Rigorous Statements and Proofs of the Lemmas in Simon's Algorithm for the Dihedral Coset Problem and Their Underlying Hypothesis

Yuchen Guo [1,*] and Shuo Yang [1,2,3,†]

[1] *State Key Laboratory of Low Dimensional Quantum Physics and Department of Physics, Tsinghua University, Beijing 100084, China*
[2] *Frontier Science Center for Quantum Information, Beijing 100084, China*
[3] *Hefei National Laboratory, Hefei 230088, China*

(Dated: August 17, 2026)

In a recent preprint, Simon proposed a polynomial-time quantum algorithm for the Dihedral Coset Problem and rested the analysis on four lemmas. Three of them carry only proof sketches, and this paper gives each of those three a statement that admits a single reading together with a complete proof. Lemma 1 follows from an exact second-moment computation for the subset-sum counts, and it holds with probability tending to one in place of the constant originally claimed. The amplitude bound of Lemma 3 follows from an exact Parseval identity on the cube of measurement outcomes and holds at every threshold with no *well-behavedness* hypothesis, so that predicate leaves the argument entirely. For Lemma 4, we compute both balls-in-bins covariances exactly and find that the second carries a term a fixed ball count leaves out. The assumption that the distinguished group contains no faulty samples can also be dropped. The two branch amplitudes share a signed prefactor, so the counting estimates control their difference and not the ratio the lemma states. We prove the additive form and show that the closing argument consumes nothing more than that. A single hypothesis survives all of this. It asks that the partition into the two sides be fixed independently of the measured string, and the rule the algorithm gives for choosing that partition does not supply it. Establishing these four lemmas therefore does not by itself establish the correctness of the algorithm.

I. INTRODUCTION

The Dihedral Coset Problem asks for a hidden shift. One is given a supply of two-term superpositions $|0, \xi\rangle + |1, \xi + d\rangle$ with ξ uniform on $\mathbb{Z}_N$ and d fixed, and the task is to recover d. It is the hidden subgroup problem for the dihedral group, and the standard Fourier method that solves the abelian cases in Shor's algorithm does not carry over [5, 14]. Ettinger and Høyer showed that polynomially many samples already determine d information-theoretically, so the whole difficulty lies in reading d out in polynomial time [6]. The best known algorithms are the subexponential sieves of Kuperberg [7, 8] and Regev [12]. Bacon, Childs and van Dam identified the optimal measurement for the problem and reduced its implementation to a classical subset-sum question [2].

Much of the interest in the problem comes from lattices. Ajtai's worst-case to average-case reduction placed lattice problems at the foundation of a family of cryptosystems [1]. Regev showed that a polynomial-time solver would give a polynomial-time quantum algorithm for the unique shortest vector problem [11], and Brakerski, Kirshanova, Stehlé and Wen extended the reduction to Learning with Errors at polynomial approximation factors [3, 13]. The complexity of the Dihedral Coset Problem therefore sits directly beneath a large part of lattice-based cryptography, including the key-encapsulation scheme now being standardised [9]. A claim of a polynomial-time quantum algorithm in this area invites close reading, and the recent lattice algorithm of Chen received exactly that before a step in it was found not to work [4].

Simon has recently posted a preliminary draft giving a polynomial-time quantum algorithm for the problem [15]. It works on Q samples at once and sorts the Q bits into groups of m. The groups whose measurement outcome is all zero then build an unsigned count that both branches of the final interference share. The analysis rests on four lemmas. Lemma 2 is exact and carries a complete two-line proof, while the other three are marked *Proof (Sketch)*. This paper supplies statements for those three that admit a single reading, together with complete proofs.

Lemma 1 asserts that enough all-zero groups are collected. Section III computes the first and second moments of their number exactly by means of two Fourier transforms, one on the cube of measurement outcomes and one on the group $\mathbb{Z}_N$ of subset sums. The correlation between two groups turns out to be exponentially small, and Chebyshev then gives probability tending to one in place of the constant [15] asks for.

Lemma 3 bounds the individual amplitudes. Section IV obtains the bound from an exact Parseval identity, which needs no independence property of the signs and no *well-behavedness* property either. The bound holds at every threshold, and the probabilities are taken directly on the distribution of records. These proofs ask for the A/B partition to be fixed independently of the measured string, which we state as Hypothesis (U) in Section IV.

Lemma 4 compares the two branch amplitudes, and its proof works on the counts that feed them. The skeleton is a pair of balls-in-bins estimates, and Section V computes the covariance of each one exactly. The first comes

* guo-yc23@mails.tsinghua.edu.cn
† shuoyang@tsinghua.edu.cn

out with the exponent and the failure probability [15] displays, while the second works with a random number of surviving states that the algorithm imposes. We also remove the assumption that the distinguished group g_a contains no faulty samples. Carrying the counts back up to the amplitudes raises a question of form. The two amplitudes of a pair share one signed prefactor, so the counting estimates deliver a bound on their difference while the lemma asks for their ratio. That turns out to cost nothing, because the closing argument only ever adds up differences. Section V E therefore proves the weaker statement and reaches the same exponent from it.

Table I summarizes where each is treated. The rule [15] gives for choosing A does not supply Hypothesis (U), so establishing these lemmas does not by itself establish the correctness of the algorithm. Results proved here are lettered, so a numbered lemma is always one of [15].

TABLE I. The four lemmas of [15] and where each is treated below.

In [15]	What it asserts	Here
Lemma 1	Enough all-zero groups are collected	C
Lemma 2	The B side contributes a common factor	§II
Lemma 3	Individual amplitudes are bounded	E, E′
Lemma 4	The two branch amplitudes are nearly equal	G, H

II. SETUP

A. Problem setup

Write $N = 2^n$ and $\omega = e^{2\pi i/N}$, and let the unknown value be $d = d_1 d_2 \cdots d_n$ with least significant bit d_n. A sample is the state

$$\tfrac{1}{\sqrt{2}}\big(|0, \xi\rangle + |1, \xi + d \bmod N\rangle\big), \tag{1}$$

with ξ uniform and unknown and d fixed. With probability $1/(c' \log n)$ for a constant c', a sample is faulty and is an $(n+1)$-bit classical string instead. Faulty samples play no role before Section V.

Applying the $\mathbb{Z}_N$ Fourier transform to the second register and measuring it returns a value y that is uniform on $\mathbb{Z}_N$ and known to us, and it collapses the first register to $|0\rangle + \omega^{yd}|1\rangle$. Taking Q samples and writing $Y = (y_1, \ldots, y_Q)$ for the measured weights, the tensor product of the Q qubits expands into

$$|\Psi\rangle \propto \sum_{\phi \in \{0,1\}^Q} \omega^{d\, z(\phi)}\, |\phi\rangle, \tag{2}$$

$$z(\phi) = \sum_{i=1}^{Q} \phi_i y_i \bmod N. \tag{3}$$

Every basis state carries the same modulus, so all of the information sits in the phase, and that phase depends on ϕ only through the weighted sum $z(\phi)$.

Computing $z(\phi)$ into an ancilla and measuring its low $n-1$ bits returns a value z' and leaves the top bit $h = \mathrm{msb}(z)$ in superposition, where msb denotes the most significant bit. Since $\omega^{d \cdot 2^{n-1}} = (-1)^{d_n}$, the surviving state is

$$|\Psi\rangle \propto \sum_{h \in \{0,1\}} (-1)^{h\, d_n} \Big(\sum_{\phi \in Z'_h} |\phi\rangle \Big) |h\rangle, \tag{4}$$

$$Z'_h = \big\{ \phi : z(\phi) = z' + 2^{n-1} h \big\}. \tag{5}$$

Here $h(\phi)$ is a function of ϕ and not a free register, and we write $Z' = Z'_0 \sqcup Z'_1$ for the set of all ϕ compatible with the measured z'. The bit d_n now sits in the relative phase of $|h\rangle$. If the register $|h\rangle$ stood alone, one Hadamard and one measurement would return d_n with certainty, and a recursion on the remaining bits of d would finish the problem. However, the two branches are attached to disjoint sets of ϕ, so they occupy orthogonal subspaces and cannot interfere. Regev's way out is to erase the ϕ register with a subset-sum oracle [11]. Simon's way out is to change basis, and the seven steps below carry that out.

B. The algorithm in seven steps

Throughout this paper the *ket* holds the registers still in superposition and the *record* holds the classical numbers already measured. The record is $(Y, z', \phi', S, W', h')$, and any of its entries may be substituted into a formula. Table II tracks both.

TABLE II. What survives each step. Erasures are reversible and produce no record.

After	In the ket	New classical data
Step 1	ϕ	$Y = (y_1, \ldots, y_Q)$
Step 2	ϕ, h	z', the low $n-1$ bits of z
Step 3	ϕ, h, $s_1 \cdots s_G$	none
Step 4	h, $s_1 \cdots s_G$	ϕ', and ϕ is gone
Step 5	h, $s_1 \cdots s_a$	$S = (\sigma'_j)_{j \in B}$
Step 6	h, h^*	$W' = (s_1, \ldots, s_{a-1}, l_{s^*})$
Step 7	h^*	$h' = h \oplus h^*$, and h is erased
Final	none	the Hadamard measurement of h^*

Steps 1 and 2 are the constructions of Section II A, which produce the record entries Y and z' and leave the ket in the state displayed there.

Step 3, the group sums. Partition $\{1, \ldots, Q\}$ into $G = Q/m$ groups $g_1, \ldots, g_G$ of $m = c \log n$ bits each, and write

$$r_j(\phi) = \sum_{i \in g_j} \phi_i y_i \bmod N, \qquad s_j(\phi) = \mathrm{top}\big(r_j(\phi)\big) \tag{6}$$

for the sum within group j and its top $L = \log n$ bits. Each s_j is computed reversibly into its own L-bit register,

giving

$$|\Psi_3\rangle \propto \sum_{\phi\in Z'} (-1)^{h(\phi)d_n}\,|\phi\rangle\,|h(\phi)\rangle\,|s_1(\phi)\cdots s_G(\phi)\rangle. \quad (7)$$

These registers survive the disappearance of ϕ at the next step, which keeps only the top L bits of each r_j.

Step 4, the change of basis. Apply a Hadamard to all Q bits and measure, obtaining a classical string $\phi' = \phi'_1\cdots\phi'_G$, written in blocks by group. The ϕ register is gone, and the amplitude of a surviving label $(h,\vec{s}\,)$ becomes a *signed count* of the ϕ compatible with that label, each one weighted by the sign $(-1)^{\phi\cdot\phi'}$,

$$|\Psi_4\rangle \propto \sum_{h,\vec{s}} \mathcal{E}[h,\vec{s}\,]\,|h,\vec{s}\,\rangle, \quad (8)$$

$$\mathcal{E}[h,\vec{s}\,] = (-1)^{hd_n} \sum_{\substack{\phi\in Z':\ h(\phi)=h\\ s_j(\phi)=s_j\ \forall j}} (-1)^{\phi\cdot\phi'}. \quad (9)$$

Step 5, the A/B split. A group whose measured block ϕ'_j is the all-zero string contributes no sign, since $(-1)^{\phi_j\cdot 0}=1$. Let A consist of the first $a = n/\log n$ such groups and let B hold the rest. Lemma 1 of [15] asserts that there are at least a all-zero groups, and Section III treats that claim. The A side holds $am = cn$ sample bits. Write $\phi=(\phi_A,\phi_B)$ and $\phi'=(\phi'_A,\phi'_B)$ for the strings cut into their A and B parts, so that $\phi'_A=0$ by construction and

$$(-1)^{\phi\cdot\phi'} \;=\; (-1)^{\phi_B\cdot\phi'_B}. \quad (10)$$

Measure s_j for every $j\in B$, obtaining $S=(\sigma'_j)_{j\in B}$, and the ket retains $(h, s_1\cdots s_a)$.

Step 6, manufacturing h^*. Here we introduce a new register that represents the top bit of the A-side partial sum. Write

$$z^*(\phi) = \sum_{i\in A}\phi_i y_i \bmod N, \qquad h^* = \mathrm{msb}(z^*) \quad (11)$$

for the A-side partial sum and its top bit. This step turns h^* into an explicit register in five moves.

1. Compute $s^* = \sum_{j\in A} s_j \bmod 2^L$ into a fresh L-bit register.
2. Erase s_a, which is now redundant, being $s^* - \sum_{j<a} s_j \bmod 2^L$.
3. Compute a check bit l_{s^*} that equals 1 exactly when positions 2 through $\log L$ of s^* are all 1, and post-select on $l_{s^*}=0$. This prohibits the carry from the bottom bits and makes $\mathrm{msb}(s^*) = \mathrm{msb}(z^*) = h^*$. The post-selection succeeds with probability about $1-2/\log n$.
4. Split s^* into its top bit h^* and its low $L-1$ bits $\bar{s}^*$, Hadamard $\bar{s}^*$, measure, and post-select on the all-zero outcome. This succeeds with probability about $2/n$ and keeps all $n/2$ values of $\bar{s}^*$ in the sum with coefficient $+1$. The bit h^* itself is never measured, since we read it at the end.
5. Measure $W=(s_1,\ldots,s_{a-1},l_{s^*})$, obtaining W'.

The ket now holds (h,h^*).

Step 7, transferring the phase. Write $\tilde{z}^*$ for the low $n-1$ bits of z^* and $z_B=\sum_{i\in B}\phi_i y_i$. From $z^*+z_B \equiv z'+2^{n-1}h$ and $-2^{n-1}\equiv 2^{n-1} \pmod{2^n}$,

$$z_B \;\equiv\; (z'-\tilde{z}^*) + 2^{n-1}h', \qquad h' \;=\; h\oplus h^*. \quad (12)$$

The two top bits reach the B-side constraint only through h'. Compute h' with one CNOT, measure it, and use h' together with h^* to erase h reversibly. Once h' is a measured constant the B side no longer sees h^*, and the two branches share one B-side factor. This is Lemma 2 of [15], and it is the only one of the four with a complete published proof. The phase survives the substitution, since

$$(-1)^{hd_n} \;=\; (-1)^{h'd_n}\,(-1)^{h^*d_n} \quad (13)$$

and the first factor is a global constant once h' is known. The surviving state is

$$|\Psi_7\rangle \;\propto\; \sum_{h^*\in\{0,1\}} (-1)^{h^*d_n}\,E_{h^*}\,|h^*\rangle, \quad (14)$$

where E_{h^*} is the signed count of the ϕ compatible with the whole record and with that value of h^*.

The final step. Hadamard h^* and measure. For a state $E_0|0\rangle + (-1)^{d_n}E_1|1\rangle$ the outcome equals d_n with probability

$$p \;=\; \frac{(E_0+E_1)^2}{2(E_0^2+E_1^2)} \;=\; \frac{1}{1+q^2}, \qquad q \;=\; \frac{|E_0-E_1|}{|E_0+E_1|}, \quad (15)$$

so the whole algorithm reduces to showing that E_0 and E_1 are close. Once p exceeds $1/2$ by a non-negligible amount, polynomially many repetitions and a majority vote return d_n, and a recursion on the remaining bits of d finishes the problem.

C. The coefficient reduces to convolution

Two families of sets record how the constraints tighten as the measurements accumulate, one family on each side of the partition. The two base members are fixed by the data known just after Step 4, namely the measured z' together with a value of the index z^* that ranges over $\mathbb{Z}_N$ as a summation variable and is never a measurement outcome,

$$\mathcal{A}_{z^*} = \Big\{\phi_A : \sum_{i\in A}\phi_i y_i = z^*\Big\}, \quad (16)$$

$$\mathcal{B}_{(z',z^*),h} = \Big\{\phi_B : z^* + \sum_{i\in B}\phi_i y_i = z' + 2^{n-1}h\Big\}. \quad (17)$$

Each later measurement discards some of the strings that remain, and the notation records this by growing a subscript. A prime marks the member that carries

TABLE III. The set families. Within each side, every entry carries the constraints of the entry above it together with the one shown.

Set	Fixed after	Constraint added
The A side, where no sign is carried		
$\mathcal{A}_{z^*}$	Step 4	$z^*(\phi_A) = z^*$
$\mathcal{A}'_{z^*,W'}$	Step 6	$s_j(\phi_A) = W'_j$ for $j < a$, together with $l_{s^*(\phi_A)} = 0$
The B side, which carries every sign		
$\mathcal{B}_{(z',z^*),h}$	Step 4	$z_B(\phi_B) = z' + 2^{n-1}h - z^*$
$\mathcal{B}_{(z',z^*),S,h}$	Step 5	$s_j(\phi_B) = \sigma'_j$ for $j \in B$
$\mathcal{B}'_{(z',\tilde{z}^*),S,h'}$	Step 7	h replaced by the measured h', so that by Eq. (12) the index z^* collapses to $\tilde{z}^*$

everything the record fixes. On the A side that member is $\mathcal{A}'_{z^*,W'}$, which collects the ϕ_A with $z^*(\phi_A) = z^*$ that also match the top blocks recorded in W' at Step 6 and satisfy the post-selection $l_{s^*} = 0$. On the B side it is $\mathcal{B}'_{(z',z^*),S,h'}$, which collects the ϕ_B that match the blocks S measured at Step 5 and obey Eq. (12) with the measured h'. Table III collects all five members with the constraint each one adds.

The A-side member is indexed by the whole of z^*, while the B-side member is indexed by its low $n-1$ bits $\tilde{z}^*$ alone. That collapse $\mathcal{B}'_{(z',z^*),S,h'} \to \mathcal{B}'_{(z',\tilde{z}^*),S,h'}$ is Lemma 2. By Eq. (12) the B side is constrained by $(z' - \tilde{z}^*)$ and h', so two values of z^* differing only in the top bit h^* give the same set of ϕ_B, and therefore the same signed count.

The B-side signed count and the normalised per-index amplitude are

$$C_{(z',\tilde{z}^*),S,h'} = \sum_{\phi_B \in \mathcal{B}'_{(z',\tilde{z}^*),S,h'}} (-1)^{\phi_B \cdot \phi'_B}, \tag{18}$$

$$\alpha_{z^*} = \lambda_4 \left|\mathcal{A}'_{z^*,W'}\right| C_{(z',\tilde{z}^*),S,h'}, \tag{19}$$

where $\lambda_4 = (E_0^2 + E_1^2)^{-1/2}$ normalises the two branches at the end of Step 7. The unnormalised amplitude of the branch h^* is

$$E_{h^*} = \sum_{\tilde{z}^* \in \{0,1\}^{n-1}} \left|\mathcal{A}'_{(h^*,\tilde{z}^*),W'}\right| C_{(z',\tilde{z}^*),S,h'}. \tag{20}$$

The A-side factor is an unsigned count and every sign in the problem has been packed into C. We write $\hat{E}_{h^*} = \lambda_4 E_{h^*}$ for the normalised branch amplitude, so that $\hat{E}_0^2 + \hat{E}_1^2 = 1$ and $\sum_{z^*\,:\,\mathrm{msb}(z^*)=h^*} \alpha_{z^*} = \hat{E}_{h^*}$. Equation (15) may be read with either pair, since q is a ratio. The unnormalised E_{h^*} is kept in what follows, since the weight of a record is proportional to $E_0^2 + E_1^2$ and that quantity is identically 1 in the hatted variables. Subtracting the two branches gives the one line [15] reduces to,

$$E_0 - E_1 = \sum_{\tilde{z}^*} \Big(\left|\mathcal{A}'_{(0,\tilde{z}^*),W'}\right| - \left|\mathcal{A}'_{(1,\tilde{z}^*),W'}\right| \Big) \cdot C_{(z',\tilde{z}^*),S,h'}. \tag{21}$$

Each lemma guards one feature of Eq. (21). Lemma 2 allows for the common factor C outside the bracket. Lemma 4 bounds the relative difference of the two counts, once they are aggregated over the low bits of z^*, by $O(n^{-((c-1)/2-1)})$. Lemma 3 bounds each product α_{z^*} on its own, so that a small relative difference implies a small absolute one.

TABLE IV. Parameters. The sample count constant is written K, where [15] writes k, so that k stays free for Fourier frequencies.

Symbol	Meaning	Value
n	bit length	$N = 2^n$
c	group size constant	≥ 12
K	sample count constant	$> c$
Q	number of samples	Kn^{c+1}
m	bits per group	$c \log n$
G	number of groups	Q/m
a	number of A groups	$n/\log n$
L	width of s_j	$\log n$
ν	length of one top block	2^{n-L}
Q_B	sample bits on the B side	$(G-a)m$
$1/(c' \log n)$	faulty sample rate	c' constant

D. Parameters

Table IV lists the parameters that are fixed throughout our proof. Our ϕ is written $b_1 \cdots b_Q$ in [15], and our ϕ' is written D. $n_j[r]$ counts the configurations of group j whose subset sum equals r,

$$n_j[r] = \#\big\{\phi_j \in \{0,1\}^m : \textstyle\sum_{i \in g_j} \phi_i y_i \equiv r\big\}. \tag{22}$$

E. The probability space

One run produces the record $(Y, z', \phi', S, W', h')$, with Y drawn at Step 1 and the rest supplied by the measurements of Steps 2 to 7. Every probability below lives in the joint law of these. The results about the samples, including Lemma F, Theorem G and Lemma H, are proved with Y uniform on $\mathbb{Z}_N^Q$. Theorem E and Theorem E′ concern

the measured string and are proved with the rest of the record held fixed and ϕ'_B varying. The bound they give is the same number for every value of the rest. A bound holding under every conditioning holds under the joint law as well, so the failure probabilities of the two kinds live in one space and may be added. The second kind deserves a warning, since ϕ' fixes the partition and the partition belongs to the record. Holding the rest of the record fixed therefore already restricts which ϕ'_B remain available. Letting ϕ'_B range over the whole cube is exactly Hypothesis (U) of Section IV, and Theorem E and Theorem E′ are stated under it.

III. LEMMA 1

Lemma 1 [15]. *If $k > c$, then with constant probability, the measured value of $D = b'_1 \dots b'_Q$ will have at least $n/\log n$ groups that consist of all zeroes.*

Here k is our K and D is our ϕ'. This section proves the lemma in the stronger form that the probability tends to one. The route is a direct moment computation on the number of all-zero groups, and the two transforms it needs are recorded first.

The argument in [15] runs in two stages. The first stage shows that zero blocks of ϕ' are at least as common as one blocks. The second maps the records carrying fewer than a all-zero groups into the records carrying at least a, by exchanging positions inside ϕ' and exchanging the low $n - 2\log n$ bits of the corresponding weights y_i. Those low bits are taken to be distinct across the sample, which makes the map invertible. Comparing the two probabilities asks for more than invertibility. Moving the low bits alone changes $\sum_i \phi_i y_i$, so the set Z'_Y of strings compatible with the measured z' is not carried onto its image counterpart. The moment computation below reaches the conclusion without any such map.

A. Two Fourier transforms

Two different transforms appear below. On the cube of measurement outcomes we use the Walsh–Hadamard transform [10],

$$\widehat{F}(\phi') = \sum_{\phi\in\{0,1\}^Q} F[\phi]\,(-1)^{\phi\cdot\phi'}, \tag{23}$$

and on the group $\mathbb{Z}_N$ of subset sums we use

$$\hat{f}(k) = \sum_{r\in\mathbb{Z}_N} f[r]\,\omega^{-kr}, \tag{24}$$

satisfying that

$$\sum_r f[r]\,\overline{g[r]} = \frac{1}{N}\sum_k \hat{f}(k)\,\overline{\hat{g}(k)}. \tag{25}$$

The second identity follows by expanding both sides and using $\frac{1}{N}\sum_k \omega^{k(r'-r)} = \delta_{rr'}$.

We also write $\langle f\rangle_k = \frac{1}{N}\sum_k f(k)$ and $\mathrm{Var}_k(f) = \langle f^2\rangle_k - \langle f\rangle_k^2$. These measure the shape of a function on the frequency line and are not probabilistic quantities.

B. One group

Immediately after Step 4 the record consists of (Y, z', ϕ') and the basis states are labelled $(h, s_1 \cdots s_G)$. The amplitude at such a label is $\gamma\sum_{\phi\in\mathcal{P}}(-1)^{\phi\cdot\phi'}$ for a record-independent γ, where the consistency class is

$$\mathcal{P}(z,\vec{s}) = \big\{\phi\in\{0,1\}^Q : z(\phi) = z,\ s_i(\phi) = s_i\ \forall i\big\}, \tag{26}$$

where $z = z' + 2^{n-1}h$. Every ϕ lies in exactly one class, so the classes partition the cube and $\sum_{\mathcal{P}}|\mathcal{P}| = 2^Q$. Writing $\mathbf{1}_{\mathcal{P}}(\phi)$ for the indicator function of a class

$$\mathbf{1}_{\mathcal{P}}[\phi] = \begin{cases} 1, & \phi\in\mathcal{P},\\ 0, & \text{otherwise,}\end{cases} \tag{27}$$

we have

$$\Pr[\phi';\phi\in\mathcal{P}] = \gamma^2\Big(\sum_{\phi\in\mathcal{P}}(-1)^{\phi\cdot\phi'}\Big)^2 = \gamma^2\big(\widehat{\mathbf{1}_{\mathcal{P}}}(\phi')\big)^2. \tag{28}$$

Lemma A. *For every group j, every Y, and every fixed outcome of the earlier measurements, $\Pr[\phi'_j = 0 \mid \phi\in\mathcal{P}] \ge 2^{-m}$. Summing over the classes, which removes the constraints imposed by z', h and $\vec{s}$,*

$$\Pr[\phi'_j = 0 \mid Y] = 2^{-2m}\sum_r n_j[r]^2. \tag{29}$$

Proof. Expanding the square gives

$$\big(\widehat{\mathbf{1}_{\mathcal{P}}}(\phi')\big)^2 = \sum_{\phi,\psi\in\mathcal{P}}(-1)^{(\phi\oplus\psi)\cdot\phi'}. \tag{30}$$

Summing over all ϕ' leaves only the terms with $\phi\oplus\psi = 0$, so

$$\Pr[\phi\in\mathcal{P}] = \sum_{\phi'}\Pr[\phi',\phi\in\mathcal{P}] = \gamma^2 2^Q|\mathcal{P}|. \tag{31}$$

Summing instead over the 2^{Q-m} strings with $\phi'_j = 0$ leaves the terms whose difference is supported inside g_j, which means that

$$\Pr[\phi'_j = 0;\phi\in\mathcal{P}] = \sum_{\phi' :\, \phi'_j = 0}\Pr[\phi';\phi\in\mathcal{P}] = \gamma^2 2^{Q-m}A_{\mathcal{P}}, \tag{32}$$

where

$$A_{\mathcal{P}} = \#\big\{(\phi,\psi)\in\mathcal{P}^2 : \phi\oplus\psi \text{ supported in } g_j\big\}. \tag{33}$$

Dividing gives

$$\Pr[\phi'_j = 0 \mid \phi \in \mathcal{P}] = 2^{-m} A_{\mathcal{P}}/|\mathcal{P}|, \tag{34}$$

which is at least 2^{-m} because the diagonal $\phi = \psi$ already contributes $|\mathcal{P}|$.

Now let us sum over the classes. Suppose ϕ and ψ lie in the same class, so that $z(\phi) = z(\psi)$, and suppose they also agree outside g_j. Then $z(\phi) - z(\psi) = r_j(\phi) - r_j(\psi)$, so $r_j(\phi) = r_j(\psi)$. The constraint on s_j is then satisfied automatically. Choosing the bits outside g_j freely,

$$\sum_{\mathcal{P}} A_{\mathcal{P}} = \underbrace{2^{Q-m}}_{\text{bits outside } g_j} \times \underbrace{\sum_{r} n_j[r]^2}_{r_j(\phi_j)=r_j(\psi_j)}, \tag{35}$$

and dividing by $\sum_{\mathcal{P}} |\mathcal{P}| = 2^Q$ gives the stated form. □

C. Two groups

Lemma B. *For $j \neq j'$ and every Y,*

$$\Pr\big[\phi'_j = \phi'_{j'} = 0 \mid Y\big] = 2^{-4m} \frac{1}{N} \sum_{k} T_j(k)\, T_{j'}(k), \tag{36}$$

where

$$T_j(k) = \sum_{s} \big|\hat{w}_j^{(s)}(k)\big|^2 \tag{37}$$

and $w_j^{(s)}[r] = n_j[r]\,\mathbf{1}[\mathrm{top}(r) = s]$ being the restriction of n_j to one top block.

Proof. The character sum of Lemma A applies with $\{\phi'_j = 0\}$ replaced by $\{\phi'_j = \phi'_{j'} = 0\}$. There are 2^{Q-2m} such strings, therefore

$$\Pr\big[\phi'_j = \phi'_{j'} = 0 \mid Y\big] = 2^{-2m} \cdot \frac{\sum_{\mathcal{P}} A_{\mathcal{P}}^{(2)}}{2^Q}, \tag{38}$$

where

$$A_{\mathcal{P}}^{(2)} = \#\big\{(\phi,\psi) \in \mathcal{P}^2 : \phi \oplus \psi \text{ supported in } g_j \cup g_{j'}\big\}. \tag{39}$$

Let ϕ and ψ agree outside $g_j \cup g_{j'}$. Every s_i with $i \neq j, j'$ then agrees automatically, and

$$z(\phi) - z(\psi) = \big[r_j(\phi) - r_j(\psi)\big] + \big[r_{j'}(\phi) - r_{j'}(\psi)\big]. \tag{40}$$

So the pair lies in one class exactly when $r_j + r_{j'}$ is preserved (from z) and $\mathrm{top}(r_j)$ and $\mathrm{top}(r_{j'})$ are each preserved (from s_j and $s_{j'}$). In this case, the two groups may trade a common offset

$$\delta := r_j(\phi) - r_j(\psi) = r_{j'}(\psi) - r_{j'}(\phi). \tag{41}$$

Define the autocorrelation

$$R_j[\delta] = \sum_{\substack{r,r' \,:\, r - r' = \delta \\ \mathrm{top}(r) = \mathrm{top}(r')}} n_j[r]\, n_j[r'] = \sum_{s} \big(w_j^{(s)} \star w_j^{(s)}\big)[\delta], \tag{42}$$

where we define

$$(f \star f)[\delta] = \sum_{r} f[r] f[r - \delta]. \tag{43}$$

Pulling out the 2^{Q-2m} choices of the bits outside the two groups and sorting the rest by δ, we obtain

$$\begin{aligned} \sum_{\mathcal{P}} A_{\mathcal{P}}^{(2)} &= 2^{Q-2m} \sum_{\delta \in \mathbb{Z}_N} R_j[\delta]\, R_{j'}[-\delta] \\ &= 2^{Q-2m} \sum_{\delta \in \mathbb{Z}_N} R_j[\delta]\, R_{j'}[\delta], \end{aligned} \tag{44}$$

because R is even. Finally, we use the fact that

$$\begin{aligned} \widehat{f \star f}(k) &= \sum_{\delta} \sum_{r} f[r] f[r - \delta]\, \omega^{-k\delta} \\ &\overset{r' = r - \delta}{=} \sum_{r, r'} f[r] f[r']\, \omega^{-k(r - r')} \\ &= \big|\hat{f}(k)\big|^2, \end{aligned} \tag{45}$$

so $\widehat{R_j}(k) = T_j(k)$, and the $T_j(k)$ of Lemma B is the Fourier transform of the autocorrelation $R_j[\delta]$. The Parseval identity then turns $\sum_\delta R_j[\delta] R_{j'}[\delta]$ into $\frac{1}{N}\sum_k T_j(k) T_{j'}(k)$ and finishes the proof. □

Since T_j is the Fourier transform of R_j, the inverse transform at $\delta = 0$ gives $\langle T_j \rangle_k = \sum_r n_j[r]^2$. Lemma A therefore also reads $\Pr[\phi'_j = 0 \mid Y] = 2^{-2m} \langle T_j \rangle_k$, and the two lemmas have the same shape.

D. Averaging over the weights

Every probability written so far has been conditional on one fixed Y, with the Hadamard measurement of Step 4 supplying the only randomness. The weights are produced by the algorithm itself at Step 1, where the Fourier measurement on the second register of a sample returns a y_i uniform on $\mathbb{Z}_N$. The Q samples are prepared independently, so

$$Y = (y_1, \ldots, y_Q) \;\sim\; \mathrm{Unif}\big(\mathbb{Z}_N^Q\big). \tag{46}$$

Theorem C below is a statement about a run of the whole algorithm, so its probability space carries Y as well, and from here on $\mathbb{E}$ denotes that average. Now we calculate the expected second moment of a single group's subset-sum count. Write

$$\beta \;=\; \frac{2^m - 1}{N} \tag{47}$$

for the chance that a fixed nonempty subset of one group hits a fixed value. The result is

$$\begin{aligned} \mathbb{E}\big[n_j[r]\, n_j[r']\big] &= \mathbf{1}[r{=}r'{=}0] + \beta\big(\mathbf{1}[r{=}0] + \mathbf{1}[r'{=}0]\big) \\ &\quad + \beta\, \delta_{rr'} + \frac{4^m - 3 \cdot 2^m + 2}{N^2}, \end{aligned} \tag{48}$$

whose four terms come from the pairs where both subsets are empty, where exactly one is, where the two coincide and are nonempty, and where they differ and neither is empty (Appendix A). Carrying Eq. (48) through $T_j(k) = \sum_s \sum_{r,r' \in s} n_j[r] n_j[r'] \, \omega^{-k(r-r')}$ gives

$$\begin{aligned} \mathbb{E}T(k) &= 2^m + \beta_2 D_\nu(k) + \beta_1 \operatorname{Re} \hat{\mathbf{1}}_{B_0}(k), \\ \beta_1 &= 2\beta, \qquad \beta_2 = \frac{(4^m - 3 \cdot 2^m + 2)\, 2^L}{N^2}, \end{aligned} \tag{49}$$

where $B_0 = \{0, 1, \ldots, \nu - 1\}$ is the bottom top-block (i.e., those with zero top bits) and $D_\nu(k) = |\hat{\mathbf{1}}_{B_0}(k)|^2$ is the Fejér kernel of order ν. Two averages of that kernel are also computed there

$$\langle D_\nu \rangle_k = \nu, \qquad \langle D_\nu^2 \rangle_k = \frac{\nu(2\nu^2 + 1)}{3} \leq \nu^3. \tag{50}$$

Theorem C. *Let $\mathcal{Z} = \sum_{j=1}^{G} \mathbf{1}[\phi'_j = 0]$ be the number of all-zero groups. Under the hypothesis $K > c$ of [15], $\Pr[\mathcal{Z} \geq a] \to 1$. This is Lemma 1 of [15] with probability tending to one in place of a constant.*

Proof. We take five steps.

Step 1. Lemmas A and B hold for each fixed Y, so

$$\Pr[\phi'_j = 0 \mid Y] = 2^{-2m} \langle T_j \rangle_k, \tag{51}$$

$$\Pr[\phi'_j = \phi'_{j'} = 0 \mid Y] = 2^{-4m} \langle T_j T_{j'} \rangle_k. \tag{52}$$

Different groups use disjoint weights and all weights are independent, so $\mathbb{E}[T_j T_{j'}](k) = \mathbb{E}T_j(k)\, \mathbb{E}T_{j'}(k)$ pointwise in k. The groups are identically distributed, so writing $\mathbb{E}T$ for the common value

$$p_1 := \Pr[\phi'_j = 0] = 2^{-2m} \langle \mathbb{E}T \rangle_k, \tag{53}$$

$$p_2 := \Pr[\phi'_j = \phi'_{j'} = 0] = 2^{-4m} \langle (\mathbb{E}T)^2 \rangle_k. \tag{54}$$

Step 2. In general $\langle (\mathbb{E}T)^2 \rangle_k \neq \langle \mathbb{E}T \rangle_k^2$, and the difference measures the correlation between two groups

$$p_2 - p_1^2 = 2^{-4m} \operatorname{Var}_k(\mathbb{E}T), \tag{55}$$

$$\eta := \frac{p_2 - p_1^2}{p_1^2} = \frac{\operatorname{Var}_k(\mathbb{E}T)}{\langle \mathbb{E}T \rangle_k^2}, \tag{56}$$

where η is the normalised correlation coefficient of two groups being simultaneously all-zero. In this sense, a question about probabilistic independence has become a question about the shape of one function.

Step 3. Write $Z_j = \mathbf{1}[\phi'_j = 0]$ for the indicator of group j. Each Z_j is Bernoulli with parameter p_1 and every pair has covariance $p_1^2 \eta$, so

$$\begin{aligned} \operatorname{Var}[\mathcal{Z}] &= \underbrace{\sum_j \operatorname{Var}[Z_j]}_{G p_1 (1 - p_1)} + \underbrace{\sum_{j \neq j'} \operatorname{Cov}(Z_j, Z_{j'})}_{G(G-1) p_1^2 \eta} \\ &\leq G p_1 + G^2 p_1^2 \eta. \end{aligned} \tag{57}$$

Since $\mathbb{E}[\mathcal{Z}] = G p_1$, dividing by $\mathbb{E}[\mathcal{Z}]^2$ gives

$$\frac{\operatorname{Var}[\mathcal{Z}]}{\mathbb{E}[\mathcal{Z}]^2} \;\leq\; \frac{1}{\mathbb{E}[\mathcal{Z}]} + \eta. \tag{58}$$

The first is the Poisson fluctuation of G nearly independent coins, and the second is the excess contributed by their dependence.

Step 4. It remains to bound η. In Eq. (49) the term 2^m is constant in k, and Var_k is invariant under adding a constant, so only the two bumps contribute. From $(x + y)^2 \leq 2x^2 + 2y^2$ and $\operatorname{Var}_k \leq \langle \cdot^2 \rangle_k$, we have

$$\operatorname{Var}_k(\mathbb{E}T) \;\leq\; 2\beta_2^2 \langle D_\nu^2 \rangle_k + 2\beta_1^2 \big\langle (\operatorname{Re} \hat{\mathbf{1}}_{B_0})^2 \big\rangle_k. \tag{59}$$

By Eq. (50) the first average is at most ν^3, and the second is at most $\langle |\hat{\mathbf{1}}_{B_0}|^2 \rangle_k = \nu$. Using $\beta_2 \leq 4^m 2^L / N^2$, $\beta_1 = 2\beta \leq 2 \cdot 2^m / N$ and $\nu = N / 2^L$, we bound the variance as

$$\begin{aligned} \beta_2^2 \nu^3 \leq \frac{16^m}{N 2^L}, \qquad \beta_1^2 \nu \leq \frac{4 \cdot 4^m}{N 2^L}, \\ \text{so} \quad \operatorname{Var}_k(\mathbb{E}T) \leq \frac{2 \cdot 16^m + 8 \cdot 4^m}{N 2^L}. \end{aligned} \tag{60}$$

For the denominator, $\langle \mathbb{E}T \rangle_k \geq 2^m$ gives $\langle \mathbb{E}T \rangle_k^2 \geq 4^m$. This gives

$$\begin{aligned} \eta \;&\leq\; \frac{2 \cdot 4^m + 8}{N 2^L} \;=\; \frac{2n^{2c} + 8}{n\, 2^n} \\ &\leq\; \frac{2n^{2c-1} + 8}{2^n} \;=\; 2^{-n + O(\log n)}, \end{aligned} \tag{61}$$

using $2^m = n^c$, $2^L = n$ and $N = 2^n$. The correlation between two groups is therefore exponentially small.

Step 5. By Lemma A, $\mathbb{E}[\mathcal{Z}] \geq G 2^{-m} = (K/c)a$, so $a \leq (c/K)\mathbb{E}[\mathcal{Z}]$ with $c/K < 1$. Chebyshev's inequality together with Eq. (58) gives

$$\begin{aligned} \Pr[\mathcal{Z} \leq a] &\leq \Pr\Big[\big|\mathcal{Z} - \mathbb{E}\mathcal{Z}\big| \geq \big(1 - \tfrac{c}{K}\big)\mathbb{E}\mathcal{Z}\Big] \\ &\leq \Big(1 - \frac{c}{K}\Big)^{-2} \Big(\frac{c}{Ka} + \eta\Big), \end{aligned} \tag{62}$$

where the last step also uses $1/\mathbb{E}[\mathcal{Z}] \leq c/(Ka)$. Here c and K are constants, $a = n / \log n \to \infty$, and $\eta \to 0$ exponentially, so the right side tends to zero. □

Therefore, Lemma 1 holds with a conclusion stronger than [15] claimed.

IV. LEMMA 3

Definition 1 [15]. *For a given measured z', and a resulting superposition ψ of values of h^* at the end of step 7, let M be the set of values (Y, D, W', S, h') measured (apart from z') during the algorithm, let T_M^+ (resp., T_M^-) be the set of states ϕ consistent with M and*

h^* and with phase 1 (resp., -1) resulting from measurement of M (ignoring the constant phase $\omega^{yz'd}$). Let $T_M = T_M^+ \cup T_M^-$, and let t_M (resp., t_M^+, t_M^-) be $|T_M|$ (resp., $|T_M^+|$, $|T_M^-|$). Then ψ is well-behaved for a value of h^* if $|t_M^+ - t_M^-| \geq \Omega(2^{-n/2}\sqrt{t_M})$.

Lemma 3 [15]. *With probability $1 - O(2^{-n})$ (over choices of M), the superposition ψ at the end of step 7 is well-behaved for at least one value of h^*. Moreover, if it's well-behaved for a particular h^*, then the probability (over choices of M) that for some z^* consistent with that value of h^* the amplitude $\alpha_{z^*} = \nu_4 \sum_{\phi_A \in A_{z^*,W'}} C_{\bar{z}^*}$ has magnitude greater than $\Omega(2^{3n/2})$ is less than $O(2^{-n})$.*

In our notation D is ϕ', ν_4 is $\lambda_4 = (E_0^2 + E_1^2)^{-1/2}$, $\bar{z}^*$ is $\tilde{z}^*$, and $A_{z^*,W'}$ is $\mathcal{A}'_{z^*,W'}$, all of which appear in Table III. The set T_M of Definition 1 and its size t_M have nothing to do with the Fourier-side quantity $T_j(k)$ of Lemma B.

Both halves in [15] rest on a variance computation for $t_M^{\pm}$, in which the signs $(-1)^{\phi\cdot\phi'}$ are taken to be pairwise independent as ϕ' varies. However, a sign depends on ϕ only through ϕ_B since $\phi'_A = 0$, so two states of T_M that share a ϕ_B carry the same sign. Lemma D below reads the same object as an exact Walsh transform, which asks nothing of the signs.

Separately from the signs, the proof of Lemma 3 in [15] implies a hypothesis that the algorithm steps alone do not give. It sums ϕ'_B over the whole cube $\{0,1\}^{Q_B}$ with the rest of the record held fixed. The set A consists of the *first a* all-zero groups, so the partition is itself a function of ϕ' and the admissible strings form a proper subset of the cube. We take the freedom as a hypothesis and carry it in the statements. Theorem E and Theorem E′ are both stated under Hypothesis (U), as is the result of Appendix B.

Hypothesis (U). *The A/B partition is fixed and does not vary with ϕ'. With the rest of the record held fixed, ϕ'_B therefore ranges over all of $\{0,1\}^{Q_B}$.*

Here we summarize our proof of Lemma 3. It is the only one of the four that involves both the size of an amplitude and the probability of a record. Its second half is the one the argument consumes, and [15] states that half conditionally for an h^* at which the superposition is *well-behaved*. The first half supplies that predicate with high probability. The proof below carries no such condition, because it runs a Parseval identity directly on the distribution of records. The normalisation $\sqrt{E_0^2 + E_1^2}$ that defines α_{z^*} is the weight of the record itself. Appendix B records the first half regardless, in the sharper form that the failure probability is ε^2 at every threshold ε.

The second half is a statement about the probability of a record. Once ϕ' is measured, the state left by Section II is

$$E_0(\phi')\,|0\rangle \;+\; (-1)^{d_n} E_1(\phi')\,|1\rangle. \tag{63}$$

The probability of that record is the squared norm of this unnormalised vector, so it is proportional to $E_0(\phi')^2 + E_1(\phi')^2$. Theorem E is proved directly on the record distribution, whose weights $E_0(\phi')^2 + E_1(\phi')^2$ are not uniform in ϕ'. The constant of proportionality is fixed by Q, N, L and the number of faulty samples, so it does not depend on the record and cancels from every ratio below. That non-uniformity is not itself an obstacle, since the weight enters the numerator and the denominator of Step 4 alike and leaves the value unchanged.

A. The signed count as a Walsh transform

Recall that $\phi'_A = 0$ and $(-1)^{\phi\cdot\phi'} = (-1)^{\phi_B\cdot\phi'_B}$, so the whole content of ϕ' sits in its B-side part ϕ'_B. Define

$$\mathcal{N}_{h^*}(\phi_B) \;=\; \#\big\{\phi_A : (\phi_A, \phi_B) \in T_M \text{ for that } h^*\big\} \tag{64}$$

as the number of A-side completions of a given B-side configuration, which takes nonnegative integer values.

Lemma D. *$E_{h^*}(\phi') = \widehat{\mathcal{N}_{h^*}}(\phi'_B)$ (the Walsh–Hadamard transform on $\{0,1\}^{Q_B}$), and consequently*

$$\sum_{\phi'_B \in \{0,1\}^{Q_B}} E_{h^*}(\phi')^2 \;=\; 2^{Q_B} \sum_{\phi_B} \mathcal{N}_{h^*}(\phi_B)^2. \tag{65}$$

Proof. We first group the states of T_M by their B-side part. States in one group share a ϕ_B and therefore the same sign, which may be pulled outside the inner sum,

$$\begin{aligned} E_{h^*}(\phi') &= \sum_{\phi \in T_M} (-1)^{\phi_B\cdot\phi'_B} \\ &= \sum_{\phi_B} \underbrace{\Big(\sum_{\phi_A} \mathbf{1}\big[(\phi_A,\phi_B) \in T_M\big]\Big)}_{=\,\mathcal{N}_{h^*}(\phi_B)} (-1)^{\phi_B\cdot\phi'_B}, \end{aligned} \tag{66}$$

and the right side is the Walsh–Hadamard transform of $\mathcal{N}_{h^*}$ evaluated at ϕ'_B.

Next, square and sum over ϕ'_B,

$$\begin{aligned} \sum_{\phi'_B} E_{h^*}(\phi')^2 &= \sum_{\phi_B,\psi_B} \mathcal{N}_{h^*}(\phi_B)\,\mathcal{N}_{h^*}(\psi_B) \\ &\qquad \times \sum_{\phi'_B \in \{0,1\}^{Q_B}} (-1)^{(\phi_B \oplus \psi_B)\cdot\phi'_B} \\ &= 2^{Q_B} \sum_{\phi_B} \mathcal{N}_{h^*}(\phi_B)^2, \end{aligned} \tag{67}$$

where the inner sum is a character sum and only the diagonal term $\phi_B = \psi_B$ survives. □

B. The amplitude bound

Write Δ_{z^*} for the single term of Eq. (20) carried by one value of z^*

$$E_{h^*}(\phi') = \sum_{\tilde{z}^* \in \{0,1\}^{n-1}} \Delta_{z^*}(\phi'), \tag{68}$$

$$\Delta_{z^*}(\phi') = \left|\mathcal{A}'_{z^*,W'}\right| C_{(z',\tilde{z}^*),S,h'}, \tag{69}$$

and after normalisation the amplitude carried by z^* is $\alpha_{z^*}(\phi') = \Delta_{z^*}(\phi')/\sqrt{E_0(\phi')^2 + E_1(\phi')^2}$, which is the α_{z^*} of [15].

Theorem E. *Assume Hypothesis (U), then for every threshold $\theta > 0$,*

$$\begin{aligned} \mathbb{E}\Big[\sum_{z^*} \alpha_{z^*}^2\Big] \;&\le\; 1, \\ \text{and hence}\quad \Pr\Big[\max_{z^*} |\alpha_{z^*}| > \theta\Big] \;&\le\; \theta^{-2}. \end{aligned} \tag{70}$$

Here $\mathbb{E}$ and $\Pr$ are defined over $\phi'_B \in \{0,1\}^{Q_B}$. Taking $\theta = 2^{3n/2}$ gives 2^{-3n}, stronger than the $O(2^{-n})$ of Lemma 3 of [15].

Proof. We take five steps.

Step 1. Once the total sum is determined, $z^* = z - \sum_{i\in B} \phi_i y_i$, where the right side depends only on ϕ_B beyond the fixed z. So z^* is a function of ϕ_B, written as $z^*(\phi_B)$.

Step 2. Let

$$\begin{aligned} \mathcal{N}_{z^*}(\phi_B) = \#\Big\{\phi_A : (\phi_A, \phi_B) \in T_M \\ \text{and } \textstyle\sum_{i\in A} \phi_i y_i = z^*\Big\}, \end{aligned} \tag{71}$$

which adds to $\mathcal{N}_{h^*}(\phi_B)$ the requirement that the A-side sum equal this particular z^*. The question is whether the extra requirement removes anything. If $z^* \neq z^*(\phi_B)$ then no ϕ_A qualifies and $\mathcal{N}_{z^*}(\phi_B) = 0$. If instead $z^* = z^*(\phi_B)$, then every ϕ_A making (ϕ_A, ϕ_B) lie in T_M satisfies $\sum_{i\in A} \phi_i y_i = z - \sum_{i\in B} \phi_i y_i = z^*$ automatically. Therefore

$$\mathcal{N}_{z^*}(\phi_B) = \mathcal{N}_{h^*}(\phi_B) \cdot \mathbf{1}[z^*(\phi_B) = z^*], \tag{72}$$

which means that

$$\sum_{z^*} \mathcal{N}_{z^*}(\phi_B) = \mathcal{N}_{h^*}(\phi_B), \tag{73}$$

$$\sum_{z^*} \mathcal{N}_{z^*}(\phi_B)^2 = \mathcal{N}_{h^*}(\phi_B)^2. \tag{74}$$

Step 3. Equation (66) applies to $\mathcal{N}_{z^*}$, giving $\Delta_{z^*}(\phi') = \widehat{\mathcal{N}_{z^*}}(\phi'_B)$, and then

$$\begin{aligned} \sum_{\phi'_B} \sum_{z^*} \Delta_{z^*}(\phi')^2 &= 2^{Q_B} \sum_{z^*} \sum_{\phi_B} \mathcal{N}_{z^*}(\phi_B)^2 \\ &\overset{\text{Step 2}}{=} 2^{Q_B} \sum_{\phi_B} \mathcal{N}_{h^*=0}(\phi_B)^2 \\ &\quad + 2^{Q_B} \sum_{\phi_B} \mathcal{N}_{h^*=1}(\phi_B)^2 \\ &= \sum_{\phi'_B} \big(E_0(\phi')^2 + E_1(\phi')^2\big), \end{aligned} \tag{75}$$

where the last equality is Lemma D again.

Step 4. The weight of a record is proportional to $E_0(\phi')^2 + E_1(\phi')^2$ and the constant cancels. The quantity α_{z^*} carries $\sqrt{E_0^2 + E_1^2}$ in its denominator and is defined only where the weight is positive, so the average runs over those records alone,

$$\mathbb{E}\Big[\sum_{z^*} \alpha_{z^*}^2\Big] = \frac{\displaystyle\sum_{\phi'_B \,:\, E_0^2+E_1^2>0} \underbrace{\big(E_0(\phi')^2 + E_1(\phi')^2\big)}_{\text{weight}} \cdot \underbrace{\frac{\sum_{z^*} \Delta_{z^*}(\phi')^2}{E_0(\phi')^2 + E_1(\phi')^2}}_{=\,\sum_{z^*} \alpha_{z^*}(\phi')^2}}{\displaystyle\sum_{\phi'_B \in \{0,1\}^{Q_B}} \big(E_0(\phi')^2 + E_1(\phi')^2\big)}. \tag{76}$$

The two factors in each numerator term cancel, leaving $\sum_{\phi'_B \,:\, E_0^2+E_1^2>0} \sum_{z^*} \Delta_{z^*}(\phi')^2$. The identity of Step 3 runs over every ϕ'_B, and the two sums differ by the nonnegative contribution of the zero-weight records. Dropping that contribution only decreases the value, so

$$\mathbb{E}\Big[\sum_{z^*} \alpha_{z^*}^2\Big] \;\le\; \frac{\sum_{\phi'_B} \sum_{z^*} \Delta_{z^*}(\phi')^2}{\sum_{\phi'_B} \big(E_0(\phi')^2 + E_1(\phi')^2\big)} \;=\; 1. \tag{77}$$

Step 5. Write $V(\phi') = \sum_{z^*} \alpha_{z^*}(\phi')^2$, which is nonnegative with $\mathbb{E}[V] \le 1$ by Step 4. Pointwise $\max_{z^*} \alpha_{z^*}(\phi')^2 \le V(\phi')$, so the event $\{\max_{z^*} |\alpha_{z^*}| > \theta\}$ is contained in $\{V > \theta^2\}$. Splitting $\mathbb{E}[V]$ according to whether V exceeds θ^2, we have

$$\mathbb{E}[V] \;\ge\; \theta^2 \Pr\big[V > \theta^2\big], \tag{78}$$

so $\Pr[V > \theta^2] \le \mathbb{E}[V]/\theta^2 \le \theta^{-2}$. □

C. The corollary

The corollary to Lemma 3 asks the same question of the aggregate

$$\alpha_{z_h^*} \;=\; \sum_{z^*\,:\,\mathrm{top}(z^*)=z_h^*} \alpha_{z^*}, \tag{79}$$

a class containing $2^{\,n-L}$ values of z^*. The proof of Theorem E used only that z^* is a function of ϕ_B, so any coarser classification carries the same argument.

Theorem E′. *Assume Hypothesis (U). Let ζ be any function of ϕ_B, with values in an arbitrary set. Write $\Delta_\zeta(\phi')$ for the part of $E_{h^*}(\phi')$ contributed by the terms with $\zeta(\phi_B)=\zeta$, and $\alpha_\zeta = \Delta_\zeta/\sqrt{E_0^2+E_1^2}$. Then for every $\theta>0$,*

$$\mathbb{E}\Big[\,\sum_{\zeta}\alpha_\zeta^2\,\Big] \le 1, \tag{80}$$

$$\Pr\Big[\,\max_{\zeta}|\alpha_\zeta| > \theta\,\Big] \le \theta^{-2}. \tag{81}$$

Proof. Repeat the five steps of Theorem E with z^* replaced by ζ. Step 1 is now the hypothesis on ζ. Step 2 uses only that hypothesis and gives $\mathcal{N}_\zeta(\phi_B) = \mathcal{N}_{h^*}(\phi_B)\mathbf{1}[\zeta(\phi_B)=\zeta]$, which has at most one nonzero entry for fixed ϕ_B. Steps 3 to 5 never refer to the meaning of ζ. □

Taking $\zeta(\phi_B)=z^*(\phi_B)$ recovers Theorem E. Taking $\zeta(\phi_B)=\mathrm{top}(z^*(\phi_B))$ gives

$$\mathbb{E}\Big[\sum_{z_h^*}\alpha_{z_h^*}^2\Big] \le 1, \tag{82}$$

$$\Pr\Big[\max_{z_h^*}\big|\alpha_{z_h^*}\big| > n^{3/2}\Big] \le n^{-3}, \tag{83}$$

well beyond the $O(1/n)$ the corollary asks for.

Theorem E and Theorem E′ give the second half of Lemma 3 and its corollary under Hypothesis (U), at every threshold and for either value of h^*. Both of these go beyond the published statement. The first half is in Appendix B and nothing above or below draws on it.

V. LEMMA 4

Lemma 4 [15]. *For any z^*, let z_h^* be the most significant $\log n$ bits of z^*, and z_l^* be the remaining bits. Then for any two values $z_{h,0}^*$ and $z_{h,1}^*$ of z_h^* that differ only in the most significant bit h^*, and for $c\ge 12$, the amplitudes $\alpha_{z_{h,0}}$ and $\alpha_{z_{h,1}}$ (defined as in the corollary to Lemma 3) differ by an expected multiplicative factor $1\pm O(n^{-(((c-1)/2)-1)})$ with probability at least $1-1/n$.*

The two amplitudes being compared are built out of the A-side counts. Substituting $\alpha_{z^*}=\lambda_4|\mathcal{A}'_{z^*,W'}|C_{(z',\tilde z^*),S,h'}$ into the aggregate of Eq. (79),

$$\alpha_{z_h^*} \;=\; \lambda_4 \sum_{z_l^*\in\mathbb{Z}_\nu} \big|\mathcal{A}'_{(z_h^*,z_l^*),W'}\big|\; C_{(z',\tilde z^*),S,h'}. \tag{84}$$

Two values of z_h^* differing only in the top bit share every bit of $\tilde z^*$ that the B-side factor sees, so C is the same function of z_l^* for both, which is Lemma 2. The comparison therefore sits entirely in the unsigned counts $|\mathcal{A}'|$, and the rest of this section is about them.

The skeleton in [15] is a pair of balls-in-bins models, one at the level of z_h^* and one at the level of z_l^*. A mean and a variance are read off each model and Chebyshev is applied to those two numbers. The balls are subset sums and are pairwise independent without being independent, which is all a variance needs, so both numbers the sketch reads off come out right. Sections V A to V D replace the two models by exact computations and recover the exponents and failure probabilities the paper displays. The assumption that g_a carries no faulty samples is removed along the way.

The A side factorises over groups, and g_a sits differently in the record from the other $a-1$ groups

$$\phi_A = \big(\;\underbrace{\phi_{g_1},\ldots,\phi_{g_{a-1}}}_{\text{top blocks pinned by } W'}\;,\;\underbrace{\phi_{g_a}}_{\text{top block free}}\;\big). \tag{85}$$

Three constraints define $\mathcal{A}'_{z^*,W'}$, namely $\mathrm{top}(r_j)=W'_j$ for $j<a$, $l_{s^*}=0$, and $z^*(\phi_A)=z^*$. Split the third one along $z^*=z_h^*\nu+z_l^*$. For the high half, the top L bits s_a of the sum over the distinguished group g_a determine z_h^* by

$$z_h^* = \sum_{j<a} W'_j + s_a(\phi_{g_a}) \bmod 2^L. \tag{86}$$

Every W'_j here is a measured constant, so fixing z_h^* is the same as fixing s_a and confines ϕ_{g_a} to the block $\mathrm{top}(r_{g_a})=s_a$. For the low half, writing $r_j=s_j\nu+\rho_j$ for each group j as in Section V D

$$z_l^* \equiv \rho_{g_a}(\phi_{g_a}) + \sum_{j<a}\rho_j \pmod\nu. \tag{87}$$

Enumerating ϕ_{g_a} first and counting the remaining groups afterwards turns the count into a double sum

$$\big|\mathcal{A}'_{(z_h^*,z_l^*),W'}\big| \;=\; \sum_{\phi_{g_a}\,:\,\mathrm{top}(r_{g_a})=s_a} X_{z_l^*-\rho_{g_a}}, \tag{88}$$

where

$$X_\rho \;=\; \#\Big\{(\phi_{g_j})_{j<a}\,:\,\mathrm{top}(r_j)=W'_j\;\forall j<a, \textstyle\sum_{j<a}\rho_j\equiv\rho\Big\} \tag{89}$$

counts the configurations of the pinned groups whose low parts add up to ρ. That bin count is taken up again in Section V D.

A. An exact covariance

We start from the covariance computation using Eq. (48).

Lemma F. *For any group j and any $r, r' \in \mathbb{Z}_N$,*

$$\mathrm{Cov}\big(n_j[r],\, n_j[r']\big) = \frac{2^m-1}{N}\Big(\delta_{rr'} - \frac{1}{N}\Big). \tag{90}$$

Consequently the top-block count $X_\sigma = \sum_{r\,:\,\mathrm{top}(r)=\sigma} n_j[r]$ has variance exactly $(2^m-1)2^{-L}(1-2^{-L})$.

Proof. The first moment behind Eq. (48) is $\mathbb{E}[n_j[r]] = \mathbf{1}[r{=}0] + \beta$, with $\beta = (2^m-1)/N$ as in Section III D, so

$$\begin{aligned}\mathbb{E}\big[n_j[r]\big]\mathbb{E}\big[n_j[r']\big] &= \mathbf{1}[r{=}0]\mathbf{1}[r'{=}0] \\ &\quad + \beta\big(\mathbf{1}[r{=}0] + \mathbf{1}[r'{=}0]\big) + \beta^2.\end{aligned} \tag{91}$$

Subtracting this from Eq. (48) leads to

$$\beta\,\delta_{rr'} + \frac{(4^m - 3\cdot 2^m + 2) - (2^m-1)^2}{N^2} = \beta\,\delta_{rr'} - \frac{2^m-1}{N^2}. \tag{92}$$

Summing over r and r' inside one block $\mathrm{top}(r) = \mathrm{top}(r') = \sigma$ gives

$$\begin{aligned}\mathrm{Var}(X_\sigma) &= \frac{2^m-1}{N}\Big(\nu - \frac{\nu^2}{N}\Big) = (2^m-1)\,\frac{\nu}{N}\Big(1-\frac{\nu}{N}\Big) \\ &= (2^m-1)\,2^{-L}\big(1-2^{-L}\big).\end{aligned}$$

□

B. Concentration of the block counts

Theorem G. *Let $\kappa > 0$ and let g_j be any group. With probability at least $1 - 2^L/\kappa^2$ over the samples of that group, every top block satisfies*

$$\big|X_\sigma - \mathbb{E}X_\sigma\big| \;\le\; \kappa\sqrt{(2^m-1)\,2^{-L}\big(1-2^{-L}\big)}. \tag{93}$$

For $\sigma \neq 0$ the mean is $\mathbb{E}X_\sigma = (2^m-1)2^{-L}$, and the bound takes the relative form

$$\frac{\big|X_\sigma - \mathbb{E}X_\sigma\big|}{\mathbb{E}X_\sigma} \;\le\; \kappa\sqrt{\frac{2^L\big(1-2^{-L}\big)}{2^m-1}}. \tag{94}$$

At the parameters $2^m = n^c$ and $2^L = n$ of [15], taking $\kappa = n$ gives a relative deviation at most $n^{-((c-1)/2-1)}$ with failure probability at most $1/n$. Those are the exponent and the failure probability displayed in Lemma 4.

Proof. We take five steps.

Step 1. The first moment behind Eq. (48) is $\mathbb{E}[n_j[r]] = \mathbf{1}[r{=}0] + \beta$, and a top block holds $\nu = N/2^L$ residues, so

$$\mathbb{E}X_\sigma = \sum_{r\,:\,\mathrm{top}(r)=\sigma}\big(\mathbf{1}[r{=}0] + \beta\big) = \mathbf{1}[\sigma{=}0] + \nu\beta, \tag{95}$$

$$\nu\beta = (2^m-1)2^{-L}. \tag{96}$$

The term $\mathbf{1}[\sigma{=}0]$ comes from the atom at $r=0$, since $\mathrm{top}(0) = 0$ places the empty subset in the block $\sigma = 0$. The $2^m - 1$ nonempty subsets divide evenly among the 2^L blocks, and the block $\sigma = 0$ holds one more.

Step 2. Lemma F gives $\mathrm{Var}(X_\sigma) = (2^m{-}1)2^{-L}(1{-}2^{-L})$ for every σ.

Step 3. Fix one σ. The standard deviation $\mathrm{sd}(X_\sigma) = \sqrt{\mathrm{Var}(X_\sigma)}$ does not depend on σ, so one threshold serves all blocks, and Chebyshev gives

$$\Pr\Big[\big|X_\sigma - \mathbb{E}X_\sigma\big| > \kappa\,\mathrm{sd}(X_\sigma)\Big] \;\le\; \frac{\mathrm{Var}(X_\sigma)}{\kappa^2\mathrm{Var}(X_\sigma)} \;=\; \frac{1}{\kappa^2}. \tag{97}$$

Only finiteness of the variance is used, and nothing about the shape of the distribution of X_σ.

Step 4. Let B_σ be the event that block σ leaves the range, for $\sigma = 0, 1, \ldots, 2^L-1$. We need to exclude the union. Pointwise $\mathbf{1}[\bigcup_\sigma B_\sigma] \le \sum_\sigma \mathbf{1}[B_\sigma]$, since the left side taking the value 1 forces at least one term on the right to do the same and the terms are nonnegative. Taking expectations and using Step 3,

$$\Pr\Big[\bigcup_\sigma B_\sigma\Big] \;\le\; \sum_{\sigma=0}^{2^L-1}\Pr[B_\sigma] \;\le\; \frac{2^L}{\kappa^2}, \tag{98}$$

and the complementary event is Eq. (93).

Step 5. For $\sigma \neq 0$, Step 1 gives $\mathbb{E}X_\sigma = (2^m{-}1)2^{-L}$, and dividing the bound of Step 4 by it produces the relative form. Substituting $2^m = n^c$ and $2^L = n$, the relative standard deviation is $\sqrt{n(1-1/n)/(n^c-1)} = n^{(1-c)/2}(1 + O(n^{-1}))$. Taking $\kappa = n$ turns this into $n^{\,1-(c-1)/2} = n^{-((c-1)/2-1)}$ with failure probability $2^L/\kappa^2 = n/n^2$. □

C. Faulty samples

Suppose g_a carries f_a faulty samples. A faulty sample fixes its bit ϕ_i at a definite value and takes no part in the superposition, so the group has only $m - f_a$ free bits. The count $n_{g_a}[r]$ then ranges over $2^{\,m-f_a}$ configurations, and the frozen bits contribute one fixed shift, which moves the whole distribution without changing its shape. So Lemma F, Theorem G and Lemma H below all hold with 2^m replaced by $2^{\,m-f_a}$.

Since $f_a \sim \mathrm{Bin}(m, p_f)$ with $p_f = 1/(c'\log n)$ and $m = c\log n$, the mean $\mathbb{E}f_a = mp_f = c/c'$ is a constant. On average

$$\begin{aligned}\mathbb{E}\big[2^{\,m-f_a}\big] &= 2^m\,\mathbb{E}\big[2^{-f_a}\big] \\ &= 2^m\Big(1-\frac{p_f}{2}\Big)^m \longrightarrow 2^m e^{-c/(2c')},\end{aligned} \tag{99}$$

indicating a loss of one constant factor. Passing from a ratio of averages to an average of ratios costs one application of Cauchy–Schwarz and stays at the same order.

Theorem G needs the high probability version. From $\binom{m}{k} \le m^k/k!$,

$$\Pr\big[f_a \ge k\big] \;\le\; \binom{m}{k}p_f^k \;\le\; \frac{(mp_f)^k}{k!} \;=\; \frac{(c/c')^k}{k!}, \tag{100}$$

TABLE V. The two models. They part company in the first row, where the second one sees the record.

	First, at z_h^*	Second, at z_l^*
Balls	all 2^m strings ϕ_{g_a} of g_a	the ϕ_A **compatible with** W'
One bin	a top block of r_{g_a}	a value of z_l^*
Bins	$2^L = n$	$\nu = 2^{n-L}$

and $k = \Theta(\log n/\log\log n)$ makes the right side smaller than $1/n$. With probability $1 - 1/n$ we therefore have $2^{m-f_a} = n^c 2^{-f_a} = n^{c-o(1)}$, so the relative deviation of Theorem G picks up a factor $n^{o(1)}$ and the exponent in $n^{-((c-1)/2-1)}$ is unchanged.

So the assumption that g_a carries no faulty samples can be dropped, and no constant probability has to be paid for it. Both failure probabilities are of order $1/n$ and may be added, and the same applies to the other A groups.

D. The second model under the conditioning

The two balls-in-bins arguments are not the same construction, as Table V shows. The first model lets ϕ_{g_a} run over all 2^m values with the other groups absent, while the second fixes the value ϕ_{g_a} first and lets the remaining groups run. The A groups therefore fall into two classes in the second model. The $a-1$ groups with $j < a$ are pinned by W' to $\mathrm{top}(r_j) = W'_j$ while the m bits inside each of them stay free, and all of the balls come from there. The group g_a has its bits fixed and contributes one definite number. The record $W' = (s_1, \dots, s_{a-1}, l_{s^*})$ holds top blocks only for $j < a$, and s_a is not among them.

Write

$$X_{s_j} = \#\big\{\phi_j \in \{0,1\}^m : \mathrm{top}(r_j) = s_j\big\}, \tag{101}$$

$$P = \prod_{j<a} X_{s_j}, \tag{102}$$

so that X_{s_j} is the block count of Theorem G taken at the pinned block, and P is the total number of ϕ_A compatible with W'. Write $r_j = s_j\nu + \rho_j$ with $0 \le \rho_j < \nu$. Since $s_j\nu \equiv 0 \pmod\nu$, the low $n-L$ bits of z^* come only from the low parts,

$$z_l^* \;\equiv\; \underbrace{\rho_{g_a}}_{\text{a definite number}} + \sum_{j<a} \rho_j \pmod\nu. \tag{103}$$

The term from g_a has to be separated out. Its effect is to shift every bin cyclically by one fixed amount and is independent of the other groups.

The bin count X_ρ of Eq. (89), which runs over the pinned groups $j < a$ alone, is then a convolution of their low parts on $\mathbb{Z}_\nu$. Restricting n_j to the pinned block gives the window slice

$$u_j(\rho) \;:=\; n_j\big[s_j\nu + \rho\big], \tag{104}$$

representing the same numbers as n_j, taken on one block and reindexed. Fixing $(\rho_1, \dots, \rho_{a-1})$, the number of compatible ϕ_A is $\prod_j n_j[s_j\nu + \rho_j] = \prod_j u_j(\rho_j)$. Summing over all choices with $\sum_j \rho_j \equiv \rho$ gives

$$\begin{aligned} X_\rho &= \sum_{\rho_1+\cdots+\rho_{a-1}\equiv\rho \pmod\nu} \prod_j u_j(\rho_j) \\ &= \big(u_1 * \cdots * u_{a-1}\big)(\rho), \end{aligned} \tag{105}$$

a cyclic convolution on $\mathbb{Z}_\nu$.

Lemma H. *Suppose at least one group has $s_j \neq 0$, and write $\mu = \mathbb{E}[P]/\nu$. Split each bin count into*

$$X_\rho \;=\; \frac{P}{\nu} + Y_\rho, \tag{106}$$

the first piece being the share of each bin when there are P balls in all, and the second how far bin ρ departs from its share. Then $\mathbb{E}X_\rho = \mu$, $\sum_\rho Y_\rho = 0$, the variables Y_ρ and P are uncorrelated, and

$$\mathrm{Cov}\big(Y_\rho, Y_{\rho'}\big) = \mu\,\delta_{\rho\rho'} - \frac{\mu}{\nu}. \tag{107}$$

This is the covariance of $\nu\mu$ balls thrown uniformly into ν bins, and the proof of Lemma 4 assumes exactly that. The conditioning on W' enters only through the ball count P, and since $P = \prod_j X_{s_j}$ has independent factors,

$$\frac{\mathrm{Var}(P)}{\big(\mathbb{E}P\big)^2} = \prod_j \Big(1 + \frac{\mathrm{Var}(X_{s_j})}{\big(\mathbb{E}X_{s_j}\big)^2}\Big) - 1 \;\approx\; a\,\frac{2^L}{2^m}, \tag{108}$$

so the relative fluctuation of P is about $\sqrt{a\,2^{\,L-m}} = n^{(2-c)/2}(\log n)^{-1/2}$.

A bin count fluctuates in two ways. The ball count itself changes, which makes every bin rise and fall together. The balls also scatter differently once the count is fixed, which makes the bins trade against one another. Written separately, Eq. (107) is the scattering and Eq. (108) is the change in the count.

Proof. Bin counts are a convolution of the slices. One Fourier transform turns that into a pointwise product, and different groups use independent weights, so the whole computation factors group by group. Write $\hat f(k) = \sum_\rho f(\rho) e^{-2\pi i k\rho/\nu}$ for the transform on $\mathbb{Z}_\nu$, so that $\hat X(k) = \prod_j \hat u_j(k)$. The zero frequency is the ball count, $\hat X(0) = \sum_\rho X_\rho = P$. The proof consists of four steps.

Step 1, a slice with $s_j \neq 0$. Let $s_{j_0} \neq 0$. Of the four terms in Eq. (48), the atom and the two cross-atoms sit at $r = 0$, and $\mathrm{top}(0) = 0$ places them in the block $s_j = 0$

alone. Every r in the block s_{j_0} is nonzero, so only two terms survive,

$$\mathbb{E}\, u_{j_0}(\rho) = \frac{2^m-1}{N}, \tag{109}$$

$$\mathbb{E}\big[u_{j_0}(\rho)\, u_{j_0}(\rho')\big] = \frac{2^m-1}{N}\,\delta_{\rho\rho'} + \frac{4^m-3\cdot 2^m+2}{N^2}, \tag{110}$$

a constant together with a diagonal term. Transforming each leads to

$$\mathbb{E}\,\hat{u}_{j_0}(k) = \beta\,\nu\,\delta_{k0}, \tag{111}$$

$$\mathbb{E}\big[\hat{u}_{j_0}(k)\overline{\hat{u}_{j_0}(k')}\big] = \beta\,\nu\,\delta_{kk'} + \frac{4^m-3\cdot 2^m+2}{N^2}\,\nu^2\,\delta_{k0}\delta_{k'0}. \tag{112}$$

Equation (111) vanishes at $k\neq 0$, and Eq. (112) vanishes at $k\neq k'$, since then the two indices cannot both be 0.

Step 2, multiplying over groups. Distinct groups use disjoint weights and all weights are independent, so the expectation of a product factors into the product of the expectations,

$$\mathbb{E}\,\hat{X}(k) = \prod_j \mathbb{E}\,\hat{u}_j(k), \tag{113}$$

$$\mathbb{E}\big[\hat{X}(k)\overline{\hat{X}(k')}\big] = \prod_j \mathbb{E}\big[\hat{u}_j(k)\overline{\hat{u}_j(k')}\big]. \tag{114}$$

There are four cases in (k,k').

(a) First moment. $\mathbb{E}\,\hat{u}_{j_0}(k)=0$ by Eq. (111), so one factor of the product vanishes and $\mathbb{E}\,\hat{X}(k)=0$. Substituting into the inverse transform $X_\rho = \frac{1}{\nu}\sum_k \hat{X}(k)e^{2\pi ik\rho/\nu}$ and taking expectations, only $k=0$ survives as

$$\mathbb{E}X_\rho = \frac{1}{\nu}\,\mathbb{E}\,\hat{X}(0) = \frac{\mathbb{E}P}{\nu} = \mu \tag{115}$$

that is independent of ρ.

(b) Second moment. By Eq. (112) one factor is again zero for $k\neq k'$, so $\mathbb{E}[\hat{X}(k)\overline{\hat{X}(k')}]=0$. As for $k=k'\neq 0$, we compute $\mathbb{E}|\hat{u}_j(k)|^2$ group by group from

$$\mathbb{E}\big|\hat{u}_j(k)\big|^2 = \sum_{\rho,\rho'\in\mathbb{Z}_\nu} \mathbb{E}\big[u_j(\rho)u_j(\rho')\big]\, e^{-2\pi ik(\rho-\rho')/\nu}, \tag{116}$$

taking the four terms of Eq. (48) in turn. At $k\neq 0$ only the atom and the diagonal survive. A group with $s_j=0$ carries both and a group with $s_j\neq 0$ carries only the diagonal, so in either case

$$\mathbb{E}\big|\hat{u}_j(k)\big|^2 = (2^m-1)2^{-L} + \mathbf{1}\big[s_j{=}0\big] = \mathbb{E}X_{s_j} \qquad (k\neq 0), \tag{117}$$

the right side being the block mean of Eq. (95). Multiplying over j and using independence to turn the product back into an expectation,

$$\begin{aligned}\mathbb{E}\big|\hat{X}(k)\big|^2 &= \prod_j \mathbb{E}X_{s_j} \\ &= \mathbb{E}\Big[\prod_j X_{s_j}\Big] = \mathbb{E}P = \nu\mu \qquad (k\neq 0).\end{aligned} \tag{118}$$

If $k=k'=0$, then $\hat{X}(0)=P$ and the second moment is $\mathbb{E}[P^2]$, used only in Step 4.

These four cases also show that Y_ρ and P are uncorrelated. We have $P=\hat{X}(0)$ while Y_ρ is a linear combination of the $\hat{X}(k)$ with $k\neq 0$, and term by term

$$\begin{aligned}\mathrm{Cov}\big(\hat{X}(k),\hat{X}(0)\big) &= \mathbb{E}\big[\hat{X}(k)\overline{\hat{X}(0)}\big] - \mathbb{E}\,\hat{X}(k)\;\overline{\mathbb{E}\,\hat{X}(0)} \\ &= 0 \qquad (k\neq 0).\end{aligned} \tag{119}$$

The variance of a single bin is therefore the sum of the two pieces with no cross term,

$$\mathrm{Var}\big(X_\rho\big) = \frac{\mathrm{Var}(P)}{\nu^2} + \mu\Big(1-\frac{1}{\nu}\Big), \tag{120}$$

the first piece from the ball count moving and the second from the balls scattering.

Step 3, back to the bins. Splitting the inverse transform by frequency,

$$\begin{aligned}X_\rho &= \frac{1}{\nu}\sum_k \hat{X}(k)\, e^{2\pi ik\rho/\nu} \\ &= \underbrace{\frac{\hat{X}(0)}{\nu}}_{=\,P/\nu} + \underbrace{\frac{1}{\nu}\sum_{k\neq 0}\hat{X}(k)\, e^{2\pi ik\rho/\nu}}_{=\,Y_\rho}.\end{aligned} \tag{121}$$

Since $\mathbb{E}Y_\rho=0$, expanding both factors and using $\overline{\hat{X}(k')} = \hat{X}(-k')$ gives the covariance of two bins ρ and ρ' as

$$\mathrm{Cov}\big(Y_\rho,Y_{\rho'}\big) = \mathbb{E}\big[Y_\rho Y_{\rho'}\big] \tag{122}$$

$$= \frac{1}{\nu^2}\sum_{k\neq 0}\sum_{k'\neq 0}\mathbb{E}\big[\hat{X}(k)\overline{\hat{X}(k')}\big]\, e^{2\pi i(k\rho-k'\rho')/\nu} \tag{123}$$

$$= \frac{\nu\mu}{\nu^2}\sum_{k\neq 0} e^{2\pi ik(\rho-\rho')/\nu} \tag{124}$$

$$= \frac{\mu}{\nu}\big(\nu\,\delta_{\rho\rho'}-1\big) = \mu\,\delta_{\rho\rho'} - \frac{\mu}{\nu}, \tag{125}$$

where only the terms with $k=k'$ survive.

Step 4, the ball count. The factors of $P=\prod_j X_{s_j}$ are independent, so $\mathbb{E}[P^2]=\prod_j \mathbb{E}[X_{s_j}^2]$ and $(\mathbb{E}P)^2 = \prod_j(\mathbb{E}X_{s_j})^2$, giving

$$\frac{\mathrm{Var}(P)}{\big(\mathbb{E}P\big)^2} = \prod_j \frac{\mathbb{E}\big[X_{s_j}^2\big]}{\big(\mathbb{E}X_{s_j}\big)^2} - 1 = \prod_j\Big[1+\frac{\mathrm{Var}(X_{s_j})}{\big(\mathbb{E}X_{s_j}\big)^2}\Big] - 1. \tag{126}$$

By Lemma F the numerator is $\mathrm{Var}(X_{s_j}) = (2^m-1)2^{-L}(1-2^{-L})$, and $\mathbb{E}X_{s_j}\geq (2^m-1)2^{-L}$, so each bracket carries a term at most

$$\frac{(2^m-1)2^{-L}(1-2^{-L})}{\big((2^m-1)2^{-L}\big)^2} = \frac{2^L-1}{2^m-1} \;\approx\; 2^{\,L-m}. \tag{127}$$

With $a = n/\log n$ factors and $(1+x)^a - 1 \approx ax$, we finally obtain

$$\frac{\mathrm{Var}(P)}{\big(\mathbb{E}P\big)^2} \;\approx\; a\,2^{\,L-m} = \frac{n}{\log n}\cdot n^{\,1-c} = \frac{n^{\,2-c}}{\log n}, \qquad (128)$$

whose square root is $n^{(2-c)/2}(\log n)^{-1/2}$. □

E. Fitting the two levels together

Lemma H splits the bin count into $X_\rho = P/\nu + Y_\rho$ and shows that the ball count P fluctuates, an effect a fixed-count picture leaves out. This subsection shows that the extra term does not affect the conclusion of Lemma 4, because P takes one value across both branches and cancels from the comparison.

Following Eq. (88), we first sort the ϕ_A by the string ϕ_{g_a} of bits in g_a, and keep the convention that z_h^* is determined by $s_a = \mathrm{top}(r_{g_a})$. Fixing ϕ_{g_a} contributes a definite ρ_{g_a} to the low part, so this class holds $X_{\rho-\rho_{g_a}}$ configurations with $z_l^* = \rho$. Substituting Eq. (88) into Eq. (79) then gives

$$\alpha_{z_h^*} = \lambda_4 \sum_{\rho\in\mathbb{Z}_\nu} \Big(\sum_{\phi_{g_a}\,:\,\mathrm{top}(r_{g_a})=s_a} X_{\rho-\rho_{g_a}} \Big) C_\rho. \qquad (129)$$

Now put $X_{\rho-\rho_{g_a}} = P/\nu + Y_{\rho-\rho_{g_a}}$ into the sum. The first piece is a constant and depends on neither ρ nor ϕ_{g_a}, so the two sums separate as

$$\begin{aligned} \alpha_{z_h^*} = {} & \underbrace{\lambda_4 \frac{P}{\nu}\Big(\sum_\rho C_\rho\Big) X_{s_a}}_{\text{main term}} \\ & + \underbrace{\lambda_4 \sum_{\phi_{g_a}\,:\,\mathrm{top}(r_{g_a})=s_a} \; \sum_\rho Y_{\rho-\rho_{g_a}} C_\rho}_{=:\,R_{h^*}}, \end{aligned} \qquad (130)$$

where X_{s_a} counts the strings ϕ_{g_a} consistent with s_a, and R_{h^*} names the error term, one number per branch.

Everything in the main term except X_{s_a} takes the same value on the two branches. The coefficient $\lambda_4 = (E_0^2 + E_1^2)^{-1/2}$ normalises the two branches together and is one number per record. The ball count $P = \prod_{j<a} X_{W'_j}$ is fixed by the measured W' and by the weights of the groups $j < a$, and neither h^* nor z_h^* occurs in it. Writing $J := \lambda_4 (P/\nu) \sum_\rho C_\rho$, the main term becomes $J\,X_{s_a}$ with the same J on both branches.

For the main term, the quantities s_a and z_h^* differ by the fixed shift $s^* = \sum_{j<a} W'_j + s_a$, so flipping the top bit of z_h^* flips the top bit of s_a and Theorem G applies directly as

$$\frac{\big|X_{s_{a,0}} - X_{s_{a,1}}\big|}{\mathbb{E}X_{s_a}} \;\le\; 2\,n^{-\left(\frac{c-1}{2}-1\right)}, \qquad (131)$$

where each block departs from the mean by at most $n^{-(\frac{c-1}{2}-1)}$, with failure probability still $1/n$. If one of the two blocks is $s_a = 0$, its mean carries the extra empty subset of Eq. (95) in a relative amount $2^L/(2^m-1) \approx n^{1-c}$. For $c \ge 12$ that lies far below $n^{-((c-1)/2-1)}$ and is absorbed. The ball count P is a common factor of the main term and factors out of the difference, so the size of its fluctuation is immaterial.

The remaining quantity to control is R_{h^*}, which involves only Y. Lemma H asks that at least one pinned block be nonzero. All $a-1$ of them vanish with probability $2^{-L(a-1)} = 2^{-(n-\log n)}$, which joins the other failure probabilities and is negligible against them.

The lemma gives $\mathrm{Var}(Y_\rho) = \mu(1-1/\nu)$ with $\mu = \mathbb{E}[P]/\nu$, and $\mathbb{E}[P]$ follows from Eq. (95),

$$\mathbb{E}[P] = \prod_{j<a} \Big((2^m-1)2^{-L} + \mathbf{1}[s_j{=}0] \Big) \;\approx\; 2^{\,(a-1)(m-L)}. \qquad (132)$$

With $am = cn$ and $aL = n$, and with m replaced by $m(1-1/(c'\log n))$ to carry the faulty samples of Section V C,

$$\mu = \frac{\mathbb{E}[P]}{\nu} = 2^{\,cn(1-1/(c'\log n))-2n-(c-2)\log n}, \qquad (133)$$

$$\sigma = \sqrt{\mu}. \qquad (134)$$

Chebyshev bounds the probability that one bin violates $|Y_\rho| \le 2^n\sigma$ by 2^{-2n}. Following Section II E, the union bound runs over the $\nu = 2^{\,n-L}$ bins and over all $2^{\,L(a-1)} = 2^{\,n-\log n}$ values of the label vector W'. The exceptional event that survives has probability

$$\underbrace{2^{\,n-\log n}}_{W'} \cdot \underbrace{2^{\,n-\log n}}_{\text{bins}} \cdot 2^{-2n} \;=\; 2^{-2\log n} \;=\; n^{-2}, \qquad (135)$$

which joins the other failure probabilities of order $1/n$. Outside it every bin sits within a relative $2^n\sigma/\mu$ of its share. Taken at a single ρ, the summand of Eq. (129) is the amplitude α_{z^*} with $z_l^* = \rho$. The matching summand of R_{h^*} is at most $2^n\sigma/\mu$ times that amplitude, and Theorem E at $\theta = 2^{\,3n/2}$ bounds it. Summing over the ν values of ρ,

$$\big|R_{h^*}\big| \;\le\; R_{\max} \;:=\; \underbrace{\frac{2^n\sigma}{\mu}}_{\text{one bin}} \cdot \underbrace{\nu}_{\text{summands}} \cdot \underbrace{2^{\,3n/2}}_{\text{Lemma 3}}. \qquad (136)$$

Taking logarithms at $c = 12$,

$$\log R_{\max} \;\le\; -\frac{3n}{2} \;+\; \frac{6n}{c'\log n} \;+\; 4\log n. \qquad (137)$$

The middle term is $o(n)$ for every constant faulty rate c', so $R_{\max} \le 2^{-3n/2+o(n)}$, comfortably below the 2^{-n} the argument needs.

F. Path to the algorithm

Everything needed for the *difference* of the two aggregated amplitudes is now in place, and the difference is

the weaker of the two statements at issue. Equation (130) reads $\alpha_{z^*_h} = J\,X_{s_a} + R_{h^*}$ on each branch, and the prefactor J is the same number for the two members of a pair. Subtracting the two copies therefore leaves it outside the bracket,

$$\alpha_{z^*_{h,0}} - \alpha_{z^*_{h,1}} = J\big(X_{s_{a,0}} - X_{s_{a,1}}\big) + \big(R_0 - R_1\big). \quad (138)$$

The two terms are handled separately. Equation (131) bounds the count difference by $2n^{-((c-1)/2-1)}\mathbb{E}X_{s_a}$, and each error term obeys $|R_{h^*}| \le R_{\max}$, so

$$\big|\alpha_{z^*_{h,0}} - \alpha_{z^*_{h,1}}\big| \;\le\; 2\,n^{-\left(\frac{c-1}{2}-1\right)}\,\big|J\big|\,\mathbb{E}X_{s_a} \;+\; 2R_{\max}. \quad (139)$$

Meanwhile, Theorem G in relative form gives $|X_{s_{a,0}} - \mathbb{E}X_{s_a}| \le n^{-((c-1)/2-1)}\mathbb{E}X_{s_a}$ on the same record, leading to

$$\mathbb{E}X_{s_a} \;\le\; \frac{X_{s_{a,0}}}{1 - n^{-((c-1)/2-1)}}. \quad (140)$$

Multiplying by $|J|$ and reading Eq. (130) backwards ($J\,X_{s_{a,0}} = \alpha_{z^*_{h,0}} - R_0$), we have

$$\begin{aligned} \big|J\big|\,\mathbb{E}X_{s_a} \;&\le\; \frac{\big|J\,X_{s_{a,0}}\big|}{1 - n^{-((c-1)/2-1)}} \\ &=\; \frac{\big|\alpha_{z^*_{h,0}} - R_0\big|}{1 - n^{-((c-1)/2-1)}} \\ &\le\; \big(1 + o(1)\big)\Big(\big|\alpha_{z^*_{h,0}}\big| + R_{\max}\Big), \end{aligned} \quad (141)$$

the last line by the triangle inequality with $|R_0| \le R_{\max}$, together with $\big(1 - n^{-((c-1)/2-1)}\big)^{-1} = 1 + o(1)$ once $c \ge 12$. The corollary to Lemma 3 bounds $|\alpha_{z^*_{h,0}}|$, so every quantity on the right of Eq. (139) is already bounded. Nothing along the way needs J to be large, or even to have a definite sign.

Lemma 4 is stated as a bound on the *ratio* in the form $1 \pm O(n^{-((c-1)/2-1)})$, which is strictly stronger than Eq. (139). To be more specific,

$$\frac{\alpha_{z^*_{h,0}}}{\alpha_{z^*_{h,1}}} = \frac{X_{s_{a,0}} + R_0/J}{X_{s_{a,1}} + R_1/J}, \quad (142)$$

which is close to 1 only when $|J|\,X_{s_a}$ dominates $R_{\max}$. A pair whose prefactor satisfies $|J|\,X_{s_a} \lesssim R_{\max}$ has a ratio nowhere near 1. Such a pair can occur, since J carries the signed sum $\sum_\rho C_\rho$ over the $\nu = 2^{\,n-L}$ values of z^*_l and no argument keeps such a sum far from zero on every record.

This section therefore establishes Eq. (139) and not the ratio. The rest of the section shows that the difference is all the closing argument ever consumes, so nothing downstream is weakened. The index z^*_h carries $L = \log n$ bits and therefore takes n values, which form $n/2$ pairs differing only in the top bit h^*. Grouping Eq. (21) by pairs, the quantity to bound is a sum of $n/2$ *differences*,

$$\big|\hat{E}_0 - \hat{E}_1\big| \;\le\; \sum_{\text{pairs}} \big|\alpha_{z^*_{h,0}} - \alpha_{z^*_{h,1}}\big|, \quad (143)$$

which is the form Eq. (139) was stated in. Both sides carry the normalisation λ_4, since $\alpha_{z^*_h}$ does.

By Theorem E′ at $\theta = n^{3/2}$, every aggregated amplitude obeys $|\alpha_{z^*_h}| \le n^{3/2}$ except with probability n^{-3}, so Eqs. (139) and (141) bound each summand by $2n^{-((c-1)/2-1)}(n^{3/2} + R_{\max})(1 + o(1)) + 2R_{\max}$. Since $R_{\max} \le 2^{-3n/2+o(n)}$ is exponentially small, the $R_{\max}$ terms are negligible against everything else and

$$\begin{aligned} \big|\hat{E}_0 - \hat{E}_1\big| \;&\le\; \frac{n}{2} \cdot 2\,n^{3/2} \cdot O\big(n^{-((c-1)/2-1)}\big) \\ &=\; O\big(n^{\,7/2-(c-1)/2}\big) \;\overset{c\ge 12}{\le}\; O(n^{-2}), \end{aligned} \quad (144)$$

with failure probability $O(1/n)$, and the exponent is consistent with the one obtained from the multiplicative form, so the weaker statement loses nothing.

The two branch amplitudes are now pinned. Since $\hat{E}_0^2 + \hat{E}_1^2 = 1$, at least one of $|\hat{E}_0|$ and $|\hat{E}_1|$ is at least $1/\sqrt{2}$. Equation (144) makes the two differ by $O(n^{-2})$, so both are within $O(n^{-2})$ of $1/\sqrt{2}$ and carry the same sign. The normalisation cancels from the ratio, so

$$q \;=\; \frac{\big|E_0 - E_1\big|}{\big|E_0 + E_1\big|} \;=\; \frac{\big|\hat{E}_0 - \hat{E}_1\big|}{\big|\hat{E}_0 + \hat{E}_1\big|} \;=\; O\big(n^{-2}\big), \quad (145)$$

and Eq. (15) gives $p = 1 - O(n^{-4})$, which exceeds $1/2$ by a constant. Polynomially many repetitions and a majority vote then return d_n, and the recursion on the remaining bits of d finishes the algorithm.

VI. NUMERICAL CHECKS

The parameters of [15] put a direct simulation of the algorithm out of reach, since $c \ge 12$ makes $Q = Kn^{c+1}$ exceed 10^{16} already at $n = 16$. The three results proved above are statements about classical counting quantities, and those can be evaluated in full at small parameters. Figure 1 collects one check for each.

Panel (a) plots the two terms of Eq. (58) against n at the parameters of [15]. The correlation coefficient η is obtained in closed form from Eq. (49) in exact rational arithmetic, using the averages of Eqs. (50), (A9) and (A10). The value agrees with a direct sum over k for $n \le 16$. The curve rises while 2^m exceeds N and falls once 2^m drops below it, crossing the Poisson term $c/(Ka)$ at $n \approx 164$. Beyond that point η contributes nothing and the bound of Theorem C is carried by $1/\mathbb{E}[\mathcal{Z}]$ alone, which leaves $\Pr[\mathcal{Z} \ge a]$ at least 0.81 for $n = 256$ and at least 0.96 for $n = 2048$.

Panel (b) takes a toy instance small enough to enumerate, with $n = 6$ and $Q = 20$ sample bits cut into $G = 4$ groups of $m = 5$, of which $a = 2$ sit on the A side. Every one of the 2^{10} strings ϕ'_B is summed over, which imposes Hypothesis (U) by construction. The measured tail runs a factor of 20 below θ^{-2} at $\theta = 1$ and a factor of 10^3 below it at $\theta = 16$, so the bound of Theorem E is comfortable as well as valid.

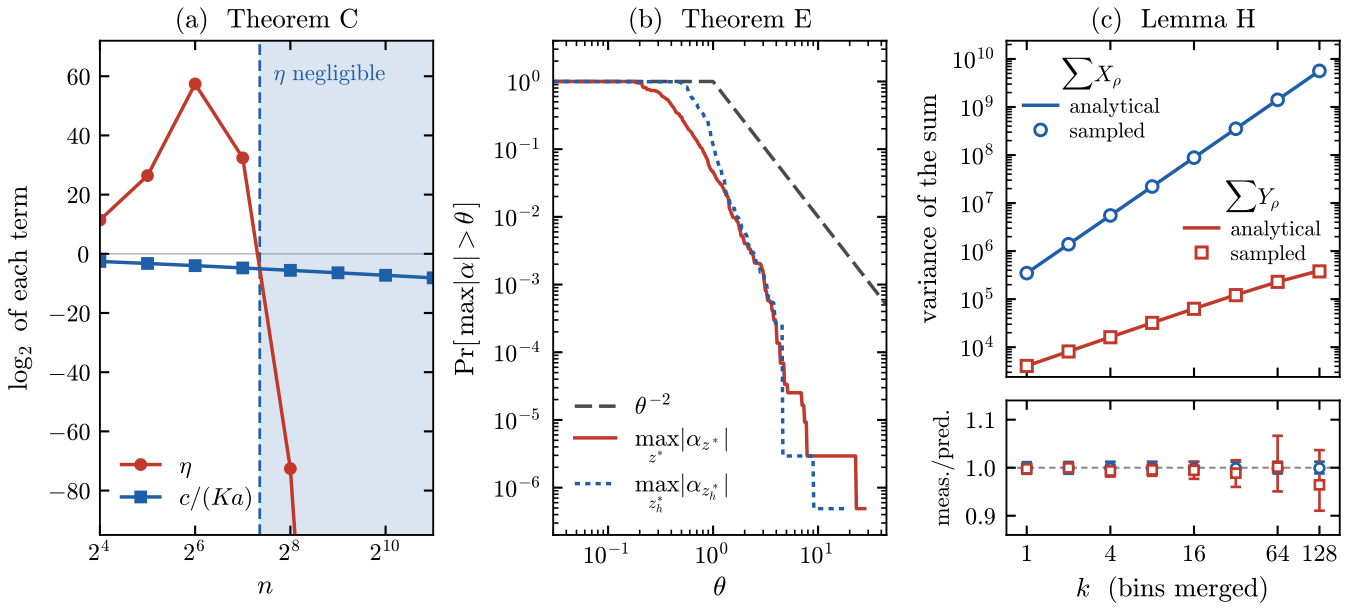

FIG. 1. One numerical check for each of the three results. (a) The two terms that bound $\mathrm{Var}[\mathcal{Z}]/\mathbb{E}[\mathcal{Z}]^2$ in Eq. (58), at $c = 12$ and $K = 1.5c$. The correlation term η is computed exactly and drops below the Poisson term at $n \approx 164$. (b) Measured tail of the amplitude against the bound of Theorem E, taken over the full cube of 2^{10} measurement outcomes of a toy instance with $n = 6$ and $Q = 20$. (c) Variance of k merged bins in the second balls-in-bins model, at $n = 12$, $m = 10$, $L = 3$, $\nu = 512$ and three pinned groups, from 2×10^5 samples. The upper curve carries the ball-count term and the lower one does not. The lines are computed from Lemma H, with μ and $\mathrm{Var}(P)$ taken from Eq. (95) and Eq. (108), while the points are measured from the samples. The strip below gives the measured value over the computed one, with 95% bootstrap intervals.

Panel (c) tests Lemma H by merging k adjacent bins. Since Y_ρ and P are uncorrelated, Eq. (107) predicts $\mathrm{Var}(\sum_{\rho<k} Y_\rho) = \mu k(1 - k/\nu)$, while the same sum of the bin counts themselves carries the further term $k^2\mathrm{Var}(P)/\nu^2$. Both are computed from the lemma alone, with μ taken from Eq. (95) and $\mathrm{Var}(P)$ from Eq. (108). The measured values come from 2×10^5 sampled runs and agree with both. The gap between the two curves is the term a fixed ball count leaves out, and it exceeds the scattering term by orders of magnitude. Therefore, the ball count governs the bin variance completely, and Section V E is where it cancels between the two branches.

None of these checks bears on Hypothesis (U), since panel (b) imposes it by construction and the other two never meet it. The agreement reported here therefore says nothing about the one gap recorded in Section VII.

VII. CONCLUSIONS AND DISCUSSIONS

In this paper we have restated the three sketched lemmas of [15] in a form that admits a single reading and have proved each of them, as summarized in Table VI. Only Lemma 4 comes out weaker than the original claim, with a bound on the difference of the two branch amplitudes in place of their ratio, and Section V E shows that the closing argument consumes only the difference.

One condition survives all of this. Hypothesis (U) asks that the A/B partition be fixed independently of ϕ', so that ϕ'_B ranges over the whole cube $\{0,1\}^{Q_B}$ with the rest of the record held fixed. However, the algorithm itself does not supply it. It requires A to consist of the first a all-zero groups, which makes the partition a function of ϕ'. The hypothesis enters at a single step in Lemma D,

$$\sum_{\phi'_B \in \{0,1\}^{Q_B}} E_{h^*}(\phi')^2 = 2^{Q_B} \sum_{\phi_B} \mathcal{N}_{h^*}(\phi_B)^2, \qquad (146)$$

where the character sum over the whole cube collapses to the diagonal $\phi_B = \psi_B$ and leaves the factor 2^{Q_B}.

Narrowing that range is not a perturbation of the identity. A group that the rule forces away from the all-zero block contributes

$$\sum_{\phi'_j \neq 0} (-1)^{(\phi_j \oplus \psi_j)\cdot\phi'_j} \;=\; 2^m \delta_{\phi_j \psi_j} - 1 \qquad (147)$$

in place of $2^m \delta_{\phi_j \psi_j}$, so Eq. (146) acquires one term for every subset of the forced groups and the off-diagonal ones no longer cancel. Controlling them would need the autocorrelation of $\mathcal{N}_{h^*}$, which is the one object Lemma D is built to avoid. Nor does it help that the discarded strings are few, because Eq. (146) holds only after the whole cube has been summed. Its left side collects squares of signed sums and its right side squares of counts, and the two are unrelated at a fixed ϕ'_B.

TABLE VI. The four lemmas in the forms proved here.

In [15]	Status
Lemma 1	Holds, with probability tending to one (Theorem C), stronger than the constant claimed
Lemma 2	Exact as published, with a two-line proof (Section II)
Lemma 3 and Corollary	The amplitude bound holds at every threshold (Theorem E, E′) and needs no well-behavedness hypothesis, so the first half of the lemma never enters and is recorded only in Appendix B
Lemma 4	Both counting estimates hold, the first with the constant [15] displays (Theorem G), the second together with the term a fixed ball count leaves out (Lemma H), and the assumption that g_a carries no faulty samples can be dropped (Section V C). The comparison of the two branches is proved in the weaker form of Eq. (139), which is the form the closing argument uses.

The two demands on A pull against each other, and that is why a change of rule does not remove the point. The groups of A have to be all-zero, since only then do they carry no sign and leave the A-side factor an unsigned count that both branches of the final interference share. Selecting them therefore means looking at ϕ', and that look costs the freedom Hypothesis (U) asks for. Fixing the index set before Step 4 would grant the hypothesis and give up the all-zero property. Therefore, establishing these four lemmas does not by itself establish the correctness of the algorithm.

Appendix A: The second moment of a group's subset-sum count

This appendix derives Eq. (48) and Eq. (49), together with the two averages of the Fejér kernel quoted in Eq. (50).

1. The four classes of pairs

Write $\Sigma(\phi) = \sum_{i\in g_j} \phi_i y_i \bmod N$, so that $n_j[r] = \sum_{\phi\in\{0,1\}^m} \mathbf{1}[\Sigma(\phi) = r]$. Multiplying two of these gives a double sum over pairs of subsets,

$$\mathbb{E}\big[n_j[r]\, n_j[r']\big] = \sum_{\phi,\psi} \Pr_y\big[\Sigma(\phi) = r,\ \Sigma(\psi) = r'\big]. \quad \text{(A1)}$$

For a single subset, $\Pr_y[\Sigma(\phi) = r]$ depends only on whether ϕ is empty. The empty subset gives $\Sigma \equiv 0$ with certainty, and any nonempty subset gives Σ uniform on $\mathbb{Z}_N$, since a sum of independent uniform values is uniform. Hence

$$\mathbb{E}\big[n_j[r]\big] = \underbrace{\mathbf{1}[r=0]}_{\phi=0} + \underbrace{\beta}_{\phi\neq 0}, \qquad \beta = \frac{2^m-1}{N}. \quad \text{(A2)}$$

For a pair, the same distinction applies to each member, and the four resulting classes are collected in Table VII. In the last case, ϕ and ψ are distinct and neither is empty, so the two sums are jointly uniform on $\mathbb{Z}_N^2$. Adding the four contributions gives Eq. (48).

2. Carrying the four classes through $T(k)$

Substitute Table VII into $T_j(k) = \sum_s \sum_{r,r'\in s} n_j[r] n_j[r']\, \omega^{-k(r-r')}$, where the inner sum runs over the r with the top L bits fixed and the low $n-L$ bits free.

The diagonal term contributes $\beta \sum_s \sum_{r\in s} 1 = \beta N = 2^m - 1$, and the atom contributes 1, the two together making 2^m.

For the constant term, $\mathbb{Z}_N$ is cut into 2^L blocks of length ν, and the bottom one is $B_0 = \{0, 1, \ldots, \nu-1\}$. Its indicator has transform

$$\hat{\mathbf{1}}_{B_0}(k) = \sum_{r=0}^{\nu-1} \omega^{-kr} = \begin{cases} \nu, & k = 0, \\ \dfrac{1-\omega^{-k\nu}}{1-\omega^{-k}}, & k \neq 0, \end{cases} \quad \text{(A3)}$$

and taking the squared modulus with $|1 - e^{-i\theta}|^2 = 4\sin^2(\theta/2)$ and $\nu/N = 2^{-L}$

$$D_\nu(k) := \big|\hat{\mathbf{1}}_{B_0}(k)\big|^2 = \frac{\sin^2(\pi k\nu/N)}{\sin^2(\pi k/N)} = \frac{\sin^2(\pi k/2^L)}{\sin^2(\pi k/N)}, \quad \text{(A4)}$$

the unnormalised Fejér kernel of order ν. Every block is a translate of B_0 and translation contributes only a phase, so $\sum_s |\sum_{r\in s} \omega^{-kr}|^2 = 2^L D_\nu(k)$, and the constant term contributes $\beta_2 D_\nu(k)$.

For the cross-atom terms, one of the two indices is pinned to 0 and the other runs over B_0. The two cases give complex conjugate expressions, so together they contribute $2\beta \operatorname{Re} \hat{\mathbf{1}}_{B_0}(k) = \beta_1 \operatorname{Re} \hat{\mathbf{1}}_{B_0}(k)$. Collecting the four gives Eq. (49).

3. Averages over the frequency line

Parseval with $f = g = \mathbf{1}_{B_0}$ gives the first

$$\big\langle D_\nu \big\rangle_k = \frac{1}{N} \sum_k \big|\hat{\mathbf{1}}_{B_0}(k)\big|^2 = \sum_r \mathbf{1}_{B_0}[r]^2 = \nu. \quad \text{(A5)}$$

For the second, we use $\widehat{f \star f}(k) = |\hat{f}(k)|^2$ from the proof of Lemma B to give

$$\frac{1}{N} \sum_k D_\nu(k)\, \omega^{k\delta} = \big(\mathbf{1}_{B_0} \star \mathbf{1}_{B_0}\big)[\delta] =: \Lambda[\delta], \quad \text{(A6)}$$

TABLE VII. The four classes of subset pairs and what each contributes to Eq. (48).

Class	Condition	Pairs	$\Pr[\Sigma_\phi{=}r, \Sigma_\psi{=}r']$	Contribution
atom	$\phi=\psi=0$	1	$\mathbf{1}[r{=}0]\mathbf{1}[r'{=}0]$	$\mathbf{1}[r{=}r'{=}0]$
cross-atom	exactly one empty	$2(2^m-1)$	$\frac{1}{N}\mathbf{1}[r{=}0]$ or $\frac{1}{N}\mathbf{1}[r'{=}0]$	$\beta\big(\mathbf{1}[r{=}0]+\mathbf{1}[r'{=}0]\big)$
diagonal	$\phi=\psi\neq 0$	2^m-1	$\frac{1}{N}\delta_{rr'}$	$\beta\,\delta_{rr'}$
constant	$\phi\neq\psi$, neither empty	$(2^m-1)(2^m-2)$	$\dfrac{1}{N^2}$	$\dfrac{4^m-3\cdot 2^m+2}{N^2}$

and since $\mathbf{1}_{B_0}$ takes values in $\{0,1\}$,

$$\begin{aligned}\Lambda[\delta] &= \#\big\{r : r\in B_0 \text{ and } r-\delta\in B_0\big\}\\ &= \big|B_0\cap(B_0+\delta)\big| = (\nu-|\delta|)_+,\end{aligned} \tag{A7}$$

the overlap of two intervals of length ν offset by δ, with $(x)_+=\max(x,0)$. Applying Parseval once more to Λ gives

$$\begin{aligned}\big\langle D_\nu^2\big\rangle_k &= \frac{1}{N}\sum_k\big|\hat\Lambda(k)\big|^2=\sum_\delta\Lambda[\delta]^2\\ &= \underbrace{\nu^2}_{\delta=0}+\underbrace{2\sum_{j=1}^{\nu-1}j^2}_{\delta=\pm1,\dots,\pm(\nu-1)} = \frac{\nu\big(2\nu^2+1\big)}{3}\ \le\ \nu^3,\end{aligned} \tag{A8}$$

which is Eq. (50).

Three further averages are used by the numerical check of Section VI, and the same two tools deliver them. Writing $f=\mathbf{1}_{B_0}$ and reading the inverse transform at $\delta=0$ gives $\langle\hat f\rangle_k=f[0]=1$, so $\langle\operatorname{Re}\hat f\rangle_k=1$. Since $\hat f(k)^2=\widehat{f\star f}(k)$ up to the reflection $r\mapsto -r$, the same reading gives $\langle\hat f^2\rangle_k=\#\{r\in B_0:-r\in B_0\}=1$, where the only such r is 0 because $\nu\le N/2$. Expanding $(\operatorname{Re}\hat f)^2=\frac14(\hat f^2+2|\hat f|^2+\overline{\hat f}^{\,2})$ and using $\langle|\hat f|^2\rangle_k=\nu$, we have

$$\big\langle(\operatorname{Re}\hat{\mathbf{1}}_{B_0})^2\big\rangle_k=\frac{1+2\nu+1}{4}=\frac{\nu+1}{2}. \tag{A9}$$

For the last one, $\hat f^2\overline{\hat f}$ at $\delta=0$ counts the triples of B_0 with $r_1+r_2=r_3$, so

$$\begin{aligned}\big\langle D_\nu\operatorname{Re}\hat{\mathbf{1}}_{B_0}\big\rangle_k &= \#\big\{(r_1,r_2)\in B_0^2 : r_1+r_2\le\nu-1\big\}\\ &= \frac{\nu(\nu+1)}{2}.\end{aligned} \tag{A10}$$

Appendix B: The first half of Lemma 3

This appendix records the other half of the statement quoted at the start of Section IV. Its role there is to make the second half non-vacuous, and the proof given in Section IV B does not need that service.

The signed count of Section II is $E_{h^*}=t_M^+-t_M^-$, so being well-behaved for h^* says $\big|E_{h^*}\big|\ge\varepsilon\sqrt{t_{M,h^*}}$. The dependence of t_M on h^* is written explicitly, since T_M is by definition the set consistent with the record and a given h^*, and the two branches carry their own counts. Failing to be well-behaved is the same as carrying little weight. The predicate of Definition 1 contains $\Omega(\cdot)$, which names a class of functions and not a number, so the clause becomes a definition only once a constant is fixed. That constant matters, because the failure probability scales as its square. The threshold is therefore written as $\varepsilon\sqrt{t_{M,h^*}}$ and the conclusion as ε^2.

One consequence of Lemma D is needed and was not drawn there. The function $\mathcal{N}_{h^*}(\phi_B)$ takes nonnegative integer values, so $\mathcal{N}_{h^*}^2\ge\mathcal{N}_{h^*}$ pointwise, and summing with $\sum_{\phi_B}\mathcal{N}_{h^*}(\phi_B)=t_{M,h^*}$ turns Eq. (65) into

$$\sum_{\phi'_B\in\{0,1\}^{Q_B}}E_{h^*}(\phi')^2\ \ge\ 2^{Q_B}\,t_{M,h^*}. \tag{B1}$$

Theorem I. *Assume Hypothesis (U). Fix $\varepsilon>0$ and call the superposition well-behaved for h^* when $|E_{h^*}(\phi')|\ge\varepsilon\sqrt{t_{M,h^*}}$. The probability that it is well-behaved for neither value of h^* is then at most ε^2. Taking $\varepsilon=2^{-n/2}$ gives 2^{-n}, which is the conclusion of the first half of Lemma 3 of [15].*

Proof. Fix everything in the record except ϕ', i.e., Y, z', W', S, h' and the A/B partition, and let ϕ'_B vary. The weight of a record is proportional to $E_0(\phi')^2+E_1(\phi')^2$, so the quantity to bound is

$$\begin{aligned}&\Pr\big[\text{neither } h^* \text{ well-behaved}\big]\\ &= \frac{\sum_{\phi'_B\ \text{failing both}}\big(E_0(\phi')^2+E_1(\phi')^2\big)}{\sum_{\phi'_B\in\{0,1\}^{Q_B}}\big(E_0(\phi')^2+E_1(\phi')^2\big)}.\end{aligned} \tag{B2}$$

A record failing both satisfies $E_0(\phi')^2<\varepsilon^2 t_{M,0}$ and $E_1(\phi')^2<\varepsilon^2 t_{M,1}$, so $E_0(\phi')^2+E_1(\phi')^2<\varepsilon^2(t_{M,0}+t_{M,1})$. There are at most 2^{Q_B} such strings, so the numerator is below $2^{Q_B}\varepsilon^2(t_{M,0}+t_{M,1})$. Applying Eq. (B1) to each branch, the denominator is

$$2^{Q_B}\sum_{\phi_B}\big(\mathcal{N}_0(\phi_B)^2+\mathcal{N}_1(\phi_B)^2\big)\ \ge\ 2^{Q_B}\big(t_{M,0}+t_{M,1}\big). \tag{B3}$$

Dividing the numerator by the denominator, both 2^{Q_B} and $t_{M,0}+t_{M,1}$ cancel and ε^2 remains. □

Note added. Since [15] appeared, its lemmas have been the subject of active public scrutiny. We became aware of parts of that discussion while this work was in preparation. Where our findings overlap with analyses circulated elsewhere we claim no priority. Our results were obtained independently and are stated so that each can be checked against [15] directly.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (NSFC) (Grant No. 12475022 and No. 125B2100) and the Quantum Science and Technology - National Science and Technology Major Project (Grant No. 2021ZD0302100).

[1] M. Ajtai, *Generating Hard Instances of Lattice Problems*, STOC 1996, pp. 99–108.

[2] D. Bacon, A. M. Childs and W. van Dam, *From Optimal Measurement to Efficient Quantum Algorithms for the Hidden Subgroup Problem over Semidirect Product Groups*, FOCS 2005, pp. 469–478.

[3] Z. Brakerski, E. Kirshanova, D. Stehlé and W. Wen, *Learning with Errors and Extrapolated Dihedral Cosets*, PKC 2018, pp. 702–727.

[4] Y. Chen, *Quantum Algorithms for Lattice Problems*, Cryptology ePrint Archive, Paper 2024/555, 2024.

[5] A. M. Childs and W. van Dam, *Quantum Algorithms for Algebraic Problems*, Rev. Mod. Phys. **82**, 1–52, 2010.

[6] M. Ettinger and P. Høyer, *On Quantum Algorithms for Noncommutative Hidden Subgroup Problems*, Adv. Appl. Math. **25**(3), 239–251, 2000.

[7] G. Kuperberg, *A Subexponential-Time Quantum Algorithm for the Dihedral Hidden Subgroup Problem*, SIAM J. Comput. **35**(1), 170–188, 2005.

[8] G. Kuperberg, *Another Subexponential-Time Quantum Algorithm for the Dihedral Hidden Subgroup Problem*, TQC 2013, LIPIcs **22**, pp. 20–34.

[9] National Institute of Standards and Technology, *Module-Lattice-Based Key-Encapsulation Mechanism Standard*, FIPS 203, 2024.

[10] R. O'Donnell, *Analysis of Boolean Functions*, Cambridge University Press, 2014.

[11] O. Regev, *Quantum Computation and Lattice Problems*, SIAM J. Comput. **33**(3), 738–760, 2004.

[12] O. Regev, *A Subexponential Time Algorithm for the Dihedral Hidden Subgroup Problem with Polynomial Space*, arXiv:quant-ph/0406151, 2004.

[13] O. Regev, *On Lattices, Learning with Errors, Random Linear Codes, and Cryptography*, J. ACM **56**(6), Article 34, 2009.

[14] P. W. Shor, *Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer*, SIAM J. Comput. **26**(5), 1484–1509, 1997.

[15] D. R. Simon, *A Polynomial-Time Quantum Algorithm for the Dihedral Coset Problem* (Preliminary Draft), Cryptology ePrint Archive, Paper 2026/1591, 31 July 2026.