced by 1. In the two-function form (L2), where h is arbitrary, some growth in t appears to be necessary: for the "caterpillar" tree (query $z_1, z_2, \dots$ in order and stop at the first -1, depth t) the optimal constant $\lambda(t) = \sup_h \mathbb{E}\|\mathbb{E}[h \mid \text{leaf}] - \mathbb{E}h\|^2 / \sum_i \delta_i \mathrm{Inf}_i(h)$ equals 1.333, 1.676, 2.022, 2.368, 2.715, 3.061, 3.408, 3.754, 4.101 for $t = 2, \dots, 10$ (exact generalized-eigenvalue computation), with increments approaching $(\ln 2)/2$, whereas for leaf-determined h the same computation gives 1. Here $2\sum_i \delta_i = 4 - 2^{2-t}$, and $\lambda(10) = 4.101 > 3.996 = 2\sum_i \delta_i$, so in (L2) the factor t cannot be replaced by $\sum_i \delta_i$ [*numerical*]; if $\lambda(t) \to \infty$, as the increments suggest, it cannot be replaced by any constant multiple of the expected depth. Whether $\lambda(t) \to \infty$ is [*open*].

Appendix O. Formal verification

This appendix describes the Lean 4 formalization. For each formalized result it names the Lean theorem that proves it. The project `lean/AAFormal` consists of twenty-seven Lean files and a separate axiom-check file `AxiomCheck.lean` (about 8500 lines in total), formalized in Lean 4 with Mathlib (Lean `v4.36.0-rc2`, Mathlib tag `v4.36.0-rc2`); all results below are unconditional. The file `Main.lean` (namespace `Paper`) restates every formalized result under its paper label (for instance `Paper.cor_H1`, `Paper.thm_D5`, `Paper.thm_W`, `Paper.thm_TTsandwich`, `Paper.lem_NS`), each as a short application of the theorem that proves it, with the hypotheses $|f| \leq 1$ and Fourier degree at most d of $f \in \mathcal{B}_d$.

- `DimBound.christoffel_dim_bound`: for a finite probability space Ω, a subspace W of $\mathbb{R}^\Omega$ and $g_j \in W$ with $\sum_j |g_j(\omega)| \leq 1$ pointwise, $\sum_j \|g_j\|_{L^2} \leq \sqrt{\dim W}$, the case $\sup \sum_j |g_j| \leq 1$ of Proposition 4.12;
- `SubcubeCover.scl_iSup`: $\Sigma^2/M - \sum_j \mu_j^2 \leq r\,(\sup_i \mathrm{load}_i)\,\Sigma$, i.e. Lemma 5.1, built from `SubcubeCover.Subcube.helly` (pairwise intersecting subcubes have a common point) and `motzkin_straus` (the Motzkin–Straus inequality for graphs with clique number at most M).

The files `Gluing.lean`, `OSSS.lean` and `Resample.lean` formalize the gluing lemma (Theorem 4.2) on the cube $\kappa \to$ `Bool` with an inductive type of decision trees and their revealments. The vector OSSS inequality (L2) of Theorem 4.1 for $H = \mathbb{R}^\iota$ is proved for every tree as `Gluing.osss_l2` (via the resampling bound `Gluing.resSq_div_le`), and

- `Gluing.gluing_full'`: $\kappa_D - \varepsilon V \leq B_q \leq 8t \sum_i \delta_i \mathrm{Inf}_{z_i}(F) \leq 8t^2 \tau_D$, with $B_q = \mathbb{E}\sum_j (\mathbb{E}[g_j^2 \mid \text{leaf}] - \mathbb{E}g_j^2)^2$, for every $\varepsilon \geq \max_j \mathbb{E}g_j^2$ and $\tau \geq \tau_D$.

(The earlier `Gluing.gluing_full` takes the OSSS inequality as an explicit hypothesis.) Supporting lemmas: `kappa_sub_le_Bq`, `Bq_eq`, `kappa_ge` ($\kappa_D \ge V^3/R^2$ from leaf root sums at most R), `vecInf_sq_le` and `sum_reveal_le` ($\sum_i \delta_i \le t$). Corollary 4.3 and Theorem 4.2 with leaf root sums are `Paper.cor_glueR` and `Paper.thm_glue`, proved via `Gluing.gluing_max` and `Gluing.gluing_final`.

The files `GluingIneq.lean`, `SpikeAvg.lean`, `Walsh.lean` and `Degree.lean` add degree-free forms of Theorems 4.2, 4.4 and 4.5 (`Gluing.gluing_final`, `Gluing.gluing_ineq`, `Gluing.closure`), a Walsh–Fourier layer (orthonormality of characters, the expansion $f = \sum_S \widehat{f}(S)\chi_S$ and $\mathrm{Inf}_i(f) = \sum_{S\ni i} \widehat{f}(S)^2$), and:

- `Gluing.gluing_ineq_leaf_of_degLe`: Theorem 4.4 with no further hypotheses, for every f with $|f| \le 1$ and Fourier degree at most $d \ge 1$: $\mathbb{E}_l[\mathrm{MaxInf}(f|_l)^2] \le 144 d^5 t^2 M$ for every decision tree of depth $t \ge 1$ and every $M \ge \mathrm{MaxInf}(f)$. The degree enters through four facts, all proved in Lean: Parseval (`Gluing.parseval`); $I(f) \le d\mathrm{Var}f$ (`Gluing.sum_infl_le`); $\sum_j \mathrm{Inf}_i(d_j f) \le (d-1)\mathrm{Inf}_i(f)$ (`Gluing.sum_infl_bderiv_le'`, via $d_i d_j = d_j d_i$); and the real-sensitivity bound $\sum_j |d_j f(z)| \le 2d^2$ (`Gluing.sum_abs_bderiv_le`). The last one restricts the multilinear extension of f, which is bounded by 1 on $[-1,1]^n$, to the line $s \mapsto (s\,\mathrm{sgn}(z_i))_{i\in P}$ and applies the endpoint Markov inequality $P'(1) \le n^2$ from Mathlib (`Gluing.markov_endpoint`), once for $P = \{i : \mathrm{sgn}(z_i)\, d_i f(z) > 0\}$ and once for its complement. Its constant $2d^2$ is better than the $4d^2$ of Lemma 2.2 (the sharp constant is d^2, [19, Theorem 3.3]).
- `Gluing.condExp_bderiv_sq_eq_infl`: the restriction identity $\mathbb{E}[(d_j f)^2 \mid l] = \mathrm{Inf}_j(f|_l)$ for j not queried on the path to the leaf l.
- `Gluing.tt_of_aa_spikes`: the direction "conjecture ⇒ TT" of Theorem 8.7(ii), degree-free: if $\mathrm{MaxInf}(u) \ge c(\mathrm{Var}u/d)^A$ on a class of functions closed under averaging over coordinates, $0 < A \le C$, $\mathrm{Var}g \le 1$, $\theta \le \mathrm{Var}(g)^C$ and $L \ge (d/\varepsilon)^A/c$, then the average h of g over its θ/L-spikes has $\mathrm{Var}h \le \varepsilon \mathrm{Var}g$ (that is, $U_{\theta/L}(g) \le \varepsilon \mathrm{Var}g$).

`TopHeavy.lean` adds:

- `Gluing.bonami`: Bonami's lemma $\mathbb{E}g^4 \le 9^k(\mathbb{E}g^2)^2$ for real g of degree at most k (induction on the Fourier support);
- `Gluing.levelk`: the level-k inequality of Lemma 3.1(b), $W^{\le k}(g) \le 3^k \|g\|_1 \|g\|_2$ for every real g (two Cauchy–Schwarz steps and Bonami's lemma, avoiding $L^{4/3}$ norms);
- `Gluing.top_heavy_sq`: top-heaviness (Corollary 3.3) in the form $\mathrm{MaxInf}(f) \ge (W^{[1,k+1]}(f)/(2d^2\,3^k))^2$ for $|f| \le 1$ of degree at most $d \ge 1$ and all $k \ge 0$; because of the Lean sensitivity constant $2d^2$ this is 4 times stronger in MaxInf than the paper's bound with $4d^2$;
- `Gluing.closure_degLe_paper`: the closure Theorem 4.5 for $|f| \le 1$ of degree at most d with the paper's leaf hypothesis $c_0(\mathrm{Var}(f|_l)/d)^C \le \mathrm{MaxInf}(f|_l)$ at every leaf ($2C \ge 1$, $c_0 \ge 0$), concluding $\mathrm{MaxInf}(f) \ge \min\big(\mathrm{Var}f/(4t^2),\, c_0^2 d^{-2C}(\mathrm{Var}f/2)^{2C}/(d + 32t^2(d-1)d^4)\big)$.

`Restriction.lean`, `Spikes.lean` and `Chain.lean` formalize the random-restriction framework of the touching transfer. A rate-p restriction is a pair $\rho = (A, y)$ and $\mathbb{E}_\rho$ is a finite weighted sum. The files prove: the restricted spectra $\mathbb{E}_y \hat{f}_\rho(T)^2 = \sum_{S\cap A = T} \widehat{f}(S)^2$; $\mathbb{E}\mathrm{Var}(f_\rho) = \sum_S (1-(1-p)^{|S|})\widehat{f}(S)^2 \in [p\mathrm{Var}f, \mathrm{Var}f]$ and $\mathbb{E}\mathrm{Inf}_i(f_\rho) = p\mathrm{Inf}_i(f)$ (`Gluing.rexp_var_restrict`, `Gluing.rexp_infl_restrict`); preservation of $|f| \le 1$ and of degree; composition of restrictions (`Gluing.rexp_rexp_restrict`); the tail bound $\mathbb{E}W^{>m}(f_\rho) \le \mathrm{Var}f\,(edp/(m+1))^{m+1}$; spikes and untouched variance; the junta bound from Corollary 3.3; and a strengthening of the base case Lemma 8.3 (`Gluing.lemma21`, with exponent 2 and $4d^2$). `Gluing.TT` is Definition 8.2 for one fixed d. `Gluing.chain_maxInf` is the chain Theorem 8.4: under

$\mathrm{TT}(1/(16ed), L)$ with cap C, with $p = 2^{-T}$, $2ed \le 2^T \le 4ed$ and $d \le \eta p 2^{k+1}$ ($\eta = 1/(16ed)$), it gives $\mathrm{MaxInf}(f) \ge \theta_T L^{-T}$ with $\theta_T = \min\left((\eta p \mathrm{Var} f/(4d^2 3^k))^2, (p\mathrm{Var} f/(16ed))^C\right)$. This improves the base-case exponent from 4 to 2 (and $8d^2$ to $4d^2$). Admissible parameters always exist (`Gluing.exists_chain_params`). `Gluing.aa_imp_maxInf` composes `tt_of_aa_spikes` with the chain, giving the sandwich of Theorem 8.7 in formal form. `NewSpike.lean` formalizes Lemma I.9 (`Gluing.lemma_ns`), with no hypothesis on g, θ or L. `Classes.lean` formalizes two corollaries: Corollary 5.12(8) for one-statistic derivatives, in the stronger form $\mathrm{MaxInf}(f) \ge I(f)^2/(4d^5)$ (`Gluing.sq_sum_infl_le_of_statistic`, via the subspace form $I(f)^2 \le 4d^4 \dim \mathcal{W} \cdot \mathrm{MaxInf}(f)$); and Corollary 5.2, both with the paper's constant $\mathrm{MaxInf}(f) \ge c\, I(f)/(32d^4)$ (`Gluing.scl_aa`) and with $I(f)^2/(64d^6)$ in place of the paper's $128d^6$ (`Gluing.scl_aa_sq`); their paper-facing forms are `Paper.cor_classes_8` and `Paper.cor_SCLAA_sq`. `BlockSym.lean` proves Markov's inequality with a factor 2, $|P'| \le 2n^2 \sup_{[-1,1]} |P|$ on $[-1,1]$ for $\deg P \le n$ (`Paper.markov_two`), and grid discretization: if $\deg P \le n$, $N \ge 4n^2$ and $|P| \le 1$ on $\{1 - 2k/N : 0 \le k \le N\}$, then $|P| \le 2$ on $[-1,1]$ (`Paper.grid_two`). `Symmetrize.lean` proves the Minsky–Papert symmetrization [40] (`Paper.minsky_papert`): if f has degree at most d and is invariant under the transpositions of coordinates inside B, then on every slice (the coordinates outside B fixed) $f = p(\sum_{j\in B} x_j)$ with $\deg p \le d$; the proof is by induction on d, with a discrete antiderivative. Together they give Theorem 6.4 and Remark 6.1 unconditionally, with constant 16 in place of 4: `Paper.thm_D6_perm` gives $\mathrm{Var}(f)^2 \le I(f)^2 \le 16b^2d^4\mathrm{MaxInf}(f)$ for $f \in \mathcal{B}_d$ invariant under every permutation of the coordinates that preserves each of b blocks, and `Paper.rem_symmetric_perm` gives $\mathrm{Var}(f)^2 \le I(f)^2 \le 16d^4\mathrm{MaxInf}(f)$ for symmetric $f \in \mathcal{B}_d$. The factor 4 comes from the factor 2 in Markov's inequality; the sharp bound of Proposition 6.2 is not formalized.
Theorem B.1 is formalized, with a weaker constant, in nine files (`Bernstein.lean`, `Fisher.lean`, `EfronStein.lean`, `Structure.lean`, `BlockPoly.lean`, `GridRound.lean`, `WreathInfl.lean`, `OccPoly.lean` and `WreathW.lean`): `Paper.thm_W` gives $\mathrm{Var}(f)^2 \le 8(\pi e(d+1))^2 w(w+1)\mathrm{MaxInf}(f)$ for $f \in \mathcal{B}_d$ on $\{\pm 1\}^{M\times w}$ ($M, w \ge 1$) invariant under $S_w \wr S_M$, in place of $40d^2w(w+1)$. The losses are Bernstein's constant $e(n+2)$ in place of n (`Paper.bernstein_trig`, hence $\pi e(d+1)$ in place of πd in the Fisher-metric step `Paper.fisher_bernstein`), the rounding threshold $8L^2(w+1)$ in place of $4L^2(w+1)$ (`Paper.grid_rounding`), and a factor 2 in the re-randomization bound. The steps also include the multi-block Minsky–Papert theorem (`Paper.minsky_papert_blocks`) and Efron–Stein over blocks, $\mathrm{Var} f \le \sum_l \mathbb{E}(f - \mathbb{E}_{B_l} f)^2$ (`Paper.efron_stein_blocks`). Theorems 6.5, 6.6 and 6.8 are not formalized.

No file contains `sorry`, and `#print axioms` reports only the standard axioms `propext`, `Classical.choice` and `Quot.sound`. The file `AxiomCheck.lean` checks this for every `Paper.*` theorem (`lake env lean AAFormal/AxiomCheck.lean`; 38 checks). To build, run

```
cd lean/AAFormal; lake exe cache get; lake build
```

(about 15 seconds with the Mathlib cache).

Acknowledgment

This work was carried out with the assistance of Grok Bot and Cursor agents.

References

1. S. Aaronson and A. Ambainis, *Forrelation: a problem that optimally separates quantum from classical computing*, SIAM J. Comput. **47** (2018), no. 3, 982–1038, Conference version in STOC 2015, 307–316; arXiv:1411.5729.

2. Scott Aaronson and Andris Ambainis, *The need for structure in quantum speedups*, Theory of Computing **10** (2014), 133–166, Preliminary version in ICS 2011; arXiv:0911.0996.
3. *Analysis on the hypercube with applications to quantum computing*, American Institute of Mathematics workshop, problem list at aimpl.org/hypercubequantum.
4. Srinivasan Arunachalam, Jop Briët, and Carlos Palazuelos, *Quantum query algorithms are completely bounded forms*, SIAM Journal on Computing **48** (2019), no. 3, 903–925, arXiv:1711.07285.
5. A. Bačkurs, *Influences in low-degree polynomials*, 2012, MIT 6.845 student project report, 12 December 2012.
6. Arturs Bačkurs and Mohammad Bavarian, *On the sum of $L1$ influences*, IEEE 29th Conference on Computational Complexity (CCC 2014), 2014, arXiv:1302.4625.
7. Nikhil Bansal, Makrand Sinha, and Ronald de Wolf, *Influence in completely bounded block-multilinear forms and classical simulation of quantum algorithms*, 37th Computational Complexity Conference (CCC 2022), LIPIcs, vol. 234, 2022, arXiv:2203.00212, pp. 28:1–28:21.
8. M. Baran, *Bernstein type theorems for compact sets in $\mathbb{R}^n$ revisited*, J. Approx. Theory **79** (1994), 190–198.
9. Robert Beals, Harry Buhrman, Richard Cleve, Michele Mosca, and Ronald de Wolf, *Quantum lower bounds by polynomials*, Journal of the ACM **48** (2001), no. 4, 778–797.
10. Shalev Ben-David, Andrew M. Childs, András Gilyén, William Kretschmer, Supartha Podder, and Daochen Wang, *Symmetries, graph properties, and quantum speedups*, 61st Annual IEEE Symposium on Foundations of Computer Science (FOCS 2020), 2020, arXiv:2006.12760.
11. C. H. Bennett, E. Bernstein, G. Brassard, and U. Vazirani, *Strengths and weaknesses of quantum computing*, SIAM J. Comput. **26** (1997), no. 5, 1510–1523, arXiv:quant-ph/9701001.
12. S. K. Bhattacharya, *Random restrictions of bounded low degree polynomials are juntas*, 16th Innovations in Theoretical Computer Science Conference (ITCS 2025), LIPIcs, vol. 325, 2025, Full version: "Aaronson–Ambainis Conjecture Is True For Random Restrictions", arXiv:2402.13952 and ECCC TR24-035; theorem numbers refer to ECCC TR24-035 rev. 1, pp. 17:1–17:21.
13. Guy Blanc, Jordan Docter, Carmen Strassle, and Li-Yang Tan, *Quantum speedups require structure or depth*, Proceedings of the 67th IEEE Symposium on Foundations of Computer Science (FOCS 2026), 2026, arXiv:2608.19158.
14. André Chailloux, *A note on the quantum query complexity of permutation symmetric functions*, 10th Innovations in Theoretical Computer Science Conference (ITCS 2019), LIPIcs, vol. 124, 2019, arXiv:1810.01790, pp. 19:1–19:7.
15. Andreas Defant, Mieczysław Mastyło, and Antonio Pérez, *On the Fourier spectrum of functions on Boolean cubes*, Mathematische Annalen **374** (2019), 653–680, arXiv:1706.03670.
16. Irit Dinur, Ehud Friedgut, Guy Kindler, and Ryan O'Donnell, *On the Fourier tails of bounded functions over the discrete cube*, Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC), 2006, Journal version: Israel J. Math. 160 (2007) 389–412, pp. 437–446.
17. H. Ehlich and K. Zeller, *Schwankung von Polynomen zwischen Gitterpunkten*, Mathematische Zeitschrift **86** (1964), 41–44.
18. P. Erdős, L. Lovász, and J. Spencer, *Strong independence of graphcopy functions*, Graph Theory and Related Topics, 1979, pp. 165–172.
19. Yuval Filmus, Hamed Hatami, Nathan Keller, and Noam Lifshitz, *On the sum of the L1 influences of bounded functions*, Israel Journal of Mathematics **214** (2016), no. 1, 167–192, arXiv:1404.3396.
20. David A. Freedman, *On tail probabilities for martingales*, Ann. Probab. **3** (1975), 100–118.
21. Uma Girish, Kunal Mittal, Barak Nehoran, and Ran Raz, *Distributional variants of the Aaronson–Ambainis conjecture*, 2026, arXiv:2609.35327.
22. L. K. Grover, *A fast quantum mechanical algorithm for database search*, Proceedings of the 28th Annual ACM Symposium on Theory of Computing (STOC), 1996, pp. 212–219.
23. Francisco Escudero Gutiérrez, *Influences of Fourier completely bounded polynomials and classical simulation of quantum algorithms*, Chicago Journal of Theoretical Computer Science (2024), Art. 2, arXiv:2304.06713.
24. Francisco Escudero Gutiérrez, Mario Saucedo, and Carlos Palazuelos, *Optimal inequalities for completely bounded polynomials and the limitations of quantum query algorithms*, 2026, arXiv:2609.05201.
25. Hao Huang, *Induced subgraphs of hypercubes and a proof of the Sensitivity Conjecture*, Annals of Mathematics **190** (2019), no. 3, 949–955.
26. Paata Ivanisvili, *Aaronson–Ambainis conjecture*, 2019, Blog post, *Zeros and Ones*, 29 October 2019.
27. ———, *Polynomial growth of Bohnenblust–Hille constants on the Hamming cube*, 2026, arXiv:2609.12427.
28. Jeff Kahn, Gil Kalai, and Nathan Linial, *The influence of variables on Boolean functions*, Proceedings of the 29th Annual Symposium on Foundations of Computer Science (FOCS), 1988, pp. 68–80.

29. Nathan Keller and Ohad Klein, *Quantum speedups need structure*, 2019, arXiv:1911.03748; withdrawn by the authors in v2.
30. O. D. Kellogg, *On bounded polynomials in several variables*, Math. Z. **27** (1928), 55–64.
31. A. Kriebel, *Polynomial compatibility on the Boolean cube from disjoint-support separators*, 2026, Zenodo preprint, doi:10.5281/zenodo.23443300, unrefereed.
32. Longcheng Li, Qian Li, Xingjian Li, and Qipeng Liu, *Impossibility of perfectly complete many-round key agreement in the QROM*, 2026, arXiv:2608.03824.
33. Qipeng Liu and Saachi Mutreja, *Parallel quantum advantage with limited adaptivity requires structure*, 2026, arXiv:2608.20297.
34. L. Lovász and B. Szegedy, *Limits of dense graph sequences*, J. Combin. Theory Ser. B **96** (2006), no. 6, 933–957, arXiv:math/0408173.
35. Shachar Lovett and Jiapeng Zhang, *Fractional certificates for bounded functions*, 14th Innovations in Theoretical Computer Science Conference (ITCS 2023), LIPIcs, vol. 251, 2023, ECCC TR22-107, pp. 84:1–84:13.
36. V. A. Markov, *Über Polynome, die in einem gegebenen Intervalle möglichst wenig von Null abweichen*, Math. Ann. **77** (1916), 213–258.
37. Gatis Midrijanis, *Exact quantum query complexity for total Boolean functions*, 2004, arXiv:quant-ph/0403168.
38. Ashley Montanaro, *Some applications of hypercontractive inequalities in quantum information theory*, Journal of Mathematical Physics **53** (2012), 122206, arXiv:1208.0161.
39. T. S. Motzkin and E. G. Straus, *Maxima for graphs and a new proof of a theorem of Turán*, Canadian Journal of Mathematics **17** (1965), 533–540.
40. Noam Nisan and Mario Szegedy, *On the degree of Boolean functions as real polynomials*, Computational Complexity **4** (1994), 301–313.
41. Ryan O'Donnell, *Analysis of Boolean functions*, Cambridge University Press, 2014.
42. Ryan O'Donnell, Michael Saks, Oded Schramm, and Rocco A. Servedio, *Every decision tree has an influential variable*, Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS), 2005, arXiv:cs/0508071, pp. 31–39.
43. Ryan O'Donnell and Yu Zhao, *Polynomial bounds for decoupling, with applications*, 31st Conference on Computational Complexity (CCC 2016), LIPIcs, vol. 50, 2016, arXiv:1512.01603; theorem numbers refer to the arXiv version, pp. 24:1–24:18.
44. T. J. Rivlin and E. W. Cheney, *A comparison of uniform approximations on an interval and a finite subset thereof*, SIAM Journal on Numerical Analysis **3** (1966), no. 2, 311–320.
45. R. Tyrrell Rockafellar, *Convex analysis*, Princeton University Press, 1970.
46. Y. Sarantopoulos, *Bounds on the derivatives of polynomials on Banach spaces*, Math. Proc. Cambridge Philos. Soc. **110** (1991), 307–312.
47. Joseph Slote, Chun-Kai Tseng, and Alexander Volberg, *An $m^{2.943}$ Bohnenblust–Hille bound on the Boolean cube*, 2026, arXiv:2609.21144, unrefereed.
48. S. J. Szarek, *On the best constants in the Khinchin inequality*, Studia Math. **58** (1976), 197–208.
49. G. Szegő, *Bemerkungen zu einem Satz von J. H. Grace über die Wurzeln algebraischer Gleichungen*, Math. Z. **13** (1922), 28–55.
50. ———, *Über einen Satz des Herrn Serge Bernstein*, Schriften der Königsberger Gelehrten Gesellschaft, Naturwissenschaftliche Klasse **5** (1928), 59–70.
51. Avishay Tal, *On the sensitivity conjecture*, 43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016), LIPIcs, vol. 55, 2016, ECCC TR16-062, pp. 38:1–38:13.
52. Michel Talagrand, *On Russo's approximate zero-one law*, Annals of Probability **22** (1994), no. 3, 1576–1587.
53. J. G. van der Corput and G. Schaake, *Ungleichungen für Polynome und trigonometrische Polynome*, Compositio Math. **2** (1935), 321–361.
54. J. L. Walsh, *On the location of the roots of certain types of polynomials*, Trans. Amer. Math. Soc. **24** (1922), 163–180.
55. F. B. Weissler, *Logarithmic Sobolev inequalities and hypercontractive estimates on the circle*, J. Funct. Anal. **37** (1980), 218–234.
56. D. R. Wilhelmsen, *A Markov inequality in several dimensions*, J. Approx. Theory **11** (1974), 216–220.
57. Mark Zhandry, *How to construct quantum random functions*, 53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS 2012), 2012, Journal version: J. ACM 68(5), 2021.
58. ———, *A note on the quantum collision and set equality problems*, Quantum Information and Computation **15** (2015), 557–567, arXiv:1312.1027.