An explicit nonlocal game of quantum value one that needs exponential entanglement

Nidhal Mghirbi Seth Douglas

Version 1.1.0, November 2026

Abstract

How much entanglement does near-optimal play of a two-player nonlocal game require? We give an explicit binary linear system game, with 1687 equations in 1993 variables and at most three variables in each equation, whose quantum value is one, and in which every strategy that wins with probability $1-\varepsilon$ needs entanglement entropy $\varepsilon^{-1+o(1)}$, hence local dimension $\exp(\varepsilon^{-1+o(1)})$. The same bound holds for the entanglement of formation of mixed strategies, and a strategy on a maximally entangled state of local dimension $\exp(C\varepsilon^{-1}\log(1/\varepsilon))$ suffices, so the exponent is sharp. For explicit two-player games of quantum value one, the best previous lower bound was polynomial in $1/\varepsilon$.

The game comes from a finitely generated, elementary amenable group: a lamplighter whose lamps are Pauli pairs indexed by finite sets of points on the three rays on which Houghton's group acts. Exponentially many of these pairs share one central involution, yet the relation that any two of them commute can be derived from a fixed finite set of relations in a number of steps only quasi-polynomial in the number of points. Two tools may be of independent use. The first counts qubits from approximate Pauli relations measured in an arbitrary state, with the errors of different pairs adding in squares. The second identifies, in both directions and at the natural precision $\sqrt{\varepsilon}$, the near-optimal strategies of a linear system game with approximate representations of its solution group measured in the players' actual state, so that no power of ε is lost by passing to a flat state. The same tools give sharp rates for two earlier examples: the CHSH(n) games need $\Theta(\min(n, 1/\varepsilon))$ ebits, and the linear system game of Slofstra and Vidick needs local dimension $\Theta(\varepsilon^{-1/2})$.

The authors developed this work with AI models from more than one developer and take responsibility for every claim. How the work was done, and what the machine checks cover, is described in the section "How this work was done", before the appendices.

Contents

1 Introduction **2**

2 Preliminaries **8**

3 Calculus in a fixed state **12**

4 Counting qubits in a state **17**

5 Linear system games keep the state **22**

6 The Pauli configuration-lamp group **26**

7 Pair areas **31**

8 Lamps force entropy: the sector theorem **38**

9 Hash models: the upper bound **42**

10 The game **48**

11 Two sharp rates **57**

12 Open questions **68**

How this work was done **69**

A A quasi-polynomial Dehn bound for Houghton's group **71**

B Slofstra's embedding with $w \mapsto -I$ **75**

C Machine checks **81**

D Formal verification and its scope **89**

1 Introduction

In a two-player nonlocal game a referee sends a question to each of two players, who cannot communicate, and they win if their answers satisfy a fixed predicate. Players who share an entangled state can sometimes win with a probability that no classical strategy reaches, and for some games the best winning probability, the *quantum value* $\omega_q(G)$, is approached only by strategies of ever larger dimension. For such a game the basic quantitative question is how much entanglement near-optimal play needs. Lower bounds on it are dimension witnesses: from the winning statistics alone, they certify that the players share a large entangled state. Write $\mathcal{E}(G,\varepsilon)$ for the least local dimension of a strategy that wins with probability at least $\omega_q(G)-\varepsilon$. How fast does $\mathcal{E}(G,\varepsilon)$ grow as $\varepsilon \to 0$?

The question has a short history. For two-player games of quantum value one, the first explicit examples gave polynomial growth. Slofstra and Vidick built a linear system game with $c\,\varepsilon^{-1/6} \le \mathcal{E}(G,\varepsilon) \le C\varepsilon^{-1/2}$ and wrote that they did not expect the inverse polynomial scaling to be optimal [SV18]. Slofstra then reached $\exp(\varepsilon^{-c})$ for a game of commuting-operator value one, and noted that it is not clear whether its quantum value is one, in part because his group is not known to be hyperlinear and in part because of limitations of the embedding theorem of [Slo20] that he used [Slo18]. Games with super-polynomial growth are known with three players [JLV20], and for two players when the value is not one [Col20]. There is also a uniform obstruction. Deciding whether $\omega_q(G)=1$ is Π_2-complete [MNY22, Thm. 1.1], while for fixed D and k it is decidable, by Tarski–Seidenberg, whether some strategy of local dimension at most D wins with probability at least $1-1/k$; so no total computable function of the game and the precision bounds $\mathcal{E}(G,\varepsilon)$ for all games of quantum value one. This names no game and gives no rate for any one game. Dykema, Paulsen and Prakash showed, through correlation functions of graphs, that the set of quantum correlations is not closed already for five questions and two answers per player [DPP19], so that approximating some limits of finite-dimensional correlations needs unbounded dimension. Table 1 collects the explicit bounds.

1.1 The result

We give an explicit two-player game of quantum value one in which every near-optimal strategy needs exponential entanglement, and we determine the exponent. Throughout, $\beta(\varepsilon) = 2^{-4(24+\log\log(1/\varepsilon))^2}$; this is $\varepsilon^{o(1)}$ as $\varepsilon \to 0$. Logarithms are to base two.

Theorem A. *There is a binary linear system game G, with* 1687 *equations in* 1993 *variables and at most three variables in each equation, such that $\omega_q(G)=1$ and no finite-dimensional*

	players	value	needed	sufficient	remark
[SV18]	2	$\omega_q = 1$	$c\,\varepsilon^{-1/6}$	$C\varepsilon^{-1/2}$	linear system game
[MV15]	2	1	$c\,\varepsilon^{-1/2}$	$\varepsilon^{-O(1)}$	answers of unbounded length
[Slo18]	2	$\omega_{qc} = 1$	$\exp(\varepsilon^{-c})$		$\omega_q = 1$ not known
[Col20]	2	below 1	$2^{\Omega(\varepsilon^{-1/8})}$	$2^{O(\varepsilon^{-1})}$	within ε of the value
[JLV20]	3	1	$2^{\Omega(\varepsilon^{-c})}$	$2^{O(\varepsilon^{-1})}$	three players
Theorem A	2	$\omega_q = 1$	$\exp(\varepsilon^{-1+o(1)})$	$\exp(C\varepsilon^{-1}\log\frac{1}{\varepsilon})$	linear system game; also in ebits

Table 1: Local dimension needed, and known to suffice, to win with probability $\omega-\varepsilon$, for games whose value ω is not attained by finite-dimensional strategies. For Coladangelo's game the bounds concern coming within ε of its value. Gao, Yu and Zhi recently gave an explicit linear system game whose commuting-operator value is one and whose quantum value is below one [GYZ26], a different question.

strategy wins G with certainty, and there are explicit constants $c, C, \varepsilon_0 > 0$ with the following properties for $0 < \varepsilon \le \varepsilon_0$.

(i) Every strategy that wins G with probability at least $1-\varepsilon$ has entanglement entropy

$$S(\rho_B) \;\ge\; c\,\varepsilon^{-1}\,\beta(\varepsilon),$$

and hence Schmidt rank, and local dimension, at least $2^{c\,\varepsilon^{-1}\beta(\varepsilon)}$. For $0 < \varepsilon \le \varepsilon_0/2$ the same lower bound, with $c/4$ in place of c, holds for the entanglement of formation of every mixed strategy.

(ii) Some strategy that wins with probability at least $1-\varepsilon$ uses a maximally entangled state of local dimension at most $\exp\big(C\varepsilon^{-1}\log(1/\varepsilon)\big)$.

In particular $\log \mathcal{E}(G,\varepsilon) = \varepsilon^{-1+o(1)}$, and the same holds for the least entanglement entropy and for strategies on maximally entangled states.

The precise statement is Theorem 10.5, with $c = 4.9\cdot 10^{-34}$, $C = 7.4\cdot 10^{10}$ and $\varepsilon_0 = 2^{-5500}$; these values are far from optimal. Figure 2 describes the game. To our knowledge G is the first explicit two-player game of quantum value one that needs more than polynomial entanglement, and the first two-player game of value one whose entanglement is pinned to a scale $\exp(\varepsilon^{-\Theta(1)})$. The lower bound is on the entanglement *entropy*, so it counts ebits and not only dimensions: near-optimal play of G consumes $\varepsilon^{-1+o(1)}$ ebits, whatever the shape of the shared state.

The group behind the game. A linear system game is governed by its solution group Γ, which contains a central involution J: a near-optimal strategy is, up to small errors, a representation of Γ in which J acts as $-I$, and the entanglement of the game is the price of keeping J away from the identity in approximate representations [SV18]. So we need a group in which J survives in approximate representations, which gives value one, but only in very large dimension.

Our group G_{P} is a lamplighter whose lamps sit on configurations rather than on points (Section 6, Fig. 1). The points are those of the three rays of Houghton's group H_3; a configuration is a finite set of points, written as a binary vector; and the lighter is a group of affine maps of configurations that permute points, toggle a point, or add one point to another. Each lamp is a Pauli pair x_v, z_v, a copy of the dihedral group of order eight, and the centres $(x_v z_v)^2$ of all lamps are one central involution J. On N points there are 2^N configurations, hence 2^N Pauli pairs. Yet the generators of the lighter either move points along the rays, as Houghton's group does, or involve at most two points of a configuration, so any two lamps can be brought together cheaply: their commutator has area $2^{O(\log^2 N)}$ (Theorem 7.8), priced by a quasi-polynomial Dehn bound for H_3 that we prove in Appendix A. The group is elementary

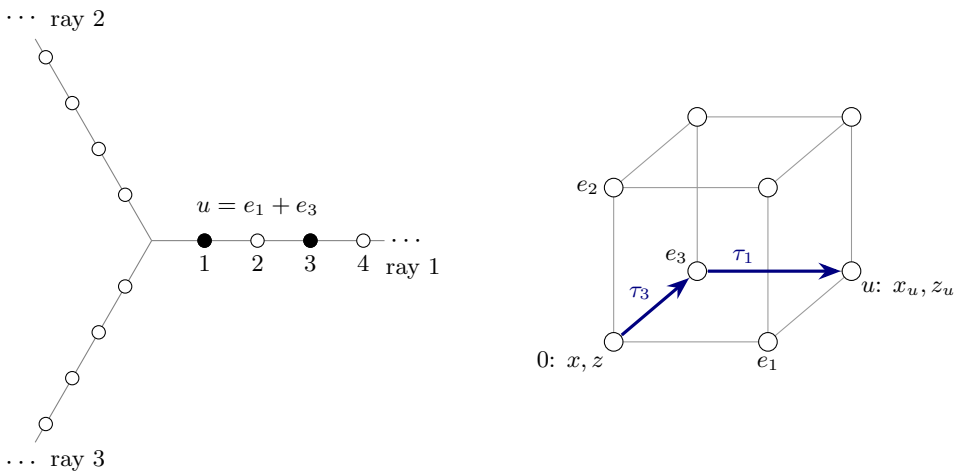


Figure 1: Left: the points of the three rays of Houghton's group and a configuration $u = e_1 + e_3$, a finite set of points (filled). Right: the 2^3 configurations supported on the points $1, 2, 3$ of ray 1, each carrying a Pauli pair (a circle), with $(x_v z_v)^2 = J$ for all v. The transporter $T_u = \tau_1 \tau_3$, a word in the generators, carries the configuration 0 to u (bold path), and conjugation by it carries the pair x, z at 0 to the pair $x_u = T_u x T_u^{-1}$, $z_u = T_u z T_u^{-1}$ at u. On N points there are 2^N pairs, and on the sector where J acts as -1 each acts as a qubit.

amenable, hence hyperlinear, which removes the first of Slofstra's obstacles. For the second we give explicitly a presentation of the type that the Slofstra–Vidick reduction needs, with one non-involutive generator, whose 29 derivations are checked by machine (Section 10).

1.2 Why the state must be kept

A strategy that wins with probability $1-\varepsilon$ controls the players' operators to precision about $\sqrt{\varepsilon}$, measured in their shared state: if ρ is Bob's reduced density matrix and $\xi = \rho^{1/2}$, his observables satisfy the relations of the solution group up to $\sqrt{\varepsilon}$ in the norm $\|Y\xi\|_2 = \mathrm{Tr}(Y^*Y\rho)^{1/2}$, and they nearly commute with ξ, again to order $\sqrt{\varepsilon}$. The usual route to a lower bound then *flattens* the state: it passes to a spectral projection of ρ [SV18], to a strategy on a maximally entangled state [Pad22], or to the operator norm. Each of these steps costs a power of ε. For linear system games the precision $\sqrt{\varepsilon}$ becomes $\varepsilon^{1/4}$ [SV18, Cor. 5.2]; for our game this would give only $\log \mathcal{E}(G, \varepsilon) \geq \varepsilon^{-1/2+o(1)}$, the exponent of maximally entangled strategies halved.

We avoid the loss by never leaving the state. The counting argument below works with the Hilbert–Schmidt vector ξ itself: it compares quantities measured in one and the same state, and it pays for every change of state with a *state commutator* $\|[Y, \xi]\|_2$, which near-optimal play supplies at precision $\sqrt{\varepsilon}$. Two results make this precise, and both may be of independent use.

The first is a robust form of the fact that M pairs of anticommuting involutions need M qubits, valid in an arbitrary state, with the errors of different pairs adding in squares rather than in norms.

Theorem C (Theorem 4.1 and Corollary 4.2)**.** *Let $X_1, \dots, X_M, Z_1, \dots, Z_M$ be involutions on $\mathbb{C}^d$, $M \geq 2$, let ρ be a density matrix and $\xi = \rho^{1/2}$. If for all $i \neq j$*

$$\|[Z_i, Z_j]\xi\|_2, \quad \|[X_i, Z_j]\xi\|_2, \quad \|\{X_i, Z_i\}\xi\|_2, \quad \|[X_i, \xi]\|_2, \quad \|[Z_i, \xi]\|_2 \;\leq\; \kappa$$

and $16M\kappa^2 \leq \frac{1}{10}$, then $S(\rho) \geq \frac{3}{4}M$.

The full statement, Theorem 4.1, bounds $S(\rho)$ from below by a sum over the pairs in which the errors of each pair enter only through sums of squares. The state commutators

cannot be dropped (Example 4.5). Stability theorems of Gowers–Hatami type [GH17] and the many-qubit tests of Natarajan and Vidick [NV17] need control of a whole Pauli group; Theorem 4.1 needs only the pairwise relations, and the ℓ^2 tolerance is what produces the exponent 1 in Theorem A. Pairwise approximate relations with a fixed error per pair allow many approximate qubits to overlap in a small space [CRSV17]; Theorem 4.1 is consistent with this, because it charges the errors of all pairs together, through a sum of squares.

The second result says that for linear system games nothing is lost in either direction when the state is kept. For a finite presentation and a word w, let $\mathrm{hlp}^{\rho}(w;e)$ be the least dimension of a homomorphism ϕ from the free group to $\mathrm{U}(d)$ together with a density matrix ρ such that every relator is e-close to the identity in the norm $\|(\cdot)\xi\|_2$, every generator has state commutator at most e, and $\|(\phi(w)+I)\xi\|_2 \le e$ (Definition 2.6).

Theorem D (Theorems 5.1 and 5.4 and Corollary 5.5)**.** *Let G be a binary linear system game of quantum value one with solution group Γ and central element J. There are constants $C, C' > 0$, depending only on the number of equations, the number of variables per equation and the question distribution, such that for every $\varepsilon > 0$*

$$\mathrm{hlp}^{\rho}_{\Gamma}\big(J; C\sqrt{\varepsilon}\big) \;\le\; \mathcal{E}(G,\varepsilon) \;\le\; \mathrm{hlp}^{\rho}_{\Gamma}\big(J; C'\sqrt{\varepsilon}\big).$$

Both directions keep the state: the lower bound uses Bob's reduced state, and the upper bound is realised on the canonical purification of the state of the model. The same holds for the least entanglement entropy.

For flat models the right-hand inequality is the classical direction of the correspondence of Slofstra and Vidick [SV18, Prop. 4.4(a)]. The left-hand one is their passage from strategies to approximate representations [SV18, Prop. 5.4], stopped before the state is flattened.

1.3 The profile of the central involution

On the side of the group, the lower bound of Theorem A is a statement about approximate representations of G_{P} in which J is visible. Let $\mathrm{hlp}(J;\delta,\varepsilon)$ be the least dimension of an ε-representation ϕ with $\|\phi(J)-I\|_f \ge \delta$, the *hyperlinear profile* of J [SV18], and let R_{P} be the finite certificate of 27 relations of G_{P} of Definition 7.1.

Theorem B (Theorem 9.6)**.** *Let $R \supseteq R_{\mathrm{P}}$ be a finite set of words in the generators p,q,t,d,x,z of G_{P} that represent 1 in G_{P}, and $J=(xz)^2$. There are constants $c, C > 0$, with C depending on the lengths of the words of R, such that for $0<\delta\le 2$ and $0<\varepsilon\le 2^{-1071}\delta$*

$$c\,\delta^2\varepsilon^{-2}\beta(\varepsilon) \;\le\; \log \mathrm{hlp}_R(J;\delta,\varepsilon) \;\le\; C\,\delta^2\varepsilon^{-2}\log(1/\varepsilon).$$

The same two-sided bound $\log \mathrm{hlp}^{\rho}_R(J;e) = e^{-2+o(1)}$ holds for the state-weighted profile.

So $\log \mathrm{hlp}(J;\delta,\varepsilon) = \varepsilon^{-2+o(1)}$. By [DM26a, Prop. 7.1] the group G_{P} is finitely presented (nothing here uses this; see Remark 6.7), so it is, to our knowledge, the first finitely presented amenable group with a central involution of superpolynomial hyperlinear profile. The upper bound comes from *hash models* (Section 9): we keep only the lamps that a short word lights, on a register of qubits indexed by a random affine hash of the configurations, and a clock handles the abelian quotient of the lighter. Slofstra's group in [Slo18] has $\log \mathrm{hlp} \ge c\,\varepsilon^{-a}$ for every $a < \frac{1}{2}$, and is not known to be hyperlinear; the sofic group of Slofstra and Vidick has a central involution of polynomial profile [SV18, Prop. 3.2].

1.4 Two sharp rates

The same two tools close the gaps in two classical examples, in which the known lower bounds lost a power of ε by flattening.

The XOR games CHSH(n) of Slofstra [Slo11] force n pairwise anticommuting observables. Ostrev showed that every nearly optimal strategy approximately contains the algebraic structure of the canonical optimal one [Ost16]; the question here is how much entanglement that forces as a function of the error. Ostrev and Vidick showed that a strategy within a factor $1-\varepsilon$ of the optimal bias needs $\Omega(\varepsilon^{-1/5})$ ebits, using $n = \Theta(\varepsilon^{-1/5})$, and that every XOR game has an ε-optimal strategy with $\lceil 1/\varepsilon \rceil$ ebits [OV18].

Theorem E (Theorem 11.3 and Corollaries 11.4 and 11.5)**.** *For $n \geq 3$ and $0 < \varepsilon \leq 1$, the least entanglement entropy of a strategy for CHSH(n) whose bias is at least $1-\varepsilon$ times the optimal bias lies between $10^{-7}\min(n, 1/\varepsilon)/6 - \frac{1}{2}$ and $\min(\lfloor n/2 \rfloor, \lceil 1/\varepsilon \rceil)$; so it is $\Theta(\min(n, 1/\varepsilon))$ as $\min(n, 1/\varepsilon) \to \infty$. The least entanglement of formation of a mixed strategy obeys the same bounds, with $10^{-7}/24$ and $\frac{1}{4}$ in place of $10^{-7}/6$ and $\frac{1}{2}$.*

This improves the exponent $\frac{1}{5}$ of [OV18] to the optimal 1. Slofstra and Vidick built a linear system game G_{SV} from a group K whose central involution c has hyperlinear profile between $\varepsilon^{-2/3}$ and ε^{-1}, and gave the bounds $c\,\varepsilon^{-1/6} \leq \mathcal{E}(G_{\mathrm{SV}}, \varepsilon) \leq C\varepsilon^{-1/2}$ [SV18].

Theorem F (Corollary 11.12 and Theorem 11.14)**.** *The hyperlinear profile of c in K is $\Theta(1/\varepsilon)$, for every fixed δ, and $\mathcal{E}(G_{\mathrm{SV}}, \varepsilon) = \Theta(\varepsilon^{-1/2})$.*

Here the relations make the spectral measure of one generator, measured in the state, nearly invariant under doubling and under a half-turn of the circle; averaging over the two inverse branches of the doubling map contracts the Wasserstein-1 distance by a factor two, so the measure is within $\frac{29\pi}{2}$ times the precision of Haar measure (Proposition 11.10), while a measure with d atoms is at distance at least $\pi/(2d)$.

The upper bound $\mathcal{E}(G_{\mathrm{SV}}, \varepsilon) \leq C\varepsilon^{-1/2}$ is that of Slofstra and Vidick [SV18, Thm. 1.1]. Its proof, through that of [SV18, Prop. 6.2], lifts approximate representations of K to the solution group by [Slo19, Prop. 4.8], which removes one non-involutive generator at a time; K has two, and after both removals the image of c is no longer balanced, so the last step of that proof does not apply as written (Section 11.2). Only the upper bound is affected: the value one of the game, the lower bounds and the nontriviality of J do not use this step, and with at most one non-involutive generator the argument is complete. For the explicit models of K we give a lift that keeps c balanced (Lemma 11.13), so [SV18, Thm. 1.1] holds as stated.

1.5 How the proof of Theorem A works

We describe the lower bound; the upper bound runs the chain backwards using flat approximate models in which the marked involution is exactly $-I$. Figure 3 in Section 10 shows both chains with their constants.

From the game to a state-weighted model. A strategy that wins with probability $1-\varepsilon$ gives, on the support of Bob's reduced state ρ, involutions for the variables of the game that satisfy the relations of the solution group Γ, and commute with ξ, up to $O(\sqrt{\varepsilon})$ in the state (Theorem 5.1). This is bookkeeping, but it is the step at which earlier arguments flattened the state.

From the game to the group. The game is built from a presentation $\mathcal{P}$ of Slofstra's type $E\Gamma_0$ through Slofstra's embedding, run so that the distinguished involution goes to $-I$ (Appendix B), and the 27 relations of the certificate R_{P} are consequences of $\mathcal{P}$ of area at most 60 (Proposition 10.3). So the model of Γ restricts to an assignment of unitaries to the generators p, q, t, d, x, z of G_{P} satisfying R_{P} to $O(\sqrt{\varepsilon})$ in the state, in which the image of J stays at distance about one from the identity.

Lamps on the sector. On the sector where (the rounded image of) J acts as -1, the transported lamps x_u, z_u, $u \subseteq \{1, \dots, N\}$, are 2^N approximate Pauli pairs measured in the state: within a pair they nearly anticommute because $(x_u z_u)^2 = J = -1$, and across pairs they nearly commute at a cost of the pair area $\Lambda_N = 2^{O(\log^2 N)}$ times the precision (Theorem 7.8). Charging conjugators through state commutators squares this cost.

Counting. Theorem C applies with $M = 2^N$ and $\kappa \approx \Lambda_N^2 \sqrt{\varepsilon}$, as long as $2^N \Lambda_N^4 \varepsilon \lesssim 1$. Choosing the largest such N gives $2^N = \varepsilon^{-1+o(1)}$ qubits of entropy on the sector (Theorem 8.1), and the sector carries a constant fraction of the state (Lemma 4.9). This is where the exponent 1 comes from: precision $\sqrt{\varepsilon}$, errors adding in squares.

The upper bound. The hash models give ε-representations of G_{P}^+ with $J \mapsto -I$ exactly in dimension $\exp(C\varepsilon^{-2} \log(1/\varepsilon))$; they give approximate representations of $\mathcal{P}$ with $w \mapsto -I$ and then of Γ with $J \mapsto -I$ (the embedding theorem, Theorem B.3 in Appendix B), and the return direction of Theorem D turns them into strategies on a maximally entangled state that lose with probability $O(\varepsilon^2)$ (Theorem 5.4).

1.6 Related work and what is new

The architecture of the group follows Slofstra's [Slo18]: a central involution J, a family of involutions whose commutators are J or 1, exponentially many of them reached by short words, and a transfer of approximate representations along the group. In [Slo18] the involutions generate, with J, an infinite Clifford group, any two of them anticommuting up to J, and they are reached through the HNN doubling $tzt^{-1} = z^2$; here they form Pauli pairs and are reached through Houghton's group acting on configurations, which keeps the group amenable. The passage from groups to games is Slofstra's embedding [Slo19, Slo20] in the explicit form of Slofstra and Vidick [SV18], and the identification of the entanglement of a linear system game with the hyperlinear profile of J is theirs. The pair-area argument and the hash models adapt our earlier construction for lamps that are copies of S_3, where they bound the sofic profile of an amenable group [DM26a]; nothing in this paper depends on that preprint.

On the analytic side, the Hilbert–Schmidt picture of a shared state, with the reduced state as a factor ξ, is the standard form of the state. Slofstra and Vidick use it for general strategies [SV18], and Marrakchi and de la Salle use it to round near-synchronous strategies [MdlS23]. The coupling of two spectral measures in one state is Connes' joint distribution of left and right multiplication [Con76]. Robust tests of many qubits in the players' state are due to Natarajan and Vidick [NV17], through a test that samples the whole Pauli group; Pauli stability for synchronous strategies is in [dlS25]; state-dependent anticommutators appear in the entropic uncertainty relations of Kaniewski, Tomamichel and Wehner [KTW14]; and rounding to maximally entangled or tracial strategies is studied in [Vid22, Pad22, Lin23, MdlS23].

What is new is the following.

(1) An explicit two-player game of quantum value one whose near-optimal strategies need $\varepsilon^{-1+o(1)}$ ebits, with a matching upper bound (Theorem A).

(2) An elementary amenable group with a central involution whose hyperlinear profile is $\exp(\varepsilon^{-2+o(1)})$, two-sided, for every finite set of valid relations that contains R_{P} (Theorem B), and a self-contained proof, from Johnson's presentation, of a quasi-polynomial Dehn bound for Houghton's group H_3 (Appendix A); our preprint [DM26b] proves a polynomial bound, which this paper does not use.

(3) A count of qubits from pairwise Pauli relations in an arbitrary state, with errors adding in squares (Theorem C).

(4) A correspondence between linear system games and state-weighted approximate representations that loses nothing in either direction (Theorem D).

(5) The optimal exponents for CHSH(n) and for the game and group of Slofstra and Vidick (Theorems E and F).

1.7 Organization and conventions

Section 2 fixes norms, approximate representations, state-weighted models and linear system games. Section 3 develops a calculus for words in unitaries measured in a fixed state, Section 4 proves Theorem C, and Section 5 proves Theorem D. Section 6 defines the group G_{P}, Section 7 proves the pair areas, Section 8 the lower bound of Theorem B and the sector theorem, and Section 9 the hash models and Theorem B. Section 10 builds the game and proves Theorem A, and Section 11 proves Theorems E and F. Section 12 lists open questions. Appendix A proves the Dehn bound for H_3, Appendix B gives Slofstra's embedding with the distinguished involution sent to $-I$, and Appendix C describes the machine checks of the finite certificates.

Logarithms are to base two unless written ln; h is the binary entropy and S the von Neumann entropy in bits. For words and group elements $[a,b] = aba^{-1}b^{-1}$; for operators $[S,T] = ST - TS$ and $\{S,T\} = ST + TS$.

2 Preliminaries

All Hilbert spaces are finite-dimensional. This section fixes the norms in which errors are measured and recalls approximate representations and their hyperlinear profile, with the two facts that make profiles computable: the defect of a word is at most linear in its area, and a change of finite presentation changes all defects by bounded factors. It then introduces state-weighted models, the objects that near-optimal strategies actually produce, and the profiles they define. Finally it recalls linear system games and fixes the measures of entanglement.

2.1 Norms and conventions

Logarithms are to base 2 unless written ln; $h(x) = -x\log x - (1-x)\log(1-x)$ is the binary entropy, S the von Neumann entropy in bits, and $D(A\|B) = \operatorname{Tr} A(\log A - \log B)$ the relative entropy of positive operators A, B with $\operatorname{supp} A \subseteq \operatorname{supp} B$. Maps compose from right to left and groups act on the left. $F(S)$ is the free group on a set S, and $|w|$ is the length of a word w. For words or group elements $[a,b] = aba^{-1}b^{-1}$; for operators $[S,T] = ST - TS$ and $\{S,T\} = ST + TS$. An *involution* is a Hermitian unitary.

For an operator T between Hilbert spaces, $\|T\|_2 = (\operatorname{Tr} T^*T)^{1/2}$ is the unnormalized Hilbert–Schmidt norm, $\|T\|_1$ the trace norm and $\|T\|_{\mathrm{op}}$ the operator norm. We use $\|XTY\|_2 \le \|X\|_{\mathrm{op}}\|T\|_2\|Y\|_{\mathrm{op}}$, the invariance of $\|\cdot\|_2$ under multiplication by unitaries on either side, and Hölder's inequality $\|ST\|_1 \le \|S\|_2\|T\|_2$. On $\mathbb{C}^d$ we write $\operatorname{tr} = \frac{1}{d}\operatorname{Tr}$ and $\|T\|_f = \|T\|_2/\sqrt{d}$ for the normalized norm, so $\|I\|_f = 1$. For unitaries U, W we have $\|U - W\|_f = \|UW^* - I\|_f$, $\|U^{-1} - I\|_f = \|U - I\|_f$, and

$$\|U - I\|_f^2 = 2 - 2\operatorname{Re}\operatorname{tr} U \le 4, \qquad \text{with equality if and only if } U = -I, \tag{1}$$

since $\operatorname{Re}\operatorname{tr} U = -1$ forces every eigenvalue of U to be -1. For unitaries U, V, $UV - VU = (UVU^{-1}V^{-1} - I)VU$, so $\|UVU^{-1}V^{-1} - I\|_f = \|UV - VU\|_f$: inside the normalized norm the group commutator and the operator commutator are interchangeable. In a fixed state they differ by state commutators (Section 2.3).

2.2 Approximate representations

An approximate representation assigns unitary matrices to the generators of a finite presentation so that every relator is mapped close to the identity. An element that is nontrivial in the

group may still be forced close to the identity, and the hyperlinear profile measures how large the matrices must be before it can stay away.

Definition 2.1 (Approximate representations)**.** Let S be finite, $R \subset F(S)$ finite and $\varepsilon \geq 0$. An *ε-representation* of $\langle S \mid R\rangle$ of dimension d is a homomorphism $\phi : F(S) \to \mathrm{U}(d)$ with $\|\phi(r) - I\|_f \leq \varepsilon$ for every $r \in R$. We call $\|\phi(w) - I\|_f$ the *defect* of w.

Definition 2.2 (Hyperlinear profile; [SV18, Definition 2.3])**.** For $w \in F(S)$, $0 < \delta \leq 2$ and $\varepsilon > 0$, the *hyperlinear profile* $\mathrm{hlp}(w; \delta, \varepsilon)$ is the least d for which some ε-representation of $\langle S \mid R\rangle$ of dimension d has $\|\phi(w) - I\|_f \geq \delta$, and ∞ if there is none. We write hlp_R when the presentation matters, and say that w is *nontrivial in approximate representations* if $\mathrm{hlp}(w; \delta, \varepsilon) < \infty$ for some $\delta > 0$ and all $\varepsilon > 0$.

The profile is nonincreasing in ε and nondecreasing in δ. By (1), $\delta = 2$ asks for $\phi(w) = -I$ exactly; this is the form used by Slofstra and Vidick. If a finitely presented group is hyperlinear, every element $g \neq 1$ is nontrivial in approximate representations, for every finite presentation and every word representing g; amenable groups are sofic, hence hyperlinear [Gro99, Wei00, ES05, Pes08].

A word w that represents 1 in $G = \langle S \mid R\rangle$ is freely equal to a product of k conjugates of elements of $R^{\pm 1}$; the least such k is its *area* $\mathrm{Area}_R(w)$, and the *Dehn function* of the presentation is $n \mapsto \max\{\mathrm{Area}_R(w) : |w| \leq n,\ w = 1 \text{ in } G\}$. For a set of words W we say that w has *area at most k in W* if it is freely equal to a product of k conjugates of elements of $W^{\pm 1}$.

Lemma 2.3 (Defect is bounded by area)**.** *If $\|\phi(r) - I\|_f \leq \varepsilon$ for every $r \in W$ and w has area at most k in W, then $\|\phi(w) - I\|_f \leq k\varepsilon$. In particular, in an ε-representation, $\|\phi(w) - \phi(w')\|_f \leq \mathrm{Area}_R(w'^{-1}w)\,\varepsilon$ whenever w, w' represent the same element.*

Proof. For unitaries $U_1, \dots, U_k$ and $W' = U_2 \cdots U_k$, unitary invariance gives $\|U_1 W' - I\|_f \leq \|U_1 W' - W'\|_f + \|W' - I\|_f = \|U_1 - I\|_f + \|W' - I\|_f$, so $\|U_1 \cdots U_k - I\|_f \leq \sum_i \|U_i - I\|_f$ by induction. Write w, up to free reduction, as $\prod_{i=1}^{k} u_i r_i^{\pm 1} u_i^{-1}$ with $r_i \in W$. Since ϕ is a homomorphism of the free group, each factor has $\|\phi(u_i)\phi(r_i)^{\pm 1}\phi(u_i)^{-1} - I\|_f = \|\phi(r_i)^{\pm 1} - I\|_f = \|\phi(r_i) - I\|_f \leq \varepsilon$, and summing gives the first claim. For the second, $w'^{-1}w$ represents 1 and $\|\phi(w) - \phi(w')\|_f = \|\phi(w'^{-1}w) - I\|_f$. □

The lemma is the only way in which approximate representations see the geometry of the group: a short filling forces a word to be nearly trivial. Its counterpart in a fixed state, where conjugators are no longer free, is Lemma 3.3(d).

Lemma 2.4 (Change of presentation)**.** *Let $\langle S \mid R\rangle$ and $\langle S' \mid R'\rangle$ present the same group, fix words $\hat{s}' \in F(S)$ representing the generators $s' \in S'$, and let $w \in F(S)$, $w' \in F(S')$ represent the same element. There is $C \geq 1$ with $\mathrm{hlp}_{R'}(w'; \delta - C\varepsilon, C\varepsilon) \leq \mathrm{hlp}_R(w; \delta, \varepsilon)$ whenever $C\varepsilon < \delta$; if $w' = s'$ is a letter and $\hat{s}' = w$, the loss in δ disappears.*

Proof. Let $\vartheta : F(S') \to F(S)$ be the homomorphism $s' \mapsto \hat{s}'$. The words $\vartheta(r')$, $r' \in R'$, and $\vartheta(w')^{-1}w$ represent 1; let $C \geq 1$ bound their areas in R. If ϕ is an ε-representation of $\langle S \mid R\rangle$ with $\|\phi(w) - I\|_f \geq \delta$, then $\phi' = \phi \circ \vartheta$ is a $C\varepsilon$-representation of $\langle S' \mid R'\rangle$ of the same dimension by Lemma 2.3, and $\|\phi'(w') - \phi(w)\|_f \leq C\varepsilon$, so $\|\phi'(w') - I\|_f \geq \delta - C\varepsilon$. In the special case $\phi'(w') = \phi(\hat{s}') = \phi(w)$. □

So lower bounds of the form $\exp(c\,(\delta/\varepsilon)^{\alpha})$, and finiteness of the profile, pass between finite presentations, with new constants. The same holds for the state-weighted profiles below (Lemma 3.5).

2.3 State-weighted models

Let ρ be a density matrix on $\mathbb{C}^d$ and $\xi = \rho^{1/2}$, so that $\|\xi\|_2 = 1$. For an operator Y on $\mathbb{C}^d$ we call

$$\|Y\xi\|_2 = \mathrm{Tr}(Y^*Y\rho)^{1/2} \qquad \text{and} \qquad \|[Y,\xi]\|_2$$

the *state-weighted norm* of Y and the *state commutator* of Y. For the *flat state* $\rho = I/d$ the first is $\|Y\|_f$ and the second vanishes. We use without comment that $\|XY\xi\|_2 \le \|X\|_{\mathrm{op}}\|Y\xi\|_2$, that $\|Y\xi U\|_2 = \|Y\xi\|_2$ for unitary U, that $\|\xi Y\|_2 = \|Y^*\xi\|_2$, that $\|(U^* - I)\xi\|_2 = \|(U - I)\xi\|_2$ for unitary U, and that $[U^*,\xi] = -U^*[U,\xi]U^*$, so $\|[U^*,\xi]\|_2 = \|[U,\xi]\|_2$. For contractions S, T the Leibniz rule $[ST,\xi] = S[T,\xi] + [S,\xi]T$ gives $\|[ST,\xi]\|_2 \le \|[S,\xi]\|_2 + \|[T,\xi]\|_2$. For unitaries U, V the identities $[U,V] = (UVU^{-1}V^{-1} - I)VU$ and $UVU^{-1}V^{-1} - I = [U,V]U^{-1}V^{-1}$ show that $\|[U,V]\xi\|_2$ and $\|(UVU^{-1}V^{-1} - I)\xi\|_2$ differ by at most $2\|[U,\xi]\|_2 + 2\|[V,\xi]\|_2$.

Definition 2.5 (State-weighted models)**.** Let $\langle S \mid R\rangle$ be a finite presentation, $w \in F(S)$ and $e \ge 0$. A *state-weighted model* of $\langle S \mid R\rangle$ at precision e on $\mathbb{C}^d$ is a pair (ϕ, ρ) of a homomorphism $\phi : F(S) \to \mathrm{U}(d)$ and a density matrix ρ on $\mathbb{C}^d$ such that, with $\xi = \rho^{1/2}$,

(S1) $\|(\phi(r) - I)\xi\|_2 \le e$ for every $r \in R$;

(S2) $\|[\phi(s),\xi]\|_2 \le e$ for every $s \in S$.

It is a model of the pair $(\langle S \mid R\rangle, w)$ if moreover

(S3) $\|(\phi(w) + I)\xi\|_2 \le e$.

We call $\|(\phi(u) - I)\xi\|_2$ the *defect of u in the state*.

Condition (S2) has no flat analogue: a trace ignores conjugation, a state does not, and (S2) is what pays for conjugators (Lemma 3.3). Condition (S3) is the state-weighted form of $\phi(w) = -I$.

Definition 2.6 (State-weighted profiles)**.** Let $\langle S \mid R\rangle$ be a finite presentation, $w \in F(S)$ and $e > 0$. The *state-weighted profile* $\mathrm{hlp}^\rho(w; e)$ is the least d for which a model of $(\langle S \mid R\rangle, w)$ at precision e on $\mathbb{C}^d$ exists, that is, a pair (ϕ, ρ) satisfying (S1)–(S3); it is ∞ if there is none. The *entropy profile* $h^S(e)$ is the infimum of $S(\rho)$ over all such models, in any dimension. We write hlp^ρ_R and h^S_R when the presentation matters.

Both profiles are nonincreasing in e, and $h^S(e) \le \log \mathrm{hlp}^\rho(w; e)$ since $S(\rho) \le \log d$. If ϕ is an e-representation with $\|\phi(w) - I\|_f \ge 2$, that is with $\phi(w) = -I$, then $(\phi, I/d)$ satisfies (S1)–(S3) at precision e. Hence

$$\mathrm{hlp}^\rho(w; e) \le \mathrm{hlp}(w; 2, e),$$

and a lower bound on hlp^ρ is a stronger statement than the same lower bound on $\mathrm{hlp}(\,\cdot\,; 2, \cdot)$.

2.4 Linear system games

Let $A \in \mathbb{F}_2^{m\times n}$ and $b \in \mathbb{F}_2^m$, and let $V_i = \{j : A_{ij} = 1\}$ be the set of variables of the i-th equation, assumed nonempty; put $s = \max_i |V_i|$. We write answers as signs, the bit 0 as $+1$ and the bit 1 as -1.

Definition 2.7 (Linear system games; [CM14, CLS17, SV18])**.**

(a) The *solution group* $\Gamma = \Gamma(A, b)$ is generated by $x_1, \dots, x_n, J$ with the relators

$$x_j^2, \quad J^2, \quad [x_j, J] \ \ (1 \le j \le n), \qquad [x_j, x_k] \ \ (j, k \in V_i \text{ for some } i),$$
$$J^{b_i} x_{j_1} x_{j_2} \cdots x_{j_r} \ \ (1 \le i \le m,\ V_i = \{j_1 < \cdots < j_r\}).$$

The element J is a central involution. Approximate representations and models of Γ are taken with respect to this presentation, and with $w = J$; we write hlp_Γ, hlp^ρ_Γ and h^S_Γ.

(b) In the game $G = G(A, b)$ with *question distribution* π on $\{1, \dots, m\} \times \{1, \dots, n\}$, a referee draws (i, j) from π. Alice receives i and answers $a \in \{\pm 1\}^{V_i}$ with $\prod_{j \in V_i} a_j = (-1)^{b_i}$; Bob receives j and answers $c \in \{\pm 1\}$. They win if $j \notin V_i$ or $a_j = c$. We put $\pi_{\min} = \min\{\pi(i, j) : j \in V_i\}$ and assume $\pi_{\min} > 0$. The *default distribution* draws i uniformly and then $j \in V_i$ uniformly, so $\pi_{\min} \geq 1/(ms)$; Slofstra and Vidick draw i and j independently and uniformly [SV18], so $\pi_{\min} = 1/(mn)$.

(c) A *strategy* consists of a unit vector $\psi \in H_A \otimes H_B$, a POVM $\{P_a^i\}_a$ on H_A for each i, indexed by the answers to i, and a POVM $\{Q_c^j\}_{c=\pm 1}$ on H_B for each j. Its *observables* are the Hermitian contractions $A_{ij} = \sum_a a_j P_a^i$ $(j \in V_i)$ and $B_j = Q_{+1}^j - Q_{-1}^j$, and the pair (i, j) with $j \in V_i$ is won with probability

$$\sum_a \big\langle \psi, (P_a^i \otimes Q_{a_j}^j)\psi \big\rangle = \tfrac{1}{2}\big(1 + \langle \psi, (A_{ij} \otimes B_j)\psi \rangle\big).$$

The winning probability $\omega(G; \mathcal{S})$ of a strategy $\mathcal{S}$ is the π-average of these probabilities, a pair with $j \notin V_i$ counting as won. The *quantum value* $\omega_q(G)$ is the supremum of $\omega(G; \mathcal{S})$ over all strategies, and $\omega_{qc}(G)$ the supremum over commuting-operator strategies, in which ψ lies in one Hilbert space and Alice's operators commute with Bob's.

The following facts are standard. A perfect commuting-operator strategy exists if and only if $J \neq 1$ in Γ [CLS17, Theorem 4]. A perfect strategy on a tensor product of Hilbert spaces, of finite or infinite dimension, exists if and only if some finite-dimensional representation of Γ sends J to $-I$ [CM14, Theorem 1], [CLS17, Theorem 5], [Slo19, Theorem 3.2]; a finite-dimensional representation in which $J \neq 1$ restricts to the (-1)-eigenspace of the image of J, where J acts as $-I$. Finally, $\omega_q(G) = 1$ if and only if J is nontrivial in approximate representations of Γ: the "if" direction is [Slo19, Proposition 3.4] (restated as [SV18, Proposition 4.3]), and the "only if" direction follows from the lower bound $\mathrm{hlp}_\Gamma(J; 2, C\varepsilon^{1/4}) \leq \mathcal{E}(G, \varepsilon)$ of [SV18, Corollary 5.2], and [SV18, Corollary 5.3] states the equivalence. Whether $\omega_q(G) = 1$ does not depend on π, because a strategy that wins with probability $1 - \varepsilon$ loses each pair (i, j) with $j \in V_i$ with probability at most $\varepsilon/\pi_{\min}$. Section 5 proves quantitative and state-weighted forms of these correspondences.

2.5 Entanglement

A unit vector $\psi \in H_A \otimes H_B$ has a Schmidt decomposition $\psi = \sum_{k=1}^{r} \sqrt{\lambda_k}\, e_k \otimes f_k$ with $\lambda_k > 0$ and orthonormal families (e_k), (f_k). Its reduced density matrices ρ_A and ρ_B have the same nonzero eigenvalues λ_k. We keep apart the *Schmidt rank* $r = \operatorname{rank} \rho_B$, which is at most both local dimensions, and the *entanglement entropy* $S(\rho_A) = S(\rho_B) \leq \log r$, the number of ebits of ψ. For a mixed state σ on $H_A \otimes H_B$ the *entanglement of formation* is [BDSW96]

$$E_F(\sigma) = \inf\Big\{ \sum_j p_j\, S\big(\operatorname{Tr}_B |\psi_j\rangle\langle\psi_j|\big) : \ \sigma = \sum_j p_j |\psi_j\rangle\langle\psi_j| \Big\}.$$

The distillable entanglement is at most E_F, so lower bounds on E_F say nothing about it.

Definition 2.8 (Dimension needed to play; [SV18])**.** For a nonlocal game G and $\varepsilon > 0$, $\mathcal{E}(G, \varepsilon)$ is the least d such that some strategy with $H_A = H_B = \mathbb{C}^d$ wins with probability at least $\omega_q(G) - \varepsilon$, and $\mathcal{E}^{\mathrm{me}}(G, \varepsilon)$ is the least d for which such a strategy exists with the maximally entangled state $\Phi_d = d^{-1/2} \sum_k e_k \otimes e_k$ of $\mathbb{C}^d \otimes \mathbb{C}^d$. So $\mathcal{E} \leq \mathcal{E}^{\mathrm{me}}$, and when $\omega_q(G) = 1$ both count strategies that win with probability at least $1 - \varepsilon$.

Remark 2.9 (Local dimension, Schmidt rank and projective measurements). Two normalizations are free. *Schmidt-rank compression:* if $\psi = \sum_{k \leq r} \sqrt{\lambda_k} e_k \otimes f_k$ and Π_A, Π_B project onto

the spans of the e_k and of the f_k, then compressing every P_a^i by Π_A and every Q_c^j by Π_B gives POVMs on two spaces of dimension r with the same probabilities, since $\psi = (\Pi_A \otimes \Pi_B)\psi$. Hence $\mathcal{E}(G,\varepsilon)$ is also the least Schmidt rank, and the least $\min(\dim H_A, \dim H_B)$, of a strategy that wins with probability at least $\omega_q(G) - \varepsilon$; a lower bound on the Schmidt rank of every such strategy is a lower bound on $\mathcal{E}(G,\varepsilon)$, and a lower bound on its entanglement entropy is a lower bound on $\log \mathcal{E}(G,\varepsilon)$. *Naimark dilation:* if K is the largest number of answers to a question, dilating each of Alice's POVMs to a projective measurement on $H_A \otimes \mathbb{C}^K$, with the state $\psi \otimes |0\rangle$, changes neither Bob's reduced density matrix nor any probability, and multiplies Alice's local dimension by at most K.

3 Calculus in a fixed state

Winning a game with probability $1-\varepsilon$ controls the players' operators to precision about $\sqrt{\varepsilon}$, but in the players' shared state, not in a normalized trace. This section develops the calculus in which such errors are measured and propagated without ever leaving the state.

The picture is as follows. A shared pure state $\psi \in H_A \otimes H_B$ is a Hilbert–Schmidt operator in disguise: in a Schmidt basis it becomes $\xi = \rho_B^{1/2}$, Bob's operators act on it from the left and Alice's, transposed, from the right (Lemma 3.1). So $\|(I \otimes Y)\psi\| = \|Y\xi\|_2$, and every error of a strategy becomes a state-weighted norm carrying the factor ξ. A normalized trace ignores conjugation, $\|gUg^{-1} - I\|_f = \|U - I\|_f$, and this is why defects are bounded by areas (Lemma 2.3). A fixed state does not: conjugating by g moves the state, and $\|(gUg^{-1} - I)\xi\|_2$ may exceed $\|(U-I)\xi\|_2$ by $2\|[g,\xi]\|_2$. So every conjugator is charged through state commutators (Lemma 3.3), and a word of area k is charged its area times the length of its conjugators (Lemma 3.4). The charge is affordable because near-optimal play supplies the state commutators at the same precision $\sqrt{\varepsilon}$ as the relations: an observable of Bob that agrees with an operator of Alice on ψ nearly commutes with ξ (Lemma 3.2). No power of ε is lost at any step.

The other tools are the rounding of unitaries to involutions and the rotation of the state by a unitary (Section 3.4); twirls, which act on Hilbert–Schmidt vectors and therefore remain orthogonal projections whatever the state (Section 3.5); the passage from a state to its restriction to the sector where a central involution acts as -1 (Section 3.6); and three facts about relative entropy (Section 3.7). They are used in Sections 4, 5, 8 and 11.

Throughout, ρ is a density matrix on $\mathbb{C}^d$ and $\xi = \rho^{1/2}$, unless the state comes from a strategy.

3.1 The standard form

Lemma 3.1 (Standard form)**.** *Let $\psi \in H_A \otimes H_B$ be a unit vector, $\rho = \rho_B$ Bob's reduced density matrix and $\xi = \rho^{1/2}$.*

(a) For every operator Y on H_B, $\|(I \otimes Y)\psi\| = \|Y\xi\|_2$.

(b) For every Hermitian operator A on H_A there is a Hermitian operator $\tilde{A}$ on H_B with $\|\tilde{A}\|_{\mathrm{op}} \le \|A\|_{\mathrm{op}}$ such that, for every operator Y on H_B,

$$\|Y\xi - \xi\tilde{A}\|_2 \le \|(A \otimes I - I \otimes Y)\psi\|.$$

The same holds with the roles of the two factors exchanged.

Proof. Let $\psi = \sum_{k=1}^r \sqrt{\lambda_k}\, e_k \otimes f_k$ be a Schmidt decomposition, so that $\xi = \sum_k \sqrt{\lambda_k}\, |f_k\rangle\langle f_k|$, and let Π_A be the projection onto $\mathrm{span}\{e_k\}$. The linear map Θ from $\Pi_A H_A \otimes H_B$ to operators on H_B with $\Theta(e_k \otimes v) = |v\rangle\langle f_k|$ is an isometry for $\|\cdot\|_2$, since $\|\sum_k e_k \otimes v_k\|^2 = \sum_k \|v_k\|^2 = \|\sum_k |v_k\rangle\langle f_k|\|_2^2$. It satisfies $\Theta(\psi) = \xi$ and $\Theta((I \otimes Y)x) = Y\Theta(x)$, which gives (a).

For (b) put $A_{lk} = \langle e_l, Ae_k\rangle$ and $\tilde{A} = \sum_{k,l} A_{lk}|f_k\rangle\langle f_l|$. Since A is Hermitian, so is $\tilde{A}$; in the basis (f_k) of its support, $\tilde{A}$ has the transposed matrix of the compression $\Pi_A A \Pi_A$, so $\|\tilde{A}\|_{\mathrm{op}} \le \|A\|_{\mathrm{op}}$. Now $(\Pi_A A \otimes I)\psi = \sum_{k,l} \sqrt{\lambda_k} A_{lk}\, e_l \otimes f_k$, which Θ maps to $\sum_{k,l} \sqrt{\lambda_k} A_{lk}|f_k\rangle\langle f_l| = \xi\tilde{A}$. The vector $((I-\Pi_A)A\otimes I)\psi$ lies in $(I-\Pi_A)H_A \otimes H_B$, which is orthogonal to $\Pi_A H_A \otimes H_B$, where both $(\Pi_A A\otimes I)\psi$ and $(I\otimes Y)\psi$ lie. Hence $\|(A\otimes I - I\otimes Y)\psi\| \ge \|(\Pi_A A\otimes I - I\otimes Y)\psi\| = \|\xi\tilde{A} - Y\xi\|_2$. Exchanging the two tensor factors gives the last sentence. □

This is the standard form of the state. Slofstra and Vidick start their analysis of general strategies from it [SV18, proof of Proposition 5.4], and Marrakchi and de la Salle use it to round near-synchronous strategies [MdlS23].

3.2 Consistency gives state commutation

If an observable of one player agrees with a Hermitian operator of the other on the shared state, the observable nearly commutes with the square root of the reduced state, with no loss in the exponent.

Lemma 3.2 (Consistency)**.** *Let B be an involution and T a Hermitian operator on $\mathbb{C}^d$, of any norm, with $\|B\xi - \xi T\|_2 \le \theta$. Then $\|[B,\xi]\|_2 \le \sqrt{2}\,\theta$.*

Proof. The form $\langle X, Y\rangle_\xi = \mathrm{Tr}(\xi X^* \xi Y)$ is the Hilbert–Schmidt inner product of $\xi^{1/2}X\xi^{1/2}$ and $\xi^{1/2}Y\xi^{1/2}$, so it is positive semidefinite and satisfies the Cauchy–Schwarz inequality. Put $x = \langle B, B\rangle_\xi = \mathrm{Tr}(\xi B \xi B) \ge 0$ and $y = \|\xi T\|_2$. Since T is Hermitian, $\|T\xi\|_2^2 = \mathrm{Tr}(T^2\rho) = \|\xi T\|_2^2$, and $\langle T, T\rangle_\xi = \langle T\xi, \xi T\rangle_{\mathrm{HS}} \le \|T\xi\|_2 \|\xi T\|_2 = y^2$. Expanding, with $\|B\xi\|_2 = 1$,

$$\theta^2 \;\ge\; \|B\xi - \xi T\|_2^2 = 1 + y^2 - 2\,\mathrm{Re}\langle B, T\rangle_\xi \;\ge\; 1 + y^2 - 2\sqrt{x}\,y \;\ge\; 1 + y^2 - (x + y^2) = 1 - x.$$

Finally $\|[B,\xi]\|_2^2 = \|B\xi\|_2^2 + \|\xi B\|_2^2 - 2\,\mathrm{Tr}(\xi B \xi B) = 2 - 2x \le 2\theta^2$. □

With Lemma 3.1(b): if A is a Hermitian operator on Alice's side and B an involution on Bob's with $\|(A\otimes I - I\otimes B)\psi\| \le \theta$, then $\|[B,\xi]\|_2 \le \sqrt{2}\,\theta$ for $\xi = \rho_B^{1/2}$. This is the Cauchy–Schwarz argument of [SV18, proof of Proposition 5.4, (5.1)], with an explicit constant and a Hermitian partner of arbitrary norm; the second point matters in Section 11, where the partner has norm $\sqrt{2}$. The constant $\sqrt{2}$ is sharp: $\xi = |0\rangle\langle 0|$, $B = \sigma_X$ and $T = 0$ give $\theta = 1$ and $\|[B,\xi]\|_2 = \sqrt{2}$.

3.3 Words in a fixed state

Let $\phi : F(S) \to \mathrm{U}(d)$ be a homomorphism, so that $\phi(w)$ depends only on the free reduction of a word w. Unlike the normalized trace, a fixed state does not ignore conjugation, so conjugators are charged through state commutators.

Lemma 3.3 (Calculus in a state)**.** *Suppose that $\|[\phi(s),\xi]\|_2 \le e_c$ for every $s \in S$.*

(a) $\|[\phi(u),\xi]\|_2 \le |u|\, e_c$ for every word u.

(b) For unitaries $U_1, \dots, U_k$, $\|(U_1 \cdots U_k - I)\xi\|_2 \le \sum_l \|(U_l - I)\xi\|_2$.

(c) For unitaries g, U, $\|(gUg^* - I)\xi\|_2 \le \|(U - I)\xi\|_2 + 2\|[g,\xi]\|_2$.

(d) If w is freely equal to $\prod_{l=1}^{k} u_l r_l^{\pm 1} u_l^{-1}$ for words u_l, r_l with $\|(\phi(r_l) - I)\xi\|_2 \le e$ for all l, then

$$\|(\phi(w) - I)\xi\|_2 \le \sum_{l=1}^{k} \big(e + 2|u_l|\, e_c\big).$$

(e) For contractions $Y_1, \dots, Y_m$ and unitaries $Y'_1, \dots, Y'_m$,

$$\|(Y_1 \cdots Y_m - Y'_1 \cdots Y'_m)\xi\|_2 \le \sum_{l=1}^{m} \Big(\|(Y_l - Y'_l)\xi\|_2 + 2 \sum_{l'>l} \|[Y'_{l'}, \xi]\|_2 \Big).$$

Parts (b), (c) and (e) do not use the hypothesis.

Proof. (a) is the Leibniz rule, with $\|[\phi(s)^{-1}, \xi]\|_2 = \|[\phi(s), \xi]\|_2$. (b) Telescoping, $U_1 \cdots U_k - I = \sum_l U_1 \cdots U_{l-1}(U_l - I)$, and the unitary left factors do not change the norm. (c) $(gUg^* - I)\xi = g(U - I)\xi g^* + g(U - I)[g^*, \xi]$, $\|U - I\|_{\mathrm{op}} \le 2$ and $\|[g^*, \xi]\|_2 = \|[g, \xi]\|_2$. (d) Since ϕ is a homomorphism of the free group, $\phi(w) = \prod_l \phi(u_l)\phi(r_l)^{\pm 1}\phi(u_l)^{-1}$; apply (b), then (c) with $\|(\phi(r_l)^{-1} - I)\xi\|_2 = \|(\phi(r_l) - I)\xi\|_2$, then (a). (e) Telescoping,

$$Y_1 \cdots Y_m - Y'_1 \cdots Y'_m = \sum_l Y_1 \cdots Y_{l-1}(Y_l - Y'_l)R_l, \qquad R_l = Y'_{l+1} \cdots Y'_m \text{ unitary.}$$

Now $(Y_l - Y'_l)R_l\xi = (Y_l - Y'_l)\xi R_l + (Y_l - Y'_l)[R_l, \xi]$. The first term has norm $\|(Y_l - Y'_l)\xi\|_2$, and the second at most $2\sum_{l'>l}\|[Y'_{l'}, \xi]\|_2$ by the Leibniz rule, since $\|Y_l - Y'_l\|_{\mathrm{op}} \le 2$. The left factors are contractions. □

Part (d) needs short conjugators. For relations of bounded area the next lemma supplies them; it is the usual relation between area and conjugator length, read off a van Kampen diagram.

Lemma 3.4 (Conjugator lengths)**.** *Let every word of a finite set R have length at most K. If w has area at most k in R, then w is freely equal to a product $\prod_{l=1}^{k'} u_l r_l^{\pm 1} u_l^{-1}$ with $k' \le k$, $r_l \in R$ and $|u_l| \le |w| + kK$ for all l.*

Proof. Let w_0 be the free reduction of w. By van Kampen's lemma [LS77, Chapter V, Theorem 1.1] (see also [Bri02, Section 4]), w_0 has a van Kampen diagram with at most k faces: a finite, connected and simply connected planar 2-complex whose edges are oriented and labelled by letters, whose boundary walk read from a base vertex is w_0, and whose faces are the bounded components of the complement of its 1-skeleton, the boundary walk of each face reading a cyclic permutation of an element of $R^{\pm 1}$. Such a diagram may be singular: it may contain edges and vertices not on any face. Van Kampen's lemma is stated in [LS77] for cyclically reduced relators. If the free reduction of $r \in R$ is $vr'v^{-1}$ with r' cyclically reduced, we use r' for the faces. A cyclic permutation of $r'^{\pm 1}$ is then freely $\tau r^{\pm 1}\tau^{-1}$ with $|\tau| \le |r'| - 1 + |v| \le K - 1$, which is all that the argument below uses. We show by induction on the number f of faces that if a van Kampen diagram D with f faces has boundary word u, then u is freely equal to a product of f conjugates of elements of $R^{\pm 1}$ with conjugators of length at most $|u| + fK$. Applied to w_0, with $|w_0| \le |w|$, this gives the lemma.

If $f = 0$, D is a finite tree, its boundary walk traverses each edge once in each direction, and u is freely trivial. Let $f \ge 1$. Choose a ray from an interior point of a face to infinity that avoids the vertices. Its last point in the closure of the union of the faces lies in the interior of an edge e with a face Φ on one side and the unbounded region on the other. The boundary walk of D traverses e exactly once, so $u = \alpha e \gamma$, where we write e also for its label read in the direction of traversal. Read from the initial vertex of e, starting along e, the boundary of Φ reads $e\delta^{-1}$, a cyclic permutation of some $r^{\pm 1}$ with $r \in R$; so $|\delta| \le K - 1$ and $e\delta^{-1} = \tau r^{\pm 1}\tau^{-1}$ freely, with $|\tau| \le K - 1$. Removing the open face Φ and the open edge e is an elementary collapse, so it leaves a van Kampen diagram D' with $f - 1$ faces, whose boundary word is $u' = \alpha\delta\gamma$, of length at most $|u| + K - 2$. Freely,

$$u = \alpha e \gamma = (\alpha\tau)\, r^{\pm 1} (\alpha\tau)^{-1} \cdot \alpha\delta\gamma,$$

and $|\alpha\tau| \le |u| - 1 + K - 1 \le |u| + K$. By induction u' is a product of $f - 1$ conjugates with conjugators of length at most $|u'| + (f - 1)K \le |u| + fK$. □

Together, Lemma 3.3(d) and Lemma 3.4 say that a word of area k in relators of defect at most e in the state, and of length ℓ, has defect at most $k\big(e + 2(\ell + kK)e_c\big)$: the area enters squared once state commutators are charged. As in the flat case, the state-weighted profiles therefore pass between finite presentations.

Lemma 3.5 (Change of presentation in a state)**.** *In the setting of Lemma 2.4, there is $C \geq 1$, depending only on the two presentations and the words $\hat{s}'$, w, w', such that for every model (ϕ, ρ) of $(\langle S \mid R\rangle, w)$ at precision e, the pair $(\phi \circ \vartheta, \rho)$ is a model of $(\langle S' \mid R'\rangle, w')$ at precision Ce, where $\vartheta(s') = \hat{s}'$. Hence $\mathrm{hlp}^{\rho}_{R'}(w'; Ce) \leq \mathrm{hlp}^{\rho}_{R}(w; e)$ and $h^{S}_{R'}(Ce) \leq h^{S}_{R}(e)$. Without (S3), the same holds for models of the presentations alone.*

Proof. Fix, once and for all, expressions of the words $\vartheta(r')$, $r' \in R'$, and $u = \vartheta(w')w^{-1}$, which represent 1, as products of conjugates of elements of $R^{\pm 1}$, up to free reduction. Let k bound the numbers of factors and L the lengths of the conjugators, and put $C = \max\big(\max_{s'}|\hat{s}'|,\, k(1 + 2L) + 2|w| + 1\big)$. By Lemma 3.3(a), $\|[\phi(\hat{s}'), \xi]\|_2 \leq |\hat{s}'|e$, which is (S2). By Lemma 3.3(d) with $e_c = e$, $\|(\phi(\vartheta(r')) - I)\xi\|_2 \leq k(1 + 2L)e$, which is (S1). For (S3) write $\phi(\vartheta(w')) + I = (\phi(u) - I)\phi(w) + (\phi(w) + I)$; since $\|(\phi(u) - I)\phi(w)\xi\|_2 \leq \|(\phi(u) - I)\xi\|_2 + 2\|[\phi(w), \xi]\|_2 \leq k(1 + 2L)e + 2|w|e$, we get $\|(\phi(\vartheta(w')) + I)\xi\|_2 \leq Ce$. The dimension and the state are unchanged. □

3.4 Rounding and rotating

Lemma 3.6 (Rounding)**.** *For a unitary U let $\mathrm{r}(U)$ be the involution obtained by replacing each eigenvalue λ of U by 1 if $\operatorname{Re}\lambda \geq 0$ and by -1 otherwise, keeping the eigenvectors. Then*

$$\|(\mathrm{r}(U) - U)\xi\|_2 \leq \|(U^2 - I)\xi\|_2 \qquad \text{and} \qquad \|[\mathrm{r}(U), \xi]\|_2 \leq \|[U, \xi]\|_2 + 2\|(U^2 - I)\xi\|_2.$$

Proof. For $\lambda = e^{i\vartheta}$ with $|\vartheta| \leq \pi/2$ we have $|\cos(\vartheta/2)| \geq 1/\sqrt{2}$, so $|\lambda - 1| = 2|\sin(\vartheta/2)| \leq 4|\sin(\vartheta/2)\cos(\vartheta/2)| = |\lambda^2 - 1|$; for $\operatorname{Re}\lambda < 0$ apply this to $-\lambda$. So $|\mathrm{r}(\lambda) - \lambda| \leq |\lambda^2 - 1|$ on the unit circle, and $\mathrm{r}(U) - U = g(U)(U^2 - I)$ for a function g with $|g| \leq 1$ on the spectrum of U (put $g(\pm 1) = 0$). As U is normal, $\|g(U)\|_{\mathrm{op}} \leq 1$ and $\|(\mathrm{r}(U) - U)\xi\|_2 \leq \|(U^2 - I)\xi\|_2$. Next $[\mathrm{r}(U), \xi] = [U, \xi] + g(U)(U^2 - I)\xi - \xi g(U)(U^2 - I)$, and, as functions of U commute, $\|\xi g(U)(U^2 - I)\|_2 = \|g(U)^*(U^{*2} - I)\xi\|_2 \leq \|(U^{*2} - I)\xi\|_2 = \|(U^2 - I)\xi\|_2$. □

For the flat state the first bound is the unitary case of [SV18, Lemma 2.8(c)], with the constant 2 there improved to 1.

Lemma 3.7 (Rotating the state)**.** *For a unitary U, $\|U^*\rho U - \rho\|_1 \leq 2\|[U, \xi]\|_2$.*

Proof. For Hermitian H, K, $H^2 - K^2 = \frac{1}{2}\big((H - K)(H + K) + (H + K)(H - K)\big)$, so $\|H^2 - K^2\|_1 \leq \|H - K\|_2\|H + K\|_2$ by Hölder. Take $H = U^*\xi U$ and $K = \xi$: then $H^2 = U^*\rho U$, $\|H - K\|_2 = \|\xi U - U\xi\|_2$ and $\|H + K\|_2 \leq 2$. □

3.5 Hilbert–Schmidt twirls

Twirling averages conjugates by a family of involutions. It is useful because each twirl is an orthogonal projection, and products of orthogonal projections are controlled by the following inequality. A related bound for sequential projective measurements is Gao's quantum union bound [Gao15].

Lemma 3.8 (Projection products)**.** *Let $P_1, \dots, P_k$ be orthogonal projections on a Hilbert space. Every vector y satisfies*

$$\|y\|^2 - \operatorname{Re}\langle y, P_k \cdots P_1 y\rangle \leq 2\sum_{l=1}^{k}\|(I - P_l)y\|^2.$$

Proof. Put $y_0 = y$ and $y_l = P_l y_{l-1}$, and let X be the left side. Since $y - y_k = \sum_l (I - P_l) y_{l-1}$ and $I - P_l$ is a self-adjoint idempotent, $X = \sum_l \operatorname{Re}\langle (I-P_l)y, (I-P_l)y_{l-1}\rangle$, and $\|(I-P_l)y_{l-1}\|^2 = \|y_{l-1}\|^2 - \|y_l\|^2$. By Cauchy–Schwarz, $X \le \big(\sum_l \|(I-P_l)y\|^2\big)^{1/2} \big(\|y\|^2 - \|y_k\|^2\big)^{1/2}$, and $\|y\|^2 - \|y_k\|^2 = 2X - \|y - y_k\|^2 \le 2X$. If $X > 0$, squaring gives the claim; otherwise there is nothing to prove. □

The inequality is [DM26b, Lemma 4.1]; we have included its short proof. For an involution Z on $\mathbb{C}^d$, the map $\alpha_Z(v) = ZvZ$ on $M_d(\mathbb{C})$ with the Hilbert–Schmidt inner product is unitary and equal to its inverse, hence self-adjoint. So

$$\mathcal{A}_Z^{\pm} = \tfrac{1}{2}(I \pm \alpha_Z)$$

are orthogonal projections on $M_d(\mathbb{C})$, whatever state is in use; the state enters only through the vectors that are twirled.

Lemma 3.9 (Two-sided twirls)**.** *Let $Z_1, \dots, Z_k$ be involutions on $\mathbb{C}^d$, $v \in M_d(\mathbb{C})$, and fix signs $\sigma_j \in \{\pm 1\}$. For $g \in \mathbb{Z}_2^k$ put $s(g) = \prod_j \sigma_j^{g_j}$ and $v_g = Z_k^{g_k} \cdots Z_1^{g_1}\, v\, Z_1^{g_1} \cdots Z_k^{g_k}$. Then, for uniformly random g,*

$$\mathbb{E}_g \|s(g) v_g - v\|_2^2 \;\le \sum_{j:\,\sigma_j = +1} \|[Z_j, v]\|_2^2 + \sum_{j:\,\sigma_j = -1} \|\{Z_j, v\}\|_2^2.$$

Proof. Let $\mathcal{B}_j = \mathcal{A}_{Z_j}^{\sigma_j}$. Averaging over g_1 first, then over g_2, and so on, $\mathbb{E}_g[s(g)v_g] = \mathcal{B}_k \cdots \mathcal{B}_1 v$. Since $\|v_g\|_2 = \|v\|_2$, the left side equals $2\|v\|_2^2 - 2\operatorname{Re}\langle v, \mathcal{B}_k \cdots \mathcal{B}_1 v\rangle$, which is at most $4\sum_j \|(I - \mathcal{B}_j)v\|_2^2$ by Lemma 3.8. Finally $(I - \mathcal{A}_Z^{+})v = \frac{1}{2}(vZ - Zv)Z$ and $(I - \mathcal{A}_Z^{-})v = \frac{1}{2}(vZ + Zv)Z$, and right multiplication by Z preserves $\|\cdot\|_2$. □

For the flat state one may twirl the operators themselves. With a non-flat state, the map $T \mapsto \frac{1}{2}(T + ZTZ)$ is not an orthogonal projection for the weighted inner product $\operatorname{Tr}(\rho S^* T)$; twirling the Hilbert–Schmidt vectors $v = Z_i\xi$, $X_i\xi$ and ξ, as in Section 4, avoids this.

3.6 Changing the state to a sector

The counting arguments compress the state, not the operators, to the eigenspace where a central involution acts as -1.

Lemma 3.10 (Sector states)**.** *Let J' be an involution on $\mathbb{C}^d$, $P_{\pm} = \frac{1}{2}(I \pm J')$, $p_- = \operatorname{Tr}(P_-\rho) > 0$, $\rho_- = P_-\rho P_-/p_-$, $\xi_- = \rho_-^{1/2}$ and $\eta = P_-\xi P_-/\sqrt{p_-}$.*

(a) $\|\xi_- - \eta\|_2 \le \|P_+\xi P_-\|_2/\sqrt{p_-}$, *and* $\|P_+\xi P_-\|_2 \le \frac{1}{2}\|[J', \xi]\|_2$.

(b) For every operator R,

$$\|R\xi_-\|_2 \le \big(\|RP_-\xi\|_2 + \|R\|_{\mathrm{op}}\|P_+\xi P_-\|_2\big)/\sqrt{p_-},$$
$$\|R\xi_-\|_2 \le \big(\|R\xi\|_2 + \|R\|_{\mathrm{op}}\|[J', \xi]\|_2\big)/\sqrt{p_-}.$$

(c) For every Hermitian contraction Y and every unitary Y',

$$\|[Y, \xi_-]\|_2 \le \big(2\|[Y, P_-]\xi\|_2 + \|[Y, \xi]\|_2 + 2\|P_+\xi P_-\|_2\big)/\sqrt{p_-},$$
$$\|[Y', \xi_-]\|_2 \le \big(\|[Y', J']\xi\|_2 + 3\|[Y', \xi]\|_2 + \|[J', \xi]\|_2\big)/\sqrt{p_-}.$$

Proof. (a) Let $A = P_-\rho P_-$ and $B = (P_-\xi P_-)^2$. Since $\rho = \xi(P_+ + P_-)\xi$, $A - B = (P_+\xi P_-)^*(P_+\xi P_-) \geq 0$, of trace $\|P_+\xi P_-\|_2^2$. By the Powers–Størmer inequality $\|A^{1/2} - B^{1/2}\|_2^2 \leq \|A - B\|_1$ [PS70, Lemma 4.1], and $A^{1/2} = \sqrt{p_-}\,\xi_-$, $B^{1/2} = P_-\xi P_-$, so $p_-\|\xi_- - \eta\|_2^2 \leq \|P_+\xi P_-\|_2^2$. Also $P_+\xi P_- = P_+[\xi, P_-]$ and $[\xi, P_-] = -\frac{1}{2}[\xi, J']$.

(b) $\|R\eta\|_2 \leq \|RP_-\xi\|_2/\sqrt{p_-}$ and $\|R(\xi_- - \eta)\|_2 \leq \|R\|_{\mathrm{op}}\|\xi_- - \eta\|_2$, which with (a) gives the first bound. For the second, $RP_-\xi = R\xi P_- + R[P_-, \xi]$ and $\|[P_-, \xi]\|_2 = \frac{1}{2}\|[J', \xi]\|_2$; with (a) again, $\|P_+\xi P_-\|_2 \leq \frac{1}{2}\|[J', \xi]\|_2$.

(c) Let Y be a Hermitian contraction or a unitary. Expanding, $[Y, P_-\xi P_-] = [Y, P_-]\xi P_- + P_-[Y, \xi]P_- + P_-\xi[Y, P_-]$. The first two terms have norms at most $\|[Y, P_-]\xi\|_2$ and $\|[Y, \xi]\|_2$, and the third at most $\|\xi[Y, P_-]\|_2 = \|[Y, P_-]^*\xi\|_2$. If Y is Hermitian, $[Y, P_-]^* = -[Y, P_-]$. If Y is unitary, $[Y, P_-]^* = [P_-, Y^*] = Y^*[Y, P_-]Y^*$, and $\|[Y, P_-]Y^*\xi\|_2 \leq \|[Y, P_-]\xi\|_2 + 2\|[Y, \xi]\|_2$. In both cases $\|[Y, \xi_- - \eta]\|_2 \leq 2\|Y\|_{\mathrm{op}}\|\xi_- - \eta\|_2$, and (a) applies; for unitary Y we also use $2\|[Y, P_-]\xi\|_2 = \|[Y, J']\xi\|_2$ and $2\|P_+\xi P_-\|_2 \leq \|[J', \xi]\|_2$. □

For a Hermitian Y of arbitrary norm the proof of (c) gives the same bound with $2\|Y\|_{\mathrm{op}}\|P_+\xi P_-\|_2$ as its last term.

3.7 Entropy

We use three standard facts. If $\mathcal{P}(\sigma) = \sum_k P_k\sigma P_k$ is the pinching by orthogonal projections P_k with $\sum_k P_k = I$, then $\log\mathcal{P}(\sigma)$ is block diagonal, so $\operatorname{Tr}\sigma\log\mathcal{P}(\sigma) = \operatorname{Tr}\mathcal{P}(\sigma)\log\mathcal{P}(\sigma)$ and

$$S(\mathcal{P}\sigma) - S(\sigma) = D(\sigma \,\|\, \mathcal{P}\sigma) \;\geq 0. \tag{2}$$

(The support condition holds: if $\mathcal{P}(\sigma)v = 0$, then $\sigma^{1/2}P_kv = 0$ for every k, so $\sigma v = 0$.) The relative entropy does not increase under quantum channels [Lin75, Uhl77]. And $D(A\|B)$ is jointly convex in pairs of positive operators [Lie73, Lin74]; since it is also positively homogeneous, $D(A_1 + A_2\|B_1 + B_2) \leq D(A_1\|B_1) + D(A_2\|B_2)$.

4 Counting qubits in a state

Let $X_1, \ldots, X_M, Z_1, \ldots, Z_M$ be involutions. If the Z_i commute, and X_i anticommutes with Z_i and commutes with Z_j for $j \neq i$, then the joint eigenspaces of the Z_i carry labels $z \in \{\pm1\}^M$, and X_i maps the eigenspace with label z onto the one with label $z^{(i)}$, the label with its i-th sign flipped. All 2^M eigenspaces then have the same dimension, so the dimension of the space is a multiple of 2^M. This section proves a robust form of this count in an arbitrary state ρ, with every error measured in that state (Theorem 4.1), and its two special cases: uniformly small errors (Corollary 4.2) and the flat state (Corollary 4.3). The errors of the pairs enter only through sums of squares, so M pairs tolerate errors of order $M^{-1/2}$ each; this ℓ^2 tolerance is what gives the exponent 2 of $1/e$ in the sector theorem (Theorem 8.1). The price of a non-flat state is the state commutators $\|[Z_i, \xi]\|_2$ and $\|[X_i, \xi]\|_2$, which cannot be dropped (Example 4.5); in games they are supplied by consistency (Lemma 3.2). The section ends with the entropy of a sector state (Lemma 4.9), which carries the count back from the sector where a central involution is -1 to the original state.

4.1 The count

Theorem 4.1 (Qubit count in a state)**.** *Let $X_1, \ldots, X_M, Z_1, \ldots, Z_M$ be involutions on $\mathbb{C}^d$, ρ a density matrix on $\mathbb{C}^d$ and $\xi = \rho^{1/2}$. Put $c_i^Z = \|[Z_i, \xi]\|_2$ and $c_i^X = \|[X_i, \xi]\|_2$, let $f(g) = Z_1^{g_1}Z_2^{g_2}\cdots Z_M^{g_M}$ for $g \in \mathbb{Z}_2^M$, let $e_i \in \mathbb{Z}_2^M$ be the i-th unit vector, and, for uniformly random g,*

$$a_i^2 = \mathbb{E}_g\|\big(f(g + e_i) - f(g)Z_i\big)\xi\|_2^2, \qquad b_i^2 = \mathbb{E}_g\|\big((-1)^{g_i}X_if(g) - f(g)X_i\big)\xi\|_2^2.$$

Then

(i) $a_i^2 \le 2\sum_{j>i}\big(\|[Z_i,Z_j]\xi\|_2 + c_j^Z\big)^2 + 2\sum_{j>i}(c_j^Z)^2$;

(ii) $b_i^2 \le 2\big(\|\{X_i,Z_i\}\xi\|_2 + c_i^Z\big)^2 + 2\sum_{j\neq i}\big(\|[X_i,Z_j]\xi\|_2 + c_j^Z\big)^2 + 2\sum_{j=1}^{M}(c_j^Z)^2$;

(iii) $S(\rho) \;\ge\; \sum_{i=1}^{M}\Big[1 - \frac{(b_i+c_i^X)^2}{2\ln 2} - h\Big(\min\Big(\frac{(a_i+c_i^Z)^2}{4}, \frac{1}{2}\Big)\Big)\Big]$.

For exact Pauli operators on $(\mathbb{C}^2)^{\otimes M}$ and $\rho = I/2^M$ every error vanishes, and (iii) is the equality $S(\rho) = M$. Since $S(\rho) \le \log \operatorname{rank}\rho \le \log d$, the theorem also bounds the rank and the dimension. When all errors are uniformly small it takes the following form.

Corollary 4.2 (Uniform errors)**.** *In the setting of Theorem 4.1, let $M \ge 2$ and suppose that, for all $i \neq j$,*

$$\|[Z_i,Z_j]\xi\|_2, \quad \|[X_i,Z_j]\xi\|_2, \quad \|\{X_i,Z_i\}\xi\|_2, \quad c_i^X, \quad c_i^Z \;\le\; \kappa.$$

If $16M\kappa^2 \le \frac{1}{10}$, then $S(\rho) \ge \frac{3}{4}M$; in particular $\operatorname{rank}\rho \ge 2^{3M/4}$.

For the flat state $\rho = I/d$ all state commutators vanish, $\|Y\xi\|_2 = \|Y\|_f$ and $S(\rho) = \log d$. The theorem then says that M approximate Pauli pairs whose errors are small in ℓ^2 need about M qubits.

Corollary 4.3 (Flat count)**.** *Let $X_1,\dots,X_M,Z_1,\dots,Z_M$ be involutions on $\mathbb{C}^D$, let $0 \le \eta \le 1$, and suppose that for every i*

$$\sum_{j>i}\|Z_iZ_j - Z_jZ_i\|_f^2 \;\le\; \eta \qquad \text{and} \qquad \|X_iZ_i + Z_iX_i\|_f^2 + \sum_{j\neq i}\|X_iZ_j - Z_jX_i\|_f^2 \;\le\; \eta.$$

Then $\log D \ge M\big(1 - \eta/\ln 2 - h(\eta/2)\big)$. In particular $D \ge 2^{M/2}$ when $\eta \le \frac{1}{10}$.

Remark 4.4 (Comparison and credits). The stability theorem of Gowers and Hatami [GH17] needs a small defect for an approximate homomorphism of the whole finite group that the pairs generate, and the many-qubit tests of Natarajan and Vidick [NV17] certify n EPR pairs from a test that samples the whole Pauli group, through a state-dependent form of that theorem. Theorem 4.1 uses only the pairwise relations of $2M$ involutions, measured in an arbitrary state, and the twirl of Lemma 3.9 manufactures the group average from them. Pauli stability for synchronous strategies with a trace is due to de la Salle [dlS25], and state-dependent anticommutators appear in the entropic uncertainty relations of Kaniewski, Tomamichel and Wehner [KTW14]. For lamps that are copies of S_3, the role of X_i is played by an involution that inverts an element of order 3; it exchanges the eigenvalues ω, ω^2 but fixes the eigenvalue 1. The analogue of the flat count for such lamps is [DM26a, Theorem B.3].

Example 4.5 (The state commutators are needed)**.** Exact Pauli operators X_i, Z_i on $(\mathbb{C}^2)^{\otimes M}$ with the pure state $\rho = \xi = |0\cdots0\rangle\langle 0\cdots0|$ satisfy the Pauli relations exactly, so $a_i = b_i = 0$ and $c_i^Z = 0$, but $S(\rho) = 0$: the bound (iii) survives only because $c_i^X = \sqrt{2}$. The commutators are needed already in (i). On $\mathbb{C}^4$ with basis $e_1,\dots,e_4$ let $\rho = \xi = |e_1\rangle\langle e_1|$, $Z_1 = \operatorname{diag}(1,-1,1,-1)$, $Z_2 = \operatorname{diag}(1,-1)\oplus\sigma_X$, and let Z_3 exchange $e_1 \leftrightarrow e_3$ and $e_2 \leftrightarrow e_4$. These are involutions with $[Z_1,Z_3] = 0$ and $[Z_1,Z_2]\xi = 0$, so every state-weighted relation in (i) for $i = 1$ vanishes. In the proof below $R = Z_2^{g_2}Z_3^{g_3}$; for $g_2 = g_3 = 1$, $R^*Z_1R - Z_1 = Z_3Z_2Z_1Z_2Z_3 - Z_1 = \operatorname{diag}(-2,2,0,0)$, which maps e_1 to $-2e_1$, and for the other values of (g_2,g_3), $(R^*Z_1R - Z_1)\xi = 0$. So $a_1^2 = 1$, consistent with (i) only because $c_3^Z = \|[Z_3,\xi]\|_2 = \sqrt{2}$.

4.2 Three entropy lemmas

Lemma 4.6 (Zero-diagonal basis)**.** *A traceless Hermitian operator H on $\mathbb{C}^d$ has an orthonormal basis $e_1, \dots, e_d$ with $\langle e_k, H e_k\rangle = 0$ for all k.*

Proof. If $H = 0$ any basis works. Otherwise H has unit eigenvectors $u_\pm$ with eigenvalues $l_+ > 0 > l_-$, and the unit vector $v = (\sqrt{-l_-}\,u_+ + \sqrt{l_+}\,u_-)/\sqrt{l_+ - l_-}$ has $\langle v, Hv\rangle = (-l_-l_+ + l_+l_-)/(l_+ - l_-) = 0$. The compression of H to $v^\perp$ is traceless; induct on d. □

The next inequality says that a two-point distribution whose masses are nearly equal in the Hellinger sense carries nearly one bit.

Lemma 4.7 (Binary Hellinger bound)**.** *For $0 \le x \le 1$, $1 - h(x) \le \big(1 - 2\sqrt{x(1-x)}\big)/\ln 2$.*

Proof. Write $x = (1+y)/2$ with $|y| \le 1$. Then $\ln 2\,(1 - h(x)) = \frac{1}{2}\big[(1+y)\ln(1+y) + (1-y)\ln(1-y)\big]$; the derivative of the bracket is $\ln(1+y) - \ln(1-y) = 2\sum_{m\ge 0} y^{2m+1}/(2m+1)$, and integrating from 0 gives $\ln 2\,(1 - h(x)) = \sum_{k\ge 1} y^{2k}/(2k(2k-1))$. The binomial series gives $1 - 2\sqrt{x(1-x)} = 1 - \sqrt{1-y^2} = \sum_{k\ge 1} \binom{2k}{k} y^{2k}/((2k-1)4^k)$. Both series have nonnegative terms and converge for $|y| \le 1$, so it suffices to show $4^k \le 2k\binom{2k}{k}$. This is an equality for $k = 1$, and from k to $k+1$ the left side grows by the factor 4 and the right side by $\frac{k+1}{k} \cdot \frac{(2k+2)(2k+1)}{(k+1)^2} = \frac{2(2k+1)}{k} \ge 4$. □

The third lemma bounds the entropy gained by pinching a state by one involution. It is a Jensen–Shannon bound in terms of a Hilbert–Schmidt distance between square-root factors.

Lemma 4.8 (Jensen–Shannon bound)**.** *Let $L, L' : \mathbb{C}^d \to \mathcal{K}$ be operators with $\|L\|_2 = \|L'\|_2 = 1$, and put $\rho = LL^*$ and $\sigma = L'L'^*$. Then*

$$D\big(\rho \,\big\|\, \tfrac{1}{2}(\rho+\sigma)\big) \;\le\; h\big(\min(\|L - L'\|_2^2/4,\, 1/2)\big).$$

Proof. The operator $L^*L - L'^*L'$ is Hermitian and traceless, so by Lemma 4.6 there is an orthonormal basis (e_k) of $\mathbb{C}^d$ with $\|Le_k\| = \|L'e_k\| =: \sqrt{w_k}$; then $\sum_k w_k = \|L\|_2^2 = 1$. Put $a_k = Le_k$ and $b_k = L'e_k$, so that $\rho = \sum_k a_k a_k^*$ and $\frac{1}{2}(\rho+\sigma) = \sum_k \frac{1}{2}(a_ka_k^* + b_kb_k^*)$. By joint convexity and homogeneity (Section 3.7),

$$D\big(\rho \,\big\|\, \tfrac{1}{2}(\rho+\sigma)\big) \le \sum_k D\big(a_ka_k^* \,\big\|\, \tfrac{1}{2}(a_ka_k^* + b_kb_k^*)\big).$$

Terms with $w_k = 0$ vanish. If $w_k > 0$, write $a_k = \sqrt{w_k}\,u$ and $b_k = \sqrt{w_k}\,u'$ with unit vectors u, u', and $r_k = |\langle u, u'\rangle|$. The operator $\frac{1}{2}(uu^* + u'u'^*)$ has eigenvalues $\frac{1}{2}(1 \pm r_k)$, and u has weights $\frac{1}{2}(1 \pm r_k)$ on the corresponding eigenvectors (compare $\langle u, \frac{1}{2}(uu^* + u'u'^*)u\rangle = \frac{1}{2}(1 + r_k^2)$), so the k-th term equals $w_k\, h(\frac{1}{2}(1 - r_k))$. Since $\|a_k - b_k\|^2 = w_k(2 - 2\,\mathrm{Re}\langle u, u'\rangle) \ge 2w_k(1 - r_k)$, we get $\frac{1}{2}(1 - r_k) \le \min(\|a_k - b_k\|^2/(4w_k), 1/2)$, and h increases on $[0, \frac{1}{2}]$. The function $y \mapsto h(\min(y/4, 1/2))$ is concave and nondecreasing on $[0, \infty)$, so by Jensen's inequality with the weights w_k,

$$\sum_k w_k\, h\Big(\min\Big(\frac{\|a_k - b_k\|^2}{4w_k}, \frac{1}{2}\Big)\Big) \le h\Big(\min\Big(\sum_k \frac{\|a_k - b_k\|^2}{4}, \frac{1}{2}\Big)\Big) = h\big(\min(\|L - L'\|_2^2/4, 1/2)\big).$$

□

Equality holds when $L = |x\rangle\langle y|$ and $L' = |x'\rangle\langle y|$ with $\langle x, x'\rangle \ge 0$. The equal-norm basis is essential: in a basis where $\|Le_k\| \ne \|L'e_k\|$ the termwise bound fails. For a single column with $b = (1 - \sqrt{x})\,a$ and $\|a\| = 1$, the term is $\log\big(2/(1 + (1-\sqrt{x})^2)\big) = \sqrt{x}/\ln 2 + O(x)$, while $h(\|a - b\|^2/4) = h(x/4) = O(x \log(1/x))$.

4.3 Proof of the count

The idea is to replace the Z_i, which commute only approximately, by exactly commuting involutions C_i on a larger space, and the X_i by exact twists $\hat{X}_i$ that anticommute with C_i and commute with C_j for $j \neq i$. The isometry $W = 2^{-M/2}\sum_g f(g) \otimes |g\rangle$ intertwines Z_i with C_i and X_i with $\hat{X}_i$ up to the averaged errors a_i and b_i, and the twirls of Lemma 3.9 bound these averages by the pairwise relations. In the embedded state, pinching by the joint eigenspaces of the C_i gives a distribution on $\{\pm 1\}^M$ in which $\hat{X}_i$ nearly exchanges the two values of z_i, so each coordinate carries nearly a full bit; and each pinching raises the entropy only a little, because C_i nearly commutes with the state.

Proof of Theorem 4.1. Step 1: twirling the Z's, (i). Fix g and write $f(g) = LZ_i^{g_i}R$ with $R = Z_{i+1}^{g_{i+1}} \cdots Z_M^{g_M}$. Since $Z_i^2 = I$, $f(g+e_i) - f(g)Z_i = LZ_i^{g_i}(Z_iR - RZ_i)$, so $\|(f(g+e_i) - f(g)Z_i)\xi\|_2 = \|(R^*Z_iR - Z_i)\xi\|_2$. Put $v = Z_i\xi$. Then $R^*Z_iR\xi = R^*vR + R^*Z_iR(\xi - R^*\xi R)$, hence

$$\|(R^*Z_iR - Z_i)\xi\|_2 \leq \|R^*vR - v\|_2 + \|R^*\xi R - \xi\|_2.$$

Square, use $(x+y)^2 \leq 2x^2 + 2y^2$, average over g, and apply Lemma 3.9 twice, with all signs $+1$ and the involutions $Z_{i+1}, \ldots, Z_M$ (the vector R^*vR is v_g of that lemma, with Z_{i+1} innermost):

$$a_i^2 \leq 2\sum_{j>i}\|[Z_j, Z_i\xi]\|_2^2 + 2\sum_{j>i}\|[Z_j, \xi]\|_2^2.$$

Finally $[Z_j, Z_i\xi] = [Z_j, Z_i]\xi + Z_i[Z_j, \xi]$.

Step 2: twirling around X_i, (ii). Multiplying on the left by the unitary $f(g)^*$, $\|((-1)^{g_i}X_if(g) - f(g)X_i)\xi\|_2 = \|((-1)^{g_i}f(g)^*X_if(g) - X_i)\xi\|_2$. With $v = X_i\xi$ and $f = f(g)$, $(-1)^{g_i}f^*X_if\xi = (-1)^{g_i}f^*vf + (-1)^{g_i}f^*X_if(\xi - f^*\xi f)$, so the left side is at most $\|(-1)^{g_i}f^*vf - v\|_2 + \|f^*\xi f - \xi\|_2$. By Lemma 3.9 with the sign -1 at i and $+1$ elsewhere, $\mathbb{E}_g\|(-1)^{g_i}f^*vf - v\|_2^2 \leq \|\{Z_i, v\}\|_2^2 + \sum_{j\neq i}\|[Z_j, v]\|_2^2$, where $\{Z_i, X_i\xi\} = \{X_i, Z_i\}\xi - X_i[Z_i, \xi]$ and $[Z_j, X_i\xi] = [Z_j, X_i]\xi + X_i[Z_j, \xi]$. By Lemma 3.9 with all signs $+1$, $\mathbb{E}_g\|f^*\xi f - \xi\|_2^2 \leq \sum_j (c_j^Z)^2$. Squaring as in Step 1 gives (ii).

Step 3: the embedding. Let $\mathcal{H} = \mathbb{C}^d \otimes \ell^2(\mathbb{Z}_2^M)$ and define $W : \mathbb{C}^d \to \mathcal{H}$ by $Wy = 2^{-M/2}\sum_g f(g)y \otimes |g\rangle$; then $W^*W = 2^{-M}\sum_g f(g)^*f(g) = I$. Let $S_i|g\rangle = |g + e_i\rangle$ and $F_i|g\rangle = (-1)^{g_i}|g\rangle$, and put $C_i = I \otimes S_i$ and $\hat{X}_i = X_i \otimes F_i$. These are involutions, the C_i commute, and $F_iS_iF_i = -S_i$, $F_iS_jF_i = S_j$ for $j \neq i$, so

$$\hat{X}_iC_i\hat{X}_i = -C_i, \qquad \hat{X}_iC_j\hat{X}_i = C_j \quad (j \neq i).$$

Reindexing the sum, $(C_iW - WZ_i)\xi = 2^{-M/2}\sum_g (f(g+e_i) - f(g)Z_i)\xi \otimes |g\rangle$ and $(\hat{X}_iW - WX_i)\xi = 2^{-M/2}\sum_g((-1)^{g_i}X_if(g) - f(g)X_i)\xi \otimes |g\rangle$, so

$$\|(C_iW - WZ_i)\xi\|_2 = a_i, \qquad \|(\hat{X}_iW - WX_i)\xi\|_2 = b_i,$$

the Hilbert–Schmidt norm of an operator $\mathbb{C}^d \to \mathcal{H}$ summing over the 2^M blocks. Put $K = W\xi$, an operator $\mathbb{C}^d \to \mathcal{H}$ with $\|K\|_2 = 1$, and $\tilde{\rho} = KK^* = W\rho W^*$. Since $K^*K = \rho$, $S(\tilde{\rho}) = S(\rho)$.

Step 4: each coordinate carries nearly a bit. For $z \in \{\pm 1\}^M$ let $P_z = \prod_i \frac{1}{2}(I + z_iC_i)$ be the joint eigenprojections of the C_i, $\bar{\rho} = \sum_z P_z\tilde{\rho}P_z$ and $p(z) = \mathrm{Tr}(P_z\tilde{\rho}) = \|P_zK\|_2^2$. Since $\bar{\rho}$ is block diagonal with blocks of weights $p(z)$, the chain rule, and the fact that conditioning does not increase entropy, give

$$S(\bar{\rho}) \;\geq\; H(p) = \sum_i H(z_i \mid z_1, \ldots, z_{i-1}) \;\geq\; \sum_i H(z_i \mid z_{-i}),$$

where z_{-i} is z without its i-th coordinate and z has law p. Fix i and consider the unit vectors $u = K$ and $v = \hat{X}_i K X_i$ in the Hilbert–Schmidt space of operators $\mathbb{C}^d \to \mathcal{H}$. Since $\hat{X}_i P_z \hat{X}_i = P_{z^{(i)}}$, $\|P_z v\|_2^2 = \|P_z \hat{X}_i K\|_2^2 = p(z^{(i)})$; and, multiplying by X_i on the right,

$$\|u - v\|_2 = \|KX_i - \hat{X}_i K\|_2 \le \|W(\xi X_i - X_i \xi)\|_2 + \|(WX_i - \hat{X}_i W)\xi\|_2 = c_i^X + b_i.$$

By the reverse triangle inequality and the orthogonality of the P_z, $\sum_z \big(\sqrt{p(z)} - \sqrt{p(z^{(i)})}\big)^2 \le \sum_z \|P_z(u-v)\|_2^2 = \|u - v\|_2^2$. For a value y of z_{-i} let m_y and m'_y be the masses of $z_i = +1$ and $z_i = -1$, and $s_y = m_y + m'_y$. Then $1 - \sum_y 2\sqrt{m_y m'_y} = \frac{1}{2}\sum_z \big(\sqrt{p(z)} - \sqrt{p(z^{(i)})}\big)^2$, and Lemma 4.7 with $x = m_y/s_y$, multiplied by s_y and summed over y (terms with $s_y = 0$ vanish), gives

$$H(z_i \mid z_{-i}) = \sum_y s_y\, h(m_y/s_y) \;\ge\; 1 - \frac{1 - \sum_y 2\sqrt{m_y m'_y}}{\ln 2} \;\ge\; 1 - \frac{(b_i + c_i^X)^2}{2 \ln 2}.$$

Step 5: pinching costs little entropy. Let $\mathcal{T}_i(\sigma) = \frac{1}{2}(\sigma + C_i \sigma C_i)$, the pinching by the eigenprojections of C_i. The $\mathcal{T}_i$ commute, and $\bar{\rho} = \mathcal{T}_M \cdots \mathcal{T}_1 \tilde{\rho}$. Put $\sigma_i = \mathcal{T}_i \cdots \mathcal{T}_1 \tilde{\rho}$. By (2), commutation, and monotonicity of D under the channel $\mathcal{T}_{i-1} \cdots \mathcal{T}_1$,

$$S(\sigma_i) - S(\sigma_{i-1}) = D\big(\mathcal{T}_{i-1} \cdots \mathcal{T}_1 \tilde{\rho} \,\big\|\, \mathcal{T}_{i-1} \cdots \mathcal{T}_1 \mathcal{T}_i \tilde{\rho}\big) \le D(\tilde{\rho} \,\|\, \mathcal{T}_i \tilde{\rho}),$$

so $S(\bar{\rho}) \le S(\tilde{\rho}) + \sum_i D(\tilde{\rho} \| \mathcal{T}_i \tilde{\rho})$. Now $\tilde{\rho} = (KZ_i)(KZ_i)^*$ and $C_i \tilde{\rho} C_i = (C_i K)(C_i K)^*$, both factors of unit Hilbert–Schmidt norm, and

$$\|KZ_i - C_i K\|_2 \le \|W(\xi Z_i - Z_i \xi)\|_2 + \|(WZ_i - C_i W)\xi\|_2 = c_i^Z + a_i.$$

Lemma 4.8 gives $D(\tilde{\rho} \| \mathcal{T}_i \tilde{\rho}) \le h\big(\min((a_i + c_i^Z)^2/4, 1/2)\big)$.

Conclusion. Combining the bounds of Steps 4 and 5 on $S(\bar{\rho})$ with $S(\tilde{\rho}) = S(\rho)$ proves (iii). □

Proof of Corollary 4.2. By Theorem 4.1(i) and (ii), $a_i^2 \le 2(M-1)(2\kappa)^2 + 2(M-1)\kappa^2 \le 10M\kappa^2$ and $b_i^2 \le 2(2\kappa)^2 + 2(M-1)(2\kappa)^2 + 2M\kappa^2 = 10M\kappa^2$. For $M \ge 2$, $(\sqrt{10M} + 1)^2 = 10M + 2\sqrt{10M} + 1 \le 16M$, so $(a_i + c_i^Z)^2$ and $(b_i + c_i^X)^2$ are at most $16M\kappa^2 \le \frac{1}{10}$. Since h increases on $[0, \frac{1}{2}]$, each bracket in Theorem 4.1(iii) is at least $1 - \frac{0.1}{2\ln 2} - h(0.025)$. Here $\frac{0.1}{2\ln 2} < 0.0722$, and $h(0.025) = 0.025 \log 40 + 0.975 \log \frac{40}{39} < 0.1334 + 0.0361 = 0.1695$, using $\log 40 < \frac{16}{3}$ (from $5^3 < 2^7$) and $0.975 \ln \frac{40}{39} \le \frac{0.975}{39} = 0.025$. So each bracket exceeds $1 - 0.0722 - 0.1695 > \frac{3}{4}$. □

Proof of Corollary 4.3. Apply Theorem 4.1 with $\rho = I/D$. All state commutators vanish and $\|Y\xi\|_2 = \|Y\|_f$, so (i) and (ii) give $a_i^2 \le 2\eta$ and $b_i^2 \le 2\eta$, and (iii) gives

$$\log D = S(I/D) \;\ge\; \sum_{i=1}^{M} \Big[1 - \frac{2\eta}{2\ln 2} - h\big(\min(\eta/2, 1/2)\big)\Big] = M\big(1 - \eta/\ln 2 - h(\eta/2)\big),$$

since h increases on $[0, \frac{1}{2}]$ and $\eta \le 1$. The bracket decreases in η. For $\eta = \frac{1}{10}$, $0.1/\ln 2 < 0.1443$, and $h(0.05) = 0.05 \log 20 + 0.95 \log \frac{20}{19} < 0.2167 + 0.0722 = 0.2889$, using $\log 20 < \frac{13}{3}$ (from $5^3 < 2^7$) and $\log \frac{20}{19} \le \frac{1}{19 \ln 2}$. So the bracket exceeds $1 - 0.1443 - 0.2889 > \frac{1}{2}$, and $D \ge 2^{M/2}$. □

For the flat state, ξ commutes with every operator, so the second terms in Steps 1 and 2 vanish and (i), (ii) hold without the factor 2; the bracket in Corollary 4.3 then improves to $1 - \eta/(2\ln 2) - h(\eta/4)$. We do not need the improvement.

4.4 Entropy of a sector

The count is applied on the sector where a central involution acts as -1, to the sector state of Lemma 3.10. The entropy of the sector state controls the entropy of the original state.

Lemma 4.9 (Entropy of a sector)**.** *In the setting of Lemma 3.10,*

$$S(\rho) \;\geq\; p_-\, S(\rho_-) - h\big(\min(\|[J',\xi]\|_2^2/4,\, 1/2)\big) \qquad \textit{and} \qquad \operatorname{rank}\rho \geq \operatorname{rank}\rho_-.$$

Proof. The pinching $\rho^\natural = P_+\rho P_+ + P_-\rho P_- = \frac{1}{2}(\rho + J'\rho J')$ is the direct sum of $(1-p_-)\rho_+$ and $p_-\rho_-$, where $\rho_+ = P_+\rho P_+/(1-p_-)$ if $p_- < 1$, so $S(\rho^\natural) \geq p_-\, S(\rho_-)$. By (2) and Lemma 4.8 with $L = \xi J'$ and $L' = J'\xi$, for which $LL^* = \rho$ and $L'L'^* = J'\rho J'$,

$$S(\rho^\natural) - S(\rho) = D(\rho \,\|\, \rho^\natural) \leq h\big(\min(\|\xi J' - J'\xi\|_2^2/4, 1/2)\big).$$

Finally $\operatorname{rank}(P_-\rho P_-) \leq \operatorname{rank}\rho$. □

5 Linear system games keep the state

Slofstra and Vidick showed that near-optimal strategies for a linear system game give approximate representations of its solution group, and that the entanglement of the game is controlled by the hyperlinear profile of J [SV18]. For a general state their representation is accurate only to order $\varepsilon^{1/4}$: winning with probability $1-\varepsilon$ controls the relations, and the commutators of Bob's observables with $\rho^{1/2}$, to order $\sqrt{\varepsilon}$ in the state [SV18, Proposition 5.4], and the passage to a spectral projection of ρ halves the exponent again [SV18, Theorem 5.1 and Corollary 5.2]. This section keeps the state, in both directions. Theorem 5.1 records, with explicit constants, that a strategy gives a state-weighted model of the solution group at precision $O(\sqrt{\varepsilon})$ in Bob's reduced state; lower bounds on the state-weighted profile then transfer to the game with no loss. Theorem 5.4 proves the converse: a model at precision e gives a strategy that loses with probability $O(e^2)$ on the canonical purification of the same state. So the game and the state-weighted profile determine each other up to constant factors in the precision, in dimension and in entropy (Corollary 5.5). In this paper Theorem 5.1 feeds the lower bounds of Sections 10 and 11, and Theorem 5.4 with the flat state gives the upper bound of Section 10.

Throughout, $G = G(A,b)$ is a linear system game with question distribution π (Definition 2.7), $s = \max_i |V_i|$, and $\Gamma = \Gamma(A,b)$ is presented as in Definition 2.7(a), with $w = J$.

5.1 From strategies to models

Theorem 5.1 (Transfer)**.** *Let a strategy win G with probability at least $1-\varepsilon$, let r be the Schmidt rank of its state, and put $\theta = 2\sqrt{\varepsilon/\pi_{\min}}$. There are a homomorphism $\phi : F(x_1,\dots,x_n,J) \to \mathrm{U}(r)$ and a density matrix ρ of full rank on $\mathbb{C}^r$ such that $\phi(J) = -I$, every $\phi(x_j)$ is an involution, and, with $\xi = \rho^{1/2}$,*

(a) $\|(\phi(R) - I)\xi\|_2 \leq \max(s,4)\,\theta$ *for every relator R of Γ;*

(b) $\|[\phi(x_j),\xi]\|_2 \leq \sqrt{2}\,\theta$ *for every j.*

The density matrix ρ is Bob's reduced density matrix restricted to its support. In particular (ϕ,ρ) is a model of (Γ,J) at precision $\max(s,4)\theta$, so $r \geq \mathrm{hlp}^{\rho}_{\Gamma}(J;\max(s,4)\theta)$, and, if $\omega_q(G) = 1$,

$$\mathcal{E}(G,\varepsilon) \;\geq\; \mathrm{hlp}^{\rho}_{\Gamma}\big(J;\, 2\max(s,4)\sqrt{\varepsilon/\pi_{\min}}\big).$$

Proof. Step 1: reductions. Let Π be the support projection of Bob's reduced density matrix, of rank r. Since $\psi = (I\otimes\Pi)\psi$, replacing each Q_c^j by $\Pi Q_c^j \Pi$ on the range of Π,

identified with $\mathbb{C}^r$, gives POVMs with the same probabilities (Remark 2.9). With everything else fixed, the winning probability is $\mathrm{Tr}(B_j M_j)$ + const for the Hermitian operator $M_j = \sum_{i:\, j\in V_i} \frac{\pi(i,j)}{2} \mathrm{Tr}_A\big((A_{ij}\otimes I)|\psi\rangle\langle\psi|\big)$ on $\mathbb{C}^r$, so replacing B_j by the involution $\mathrm{sgn}(M_j)$ (with $\mathrm{sgn}(0)=1$), that is, Bob's measurement by $\{\frac{1}{2}(I \pm \mathrm{sgn}(M_j))\}$, does not lower it; we do this for each j in turn. Finally we replace Alice's POVMs by Naimark dilations (Remark 2.9), which changes neither Bob's reduced state nor any probability. Now Bob's observables B_j are involutions on $\mathbb{C}^r$, Bob's reduced density matrix ρ has full rank on $\mathbb{C}^r$, and, for each i, the A_{ij} ($j \in V_i$) are commuting involutions with $\prod_{j\in V_i} A_{ij} = (-1)^{b_i} I$, because every answer satisfies the equation. The strategy still wins with probability at least $1-\varepsilon$.

Step 2: consistency. A pair (i,j) with $j \in V_i$ is lost with probability $\ell_{ij} \le \varepsilon/\pi_{\min}$, and, as A_{ij} and B_j are involutions, $\|(A_{ij}\otimes I - I\otimes B_j)\psi\|^2 = 2 - 2\langle\psi,(A_{ij}\otimes B_j)\psi\rangle = 4\ell_{ij} \le \theta^2$. If $Y_1,\dots,Y_q$ are contractions on Alice's space with $\|(Y_l\otimes I - I\otimes B_{j_l})\psi\| \le \theta$, then

$$\big\|(I\otimes B_{j_1}\cdots B_{j_q})\psi - (Y_q\cdots Y_1\otimes I)\psi\big\| \le q\theta :$$

replace $(I\otimes B_{j_q})\psi$ by $(Y_q\otimes I)\psi$, at cost θ since the B's are contractions, move $Y_q\otimes I$ to the left past $I\otimes B_{j_1}\cdots B_{j_{q-1}}$, and induct.

Step 3: the relations, (a). Put $\phi(x_j) = B_j$ if j lies in some V_i, $\phi(x_j) = I$ otherwise, and $\phi(J) = -I$. The relators x_j^2, J^2 and $[x_j, J]$ are mapped to I. By Lemma 3.1(a), $\|(\phi(R) - I)\xi\|_2 = \|(I\otimes(\phi(R)-I))\psi\|$. For $j,k \in V_i$, the estimate of Step 2 with $(j_1,j_2,j_3,j_4) = (j,k,j,k)$, $Y_l = A_{ij_l}$ and $Y_4Y_3Y_2Y_1 = A_{ik}A_{ij}A_{ik}A_{ij} = I$ gives $\|(B_jB_kB_jB_k - I)\xi\|_2 \le 4\theta$, and $\phi([x_j,x_k]) = B_jB_kB_jB_k$. For $V_i = \{j_1 < \cdots < j_r\}$, the estimate with $Y_l = A_{ij_l}$ and $A_{ij_r}\cdots A_{ij_1} = (-1)^{b_i} I$ gives $\|((-1)^{b_i}B_{j_1}\cdots B_{j_r} - I)\xi\|_2 \le r\theta \le s\theta$, and $\phi(J^{b_i}x_{j_1}\cdots x_{j_r}) = (-1)^{b_i}B_{j_1}\cdots B_{j_r}$.

Step 4: the state commutators, (b). For $j \in V_i$, Lemma 3.1(b) gives a Hermitian $\tilde{A}$ on $\mathbb{C}^r$ with $\|B_j\xi - \xi\tilde{A}\|_2 \le \|(A_{ij}\otimes I - I\otimes B_j)\psi\| \le \theta$, and Lemma 3.2 gives $\|[B_j,\xi]\|_2 \le \sqrt{2}\,\theta$. For j in no V_i, $\phi(x_j) = I$.

Conclusion. Since $\phi(J) = -I$ commutes with ξ and $\sqrt{2} \le \max(s,4)$, the pair (ϕ,ρ) satisfies (S1)–(S3) for $w = J$ at precision $e = \max(s,4)\theta$, so $r \ge \mathrm{hlp}^{\rho}_{\Gamma}(J;e)$. Every strategy with $H_A = H_B = \mathbb{C}^d$ has Schmidt rank at most d, which gives the bound on $\mathcal{E}(G,\varepsilon)$. □

Remark 5.2 (Credits). For the default distribution $\pi_{\min} \ge 1/(ms)$, so the precision is at most $2\max(s,4)\sqrt{ms\,\varepsilon}$. For the flat state, (a) is [SV18, Proposition 4.4(b)] up to constants. For a general state, [SV18, Proposition 5.4] proves (a) and (b) with unspecified constants and then passes to a spectral projection of ρ, which gives $\mathrm{hlp}_{\Gamma}(J;2,C\varepsilon^{1/4}) \le \mathcal{E}(G,\varepsilon)$ [SV18, Corollary 5.2]. Theorem 5.1 stops before that step. It is bookkeeping; what it adds is that lower bounds on hlp^{ρ}, such as those of Theorems 8.1 and 11.14, apply at the natural precision $\sqrt{\varepsilon}$. A square root is also lost when near-optimal strategies for boolean constraint system games are rounded to strategies on a maximally entangled state [Pad22, Corollary 1.3].

5.2 From models to strategies

Conversely, a model that satisfies (S1)–(S3) at precision e gives a strategy that loses with probability $O(e^2)$, on a purification of the same state. For the flat state this is the classical direction of the correspondence [SV18, Proposition 4.4(a)]. The point here is that the state, and with it its spectrum, is kept.

Let ρ be a density matrix on $\mathbb{C}^d$ and $\xi = \rho^{1/2}$, and let ^{T} be the transpose in the standard basis. The map $\Theta : \mathbb{C}^d\otimes\mathbb{C}^d \to M_d(\mathbb{C})$ with $\Theta(v\otimes w) = wv^T$ is an isometry for $\|\cdot\|_2$ and satisfies $\Theta((Q\otimes I)x) = \Theta(x)Q^T$ and $\Theta((I\otimes Y)x) = Y\Theta(x)$. The *canonical purification* of ρ is $\psi_\rho = \Theta^{-1}(\xi)$. Bob's reduced state is ρ, Alice's is ρ^T, and we write $\eta = \xi^T$ for its square root. So $\|(C\otimes I)\psi_\rho\| = \|\xi C^T\|_2 = \|C\eta\|_2$ for operators C on Alice's side, and, for an operator B on Bob's side,

$$\|(B^T\otimes I - I\otimes B)\psi_\rho\| = \|\xi B - B\xi\|_2 = \|[B,\xi]\|_2. \tag{3}$$

Alice therefore imitates Bob's observable B by measuring B^T, at a cost equal to its state commutator. For the flat state, $\psi_{I/d} = d^{-1/2}\sum_k e_k \otimes e_k = \Phi_d$ is the maximally entangled state.

Lemma 5.3 (Sequential measurements)**.** *Let $B_1, \dots, B_k$ be involutions on $\mathbb{C}^d$, $c_j = \|[B_j, \xi]\|_2$ and $\kappa_{jl} = \|[B_j, B_l]\xi\|_2$. On ψ_ρ, let Alice measure $\{\frac{1}{2}(I + aB_1^T)\}_{a=\pm1}$, then $\{\frac{1}{2}(I + aB_2^T)\}_{a=\pm1}$, and so on up to B_k^T, recording the outcomes $a_1, \dots, a_k$, and let Bob measure one of the B_j, with outcome b_j.*

(a) Put $d_j = 2\Pr[a_j \neq b_j]^{1/2}$, so that $\Pr[a_j \neq b_j] = d_j^2/4$. Then $d_j \leq c_j + \sum_{l<j}\big(\kappa_{jl}/\sqrt{2} + \sum_{t<l} c_t\big)$.

(b) For a sign σ put $\tau = \|(B_1 \cdots B_k - \sigma I)\xi\|_2$. Then $\Pr[a_1 \cdots a_k \neq \sigma] \leq \frac{1}{4}\big(\tau + \sum_j d_j\big)^2$.

Proof. Put $Q_j = B_j^T$, $P_j^a = \frac{1}{2}(I + aQ_j)$ and, for $a \in \{\pm1\}^l$, $K_a = P_l^{a_l} \cdots P_1^{a_1}$. Then $\sum_a K_a^* K_a = \sum_a K_a K_a^* = I$, by induction on l. The coherent form of Alice's procedure is the isometry $Vy = \sum_{a\in\{\pm1\}^k} K_a y \otimes |a\rangle$ into $\mathbb{C}^d \otimes \mathbb{C}^{2^k}$, and a_j is the eigenvalue of the record observable Z_j, which multiplies $|a\rangle$ by a_j. Since Z_j and B_j are commuting involutions on different systems, $\Pr[a_j \neq b_j] = \frac{1}{4}d_j^2$ with $d_j = \|(Z_j \otimes I - I \otimes B_j)(V \otimes I)\psi_\rho\|$.

Branch commutators. Let $D_l = \big(\sum_{a\in\{\pm1\}^l}\|[K_a, \eta]\|_2^2\big)^{1/2}$, so $D_0 = 0$. Writing $K_a = P_l^{a_l} K_{a'}$, $[K_a, \eta] = P_l^{a_l}[K_{a'}, \eta] + [P_l^{a_l}, \eta]K_{a'}$ and $[P_l^{a_l}, \eta] = \frac{a_l}{2}[Q_l, \eta]$. Minkowski's inequality and the two completeness relations give $D_l \leq D_{l-1} + \|[Q_l, \eta]\|_2/\sqrt{2}$, and $\|[Q_l, \eta]\|_2 = \|(\xi B_l - B_l\xi)^T\|_2 = c_l$. Hence $D_l \leq \sum_{t\leq l} c_t/\sqrt{2}$. For an operator C on $\mathbb{C}^d$, $CK_a\eta = C\eta K_a + C[K_a, \eta]$, and $\sum_a \|C\eta K_a\|_2^2 = \mathrm{Tr}(\eta C^* C\eta \sum_a K_a K_a^*) = \|C\eta\|_2^2$, so

$$\Big(\sum_{a\in\{\pm1\}^l}\|CK_a\eta\|_2^2\Big)^{1/2} \leq \|C\eta\|_2 + \|C\|_{\mathrm{op}} D_l. \tag{4}$$

(a). The component of $(Z_j \otimes I - I \otimes B_j)(V \otimes I)\psi_\rho$ along $|a\rangle$ is $(a_j K_a \otimes I - K_a \otimes B_j)\psi_\rho$. Write $K_a = L_a R_a$ with $L_a = P_k^{a_k} \cdots P_j^{a_j}$ and $R_a = P_{j-1}^{a_{j-1}} \cdots P_1^{a_1}$; since $a_j P_j^{a_j} = P_j^{a_j} Q_j$, $a_j K_a = L_a Q_j R_a$. For fixed $a' = (a_1, \dots, a_{j-1})$, $\sum_{a_j,\dots,a_k} L_a^* L_a = I$, so

$$d_j^2 = \sum_{a'\in\{\pm1\}^{j-1}} \big\|(Q_j R_{a'} \otimes I - R_{a'} \otimes B_j)\psi_\rho\big\|^2.$$

Now $Q_j R_{a'} = R_{a'} Q_j + [Q_j, R_{a'}]$. The vectors $(R_{a'} \otimes I)(Q_j \otimes I - I \otimes B_j)\psi_\rho$ have total squared norm c_j^2, by completeness and (3), and $\|([Q_j, R_{a'}] \otimes I)\psi_\rho\| = \|[Q_j, R_{a'}]\eta\|_2$. Expanding $[Q_j, R_{a'}] = \sum_{l<j} P_{j-1}^{a_{j-1}} \cdots P_{l+1}^{a_{l+1}} \frac{a_l}{2}[Q_j, Q_l]\, K_{(a_1,\dots,a_{l-1})}$ and using Minkowski's inequality over l, completeness of the left factors and the two values of a_l,

$$\Big(\sum_{a'}\|[Q_j, R_{a'}]\eta\|_2^2\Big)^{1/2} \leq \sum_{l<j}\frac{1}{\sqrt{2}}\Big(\sum_{a''\in\{\pm1\}^{l-1}}\|[Q_j, Q_l]K_{a''}\eta\|_2^2\Big)^{1/2} \leq \sum_{l<j}\frac{\kappa_{jl} + 2D_{l-1}}{\sqrt{2}},$$

by (4) with $\|[Q_j, Q_l]\|_{\mathrm{op}} \leq 2$ and $\|[Q_j, Q_l]\eta\|_2 = \|(\xi[B_l, B_j])^T\|_2 = \kappa_{jl}$. With Minkowski's inequality and the bound on D_{l-1} this proves (a).

(b). $\Pr[a_1 \cdots a_k \neq \sigma] = \frac{1}{4}\|(Z_1 \cdots Z_k - \sigma)(V \otimes I)\psi_\rho\|^2$. The record observables commute with each other and with Bob's operators, so

$$Z_1 \cdots Z_k - B_1 \cdots B_k = \sum_j B_1 \cdots B_{j-1} Z_{j+1} \cdots Z_k (Z_j - B_j),$$

which has norm at most $\sum_j d_j$ on $(V \otimes I)\psi_\rho$. Bob's operators commute with $V \otimes I$, so $\|(B_1 \cdots B_k - \sigma)(V \otimes I)\psi_\rho\| = \|(I \otimes (B_1 \cdots B_k - \sigma))\psi_\rho\| = \tau$. □

Theorem 5.4 (Return). *Let (ϕ,ρ) be a model of (Γ,J) at precision e on $\mathbb{C}^d$, and put $B_j=\mathrm{r}(\phi(x_j))$ (Lemma 3.6). Consider the strategy on ψ_ρ in which Bob, on question j, measures B_j, and Alice, on question i with $V_i=\{j_1<\cdots<j_r\}$, measures $B^T_{j_1},\dots,B^T_{j_r}$ in turn as in Lemma 5.3 and answers the recorded signs, flipping the last one if their product is not $(-1)^{b_i}$. Every question pair (i,j) with $j\in V_i$ is lost with probability at most C_se^2, where*

$$\nu_l=3\Big(1+\frac{(l-1)(l-2)}{2}\Big)+\frac{13(l-1)}{\sqrt{2}},\qquad C_s=\frac{1}{4}\Big[\nu_s^2+\Big(s^2+2+\sum_{l=1}^{s}\nu_l\Big)^2\Big];$$

for instance $C_3<788.2$. The strategy has local dimensions d, Bob's reduced state is ρ, and its entanglement entropy is $S(\rho)$. For $\rho=I/d$ the state is Φ_d: every e-representation of Γ on $\mathbb{C}^d$ with $\phi(J)=-I$ gives a strategy on Φ_d that loses every question pair with probability at most C_se^2, and hence $\mathcal{E}^{\mathrm{me}}(G,\varepsilon)\le\mathrm{hlp}_\Gamma\big(J;2,\sqrt{\varepsilon/C_s}\big)$ for every $\varepsilon>0$.

Proof. Step 1: rounding and pairs. Write $U_j=\phi(x_j)$. By Lemma 3.6, the relator x_j^2 and (S2), $\|(B_j-U_j)\xi\|_2\le e$ and $\|[B_j,\xi]\|_2\le 3e$. For j,l in one equation let $R=U_jU_lU_j^*U_l^*=\phi([x_j,x_l])$. Then $[U_j,U_l]=(R-I)U_lU_j$, so $\|[U_j,U_l]\xi\|_2\le e+2\cdot 2e=5e$. Also $B_jB_l-U_jU_l=B_j(B_l-U_l)+(B_j-U_j)U_l$ has $\|(B_jB_l-U_jU_l)\xi\|_2\le e+(e+2e)=4e$, and the same holds in the other order, so $\kappa_{jl}=\|[B_j,B_l]\xi\|_2\le 13e$.

Step 2: the parity of an equation. For equation i let $U=U_{j_1}\cdots U_{j_r}$ and $U_J=\phi(J)$. The relator gives $\|(U_J^{b_i}U-I)\xi\|_2\le e$; if $b_i=1$, then $U+I=U_J^*(U_JU-I)+U_J^*(U_J+I)$ and (S3) give $\|(U+I)\xi\|_2\le 2e$. Replacing the r factors by the B's costs at most $\sum_{l=1}^{r}(e+2(r-l)e)=r^2e$, by Lemma 3.3(e) with the state commutators $\|[U_{j_l},\xi]\|_2\le e$. So $\|(B_{j_1}\cdots B_{j_r}-(-1)^{b_i})\xi\|_2\le(r^2+2)e$.

Step 3: the losing probability. By Lemma 5.3(a), with $c_t\le 3e$ and $\kappa\le 13e$, $d_l\le\big(3+13(l-1)/\sqrt{2}+3(l-1)(l-2)/2\big)e=\nu_le$ for the l-th variable of the equation. Flipping the last record changes Alice's answer only on the event of wrong parity, so by Lemma 5.3 the pair (i,j_l) is lost with probability at most $\frac14 d_l^2+\frac14\big((r^2+2)e+\sum_{l'}d_{l'}\big)^2\le C_se^2$, as ν_l increases with l and $r\le s$. Alice's procedure is a measurement on $\mathbb{C}^d$ followed by a classical function of its outcomes, and it always answers an assignment that satisfies equation i.

The flat state. For $\rho=I/d$, $\psi_\rho=\Phi_d$, condition (S2) is empty, and (S1), (S3) say $\|\phi(R)-I\|_f\le e$ for the relators R and $\|\phi(J)+I\|_f\le e$; an e-representation with $\phi(J)=-I$ satisfies them. Given $\varepsilon>0$, let $e=\sqrt{\varepsilon/C_s}$; if $\mathrm{hlp}_\Gamma(J;2,e)=d<\infty$, the strategy just built lives on Φ_d and loses with probability at most $C_se^2=\varepsilon$, so $\mathcal{E}^{\mathrm{me}}(G,\varepsilon)\le d$. □

Numerically, $\nu_1=3$, $\nu_2=3+13/\sqrt{2}<12.193$ and $\nu_3=6+26/\sqrt{2}<24.385$, so $C_3<\frac14(24.385^2+50.578^2)<788.2$. The constant is far from optimal. Nothing is postselected: every run of Alice's procedure produces a valid answer, and the shared state is ψ_ρ itself.

5.3 The two-sided correspondence

Corollary 5.5 (Two-sided correspondence). *Let $\omega_q(G)=1$, put $C_G=2\max(s,4)/\sqrt{\pi_{\min}}$, and let C_s be as in Theorem 5.4.*

(a) For every $\varepsilon>0$, $\mathrm{hlp}^\rho_\Gamma\big(J;C_G\sqrt{\varepsilon}\big)\le\mathcal{E}(G,\varepsilon)\le\mathrm{hlp}^\rho_\Gamma\big(J;\sqrt{\varepsilon/C_s}\big)$.

(b) Let $q(\varepsilon)$ be the infimum of the entanglement entropies of the strategies that win with probability at least $1-\varepsilon$. Then $h^S_\Gamma\big(C_G\sqrt{\varepsilon}\big)\le q(\varepsilon)\le h^S_\Gamma\big(\sqrt{\varepsilon/C_s}\big)$.

(c) Let $e_\Gamma(d)$ be the infimum of the precisions e with $\mathrm{hlp}^\rho_\Gamma(J;e)\le d$, and $\ell_G(d)$ the infimum of the losing probabilities of strategies with $H_A=H_B=\mathbb{C}^d$. Then $e_\Gamma(d)^2/C_G^2\le\ell_G(d)\le C_s\,e_\Gamma(d)^2$ for every d.

Proof. The first inequalities of (a) and (b) are Theorem 5.1, whose model lives on the Schmidt support with ρ equal to Bob's reduced state there, so that $S(\rho)$ is the entanglement entropy. The second inequalities follow from Theorem 5.4 with $e = \sqrt{\varepsilon/C_s}$: the strategy on ψ_ρ has local dimension d, entanglement entropy $S(\rho)$, and losing probability at most $C_s e^2 = \varepsilon$. For (c), a strategy on $\mathbb{C}^d \otimes \mathbb{C}^d$ with losing probability ℓ has Schmidt rank $r \le d$, and Theorem 5.1 gives a model on $\mathbb{C}^r$ at precision $C_G\sqrt{\ell}$. Replacing $\phi(x_j)$ by $\phi(x_j) \oplus I$, $\phi(J)$ by $-I_d$ and ρ by $\rho \oplus 0$ gives a model on $\mathbb{C}^d$ with the same state-weighted quantities, since $\xi \oplus 0$ vanishes on the added block; so $e_\Gamma(d) \le C_G\sqrt{\ell_G(d)}$. Conversely, for every $e > e_\Gamma(d)$ there is a model on some $\mathbb{C}^{d'}$, $d' \le d$, at precision e, and Theorem 5.4 gives a strategy with local dimension d' and losing probability at most $C_s e^2$, which we pad to $\mathbb{C}^d$. □

The state commutators in (S2) are needed for the return direction, not only for counting.

Example 5.6 (The return needs the state commutators)**.** Let $B_1 = B_2 = \sigma_Z$ on $\mathbb{C}^2$, $\rho = |+\rangle\langle+|$, and consider the single equation $x_1 + x_2 = 0$, with $\phi(x_1) = \phi(x_2) = \sigma_Z$ and $\phi(J) = -I$. The relations hold exactly, $B_1 B_2 = I$, but $\|[B_j, \xi]\|_2 = \sqrt{2}$. The canonical purification is the product state $|+\rangle \otimes |+\rangle$; Alice's two sequential measurements of σ_Z agree with each other, and Bob's measurement of σ_Z is independent of them, so the strategy of Theorem 5.4 loses with probability $\frac{1}{2}$. The game itself is classically winnable; what fails is the construction without its hypothesis.

Remark 5.7 (Credits). That flat approximate representations give strategies with quadratically small loss on a maximally entangled state is due to Slofstra and Vidick [SV18, Proposition 4.4(a)], following [Slo19, Proposition 3.4]; the two-sided flat correspondence for maximally entangled strategies is [SV18, Corollary 4.6]. Paddock develops the correspondence between near-optimal strategies and approximate representations, including rounding to maximally entangled strategies [Pad22]. What Theorem 5.4 and Corollary 5.5 add is that an arbitrary state, with its spectrum, is kept in both directions at the natural precision $e \asymp \sqrt{\varepsilon}$. So the dimension, the entanglement entropy, or any other function of the spectrum that does not change when zero eigenvalues are added can be read off the state-weighted profile.

6 The Pauli configuration-lamp group

This section builds the group G_{P} behind Theorems A and B and proves the facts about it that the rest of the paper uses: J is a central involution with $J \neq 1$, the six elements p, q, t, d, x, z generate G_{P} (Proposition 6.6), and G_{P} is amenable (Proposition 6.8). Its extension G_{P}^{+} by a ray swap has the same properties (Lemma 6.9); the game of Section 10 is built on it. The construction has three layers, drawn in Fig. 1. The *points* are those of the three rays of Houghton's group H_3, which moves them by permutations that are eventually translations. The *sites* are the finite sets of points, written as binary vectors and called configurations; an affine group A moves sites by permuting points, toggling a point, or adding one point into another. Each site carries a *Pauli pair*, a copy of the dihedral group $D_4 \cong \langle \sigma_X, \sigma_Z \rangle$, and the centres of all the pairs are identified to one element J. On N points there are 2^N sites and hence 2^N Pauli pairs, any two of which commute, while the generators t and d that change sites involve only one and two points. Section 7 shows that the commutation of any two of these pairs follows from a fixed finite set of relations at a cost quasi-polynomial in N, and Section 8 turns this into the entanglement lower bound. That cost is measured by the Dehn function of a presentation of H_3, which we fix in Section 6.1 together with the bound proved in Appendix A. The host group A and the generators p, q, t, d are those of [DM26a, Section 3], where the lamps are copies of S_3.

6.1 Houghton's group

Definition 6.1 (Houghton's group)**.** Let $Y = \{1,2,3\} \times \{1,2,\dots\}$; the point (i,j) lies on *ray* i at *height* j. *Houghton's group* $H = H_3$ [Hou78] is the group of permutations g of Y that are eventually translations on each ray: there are integers $m_1(g), m_2(g), m_3(g)$ with $g(i,j) = (i, j + m_i(g))$ for all large j. Let $p \in H$ be the permutation

$$p(1,j) = (1,j+1), \qquad p(2,1) = (1,1), \qquad p(2,j) = (2,j-1) \ \ (j \geq 2), \qquad p(3,j) = (3,j),$$

let q be defined in the same way with rays 2 and 3 exchanged, and put $s = [p,q] = pqp^{-1}q^{-1}$.

The finitary permutations of Y form a subgroup $\mathrm{FSym}(Y) \leq H$. The map $g \mapsto (m_1(g), m_2(g), m_3(g))$ is a homomorphism to $\mathbb{Z}^3$, and its kernel is $\mathrm{FSym}(Y)$, because an element with all $m_i(g) = 0$ fixes every point of large height. Moreover $m_1(g)+m_2(g)+m_3(g) = 0$: if g translates every point of height above K on ray i by $m_i(g)$, and B is the set of the $3K$ points of height at most K, then $g(B)$ is the complement of $g(Y \setminus B)$, which has $\sum_i (K + m_i(g))$ points. As $m(p) = (1,-1,0)$ and $m(q) = (1,0,-1)$, recording $-m_2$ and $-m_3$ gives a surjective homomorphism

$$H \to \mathbb{Z}^2, \qquad p \mapsto (1,0), \quad q \mapsto (0,1), \qquad \text{with kernel } \mathrm{FSym}(Y). \tag{5}$$

On the first ray we write i for the point $(1,i)$ and $(i\ j)$ for the transposition of the points i and j.

Lemma 6.2 (Generators of H)**.**

(a) For $k \geq 0$ the element $\sigma_k = p^k s p^{-k}$ is the transposition $(k{+}1\ \ k{+}2)$; in particular $s = (1\ 2)$.

(b) H is generated by p and q.

Proof. (a) Composing from right to left, the letters q^{-1}, p^{-1}, q, p of s move $(1,1)$ through $(3,1), (3,1), (1,1)$ to $(1,2)$, and $(1,2)$ through $(1,1), (2,1), (2,1)$ to $(1,1)$. They move $(1,j)$ with $j \geq 3$ through $(1,j{-}1), (1,j{-}2), (1,j{-}1)$ back to $(1,j)$, a point $(2,j)$ through $(2,j), (2,j{+}1), (2,j+1)$ back to $(2,j)$, and a point $(3,j)$ through $(3,j+1), (3,j+1), (3,j)$ back to $(3,j)$. So $s = (1\ 2)$. Since p^k maps the points $1,2$ to $k+1, k+2$, the conjugate σ_k is $(k{+}1\ \ k{+}2)$.

(b) Let $H' = \langle p, q \rangle$. By (a), H' contains the adjacent transpositions of ray 1, hence every finitary permutation of Y supported on ray 1. Let $D \subset Y$ be finite and choose k, l larger than the heights of its points. The element $g = p^k q^l$ maps D into ray 1: q^l maps ray 1 into itself and the points of height at most l on ray 3 into ray 1, and fixes ray 2; then p^k maps ray 1 into itself and the points of height at most k on ray 2 into ray 1. So for a permutation π supported in D, the conjugate $g\pi g^{-1}$ is supported on ray 1 and lies in H', and so does π. Hence $\mathrm{FSym}(Y) \leq H'$, and since H' maps onto $\mathbb{Z}^2$ under (5), whose kernel is $\mathrm{FSym}(Y)$, we get $H' = H$. □

Johnson [Joh99] gave a finite presentation of H on the generators p, q and $\alpha = [p,q]$; see also [Lee12, Theorem 2.8]. Eliminating α gives the presentation

$$P_H = \langle p, q \mid \ s^2,\ (sps\,p^{-1})^3,\ [s, p^2 s p^{-2}],\ psp^{-1}qs^{-1}q^{-1} \rangle, \qquad s = [p,q]. \tag{6}$$

Its relators hold in H by Lemma 6.2(a): $s = (1\ 2)$ and $psp^{-1} = (2\ 3)$, so $sps\,p^{-1}$ is a 3-cycle; s commutes with $p^2 s p^{-2} = (3\ 4)$; and $qsq^{-1} = (2\ 3) = psp^{-1}$, because q acts on ray 1 as p does. We do not use Johnson's theorem: Appendix A reproves that P_H presents H, with an explicit bound on its Dehn function (Corollary A.5).

For a set W of words, a word w has *area at most k in W* if it is freely equal to a product of k conjugates of elements of $W^{\pm 1}$; the *area* $\mathrm{Area}_W(w)$ is the least such k, and ∞ if there

is none. Let $\delta_H(n)$ be the largest area, with respect to the four relators of P_H, of a word of length at most n in $p^{\pm1}, q^{\pm1}$ that represents 1 in H. This is the Dehn function of P_H; it is nondecreasing. Two bounds on it are available.

(H1) $\delta_H(n) \le 2^{(\log n+6)^2}$ for all $n \ge 1$. This is Corollary A.5, proved in Appendix A from Johnson's presentation alone, by splitting in half Lee's recursion for commutators of far-apart transpositions [Lee12, Section 4.2]. Lee's own bound is exponential [Lee12, Theorem D].

(H2) $\delta_H(n) \le C_H n^{p_H}$ for an absolute constant C_H, where $p_H = 4 + \log 107 \approx 10.74$. This is proved in [DM26b, Theorems 3.2 and 3.3] for Johnson's presentation, through the bound $135M^{\log 107}$ on the area of the commutator of two transpositions at distance M, and it passes to P_H by Lemma A.1. (Permutations act on the right in [DM26b], and its generators are the inverses, as maps, of ours; since $g \mapsto g^{-1}$ turns that product into composition, the two presentations match letter by letter and have the same areas.)

Every theorem of this paper uses only (H1). The bound (H2) rests on an unrefereed preprint and appears only in remarks, where it improves factors $2^{-O((\log\log(1/\varepsilon))^2)}$ to powers of $\log(1/\varepsilon)$; those remarks write $\kappa = 2 + p_H = 6 + \log 107$.

6.2 The affine group of configurations

Let $V = \mathbb{F}_2^{(Y)}$ be the space of finitely supported functions $Y \to \mathbb{F}_2$, the *configurations*, with basis $(e_y)_{y\in Y}$. We identify a configuration with its support, a finite set of points. The group H acts faithfully on V by permutation matrices, $g(e_y) = e_{g(y)}$, and we identify H with its image in $\mathrm{GL}(V)$. Let $\mathrm{GL}_{\mathrm{fin}}(Y)$ be the group of linear automorphisms of V that fix all but finitely many basis vectors. It is the directed union of the finite groups $\mathrm{GL}(\mathbb{F}_2^D)$ over finite $D \subset Y$, each extended by the identity on the other basis vectors. It is normalized by H, and $H \cap \mathrm{GL}_{\mathrm{fin}}(Y) = \mathrm{FSym}(Y)$, because a permutation matrix fixes all but finitely many basis vectors exactly when the permutation is finitary. Put

$$L = \mathrm{GL}_{\mathrm{fin}}(Y)\, H \le \mathrm{GL}(V), \qquad A = V \rtimes L,$$

so that A is the group of affine maps $v \mapsto g(v) + u$ of V with $g \in L$ and $u \in V$. Write τ_u for the translation $v \mapsto v + u$, and for distinct $y, y' \in Y$ let $E_{y\leftarrow y'}$ be the transvection that sends $e_{y'}$ to $e_{y'} + e_y$ and fixes the other basis vectors. For every permutation g of Y, acting on V by permutation matrices, and every $M \in \mathrm{GL}(V)$,

$$g\tau_{e_y}g^{-1} = \tau_{e_{g(y)}}, \qquad gE_{y\leftarrow y'}g^{-1} = E_{g(y)\leftarrow g(y')}, \qquad M\tau_u M^{-1} = \tau_{M(u)}. \tag{7}$$

On the first ray we write $e_k = e_{(1,k)}$ and $\tau_k = \tau_{e_k}$, and put

$$t = \tau_1, \qquad d = E_{(1,2)\leftarrow(1,1)}.$$

Thus t toggles the point $(1,1)$ in a configuration, and d toggles $(1,2)$ in every configuration that contains $(1,1)$. Linear maps fix the zero configuration 0.

Lemma 6.3.

(a) L is generated by p, q, d, and A is generated by p, q, t, d.

(b) A acts doubly transitively on V, and the stabilizer of 0 is L.

Proof. (a) The subgroup $\langle p, q, d\rangle$ contains H by Lemma 6.2(b). Since $\mathrm{FSym}(Y) \le H$ acts doubly transitively on Y, the second identity of (7) puts every transvection $E_{y\leftarrow y'}$ into it.

Transvections generate $\mathrm{GL}(\mathbb{F}_2^D)$ for every finite D, by row reduction, so $\langle p, q, d\rangle$ contains $\mathrm{GL}_{\mathrm{fin}}(Y)$, hence L. By the first identity of (7) the H-conjugates of t are all the τ_{e_y}, which generate the translations; so $\langle p, q, t, d\rangle = A$.

(b) An affine map that fixes 0 is linear, so the stabilizer of 0 is L. Given $v \neq w$ in V, the translation τ_v carries (v, w) to $(0, v + w)$ with $v + w \neq 0$. Choose a finite $D \subset Y$ containing $(1, 1)$ and the support of $v + w$; some $M \in \mathrm{GL}(\mathbb{F}_2^D) \leq \mathrm{GL}_{\mathrm{fin}}(Y)$ sends $v + w$ to e_1, so $M\tau_v$ carries (v, w) to $(0, e_1)$. Hence A is transitive on ordered pairs of distinct configurations. □

6.3 Pauli lamps

Let $D_4 = \langle x, z \mid x^2, z^2, (xz)^4\rangle$ be the dihedral group of order eight. Its centre is $\{1, (xz)^2\}$, and $(xz)^2 = [x, z]$ because x and z are involutions. The assignment $x \mapsto \sigma_X$, $z \mapsto \sigma_Z$ to the Pauli matrices is a faithful representation that sends $(xz)^2$ to $-I$; we call D_4 the *Pauli lamp*.

For $a, b \in \mathbb{F}_2^{(V)}$, finitely supported functions $V \to \mathbb{F}_2$, put $\langle b, a\rangle = \sum_{v\in V} b(v)a(v) \in \mathbb{F}_2$. Let E_V be the set $\mathbb{F}_2 \times \mathbb{F}_2^{(V)} \times \mathbb{F}_2^{(V)}$ with the product

$$(c, a, b)\,(c', a', b') = \big(c + c' + \langle b, a'\rangle,\ a + a',\ b + b'\big). \tag{8}$$

This is a group: by bilinearity both bracketings of a triple product have first coordinate $c + c' + c'' + \langle b, a'\rangle + \langle b, a''\rangle + \langle b', a''\rangle$, the identity is $(0, 0, 0)$, and the inverse of (c, a, b) is $(c + \langle b, a\rangle, a, b)$. Put

$$J = (1, 0, 0), \qquad x_v = (0, \mathbf{1}_v, 0), \qquad z_v = (0, 0, \mathbf{1}_v) \qquad (v \in V),$$

where $\mathbf{1}_v$ is the indicator function of $\{v\}$, and for $a, b \in \mathbb{F}_2^{(V)}$ let $X^a = \prod_v x_v^{a(v)}$ and $Z^b = \prod_v z_v^{b(v)}$. By (8) the x_v commute with each other, as do the z_v; x_v commutes with z_w for $v \neq w$; and

$$z_v x_v = (1, \mathbf{1}_v, \mathbf{1}_v) = J\, x_v z_v, \qquad (x_v z_v)^2 = J, \qquad J^c X^a Z^b = (c, a, b).$$

So every element of E_V has the unique *normal form* $J^c X^a Z^b$, and (8) is the rule $Z^b X^{a'} = J^{\langle b, a'\rangle} X^{a'} Z^b$. The element J is central in E_V, since $\langle b, 0\rangle = \langle 0, a\rangle = 0$, and $J^2 = 1$.

Lemma 6.4. *The homomorphism $\bigoplus_{v\in V} D_4 \to E_V$ that sends x and z in the summand at v to x_v and z_v is surjective, and its kernel is the subgroup Z_{ev} of products of an even number of the central elements $J_v = (xz)_v^2$. Thus E_V is the central product of Pauli lamps, one at each configuration, with all centres identified to J.*

Proof. The relations of D_4 hold for x_v, z_v, since $(x_v z_v)^4 = J^2 = 1$, and elements attached to distinct sites commute, so the homomorphism exists. Its image contains all x_v, z_v and $J = (x_0 z_0)^2$, so it is onto by the normal form. Each J_v maps to J, so Z_{ev} lies in the kernel. For a finite nonempty set $F \subset V$, the subgroup $\bigoplus_{v\in F} D_4$ has order $8^{|F|}$ and meets Z_{ev} in a group of order $2^{|F|-1}$, while its image $\{(c, a, b) : a, b \text{ supported in } F\}$ has order $2 \cdot 4^{|F|} = 8^{|F|}/2^{|F|-1}$. So the kernel on $\bigoplus_{v\in F} D_4$ is exactly $Z_{\mathrm{ev}} \cap \bigoplus_{v\in F} D_4$, and since F is arbitrary the kernel is Z_{ev}. □

The group A permutes V, and it acts on E_V by moving sites and fixing J:

$$g \cdot (c, a, b) = \big(c,\ a \circ g^{-1},\ b \circ g^{-1}\big), \qquad \text{so that} \qquad g \cdot x_v = x_{g(v)}, \quad g \cdot z_v = z_{g(v)}, \quad g \cdot J = J.$$

This is an action by automorphisms, because g is a bijection of V and so $\langle b \circ g^{-1}, a' \circ g^{-1}\rangle = \langle b, a'\rangle$. Under Lemma 6.4 it is induced by the permutation of the summands of $\bigoplus_V D_4$, which preserves Z_{ev}.

6.4 The groups G_{P} and G_{P}^{+}

Definition 6.5 (The Pauli configuration-lamp group)**.** The *Pauli configuration-lamp group* is $G_{\mathrm{P}} = E_V \rtimes A$. We write $x = x_0$ and $z = z_0$ for the Pauli pair at the zero configuration, so that $J = (xz)^2 = [x, z]$, and $p, q, t, d \in A$ are as above.

In G_{P} we have $gx_vg^{-1} = x_{g(v)}$, $gz_vg^{-1} = z_{g(v)}$ and $gJg^{-1} = J$ for $g \in A$. Equivalently, G_{P} is the quotient of the permutational wreath product $D_4 \wr_V A = \left(\bigoplus_V D_4\right) \rtimes A$ by the central subgroup Z_{ev}, which is normal because A preserves it.

Proposition 6.6. *J is a central involution of G_{P} and $J \neq 1$. It commutes with p, q, t, d, x, z, and these six elements generate G_{P}.*

Proof. J is central in E_V and fixed by the action of A, so it is central in $E_V \rtimes A$; $J^2 = 1$ by (8), and $J = (1, 0, 0) \neq (0, 0, 0)$. By Lemma 6.3(a) the subgroup generated by the six elements contains A. For $v \in V$ the translation τ_v sends 0 to v, so the subgroup contains $x_v = \tau_v x \tau_v^{-1}$, $z_v = \tau_v z \tau_v^{-1}$ and $J = (xz)^2$, which generate E_V by the normal form. □

Remark 6.7 (Finite presentation). The group G_{P} is finitely presented. By [DM26a, Proposition 7.1] the group A is finitely presented, so by Lemma 6.3(a) it has a finite presentation $\mathcal{R}_A$ on p, q, t, d. Adding the generators x, z and the relators x^2, z^2, $(xz)^4$, $[\ell, u]$ for $\ell \in \{p, q, d\}$ and $u \in \{x, z\}$, $[u, tu't^{-1}]$ for $u, u' \in \{x, z\}$, and $[(xz)^2, t]$, which are the lamp words of the certificate R_{P} of Definition 7.1, gives a presentation of G_{P}. Indeed, the stabilizer $L = \langle p, q, d\rangle$ of 0 centralizes $\langle x, z\rangle$, so the conjugate $g\langle x, z\rangle g^{-1}$ depends only on $g(0)$; double transitivity (Lemma 6.3(b)) carries the commutation of the pairs at 0 and e_1, and the identification of their centres, to every two distinct sites; and Lemma 6.4 identifies the group so presented with G_{P}. This paper never uses finite presentability. All profiles are taken with respect to explicit finite sets of words valid in G_{P} that contain R_{P}, and Theorem B holds for every such set, hence in particular for every finite presentation of G_{P} on p, q, t, d, x, z.

Proposition 6.8 (Amenability)**.** *Let $\pi : G_{\mathrm{P}} \to \mathbb{Z}^2$ be the composite $G_{\mathrm{P}} \to A \to L \to L/\operatorname{GL}_{\mathrm{fin}}(Y) \cong H/\operatorname{FSym}(Y) \cong \mathbb{Z}^2$, the last map being (5). Its kernel $E_V \rtimes \left(V \rtimes \operatorname{GL}_{\mathrm{fin}}(Y)\right)$ is locally finite. Hence G_{P} is locally finite-by-$\mathbb{Z}^2$; in particular it is elementary amenable, amenable, sofic and hyperlinear.*

Proof. The map $A \to L$ is the linear part, with kernel V. Since $L = \operatorname{GL}_{\mathrm{fin}}(Y)H$ with $\operatorname{GL}_{\mathrm{fin}}(Y)$ normal, the inclusion induces an isomorphism $H/(H \cap \operatorname{GL}_{\mathrm{fin}}(Y)) \cong L/\operatorname{GL}_{\mathrm{fin}}(Y)$, and $H \cap \operatorname{GL}_{\mathrm{fin}}(Y) = \operatorname{FSym}(Y)$; so π is well defined and onto, and its kernel is as stated. The group E_V is locally finite: a finitely generated subgroup lies in the subgroup of elements (c, a, b) with a, b supported on a fixed finite set $F \subset V$, which has order $2 \cdot 4^{|F|}$. The group $V \rtimes \operatorname{GL}_{\mathrm{fin}}(Y)$ is the directed union of the finite groups $\mathbb{F}_2^D \rtimes \operatorname{GL}(\mathbb{F}_2^D)$ over finite $D \subset Y$, so it is locally finite. An extension of a locally finite group K by a locally finite group is locally finite: a finitely generated subgroup Λ has finite image modulo K, so $\Lambda \cap K$ has finite index in Λ, is therefore finitely generated, and is finite. So $\ker \pi$ is locally finite.

The class of elementary amenable groups is the smallest class containing the finite and the abelian groups and closed under subgroups, quotients, extensions and directed unions [Cho80]. A locally finite group is the directed union of its finite subgroups, so $\ker \pi$ is elementary amenable, and so is the extension G_{P} of $\ker \pi$ by $\mathbb{Z}^2$. Elementary amenable groups are amenable, amenable groups are sofic [Gro99, Wei00], and sofic groups are hyperlinear [ES05]; see [Pes08]. □

The ray swap. Let ι be the permutation of Y that exchanges rays 2 and 3: $\iota(1, j) = (1, j)$, $\iota(2, j) = (3, j)$ and $\iota(3, j) = (2, j)$. It is not in H, but it normalizes H (conjugation by ι exchanges the translation lengths on rays 2 and 3) and $\operatorname{FSym}(Y)$. As a permutation matrix

on V it normalizes $\mathrm{GL}_{\mathrm{fin}}(Y)$, hence L, and by (7) it normalizes the translations, hence A. It does not lie in A: it fixes 0, so it would lie in L (Lemma 6.3(b)), and if $\iota = Mh$ with $M \in \mathrm{GL}_{\mathrm{fin}}(Y)$ and $h \in H$, then $M = \iota h^{-1}$ is a permutation matrix that fixes all but finitely many basis vectors, so $\iota \in \mathrm{FSym}(Y)H = H$. So $A^+ = \langle A, \iota\rangle = A \rtimes \langle\iota\rangle$ is a group of affine maps of V, with $\langle\iota\rangle \cong \mathbb{Z}/2$; it acts on E_V by moving sites, by the formula above, and ι fixes 0. Put

$$G_{\mathrm{P}}^+ = E_V \rtimes A^+ = G_{\mathrm{P}} \rtimes \langle\iota\rangle.$$

Lemma 6.9. *In G_{P}^+, $\iota p\iota = q$, and ι commutes with t, d, x, z and J. So G_{P}^+ is generated by p, ι, t, d, x, z, the element J is central in G_{P}^+, and G_{P}^+ is elementary amenable, hence hyperlinear.*

Proof. Since $\iota = \iota^{-1}$, the element $\iota p\iota$ is the conjugate $\iota p\iota^{-1}$. We compare it with q as permutations of Y, composing from right to left. On ray 1, $\iota p\iota(1, j) = \iota(1, j+1) = (1, j+1)$. On ray 3, $\iota p\iota(3, 1) = \iota p(2, 1) = (1, 1)$ and $\iota p\iota(3, j) = \iota(2, j-1) = (3, j-1)$ for $j \geq 2$. On ray 2, $\iota p\iota(2, j) = \iota p(3, j) = \iota(3, j) = (2, j)$. This is the definition of q. Since ι fixes the points $(1, 1)$ and $(1, 2)$, the identities (7) give $\iota t\iota = t$ and $\iota d\iota = d$. Since ι is linear it fixes the site 0, so it commutes with $x = x_0$ and $z = z_0$, and it fixes J. As J is central in G_{P} (Proposition 6.6) and commutes with ι, it is central in G_{P}^+. By Proposition 6.6 and $q = \iota p\iota$, the elements p, ι, t, d, x, z generate G_{P}^+. Finally G_{P}^+ is an extension of G_{P} by $\mathbb{Z}/2$, so it is elementary amenable, hence hyperlinear, as in Proposition 6.8. □

Finite-dimensional models of G_{P}^+ in which J acts as $-I$ exactly are constructed explicitly in Section 9.

7 Pair areas

The generators x and z of G_{P} act on the Pauli pair at the site 0 only. To reach the pair at a site u we conjugate by a word T_u that carries 0 to u, a *transporter* (Fig. 1). Pairs at different sites commute in G_{P}, but an approximate representation sees a relation only through a finite certificate: by Lemma 2.3, a word of area k with respect to words of defect at most ε has defect at most $k\varepsilon$. Section 8 needs the 2^N pairs at the configurations of the first N points of ray 1 to commute up to small errors, so it needs their commutators to have small area with respect to one fixed finite set of words R_{P}. This section proves that they have area at most $\Lambda_N \leq 2^{(\log N + 23)^2}$ (Theorem 7.8), quasi-polynomial in N, the logarithm of the number 2^N of pairs, and that the transported pairs square to the transported J at linear cost (Lemma 7.11).

The proof is a computation with words in which every step involves at most three points of ray 1. Two lamps are first brought to one translation word; the configuration that separates them is then cleared down to a single point by at most N transvections, with the translation word renormalized after each; what is left is the commutation of the lamps at two neighbouring sites, a word of R_{P}. Each step is a *local filling* (Lemma 7.7), whose cost is governed by the Dehn function δ_H of H_3 at scale $O(N)$: two short words in p, q that agree on the one or two points that matter differ by an element of a point stabilizer, which commutes with the transported generator. This adapts [DM26a, Theorem 5.1], where the lamps are copies of S_3, to Pauli lamps.

7.1 The certificate

We use the notation of Section 6. On the first ray we write i for the point $(1, i)$, e_i for $e_{(1,i)}$, τ_i for the translation τ_{e_i}, and $(i\ j)$ for the transposition of the points i and j; thus $t = \tau_1$, $d = E_{2\leftarrow 1}$ and $s = [p, q] = (1\ 2)$ (Lemma 6.2). By (7) and Definition 6.5, a linear map fixes

the site 0 and so commutes with $x = x_0$ and $z = z_0$, an element $h \in A$ conjugates the pair at v to the pair at $h(v)$, and J is central (Proposition 6.6). We write J also for the word $xzxz$.

Definition 7.1 (The certificate R_{P})**.** Put

$$s = [p, q], \qquad B = psp^{-1}, \qquad t_m = p^{m-1}\, t\, p^{1-m} \quad (m = 1, 2, 3).$$

The *certificate* R_{P} consists of the 27 words of Table 2 in the letters p, q, t, d, x, z: thirteen *host words* in p, q, t, d (families H, Q, S, C, A) and fourteen *lamp words* (families Z, L, P, K). Expanded literally, the words have length at most 30; we take each to be its free reduction, which has length at most 22. In particular R_{P} contains x^2, z^2 and $(xz)^4$. In this section areas are taken with respect to R_{P}, and we write $\mathrm{Area} = \mathrm{Area}_{R_{\mathrm{P}}}$.

	words	meaning	number	longest
H	s^2, $(sps\,p^{-1})^3$, $[s, p^2sp^{-2}]$, $psp^{-1}qs^{-1}q^{-1}$	the presentation P_H of H_3	4	22
Q	t^2	translations are involutions	1	2
S	$[t, sp]$, $[t, sq]$, $[d, Bsp]$, $[d, Bsq]$	stabilizers of prototypes commute with them	4	16
C	$[t_1, t_2]$	translations commute	1	8
A	$d\,t_1 d^{-1}(t_1t_2)^{-1}$, $[d, t_2]$, $[d, t_3]$	how d moves the translations at $1, 2, 3$	3	12
Z	$[\ell, a]$ for $\ell \in \{p, q, d\}$, $a \in \{x, z\}$	linear maps fix the lamp at 0	6	4
L	x^2, z^2, $(xz)^4$	the lamp at 0 is D_4	3	8
P	$[a, tbt^{-1}]$ for $a, b \in \{x, z\}$	the lamps at 0 and e_1 commute	4	8
K	$[J, t]$, with $J = xzxz$	J commutes with t	1	10
	total		27	

Table 2: The certificate R_{P}: thirteen host words in p, q, t, d and fourteen lamp words, written with the abbreviations of Definition 7.1. The last column is the largest length of a word of the family after free reduction.

Lemma 7.2. *Every word of R_{P} represents 1 in G_{P}.*

Proof. By Lemma 6.2(a), $s = (1\ 2)$ and $B = (2\ 3)$. We check the families in turn.

- H These are the relators of P_H, which hold in H (Section 6.1).
- Q t is a translation of the $\mathbb{F}_2$-space V, so $t^2 = 1$.
- S sp and sq fix the point 1, and Bsp and Bsq fix the points 1 and 2; by (7) they commute with $t = \tau_1$, respectively with $d = E_{2\leftarrow 1}$.
- C Translations commute.
- A $t_m = \tau_m$ because p^{m-1} maps the point 1 to m, and $dt_md^{-1} = \tau_{d(e_m)}$ by (7), with $d(e_1) = e_1 + e_2$, $d(e_2) = e_2$ and $d(e_3) = e_3$.
- Z p, q and d are linear, so they fix the site 0 and commute with x and z.
- L x and z are involutions, and $(xz)^4 = J^2 = 1$.
- P tbt^{-1} lies in the pair at the site $t(0) = e_1 \neq 0$, and pairs at distinct sites commute in E_V.
- K J is central. □

The 27 words are listed in the ancillary file `relators_pauli.txt`, and a script checks them independently in a faithful model of G_{P} (Appendix C).

Areas. Area is invariant under free equality, inversion and conjugation, and subadditive under products. If $\mu U \nu$ and $\mu U' \nu$ are words, then $\mu U \nu = (\mu U U'^{-1} \mu^{-1})\, \mu U' \nu$ freely, so replacing a subword U by U' changes the area by at most $\mathrm{Area}(UU'^{-1})$; we call this a *replacement of cost* $\mathrm{Area}(UU'^{-1})$. Finally, $[w, v_1 v_2] = [w, v_1] \cdot v_1 [w, v_2] v_1^{-1}$ and $[w, v^{-1}] = v^{-1} [w, v]^{-1} v$ freely, so for all words w, v and every factorization $v = v_1^{\pm 1} \cdots v_m^{\pm 1}$ into words v_l

$$\mathrm{Area}([w, v]) = \mathrm{Area}([v, w]) \le \sum_{l=1}^{m} \mathrm{Area}([w, v_l]). \tag{9}$$

The family H is the presentation P_H of (6), so a word w in p, q that represents 1 in H has $\mathrm{Area}(w) \le \delta_H(|w|)$. The function δ_H is finite and nondecreasing (Corollary A.5), and it is the only input about H beyond explicit computations with permutations.

7.2 Short words in Houghton's group

All transport takes place among the first n points of ray 1. For $n \ge 1$ let $\Sigma_n \le H$ be the group of permutations supported in $\{1, \dots, n\}$, identified with $\mathrm{Sym}(n)$. Recall that $\sigma_k = p^k s p^{-k}$ is the transposition $(k{+}1 \;\; k{+}2)$ (Lemma 6.2). Lengths of words in p, q are counted with s written out as $pqp^{-1}q^{-1}$.

Lemma 7.3 (Finite permutations)**.**

(a) For $1 \le i < j$ the word $\omega_{i,j} = p^{i-1}(sp)^{j-i-1}(sp^{-1})^{j-i-1} s\, p^{1-i}$ represents $(i\ j)$ and has length $10j - 8i - 8 < 10j$.

(b) Every element of Σ_n that moves $m \ge 2$ points is represented by a word in p, q of length less than $10(m-1)n$.

(c) Let $y_1, \dots, y_l \in \{1, \dots, n\}$ be distinct. Put $\pi_0 = 1$ and, for $m = 1, \dots, l$, $w_m = \pi_{m-1}^{-1}(y_m)$ and $\pi_m = \pi_{m-1}(m\ w_m)$, where $(m\ m) = 1$. Then $\pi[y_1, \dots, y_l] := \pi_l$ maps m to y_m for $m \le l$, and it is a product of at most l transpositions of points of $\{1, \dots, n\}$. The product of the corresponding words ω_{m,w_m}, also written $\pi[y_1, \dots, y_l]$, has length less than $10ln$.

Proof. (a) Conjugating $(i{+}1\ \ j)$ by $\sigma_{i-1} = (i\ \ i{+}1)$ gives $(i\ j)$ for $j > i+1$, so by induction on $j - i$

$$(i\ j) = \sigma_{i-1}\sigma_i \cdots \sigma_{j-3}\ \sigma_{j-2}\ \sigma_{j-3} \cdots \sigma_i \sigma_{i-1}.$$

In this product the powers of p telescope: $\sigma_k \sigma_{k+1} = p^k sps\, p^{-k-1}$ and $\sigma_{k+1}\sigma_k = p^{k+1} s p^{-1} s\, p^{-k}$, and multiplying out gives $\omega_{i,j}$. Its length is $2(i-1)+5(j-i-1)+5(j-i-1)+4 = 10j-8i-8$.

(b) A permutation that moves m points is a product of at most $m-1$ transpositions of those points, each represented by (a) by a word of length less than $10n$.

(c) By induction $\pi_{m-1}(k) = y_k$ for $k < m$. Since $y_m \ne y_k$ for $k < m$, the point $w_m = \pi_{m-1}^{-1}(y_m)$ is not among $1, \dots, m-1$, so $w_m \ge m$ and $(m\ w_m)$ fixes $1, \dots, m-1$. Hence $\pi_m(k) = y_k$ for $k < m$, and $\pi_m(m) = \pi_{m-1}(w_m) = y_m$. Each w_m lies in $\{1, \dots, n\}$, because $\pi_{m-1} \in \Sigma_n$. There are at most l nontrivial factors, each a word of length less than $10n$ by (a). □

That a point stabilizer of H is isomorphic to H is known [dC06, Example 3.6]. The next lemma gives such isomorphisms explicitly on generators, with short words.

Lemma 7.4 (Point stabilizers)**.** *For $k \ge 1$ let $H^{(k)}$ be the pointwise stabilizer in H of the points $1, \dots, k$, put $\zeta_k = \sigma_{k-1} \cdots \sigma_1 \sigma_0$ and $K_k = (\zeta_k p,\ \zeta_k q)$.*

(a) There is an isomorphism $H \to H^{(k)}$ that sends p, q to $\zeta_k p, \zeta_k q$ and the transposition $(i\ j)$ to $(i{+}k \;\; j{+}k)$. In particular K_k generates $H^{(k)}$.

(b) *Every element of* $\Sigma_n \cap H^{(k)}$ *that moves* $m \geq 2$ *points is represented by a word with fewer than* $10(m-1)n$ *letters from* $K_k^{\pm 1}$.

Here $\zeta_1 = s$ *and* $\zeta_2 = Bs$, *so* $K_1 = (sp, sq)$ *and* $K_2 = (Bsp, Bsq)$ *are the pairs of the family* S; *their letters have length at most* 5 *and* 11 *in* p, q.

Proof. (a) Let $\theta_k : Y \to Y \setminus \{1, \dots, k\}$ map $(1, j)$ to $(1, j+k)$ and fix rays 2 and 3. For $g \in H$ let $g^{[k]}$ act as $\theta_k g \theta_k^{-1}$ on the image of θ_k and fix the points $1, \dots, k$. Then $g^{[k]}$ is eventually a translation on each ray, and $g \mapsto g^{[k]}$ is an injective homomorphism $H \to H^{(k)}$. It is onto, because an element $h \in H^{(k)}$ restricts to a permutation of the image of θ_k, and $\theta_k^{-1} h \theta_k$ lies in H. Since θ_k maps the point i to $i+k$, $(i\ j)^{[k]} = (i{+}k\ \ j{+}k)$. The element ζ_k is the cycle $1 \mapsto k{+}1 \mapsto k \mapsto \cdots \mapsto 2 \mapsto 1$ of points of ray 1. Hence $\zeta_k p$ fixes the points $1, \dots, k$, maps $(1, j)$ to $(1, j+1)$ for $j > k$, maps $(2, 1)$ to $(1, k+1)$, and agrees with p on the other points; that is, $\zeta_k p = p^{[k]}$. Likewise $\zeta_k q = q^{[k]}$.

(b) An element $g \in \Sigma_n \cap H^{(k)}$ that moves m points is $g'^{[k]}$ for some $g' \in \Sigma_{n-k}$ that moves m points. Write g' as a word in p, q of length less than $10(m-1)(n-k)$ by Lemma 7.3(b), and substitute $\zeta_k p, \zeta_k q$ for p, q.

Finally $\zeta_1 = \sigma_0 = s$ and $\zeta_2 = \sigma_1 \sigma_0 = Bs$. With $s = pqp^{-1}q^{-1}$ and $B = psp^{-1}$, the words sp, sq have length 5, and Bsp, Bsq have length 11. □

7.3 Transporters and local fillings

Definition 7.5 (Transporters)**.** Let $n \geq 3$. For distinct points $i, j \in \{1, \dots, n\}$ put

$$U_i = \pi[i], \qquad \tau_i = U_i \, t \, U_i^{-1}, \qquad U_{ij} = \pi[j, i], \qquad X_{ij} = U_{ij} \, d \, U_{ij}^{-1},$$

words in p, q, t, d built from the words of Lemma 7.3(c). For $u \subseteq \{1, \dots, n\}$ with elements $i_1 < \cdots < i_m$ put $T_u = \tau_{i_1} \cdots \tau_{i_m}$, with $T_\emptyset$ the empty word. We identify u with the configuration $\sum_{i \in u} e_i \in V$.

Since U_i maps the point 1 to i, the word τ_i represents the translation $\tau_i = \tau_{e_i}$ by (7), and we use the same symbol for both. Since U_{ij} maps 1 to j and 2 to i, the word X_{ij} represents the transvection $E_{i \leftarrow j}$. Translations commute, so T_u represents τ_u, and $T_u x T_u^{-1}$ and $T_u z T_u^{-1}$ represent the Pauli pair x_u, z_u at the site u. By Lemma 7.3, $|U_i| < 10n$ and $|U_{ij}| < 20n$, and U_i is a product of at most one transposition and U_{ij} of at most two.

The elements t and d have *prototypes* $P_t = \{1\}$ and $P_d = \{1, 2\}$: by (7), an element of H that fixes P_f pointwise commutes with f. So two transporting words that agree on the prototype give the same conjugate. The next lemma prices this, and the lemma after it lists the four local identities used in the proof of Theorem 7.8. They are the Pauli versions of [DM26a, Lemmas 4.1 and 4.2]. The transporting word may end with a power of p, because the words of the family A use t_2 and t_3.

Lemma 7.6 (Change of transporter)**.** *Let* $n \geq 3$ *and* $f \in \{t, d\}$. *Let* U *and* W_0 *be words in* $p^{\pm 1}, q^{\pm 1}$ *of length less than* $30n$ *that represent elements of* Σ_n, *such that* $W_0^{-1} U$ *moves at most* 10 *points. Let* $W = W_0 p^j$, *where* $j = 0$ *if* $f = d$ *and* $j \in \{0, 1, 2\}$ *if* $f = t$. *If the permutations represented by* W *and* U *agree on* P_f, *then*

$$\mathrm{Area}\big(W f W^{-1} \cdot U f^{-1} U^{-1}\big) \ \leq \ \mathcal{C}(n) := 2\,\delta_H(1500n) + 130n.$$

Proof. Step 1: a word for the stabilizer element. Let $k = |P_f|$ and let g be the element represented by $W^{-1}U$. It fixes P_f pointwise, so $g \in H^{(k)}$. If $j = 0$, then $g \in \Sigma_n$ moves at most 10 points, and by Lemma 7.4(b) it is represented by a word Z with fewer than $90n$ letters from $K_k^{\pm 1}$. If $j \geq 1$, then $f = t$, $k = 1$, and $g = (sp)^{-j} g'$ with $g' = (sp)^j p^{-j} \cdot W_0^{-1} U$. Here $(sp)p^{-1} = s$ and $(sp)^2 p^{-2} = sB$ lie in Σ_3, so $g' \in \Sigma_n$ moves at most 13 points, and g' fixes

the point 1 because g and sp do. By Lemma 7.4(b), g' is represented by a word Z' with fewer than $120n$ letters from $K_1^{\pm 1}$, and we put $Z = (sp)^{-j} Z'$. In both cases Z has fewer than $122n$ letters from $K_k^{\pm 1}$, each of length at most 11 in p, q.

Step 2: the filling. The word $\Omega = W^{-1} U Z^{-1}$ in p, q represents 1 in H and has length less than $(30n+2) + 30n + 11 \cdot 122n \le 1500n$, so $\mathrm{Area}(\Omega) \le \delta_H(1500n)$ by the family H. For each letter ℓ of K_k the word $[f, \ell]$ is in the family S, so $\mathrm{Area}([f, Z]) < 122n$ by (9). Since $U = W \Omega Z$ freely,

$$W f W^{-1} \cdot U f^{-1} U^{-1} = W \, (f \Omega f^{-1}) \, [f, Z] \, \Omega^{-1} \, W^{-1}$$

freely, and its area is at most $2\delta_H(1500n) + 122n \le \mathcal{C}(n)$. □

Lemma 7.7 (Local fillings)**.** *Let $n \ge 3$ and*

$$L(n) = 10\, \delta_H(1500n) + 700\, n.$$

Let $i, j, k \in \{1, \dots, n\}$ with $i \ne j$, and $a, b \in \{x, z\}$. Each of the following words represents 1 in G_{P} and has area at most $L(n)$.

(a) τ_i^2, of area at most 1, and $[\tau_i, \tau_j]$, of area at most $4\mathcal{C}(n) + 1$.

(b) $X_{ij} \tau_k X_{ij}^{-1} \tau_k^{-1}$ if $k \ne j$, and $X_{ij} \tau_j X_{ij}^{-1} \tau_i^{-1} \tau_j^{-1}$; of area at most $5\mathcal{C}(n) + 1$.

(c) $[X_{ij}, a]$, of area at most $2|U_{ij}| + 1 < 40n + 1$.

(d) $[a, \tau_i b \tau_i^{-1}]$, of area at most $4|U_i| + 1 < 40n + 1$.

Proof. The words represent 1: in (a) and (b) because translations are commuting involutions and $E_{i \leftarrow j} \tau_k E_{i \leftarrow j}^{-1} = \tau_{E_{i \leftarrow j}(e_k)}$, with $E_{i \leftarrow j}(e_k) = e_k$ for $k \ne j$ and $E_{i \leftarrow j}(e_j) = e_i + e_j$; in (c) because $E_{i \leftarrow j}$ is linear; in (d) because $\tau_i b \tau_i^{-1}$ lies in the pair at the site $e_i \ne 0$. Since $5\mathcal{C}(n) + 1 \le L(n)$ and $40n + 1 \le L(n)$, it remains to prove the stated area bounds.

Changes of transporter. In (a) and (b) we replace a subword $U f^{\pm 1} U^{-1}$ by $W f^{\pm 1} W^{-1}$ as in Lemma 7.6; this is a replacement of cost at most $\mathcal{C}(n)$, since the two words compared are conjugate to the word of the lemma or to its inverse. Below, W_0 is a product of at most three transpositions, of length less than $30n$ (Lemma 7.3(c)), and $U \in \{U_i, U_j, U_k, U_{ij}\}$ is a product of at most two, so $W_0^{-1} U$ moves at most 10 points.

(a) The word τ_i^2 is freely $U_i t^2 U_i^{-1}$, a conjugate of the word t^2. For $[\tau_i, \tau_j]$ let $W_0 = \pi[i, j]$, which maps 1 to i and 2 to j. Replace both occurrences of $\tau_i^{\pm 1}$ by $(W_0 t W_0^{-1})^{\pm 1}$, and both of $\tau_j^{\pm 1}$ by $(W_0 t_2 W_0^{-1})^{\pm 1} = (W_0 p\, t\, p^{-1} W_0^{-1})^{\pm 1}$: these are four changes of transporter, since W_0 and U_i map 1 to i, and $W_0 p$ and U_j map 1 to j. What is left is $W_0 [t_1, t_2] W_0^{-1}$, a conjugate of the word of the family C. So the area is at most $4\mathcal{C}(n) + 1$.

(b) Let $W_0 = \pi[j, i, k]$, or $W_0 = \pi[j, i]$ if $k \in \{i, j\}$, so that W_0 maps 1 to j and 2 to i, and let $m \in \{1, 2, 3\}$ be the point with $W_0(m) = k$; thus $m = 1$ exactly when $k = j$, and $m = 2$ exactly when $k = i$. Replace $X_{ij}^{\pm 1}$ by $(W_0 d W_0^{-1})^{\pm 1}$, a change of transporter because W_0 and U_{ij} agree on $P_d = \{1, 2\}$; replace $\tau_k^{\pm 1}$ by $(W_0 t_m W_0^{-1})^{\pm 1}$, where $t_m = p^{m-1} t p^{1-m}$ and $W_0 p^{m-1}$ maps 1 to k; and, when $k = j$, replace $\tau_i^{\pm 1}$ by $(W_0 t_2 W_0^{-1})^{\pm 1}$. If $k \ne j$, four changes turn the word into $W_0 (d t_m d^{-1} t_m^{-1}) W_0^{-1}$ with $m \in \{2, 3\}$, a conjugate of a word of the family A. If $k = j$, five turn it into $W_0 (d t_1 d^{-1} t_2^{-1} t_1^{-1}) W_0^{-1}$, again a conjugate of one. So the area is at most $5\mathcal{C}(n) + 1$.

(c) The word $X_{ij} = U_{ij} d U_{ij}^{-1}$ has $2|U_{ij}| + 1$ letters, each in $\{p, q, d\}^{\pm 1}$, and $[\ell, a]$ is a word of the family Z for $\ell \in \{p, q, d\}$. By (9), $\mathrm{Area}([X_{ij}, a]) \le 2|U_{ij}| + 1 < 40n + 1$.

(d) Freely the word is $U_i \, [U_i^{-1} a U_i,\ t (U_i^{-1} b U_i) t^{-1}] \, U_i^{-1}$. By (9) and the family Z, replacing $U_i^{-1} a^{\pm 1} U_i$ by $a^{\pm 1}$ costs at most $\mathrm{Area}([U_i^{-1}, a]) \le |U_i|$, and likewise for b. Four such replacements leave $U_i [a, t b t^{-1}] U_i^{-1}$, a conjugate of a word of the family P. So the area is at most $4|U_i| + 1 < 40n + 1$. □

7.4 The pair-area theorem

Theorem 7.8 (Pair areas)**.** *Let $N \geq 1$ and $n = \max(3, N)$. For every $u \subseteq \{1, \dots, N\}$ the transporter T_u is a word in p, q, t of length $|T_u| \leq 21Nn$ that represents the translation τ_u, so that $T_u x T_u^{-1}$ and $T_u z T_u^{-1}$ represent the Pauli pair at the site u. For distinct $u, w \subseteq \{1, \dots, N\}$ and $a, b \in \{x, z\}$,*

$$\operatorname{Area}\big([T_u a T_u^{-1},\, T_w b T_w^{-1}]\big) \;\leq\; \Lambda_N := 20N^2 L(n),$$

with $L(n) = 10\delta_H(1500n) + 700n$ as in Lemma 7.7. Moreover, with $C_0 = 4.5 \cdot 10^4$,

$$14000N^2 n \;\leq\; \Lambda_N \;\leq\; C_0 N^2\big(\delta_H(C_0 N) + N\big) \;\leq\; 2^{(\log N + 23)^2}.$$

Overview of the proof. In G_{P} the statement is a short computation. Conjugation by τ_u turns the commutator of the lamps at u and w into that of the lamps at 0 and $u_0 = u + w \neq 0$. Fix a point j of u_0. The transvection $E_{i \leftarrow j}$ toggles the point i in every configuration that contains j, so the product G of the at most $N - 1$ transvections $E_{i \leftarrow j}$ over the other points i of u_0 is a linear map that carries u_0 to e_j and fixes 0. So the commutator is conjugate to that of the lamps at 0 and e_j, and these commute by the family P. The proof follows this computation word by word, and pays for each identity it uses with a local filling (Lemma 7.7). *Step 1* merges $T_u^{-1} T_w$ into the single translation word T_{u_0} by $O(N^2)$ swaps. *Step 2* applies the transvections one at a time, and after each one renormalizes the translation word, so that it never has more than $N + 1$ factors; this costs $O(N)$ local fillings per transvection. *Step 3* moves the lamp letters across G, which is linear, and ends with one word of the family P. Each local filling costs at most $L(n)$, so the total is $O(N^2 L(n))$.

Example 7.9. Let $N = 3$, $u = \{1, 3\}$ and $w = \{2\}$, so that $T_u = \tau_1 \tau_3$ (as in Fig. 1) and $T_w = \tau_2$. Up to conjugation by T_u, the commutator is $[a, Z b Z^{-1}]$ with $Z = T_u^{-1} T_w = \tau_3^{-1} \tau_1^{-1} \tau_2$. In Step 1, two words τ_i^2 turn Z into $\tau_3 \tau_1 \tau_2$, and two swaps sort it into $T_{u_0} = \tau_1 \tau_2 \tau_3$, where $u_0 = \{1, 2, 3\}$. In Step 2 we take the pivot $j = 3$. Conjugation by X_{13}, which represents $E_{1 \leftarrow 3}$, fixes τ_1 and τ_2 and turns τ_3 into $\tau_3 \tau_1$, up to local fillings:

$$X_{13}\, \tau_1 \tau_2 \tau_3\, X_{13}^{-1} \;\rightsquigarrow\; \tau_1 \tau_2 \tau_3 \tau_1 \;\rightsquigarrow\; \tau_1 \tau_1 \tau_2 \tau_3 \;\rightsquigarrow\; \tau_2 \tau_3 = T_{\{2,3\}},$$

by two swaps and one word τ_1^2. Conjugation by X_{23} turns $\tau_2 \tau_3$ into $\tau_2 \tau_3 \tau_2$, then $\tau_2 \tau_2 \tau_3$, then τ_3. So $G = X_{23} X_{13}$ carries the site u_0 to e_3; indeed $E_{2 \leftarrow 3} E_{1 \leftarrow 3}(e_1 + e_2 + e_3) = e_3$. In Step 3 the letters $a^{\pm 1}, b^{\pm 1}$ are moved across G, which represents a linear map, and what is left is $[a, \tau_3 b \tau_3^{-1}]$, a local filling of type (d). Every identity used involves at most three points.

Proof of Theorem 7.8. Step 0: the transporters. Translations commute, so T_u represents τ_u. Each τ_i is a word in p, q, t of length $2|U_i| + 1 < 20n + 1$, so $|T_u| \leq 20Nn \leq 21Nn$. Write $L = L(n)$; by Lemma 7.7 each of its words has area at most L, and $L \geq 700n$.

Step 1: one translation word. Freely $[T_u a T_u^{-1}, T_w b T_w^{-1}] = T_u [a, Z b Z^{-1}] T_u^{-1}$ with $Z = T_u^{-1} T_w$. Replace each of the at most N factors τ_i^{-1} of T_u^{-1} by τ_i, at cost $\operatorname{Area}(\tau_i^{-2}) \leq 1$ each. Sort the resulting product of at most $2N$ factors τ_i into increasing order of i by at most $\binom{2N}{2}$ swaps of adjacent distinct factors, each of cost at most L (Lemma 7.7(a)), and cancel the at most N adjacent equal pairs, at cost 1 each. The result is T_{u_0} with $u_0 = u + w \neq 0$, so $\operatorname{Area}(Z T_{u_0}^{-1}) \leq 2N + N(2N - 1)L \leq 2N^2 L$. Since $Z^{\pm 1}$ occurs four times in $[a, Z b Z^{-1}]$,

$$\operatorname{Area}\big([T_u a T_u^{-1}, T_w b T_w^{-1}]\big) \leq 8N^2 L + \operatorname{Area}\big([a, T_{u_0} b T_{u_0}^{-1}]\big).$$

Step 2: normalized clearing. Let $r \leq N$ be the number of points of u_0, choose one of them, j, as the *pivot*, and list the others as $i_1, \dots, i_{r-1}$. Put $G_l = X_{i_l j} \cdots X_{i_1 j}$ and $u_l = u_0 + e_{i_1} + \cdots + e_{i_l}$, which is u_0 with $i_1, \dots, i_l$ removed; so $u_{r-1} = \{j\}$. We claim

$$\operatorname{Area}\big(G_l T_{u_0} G_l^{-1}\, T_{u_l}^{-1}\big) \leq 2lrL \qquad (0 \leq l \leq r - 1).$$

For $l = 0$ the word is freely trivial. Given the claim for l, put $i = i_{l+1}$ and $X = X_{ij}$. Freely $G_{l+1}T_{u_0}G_{l+1}^{-1} = X(G_lT_{u_0}G_l^{-1})X^{-1}$, which becomes $XT_{u_l}X^{-1}$ at cost at most $2lrL$. Freely $XT_{u_l}X^{-1}$ is the product, in order, of the words $X\tau_{i'}X^{-1}$ over the points i' of u_l. By Lemma 7.7(b), at a total cost of at most rL, each of them becomes $\tau_{i'}$, except that $X\tau_jX^{-1}$ becomes $\tau_j\tau_i$. The new factor τ_i has a partner τ_i in T_{u_l}, since i is a point of u_l; at most $r-1$ swaps make the two adjacent, and τ_i^2 cancels them at cost 1. What remains is $T_{u_{l+1}}$, and this step costs at most $rL + (r-1)L + 1 \le 2rL$, which proves the claim. The translation word is back in normal form after every transvection and never has more than $r+1$ factors.

Step 3: a single point. Let $G = G_{r-1}$ and $\Theta = GT_{u_0}G^{-1}\tau_j^{-1}$, of area at most $2r(r-1)L \le 2N^2L$ by Step 2. Freely $T_{u_0} = G^{-1}\Theta\tau_jG$. Substituting this into the four occurrences of $T_{u_0}^{\pm1}$ in $[a, T_{u_0}bT_{u_0}^{-1}]$ and deleting the four occurrences of $\Theta^{\pm1}$ costs at most $8N^2L$, and leaves

$$a\,G^{-1}\tau_jG\,b\,G^{-1}\tau_j^{-1}G\,a^{-1}\,G^{-1}\tau_jG\,b^{-1}\,G^{-1}\tau_j^{-1}G.$$

The word G is a product of $r-1$ words X_{ij}, so $\mathrm{Area}([G,c]) \le (r-1)(40n+1)$ for $c \in \{x, z\}$ by (9) and Lemma 7.7(c), and likewise for $[G, c^{-1}] = c^{-1}[G,c]^{-1}c$. Replacing $Gb^{\pm1}G^{-1}$ by $b^{\pm1}$, and then, after conjugating by G, $Ga^{\pm1}G^{-1}$ by $a^{\pm1}$, costs at most $4(r-1)(40n+1)$ and leaves a conjugate of $[a, \tau_jb\tau_j^{-1}]$, of area at most $40n+1$ by Lemma 7.7(d). Altogether, since $4N(40n+1) + 40n + 1 \le 205Nn \le N^2L$,

$$\mathrm{Area}\big([a, T_{u_0}bT_{u_0}^{-1}]\big) \le 8N^2L + 4N(40n+1) + 40n + 1 \le 9N^2L,$$

and with Step 1 the area of the commutator is at most $17N^2L \le \Lambda_N$.

Step 4: the size of Λ_N. Since $L(n) \ge 700n$, $\Lambda_N \ge 14000N^2n$. Since $n \le 3N$ and δ_H is nondecreasing, $L(n) \le 10\delta_H(4500N) + 2100N$, so $\Lambda_N \le 200N^2\delta_H(4500N) + 42000N^3 \le C_0N^2(\delta_H(C_0N) + N)$. By (H1), $\delta_H(C_0N) \le 2^{(\log(C_0N)+6)^2}$, and also $N \le 2^{(\log(C_0N)+6)^2}$; so with $\ell = \log N \ge 0$ and $\log C_0 < 15.46$,

$$\log\Lambda_N \le \log(2C_0) + 2\ell + (\ell + \log C_0 + 6)^2 \le 16.46 + 2\ell + (\ell + 21.46)^2 \le \ell^2 + 45\ell + 477 \le (\ell+23)^2.$$

□

Remark 7.10 (Growth of Λ_N). By Theorem 7.8, $14000N^2n \le \Lambda_N \le 2^{(\log N+23)^2}$ for all $N \ge 1$. The factor N^2 counts local moves: at most N transvections, each followed by at most N swaps. The Dehn function of H_3 enters only through $L(n)$, at the scale $n = \max(3, N)$ of the transport. So all $\binom{2^N}{2}$ pairs of lamps on N points commute at area $O(N^2\delta_H(C_0N))$, which is $2^{O(\log^2 N)}$ by (H1). Under (H2), $\delta_H(m) \le C_Hm^{p_H}$ with $p_H \ge 1$, and then

$$\Lambda_N \le C_0N^2\big(C_H(C_0N)^{p_H} + N\big) \le C'N^{2+p_H} = C'N^{\kappa} \qquad (N \ge 1), \qquad C' = C_0\big(C_HC_0^{p_H} + 1\big),$$

polynomial in N.

7.5 Transporting the central involution

The lower bounds compare the image of J with the products $x_uz_ux_uz_u$ at the various sites. Transporting J is cheap, and the transported Pauli pair squares to the transported J with no relation at all.

Lemma 7.11 (Transporting J). *Let $N \ge 1$, $n = \max(3, N)$ and $u \subseteq \{1, \dots, N\}$. Then*

$$\mathrm{Area}\big(T_uJT_u^{-1}J^{-1}\big) \le 4|T_u| \le 84Nn,$$

and $(T_uxT_u^{-1}\,T_uzT_u^{-1})^2 = T_uJT_u^{-1}$ freely.

Proof. The word $T_uJT_u^{-1}J^{-1}$ is $[T_u, J]$, and the letters of T_u are $p^{\pm1}$, $q^{\pm1}$ and $t^{\pm1}$. For $g \in \{p, q\}$, (9) applied to the four letters of $J = xzxz$ and the family Z give $\mathrm{Area}([g, J]) \le 4$. For the letter t, $[t, J] = [J, t]^{-1}$ is the inverse of the word of the family K. By (9) applied to the letters of T_u, $\mathrm{Area}([T_u, J]) \le 4|T_u|$, and $|T_u| \le 21Nn$ by Theorem 7.8. The second claim holds because $T_u^{-1}T_u$ cancels freely, so the word is freely $T_u(xz)^2T_u^{-1} = T_uJT_u^{-1}$. □

8 Lamps force entropy: the sector theorem

This section proves the lower bound of Theorem B, and the group-theoretic half of the lower bound of Theorem A. Its main result, Theorem 8.1, is stated for an arbitrary state: if an assignment of unitaries to p, q, t, d, x, z satisfies the certificate R_{P} to precision e in a state ρ, nearly commutes with $\rho^{1/2}$, and keeps J visibly away from I in that state, then ρ has entropy at least $e^{-2+o(1)}$. Section 10 feeds into it the models that Theorem 5.1 extracts from strategies, so the theorem bounds the entanglement entropy of every near-optimal strategy. For the flat state it is the lower bound for the hyperlinear profile of J (Corollary 8.4). Throughout, areas are taken with respect to R_{P} (Definition 7.1), J also denotes the word $xzxz$, and for $0 < \varepsilon < \frac{1}{2}$ we put

$$\beta(\varepsilon) = 2^{-4(24+\log\log(1/\varepsilon))^2}. \tag{10}$$

Since $\log(1/\beta(\varepsilon)) = O((\log\log(1/\varepsilon))^2)$, we have $\beta(\varepsilon) = \varepsilon^{o(1)}$ as $\varepsilon \to 0$, and β is increasing on $(0, \frac{1}{2})$.

The picture. On the first N points of ray 1 there are 2^N configurations $u \subseteq \{1, \dots, N\}$, and the transporter T_u of Definition 7.5 carries the Pauli pair x, z at the configuration 0 to the pair $T_u x T_u^{-1}, T_u z T_u^{-1}$ at u. In G_{P} these 2^N pairs satisfy the Pauli relations: pairs at different configurations commute, and each pair squares to the same central involution, $(T_u x T_u^{-1}\, T_u z T_u^{-1})^2 = T_u J T_u^{-1} = J$. On the sector where J acts as -1, an exact representation therefore carries 2^N anticommuting pairs, that is, 2^N qubits. In an approximate model the transported lamps become 2^N approximate Pauli pairs measured in the sector state, and Corollary 4.2 counts them: they still force about 2^N bits of entropy, as long as 2^N times the square of their errors stays below a constant. The errors are priced by the pair areas of Theorem 7.8. A commutator of two transported lamps has area at most Λ_N, and in a fixed state a filling of area Λ costs its area times the length of its conjugators, which is again of order Λ (Lemmas 3.3 and 3.4); so each error is of order $\Lambda_N^2 e/\Delta$, where Δ measures how far J is from I. The proof balances 2^N against Λ_N^4; with $\Lambda_N \le 2^{(\log N + 23)^2}$ the best N has $2^N \approx \Delta^2 e^{-2} \beta(e)$.

Theorem 8.1 (The sector theorem)**.** *Let $\phi : F(p, q, t, d, x, z) \to \mathrm{U}(D)$ be a homomorphism and ρ a density matrix on $\mathbb{C}^D$, with $\xi = \rho^{1/2}$. Suppose that, for some $e > 0$ and $0 < \Delta \le 2$,*

(i) $\|(\phi(r) - I)\xi\|_2 \le e$ *for every* $r \in R_{\mathrm{P}}$*;*

(ii) $\|[\phi(s), \xi]\|_2 \le e$ *for every letter* $s \in \{p, q, t, d, x, z\}$*;*

(iii) $\|(\phi(J) - I)\xi\|_2 \ge \Delta$*.*

If $e \le 2^{-1071}\Delta$, then

$$\log \operatorname{rank} \rho \;\ge\; c_1\, \Delta^2 e^{-2}\, \beta(e), \qquad c_1 = \frac{1}{4.1 \cdot 10^7}.$$

Moreover, let $J' = \mathrm{r}(\phi(J))$ be the rounding of Lemma 3.6, $P_\pm = \frac{1}{2}(I \pm J')$, $p_- = \operatorname{Tr}(P_- \rho)$, and let $\rho_- = P_- \rho P_- / p_-$ be the sector state of Lemma 3.10. Then $p_- \ge \Delta^2/16$, $\|[J', \xi]\|_2 \le 6e$, $S(\rho_-) \ge c_1 \Delta^2 e^{-2} \beta(e)$, and

$$S(\rho) \;\ge\; \frac{c_1}{16}\, \Delta^4 e^{-2}\, \beta(e) - 1.$$

For the flat state $\rho = I/D$ the state commutators vanish, condition (ii) holds trivially, and the theorem bounds $\log D$; this is Corollary 8.4. Nothing about the state is used beyond (i)–(iii), so the theorem applies to any state in which a model is accurate.

The proof follows the picture. For a scale N we round the transported lamps to involutions (Step 1), derive their pairwise relations from the pair areas and their relation to J from Lemma 7.11 (Step 2), pass to the sector state where the rounded J is -1 (Step 3), and apply

the count (Step 4); Step 5 chooses N. Every estimate is measured in the state, and conjugators are paid for through state commutators.

Proof. Fix $N \geq 1$, to be chosen in Step 5, and put $n = \max(3, N)$, $M = 2^N$, $l = 21Nn \geq 63$ and $\Lambda = \Lambda_N$. For $u \subseteq \{1, \dots, N\}$ the transporter T_u has $|T_u| \leq l$ (Theorem 7.8). By Remark 7.10, $\Lambda \geq 14000N^2 n = \frac{2000N}{3}\, l$, so

$$l \leq \Lambda/666, \qquad l^2 \leq \Lambda^2/443556. \tag{11}$$

Every word of R_{P} has length at most 30 (in fact at most 22, Definition 7.1), so Lemma 3.4 applies with $K = 30$.

Step 1: rounding. Let $A_u = \phi(T_u x T_u^{-1})$ and $B_u = \phi(T_u z T_u^{-1})$. These words have length at most $2l + 1$, so their state commutators are at most $(2l + 1)e$ by Lemma 3.3(a). Since $A_u^2 = \phi(T_u)\phi(x^2)\phi(T_u)^{-1}$ with $x^2 \in R_{\mathrm{P}}$ and $\|[\phi(T_u), \xi]\|_2 \leq le$, Lemma 3.3(c) gives $\|(A_u^2 - I)\xi\|_2 \leq e + 2le \leq (2l + 1)e$, and likewise for B_u with $z^2 \in R_{\mathrm{P}}$. Put

$$X_u = \mathrm{r}(A_u), \qquad Z_u = \mathrm{r}(B_u), \qquad J' = \mathrm{r}(\phi(J)).$$

By Lemma 3.6, $\|(X_u - A_u)\xi\|_2 \leq (2l + 1)e$ and $\|[X_u, \xi]\|_2 \leq 3(2l + 1)e \leq 7le$, and the same holds for Z_u and B_u. We write

$$\gamma = 7le, \qquad \text{so that} \qquad \|[X_u, \xi]\|_2,\ \|[Z_u, \xi]\|_2 \leq \gamma, \qquad \gamma \geq 441e.$$

Since $\phi(J)^2 = \phi((xz)^4)$ and $(xz)^4 \in R_{\mathrm{P}}$, we have $\|(\phi(J)^2 - I)\xi\|_2 \leq e$, so Lemma 3.6 gives $\|(J' - \phi(J))\xi\|_2 \leq e$ and, as $J = xzxz$ has length 4, $\|[J', \xi]\|_2 \leq 4e + 2e = 6e$.

Step 2: relations. Different sites. Let $u \neq w$, $a, b \in \{x, z\}$, and $c = [T_u a T_u^{-1}, T_w b T_w^{-1}]$. This word has length at most $4(2l+1) \leq 9l$ and area at most Λ (Theorem 7.8). By Lemma 3.4 it is freely a product of at most Λ conjugates of words of $R_{\mathrm{P}}^{\pm 1}$ with conjugators of length at most $9l + 30\Lambda$, so by Lemma 3.3(d)

$$\|(\phi(c) - I)\xi\|_2 \ \leq\ \Lambda\big(1 + 2(9l + 30\Lambda)\big)e = \Lambda(60\Lambda + 18l + 1)e \ \leq\ 61\Lambda^2 e,$$

because $18l + 1 \leq \Lambda$ by (11). Let R_u^a be X_u if $a = x$ and Z_u if $a = z$. We replace the four factors of $\phi(c) = Y_1 Y_2 Y_1^{-1} Y_2^{-1}$, $Y_1 = \phi(T_u a T_u^{-1})$, $Y_2 = \phi(T_w b T_w^{-1})$, by the involutions $R_u^a, R_w^b, R_u^a, R_w^b$. Each factor moves by at most 3γ in the state-weighted norm: by at most $(2l + 1)e \leq \gamma$ for Y_1, Y_2, and for the inverted factors, writing $Y^{-1} - R = Y^{-1}(R - Y)R$, by at most $\|(R - Y)\xi\|_2 + 2\|[R, \xi]\|_2 \leq 3\gamma$. The replacements have state commutators at most γ, so Lemma 3.3(e) bounds the total cost by $4 \cdot 3\gamma + 2(3 + 2 + 1)\gamma = 24\gamma$. Since $[R_u^a, R_w^b] = (R_u^a R_w^b R_u^a R_w^b - I)R_w^b R_u^a$ and moving $R_w^b R_u^a$ to the right of ξ costs at most $2 \cdot 2\gamma$,

$$\|[Z_u, Z_w]\xi\|_2, \quad \|[X_u, Z_w]\xi\|_2 \ \leq\ 61\Lambda^2 e + 28\gamma \qquad (u \neq w). \tag{12}$$

One site. Let $K_u = X_u Z_u X_u Z_u$. By Lemma 7.11, $A_u B_u A_u B_u = \phi(T_u J T_u^{-1}) = \phi(j_u)\phi(J)$ with $j_u = T_u J T_u^{-1} J^{-1}$, a word of length at most $2l + 8$ and area at most $4|T_u| \leq 4l$. By Lemmas 3.3 and 3.4(d),

$$\|(\phi(j_u) - I)\xi\|_2 \leq 4l\big(1 + 2(2l + 8 + 30 \cdot 4l)\big)e = 4l(244l + 17)e \leq 984\, l^2 e.$$

Replacing X_u, Z_u by A_u, B_u in K_u costs at most $4(2l+1)e + 2(3+2+1)(2l+1)e = 16(2l+1)e \leq 34le$ by Lemma 3.3(e), with the state commutators of A_u, B_u. Moving $\phi(J)$ to the right of ξ costs at most $2 \cdot 4e$, and replacing $\phi(J)$ by J' costs at most e. Hence, for every u,

$$\|(K_u - J')\xi\|_2 \ \leq\ 34le + 984l^2 e + 8e + e \ \leq\ \gamma_0 := 1000\, l^2 e.$$

As X_u and Z_u are involutions, $X_u K_u X_u = Z_u K_u Z_u = K_u^*$, and $K_u^* - J' = K_u^*(J' - K_u)J'$, so $\|(K_u^* - J')\xi\|_2 \leq \gamma_0 + 2\|[J', \xi]\|_2 \leq \gamma_0 + 12e$. Writing $X_u J' X_u - J' = X_u(J' - K_u)X_u + (K_u^* - J')$,

we get $\|(X_uJ'X_u - J')\xi\|_2 \le (\gamma_0 + 2\gamma) + (\gamma_0 + 12e)$, and since $[X_u, J'] = (X_uJ'X_u - J')X_u$, moving X_u to the right of ξ adds 2γ. With the same argument for Z_u,

$$\|[X_u, J']\xi\|_2, \quad \|[Z_u, J']\xi\|_2 \;\le\; 2\gamma_0 + 4\gamma + 12e. \tag{13}$$

Finally $\{X_u, Z_u\} = (K_u + I)Z_uX_u$. Since $(J' + I)P_- = 0$,

$$\{X_u, Z_u\}P_- = (K_u - J')Z_uX_uP_- + (J' + I)[Z_uX_u, P_-].$$

Applied to ξ, the first term has norm at most $\gamma_0 + 2 \cdot 2\gamma + 2\|[P_-, \xi]\|_2 \le \gamma_0 + 4\gamma + 6e$, as $\|[P_-, \xi]\|_2 = \frac{1}{2}\|[J', \xi]\|_2 \le 3e$. The second has norm at most $2\|[Z_uX_u, P_-]\xi\|_2 = \|[Z_uX_u, J']\xi\|_2$, and $[Z_uX_u, J'] = Z_u[X_u, J'] + [Z_u, J']X_u$ gives the bound $2(2\gamma_0 + 4\gamma + 12e) + 2\gamma$ by (13). As $30e \le \gamma$,

$$\|\{X_u, Z_u\}P_-\xi\|_2 \;\le\; 5\gamma_0 + 14\gamma + 30e \;\le\; 5\gamma_0 + 15\gamma. \tag{14}$$

Step 3: the sector. We have $p_- = \mathrm{Tr}(P_-\rho) = \|P_-\xi\|_2^2 = \frac{1}{4}\|(I - J')\xi\|_2^2$, and $\|(I - J')\xi\|_2 \ge \|(I - \phi(J))\xi\|_2 - \|(\phi(J) - J')\xi\|_2 \ge \Delta - e \ge \Delta/2$, so $p_- \ge \Delta^2/16$. Let ρ_- and $\xi_- = \rho_-^{1/2}$ be as in Lemma 3.10; by its part (a), $\|P_+\xi P_-\|_2 \le \frac{1}{2}\|[J', \xi]\|_2 \le 3e$. We bound the five quantities that Corollary 4.2 needs, in the state ρ_-.

- For an operator R with $\|R\|_{\mathrm{op}} \le 2$, Lemma 3.10(b) gives

  $$\|R\xi_-\|_2 \le (\|R\xi\|_2 + \|R\|_{\mathrm{op}}\|[J', \xi]\|_2)/\sqrt{p_-} \le (\|R\xi\|_2 + 12e)/\sqrt{p_-}.$$

  With (12), $\|[Z_u, Z_w]\xi_-\|_2$ and $\|[X_u, Z_w]\xi_-\|_2$ are at most $(61\Lambda^2 e + 28\gamma + 12e)/\sqrt{p_-}$.
- By the first bound of Lemma 3.10(b) and (14),

  $$\|\{X_u, Z_u\}\xi_-\|_2 \le \frac{\|\{X_u, Z_u\}P_-\xi\|_2 + 2\|P_+\xi P_-\|_2}{\sqrt{p_-}} \le \frac{5\gamma_0 + 15\gamma + 6e}{\sqrt{p_-}}.$$

- By Lemma 3.10(c) for the unitaries X_u, Z_u, with (13),

  $$\begin{aligned}\|[X_u, \xi_-]\|_2 &\le \frac{\|[X_u, J']\xi\|_2 + 3\|[X_u, \xi]\|_2 + \|[J', \xi]\|_2}{\sqrt{p_-}} \\ &\le \frac{(2\gamma_0 + 4\gamma + 12e) + 3\gamma + 6e}{\sqrt{p_-}} = \frac{2\gamma_0 + 7\gamma + 18e}{\sqrt{p_-}},\end{aligned}$$

  and the same bound holds for $\|[Z_u, \xi_-]\|_2$.

By (11), $\gamma = 7le \le 0.011\,\Lambda e$ and $\gamma_0 = 1000l^2e \le 0.0023\,\Lambda^2 e$, and $\Lambda \ge 42000$. So $61\Lambda^2 e + 28\gamma + 12e \le (61\Lambda^2 + 0.31\Lambda + 12)e$, $5\gamma_0 + 15\gamma + 6e \le (0.012\Lambda^2 + 0.17\Lambda + 6)e$ and $2\gamma_0 + 7\gamma + 18e \le (0.005\Lambda^2 + 0.08\Lambda + 18)e$, and all five quantities are at most

$$\eta := \frac{63\Lambda^2 e}{\sqrt{p_-}} \;\le\; \frac{252\Lambda^2 e}{\Delta}.$$

Step 4: counting. Apply Corollary 4.2 on $\mathbb{C}^D$ to the state ρ_- and the $M \ge 2$ pairs of involutions (X_u, Z_u), $u \subseteq \{1, \dots, N\}$, ordered in any way; all its hypotheses hold with the uniform bound η. If

$$2^N\Lambda_N^4 \;\le\; \frac{\Delta^2}{1.02 \cdot 10^7\, e^2}, \tag{15}$$

then $16M\eta^2 \le 16 \cdot 252^2 \cdot 2^N\Lambda^4 e^2/\Delta^2 \le 16 \cdot 252^2/(1.02 \cdot 10^7) < 0.0997 < \frac{1}{10}$, so Corollary 4.2 gives $S(\rho_-) \ge \frac{3}{4}M \ge 2^{N-1}$. By Lemma 4.9, $\log \mathrm{rank}\,\rho \ge \log \mathrm{rank}\,\rho_- \ge S(\rho_-)$.

Step 5: the choice of N. By Remark 7.10, $\log \Lambda_N \le (\log N + 23)^2$. Put

$$Q = \log \frac{\Delta^2}{1.02 \cdot 10^7\, e^2}, \qquad g(N) = N + 4(\log N + 23)^2,$$

so that g is increasing, $g(1) = 2117$, and (15) holds whenever $g(N) \le Q$. For $e \le 2^{-1071}\Delta$ we have $Q \ge 2142 - \log(1.02 \cdot 10^7) > 2118$, since $\log(1.02 \cdot 10^7) < 23.3$, and $e \le \Delta/2$. Let N be the largest integer with $g(N) \le Q$; then $N \ge 1$. By maximality $g(N+1) > Q$, that is, $N > Q - 1 - 4(\log(N+1) + 23)^2$. Also $N + 1 < g(N) \le Q < 2\log(1/e)$, because $\Delta \le 2$ gives $Q \le 2 + 2\log(1/e) - 23$; so $\log(N+1) < 1 + \log\log(1/e)$. Therefore

$$2^{N-1} \;>\; 2^{\,Q-2-4(\log(N+1)+23)^2} \;\ge\; \frac{\Delta^2}{4.08 \cdot 10^7\, e^2}\, 2^{-4(24+\log\log(1/e))^2} \;\ge\; c_1 \Delta^2 e^{-2} \beta(e),$$

and with Step 4 this proves the bounds for $\log \operatorname{rank} \rho$ and $S(\rho_-)$. We showed $p_- \ge \Delta^2/16$ and $\|[J', \xi]\|_2 \le 6e$ in Steps 3 and 1. Finally, by Lemma 4.9 and $h \le 1$,

$$S(\rho) \;\ge\; p_-\, S(\rho_-) - h\big(\min(\|[J', \xi]\|_2^2/4, \tfrac{1}{2})\big) \;\ge\; \frac{\Delta^2}{16}\, S(\rho_-) - 1. \qquad \square$$

Remark 8.2 (Where the exponent 4 comes from). The pair area enters (15) to the fourth power because it enters the errors squared: a filling of area Λ is charged its area times the length of its conjugators, which is again of order Λ (Lemma 3.4). This is why β has the factor 4 in its exponent. In the flat state conjugation costs nothing, the words c of Step 2 have defect at most Λe (Lemma 2.3), and the same argument, with Λ_N^2 in place of Λ_N^4, gives the factor $2^{-2(24+\log\log(1/e))^2}$ for $\log D$ (with other constants). Both factors are $e^{o(1)}$, and we do not use the sharper flat one. The exponent 2 of Δ/e is the same in both cases.

Remark 8.3 (Under the polynomial Dehn bound). Assume (H2). Then the pair areas satisfy $\Lambda_N \le C' N^\kappa$ for all $N \ge 1$, with $\kappa = 6 + \log 107$ and an absolute constant $C' \ge 1$ (Remark 7.10). In Step 5 choose instead N maximal with $2^N C'^4 N^{4\kappa} \le \Delta^2/(1.02 \cdot 10^7 e^2)$; such $N \ge 1$ exists when $e \le e_0\Delta$ with $e_0 = 1/(5000\, C'^2)$, and then (15) holds. Maximality gives $2^{N-1} > \Delta^2/(4.08 \cdot 10^7 C'^4 e^2 (N+1)^{4\kappa})$, and $N + 1 \le 3\log(1/e)$ because $2^N < e^{-2}$. So, for $e \le e_0\Delta$,

$$\log \operatorname{rank} \rho,\; S(\rho_-) \;\ge\; c\, \Delta^2 e^{-2} (\log(1/e))^{-4\kappa}, \qquad c = \frac{1}{4.08 \cdot 10^7 \cdot 81^\kappa C'^4},$$

and the factor $\beta(e)$ becomes a power of $\log(1/e)$, here and in everything deduced from Theorem 8.1.

8.1 The flat profile of J

For the flat state, Theorem 8.1 is a statement about ε-representations. We record it for every finite set of relators that contains the certificate, and, with weaker constants, for every finite set of relators in which the certificate has bounded area. The latter covers finite presentations of G_{P} and of G_{P}^+ on any finite generating set and any word representing J (Remark 6.7).

Corollary 8.4 (Lower bound for the profile of J)**.**

(a) Let R be a finite set of words in p, q, t, d, x, z that represent 1 in G_{P}, with $R_{\mathrm{P}} \subseteq R$. For $0 < \delta \le 2$ and $0 < \varepsilon \le 2^{-1071}\delta$,

$$\log \operatorname{hlp}_R(J; \delta, \varepsilon) \;\ge\; c_1\, \delta^2 \varepsilon^{-2}\, \beta(\varepsilon).$$

(b) Let S be a finite alphabet, $R \subset F(S)$ finite and $w \in F(S)$. Suppose there are words $\hat{s} \in F(S)$ for $s \in \{p, q, t, d, x, z\}$ and an integer $k \geq 1$ such that, for every $r \in R_{\mathrm{P}}$, the word $\hat{r}$ obtained from r by substituting $\hat{s}$ for s has area at most k in R, and so has $w^{-1}(\hat{x}\hat{z})^2$. Then for $0 < \delta \leq 2$ and $0 < \varepsilon \leq 2^{-1072}\delta/k$,

$$\log \mathrm{hlp}_R(w; \delta, \varepsilon) \;\geq\; \frac{c_1}{4k^2}\, \delta^2 \varepsilon^{-2}\, \beta(\varepsilon).$$

Under (H2) the factor $\beta(\varepsilon)$ improves to $(\log(1/\varepsilon))^{-4\kappa}$, with other constants (Remark 8.3).

Proof. For the flat state $\rho = I/D$ we have $\xi = D^{-1/2}I$, so $\|Y\xi\|_2 = \|Y\|_f$ for every Y, every state commutator vanishes, and $\operatorname{rank}\rho = D$.

(a) Let ϕ be an ε-representation of $\langle p, q, t, d, x, z \mid R\rangle$ in $\mathrm{U}(D)$ with $\|\phi(J) - I\|_f \geq \delta$. Conditions (i)–(iii) of Theorem 8.1 hold for the flat state with $e = \varepsilon$ and $\Delta = \delta$, so $\log D \geq c_1\delta^2\varepsilon^{-2}\beta(\varepsilon)$.

(b) Let ϕ be an ε-representation of $\langle S \mid R\rangle$ in $\mathrm{U}(D)$ with $\|\phi(w) - I\|_f \geq \delta$, and let ϕ' be the homomorphism on $F(p, q, t, d, x, z)$ with $\phi'(s) = \phi(\hat{s})$. By Lemma 2.3, $\|\phi'(r) - I\|_f = \|\phi(\hat{r}) - I\|_f \leq k\varepsilon$ for $r \in R_{\mathrm{P}}$, and $\|\phi'(J) - \phi(w)\|_f = \|\phi(w^{-1}(\hat{x}\hat{z})^2) - I\|_f \leq k\varepsilon$, so $\|\phi'(J) - I\|_f \geq \delta - k\varepsilon \geq \delta/2$. Apply Theorem 8.1 to ϕ' and the flat state with $e = k\varepsilon$ and $\Delta = \delta/2$; its hypothesis $e \leq 2^{-1071}\Delta$ is our assumption on ε. Since β is increasing and $k\varepsilon \geq \varepsilon$, $\log D \geq c_1(\delta/2)^2(k\varepsilon)^{-2}\beta(k\varepsilon) \geq \frac{c_1}{4k^2}\delta^2\varepsilon^{-2}\beta(\varepsilon)$. □

When $\langle S \mid R\rangle$ presents G_{P} (or G_{P}^+) and w represents J, the hypothesis of (b) holds for any words $\hat{s}$ that represent p, q, t, d, x, z, because the words $\hat{r}$ and $w^{-1}(\hat{x}\hat{z})^2$ represent 1 (Lemma 7.2) and so have finite area; this is the argument of Lemma 2.4. The matching upper bound, and the state-weighted version, are Theorem 9.6.

9 Hash models: the upper bound

This section proves the upper bound of Theorem B: G_{P}^+ has ε-representations in which J acts as $-I$ exactly, in dimension $\exp(C\varepsilon^{-2}\log(1/\varepsilon))$ (Proposition 9.4 and Theorem 9.5). Together with Theorem 8.1 this determines the profile of J, flat and state-weighted, up to $\varepsilon^{o(1)}$ in the exponent (Theorem 9.6). In Section 10 the same models, applied to the words of the presentation $\mathcal{P}$, give the upper bound of Theorem A.

The idea. Since G_{P}^+ is amenable (Lemma 6.9), Følner sets give such models, but they record every lamp of a large part of V, so their dimension is exponential in the number of configurations they see; for S_3 lamps they are doubly exponential [DM26a, Proposition 7.3]. A model of a short word needs only the lamps that the word lights. We keep these on a register of qubits through a random affine hash η of the configurations: the lamp at the site v acts as the Pauli pair on the qubit $\eta(v)$. Affine maps act on hashes by precomposition, so the lighter moves the hash covariantly. Two lit lamps collide only if η identifies their sites, which happens with probability about $\ell^2 2^{-k}$ for k hash bits. The quotient $\mathbb{Z}^2$ of G_{P} is handled by a clock, a box of side $2m+1$, which a word of length ℓ leaves only from a boundary layer of relative size about ℓ/m. A fraction ν of bad blocks costs a defect $\sqrt{4\nu}$ in the normalized norm, so we take $m \approx \varepsilon^{-2}$ and $2^k \approx \varepsilon^{-2}$. The hashes are affine maps on a space of dimension about $6m$, so there are about 2^{6mk} of them, whence the dimension $\exp(C\varepsilon^{-2}\log(1/\varepsilon))$. Since $J = (xz)^2$ is made of the Pauli pair at the site 0, which no hash can split, J acts as $-I$ on every block. For S_3 lamps and permutation models, the analogous construction is [DM26a, Lemmas 6.1, 6.2 and Proposition 6.3]; here the register holds qubits, on which all lamp centres act by the same scalar -1, the clock also carries ι, and errors are measured in the normalized Hilbert–Schmidt norm.

9.1 Pauli registers and affine hashing

Let W be a finite-dimensional vector space over $\mathbb{F}_2$, whose elements we call sites. As in Section 6, let E_W be the set $\mathbb{F}_2 \times \mathbb{F}_2^W \times \mathbb{F}_2^W$ with the product (8): the central product of Pauli lamps x_w, z_w at the sites $w \in W$, with all centres identified with $J = (1,0,0)$. The *support* of $\lambda = (s,a,b) \in E_W$ is $\operatorname{supp}\lambda = \operatorname{supp} a \cup \operatorname{supp} b \subseteq W$, and for $\Omega \subseteq W$ we write E_Ω for the subgroup of elements supported in Ω. The affine group $\mathrm{AGL}(W)$ acts on E_W by moving sites, $g \cdot (s,a,b) = (s, a\circ g^{-1}, b \circ g^{-1})$, and $\mathcal{K}_W = E_W \rtimes \mathrm{AGL}(W)$ has the product $(\lambda, g)(\mu, h) = (\lambda \cdot (g \cdot \mu), gh)$.

The register. Fix $k \geq 1$ and let $\mathcal{H}_k = (\mathbb{C}^2)^{\otimes \mathbb{F}_2^k}$, a register of 2^k qubits indexed by $\mathbb{F}_2^k$. For $\alpha \in \mathbb{F}_2^{\mathbb{F}_2^k}$ let X^α be the tensor product of σ_X on the qubits c with $\alpha(c) = 1$ and of I on the others, and define Z^α in the same way with σ_Z. Since $\sigma_Z\sigma_X = -\sigma_X\sigma_Z$ and operators on different qubits commute,

$$Z^{\alpha'}X^{\alpha} = (-1)^{\langle \alpha', \alpha\rangle} X^\alpha Z^{\alpha'}, \qquad \langle \alpha', \alpha\rangle = \sum_{c \in \mathbb{F}_2^k} \alpha'(c)\alpha(c), \tag{16}$$

and $X^\alpha X^{\alpha'} = X^{\alpha+\alpha'}$, $Z^\alpha Z^{\alpha'} = Z^{\alpha+\alpha'}$. We write X_c and Z_c when α is the indicator function of $\{c\}$.

Hashes. Let $\mathrm{Aff}(W, \mathbb{F}_2^k)$ be the set of affine maps $W \to \mathbb{F}_2^k$; it has $2^{k(\dim W + 1)}$ elements. For $\eta \in \mathrm{Aff}(W, \mathbb{F}_2^k)$ and $a \in \mathbb{F}_2^W$ let $\eta_* a \in \mathbb{F}_2^{\mathbb{F}_2^k}$ be the push-forward, $(\eta_* a)(c) = \sum_{w:\, \eta(w) = c} a(w)$; it is additive in a. Put

$$\chi_\eta(s,a,b) = (-1)^s\, X^{\eta_* a} Z^{\eta_* b} \qquad \big((s,a,b) \in E_W\big),$$

a unitary on $\mathcal{H}_k$. Thus χ_η puts the lamp at the site w on the qubit $\eta(w)$. When two lit sites share a hash, χ_η need not be multiplicative.

Lemma 9.1 (Pauli hashing)**.** *Let $\eta \in \mathrm{Aff}(W, \mathbb{F}_2^k)$.*

(a) $\chi_\eta(J) = -I$, $\chi_\eta(x_w) = X_{\eta(w)}$ *and* $\chi_\eta(z_w) = Z_{\eta(w)}$ *for* $w \in W$.

(b) $\chi_{\eta\circ g}(\lambda) = \chi_\eta(g \cdot \lambda)$ *for* $g \in \mathrm{AGL}(W)$ *and* $\lambda \in E_W$.

(c) If η is injective on $\Omega \subseteq W$, then χ_η restricts to a homomorphism $E_\Omega \to \mathrm{U}(\mathcal{H}_k)$.

(d) If $\Omega \subseteq W$ has at most n elements and η is uniformly distributed on $\mathrm{Aff}(W, \mathbb{F}_2^k)$, then η is injective on Ω except with probability at most $\binom{n}{2}2^{-k}$.

Proof. (a) is the definition. (b) For $c \in \mathbb{F}_2^k$, substituting $u = g(w)$, $((\eta \circ g)_* a)(c) = \sum_{w:\, \eta(g(w)) = c} a(w) = \sum_{u:\, \eta(u) = c} a(g^{-1}(u)) = (\eta_*(a \circ g^{-1}))(c)$; the same holds for b, and the sign is unchanged.

(c) Let $\lambda = (s,a,b)$ and $\lambda' = (s',a',b')$ lie in E_Ω. By (16) and additivity of η_*, respectively by (8),

$$\chi_\eta(\lambda)\chi_\eta(\lambda') = (-1)^{s+s'+\langle \eta_* b, \eta_* a'\rangle} X^{\eta_*(a+a')} Z^{\eta_*(b+b')},$$
$$\chi_\eta(\lambda\lambda') = (-1)^{s+s'+\langle b, a'\rangle} X^{\eta_*(a+a')} Z^{\eta_*(b+b')}.$$

Expanding, $\langle \eta_* b, \eta_* a'\rangle = \sum b(w)a'(w')$ over the pairs (w,w') with $\eta(w) = \eta(w')$. A pair contributes only if $w \in \operatorname{supp} b$ and $w' \in \operatorname{supp} a'$, so both lie in Ω; since η is injective on Ω, only the pairs with $w = w'$ contribute, and the sum is $\langle b, a'\rangle$.

(d) Write $\eta = \eta_0 + c_0$ with η_0 linear and $c_0 \in \mathbb{F}_2^k$; then η_0 is uniform on the linear maps. For $w \neq w'$ we have $\eta(w) = \eta(w')$ if and only if $\eta_0(w - w') = 0$. Since $w - w' \neq 0$ extends to a basis of W, the linear map $\eta_0 \mapsto \eta_0(w - w')$ is onto $\mathbb{F}_2^k$, so $\eta_0(w - w')$ is uniform on $\mathbb{F}_2^k$ and vanishes with probability 2^{-k}. Sum over the $\binom{n}{2}$ pairs. □

The hash action. For $\gamma = (\lambda, g) \in \mathcal{K}_W$ define a unitary Θ_γ on $\ell^2(\mathrm{Aff}(W, \mathbb{F}_2^k)) \otimes \mathcal{H}_k$ by

$$\Theta_\gamma(\delta_\eta \otimes v) = \delta_{\eta \circ g^{-1}} \otimes \chi_{\eta \circ g^{-1}}(\lambda)\, v.$$

It is unitary because $\eta \mapsto \eta \circ g^{-1}$ is a bijection of $\mathrm{Aff}(W, \mathbb{F}_2^k)$ and each $\chi_\eta(\lambda)$ is unitary. We call the subspaces $\delta_\eta \otimes \mathcal{H}_k$ the *blocks*; each Θ_γ maps every block onto a block.

Lemma 9.2 (Products)**.** *Let $\gamma_i = (\lambda_i, g_i) \in \mathcal{K}_W$ for $i = 1, \dots, L$, put $h_i = g_1 \cdots g_{i-1}$ (so $h_1 = 1$), $g = g_1 \cdots g_L$ and $\Omega = \bigcup_i h_i(\operatorname{supp} \lambda_i)$. Then $\Theta_{\gamma_1} \cdots \Theta_{\gamma_L}$ maps the block of η onto the block of $\eta \circ g^{-1}$, and it agrees with $\Theta_{\gamma_1 \cdots \gamma_L}$ on the block of every η for which $\eta \circ g^{-1}$ is injective on Ω. In particular, if $|\operatorname{supp} \lambda| \le 1$, then $\Theta_{(\lambda,g)}^{-1} = \Theta_{(\lambda,g)^{-1}}$.*

Proof. Put $\eta' = \eta \circ g^{-1}$. The factor Θ_{γ_i} is applied to the block of $\eta \circ (g_{i+1} \cdots g_L)^{-1}$ and produces the hash $\eta \circ (g_i \cdots g_L)^{-1} = \eta' \circ h_i$. So, by Lemma 9.1(b),

$$\Theta_{\gamma_1} \cdots \Theta_{\gamma_L}(\delta_\eta \otimes v) = \delta_{\eta'} \otimes \chi_{\eta' \circ h_1}(\lambda_1) \cdots \chi_{\eta' \circ h_L}(\lambda_L)\, v = \delta_{\eta'} \otimes \chi_{\eta'}(h_1 \cdot \lambda_1) \cdots \chi_{\eta'}(h_L \cdot \lambda_L)\, v.$$

On the other hand $\gamma_1 \cdots \gamma_L = \big((h_1 \cdot \lambda_1) \cdots (h_L \cdot \lambda_L),\, g\big)$, by induction on L from the product rule of $\mathcal{K}_W$, so $\Theta_{\gamma_1 \cdots \gamma_L}(\delta_\eta \otimes v) = \delta_{\eta'} \otimes \chi_{\eta'}\big((h_1 \cdot \lambda_1) \cdots (h_L \cdot \lambda_L)\big)v$. The elements $h_i \cdot \lambda_i$ are supported in $h_i(\operatorname{supp} \lambda_i) \subseteq \Omega$, so if η' is injective on Ω the two expressions agree by Lemma 9.1(c). For the last claim apply this to $\gamma_1 = \gamma = (\lambda, g)$ and $\gamma_2 = \gamma^{-1} = (g^{-1} \cdot \lambda^{-1}, g^{-1})$: here $g_1 g_2 = 1$ and $\Omega = \operatorname{supp} \lambda$ has at most one element, so every hash is injective on it, and $\Theta_\gamma \Theta_{\gamma^{-1}} = \Theta_1 = I$. □

9.2 The clock

Let $\mathcal{Q} = \mathbb{Z}^2 \rtimes \mathbb{Z}/2$, where the generator of $\mathbb{Z}/2$ exchanges the two coordinates. We write its elements as (n, j) with $n = (n_1, n_2) \in \mathbb{Z}^2$ and $j \in \{0, 1\}$, and $\varsigma(n_1, n_2) = (n_2, n_1)$, so that $(n, j)(n', j') = (n + \varsigma^j n', j + j')$. Let $\pi : G_{\mathrm{P}} \to \mathbb{Z}^2$ be the homomorphism of Proposition 6.8, the composite $G_{\mathrm{P}} \to A \to L \to L/\operatorname{GL}_{\mathrm{fin}}(Y) \cong H/\operatorname{FSym}(Y) \cong \mathbb{Z}^2$ with the last map (5), and let $\pi^+ : G_{\mathrm{P}}^+ \to \mathcal{Q}$ be $\pi^+(g\iota^j) = (\pi(g), j)$ for $g \in G_{\mathrm{P}}$. It is a homomorphism because $\pi(\iota g \iota) = \varsigma\pi(g)$ for $g \in G_{\mathrm{P}}$, which it suffices to check on generators: by Lemma 6.9 and (5), $\iota p \iota = q$, $\iota q \iota = p$, $\pi(p) = (1, 0)$ and $\pi(q) = (0, 1)$, while t, d, x, z commute with ι and lie in $\ker \pi$. The kernel of π^+ is $\ker \pi = E_V \rtimes (V \rtimes \operatorname{GL}_{\mathrm{fin}}(Y))$, a locally finite group (Proposition 6.8). For $\mathsf{c} = (n, j) \in \mathcal{Q}$ and $g \in G_{\mathrm{P}}^+$ put

$$g_{\mathsf{c}} = p^{n_1} q^{n_2} \iota^j, \qquad \zeta(g, \mathsf{c}) = g^{-1}_{\pi^+(g)\mathsf{c}}\, g\, g_{\mathsf{c}} \ \in \ker \pi^+.$$

Then $\pi^+(g_{\mathsf{c}}) = \mathsf{c}$ and $\zeta(1, \mathsf{c}) = 1$, and ζ is a cocycle:

$$\zeta(gh, \mathsf{c}) = \zeta\big(g, \pi^+(h)\mathsf{c}\big)\, \zeta(h, \mathsf{c}), \qquad \text{in particular} \qquad \zeta(g^{-1}, \mathsf{c}) = \zeta\big(g, \pi^+(g)^{-1}\mathsf{c}\big)^{-1}. \tag{17}$$

For $r \ge 1$ let $Y_r \subset Y$ be the set of the $3r$ points of height at most r, and $W_r = \mathbb{F}_2^{Y_r} \subseteq V$, the configurations supported in Y_r. Let $\mathcal{K}_r \le \ker \pi^+$ be the set of elements whose lamps lie at sites in W_r and whose affine part $v \mapsto Mv + u$ has $u \in W_r$, $M(W_r) = W_r$ and $Me_y = e_y$ for $y \notin Y_r$. It is a finite subgroup, and restriction to W_r is an isomorphism onto $\mathcal{K}_{W_r}$: the product $(\lambda, g)(\mu, h) = (\lambda \cdot (g \cdot \mu), gh)$ involves g only on W_r, every affine map of W_r extends by the identity on the other basis vectors, and the kernel is trivial. For $m \ge 1$ let $B_m = \{-m, \dots, m\}^2$ and $F_m = B_m \times \{0, 1\} \subset \mathcal{Q}$.

Lemma 9.3 (The clock)**.** *Let $m \ge 1$, $r = 2m + 3$, $\mathsf{c} \in F_m$ and $s \in \{p, q, \iota, t, d, x, z\}^{\pm 1}$ with $\pi^+(s)\mathsf{c} \in F_m$. Then $\zeta(s, \mathsf{c}) \in \mathcal{K}_r$; its lamp part is trivial if $s \notin \{x, z\}^{\pm 1}$, and $\zeta(x^{\pm 1}, \mathsf{c}) = x$, $\zeta(z^{\pm 1}, \mathsf{c}) = z$. Moreover $\pi^+(s)F_m = F_m$ for $s \notin \{p, q\}^{\pm 1}$.*

Proof. We use three facts about the permutations p, q, ι of Y. Each of $p^{\pm1}, q^{\pm1}$ changes the height of a point by at most one, and ι preserves heights. On the points of height at least 2, $p^{\pm1}$ and $q^{\pm1}$ act as translations along the three rays, by $\pm(1,-1,0)$ and $\pm(1,0,-1)$, and these translations commute. Consequently, for integers a, b the commutator $[p^a, q^b]$ fixes every point of height at least $|a|+|b|+2$: while its letters are applied, such a point is moved on each ray by at most $|a|+|b|$ steps away from its starting height, so it stays at height at least 2, every letter acts as a translation, and the translations add up to zero. So $[p^a, q^b]$, and likewise $[q^a, p^b]$, is a finitary permutation of Y supported in heights at most $|a|+|b|+1$. A finitary permutation of Y supported in Y_r acts on V by a permutation matrix that preserves W_r and fixes e_y for $y \notin Y_r$, so it lies in $\mathcal{K}_r$, without lamps.

Let $\mathsf{c} = (n, j)$, so $|n_1|, |n_2| \le m$. Recall $\iota^{-1} = \iota$ and $\iota p^a q^b = q^a p^b \iota$.

$s = p^{\pm1}$. Then $\pi^+(s)\mathsf{c} = ((n_1 \pm 1, n_2), j)$ and $\zeta(s, \mathsf{c}) = \iota^j q^{-n_2} p^{-n_1 \mp 1}\, p^{\pm1}\, p^{n_1} q^{n_2} \iota^j = 1$.

$s = q^{\pm1}$. Then $\pi^+(s)\mathsf{c} = ((n_1, n_2 \pm 1), j)$ and

$$\zeta(s, \mathsf{c}) = \iota^j q^{-n_2 \mp 1} p^{-n_1} q^{\pm1} p^{n_1} q^{n_2} \iota^j = \iota^j\, q^{-(n_2 \pm 1)}\, [p^{-n_1}, q^{\pm1}]\, q^{n_2 \pm 1}\, \iota^j.$$

The commutator is supported in heights at most $|n_1| + 2$; conjugating by $q^{-(n_2\pm1)}$ moves its support by at most $|n_2 \pm 1| \le m$ in height, because $\pi^+(s)\mathsf{c} \in F_m$, and ι preserves heights. So $\zeta(s, \mathsf{c})$ is a finitary permutation supported in heights at most $2m + 2 < r$.

$s = \iota^{\pm1}$. As an element of G_{P}^+, $s = \iota$, and $\pi^+(\iota)\mathsf{c} = (\varsigma n, j+1)$. Using $\iota p^{n_1} q^{n_2} = q^{n_1} p^{n_2} \iota$,

$$\zeta(\iota, \mathsf{c}) = \iota^{j+1} q^{-n_1} p^{-n_2}\, \iota\, p^{n_1} q^{n_2} \iota^j = \iota^{j+1}\, [q^{-n_1}, p^{-n_2}]\, \iota^{j+1},$$

a finitary permutation supported in heights at most $2m + 1$.

$s \in \{t, d, x, z\}^{\pm1}$. Here $\pi^+(s)$ is trivial, $s^{-1} = s$ in G_{P}^+, and $\zeta(s, \mathsf{c}) = h\, s\, h^{-1}$ with $h = g_{\mathsf{c}}^{-1} = \iota^j q^{-n_2} p^{-n_1}$, a permutation of Y that moves the points $(1,1)$ and $(1,2)$ to points $y_1 = h(1,1)$ and $y_2 = h(1,2)$ of height at most $2m + 2$. By (7), which holds for every permutation of Y, $hth^{-1} = \tau_{e_{y_1}}$ and $hdh^{-1} = E_{y_2 \leftarrow y_1}$, which lie in $\mathcal{K}_r$ without lamps. The permutation h is linear on V, so it fixes the site 0, and $hxh^{-1} = x$, $hzh^{-1} = z$.

The last claim holds because $\pi^+(s)$ is trivial for $s \in \{t, d, x, z\}^{\pm1}$, $\pi^+(\iota)(n, j) = (\varsigma n, j+1)$, and ς preserves B_m. □

9.3 Hash models

Proposition 9.4 (Pauli hash models)**.** *Let $\mathcal{W}$ be a finite set of words in the letters p, q, ι, t, d, x, z, each of length at most $\ell \ge 1$ and each representing 1 in G_{P}^+. For every $0 < \varepsilon \le 1$ there are a dimension D with*

$$\log D \;\le\; 336\, \ell\, \varepsilon^{-2} \log(2\ell/\varepsilon) + 8\, \ell^2 \varepsilon^{-2} + 4 \log(2\ell/\varepsilon) + 8 \;\le\; 400\, \ell^2 \varepsilon^{-2} \log(2\ell/\varepsilon)$$

and a homomorphism $\phi : F(p, q, \iota, t, d, x, z) \to \mathrm{U}(D)$ such that $\|\phi(w) - I\|_f \le \varepsilon$ for every $w \in \mathcal{W}$, $\phi(x)^2 = \phi(z)^2 = I$, and

$$\phi\big((xz)^2\big) = -I \qquad \textit{exactly.}$$

The model is a direct sum of blocks indexed by a clock position c and a hash η; each block is a register of qubits. A letter moves the clock by π^+ and acts on the fibre by the hash action of the cocycle ζ. A word that represents 1 acts as the identity on every block whose clock position is far from the boundary of the box and whose hash separates the at most ℓ sites that the word lights; the other blocks are few.

Proof. Parameters and space. Let $m = \lceil 16 \ell \varepsilon^{-2} \rceil$, $r = 2m + 3$ and $W = W_r$, so $\dim W = 3r = 6m + 9$, and let $k = \lceil \log(4\ell^2 \varepsilon^{-2}) \rceil$. Write $\mathcal{A} = \mathrm{Aff}(W, \mathbb{F}_2^k)$, and let

$$\mathcal{H} = \ell^2(F_m) \otimes \ell^2(\mathcal{A}) \otimes \mathcal{H}_k.$$

The *blocks* of $\mathcal{H}$ are the subspaces $\delta_{\mathsf{c}} \otimes \delta_\eta \otimes \mathcal{H}_k$, $(\mathsf{c}, \eta) \in F_m \times \mathcal{A}$; each has dimension 2^{2^k}. Through the identification of $\mathcal{K}_r$ with $\mathcal{K}_W$, every $\gamma \in \mathcal{K}_r$ defines the unitary Θ_γ of Section 9.1 on $\ell^2(\mathcal{A}) \otimes \mathcal{H}_k$.

The generators. Let $s \in \{p, q, \iota, t, d, x, z\}$ and $F_s = F_m \cap \pi^+(s)^{-1} F_m$. For $\mathsf{c} \in F_s$ and $v \in \ell^2(\mathcal{A}) \otimes \mathcal{H}_k$ put

$$\phi(s)(\delta_{\mathsf{c}} \otimes v) = \delta_{\pi^+(s)\mathsf{c}} \otimes \Theta_{\zeta(s,\mathsf{c})} v,$$

which makes sense by Lemma 9.3. The sets $F_m \setminus F_s$ and $F_m \setminus \pi^+(s) F_m$ have the same size; fix a bijection b_s between them and put $\phi(s)(\delta_{\mathsf{c}} \otimes v) = \delta_{b_s(\mathsf{c})} \otimes v$ for $\mathsf{c} \in F_m \setminus F_s$. Then $\phi(s)$ permutes the clock basis (δ_{c}) and acts unitarily on each fibre, so it is unitary, and it maps blocks onto blocks. By Lemma 9.3, $F_s = F_m$, and no b_s is needed, unless $s \in \{p, q\}$. Extend ϕ to $F(p, q, \iota, t, d, x, z)$.

Step 1: J is exact. By Lemma 9.3, $\zeta(x, \mathsf{c}) = x = (x_0, 1)$ for every $\mathsf{c} \in F_m$, so by Lemma 9.1(a) $\phi(x)(\delta_{\mathsf{c}} \otimes \delta_\eta \otimes v) = \delta_{\mathsf{c}} \otimes \delta_\eta \otimes X_{\eta(0)} v$, and likewise $\phi(z)$ acts by $Z_{\eta(0)}$. Hence $\phi(x)^2 = \phi(z)^2 = I$, and on every block $\phi((xz)^2)$ acts by $(X_{\eta(0)} Z_{\eta(0)})^2 = -I$, since $(\sigma_X \sigma_Z)^2 = -I$. So $\phi((xz)^2) = -I$.

Step 2: letters and inverse letters. Let $s \in \{p, q, \iota, t, d, x, z\}^{\pm 1}$ and $\mathsf{c} \in F_m$ with $\pi^+(s)\mathsf{c} \in F_m$. We claim

$$\phi(s)(\delta_{\mathsf{c}} \otimes v) = \delta_{\pi^+(s)\mathsf{c}} \otimes \Theta_{\zeta(s,\mathsf{c})} v. \tag{18}$$

For a letter this is the definition. For an inverse letter $s = u^{-1}$ let $\mathsf{c}' = \pi^+(u)^{-1}\mathsf{c}$; it lies in F_m and $\pi^+(u)\mathsf{c}' = \mathsf{c} \in F_m$, so $\mathsf{c}' \in F_u$. Since $\phi(u)$ is unitary and maps the fibre over c' onto the fibre over c by $\Theta_{\zeta(u,\mathsf{c}')}$, we get $\phi(u)^{-1}(\delta_{\mathsf{c}} \otimes v) = \delta_{\mathsf{c}'} \otimes \Theta^{-1}_{\zeta(u,\mathsf{c}')} v$. The lamp part of $\zeta(u, \mathsf{c}')$ is supported in at most one site (Lemma 9.3), so $\Theta^{-1}_{\zeta(u,\mathsf{c}')} = \Theta_{\zeta(u,\mathsf{c}')^{-1}}$ by Lemma 9.2, and $\zeta(u, \mathsf{c}')^{-1} = \zeta(u^{-1}, \mathsf{c})$ by (17).

Step 3: a word on a good block. Let $w = s_1 \cdots s_L \in \mathcal{W}$, with $s_i \in \{p, q, \iota, t, d, x, z\}^{\pm 1}$ and $L \le \ell$. For $\mathsf{c} \in F_m$ put $\mathsf{c}_{L+1} = \mathsf{c}$ and $\mathsf{c}_i = \pi^+(s_i)\mathsf{c}_{i+1}$. Each letter changes the $\mathbb{Z}^2$-coordinate of the clock by a unit vector, by the swap ς, or not at all, so it changes its sup norm by at most one. Hence if $\mathsf{c} = (n, j)$ with $n \in B_{m-\ell}$, all c_i lie in F_m; call such c *interior*. For interior c, (18) and induction give

$$\phi(w)(\delta_{\mathsf{c}} \otimes v) = \delta_{\mathsf{c}_1} \otimes \Theta_{\gamma_1} \cdots \Theta_{\gamma_L} v, \qquad \gamma_i = \zeta(s_i, \mathsf{c}_{i+1}) \in \mathcal{K}_r.$$

Since w represents 1 in G^+_{P}, we have $\mathsf{c}_1 = \pi^+(w)\mathsf{c} = \mathsf{c}$, and by (17) $\gamma_1 \cdots \gamma_L = \zeta(w, \mathsf{c}) = g_{\mathsf{c}}^{-1} g_{\mathsf{c}} = 1$. Write $\gamma_i = (\lambda_i, g_i)$; then $g_1 \cdots g_L = 1$. Each λ_i is trivial or the lamp x_0 or z_0 (Lemma 9.3), so the set $\Omega_{\mathsf{c}} = \bigcup_i h_i(\operatorname{supp} \lambda_i)$ of Lemma 9.2 consists of points $h_i(0)$, one for each letter $x^{\pm 1}, z^{\pm 1}$ of w, and has at most ℓ elements. By Lemma 9.2, $\phi(w)$ acts as the identity on the block of (c, η) whenever c is interior and η is injective on Ω_{c}.

Step 4: the defect. Call a block bad if $\phi(w)$ does not act on it as the identity. The fraction of non-interior $\mathsf{c} \in F_m$ is $1 - \left(\frac{2m-2\ell+1}{2m+1}\right)^2 \le \frac{4\ell}{2m+1} < \frac{2\ell}{m}$, and for each interior c the fraction of η that are not injective on Ω_{c} is at most $\binom{\ell}{2} 2^{-k} < \ell^2 2^{-k-1}$ by Lemma 9.1(d). So the fraction ν of bad blocks satisfies $\nu < 2\ell/m + \ell^2 2^{-k-1} \le \varepsilon^2/8 + \varepsilon^2/8 = \varepsilon^2/4$, by the choice of m and k. The operator $\phi(w)$ maps blocks onto blocks. On a block b that is not bad, $(\phi(w) - I)|_{\mathsf{b}}$ vanishes, while on a bad block $\|(\phi(w) - I)v\| \le 2\|v\|$, so $\|(\phi(w) - I)|_{\mathsf{b}}\|_2 \le 2(\dim \mathsf{b})^{1/2}$. Summing over the blocks, which all have the same dimension,

$$\|\phi(w) - I\|_f^2 = \frac{1}{\dim \mathcal{H}} \sum_{\mathsf{b}} \|(\phi(w) - I)|_{\mathsf{b}}\|_2^2 \le 4\nu \le \varepsilon^2.$$

Step 5: the dimension. We have $\dim \mathcal{H} = |F_m| \cdot |\mathcal{A}| \cdot 2^{2^k} = 2(2m+1)^2 \cdot 2^{k(6m+10)} \cdot 2^{2^k}$. Put $\Lambda_\varepsilon = \log(2\ell/\varepsilon) \ge 1$. Since $m \le 17\ell\varepsilon^{-2}$, we have $6m + 10 \le 112\ell\varepsilon^{-2}$ and $2m + 1 \le 35\ell\varepsilon^{-2} \le 9(2\ell/\varepsilon)^2$; and $2^k < 8\ell^2\varepsilon^{-2}$, $k < 1 + 2\Lambda_\varepsilon \le 3\Lambda_\varepsilon$. Hence

$$\log \dim \mathcal{H} \le 1 + 2(\log 9 + 2\Lambda_\varepsilon) + 336\, \ell \Lambda_\varepsilon \varepsilon^{-2} + 8\ell^2\varepsilon^{-2},$$

which is the first bound, since $1 + 2\log 9 < 8$. As $\ell, \Lambda_\varepsilon, \varepsilon^{-1} \geq 1$, it is at most $(8 + 4 + 336 + 8)\,\ell^2\Lambda_\varepsilon\varepsilon^{-2} \leq 400\,\ell^2\Lambda_\varepsilon\varepsilon^{-2}$. Take $D = \dim \mathcal{H}$. □

The main term of $\log D$ is $k \dim W \leq 336\,\ell\varepsilon^{-2}\log(2\ell/\varepsilon)$, the logarithm of the number of hashes; ℓ^2 enters only through the register, as $8\ell^2\varepsilon^{-2}$ without a logarithm. The factor $\log(1/\varepsilon)$ comes from the $k \approx 2\log(1/\varepsilon)$ bits of the covariant hash family.

9.4 The profile of J

Theorem 9.5 (Upper bound for the profile)**.** *Let $S' \subseteq \{p, q, \iota, t, d, x, z\}$ contain x and z, and let R be a finite set of words in S', each representing 1 in G_{P}^+, whose free reductions have length at most $\ell \geq 1$. Write $\mathrm{hlp} = \mathrm{hlp}_R$ for the profile of $\langle S' \mid R\rangle$ and $J = (xz)^2$.*

(a) For $0 < \varepsilon \leq 1$, $\mathrm{hlp}(J; 2, \varepsilon) \leq 2^{\,400\,\ell^2\varepsilon^{-2}\log(2\ell/\varepsilon)}$, realized by an ε-representation with $\phi(J) = -I$.

(b) For $0 < \delta \leq 2$ and $0 < \varepsilon \leq \delta/\sqrt{2}$,

$$\mathrm{hlp}(J; \delta, \varepsilon) \;\leq\; \frac{4}{\delta^2} \cdot 2^{\,200\,\ell^2\,\delta^2\varepsilon^{-2}\log(3\ell/\varepsilon)}.$$

This applies to $R = R_{\mathrm{P}}$ with $\ell = 22$ (Definition 7.1 and Lemma 7.2), and to every finite set of relators of G_{P} on p, q, t, d, x, z or of G_{P}^+ on p, ι, t, d, x, z or p, q, ι, t, d, x, z.

Proof. (a) Since $\phi(r)$ depends only on the free reduction of r, we may replace each word of R by its free reduction. Apply Proposition 9.4 to $\mathcal{W} = R$ and restrict ϕ to $F(S')$. The defects are at most ε, and $\|\phi(J) - I\|_f = \|-2I\|_f = 2$.

(b) Let $N = \lfloor 4/\delta^2 \rfloor \geq 1$ and $\varepsilon_1 = \sqrt{2}\,\varepsilon/\delta \leq 1$. Let ϕ_1 be the ε_1-representation of (a), of dimension D_1 with $\log D_1 \leq 400\ell^2\varepsilon_1^{-2}\log(2\ell/\varepsilon_1)$, and let $\phi = \phi_1 \oplus \mathbf{1}$, where $\mathbf{1}$ sends every generator to the identity of $\mathbb{C}^{(N-1)D_1}$. Every word is sent to I by $\mathbf{1}$, so $\|\phi(u) - I\|_f^2 = \frac{1}{N}\|\phi_1(u) - I\|_f^2$ for every word u. Hence $\|\phi(J) - I\|_f^2 = 4/N \geq \delta^2$, and, since $N \geq 2/\delta^2$ (as $\lfloor y \rfloor \geq y/2$ for $y \geq 1$), every relator has defect at most $\varepsilon_1/\sqrt{N} \leq \varepsilon_1\delta/\sqrt{2} = \varepsilon$. The dimension is $ND_1 \leq (4/\delta^2)D_1$, and $400\ell^2\varepsilon_1^{-2} = 200\ell^2\delta^2\varepsilon^{-2}$, $\log(2\ell/\varepsilon_1) = \log(\sqrt{2}\,\ell\delta/\varepsilon) \leq \log(3\ell/\varepsilon)$.

The words of R_{P} represent 1 in $G_{\mathrm{P}} \leq G_{\mathrm{P}}^+$ (Lemma 7.2), and their free reductions have length at most 22 (Definition 7.1); relators of G_{P} or G_{P}^+ represent 1 in G_{P}^+. □

We can now state and prove Theorem B. Recall that $\mathrm{hlp}_R^\rho(J; e)$ is the least dimension of a state-weighted model (ϕ, ρ) of $\langle p, q, t, d, x, z \mid R\rangle$ satisfying (S1)–(S3) of Definition 2.5 for $w = J$ (Definition 2.6).

Theorem 9.6 (The profile of J; Theorem B)**.** *Let R be a finite set of words in p, q, t, d, x, z that represent 1 in G_{P}, with $R_{\mathrm{P}} \subseteq R$, let $\ell \geq 22$ bound the lengths of their free reductions, and put $C_\ell = 400\,\ell^2(2 + \log \ell)$.*

(a) For $0 < \delta \leq 2$ and $0 < \varepsilon \leq 2^{-1071}\delta$,

$$c_1\,\delta^2\varepsilon^{-2}\,\beta(\varepsilon) \;\leq\; \log \mathrm{hlp}_R(J; \delta, \varepsilon) \;\leq\; C_\ell\,\delta^2\varepsilon^{-2}\log(1/\varepsilon).$$

(b) For $0 < e \leq 2^{-1071}$,

$$3c_1\,e^{-2}\,\beta(e) \;\leq\; \log \mathrm{hlp}_R^\rho(J; e) \;\leq\; \log \mathrm{hlp}_R(J; 2, e) \;\leq\; C_\ell\,e^{-2}\log(1/e).$$

Here $c_1 = 1/(4.1 \cdot 10^7)$ and $\beta(\varepsilon) = 2^{-4(24+\log\log(1/\varepsilon))^2} = \varepsilon^{o(1)}$. In particular $\log \mathrm{hlp}_R(J; \delta, \varepsilon) = \varepsilon^{-2+o(1)}$ for every fixed $\delta \in (0, 2]$, and $\log \mathrm{hlp}_R^\rho(J; e) = e^{-2+o(1)}$.

Proof. Lower bounds. The lower bound in (a) is Corollary 8.4(a). For (b), let (ϕ, ρ) be a model on $\mathbb{C}^d$ satisfying (S1)–(S3) at precision e. Then (i) and (ii) of Theorem 8.1 hold, because $R_{\mathrm{P}} \subseteq R$, and by (S3)

$$\|(\phi(J) - I)\xi\|_2 \;\geq\; \|2\xi\|_2 - \|(\phi(J) + I)\xi\|_2 \;\geq\; 2 - e =: \Delta,$$

so (iii) holds with $1 \leq \Delta \leq 2$. The hypothesis $e \leq 2^{-1071}\Delta$ holds because $e \leq 2^{-1071}$ and $\Delta \geq 1$. So $\log d \geq \log \operatorname{rank} \rho \geq c_1 \Delta^2 e^{-2} \beta(e) \geq 3 c_1 e^{-2} \beta(e)$, as $\Delta^2 \geq (2 - 2^{-1071})^2 \geq 3$.

Comparison. An e-representation ϕ with $\|\phi(J) - I\|_f \geq 2$ has $\phi(J) = -I$ by (1); with the flat state $\rho = I/d$ it satisfies (S1)–(S3) at precision e, since $\|Y\xi\|_2 = \|Y\|_f$ and all state commutators vanish. So $\operatorname{hlp}_R^{\rho}(J; e) \leq \operatorname{hlp}_R(J; 2, e)$.

Upper bounds. By Theorem 9.5(a), $\log \operatorname{hlp}_R(J; 2, e) \leq 400 \ell^2 e^{-2} \log(2\ell/e)$, and $\log(2\ell/e) \leq (1 + \log(2\ell)) \log(1/e)$ because $\log(1/e) \geq 1$; this gives (b). For (a), Theorem 9.5(b) applies since $\varepsilon \leq \delta/\sqrt{2}$, and gives $\log \operatorname{hlp}_R(J; \delta, \varepsilon) \leq \log(4/\delta^2) + 200 \ell^2 \delta^2 \varepsilon^{-2} \log(3\ell/\varepsilon)$. Here $\log(3\ell/\varepsilon) \leq (1 + \log(3\ell)) \log(1/\varepsilon)$, and, since $\delta \geq 2^{1071}\varepsilon$, $4/\delta^2 < \varepsilon^{-2}$, so $\log(4/\delta^2) < 2 \log(1/\varepsilon) \leq \delta^2 \varepsilon^{-2} \log(1/\varepsilon)$. Altogether $\log \operatorname{hlp}_R(J; \delta, \varepsilon) \leq \big(200 \ell^2 (1 + \log(3\ell)) + 1\big) \delta^2 \varepsilon^{-2} \log(1/\varepsilon) \leq C_\ell \delta^2 \varepsilon^{-2} \log(1/\varepsilon)$. □

Remark 9.7 (Other presentations).

(a) *Other generators.* Let S be a finite alphabet, fix words $\hat{s} \in F(p, q, \iota, t, d, x, z)$ for $s \in S$, and let $R \subset F(S)$ be finite and $w \in F(S)$ be such that the substituted words $\hat{r}$, $r \in R$, represent 1 in G_{P}^+ and $\hat{w}$ represents J. Apply Proposition 9.4 to the words $\hat{r}$ and $\hat{w}^{-1}(xz)^2$: the composite $s \mapsto \phi(\hat{s})$ is an ε-representation of $\langle S \mid R \rangle$ with $\|\phi(\hat{w}) + I\|_f \leq \varepsilon$, hence $\|\phi(\hat{w}) - I\|_f \geq 1$. Padding as in Theorem 9.5(b), with $N = \lfloor 1/\delta^2 \rfloor$, gives $\log \operatorname{hlp}_R(w; \delta, \varepsilon) \leq C \delta^2 \varepsilon^{-2} \log(1/\varepsilon)$ for $0 < \delta \leq 1$ and $0 < \varepsilon \leq \varepsilon_0 \delta$, with $C, \varepsilon_0 > 0$ depending on the words. If moreover there are words in $F(S)$ for p, q, t, d, x, z as in Corollary 8.4(b), that corollary gives the matching lower bound. This covers every finite presentation of G_{P} or of G_{P}^+ and every word representing J; such presentations exist (Remark 6.7), but Theorem B does not need them.

(b) *The extension G_{P}^+.* For a finite set R^+ of words in p, ι, t, d, x, z valid in G_{P}^+ in which the words of R_{P}, with q replaced by $\iota p \iota$, have bounded area, both bounds of Theorem 9.6(a) hold with other constants: the upper bound by Theorem 9.5, the lower bound by Corollary 8.4(b).

(c) *State-weighted version.* The lower bound of Theorem 9.6(b) also holds, with other constants, for every finite set R of words in p, q, t, d, x, z in which the words of R_{P} have finite area. Fix expressions of these words as products of conjugates of words of $R^{\pm 1}$; there are finitely many, so by Lemma 3.3(d),(a) a model at precision e satisfies (i) and (ii) of Theorem 8.1 at precision $C_R e$, with C_R depending only on the expressions, and (iii) with $\Delta = 2 - e$ as in the proof of Theorem 9.6.

10 The game

A linear system game sees a group only through its solution group. To turn the Pauli configuration-lamp group into a game we therefore embed a group that maps onto it into a solution group, with J in the role of the central involution of the game. Slofstra's embedding theorem does this for groups given by presentations of a special shape, called $E\Gamma_0$ [Slo19, §4]: each relation says that two involutions commute, that some commuting involutions multiply to the identity, or that a generator conjugates one involution to another. Slofstra and Vidick showed that the embedding keeps a distinguished involution w of the presentation visible in approximate representations, as the element J of the solution group [SV18, §6]. This section writes down such a presentation $\mathcal{P}$ for our group: its group maps onto G_{P}^+ with $w \mapsto J$, and

every word of R_{P} follows from its relations with area at most 60 (Proposition 10.3). It then proves Theorem A (Theorem 10.5), by composing the transfer of Section 5, the embedding of Appendix B, the sector theorem of Section 8 and the hash models of Section 9, once in each direction.

The presentation has only one non-involutive generator, p, and this is forced. Houghton's commutator $s = [p, q]$ is an involution, but in an $E\Gamma_0$ group with two non-involutive generators p, q the relation $[p, q]^2 = e$ already implies $[p, q] = e$ (Proposition 10.2). We therefore adjoin the ray swap ι of Lemma 6.9 and write $q = \iota p \iota$. Then the commutator becomes a product of p-conjugates of the single involution ι (equation (19)), which the $E\Gamma_0$ format can express.

What has to be checked is finite. Every relation of $\mathcal{P}$ must hold in G_{P}^{+}, so that the hash models of G_{P}^{+} give approximate representations of $\mathcal{P}$ with $w \mapsto -I$. Every word of R_{P} must follow from the relations of $\mathcal{P}$ with bounded area, so that approximate representations of $\mathcal{P}$ give approximate representations of R_{P}. And Slofstra's construction, applied to $\mathcal{P}$, must produce the linear system of the game. These three facts are verified by machine, each by at least two programs that share no code (Appendix C); every derivation can also be replayed by hand.

10.1 Presentations of type $E\Gamma_0$

Definition 10.1 ($E\Gamma_0$ presentations; [Slo19, Def. 4.7], [SV18, §6])**.** An $E\Gamma_0$ *presentation* $\mathcal{P} = \langle S_0 \sqcup S_1 \mid R \rangle$ consists of two finite sets of generators, the *involutive* generators S_0 and the linearly ordered *non-involutive* generators $S_1 = \{y_1, \dots, y_\ell\}$, a finite family $\mathcal{V}$ of subsets of S_0 (the *linear sets*), and the following relations, and no others:

(E1) $x^2 = e$ for every $x \in S_0$;

(E2) for every $V \in \mathcal{V}$, the *linear relation* $\prod_{x \in V} x = e$, the product taken in a fixed order, and the commutations $xy = yx$ for $x, y \in V$, which are listed among the relations (E3) as $xyx = y$;

(E3) conjugacy relations $xyx = z$ for the triples (x, y, z) of a finite set $C_0 \subseteq S_0^3$;

(E4) conjugacy relations $yxy^{-1} = z$ for the triples (y, x, z) of a finite set $C_1 \subseteq S_1 \times S_0 \times S_0$;

(E5) for some pairs $i > j$, one relation $y_i y_j y_i^{-1} = y_j^{L_{ij}}$ for each such pair, with an integer $L_{ij} \geq 1$.

We write $\langle \mathcal{P} \rangle$ for the group presented, and R for its relators x^2, $\prod_{x \in V} x$, $xyxz^{-1}$, $yxy^{-1}z^{-1}$ and $y_i y_j y_i^{-1} y_j^{-L_{ij}}$. Areas, approximate representations and state-weighted models of $\mathcal{P}$ are taken with respect to R.

Slofstra calls the groups so presented extended homogeneous-linear-plus-conjugacy groups. Because the generators in S_0 are involutions, the relation $xyx = y$ says the same as $xy = yx$, so listing the commutations of a linear set as triples $(x, y, y) \in C_0$, instead of as the separate relations $xy = yx$ of [Slo19, Def. 4.7] and [SV18, §6], presents the same group. When $\ell \leq 1$ there are no pairs $i > j$, so (E5) is empty, and if moreover every linear set has at most three elements, every relator has length at most 4.

10.2 Why one non-involutive generator

In Houghton's group the commutator $s = [p, q]$ is the transposition of $(1, 1)$ and $(1, 2)$, and the first word of R_{P} is s^2. In an $E\Gamma_0$ group, the quotient by the normal subgroup generated by the involutive generators sees only the relations (E5), and with two non-involutive generators the group these define is torsion-free.

Proposition 10.2 (Two non-involutive generators are too many)**.** *Let $\mathcal{P}$ be an $E\Gamma_0$ presentation with $S_1 = \{p, q\}$, and let $G' = \langle \mathcal{P} \rangle$. If $[p, q]^2 = e$ in G', then $[p, q] = e$ in G'. In particular, if G' has a homomorphism to G_{P} that sends p and q to the generators p and q of G_{P}, then the word $s^2 = [p, q]^2$ of R_{P} is not trivial in G', so R_{P} does not follow from the relations of $\mathcal{P}$ under the substitution $p \mapsto p$, $q \mapsto q$.*

Proof. Let N be the normal subgroup of G' generated by S_0. Every relation of types (E1)–(E4) becomes trivial in G'/N, so $G'/N = \langle p, q \mid R_5 \rangle$, where R_5 is the set of relations of type (E5). There is at most one of them, since there is one pair $i > j$. If R_5 is empty, G'/N is free of rank 2. Otherwise G'/N is the Baumslag–Solitar group $\langle a, b \mid bab^{-1} = a^L \rangle$ for some $L \geq 1$, which is $\mathbb{Z}[1/L] \rtimes \mathbb{Z}$, and equals $\mathbb{Z}^2$ for $L = 1$. These groups are torsion-free. Since $[p, q]^2 = e$ in G'/N, we get $[p, q] = e$ in G'/N, so G'/N is abelian. A free group of rank two and $\mathbb{Z}[1/L] \rtimes \mathbb{Z}$ with $L \geq 2$ are not abelian, so R_5 consists of the single relation $qpq^{-1} = p$ or $pqp^{-1} = q$. That relation belongs to the presentation, so $[p, q] = e$ in G'.

For the second statement, such a homomorphism sends $[p, q]$ to s, which moves the configuration $e_{(1,1)}$ and so is not trivial in G_{P}. Hence $[p, q] \neq e$ in G', and by the first statement $[p, q]^2 \neq e$ in G'. □

We keep p as the only non-involutive generator and replace q by $\iota p \iota$, which equals q in G_{P}^+ (Lemma 6.9). Write $\iota_k = p^k \iota p^{-k}$. In the free product $\langle p \rangle * \langle \iota \mid \iota^2 \rangle$,

$$[p, \iota p \iota] = p \iota p \iota p^{-1} \iota p^{-1} \iota = \iota_1 \, \iota_2 \, \iota_1 \, \iota_0.$$

So the commutator, which carried the torsion obstruction, is a product of p-conjugates of one involution, each reached from ι by conjugacy relations $p \iota_k p^{-1} = \iota_{k+1}$ of type (E4); and $\iota_1 \iota_2 \iota_1$ is again an involution, which a computation in Y identifies. From the definition of p, the involution ι_1 exchanges $(1, 1) \leftrightarrow (3, 1)$ and $(2, k) \leftrightarrow (3, k + 1)$ for $k \geq 1$, and fixes $(1, j)$ for $j \geq 2$; and ι_2 exchanges $(1, 2) \leftrightarrow (3, 1)$, $(1, 1) \leftrightarrow (3, 2)$ and $(2, k) \leftrightarrow (3, k + 2)$ for $k \geq 1$, and fixes $(1, j)$ for $j \geq 3$. Conjugating ι_2 by ι_1 transports these pairs to $(1, 2) \leftrightarrow (1, 1)$, $(3, 1) \leftrightarrow (2, 1)$ and $(3, k + 1) \leftrightarrow (2, k + 1)$, and fixes $(1, j)$ for $j \geq 3$. Hence

$$\iota_1 \iota_2 \iota_1 = \sigma \, \iota_0, \qquad \sigma = \big((1, 1)\ (1, 2)\big), \tag{19}$$

where σ and $\iota_0 = \iota$ have disjoint supports; so $[p, \iota p \iota] = \sigma = s$ in G_{P}^+. This is the *s-lemma*: one conjugacy relation $\iota_1 \iota_2 \iota_1 = g_3$ and one linear relation on $\{g_3, \iota, \sigma\}$ express $[p, \iota p \iota] = \sigma$ with $\sigma \in S_0$. From then on s is an involutive generator, and every other word of R_{P} says that a conjugate of an involution is an involution, that two such conjugates commute, or that three commuting ones multiply to e, which an $E\Gamma_0$ presentation can express. The presentation below avoids the obstruction of Proposition 10.2: its quotient by the normal subgroup generated by S_0 is $\langle p \rangle \cong \mathbb{Z}$, which has no nontrivial commutators, while by (19) the commutator s lies in that subgroup.

10.3 The presentation $\mathcal{P}$

Proposition 10.3 (The presentation)**.** *There is an $E\Gamma_0$ presentation $\mathcal{P} = \langle S_0 \sqcup S_1 \mid R \rangle$ with $S_1 = \{p\}$ and a distinguished involutive generator w, listed in full in Appendix C.5, with the following properties.*

(a) Size. *S_0 has 26 elements: ι, t, d, x, z, an involution σ, and $g_1, \ldots, g_{20}$, with $w = g_{14}$. There are 8 linear sets, each with three elements; $|C_0| = 76$, consisting of 32 triples used by the derivations of (d), the 25 triples (g, w, w) for $g \in S_0 \setminus \{w\}$, and the 19 commutations inside linear sets not already among these; and $|C_1| = 10$. Every relator has length at most 4.*

(b) A model. *Each $g \in S_0$ comes with a defining word $\vartheta(g)$ in p, ι, t, d, x, z, among them $\vartheta(\iota) = \iota$, $\vartheta(t) = t$, $\vartheta(d) = d$, $\vartheta(x) = x$, $\vartheta(z) = z$, $\vartheta(\sigma) = [p, \iota p \iota]$, $\vartheta(g_1) = p\iota p^{-1}$, $\vartheta(g_7) = ptp^{-1}$ and $\vartheta(w) = xzxz$. Put $\vartheta(p) = p$. For every relator $r \in R$ the word $\vartheta(r)$ represents 1 in G_{P}^{+}, and its free reduction has length at most $\ell_E = 44$. Hence ϑ induces a homomorphism $\pi : \langle \mathcal{P} \rangle \to G_{\mathrm{P}}^{+}$, which is onto, with $\pi(w) = J$.*

(c) Centrality. *R contains $gwg = w$ for every $g \in S_0 \setminus \{w\}$, and $pwp^{-1} = w$.*

(d) The certificate. *Put $\hat{a} = a$ for $a \in \{p, t, d, x, z\}$ and $\hat{q} = \iota p \iota$. For $r \in R_{\mathrm{P}}$ let $\hat{r}$ be the word in p, ι, t, d, x, z obtained by replacing every letter $a^{\pm 1}$ of r by $\hat{a}^{\pm 1}$. Then every $\hat{r}$, $r \in R_{\mathrm{P}}$, and the word $xzxz\, w^{-1}$ are trivial in $\langle \mathcal{P} \rangle$, each with area at most $A_* = 60$ with respect to R.*

The finite statements in (b) and (d) are checked by machine (Appendix C).

The eight linear sets are $\{\iota_1\iota_2\iota_1, \iota, \sigma\}$ (the s-lemma), $\{t, t_2, tt_2\}$, $\{xzx, z, w\}$, $\{zxz, x, w\}$, and the four sets $\{uvu, v, uvuv\}$ and $\{vuv, u, uvuv\}$ for $(u, v) = (x, t)$ and (z, t), which express the commutation of the Pauli lamp at 0 with the one at $e_{(1,1)}$; here $t_2 = ptp^{-1}$, and we name generators by their defining words. The ten triples of C_1 say that conjugation by p maps $\iota \mapsto \iota_1 \mapsto \iota_2$, $\sigma \mapsto p\sigma p^{-1} \mapsto p^2\sigma p^{-2}$, $t \mapsto t_2 \mapsto t_3$ and $d \mapsto pdp^{-1}$, and fixes x, z and w.

Proof. The proposition is a finite list of identities. The programs that check them are described in Appendix C; here we say what is checked and why that proves the statement.

Format and size. The relations are stored as the lists C_0, C_1 and $\mathcal{V}$, so each has one of the shapes (E1)–(E4); the programs check $S_1 = \{p\}$, the counts in (a), and that every pair u, v inside a linear set has (u, v, v) or (v, u, u) in C_0. The length bound is read off the shapes.

The model, (b). Each defining word, and each relator translated through ϑ, is evaluated exactly in a faithful model of G_{P}^{+} (Appendix C.2): every defining word of an involutive generator squares to the identity, every translated relator gives the identity, and after free reduction the longest translated relator, that of the triple (g_5, w, w), has 44 letters. So ϑ induces π. It is onto because G_{P}^{+} is generated by p, ι, t, d, x, z (Lemma 6.9), and $\pi(w) = J$ because $\vartheta(w) = xzxz$.

Centrality, (c). The 25 triples (g, w, w) and the triple $pwp^{-1} = w$ are in the lists.

The certificate, (d). This is a statement about the abstract group $\langle \mathcal{P} \rangle$ and does not use the model. For each of the 29 target words, namely the s-lemma $[p, \iota p \iota]\sigma^{-1}$, the 27 words $\hat{r}$ with each occurrence of $[p, \iota p \iota]$ replaced by σ, and $xzxz\, w^{-1}$, the file records a derivation, which is replayed in three stages.

(i) The word, which has p-exponent 0, is written as a product $\prod_k p^{n_k} g_k^{\pm 1} p^{-n_k}$ of conjugates of involutive letters, n_k being the p-exponent of the prefix before g_k. This is an identity in the free group, checked by free reduction.

(ii) Each $p^{n_k} g_k p^{-n_k}$ is replaced by an involutive generator h_k through $|n_k|$ relations of C_1, and g_k^{-1} by g_k through (E1). The result is the *resolved sequence* $h_1 \cdots h_M$.

(iii) The resolved sequence is reduced to the empty word by recorded moves, each an instance of a listed relation: cancel aa (E1); replace aba by c for $(a, b, c) \in C_0$; replace ab by ba for $(a, b, b) \in C_0$; replace two letters of a linear set by the third, or delete the three letters of a linear set, in any order (an order that is not a cyclic rotation of the listed one costs three uses of (E1) more).

Each step multiplies the word by a conjugate of one relator and a bounded number of relators x^2, so counting relator applications bounds the area. Each occurrence of $[p, \iota p \iota]$ is charged the area of the s-lemma, whose resolved sequence is $\iota_1\, \iota_2\, \iota_1\, \iota\, \sigma$: one move of C_0 gives $(\iota_1\iota_2\iota_1)\, \iota\, \sigma$, and one linear relation deletes it. The last derivation is $xzxz\, w \mapsto (xzx)\, z\, w \mapsto e$, by the triple (x, z, xzx) and the linear set $\{xzx, z, w\}$. An exact count of cells in the free group on $S_0 \sqcup S_1$,

charging every relator and every use of (E1), gives area at most 60 for each of the 29 words, attained by $(sps\,p^{-1})^3$: six uses of the s-lemma at 9 cells each, and 6 more (Appendix C.3). Two programs that share no code establish this bound, one by a fresh search and one by writing each recorded derivation out as an explicit product of conjugates of relators and checking it by free reduction. A third re-justifies every move as a rewriting along a cyclic rotation of a relator and confirms the derivations with a more generous count, at most 335 cells per word; this weaker bound would serve equally well, as A_* enters only constants. No derivation uses the 25 triples (g, w, w) or the 19 added commutations. □

Any derivation can be replayed by hand. For example, $qxq^{-1}x^{-1}$ becomes $\iota p\iota\, x\, \iota p^{-1}\iota\, x^{-1}$, with resolved sequence $\iota\,\iota_1\, x\, \iota_1\,\iota\, x$ (the factor pxp^{-1} resolves to x by C_1). The triple (ι_1, x, x) turns it into $\iota\, x\, \iota\, x$, the triple (x, ι, ι) into $\iota\,\iota$, and (E1) cancels it.

What the presentation is for. Three properties of $\mathcal{P}$ and w matter. The first two are the hypotheses of the embedding theorem of Slofstra and Vidick [SV18, Prop. 6.2]; the third ties $\mathcal{P}$ to the certificate.

(R1) *w is a central involution* of $\langle\mathcal{P}\rangle$: it commutes with every generator by Proposition 10.3(c), and $w^2 = e$ by (E1). Its image $\pi(w) = J$ is the central involution of G_{P}^{+}.

(R2) *w survives*: for every $\varepsilon > 0$, $\mathcal{P}$ has a finite-dimensional ε-representation with $w \mapsto -I$ (Lemma 10.4 below), so w is nontrivial in approximate representations. This comes from the amenability of G_{P}^{+} through its hash models, and gives the upper bound and the value one.

(R3) *R_{P} follows from $\mathcal{P}$*, with area at most $A_* = 60$ (Proposition 10.3(d)). This gives the lower bound.

Property (R1) is what the flat embedding theorem of Slofstra and Vidick needs: there a visible central involution is amplified to $-I$ [SV18, Prop. 2.7]. The state-weighted proof below never amplifies, and uses only (R2) and (R3).

Lemma 10.4 (w survives)**.** *For every $0 < \varepsilon \le 1$ there is an ε-representation ϕ of $\mathcal{P}$ with $\phi(w) = -I$ in a dimension D with*

$$\log_2 D \;\le\; 400\,\ell_E^2\,\varepsilon^{-2}\log_2(2\ell_E/\varepsilon), \qquad \ell_E = 44.$$

Proof. Let $\vartheta : F(S_0 \sqcup S_1) \to F(p, \iota, t, d, x, z)$ be the homomorphism of Proposition 10.3(b), and let $\mathcal{W}$ be the finite set of the free reductions of the words $\vartheta(r)$, $r \in R$. They represent 1 in G_{P}^{+} and have length at most ℓ_E. By Proposition 9.4 there are D as stated and $\phi_0 : F(p, q, \iota, t, d, x, z) \to \mathrm{U}(D)$ with $\|\phi_0(u) - I\|_f \le \varepsilon$ for $u \in \mathcal{W}$ and $\phi_0((xz)^2) = -I$. Since $\phi_0(\vartheta(r))$ depends only on the free reduction of $\vartheta(r)$, $\phi = \phi_0 \circ \vartheta$ is an ε-representation of $\mathcal{P}$, and $\phi(w) = \phi_0(xzxz) = -I$. □

10.4 Theorem A

Let (A, b) be the binary linear system that Lemma B.1 builds from $\mathcal{P}$ and w. It is listed in `anc/game_Ab.json`; Fig. 2 describes it and Remark 10.7 counts its parts. Let $G = G(A, b)$ be its game with the default question distribution of Definition 2.7, and $\Gamma = \Gamma(A, b)$ its solution group. Recall $\beta(\varepsilon) = 2^{-4(24+\log\log(1/\varepsilon))^2}$, which is nondecreasing on $(0, \frac{1}{2})$ and equals $\varepsilon^{o(1)}$ as $\varepsilon \to 0$.

Theorem 10.5 (Theorem A)**.** *The game G has $m = 1687$ equations in $n = 1993$ variables, with at most $s = 3$ variables in each equation, and:*

The game at a glance. We write x_a for the variable of a generator a.

- *Size.* $m = 1687$ equations over $\mathbb{F}_2$ in $n = 1993$ variables, each equation in at most $s = 3$ variables. The system is in `anc/game_Ab.json`; it is generated from $\mathcal{P}$ and checked by machine (Remark 10.7 and Appendix C).
- *Questions.* The referee picks an equation i uniformly at random and then one of its variables j uniformly at random. Alice receives i and Bob receives j.
- *Answers.* Alice answers values of the variables of equation i that satisfy it, one of 4 answers for three variables, and Bob answers one bit. They win if they give x_j the same value.
- *Equations.* All are homogeneous except $x_w + x_u = 1$. There are four kinds (Lemma B.1): the 8 linear relations of $\mathcal{P}$, such as $x_{xzx}+x_z+x_w = 0$, where w stands for $xzxz$; the equation $x_w+x_u = 1$, which makes J a commutator; 274 equations in three variables, such as $x_a + x_{a'} + x_{a''} = 0$, that put the 97 conjugacy relations in a special form; and, for each of the resulting 234 relations $aba = c$, six equations in seven new variables, $x_a + y_1 + y_2 = x_b + y_2 + y_3 = y_3 + y_4 + y_5 = x_a + y_5 + y_6 = x_c + y_6 + y_7 = y_1 + y_4 + y_7 = 0$.
- *Value.* $\omega_q(G) = 1$, but no strategy on finite-dimensional spaces wins with certainty, and hence none on a tensor product of Hilbert spaces of any dimension [Slo19, Thm. 3.2]. The system has no solution over $\mathbb{F}_2$.
- *Entanglement.* Every strategy that wins with probability $1 - \varepsilon$ has entanglement entropy at least $\varepsilon^{-1+o(1)}$, and a maximally entangled strategy with $\varepsilon^{-1+o(1)}$ ebits suffices (Theorem 10.5).

Figure 2: The game G of Theorem A.

(a) $\omega_q(G) = 1$, *and no strategy on finite-dimensional spaces wins* G *with probability one; hence, by [Slo19, Thm. 3.2], no strategy on a tensor product of Hilbert spaces of any dimension does.*

(b) Let $c = 4.9 \cdot 10^{-34}$ *and* $\varepsilon_0 = 2^{-5500}$. *For* $0 < \varepsilon \le \varepsilon_0$, *every strategy that wins* G *with probability at least* $1 - \varepsilon$ *satisfies*

$$\log \operatorname{rank} \rho_B \;\ge\; S(\rho_B) \;\ge\; c\,\varepsilon^{-1}\beta(\varepsilon),$$

where ρ_B *is Bob's reduced density matrix.*

(c) For $0 < \varepsilon \le \varepsilon_0/2$, *every mixed state* σ *and measurements that win* G *with probability at least* $1 - \varepsilon$ *satisfy* $E_F(\sigma) \ge \frac{c}{4}\,\varepsilon^{-1}\beta(\varepsilon)$ *(Corollary 10.6).*

(d) For $0 < \varepsilon \le \frac{1}{2}$, $\mathcal{E}^{\mathrm{me}}(G,\varepsilon) \le \exp\bigl(C\varepsilon^{-1}\log(1/\varepsilon)\bigr)$ *with* $C = 7.4 \cdot 10^{10}$.

Consequently $\log \mathcal{E}(G,\varepsilon) = \varepsilon^{-1+o(1)} = \log \mathcal{E}^{\mathrm{me}}(G,\varepsilon)$ *as* $\varepsilon \to 0$.

All constants are explicit and far from optimal. The constants c and ε_0 come from the constant $A_0 = 42$ of Lemma B.1, the largest area in Γ of finitely many words that the construction makes trivial (Remark B.2), and from the sector theorem (Remark 10.8).

10.5 The proof at a glance

Figure 3 shows the two chains of the proof.

The lower bound. A strategy that wins with probability $1-\varepsilon$ gives, on its Schmidt support, a state-weighted model (ϕ, ρ) of the solution group Γ with $\phi(J) = -I$ (Theorem 5.1): every relator has defect at most 4θ in Bob's reduced state ρ, and every generator nearly commutes with $\xi = \rho^{1/2}$, where $\theta = 2\sqrt{ms\varepsilon}$. So the precision is of order $\sqrt{\varepsilon}$, and it is never flattened. The substitution Ψ of the embedding pulls the model back to a model of $\mathcal{P}$ at a constant factor worse precision, and the commutator relation $[x_w, x_{j_0}]J = e$ built into the game forces $\phi(x_w)$ to stay at distance about 1 from I in the state (Theorem B.3(b)): this is where $J = -I$

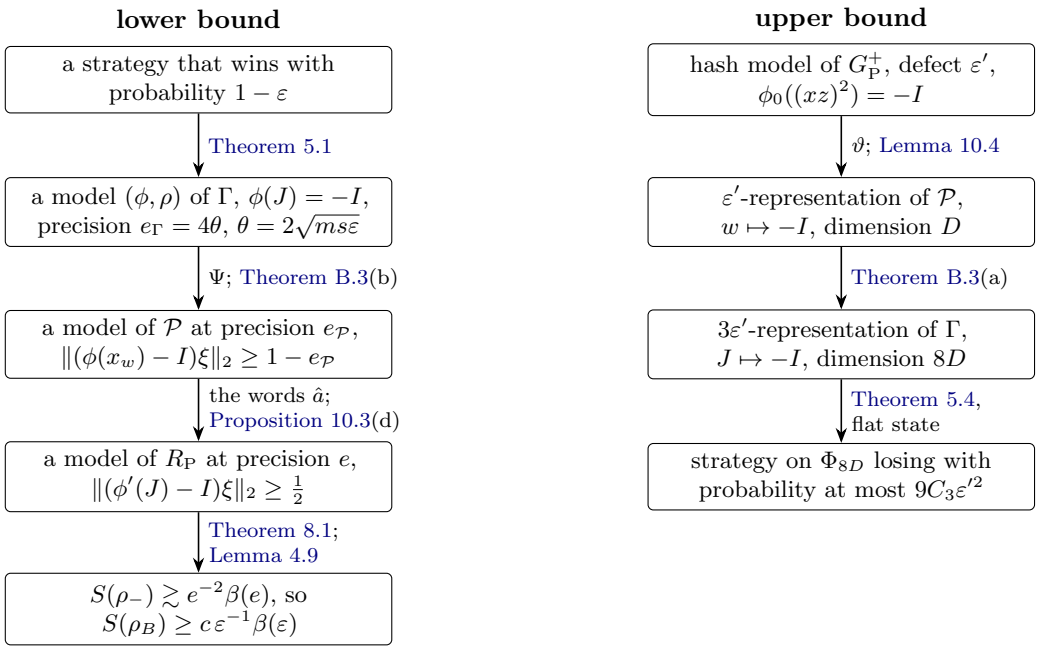


Figure 3: The two directions of the proof of Theorem A. Here $e_\mathcal{P} = C_\Psi e_\Gamma$ and $e = 36780\, e_\mathcal{P}$; on the right, $\varepsilon' = \frac{1}{3}\sqrt{\varepsilon/C_3}$ with the constant $C_3 < 788.2$ of Theorem 5.4.

becomes visible on the side of $\mathcal{P}$. The words $\hat{a}$ of Proposition 10.3(d) pull the model back once more, to an assignment of unitaries to p, q, t, d, x, z that satisfies R_P in the state and in which $J = (xz)^2$ stays visible. Then the sector theorem counts approximate Pauli pairs on the sector where J acts as -1, and returns entropy e^{-2}, up to the factor β, at precision e (Theorem 8.1); Lemma 4.9 moves the entropy from the sector state to ρ. Every step costs a constant factor in the precision, so $e^{-2} \asymp (\sqrt{\varepsilon})^{-2} = \varepsilon^{-1}$ survives.

The upper bound. The hash models give approximate representations of G_P^+ with $J \mapsto -I$ exactly, of defect ε' in dimension $\exp(O(\varepsilon'^{-2}\log(1/\varepsilon')))$ (Proposition 9.4). Composed with the defining words they are ε'-representations of $\mathcal{P}$ with $w \mapsto -I$ (Lemma 10.4). Slofstra's gadgets turn these into $3\varepsilon'$-representations of Γ with $J \mapsto -I$, in eight times the dimension (Theorem B.3(a)). With the flat state, the return direction of the transfer turns these into strategies on a maximally entangled state that lose with probability $O(\varepsilon'^2)$ (Theorem 5.4). Taking $\varepsilon' \asymp \sqrt{\varepsilon}$ gives local dimension $\exp(O(\varepsilon^{-1}\log(1/\varepsilon)))$.

10.6 Proof of the lower bound

Throughout, $m = 1687$ and $s = 3$; the default distribution has $\pi_\mathrm{min} = 1/(ms)$. Let Ψ, x_{j_0} and $A_0 = 42$ be as in Lemma B.1 and Remark B.2, and $C_\Psi = A_0(8A_0+17) = 14826$ as in Theorem B.3. Put

$$C_* = 294240\sqrt{ms}\,C_\Psi < 3.11\cdot 10^{11}, \qquad c_* = \frac{c_1}{512\,C_*^2} > 4.9\cdot 10^{-34}, \qquad c_1 = \frac{1}{4.1\cdot 10^7},$$

where c_1 is the constant of Theorem 8.1. We prove (b) with c_* in place of c, which is stronger, and for every $\varepsilon_0 \in (0, \frac{1}{4}]$ such that, for $0 < \varepsilon \leq \varepsilon_0$,

$$C_*\sqrt{\varepsilon} \leq 2^{-1072} \qquad \text{and} \qquad c_*\,\varepsilon^{-1}\beta(\varepsilon) \geq 1; \tag{20}$$

such an ε_0 exists because $\varepsilon^{-1}\beta(\varepsilon) \to \infty$ as $\varepsilon \to 0$. The value $\varepsilon_0 = 2^{-5500}$ of Theorem 10.5 qualifies. Write $L = \log(1/\varepsilon) \geq 5500$. The first condition says $L \geq 2144 + 2\log C_*$, and $2144 + 2\log C_* < 2221$. The second says $L - 4(24+\log L)^2 \geq \log(1/c_*)$, where $\log(1/c_*) < 111$. Its left side is increasing for $L \geq 5500$ and exceeds 190 at $L = 5500$.

Proof of Theorem 10.5(b). Let $0 < \varepsilon \le \varepsilon_0$, fix a strategy that wins with probability at least $1-\varepsilon$, and put $\theta = 2\sqrt{\varepsilon/\pi_{\min}} = 2\sqrt{ms\varepsilon}$.

Step 1: the transfer. By Theorem 5.1 there are a homomorphism $\phi : F(x_1,\dots,x_n,J) \to \mathrm{U}(r)$, where r is the Schmidt rank of the state, and a density matrix ρ of full rank on $\mathbb{C}^r$, Bob's reduced density matrix restricted to its support, such that, with $\xi = \rho^{1/2}$, $\phi(J) = -I$, every relator R' of Γ has $\|(\phi(R')-I)\xi\|_2 \le \max(s,4)\theta = 4\theta$, and every x_j has $\|[\phi(x_j),\xi]\|_2 \le \sqrt{2}\theta$. Since $\phi(J) = -I$ commutes with ξ, (ϕ,ρ) is a state-weighted model of Γ (Definition 2.5) at precision $e_\Gamma = 4\theta$.

Step 2: from Γ to $\mathcal{P}$. By Theorem B.3(b), $(\phi_{\mathcal{P}},\rho)$ with $\phi_{\mathcal{P}} = \phi\circ\Psi$ is a state-weighted model of $\mathcal{P}$ at precision $e_{\mathcal{P}} = C_\Psi e_\Gamma$, and $W = \phi_{\mathcal{P}}(w) = \phi(x_w)$ satisfies

$$\|(W-I)\xi\|_2 \;\ge\; 1 - e_{\mathcal{P}}.$$

Step 3: from $\mathcal{P}$ to R_{P}. Put $\phi'(a) = \phi_{\mathcal{P}}(\hat{a})$ for $a \in \{p,q,t,d,x,z\}$, so that $\phi'(r) = \phi_{\mathcal{P}}(\hat{r})$ for every word r in these letters. Let $r \in R_{\mathrm{P}}$. It has at most 22 letters (Definition 7.1), so $|\hat{r}| \le 66$, and $\hat{r}$ has area at most 60 with respect to R by Proposition 10.3(d). The relators of $\mathcal{P}$ have length at most 4, so by Lemma 3.4 the word $\hat{r}$ is freely equal to a product of at most 60 conjugates of relators of $\mathcal{P}$ and their inverses, with conjugators of length at most $66 + 60\cdot 4 = 306$. By Lemma 3.3(d), with (S1) and (S2) for $\mathcal{P}$ at precision $e_{\mathcal{P}}$,

$$\|(\phi'(r)-I)\xi\|_2 \;\le\; 60\,(1+2\cdot 306)\,e_{\mathcal{P}} \;=\; 36780\,e_{\mathcal{P}}.$$

The same holds for the word $xzxz\,w^{-1}$, which has 5 letters and area at most 60. Each $\hat{a}$ has at most 3 letters, so $\|[\phi'(a),\xi]\|_2 \le 3e_{\mathcal{P}}$ by Lemma 3.3(a). Put

$$e = 36780\,e_{\mathcal{P}} = 147120\,C_\Psi\,\theta = C_*\sqrt{\varepsilon}.$$

Then (ϕ',ρ) satisfies conditions (i) and (ii) of Theorem 8.1 at precision e, and $e \le 2^{-1072}$ by (20).

Step 4: visibility. Here J denotes the word $xzxz$, so $\phi'(J) = \phi_{\mathcal{P}}(xzxz)$ and $\phi'(J)W^* = \phi_{\mathcal{P}}(xzxz\,w^{-1})$. Since $\phi'(J) - W = (\phi'(J)W^* - I)W$ and $W\xi = \xi W + [W,\xi]$,

$$\|(\phi'(J)-W)\xi\|_2 \;\le\; \|(\phi'(J)W^*-I)\xi\|_2 + 2\|[W,\xi]\|_2 \;\le\; e + 2e_{\mathcal{P}},$$

using Step 3 and (S2) for the generator w. With Step 2 and $3e_{\mathcal{P}} \le e \le \frac14$,

$$\|(\phi'(J)-I)\xi\|_2 \;\ge\; \|(W-I)\xi\|_2 - \|(\phi'(J)-W)\xi\|_2 \;\ge\; 1 - e - 3e_{\mathcal{P}} \;\ge\; \tfrac12.$$

Step 5: the sector. By Steps 3 and 4, Theorem 8.1 applies with $\Delta = \frac12$, since $e \le 2^{-1072} = 2^{-1071}\Delta$. Let $J' = \mathrm{r}(\phi'(J))$ be the rounding of Lemma 3.6, $P_- = \frac12(I - J')$, $p_- = \mathrm{Tr}(P_-\rho)$, and ρ_- the sector state of Lemma 3.10 for J'. By Theorem 8.1, $p_- \ge \Delta^2/16 = \frac{1}{64}$ and

$$S(\rho_-) \;\ge\; \tfrac{c_1}{4}\,e^{-2}\,\beta(e) \;\ge\; \frac{c_1}{4C_*^2}\,\varepsilon^{-1}\beta(\varepsilon),$$

because $e^{-2} = C_*^{-2}\varepsilon^{-1}$, and $\varepsilon \le \sqrt{\varepsilon} \le e < \frac12$ (as $C_* \ge 1$) with β nondecreasing on $(0,\frac12)$.

Step 6: back to ρ. By Lemma 4.9, since $h \le 1$,

$$S(\rho) \;\ge\; p_-\,S(\rho_-) - 1 \;\ge\; \frac{c_1}{256\,C_*^2}\,\varepsilon^{-1}\beta(\varepsilon) - 1 \;=\; 2c_*\,\varepsilon^{-1}\beta(\varepsilon) - 1 \;\ge\; c_*\,\varepsilon^{-1}\beta(\varepsilon),$$

the last step by (20). Finally ρ and ρ_B have the same nonzero eigenvalues, so $\log\operatorname{rank}\rho_B = \log\operatorname{rank}\rho \ge S(\rho) = S(\rho_B)$. □

The proof uses the strategy only through Theorem 5.1, and the state only through (S1) and (S2); no spectral projection of ρ is taken. Since $\omega_q(G) = 1$ by (a), a strategy on $\mathbb{C}^d \otimes \mathbb{C}^d$ that wins with probability at least $\omega_q(G) - \varepsilon$ has Schmidt rank at most d, and (b) applies to it; so $\log \mathcal{E}(G, \varepsilon) \geq c\,\varepsilon^{-1}\beta(\varepsilon)$ for $0 < \varepsilon \leq \varepsilon_0$.

Corollary 10.6 (Entanglement of formation)**.** *With c and ε_0 as in Theorem 10.5(b), let $0 < \varepsilon \leq \varepsilon_0/2$. Every mixed state σ on $H_A \otimes H_B$ and measurements that win G with probability at least $1 - \varepsilon$ satisfy*

$$E_F(\sigma) \;\geq\; \tfrac{c}{4}\,\varepsilon^{-1}\beta(\varepsilon).$$

Proof. Let $\sigma = \sum_j p_j |\psi_j\rangle\langle\psi_j|$ with $p_j > 0$ and unit vectors ψ_j, and let ℓ_j be the losing probability of the pure strategy that uses ψ_j with the given measurements. The losing probability is linear in the state, so $\sum_j p_j \ell_j \leq \varepsilon$, and the components with $\ell_j > 2\varepsilon$ carry total weight less than $\frac{1}{2}$. Each component with $\ell_j \leq 2\varepsilon$ is a strategy that wins with probability at least $1 - 2\varepsilon$, with $2\varepsilon \leq \varepsilon_0$, so by Theorem 10.5(b) and the monotonicity of β,

$$S\big(\mathrm{Tr}_A |\psi_j\rangle\langle\psi_j|\big) \;\geq\; c\,(2\varepsilon)^{-1}\beta(2\varepsilon) \;\geq\; \tfrac{c}{2}\,\varepsilon^{-1}\beta(\varepsilon).$$

These components carry weight at least $\frac{1}{2}$, so $\sum_j p_j S(\mathrm{Tr}_A |\psi_j\rangle\langle\psi_j|) \geq \frac{c}{4}\varepsilon^{-1}\beta(\varepsilon)$. The infimum over decompositions is $E_F(\sigma)$ (Definition 2.8). □

10.7 Proof of the upper bound and of value one

Proof of Theorem 10.5(d) and (a). Let $C_3 < 788.2$ be the constant C_s of Theorem 5.4 for $s = 3$. Let $0 < \varepsilon \leq \frac{1}{2}$ and put $\varepsilon' = \frac{1}{3}\sqrt{\varepsilon/C_3} \leq 1$.

Step 1: a representation of $\mathcal{P}$. By Lemma 10.4 there is an ε'-representation of $\mathcal{P}$ with $w \mapsto -I$ in a dimension D with $\log_2 D \leq 400 \cdot 44^2\, \varepsilon'^{-2} \log_2(88/\varepsilon')$.

Step 2: a representation of Γ. By Theorem B.3(a) there is a $3\varepsilon'$-representation ϕ_Γ of Γ on $\mathbb{C}^{8D}$ with $\phi_\Gamma(J) = -I$.

Step 3: a maximally entangled strategy. Let $\rho = I/(8D)$ be the flat state. Then $\|Y\xi\|_2 = \|Y\|_f$ for every operator Y, all state commutators vanish, and $\phi_\Gamma(J) + I = 0$, so (ϕ_Γ, ρ) is a model of (Γ, J) at precision $3\varepsilon'$ (Definition 2.5). By Theorem 5.4, the strategy built there on the canonical purification of ρ, which is the maximally entangled state Φ_{8D}, has local dimension $8D$ and loses every question pair, hence a random one, with probability at most $C_3(3\varepsilon')^2 = \varepsilon$.

Step 4: the count. Since $\varepsilon'^{-2} = 9C_3\varepsilon^{-1}$ and $88/\varepsilon' = 264\sqrt{C_3/\varepsilon}$,

$$\log_2(8D) \;\leq\; 3 + 400 \cdot 44^2 \cdot 9\, C_3\, \varepsilon^{-1}\Big(\log_2\big(264\sqrt{C_3}\big) + \tfrac{1}{2}\log_2(1/\varepsilon)\Big).$$

Here $400 \cdot 44^2 \cdot 9\, C_3 < 5.50 \cdot 10^9$ and $\log_2(264\sqrt{C_3}) < 12.86$, and for $\varepsilon \leq \frac{1}{2}$ we have $\log_2(1/\varepsilon) \geq 1$ and $3 \leq \frac{3}{2}\varepsilon^{-1}\log_2(1/\varepsilon)$. So $\log_2(8D) \leq 7.4 \cdot 10^{10}\,\varepsilon^{-1}\log_2(1/\varepsilon)$. Since $2^y \leq e^y$ for $y \geq 0$, the strategy of Step 3 shows $\mathcal{E}^{\mathrm{me}}(G, \varepsilon) \leq \exp(C\varepsilon^{-1}\log(1/\varepsilon))$ with $C = 7.4 \cdot 10^{10}$, once we know $\omega_q(G) = 1$.

Step 5: value one. Step 3 gives, for every $0 < \varepsilon \leq \frac{1}{2}$, a strategy that wins with probability at least $1 - \varepsilon$. So $\omega_q(G) = 1$, which completes (d). Suppose a strategy on finite-dimensional spaces won with probability one. It would win with probability at least $1 - \varepsilon$ for every ε, so by (b) its reduced density matrix would satisfy $\log \mathrm{rank}\,\rho_B \geq c\,\varepsilon^{-1}\beta(\varepsilon)$ for every $0 < \varepsilon \leq \varepsilon_0$. The right side is unbounded as $\varepsilon \to 0$, while $\mathrm{rank}\,\rho_B \leq \dim H_B$. So there is no such strategy. By [Slo19, Thm. 3.2], a linear system game has a perfect strategy on a tensor product of Hilbert spaces of arbitrary dimension if and only if it has one on finite-dimensional spaces; so there is none of that kind either. □

Proof of the last sentence of Theorem 10.5. For $0 < \varepsilon \le \varepsilon_0$, the paragraph after the proof of (b), $\mathcal{E} \le \mathcal{E}^{\mathrm{me}}$ and (d) give

$$c\,\varepsilon^{-1}\beta(\varepsilon) \;\le\; \log \mathcal{E}(G,\varepsilon) \;\le\; \log \mathcal{E}^{\mathrm{me}}(G,\varepsilon) \;\le\; C\log_2(e)\,\varepsilon^{-1}\log(1/\varepsilon).$$

Since $\beta(\varepsilon) = \varepsilon^{o(1)}$ and $\log(1/\varepsilon) = \varepsilon^{-o(1)}$, both logarithms are $\varepsilon^{-1+o(1)}$. □

10.8 Remarks

Remark 10.7 (The size of the game). The four steps of Lemma B.1, applied to $\mathcal{P}$, give the following counts, which the generating program and the independent checker of Appendix C both reproduce. Step 1 adds 12 involutions (f, v, and one Z_{bc} for each of the 10 triples of C_1) and 20 conjugacy relations to the 76 of $\mathcal{P}$. Step 2 adds the involutions t^*, u, the linear set $\{w, u\}$ and one conjugacy relation; this leaves $N = 40$ generators besides J and 97 conjugacy relations. Step 3 adds $3N + 1 + 2 \cdot 97 = 315$ involutions and $2N + 2 \cdot 97 = 274$ linear sets, and turns the 97 conjugacy relations into $N + 2 \cdot 97 = 234$ nice ones. Step 4 replaces each nice relation by 6 equations in 7 new variables. So $8 + 1 + 274 = 283$ equations come from linear sets and $6 \cdot 234 = 1404$ from gadgets, 1687 in all, in $26 + 12 + 2 + 315 + 1638 = 1993$ variables. All equations have three variables except $x_w + x_u = 1$, which has two and is the only inhomogeneous one. Gaussian elimination shows that $Ax = b$ has no solution over $\mathbb{F}_2$ (A has rank 1512), as it must, since J is a commutator in Γ: no deterministic strategy wins with certainty. A program independent of the generator also evaluates every relator of Γ in the representations $\gamma, \gamma_1, \gamma_2$ of the proof of Lemma B.1, built from the exact model of G_{P}^{+} with its central involution J in place of -1, and finds a homomorphism $\Gamma \to G_{\mathrm{P}}^{+} \wr \mathrm{Sym}(8)$ in which $J \neq 1$ (Appendix C.3). None of this is used in the proof of Theorem 10.5, which needs only that (A, b) is the output of Lemma B.1.

Remark 10.8 (The constants). Only the constants c and ε_0 depend on the area constant $A_0 = 42$ of Remark B.2, and they are astronomically small: already the precision condition $e \le 2^{-1071}\Delta$ of the sector theorem forces $\varepsilon_0 < 2^{-2144}$, and the factor $\beta(\varepsilon)$ pushes the admissible ε_0 down to about 2^{-5420}. The constant C of (d) is explicit, and improving it is a matter of the hash models (the factor $400\,\ell_E^2$) and of the constant of Theorem 5.4. The bounds (b) and (d) have the same power of ε; only the factors $\beta(\varepsilon)$ and $\log(1/\varepsilon)$ separate them.

Remark 10.9 (Under the polynomial Dehn bound). If the polynomial bound (H2) on the Dehn function of H_3 is used, Remark 8.3 replaces the factor $2^{-4(24+\log\log(1/e))^2}$ of Theorem 8.1 by a constant multiple of $(\log(1/e))^{-4\kappa}$, for $e \le e_0\Delta$. The proof above then gives Theorem 10.5(b),(c) and Corollary 10.6 with $\beta(\varepsilon)$ replaced by $(\log(1/\varepsilon))^{-4\kappa}$, with new constants c, ε_0, since $e \ge \varepsilon$ gives $(\log(1/e))^{-4\kappa} \ge (\log(1/\varepsilon))^{-4\kappa}$.

11 Two sharp rates

The state-weighted method of Sections 3 to 5 is not specific to lamp groups. This section applies it to two games for which the earlier lower bounds lose a power of ε on the way from the strategy to a count. For the XOR games CHSH(n) the qubit count of Theorem 4.1, applied once in the reduced state of the players, gives the ebit cost $\Theta(\min(n, 1/\varepsilon))$ (Theorem E, Section 11.1). For the linear system game of Slofstra and Vidick, a transport argument for spectral measures in one state gives the profile $\Theta(1/\varepsilon)$ of its group and, through Theorem 5.1, the local dimension $\Theta(\varepsilon^{-1/2})$ (Theorem F, Section 11.2). In both cases the new lower bound meets an upper bound of the same order: for CHSH(n) a known one, and for the game of Slofstra and Vidick theirs, through a lift that keeps the distinguished involution balanced (Lemma 11.13).

11.1 The CHSH(n) games

The XOR games CHSH(n) of Slofstra [Slo11] force n pairwise anticommuting observables on one side. A strategy within a factor $1-\varepsilon$ of the optimal bias pays a CHSH deficit for each of the $\binom{n}{2}$ pairs, and these deficits add up to $O(n^2\varepsilon)$. From triples of anticommuting observables one builds $\Theta(n)$ approximate qubits [OV18], and Theorem 4.1 charges each of them only the sum of the squares of its pair errors, which is $O(n\varepsilon)$. So a constant fraction of the $n/3$ qubits survives as long as $n\varepsilon$ is small.

Definition 11.1 (CHSH(n); [Slo11], [OV18, Section 2.3])**.** In CHSH(n), $n \geq 2$, the referee picks $i<j$ uniformly, sends Alice i or j and Bob (i,j) or (j,i), independently and uniformly, and accepts if and only if $a \oplus b = 1$ when Alice received j and Bob (j,i), and $a \oplus b = 0$ otherwise. A strategy is a unit vector $\psi \in H_A \otimes H_B$, with H_A, H_B finite-dimensional, together with binary observables (Hermitian contractions) A_k for Alice, $k \in [n]$, and B_{ij} for Bob, $i \neq j$. Its *bias*, the probability of winning minus that of losing, is

$$\text{bias} = \frac{1}{2n(n-1)} \sum_{i<j} C_{ij}, \qquad C_{ij} = \langle \psi, (A_i \otimes (B_{ij} + B_{ji}))\psi \rangle + \langle \psi, (A_j \otimes (B_{ij} - B_{ji}))\psi \rangle.$$

The optimal bias is $\text{bias}^* = \sqrt{2}/2$, and a strategy is *$\varepsilon$-optimal* if $\text{bias} \geq (1-\varepsilon)\,\text{bias}^*$.

Ostrev and Vidick measure the bias by its absolute value; replacing every A_k by $-A_k$ changes its sign and nothing else, so we may take it nonnegative. The next lemma is the robust form of Tsirelson's bound for one pair, with every error measured in Alice's reduced state.

Lemma 11.2 (One pair)**.** *Let A, A' be involutions on H_A, B, B' involutions on H_B, $\psi \in H_A \otimes H_B$ a unit vector and $\xi_A = \rho_A^{1/2}$. Put $T_0 = (B+B')/\sqrt{2}$, $T_1 = (B-B')/\sqrt{2}$, $C = \langle \psi, (A \otimes (B+B'))\psi \rangle + \langle \psi, (A' \otimes (B-B'))\psi \rangle$ and $\delta = 2\sqrt{2} - C$. Then*

$$\delta = \frac{1}{\sqrt{2}} \Big(\|(A \otimes I - I \otimes T_0)\psi\|^2 + \|(A' \otimes I - I \otimes T_1)\psi\|^2 \Big).$$

Consequently $\delta \geq 0$, and

$$\|\{A, A'\}\xi_A\|_2^2 \leq 33\,\delta, \qquad \|[A, \xi_A]\|_2^2 \leq 2\sqrt{2}\,\delta, \qquad \|[A', \xi_A]\|_2^2 \leq 2\sqrt{2}\,\delta.$$

Proof. Write $\langle \cdot \rangle$ for the expectation in ψ. Since $A^2 = I$, $\|(A \otimes I - I \otimes T_0)\psi\|^2 = 1 + \langle T_0^2 \rangle - \sqrt{2}\langle A \otimes (B+B') \rangle$, and likewise for the second norm. As $T_0^2 + T_1^2 = B^2 + B'^2 = 2I$, the sum of the two norms is $4 - \sqrt{2}\,C$, which is the identity. So both norms are at most $\tau = (\sqrt{2}\delta)^{1/2}$. Since $\{T_0, T_1\} = B^2 - B'^2 = 0$ and $\|T_0\|_{\text{op}}, \|T_1\|_{\text{op}} \leq \sqrt{2}$, and since $A \otimes I$ commutes with $I \otimes T_1$,

$$(AA' \otimes I)\psi = (A \otimes I)(I \otimes T_1)\psi + (A \otimes I)\big((A' \otimes I) - (I \otimes T_1)\big)\psi = (I \otimes T_1 T_0)\psi + \epsilon_1,$$

with $\|\epsilon_1\| \leq (1+\sqrt{2})\tau$. In the same way $(A'A \otimes I)\psi = (I \otimes T_0 T_1)\psi + \epsilon_2$ with $\|\epsilon_2\| \leq (1+\sqrt{2})\tau$. Adding, and using Lemma 3.1(a) for Alice's side, $\|\{A, A'\}\xi_A\|_2 \leq 2(1+\sqrt{2})\tau$; and $4(1+\sqrt{2})^2\sqrt{2} < 33$. Finally Lemma 3.1(b), with the roles of the two factors exchanged, gives Hermitian operators $\tilde{T}_0, \tilde{T}_1$ on H_A with $\|A\xi_A - \xi_A\tilde{T}_0\|_2$, $\|A'\xi_A - \xi_A\tilde{T}_1\|_2 \leq \tau$, and Lemma 3.2 gives the last two bounds. □

With all observables involutions, the lemma gives $C_{ij} \leq 2\sqrt{2}$, which is Tsirelson's bound [Cir80], and hence bias $\leq$ bias*; the reduction to involutions in the next proof shows that this holds for all strategies.

Theorem 11.3 (Ebits for CHSH(n)). *Let $n \geq 3$ and $\varepsilon > 0$, and put $M = \lfloor \min(n, 10^{-7}/\varepsilon)/3 \rfloor$. Every ε-optimal strategy for CHSH(n) satisfies*

$$S(\rho_A) \;\geq\; M/2.$$

In particular $S(\rho_A) \geq 10^{-7} \min(n, 1/\varepsilon)/6 - 1/2$.

The proof selects $n' = 3M$ questions whose pairwise CHSH deficits are small on average, splits them into M triples, turns each triple into one approximate Pauli pair as in [OV18], and applies Theorem 4.1 to the M pairs in Alice's reduced state. The pair errors enter only through their squares, which add up to $O(n'^2\varepsilon)$, that is, $O(n'\varepsilon)$ per qubit.

Proof. If $M = 0$ there is nothing to prove, so let $M \geq 1$ and $n' = 3M \leq \min(n, 10^{-7}/\varepsilon)$.

Step 0: reductions. With all other observables fixed, the bias is an affine function $\mathrm{Tr}(A_k N_k) + \text{const}$ of Alice's observable A_k, with N_k Hermitian, so replacing A_k by $\mathrm{sgn}(N_k)$ (with $\mathrm{sgn}(0) = 1$) does not lower it; the same holds for Bob's observables. Doing this for each observable in turn, we may assume that all observables are involutions; the state is unchanged. Let $\delta_{ij} = 2\sqrt{2} - C_{ij} \geq 0$ (Lemma 11.2). From bias $\geq (1-\varepsilon)\sqrt{2}/2$,

$$\sum_{i<j} \delta_{ij} \;\leq\; 2\sqrt{2}\,\varepsilon \binom{n}{2}.$$

Averaging over uniformly random n'-subsets of $[n]$, fix a subset $\mathcal{N}$ whose pair average $\bar{\delta} = \binom{n'}{2}^{-1} \sum_{x<y\in\mathcal{N}} \delta_{xy}$ is at most $2\sqrt{2}\,\varepsilon$. A fixed pair of $\mathcal{N}$ lies in a common block of a uniformly random partition of $\mathcal{N}$ into M triples with probability $2/(n'-1)$; fix a partition $\mathcal{T}_1, \ldots, \mathcal{T}_M$ for which the sum of δ_{xy} over the $3M$ pairs inside the triples is at most $\frac{2}{n'-1}\binom{n'}{2}\bar{\delta} = n'\bar{\delta}$.

Step 1: pair errors. Write $\xi = \rho_A^{1/2}$, $\eta_{xy} = \|\{A_x, A_y\}\xi\|_2$ and $c_x = \|[A_x, \xi]\|_2$. Applying Lemma 11.2 to the pair $\{x, y\}$, with $(A, A', B, B') = (A_x, A_y, B_{xy}, B_{yx})$ for $x < y$,

$$\eta_{xy}^2 \leq 33\,\delta_{xy}, \qquad c_x^2 \leq 2\sqrt{2}\,\delta_{xy} \quad \text{for every } y \neq x.$$

Choosing for each $x \in \mathcal{N}$ the partner $y \in \mathcal{N}$ with the least δ_{xy},

$$\sum_{x\in\mathcal{N}} c_x^2 \;\leq\; 2\sqrt{2} \sum_{x\in\mathcal{N}} \frac{1}{n'-1} \sum_{y\in\mathcal{N}\setminus x} \delta_{xy} \;=\; 2\sqrt{2}\,n'\bar{\delta}.$$

Step 2: one qubit from each triple. The construction of pairs from triples is that of [OV18]. For $\mathcal{T}_k = \{a, b, c\}$, labelled in any order, put $U_k = iA_aA_b$ and $W_k = iA_bA_c$, which are unitaries, and let $X_k = \mathrm{r}(U_k)$ and $Z_k = \mathrm{r}(W_k)$ be their roundings (Lemma 3.6). Since $U_k^2 - I = -A_aA_b\{A_a, A_b\}$, we have $\|(U_k^2 - I)\xi\|_2 = \eta_{ab}$, so

$$\|(X_k - U_k)\xi\|_2 \leq \eta_{ab}, \qquad \|[X_k, \xi]\|_2 \leq c_a + c_b + 2\eta_{ab},$$

and likewise for Z_k with (b, c) in place of (a, b). Put $\lambda_k = \eta_{ab} + \eta_{bc} + \eta_{ac} + c_a + c_b + c_c$ and, for $k \neq l$, $\mu_{kl} = \sum_{x\in\mathcal{T}_k, y\in\mathcal{T}_l} \eta_{xy}$. For distinct indices b, c, b', c', and for one triple $\{a, b, c\}$, we have the identities, which use only $A_x^2 = I$,

$$\begin{aligned} A_bA_cA_{b'}A_{c'} - A_{b'}A_{c'}A_bA_c &= A_b\{A_c, A_{b'}\}A_{c'} - \{A_b, A_{b'}\}A_cA_{c'} \\ &\quad + A_{b'}A_b\{A_c, A_{c'}\} - A_{b'}\{A_b, A_{c'}\}A_c, \\ \{U_k, W_k\} &= -A_b\{A_a, A_c\}A_b + \{A_a, A_b\}A_cA_b - A_aA_b\{A_b, A_c\}. \end{aligned}$$

Apply them to ξ and move the trailing observables to the right of ξ, at cost $2c_x$ for each, the anticommutators having norm at most 2. For $k \neq l$ this gives

$$\|[W_k, W_l]\xi\|_2,\ \|[U_k, W_l]\xi\|_2 \;\leq\; \mu_{kl} + 4\lambda_k + 4\lambda_l, \qquad \|\{U_k, W_k\}\xi\|_2 \leq 4\lambda_k.$$

Lemma 3.3(e), with the primed operators the unitaries U and W, whose state commutators are at most λ_k, replaces U, W by X, Z: for $k \neq l$,

$$\|[Z_k, Z_l]\xi\|_2,\ \|[X_k, Z_l]\xi\|_2 \le \mu_{kl} + 8\lambda_k + 8\lambda_l,$$
$$\|\{X_k, Z_k\}\xi\|_2 \le 12\lambda_k, \qquad \|[X_k, \xi]\|_2,\ \|[Z_k, \xi]\|_2 \le 2\lambda_k.$$

Step 3: counting. Apply Theorem 4.1 to the M pairs (X_k, Z_k) in the state ρ_A. With $c_k^X = \|[X_k, \xi]\|_2$ and $c_k^Z = \|[Z_k, \xi]\|_2$, its parts (i)–(iii) give numbers $a_k, b_k \ge 0$ with $a_k^2 \le 2\sum_{l>k}(\|[Z_k, Z_l]\xi\|_2 + c_l^Z)^2 + 2\sum_{l>k}(c_l^Z)^2$, $b_k^2 \le 2(\|\{X_k, Z_k\}\xi\|_2 + c_k^Z)^2 + 2\sum_{l\neq k}(\|[X_k, Z_l]\xi\|_2 + c_l^Z)^2 + 2\sum_l (c_l^Z)^2$ and

$$S(\rho_A) \ \ge\ \sum_{k=1}^{M}\Big[1 - \frac{(b_k + c_k^X)^2}{2\ln 2} - h\Big(\min\Big(\frac{(a_k + c_k^Z)^2}{4}, \frac{1}{2}\Big)\Big)\Big].$$

Put $\Sigma_\lambda = \sum_k \lambda_k^2$. By Step 2 and $(x+y+z)^2 \le 3(x^2+y^2+z^2)$,

$$a_k^2 \le 6\sum_{l\neq k}\mu_{kl}^2 + 384M\lambda_k^2 + 608\,\Sigma_\lambda, \qquad b_k^2 \le 392\lambda_k^2 + 6\sum_{l\neq k}\mu_{kl}^2 + 384M\lambda_k^2 + 608\,\Sigma_\lambda.$$

By $(x+y)^2 \le 2x^2+2y^2$, the monotonicity of h on $[0, \frac12]$, and Jensen's inequality for the concave nondecreasing function $t \mapsto h(\min(t, \frac12))$,

$$S(\rho_A) \ \ge\ M - \frac{B}{\ln 2} - M\,h\Big(\min\Big(\frac{A}{2M}, \frac12\Big)\Big), \qquad A = \sum_k (a_k^2 + 4\lambda_k^2), \quad B = \sum_k (b_k^2 + 4\lambda_k^2).$$

Now $\mu_{kl}^2 \le 9\sum_{x\in\mathcal{T}_k, y\in\mathcal{T}_l}\eta_{xy}^2$, so $\sum_{k\neq l}\mu_{kl}^2 \le 18\sum_{x<y\in\mathcal{N}}\eta_{xy}^2 \le 594\binom{n'}{2}\bar\delta \le 297\,n'^2\bar\delta$, and $\Sigma_\lambda \le 6\big(\sum_{\text{internal}}\eta_{xy}^2 + \sum_{x\in\mathcal{N}}c_x^2\big) \le 6(33 + 2\sqrt2)\,n'\bar\delta \le 215\,n'\bar\delta$, where the internal pairs are those inside the triples. Hence, with $M \ge 1$ and $M = n'/3$,

$$A \le 6\cdot 297\,n'^2\bar\delta + 996\,M\Sigma_\lambda \le 73162\,n'^2\bar\delta, \qquad B \le 6\cdot 297\,n'^2\bar\delta + 1388\,M\Sigma_\lambda \le 101256\,n'^2\bar\delta.$$

With $\bar\delta \le 2\sqrt2\,\varepsilon$ this gives

$$S(\rho_A) \ \ge\ M\Big[1 - 1.24\cdot 10^6\, n'\varepsilon - h\big(\min(3.11\cdot 10^5\, n'\varepsilon, \tfrac12)\big)\Big].$$

Since $n'\varepsilon \le 10^{-7}$, the bracket is at least $1 - 0.124 - h(0.0311) > 0.67$, so $S(\rho_A) \ge M/2$. Finally $M \ge \min(n, 10^{-7}/\varepsilon)/3 - 1 \ge 10^{-7}\min(n, 1/\varepsilon)/3 - 1$. □

Corollary 11.4 (The ebit cost of CHSH(n))**.** *Let $S_{\min}(n,\varepsilon)$ be the least entanglement entropy $S(\rho_A)$ of an ε-optimal strategy for CHSH(n). For $n \ge 3$ and $0 < \varepsilon \le 1$,*

$$\frac{10^{-7}}{6}\min(n, 1/\varepsilon) - \frac12 \ \le\ S_{\min}(n,\varepsilon) \ \le\ \min\big(\lfloor n/2\rfloor, \lceil 1/\varepsilon\rceil\big).$$

So $S_{\min}(n,\varepsilon) = \Theta(\min(n, 1/\varepsilon))$ as $\min(n, 1/\varepsilon) \to \infty$.

Proof. The lower bound is Theorem 11.3. For the upper bound $\lceil 1/\varepsilon\rceil$, every XOR game has an ε-optimal strategy on a maximally entangled state of local dimension $2^{\lceil 1/\varepsilon\rceil}$ [OV18, Theorem 2]. For the upper bound $\lfloor n/2\rfloor$, let $m = \lfloor n/2\rfloor$; the Jordan–Wigner construction gives $2m+1 \ge n$ pairwise anticommuting involutions $A_1, \dots, A_n$ on $(\mathbb{C}^2)^{\otimes m}$. On the maximally entangled state Φ of local dimension 2^m, for which $\langle\Phi, (Y\otimes Z^T)\Phi\rangle = 2^{-m}\operatorname{Tr}(YZ)$, let Bob use $B_{ij} = ((A_i + A_j)/\sqrt2)^T$ and $B_{ji} = ((A_i - A_j)/\sqrt2)^T$ for $i < j$, which are involutions. Then $B_{ij} + B_{ji} = \sqrt2 A_i^T$ and $B_{ij} - B_{ji} = \sqrt2 A_j^T$, so $C_{ij} = 2\sqrt2$ for every pair and bias = bias*, with m ebits [Slo11, OV18]. □

The entropy bound does not use purity in an essential way: shared randomness cannot lower the entanglement of formation.

Corollary 11.5 (Entanglement of formation)**.** *Let σ be a state on $H_A \otimes H_B$ and let binary observables for CHSH(n), $n \geq 3$, achieve bias at least $(1-\varepsilon)\,\mathrm{bias}^*$ on σ. Then*

$$E_F(\sigma) \;\geq\; \frac{1}{4}\Big\lfloor \frac{1}{3}\min\Big(n, \frac{10^{-7}}{2\varepsilon}\Big)\Big\rfloor.$$

Proof. Let $\sigma = \sum_j p_j |\psi_j\rangle\langle\psi_j|$ be any decomposition into pure states. The bias is linear in the state, so $\mathrm{bias}(\sigma) = \sum_j p_j \,\mathrm{bias}(\psi_j)$ with the same observables, and $\mathrm{bias}(\psi_j) \leq \mathrm{bias}^*$. Put $\varepsilon_j = 1 - \mathrm{bias}(\psi_j)/\,\mathrm{bias}^* \geq 0$; then $\sum_j p_j \varepsilon_j \leq \varepsilon$, so the components with $\varepsilon_j \leq 2\varepsilon$ have total weight at least $\frac{1}{2}$. Each of them is a 2ε-optimal pure strategy, so its reduced state on H_A has entropy at least half the floor in the statement, by Theorem 11.3. Averaging over j, with entropy ≥ 0 for the other components, and taking the infimum over decompositions gives the claim. □

The strategies of Corollary 11.4 use pure states, for which E_F is the entanglement entropy, so the entanglement of formation needed is also $\Theta(\min(n, 1/\varepsilon))$. *Theorem E is Theorem 11.3 with Corollaries 11.4 and 11.5.*

Remark 11.6. The rate is linear in $1/\varepsilon$, which [OV18, Theorem 2] shows is optimal; the lower bound of [OV18, Theorem 1] is $\Omega(\varepsilon^{-1/5})$, with the same question sets of sizes n and $n(n-1)$. The argument of [OV18] also builds about $n/3$ pairs from triples, but converts approximate anticommutation into a qubit count with polynomial losses at each step. Here every error stays in the state-weighted ℓ^2 norm, and Theorem 4.1 is applied once. The threshold $n\varepsilon \approx 1$ at which the count stops is the scale at which the maximally entangled strategy of [OV18, Theorem 2], with $\lceil 1/\varepsilon \rceil$ ebits, becomes cheaper than the exact strategy with $\lfloor n/2 \rfloor$.

11.2 The Slofstra–Vidick group and game

Slofstra and Vidick built a linear system game G_{SV} with $C\varepsilon^{-1/6} \leq \mathcal{E}(G_{\mathrm{SV}}, \varepsilon) \leq C'\varepsilon^{-1/2}$ [SV18, Theorem 1.1] from the group [SV18, Proposition 3.2]

$$K = \big\langle a, b, c, x, y \;\big|\; xyx^{-1}y^{-2},\; xcx^{-1}c^{-1},\; yay^{-1}b^{-1},\; yby^{-1}a^{-1},\; c^{-1}ab,\; a^2,\; b^2,\; c^2 \big\rangle, \tag{21}$$

in which c is a central involution, and whose profile satisfies $C\varepsilon^{-2/3} \leq \mathrm{hlp}_K(c; \delta, \varepsilon) \leq C'/\varepsilon$. We show that both lower exponents are the upper ones: $\mathrm{hlp}_K(c; \delta, \varepsilon) = \Theta(1/\varepsilon)$ (Corollary 11.12) and $\mathcal{E}(G_{\mathrm{SV}}, \varepsilon) = \Theta(\varepsilon^{-1/2})$ (Theorem 11.14).

The picture is as follows. Let μ be the spectral measure of $\phi(y)$ in the state. The relation $xyx^{-1} = y^2$ makes μ nearly invariant under the doubling map $\vartheta \mapsto 2\vartheta$ of the circle, and, where c acts as -1, the relations $yay^{-1} = b$ and $c = ab$ make it nearly invariant under the half-turn $\vartheta \mapsto \vartheta + \pi$. Averaging over the two inverse branches of the doubling map contracts the Wasserstein distance by a factor 2, so a measure that is nearly invariant under both maps is close to Haar measure (Proposition 11.10), while a measure with d atoms is at distance at least $\pi/(2d)$ from it. All quantities are measured in one state; the Wasserstein-1 distance, rather than a quadratic one, is what makes a change of state cost only a linear error.

Measures on the circle

Let $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$ with the geodesic distance $d(\vartheta, \vartheta') \in [0, \pi]$, and let h be the normalized Haar measure. For probability measures μ, ν on $\mathbb{T}$ put

$$W_1(\mu, \nu) = \sup\Big\{ \int f\, d\mu - \int f\, d\nu \;:\; f : \mathbb{T} \to \mathbb{R} \text{ is 1-Lipschitz} \Big\}.$$

By Kantorovich–Rubinstein duality [Vil09, Chapter 5] this is the least mean distance of a coupling of μ and ν; we use only the triangle inequality and the elementary bound $W_1(\mu,\nu) \le \int d\,d\gamma$ for every coupling γ of μ and ν, which holds because $\int f\,d\mu - \int f\,d\nu = \int (f(\vartheta)-f(\vartheta'))\,d\gamma$. Let $\mathsf{D}(\vartheta) = 2\vartheta$ and $\mathsf{S}(\vartheta) = \vartheta + \pi$, with push-forwards D_*, S_*. The two square roots of $e^{i\vartheta}$ have angles $\vartheta/2$ and $\vartheta/2+\pi$, a set that does not depend on the representative of ϑ. Let P replace each point by the uniform measure on its two square roots: $\int f\,d(\mathsf{P}\mu) = \int \mathsf{P}^* f\,d\mu$ with $\mathsf{P}^* f(\vartheta) = \frac{1}{2}\big(f(\vartheta/2) + f(\vartheta/2+\pi)\big)$.

Lemma 11.7 (Doubling and half-turn)**.** *For probability measures μ,ν on $\mathbb{T}$:*

(a) $\mathsf{P}\mathsf{h} = \mathsf{h}$ *and* $\mathsf{P}(\mathsf{D}_*\mu) = \frac{1}{2}(\mu + \mathsf{S}_*\mu)$*;*

(b) $W_1(\mathsf{P}\mu, \mathsf{P}\nu) \le \frac{1}{2} W_1(\mu,\nu)$*;*

(c) $W_1(\mu, \mathsf{h}) \le W_1(\mathsf{D}_*\mu, \mu) + W_1(\mathsf{S}_*\mu, \mu)$*;*

(d) if μ is supported on at most d points, then $W_1(\mu,\mathsf{h}) \ge \pi/(2d)$*.*

Proof. (a) The maps $\vartheta \mapsto \vartheta/2$ and $\vartheta \mapsto \vartheta/2+\pi$, on a fundamental interval of length 2π, carry h to the uniform measures on two complementary half-circles, so $\int \mathsf{P}^* f\,d\mathsf{h} = \int f\,d\mathsf{h}$. The square roots of $e^{2i\vartheta}$ are ϑ and $\vartheta+\pi$.

(b) Let f be 1-Lipschitz and $d(\vartheta,\vartheta') = t$. Choose real representatives with $\vartheta' = \vartheta + s$ and $|s| = t \le \pi$. The roots $\vartheta/2$ and $(\vartheta+s)/2$, and the roots $\vartheta/2+\pi$ and $(\vartheta+s)/2+\pi$, are at distance at most $|s|/2$, so $|\mathsf{P}^* f(\vartheta) - \mathsf{P}^* f(\vartheta')| \le t/2$. Thus $2\mathsf{P}^* f$ is 1-Lipschitz and $\int f\,d(\mathsf{P}\mu - \mathsf{P}\nu) = \int \mathsf{P}^* f\,d(\mu-\nu) \le \frac{1}{2} W_1(\mu,\nu)$.

(c) By the triangle inequality, (a) and (b),

$$\begin{aligned} W_1(\mu,\mathsf{h}) &\le W_1(\mu, \mathsf{P}\mathsf{D}_*\mu) + W_1(\mathsf{P}\mathsf{D}_*\mu, \mathsf{P}\mu) + W_1(\mathsf{P}\mu, \mathsf{P}\mathsf{h}) \\ &\le \tfrac{1}{2} W_1(\mu, \mathsf{S}_*\mu) + \tfrac{1}{2} W_1(\mathsf{D}_*\mu, \mu) + \tfrac{1}{2} W_1(\mu,\mathsf{h}), \end{aligned}$$

where the first term was computed directly: $\int f\,d(\mu - \frac{1}{2}(\mu+\mathsf{S}_*\mu)) = \frac{1}{2}\int f\,d(\mu - \mathsf{S}_*\mu)$. Since $W_1(\mu,\mathsf{h}) \le \pi$ is finite, (c) follows.

(d) Let $k \le d$ be the number of points of $\operatorname{supp}\mu$. They cut $\mathbb{T}$ into k arcs of lengths $l_1,\dots,l_k$ with $\sum_j l_j = 2\pi$. The function $f(\vartheta) = d(\vartheta, \operatorname{supp}\mu)$ is 1-Lipschitz, $\int f\,d\mu = 0$, and, by Cauchy–Schwarz,

$$\int f\,d\mathsf{h} = \frac{1}{2\pi}\sum_j \int_0^{l_j} \min(t, l_j - t)\,dt = \sum_j \frac{l_j^2}{8\pi} \;\ge\; \frac{(2\pi)^2}{8\pi k} = \frac{\pi}{2k}. \qquad \square$$

The operator P is the transfer operator of the doubling map, and (b) is an averaged form of the square-root lifting that Slofstra and Vidick use in operator norm [SV18, Claims 3.6 and 3.7]. Part (d) needs no lower bound on the masses of the atoms.

For a unitary Y and a density matrix ρ on $\mathbb{C}^d$ let μ_Y^ρ be the *spectral measure of Y in ρ*, the probability measure on $\mathbb{T}$ with $\int f\,d\mu_Y^\rho = \operatorname{Tr}(\rho f(Y))$, where $f(Y)$ applies f to the angles of the eigenvalues. It is supported on the angles of the at most d eigenvalues of Y, and

$$\mu_{Y^2}^\rho = \mathsf{D}_*\mu_Y^\rho, \qquad \mu_{-Y}^\rho = \mathsf{S}_*\mu_Y^\rho, \qquad \mu_{UYU^*}^\rho = \mu_Y^{U^*\rho U} \quad (U \text{ unitary}). \tag{22}$$

Lemma 11.8 (Coupling in one state)**.** *Let P,Q,Y,U be unitaries on $\mathbb{C}^d$ and $\xi = \rho^{1/2}$.*

(a) There is a coupling ν of μ_P^ρ and μ_Q^ρ, viewed on the unit circle, with $\int |z-z'|^2\,d\nu(z,z') = \|P\xi - \xi Q\|_2^2$. Hence $W_1(\mu_P^\rho, \mu_Q^\rho) \le \frac{\pi}{2}\|P\xi - \xi Q\|_2$.

(b) $W_1(\mu_Y^{U^*\rho U}, \mu_Y^\rho) \le \pi \|[U,\xi]\|_2$*.*

Proof. (a) Let $P = \sum_i z_i E_i$ and $Q = \sum_j z'_j F_j$ be spectral decompositions with distinct eigenvalues, and $\nu = \sum_{i,j} \|E_i \xi F_j\|_2^2 \, \delta_{(z_i, z'_j)}$. The operators $E_i \xi F_j$ are mutually orthogonal in the Hilbert–Schmidt inner product. Since $\sum_j F_j = I$, $\sum_j \|E_i \xi F_j\|_2^2 = \|E_i \xi\|_2^2 = \mathrm{Tr}(E_i \rho)$, and likewise $\sum_i \|E_i \xi F_j\|_2^2 = \mathrm{Tr}(\rho F_j)$, so ν couples μ_P^ρ and μ_Q^ρ; and $\int |z - z'|^2 d\nu = \|\sum_{i,j} (z_i - z'_j) E_i \xi F_j\|_2^2 = \|P\xi - \xi Q\|_2^2$. The arc length between two points of the unit circle is at most $\frac{\pi}{2}$ times their chordal distance, and $\int |z - z'| \, d\nu \le (\int |z - z'|^2 d\nu)^{1/2}$.

(b) Let f be 1-Lipschitz and $f_0 = f - c_f$, with c_f the midpoint of the range of f, so that $|f_0| \le \pi/2$, the diameter of $\mathbb{T}$ being π. Both measures have mass one, so $\int f \, d(\mu_Y^{U^* \rho U} - \mu_Y^\rho) = \mathrm{Tr}\big((U^* \rho U - \rho) f_0(Y)\big) \le \frac{\pi}{2} \|U^* \rho U - \rho\|_1$, which is at most $\pi \|[U, \xi]\|_2$ by Lemma 3.7. □

The coupling in (a) is Connes' joint distribution trick [Con76, Lemma 1.2.6], in the form used in [SV18, Lemma 5.5]; for $\rho = I/d$ it gives the Hoffman–Wielandt inequality. Part (b) is the reason for working with W_1: a change of state by a unitary costs the state commutator linearly.

The profile of K

Theorem 11.9 (The profile of c in a state)**.** *Let ϕ assign unitaries A, B, C, X, Y on $\mathbb{C}^d$ to a, b, c, x, y, and let ρ be a density matrix on $\mathbb{C}^d$ with $\xi = \rho^{1/2}$. Suppose that, for some $e > 0$,*

(i) every relator r in (21) has $\|(\phi(r) - I)\xi\|_2 \le e$;

(ii) $\|[\phi(s), \xi]\|_2 \le e$ for $s \in \{a, b, c, x, y\}$;

(iii) $\|(C + I)\xi\|_2 \le e$.

Then $d \ge 1/(29e)$. In the notation of Definition 2.6, $\mathrm{hlp}_K^\rho(c; e) \ge 1/(29e)$.

The two relations are read off as near-invariances of $\mu = \mu_Y^\rho$, each up to a coupling in the state (Lemma 11.8(a)) and a change of state by a unitary (Lemma 11.8(b)); Lemma 11.7 then compares μ with Haar measure. That comparison does not involve the dimension, and we state it on its own.

Proposition 11.10 (The spectral measure of y in a state)**.** *Under the hypotheses of Theorem 11.9, the spectral measure of $Y = \phi(y)$ in ρ satisfies*

$$W_1(\mu_Y^\rho, \mathsf{h}) \;\le\; \frac{29\pi}{2}\, e.$$

The bound uses all three hypotheses: the state commutators (ii) pay for the changes of state in Lemma 11.8(b), and (iii), which says that c acts as -1 in the state, turns the relations $yay^{-1} = b$ and $c = ab$ into the half-turn.

Proof. Let $\mu = \mu_Y^\rho$, $\delta_{\mathsf{D}} = W_1(\mathsf{D}_* \mu, \mu)$ and $\delta_{\mathsf{S}} = W_1(\mathsf{S}_* \mu, \mu)$.

Doubling. Put $P = Y^2$ and $Q = XYX^*$. The relator $r = xyx^{-1}y^{-2}$ has $\phi(r) = QP^*$, so $(Q - P)\xi = (\phi(r) - I)P\xi = (\phi(r) - I)\xi P + (\phi(r) - I)[P, \xi]$, and $\|[P, \xi]\|_2 \le 2e$; hence $\|(Q - P)\xi\|_2 \le 5e$. Also $\|[Q, \xi]\|_2 \le 3e$, so $\|P\xi - \xi Q\|_2 \le \|(P - Q)\xi\|_2 + \|[Q, \xi]\|_2 \le 8e$. By Lemma 11.8(a) and (22), $W_1(\mathsf{D}_* \mu, \mu_Y^{X^* \rho X}) \le 4\pi e$, and by Lemma 11.8(b), $W_1(\mu_Y^{X^* \rho X}, \mu) \le \pi e$. So $\delta_{\mathsf{D}} \le 5\pi e$.

Half-turn. The word $w = a^{-1} y a y^{-1} c^{-1}$ is freely equal to

$$\big(a^{-1}(yay^{-1}b^{-1})a\big)\,(a^2)^{-1}\,\big(c(c^{-1}ab)c^{-1}\big),$$

a product of three conjugates of relators with conjugators of length 1, 0 and 1, so $\|(\phi(w) - I)\xi\|_2 \le 3e + e + 3e = 7e$ by Lemma 3.3(d). Since $A^* Y A = \phi(w) C Y$,

$$\|(A^* Y A - CY)\xi\|_2 = \|(\phi(w) - I) CY \xi\|_2 \le 7e + 2\|[CY, \xi]\|_2 \le 11e,$$

and $\|(CY+Y)\xi\|_2 = \|(C+I)Y\xi\|_2 \le \|(C+I)\xi\|_2 + 2\|[Y,\xi]\|_2 \le 3e$. So $\|(A^*YA+Y)\xi\|_2 \le 14e$ and $\|(-Y)\xi - \xi(A^*YA)\|_2 \le 14e + \|[A^*YA,\xi]\|_2 \le 17e$. By Lemma 11.8(a) and (22), $W_1(\mathsf{S}_*\mu, \mu_Y^{A\rho A^*}) \le \frac{17}{2}\pi e$, and by Lemma 11.8(b) with $U = A^*$, $W_1(\mu_Y^{A\rho A^*}, \mu) \le \pi e$. So $\delta_{\mathsf{S}} \le \frac{19}{2}\pi e$.

Conclusion. By Lemma 11.7(c), $W_1(\mu, \mathsf{h}) \le \delta_{\mathsf{D}} + \delta_{\mathsf{S}} \le \big(5 + \frac{19}{2}\big)\pi e = \frac{29}{2}\pi e$. □

Proof of Theorem 11.9. The measure μ_Y^ρ is supported on the angles of the at most d eigenvalues of Y, so by Lemma 11.7(d) and Proposition 11.10, $\frac{\pi}{2d} \le W_1(\mu_Y^\rho, \mathsf{h}) \le \frac{29}{2}\pi e$, that is $d \ge 1/(29e)$. Conditions (i)–(iii) are (S1)–(S3) for K and $w = c$, which gives the second sentence. □

For $\rho = I/d$ the state commutators vanish, and (i)–(iii) say that ϕ is an e-representation with $\|\phi(c) + I\|_f \le e$. Together with an explicit model this pins down the profile of c.

Example 11.11 (A model)**.** Let $d = 2m$, $z_j = e^{2\pi i j/d}$ for $0 \le j < d$, and on $\mathbb{C}^d$ put $Y = \mathrm{diag}(z_j)$, let A exchange e_j and e_{j+m} (indices mod d), $B = -A$, $C = -I$, and let X be the permutation matrix with $Xe_{\varpi(j)} = e_j$, where $\varpi(j) = 2(j \bmod m) + \lfloor j/m \rfloor$. Then $A^2 = B^2 = C^2 = I$ and $C^{-1}AB = A^2 = I$; $YAY^* = -A = B$ and $YBY^* = A$, because $z_{j+m}/z_j = -1$; and C is central. Finally $XYX^* = \mathrm{diag}(z_{\varpi(j)})$, so XYX^*Y^{-2} is I on the first m coordinates and $e^{2\pi i/d}I$ on the last m. With the flat state, every relator of (21) but $xyx^{-1}y^{-2}$ has defect 0, that one has defect $\|XYX^*Y^{-2} - I\|_f = \sqrt{2}\sin(\pi/d)$, and (ii), (iii) hold with 0. This is a variant of the model of [SV18, Lemma 3.8].

Corollary 11.12 (The profile of c)**.** *For every $e > 0$,*

$$\frac{1}{29e} \;\le\; \mathrm{hlp}_K^\rho(c;e) \;\le\; \mathrm{hlp}_K(c;2,e) \;\le\; 2\Big\lceil \frac{\pi}{\sqrt{2}\,e} \Big\rceil.$$

For each $0 < \delta \le 2$ there are constants $c_\delta, C_\delta > 0$ with $c_\delta/\varepsilon \le \mathrm{hlp}_K(c;\delta,\varepsilon) \le C_\delta/\varepsilon$ for $0 < \varepsilon \le 2$.

Proof. The first inequality is Theorem 11.9. The second holds for every presentation: an e-representation with $\|\phi(c) - I\|_f \ge 2$ has $\phi(c) = -I$ by (1), and with the flat state it satisfies (S1)–(S3). For the third, Example 11.11 with $m = \lceil \pi/(\sqrt{2}e) \rceil$ has $\phi(c) = -I$ and defects at most $\sqrt{2}\sin(\pi/(2m)) \le \sqrt{2}\pi/(2m) \le e$. For $\delta < 2$, $\mathrm{hlp}_K(c;2,C\varepsilon) \le \mathrm{hlp}_K(c;\delta,\varepsilon) \le \mathrm{hlp}_K(c;2,\varepsilon)$ for a constant $C \ge 1$ depending on δ [SV18, Proposition 2.7], since c is a central involution represented by a generator. So $\mathrm{hlp}_K(c;\delta,\varepsilon) \ge 1/(29C\varepsilon)$, and for $\varepsilon \le 2$, $\mathrm{hlp}_K(c;\delta,\varepsilon) \le 2(\pi/(\sqrt{2}\varepsilon) + 1) \le (\sqrt{2}\pi + 4)/\varepsilon$. □

Corollary 11.12 improves the lower bound $C\varepsilon^{-2/3}$ of [SV18, Lemma 3.5 and Proposition 3.2] to the order of their upper bound. Their argument passes from the normalized norm to the operator norm, which costs a factor $\sqrt{d}$; the argument above never uses the operator norm.

The game

Let $G_{\mathrm{SV}} = G(A,b)$ be the linear system game of [SV18, Theorem 1.1], with m equations, n variables, $s = \max_i |V_i|$ and solution group $\Gamma = \Gamma(A,b)$, the solution group of [SV18, Proposition 6.1]. Its question distribution is that of [SV18, Section 4]: the equation i and the variable j are independent and uniform, and a pair with $j \notin V_i$ is always won; so $\pi_{\min} = 1/(mn)$ (Definition 2.7). The construction [SV18, Propositions 6.1 and 6.2] applies Slofstra's embedding theorem [Slo20], in the form of [Slo19, §4], to K, with distinguished central involution c. It supplies a homomorphism $\Psi : F(a,b,c,x,y) \to F(x_1,\dots,x_n,J)$ that descends to an embedding of K into Γ, and a word z in $x_1,\dots,x_n,J$, the image of an added involution t with $tct = cJ$, such that $[\Psi(c), z] = J$ in Γ [SV18, proof of Proposition 6.2]. (The proof there prints

the relation as $twt = J$, for the distinguished involution w; the relation used is $twt = wJ$, that of [Slo19, Lemma 5.2], which is what gives $[w, t] = J$.) For the lower bound we use only that Ψ descends to a homomorphism $K \to \Gamma$ and this identity. Since J is nontrivial in approximate representations of Γ [SV18, Proposition 6.1], $\omega_q(G_{\mathrm{SV}}) = 1$ (see the discussion after Definition 2.7), so $\mathcal{E}(G_{\mathrm{SV}}, \varepsilon)$ counts strategies that win with probability at least $1 - \varepsilon$. Qualitatively, this nontriviality comes from Slofstra's embeddings and his tensor-power argument [Slo19, Lemma 5.2]; Lemma 11.13 below gives it again, with rates.

The upper bound $\mathcal{E}(G_{\mathrm{SV}}, \varepsilon) \le C'\varepsilon^{-1/2}$ of [SV18, Theorem 1.1] rests on the second half of the proof of [SV18, Proposition 6.2]. That argument takes from [Slo19, Proposition 4.8] a $2d$-dimensional lift of a d-dimensional approximate representation of K with $c \mapsto (-I_d) \oplus I_d$, and then lets t exchange the two blocks. But [Slo19, Proposition 4.8] removes the non-involutive generators one at a time, and each removal doubles the dimension and sends every involutive generator s to the image of s plus an identity block. The group K has two non-involutive generators, $y < x$, and no presentation of K in the sense of [SV18, §6] has fewer: the quotient of K by a, b, c is $\langle x, y \mid xyx^{-1} = y^2 \rangle$, which is torsion-free and not cyclic, so it is also the quotient of K by all its involutions, and that quotient is generated by the non-involutive generators of any such presentation. So the two lifts give $c \mapsto W = (-I_d) \oplus I_{3d}$. For every unitary T on $\mathbb{C}^{4d}$ the $(+1)$-eigenspaces of W and TWT^* have dimension $3d$, so they meet in dimension at least $2d$, where $TWT^* + W = 2I$; hence $\|TWT^* + W\|_f^2 \ge 2$, and the relation $tct = cJ$ with $J \mapsto -I$ fails by at least $\sqrt{2}$, whatever the precision of the representation. So for K this last step, as written in [SV18] (arXiv:1711.10676v2), does not produce the representation it needs; with at most one non-involutive generator it is complete, and that is the case of Theorem B.3. This affects only the upper bound $\mathrm{hlp}(J; 2, \varepsilon) \le N\ \mathrm{hlp}(c; 2, C'\varepsilon)$ of [SV18, Prop. 6.2]. Its lower bound, the nontriviality of J and the value one of the game do not use this step: they use only that the embedding of [Slo19, Prop. 4.8] keeps c nontrivial in approximate representations, then [Slo19, Lemma 5.2 and Prop. 4.2] and, for the value, [SV18, Cor. 5.3]. For K the upper bound itself holds: Lemma 11.13 below supplies a balanced lift for the explicit models of K, and with it the upper bounds of [SV18, Prop. 6.1 and Thm. 1.1] hold as stated. For the models of Example 11.11, in which X is a permutation matrix, the second removal can be done without doubling, and the marked involution stays balanced.

Lemma 11.13 (A balanced lift for K)**.** *Let $m \ge 1$, $d = 2m$, and let ϕ be the model of Example 11.11 on $\mathbb{C}^d$, whose relators have defect at most $e_m = \sqrt{2}\sin(\pi/(2m))$. Then the solution group Γ of G_{SV} has a $C_\Gamma e_m$-representation γ on $\mathbb{C}^{8d}$ with $\gamma(J) = -I$, where $C_\Gamma \le 75$ is the constant of [Slo19, Remark 4.5].*

Proof. We follow the construction of Γ in [SV18, proof of Proposition 6.2] and lift ϕ through each of its steps. Write A, B, C, X, Y for the images of a, b, c, x, y, and use the notation Ω, $U \oplus V$ and $\mathrm{adiag}(U, V)$ of Appendix B.1. As an extended presentation in the sense of [SV18, §6], K has the involutive generators a, b, c, the linear set $\{a, b, c\}$ with its commutations, which Slofstra and Vidick add and the model satisfies exactly, the conjugacy relations $yay^{-1} = b$, $yby^{-1} = a$ and $xcx^{-1} = c$, and the non-involutive generators $y < x$ with the relation $xyx^{-1} = y^2$. In the model only the relator $r = xyx^{-1}y^{-2}$ has nonzero defect, and that defect is e_m.

Step 1: removing y [Slo19, Proposition 4.8]. Add involutions f, v, with $y = fv$ in mind, and the relation that f commutes with x. Replace $yay^{-1} = b$ by $vav = Z_a$ and $fZ_af = b$, and $yby^{-1} = a$ by $vbv = Z_b$ and $fZ_bf = a$, with ancilla involutions Z_a, Z_b. Once $y = fv$ and $fx = xf$, the relation $xyx^{-1} = y^2$ reads $xvx^{-1} = vfv$; replace it by $vfv = W_0$ and $xvx^{-1} = W_0$, with an ancilla involution W_0. Then y occurs only in $y = fv$, and is deleted. On $\mathbb{C}^d \oplus \mathbb{C}^d$ put

$$\gamma_1(s) = \phi(s) \oplus I \quad (s = a, b, c), \qquad \gamma_1(f) = \Omega, \qquad \gamma_1(v) = \mathrm{adiag}(Y^*, Y), \qquad \gamma_1(x) = X \oplus X,$$

and give Z_a, Z_b, W_0 their defining values $\gamma_1(vav)$, $\gamma_1(vbv)$, $\gamma_1(vfv)$. This is the lift of [Slo19, Proposition 4.8], with $\gamma_1(f)\gamma_1(v) = Y \oplus Y^*$. The relations that define the ancillas hold exactly, and so do $f^2 = v^2 = e$ and $fx = xf$. Every other relator is mapped to a block-diagonal matrix each of whose two blocks is the identity or a conjugate of $\phi(r')^{\pm 1}$ for a relator r' of K: for instance $\gamma_1(fZ_afb^{-1}) = \phi(yay^{-1}b^{-1}) \oplus I$, and $\gamma_1(xvx^{-1}W_0^{-1})$ is conjugate to $\phi(r) \oplus Y^{-2}\phi(r)^{-1}Y^2$. So γ_1 is an e_m-representation of the presentation reached, and $\gamma_1(c) = (-I_d) \oplus I_d$.

Step 2: removing x without doubling. Now x is the only non-involutive generator. The construction of [Slo19, Proposition 4.8] adds involutions f', v', with $x = f'v'$ in mind, and replaces each of the three relations $xsx^{-1} = s'$, namely $(s, s') = (c, c)$, (f, f) and (v, W_0), by $v'sv' = Z'_s$ and $f'Z'_sf' = s'$, with an ancilla involution Z'_s; it adds no commutation relations, since no non-involutive generator follows x, and then deletes x. The lift used there would double the dimension again. Instead we use that X is a permutation matrix. Reflecting each cycle of the permutation, $a_i \mapsto a_{-i}$ on a cycle $(a_0\, a_1 \cdots a_{k-1})$, gives an involutive permutation matrix τ with $\tau X \tau = X^{-1}$. Put

$$\gamma_2(f') = \tau \oplus \tau, \qquad \gamma_2(v') = \tau X \oplus \tau X, \qquad \gamma_2(Z'_s) = \gamma_2(v')\gamma_1(s)\gamma_2(v'),$$

and $\gamma_2(s) = \gamma_1(s)$ for the other generators. The matrix τX is real, and $(\tau X)^* = X^{-1}\tau = \tau X$ and $(\tau X)^2 = X^{-1}X = I$, so $\gamma_2(f')$ and $\gamma_2(v')$ are involutions, and $\gamma_2(f')\gamma_2(v') = X \oplus X = \gamma_1(x)$. Hence $\gamma_2(f'Z'_sf's'^{-1}) = \gamma_1(xsx^{-1}s'^{-1})$: every relation keeps its residual. The relations $v'sv' = Z'_s$ hold exactly, and ${Z'_s}^2$ is conjugate to s^2. So γ_2 is an e_m-representation, on $\mathbb{C}^{2d}$, of a presentation all of whose generators are involutions, and still $\gamma_2(c) = (-I_d) \oplus I_d$.

Step 3: the commutator [SV18, proof of Proposition 6.2]. Adjoin the central involution J and an involution t with $tct = cJ$. In linear-plus-conjugacy form this takes one more involution u, with the linear relation $cu = J$ and the conjugacy relation $tct = u$, as in Step 2 of the proof of Lemma B.1; Slofstra and Vidick leave this encoding implicit, and any encoding through auxiliary involutions that are determined by t, c and J works the same way. Put $\gamma_3(J) = -I$, let $\gamma_3(t) = \Omega$ exchange the two copies of $\mathbb{C}^d$, and put $\gamma_3(u) = -\gamma_2(c) = I_d \oplus (-I_d)$. Because $\gamma_2(c) = (-I_d) \oplus I_d$ has its two eigenvalues with equal multiplicities, $\gamma_3(t)\gamma_3(c)\gamma_3(t) = \gamma_3(u)$; also $\gamma_3(c)\gamma_3(u) = -I = \gamma_3(J)$, and $-I$ commutes with everything. So all the new relations hold exactly, and γ_3 is an e_m-representation of the resulting linear-plus-conjugacy presentation on $\mathbb{C}^{2d}$ with $J \mapsto -I$.

Step 4: the solution group [Slo19, Lemma 4.4 and Proposition 4.2]. These two steps turn the linear-plus-conjugacy presentation into the solution group Γ. By [Slo19, Remark 4.5] they turn every $2d$-dimensional e-representation of it into an $8d$-dimensional $C_\Gamma e$-representation of Γ that restricts to four copies of the given one on its generators, with $C_\Gamma \le 75$. Applied to γ_3, this gives γ, and $\gamma(J) = -I$. □

The program `anc/sv_balanced_lift.py` builds these four steps for $m \le 8$ and evaluates every relator. Steps 1–3 keep every defect at most e_m, and all 2373 relators of the solution group so obtained, with the products of the equations taken in every order, have defect at most $e_m/\sqrt{2}$, far below $75e_m$. As a negative control it builds the two literal lifts and confirms the bound $\|TWT^* + W\|_f^2 \ge 2$, which a block permutation attains (Appendix C.3).

Theorem 11.14 (Entanglement of the Slofstra–Vidick game)**.** *There are constants $c_0, C_0 > 0$ such that $c_0\,\varepsilon^{-1/2} \le \mathcal{E}(G_{\mathrm{SV}}, \varepsilon) \le C_0\,\varepsilon^{-1/2}$ for $0 < \varepsilon \le 1$.*

For the upper bound, Lemma 11.13 and the return direction with the flat state give strategies on a maximally entangled state of local dimension $16m$ that lose with probability $O(1/m^2)$. For the lower bound, a near-optimal strategy gives, by Theorem 5.1, a model of Γ in Bob's reduced state at precision $O(\sqrt{\varepsilon})$, in which $J = -I$. Composed with Ψ it is a model of K in

which c is visible, by the identity $[\Psi(c), z] = J$; on the sector where the rounding of $\phi(\Psi(c))$ is -1 it satisfies the hypotheses of Theorem 11.9, which bounds the Schmidt rank by $\Omega(\varepsilon^{-1/2})$.

Proof. The upper bound. Let C_s be the constant of Theorem 5.4 and C_Γ that of Lemma 11.13. For $m \geq 1$, Lemma 11.13 gives a $C_\Gamma e_m$-representation of Γ on $\mathbb{C}^{16m}$ with $J \mapsto -I$, where $e_m = \sqrt{2}\sin(\pi/(2m)) \leq \pi/(\sqrt{2}\,m)$. By Theorem 5.4 with the flat state, the strategy built there on Φ_{16m} loses every question pair, hence a random one, with probability at most $C_s C_\Gamma^2 e_m^2 \leq \pi^2 C_s C_\Gamma^2/(2m^2)$. Given $0 < \varepsilon \leq 1$, take $m = \lceil \pi C_\Gamma \sqrt{C_s/(2\varepsilon)}\,\rceil$. Then the losing probability is at most ε, and since $\varepsilon \leq 1$,

$$\mathcal{E}(G_{\mathrm{SV}}, \varepsilon) \;\leq\; \mathcal{E}^{\mathrm{me}}(G_{\mathrm{SV}}, \varepsilon) \;\leq\; 16m \;\leq\; 16\Big(\pi C_\Gamma \sqrt{C_s/2} + 1\Big)\varepsilon^{-1/2}.$$

This is the upper bound of [SV18, Theorem 1.1], with $C_0 = 16(\pi C_\Gamma \sqrt{C_s/2} + 1)$; it also shows $\omega_q(G_{\mathrm{SV}}) = 1$ once more.

The lower bound. Take a strategy that wins with probability at least $1 - \varepsilon$, with Schmidt rank r, and let ϕ and ρ on $\mathbb{C}^r$ be as in Theorem 5.1, with $\theta = 2\sqrt{mn\varepsilon}$ and $e_0 = \max(s, 4)\theta$. Put $\phi_K = \phi \circ \Psi$ and $W = \phi_K(c)$.

Constants. The following words are trivial in Γ: $\Psi(r)$ for each relator r of (21); $\Psi([u, c])$ for $u \in \{a, b, c, x, y\}$, as c is central in K; and $[\Psi(c), z]J^{-1}$. Fix for each of them an expression as a product of conjugates of relators of Γ. These finitely many expressions do not depend on ε or on the strategy. By Lemma 3.3(d),(a) and Theorem 5.1, which bounds the defects of the relators of Γ by e_0 and the state commutators of the letters x_j by $\sqrt{2}\theta$ (and $\phi(J) = -I$ commutes with ξ), there is a constant C_1, depending only on G_{SV} and the chosen expressions, such that each of these words has defect $\|(\phi(\cdot) - I)\xi\|_2 \leq C_1\theta$, and $\|[\phi_K(u), \xi]\|_2,\ \|[\phi(z), \xi]\|_2 \leq C_1\theta$ for $u \in \{a, b, c, x, y\}$.

Visibility. Let $V = \phi(z)$ and $\tau = \|(W - I)\xi\|_2$. Since $\phi(J) = -I$, $\phi([\Psi(c), z]J^{-1}) = -WVW^*V^*$, so $\|(WVW^*V^* + I)\xi\|_2 \leq C_1\theta$. From $WVW^*V^* - I = (W - I) + WV(W^* - I)V^*$,

$$\|(WVW^*V^* - I)\xi\|_2 \leq \tau + \|(W^* - I)(\xi V^* + [V^*, \xi])\|_2 \leq 2\tau + 2\|[V, \xi]\|_2 \leq 2\tau + 2C_1\theta.$$

As $2 = \|2\xi\|_2$, we get $\tau \geq 1 - \frac{3}{2}C_1\theta$.

The sector. Let $J' = \mathrm{r}(W)$. By Lemma 3.6 and the relator c^2, $\|(J' - W)\xi\|_2 \leq C_1\theta$ and $\|[J', \xi]\|_2 \leq 3C_1\theta$. Let $P_- = \frac{1}{2}(I - J')$ and $p_- = \mathrm{Tr}(P_-\rho) = \frac{1}{4}\|(I - J')\xi\|_2^2$. Then $\|(I - J')\xi\|_2 \geq \tau - C_1\theta \geq 1 - \frac{5}{2}C_1\theta$, so $p_- \geq \frac{1}{16}$ when $\theta \leq 1/(5C_1)$; assume this. For $u \in \{a, b, c, x, y\}$ and $U = \phi_K(u)$, the group commutator $UWU^*W^* = \phi(\Psi([u, c]))$ has defect at most $C_1\theta$, so $\|[U, W]\xi\|_2 = \|(UWU^*W^* - I)WU\xi\|_2 \leq C_1\theta + 4C_1\theta$. With $N = J' - W$, $\|N\|_{\mathrm{op}} \leq 2$ and $\|[U, N]\xi\|_2 \leq \|N\xi\|_2 + \|N\xi\|_2 + \|N\|_{\mathrm{op}}\|[U, \xi]\|_2 \leq 4C_1\theta$, so $\|[U, J']\xi\|_2 \leq 9C_1\theta$.

The profile in the sector. Let ρ_- and ξ_- be as in Lemma 3.10, with $\sqrt{p_-} \geq \frac{1}{4}$. By Lemma 3.10(c), $\|[U, \xi_-]\|_2 \leq 4(9 + 3 + 3)C_1\theta = 60C_1\theta$ for $U = \phi_K(u)$. By Lemma 3.10(b) with $\|\phi_K(r) - I\|_{\mathrm{op}} \leq 2$, every relator of (21) has $\|(\phi_K(r) - I)\xi_-\|_2 \leq 4(C_1 + 6C_1)\theta = 28C_1\theta$. Finally $(J' + I)P_- = 0$, so $(J' + I)\xi_- = (J' + I)(\xi_- - P_-\xi P_-/\sqrt{p_-})$, of norm at most $\|[J', \xi]\|_2/\sqrt{p_-}$ by Lemma 3.10(a), and

$$\|(W + I)\xi_-\|_2 \leq \|(W - J')\xi_-\|_2 + \|(J' + I)\xi_-\|_2 \leq 4(C_1 + 6C_1)\theta + 12C_1\theta = 40C_1\theta.$$

So ϕ_K and ρ_- satisfy (i)–(iii) of Theorem 11.9 on $\mathbb{C}^r$ with $e = 60C_1\theta$, and $r \geq 1/(1740\,C_1\theta) = \varepsilon^{-1/2}/(3480\,C_1\sqrt{mn})$.

If $\theta > 1/(5C_1)$, that is $\varepsilon > \varepsilon_0 = 1/(100C_1^2 mn)$, then $\mathcal{E}(G_{\mathrm{SV}}, \varepsilon) \geq 1 \geq \sqrt{\varepsilon_0}\,\varepsilon^{-1/2}$. So the lower bound holds for all $\varepsilon > 0$ with $c_0 = \min\big(1/(3480C_1\sqrt{mn}), \sqrt{\varepsilon_0}\big)$, since a strategy of local dimension d has Schmidt rank at most d. □

Remark 11.15. The lower bound $C\varepsilon^{-1/6}$ of [SV18, Theorem 1.1] passes from the state to a spectral projection of it, and from the normalized norm to the operator norm; the proof

above keeps the state throughout. The bound is on the Schmidt rank, and so on the local dimension: $\log \mathcal{E}(G_{\mathrm{SV}}, \varepsilon) = \frac{1}{2}\log(1/\varepsilon) + O(1)$. It is not a bound on the entanglement entropy, since Theorem 11.9 counts the atoms of a spectral measure, whose masses may be unequal.

Theorem F is Corollary 11.12 with Theorem 11.14.

12 Open questions

The exact scale. For the game of Theorem A,

$$c\,\varepsilon^{-1}\beta(\varepsilon) \;\le\; \log \mathcal{E}(G,\varepsilon) \;\le\; \log \mathcal{E}^{\mathrm{me}}(G,\varepsilon) \;\le\; C\varepsilon^{-1}\log(1/\varepsilon).$$

The factor $\beta(\varepsilon) = 2^{-O((\log\log(1/\varepsilon))^2)}$ comes only from the pair areas Λ_N, that is from the Dehn function of H_3, and the factor $\log(1/\varepsilon)$ only from the number of affine hashes over which the models of the upper bound average. Under the polynomial bound (H2), which we do not use, $\beta(\varepsilon)$ improves to a power of $\log(1/\varepsilon)$ (Remark 8.3).

Question 12.1. *Is $\log \mathcal{E}(G,\varepsilon)$ of order $\varepsilon^{-1}\log(1/\varepsilon)$, or of order ε^{-1}? In G_{P}, is $\log \mathrm{hlp}(J;2,\varepsilon)$ of order $\varepsilon^{-2}\log(1/\varepsilon)$?*

The lower bound sees $2^N \approx \varepsilon^{-1}$ lamps on $N \approx \log(1/\varepsilon)$ points, one qubit each. The upper bound keeps about ε^{-1} qubits as well, but averages over a family of about 2^{6mk} affine hashes, with $k \approx \log(1/\varepsilon)$ bits, and the size of that family is where the factor $\log(1/\varepsilon)$ comes from. Deciding which count is right seems to need either a lower bound that also charges for the points a lamp uses, or models built from a much smaller family of hashes.

The Dehn function of Houghton's group. The proof of Corollary A.5 halves Lee's recursion for the commutator of two transpositions at distance M and obtains $M^{(1+o(1))\log M}$. The Dehn function of H_3 is at least quadratic: H_3 is not hyperbolic, since its finitary permutations have unbounded finite orders, and a finitely presented group with subquadratic Dehn function is hyperbolic [Bri02].

Question 12.2. *Is the Dehn function of H_3 quadratic, improving the polynomial bound (H2) of our preprint [DM26b]? In particular, do the commutators $[\sigma_0, \sigma_M]$ of far-apart transpositions have area $O(M^2)$?*

A quadratic bound would give $\Lambda_N = O(N^4)$ and replace $\beta(\varepsilon)$ by $(\log(1/\varepsilon))^{-O(1)}$; it would not touch the factor $\log(1/\varepsilon)$ of Question 12.1.

Faster growth. By Π_2-completeness [MNY22], every total computable bound on $\mathcal{E}(G,\varepsilon)$ in terms of the game and the precision is exceeded by some game of quantum value one, but the argument exhibits no such game. The exponent 1 of Theorem A is the natural limit of the present method: a strategy can only certify about ε^{-1} independent qubits at precision $\sqrt{\varepsilon}$ when the errors add in squares, so faster growth must come from a mechanism other than counting commuting Pauli pairs, for instance from lamps whose state space itself grows with the precision.

Question 12.3. *Is there an explicit two-player game G of quantum value one such that $\log\log \mathcal{E}(G,\varepsilon) \ge \varepsilon^{-c}$ for some $c > 0$? Can it be a linear system game whose solution group is amenable?*

For amenable solution groups the second part asks for a central involution J whose state-weighted profile is at least doubly exponential; by Theorem D this is exactly what such a game needs.

Keeping the state. By Theorem D the entanglement of a linear system game is the state-weighted profile of J, and the entanglement of maximally entangled strategies is the flat profile $\mathrm{hlp}(J;2,\cdot)$, at the same precision $\sqrt{\varepsilon}$ up to constants. In all examples we know, the logarithms of the two profiles agree up to factors $(1/e)^{o(1)}$ as the precision $e \to 0$: for G_{P} by Theorem B, and for the group K of Slofstra and Vidick, where the profiles themselves agree up to constant factors, by Theorem F.

Question 12.4. *Is there a finitely presented group with a central involution J for which $\log \mathrm{hlp}^{\rho}(J;e)$ is much smaller than $\log \mathrm{hlp}(J;2,e)$, say $\log \mathrm{hlp}^{\rho}(J;e) \le (\log \mathrm{hlp}(J;2,e))^{1-c}$? If such a group can be the solution group of a linear system game of quantum value one, then, by Theorem D and [SV18, Cor. 4.6], strategies of that game far from maximally entangled save a power of the entanglement.*

A smaller game. The game has 1687 equations because it encodes, through Slofstra's embedding, an $E\Gamma_0$ presentation, with 27 generators, of a group that maps onto G_{P}^{+}. Most of the equations are gadgets: each conjugacy relation of the presentation costs fourteen equations. The constants of Theorem A are explicit but astronomically small, and a smaller game would improve them too.

Question 12.5. *What is the least number of equations of a linear system game, or of questions of a two-player game, of quantum value one in which $\mathcal{E}(G,\varepsilon)$ grows faster than every polynomial in $1/\varepsilon$?*

Other measures of entanglement. The lower bounds of Theorems A and E hold for the entanglement entropy of pure strategies and for the entanglement of formation of mixed ones. For mixed states the entanglement of formation is at least the entanglement cost, the distillable entanglement and the relative entropy of entanglement, so lower bounds on it do not bound these.

Question 12.6. *Do near-optimal mixed strategies for CHSH(n), or for the game of Theorem A, have distillable entanglement of the same order as their entanglement of formation?*

How this work was done

Roles. This paper is the work of its two authors together with AI models from more than one developer. The authors set the direction: they chose the question, fixed the scope, decided which lines to pursue and which to drop, set the verification protocol, and take full responsibility for every claim. The models did most of the exploration, derivation, computation and drafting. For the group and the game they proposed constructions and proof strategies, wrote and checked proofs, and carried out the computations. They found the qubit count of Section 4, the transfer and return arguments of Section 5, the rate for CHSH(n) and the Wasserstein-1 argument for the group K of Slofstra and Vidick. They also wrote the verification programs, searched the literature, and drafted and edited the text.

Sources. The paper merges and supersedes two earlier unpublished drafts of ours. The first built the game and proved that every strategy needs local dimension $\exp(\varepsilon^{-1/2+o(1)})$, and every strategy on a maximally entangled state $\exp(\varepsilon^{-1+o(1)})$; the second developed the state-weighted method and closed that gap. This version proves every step from the definitions, apart from results of other authors that are cited with their numbers. In particular it reproves the facts about Houghton's group that it uses (Appendix A), and it does not depend on our preprints [DM26a, DM26b], which appear only in remarks and credits. Material of the drafts

that the main results do not need has been left out: the general question of when Slofstra's embedding transfers approximate representations, a comparison map between the solution groups of different games, the entropy needed by the Slofstra–Vidick game, and a discussion of the three-player game of Ji, Leung and Vidick.

Verification. The checks fall into three kinds.

- *Machine checks of finite statements.* The programs described in Appendix C check the following:

  - the presentation $\mathcal{P}$ and its model in G_{P}^{+};
  - the 29 derivations behind Proposition 10.3(d), with the area bound $A_* = 60$;
  - the 27 words of R_{P};
  - the linear system of the game;
  - the area certificates behind $A_0 = 42$.

  Each statement is checked by at least two programs that share no code, and every checking program runs negative controls; Appendix C says which program establishes what. Their recorded output is in `anc/`.

- *Numerical attacks on the analytic statements.* A program in the ancillary files, `anc/numerics/analytic_checks.py`, evaluates both sides of the qubit count (Theorem 4.1), the rounding and sequential-measurement lemmas, the loss bound of the return theorem (Theorem 5.4) on perturbed models of the magic square game, and the bound of Theorem 11.9, on random and optimized instances with flat, full-rank, low-rank and pure states, and runs a negative control showing that the state commutators of (S2) cannot be dropped from the return theorem. No inequality was violated. In the earlier drafts the triple construction for CHSH(n) and the Wasserstein-1 contraction for K were attacked in the same way. A further program, `anc/sv_balanced_lift.py`, builds the lift of Lemma 11.13 for small m, evaluates every relator, and confirms, as a negative control, that the literal lifts cannot be balanced.

- *Reading in the earlier drafts.* The qubit count, the transfer theorem, the argument for CHSH(n) and a first form of the argument for K were proof-read by a model instance other than the one that derived them. They were then reviewed blind, refutation first, by a model of a different developer that had not seen our working files. That review found no counterexample to any principal statement and two repairable gaps, which are repaired here. The group-theoretic part was checked adversarially in the first draft.

- *Reading of this version.* Two model instances that had not taken part in writing this version read it in full, refutation first: one read the analytic half and one the group-theoretic half. They re-derived the proofs, recomputed the constants, checked the cited results of other papers against their sources, and attacked the analytic lemmas numerically. They also rebuilt the presentation and the game independently of the shipped programs. They found no error in a main statement. They corrected the following slips, none of which affects the main results:

  - an intermediate defect bound in the proof of Lemma B.1;
  - the range of ε in Theorem 11.14 and in Theorem E;
  - a remark claiming a case of equality;
  - the base of a logarithm in the introduction.

  They also found that 17 generators of an earlier version of $\mathcal{P}$ played no role. These have been removed, which shrinks the game from 2061 to 1687 equations.

- *A blind audit.* A model of another developer, which had seen none of our working files, then audited this version refutation first. It re-derived the main arguments, including the return direction (Theorem 5.4), which no model of another developer had read before; reran every program; and checked every reference against its primary source. It found no counterexample to a main statement. It found that the proof of the upper bound of [SV18, Thm. 1.1], which Theorem F cites, uses a lift from [Slo19, Prop. 4.8] that, for a group with two non-involutive generators such as K, does not leave the marked involution balanced as the last step of that proof needs, and Lemma 11.13 now supplies a balanced lift for the explicit models of K; that our account of the machine checks claimed more than the programs established, which led to the second cell-count verifier for $A_* = 60$ and to the cap of 42 factors in both verifiers of (C4); and several smaller points. All are corrected here.

Formal verification. The accompanying Lean 4.35.0 development in `lean/` checks the aggregate Theorem A, the conjugator-length bound of Lemma 3.4, the finite-dimensional qubit count and its uniform-error consequence (Theorem 4.1 and Corollary 4.2), transfer and prescribed-measurement return (Theorems 5.1 and 5.4), the three profile comparisons of Corollary 5.5, and the stated aggregates for Theorems E and F. The sole parameter of the Theorem A aggregate is the implication from a perfect projective tensor-product strategy to a finite-dimensional operator solution, restricted to the linear systems defined in the development: Definition 1 and Theorem 5, (1)⇒(2), of Cleve–Liu–Slofstra, arXiv:1606.02278v1 [CLS17]. Its proof is not formalized here; the changes from spatial POVMs to this projective formulation are proved in Lean. The certified roots have only the standard axioms of propositional extensionality, classical choice and quotient soundness; their selected proof export was replayed by `leanchecker` and `nanoda`. The release packaging itself does not constitute a fresh replay. Appendix D gives the declaration map and scope, including the absence of a standalone certified Theorem B and of a formalization of the whole manuscript. Build and replay instructions are in `lean/README.md` of the accompanying repository.

A A quasi-polynomial Dehn bound for Houghton's group

This appendix proves the bound (H1), $\delta_H(n) \le 2^{(\log n+6)^2}$, on the Dehn function of the presentation P_H of Houghton's group in (6), and with it that P_H presents $H = H_3$. We work in Johnson's presentation P_J, which has the extra generator $\alpha = [p, q]$, and first check that its Dehn function bounds that of P_H (Lemma A.1). Every null word is then collected, at polynomial cost, into a word in the adjacent transpositions σ_j of ray 1, which normal forms in a symmetric group reduce to the empty word (Proposition A.4). The only expensive moves are commutations of far-apart transpositions $[\sigma_0, \sigma_m]$, $m \ge 2$. Lee's recursion [Lee12, Section 4.2] bounds their areas through areas at smaller distances and gives an exponential bound; splitting the distance in half gives $m^{(1+o(1))\log m}$ (Proposition A.3). The recursion is that of [DM26b, Section 3] and the collection argument that of [DM26b, Theorem 3.3]; both are proved here in full. We use the notation of Section 6.1.

A.1 Johnson's presentation

Words in the letters $p^{\pm1}, q^{\pm1}, \alpha^{\pm1}$ are evaluated in H with $\alpha \mapsto s = [p, q]$, and such a word is *null* if it represents 1. Johnson's presentation [Joh99] (see [Lee12, Theorem 2.8]), written in our generators, is

$$P_J = \big\langle p, q, \alpha \;\big|\; r_1 = \alpha^2,\ r_2 = (\alpha p \alpha p^{-1})^3,\ r_3 = [\alpha, p^2 \alpha p^{-2}],\ r_4 = [p, q]\alpha^{-1},\ r_5 = p\alpha p^{-1} q \alpha^{-1} q^{-1} \big\rangle. \quad (23)$$

Its relators are null: r_4 by definition, and r_1, r_2, r_3, r_5 are the relators of P_H with α in place of s, which hold in H (Section 6.1). Let Area_J be the area with respect to $\{r_1, \dots, r_5\}$, and let $\delta_J(n)$ be the largest Area_J of a null word of length at most n; a priori it might be infinite. We do not assume that P_J presents H.

Lemma A.1 (From P_J to P_H)**.** *Every word w in $p^{\pm1}, q^{\pm1}$ that represents 1 in H satisfies $\mathrm{Area}_{P_H}(w) \le \mathrm{Area}_J(w)$. Hence $\delta_H(n) \le \delta_J(n)$ for every n.*

Proof. Let θ be the homomorphism from the free group on p, q, α to the free group on p, q that fixes p and q and sends α to $s = [p, q]$. Relator by relator,

$$\theta(r_1) = s^2, \quad \theta(r_2) = (sps\,p^{-1})^3, \quad \theta(r_3) = [s, p^2 s p^{-2}], \quad \theta(r_5) = psp^{-1}qs^{-1}q^{-1}$$

are the four relators of P_H, and $\theta(r_4) = [p, q][p, q]^{-1}$ is freely trivial. If $w = \prod_{i=1}^{k} u_i r_{c_i}^{\pm1} u_i^{-1}$ freely, then applying θ, which fixes w, gives $w = \prod_{i=1}^{k} \theta(u_i)\theta(r_{c_i})^{\pm1}\theta(u_i)^{-1}$ freely; deleting the freely trivial factors with $c_i = 4$ writes w as a product of at most k conjugates of relators of P_H and their inverses. The second claim follows, because a word of length at most n in $p^{\pm1}, q^{\pm1}$ is also one in $p^{\pm1}, q^{\pm1}, \alpha^{\pm1}$. □

From now on areas are taken in P_J, $\mathrm{Area} = \mathrm{Area}_J$. Freely $[g, hk] = [g, h] \cdot h[g, k]h^{-1}$ and $[g^{-1}, h] = g^{-1}[h, g]g$, and area is invariant under free conjugation and inversion; so the area of a commutator $[g, h_1 \cdots h_m]$ is at most the sum of the areas of the $[g, h_l]$, and inverting g or h does not change the area of $[g, h]$.

A.2 Far commutators

For $j \in \mathbb{Z}$ and $k \ge 0$ put

$$\sigma_j = p^j \alpha p^{-j}, \qquad \sigma'_j = q^j \alpha q^{-j}, \qquad \eta = q^{-1}p, \qquad \eta_k = p^k \eta p^{-k}.$$

For $j \ge 0$ the words σ_j and σ'_j represent the transposition of $(1, j+1)$ and $(1, j+2)$, by Lemma 6.2(a) and because q acts on ray 1 as p does. Since p^{-k} maps $(1,1), (1,2)$ to $(2,k), (2,k-1)$ for $k \ge 2$, the word σ_{-k} represents the transposition of $(2, k-1)$ and $(2, k)$. The element η fixes ray 1 pointwise, and q fixes ray 2 pointwise. So the following words are null, and for $m \ge 2$, $i, j \ge 0$ and $k \ge 2$ we put

$$\mathsf{A}(m) = \mathrm{Area}[\sigma_0, \sigma_m], \quad \mathsf{E}(i) = \mathrm{Area}[\sigma_i, \eta], \quad \mathsf{U}(j) = \mathrm{Area}(\sigma'_j \sigma_j^{-1}), \quad \mathsf{B}(k) = \mathrm{Area}[\sigma_{-k}, q],$$

and $\mathsf{A}^*(M) = \max_{2 \le m \le M} \mathsf{A}(m)$. Since $[\sigma_i, \sigma_j] = p^i[\sigma_0, \sigma_{j-i}]p^{-i}$ freely, $\mathsf{A}^*(M)$ bounds the area of every $[\sigma_i^{\pm1}, \sigma_j^{\pm1}]$ with $i, j \in \mathbb{Z}$ and $2 \le |i - j| \le M$.

Cells. The following moves cost one relator each, or four for the last.

(c1) $[\alpha, \eta] = q^{-1}r_5^{-1}q$ freely, so $\mathsf{E}(0) \le 1$.

(c2) Replacing pqp^{-1} by αq, or qpq^{-1} by $\alpha^{-1}p$: the words $pqp^{-1}(\alpha q)^{-1} = r_4$ and $qpq^{-1}(\alpha^{-1}p)^{-1} = qpq^{-1}p^{-1}\alpha$ are cyclic conjugates of $r_4^{\pm1}$.

(c3) $[\sigma_i, \sigma_{i+2}] = p^i r_3 p^{-i}$ freely, so $\mathsf{A}(2) \le 1$.

(c4) Replacing σ_j^{-1} by σ_j: $\sigma_j^{-1}\sigma_j^{-1} = p^j r_1^{-1} p^{-j}$.

(c5) The braid word $\sigma_i\sigma_{i+1}\sigma_i\sigma_{i+1}^{-1}\sigma_i^{-1}\sigma_{i+1}^{-1}$ has area at most 4: replacing its three inverted letters by (c4) leaves $(\sigma_i\sigma_{i+1})^3 = p^i r_2 p^{-i}$.

Lemma A.2 (Lee's recursion)**.** *For $i \ge 1$, $j \ge 1$ and $k \ge 2$:*

(i) $\mathsf{E}(0) \le 1$ *and* $\mathsf{E}(i) \le 2i + 1 + \sum_{m=2}^{i+1} \mathsf{A}(m)$;

(ii) $\mathsf{U}(0) = 0$ *and* $\mathsf{U}(j) \le \sum_{l=0}^{j-1} \mathsf{E}(l)$;

(iii) $\mathsf{B}(k) \le 2k + 5 + \sum_{m=2}^{k-1} \mathsf{A}(m)$;

(iv) $\mathsf{A}(k+j) \le \mathsf{A}(k) + 2j\,\mathsf{B}(k) + 2\,\mathsf{U}(j)$.

Proof. (i) The bound on $\mathsf{E}(0)$ is (c1). For $k \ge 0$, the word $\eta_{k+1}\sigma_{k-1}\eta_k^{-1}$ is freely $p^k(pq^{-1}p^{-1}\alpha q)p^{-k}$, a conjugate of r_4^{-1}; so η_k becomes $\eta_{k+1}\sigma_{k-1}$ at the cost of one cell, and $\eta = \eta_0$ becomes $\eta_i S$ with $S = \sigma_{i-2}\cdots\sigma_0\sigma_{-1}$ at the cost of i cells. Doing this for η and for η^{-1} in $[\sigma_i, \eta]$ costs $2i$ cells. Then $[\sigma_i, \eta_i S] = [\sigma_i, \eta_i] \cdot \eta_i[\sigma_i, S]\eta_i^{-1}$ freely. The first factor is $p^i[\alpha, \eta]p^{-i}$, of area at most 1. The second is a product of conjugates of the $[\sigma_i, \sigma_l]$, $l = i-2, \dots, -1$, whose distances $i - l$ run through $2, \dots, i+1$.

(ii) $\sigma_0' = \sigma_0 = \alpha$. Freely $\sigma_{l+1}' = q\sigma_l' q^{-1}$. In $\sigma_{l+1}'\sigma_{l+1}^{-1}$, replacing σ_l' by σ_l costs $\mathsf{U}(l)$ and leaves $q\sigma_l q^{-1}\sigma_{l+1}^{-1}$, which is freely $q[\sigma_l, \eta]q^{-1}$. So $\mathsf{U}(l+1) \le \mathsf{U}(l) + \mathsf{E}(l)$.

(iii) Freely $[\sigma_{-k}, q] = p^{-k}[\alpha, p^k q p^{-k}]p^k$. By (c2), $p^k q p^{-k}$ becomes $\sigma_{k-1}\cdots\sigma_1\sigma_0 q$ at the cost of k cells: if $p^{k-1}qp^{1-k}$ has become $\sigma_{k-2}\cdots\sigma_0 q$, then $p^k qp^{-k}$ has become $\sigma_{k-1}\cdots\sigma_1\, pqp^{-1}$ freely, and one cell replaces pqp^{-1} by $\sigma_0 q$. This word occurs twice in the commutator, which costs $2k$ cells. Then $[\alpha, \sigma_{k-1}\cdots\sigma_2 \cdot \sigma_1\sigma_0 q]$ has area at most $\sum_{l=2}^{k-1} \mathsf{A}(l) + \mathrm{Area}[\alpha, \sigma_1\sigma_0 q]$. Finally, in $[\alpha, \sigma_1\sigma_0 q] = \alpha\sigma_1\sigma_0\, q\alpha^{-1}q^{-1}\,\sigma_0^{-1}\sigma_1^{-1}$, replacing $q\alpha^{-1}q^{-1} = \sigma_1'^{-1}$ by σ_1^{-1} costs $\mathsf{U}(1) \le \mathsf{E}(0) \le 1$ by (ii), and leaves the braid word $\sigma_0\sigma_1\sigma_0\sigma_1^{-1}\sigma_0^{-1}\sigma_1^{-1}$, of area at most 4 by (c5). So $\mathrm{Area}[\alpha, \sigma_1\sigma_0 q] \le 5$.

(iv) Conjugating by p^{-k} turns $[\sigma_0, \sigma_{k+j}]$ into $[\sigma_{-k}, \sigma_j]$. Replacing $\sigma_j^{\pm 1}$ by $\sigma_j'^{\pm 1}$ costs $2\mathsf{U}(j)$ and gives $\sigma_{-k}\, q^j \alpha q^{-j}\, \sigma_{-k}^{-1}\, q^j\alpha^{-1}q^{-j}$. Moving σ_{-k} to the right across q^j, and σ_{-k}^{-1} to the left across q^{-j}, costs $2j\mathsf{B}(k)$, since each move across one letter is a replacement whose cost is the area of a conjugate of $[\sigma_{-k}, q]^{\pm 1}$. What remains is $q^j[\sigma_{-k}, \alpha]q^{-j}$, and $[\sigma_{-k}, \alpha] = p^{-k}[\sigma_0, \sigma_k]p^k$ freely. □

With $k = 2$, part (iii) gives $\mathsf{B}(2) \le 9$, and part (iv) gives $\mathsf{A}(m) \le 2\mathsf{U}(m-2) + 18m - 35$ for $m \ge 3$. Iterating this, as in [Lee12], bounds $\mathsf{A}(m)$ only exponentially in m. Splitting the distance in half gives a quasi-polynomial bound.

Proposition A.3 (Far commutators)**.** *For* $M \ge 2$ *and* $\lambda = \lceil \log M \rceil$,

$$\mathsf{A}^*(M) \;\le\; 5 \cdot 3^{\lambda-1} 2^{\lambda(\lambda-1)} - 4 \;\le\; 5\,(6M)^{\log M}.$$

Proof. Step 1: one halving. Let $m \ge 3$, and put $k = \lceil m/2 \rceil \ge 2$ and $j = \lfloor m/2 \rfloor \ge 1$, so that $j \le k < m$ and $k + j = m$. By Lemma A.2, $\mathsf{B}(k) \le 2k + 5 + (k-2)\mathsf{A}^*(k)$, and $\mathsf{E}(l) \le 2l + 1 + l\,\mathsf{A}^*(l+1)$ for $l \ge 1$, so that

$$\mathsf{U}(j) \le 1 + \sum_{l=1}^{j-1}(2l+1) + \mathsf{A}^*(k)\sum_{l=1}^{j-1} l = j^2 + \tfrac{1}{2}j(j-1)\,\mathsf{A}^*(k).$$

Part (iv) and $j \le k$ give

$$\mathsf{A}(m) \le \mathsf{A}^*(k)\big(1 + 2j(k-2) + j(j-1)\big) + 4jk + 10j + 2j^2 \le (3k^2 - 5k + 1)\mathsf{A}^*(k) + 6k^2 + 10k.$$

Step 2: the recursion. Put $G(M) = \mathsf{A}^*(M) + 4$. Since $k \ge 2$,

$$3k^2 G(k) - (3k^2 - 5k + 1)\mathsf{A}^*(k) - 6k^2 - 10k - 4 = (5k-1)\mathsf{A}^*(k) + 6k^2 - 10k - 4 \ge 0,$$

so $\mathsf{A}(m) + 4 \le 3k^2 G(k)$ with $k = \lceil m/2 \rceil$. The right side is nondecreasing in m, and $G(2) = \mathsf{A}(2) + 4 \le 5$ by (c3); hence $G(m) \le 3\lceil m/2 \rceil^2 G(\lceil m/2 \rceil)$ for every $m \ge 3$.

Step 3: iteration. Let $m_0 = M$ and $m_{l+1} = \lceil m_l/2 \rceil$. Then $m_l \le 2^{\lambda-l}$, so the first index l_0 with $m_{l_0} \le 2$ is at most $\lambda-1$, and $m_{l_0} = 2$. Applying Step 2 at $m_0, \dots, m_{l_0-1}$, with $m_l^2 \le 4^{\lambda-l}$, gives

$$G(M) \le G(2) \prod_{l=1}^{l_0} 3 \cdot 4^{\lambda-l} \le 5 \prod_{l=1}^{\lambda-1} 3 \cdot 4^{\lambda-l} = 5 \cdot 3^{\lambda-1} 2^{\lambda(\lambda-1)}.$$

The last bound of the proposition follows from $3^{\lambda-1}2^{\lambda(\lambda-1)} = (3 \cdot 2^{\lambda})^{\lambda-1}$, $2^{\lambda} \le 2M$ and $\lambda - 1 \le \log M$. □

A.3 Collection

Proposition A.4 (Collection)**.** *For $n \ge 1$, $\delta_J(n) \le 400n^4\big(1 + \mathsf{A}^*(4n+2)\big)$.*

Proof. Fix n and put $F = 1 + \mathsf{A}^*(4n+2)$. By Lemma A.2(i),(ii), $\mathsf{E}(l) \le 2Fl + 1$ and $\mathsf{U}(l) \le Fl(l-1) + l \le Fl^2$ for $0 \le l \le 2n$. Let w be a null word of length at most n, put $g_{m,l} = p^m q^l$ and $h = g_{n,n}$, and consider hwh^{-1}.

Step 1: collecting onto ray 1. While the letters of w are read after h, the exponent sums (m, l) of p and q stay in $[0, 2n]^2$. We keep the prefix read so far in the form $v\, g_{m,l}$, with v a word in the $\sigma_j^{\pm 1}$, starting from the empty v and $g_{n,n} = h$, and use the rules

$$g_{m,l}\, p = (\sigma_{m+l-1}^{-1} \cdots \sigma_m^{-1})\, g_{m+1,l}, \qquad g_{m,l}\, p^{-1} = (\sigma_{m-1} \cdots \sigma_{m+l-2})\, g_{m-1,l},$$
$$g_{m,l}\, q^{\pm 1} = g_{m,l\pm 1}, \qquad g_{m,l}\, \alpha^{\pm 1} = \sigma_{m+l}^{\pm 1}\, g_{m,l}.$$

For the first, $q^l p q^{-l} = \sigma'^{-1}_{l-1} \cdots \sigma'^{-1}_0 p$ by l cells (c2) replacing qpq^{-1} by $\alpha^{-1}p$; replacing each σ'_i by σ_i costs $\mathsf{U}(i)$; and $p^m \sigma_i p^{-m} = \sigma_{m+i}$ freely. The second is the first at $(m-1, l)$, rearranged, and is used only with $m \ge 1$. The third is free, and the fourth costs $\mathsf{U}(l)$. So each rule costs at most $l + \sum_{i<l} \mathsf{U}(i) \le 2n + F\sum_{i<2n} i^2 \le 2n + \frac{8}{3}Fn^3 \le 5Fn^3$. Since w has exponent sums 0 in p and in q (its image under (5) is trivial), the final form is $v\, h$, so hwh^{-1} has become v at a cost of at most $5Fn^4$. Each rule adds at most $2n$ letters, so v has at most $2n^2$ letters $\sigma_j^{\pm 1}$, all with $0 \le j \le 4n$. The rules hold in H, so v represents 1.

Step 2: normal forms. Replace each σ_j^{-1} in v by σ_j, one cell each by (c4). Let $K = 4n+1$. The word v now represents 1 in the symmetric group of the points $(1,1), \dots, (1, K+1)$, in which $\sigma_0, \dots, \sigma_{K-1}$ are the adjacent transpositions. For $0 \le k \le K$ let $c_k = \sigma_{K-1}\sigma_{K-2} \cdots \sigma_k$, with c_K empty; it maps $(1, k+1)$ to $(1, K+1)$. So every permutation π of these points is uniquely $u\, c_k$ with u fixing $(1, K+1)$, namely with $(1, k+1) = \pi^{-1}(1, K+1)$; applying this to u and so on gives a unique normal form, and the normal form of 1 is the empty word. A generator σ_i is inserted on the right of $u\, c_k$ by

$$c_k\sigma_i = \sigma_i c_k \ \ (i \le k-2), \qquad c_k\sigma_{k-1} = c_{k-1}, \qquad c_k\sigma_k = c_{k+1}, \qquad c_k\sigma_i = \sigma_{i-1}c_k \ \ (i \ge k+1),$$

and the new letter σ_i or σ_{i-1} on the left, whose index is at most $K-2$, is then inserted into u. The first rule moves σ_i across $\sigma_k, \dots, \sigma_{K-1}$ by commutations of transpositions at distance at least 2. The second concatenates, and the third cancels $\sigma_k\sigma_k$ by one cell (c4). The last moves σ_i left across $\sigma_k, \dots, \sigma_{i-2}$, applies one braid $\sigma_i\sigma_{i-1}\sigma_i = \sigma_{i-1}\sigma_i\sigma_{i-1}$, and moves σ_{i-1} left across $\sigma_{i+1}, \dots, \sigma_{K-1}$. A level costs at most $K+1$ moves and an insertion passes through at most K levels, so a word of length ℓ reaches its normal form, here the empty word, in at most $\ell K(K+1)$ moves. A commutation costs at most $\mathsf{A}^*(4n) \le F$, a braid at most 4 by (c5), and a cancellation 1, so each move costs at most $4F$. Since $K(K+1) = (4n+1)(4n+2) \le 30n^2$, this step costs at most $2n^2 + 2n^2 \cdot 30n^2 \cdot 4F \le 242Fn^4$.

In total $\mathrm{Area}(hwh^{-1}) \le (5 + 242)Fn^4 \le 400Fn^4$, and area is invariant under conjugation by h. □

Corollary A.5 (The bound (H1))**.** *For $n \geq 1$, $\delta_H(n) \leq \delta_J(n) \leq 2^{(\log n+6)^2}$. In particular P_J and P_H are presentations of H.*

Proof. The first inequality is Lemma A.1. By Propositions A.3 and A.4 with $\lambda = \lceil \log(4n+2) \rceil \leq \log n + 3.59$ and $\ell = \log n \geq 0$,

$$\log \delta_J(n) \leq \log 2000 + 4\ell + (\lambda - 1)\log 3 + \lambda(\lambda - 1) \leq \ell^2 + 11.8\ell + 24.4 \leq (\ell + 6)^2.$$

Here $\log 2000 < 10.97$, $(\lambda - 1)\log 3 \leq 1.585(\ell + 2.59) \leq 1.585\ell + 4.11$ and $\lambda(\lambda - 1) \leq (\ell + 3.59)(\ell + 2.59) \leq \ell^2 + 6.18\ell + 9.30$. So every null word in p, q, α is a product of conjugates of relators of P_J, and by Lemma A.1 every word in p, q that represents 1 in H is a product of conjugates of relators of P_H. Since p and q generate H (Lemma 6.2(b)) and the relators hold in H, P_J and P_H present H. □

The proof does not use that P_J presents H, so it reproves Johnson's theorem [Joh99].

Remark A.6 (Comparison with the polynomial bound). The bound of Proposition A.3 is $M^{(1+o(1))\log M}$. A doubling endomorphism of H_3 gives $\mathsf{A}^*(M) \leq 135M^{\log 107}$ [DM26b, Theorem 3.2], which with Proposition A.4 is the bound (H2); it is not used in this paper. The two bounds are not comparable for all M, but the values of the recursion in the proof of Proposition A.3, computed from Lemma A.2 with equality and $\mathsf{A}(2) = 1$, lie below $135M^{\log 107}$ for every $M \leq 750$; for instance they give $\mathsf{A}(8) \leq 543$ and $\mathsf{A}(28) \leq 536\,559$.

B Slofstra's embedding with $w \mapsto -I$

This appendix proves the embedding behind the game of Section 10. An $E\Gamma_0$ presentation $\mathcal{P}$ with at most one non-involutive generator and a distinguished involutive generator w is turned into a binary linear system (A, b), together with a substitution Ψ of words, so that approximate representations pass in both directions with linear loss and w corresponds to J (Lemma B.1 and Theorem B.3). The construction is Slofstra's [Slo19, §4]: the non-involutive generator is removed with ancilla involutions [Slo19, Prop. 4.8], a commutator device makes $J = [w, t^*]$ [Slo19, Lemma 5.2], the conjugacy relations are put in a special form [Slo19, Lemma 4.4], and each of them is replaced by six linear equations [Slo19, Prop. 4.2]. The two-sided statement follows Slofstra and Vidick [SV18, Prop. 6.2]. What we add is bookkeeping: in the direction from $\mathcal{P}$ to Γ the dimension grows by the factor 8 and the defect by the factor 3; the direction from Γ to $\mathcal{P}$ keeps an arbitrary state; and we count the variables and equations. We treat only $|S_1| \leq 1$, the case of Proposition 10.3.

B.1 Conventions

A *linear-plus-conjugacy presentation* has a finite set T of generators, a further generator $J \notin T$, and the following relators: x^2 for $x \in T \cup \{J\}$; $[x, J]$ for $x \in T$; for each V in a finite family $\mathcal{V}$ of nonempty subsets of T (the *linear sets*), the commutators $[x, x']$ for $x, x' \in V$ and one product relator $J^{b_V} \prod_{x \in V} x$, with $b_V \in \{0, 1\}$ and the product taken in a fixed order; and *conjugacy relators* $xyxz^{-1}$ for the triples (x, y, z) of a finite set $\mathcal{C} \subseteq T^3$. With $\mathcal{C} = \emptyset$ this is the presentation of a solution group (Definition 2.7), up to the order of the products. The presentation is *nice* if for every $(x, y, z) \in \mathcal{C}$ either $y = z$, or y and z lie in a common linear set; then $[y, z]$ is freely trivial or a relator.

All assignments below send each generator to a *block-monomial* unitary: an operator on $(\mathbb{C}^D)^{\oplus k}$ with exactly one nonzero block in each block row and each block column, every nonzero block being a unitary on $\mathbb{C}^D$. Products and inverses of block-monomial unitaries are block-monomial. If $M = \mathrm{diag}(M_1, \ldots, M_k)$ is block-diagonal, then $\|M - I\|_f^2 = \frac{1}{k}\sum_l \|M_l - I\|_f^2 \leq \max_l \|M_l - I\|_f^2$. In our formulas every nonzero block of the image of a generator is the image

of a word under a given assignment γ; so when the image of a relator r is block-diagonal, as it is in every case below, its diagonal blocks are images $\gamma(r_1), \dots, \gamma(r_k)$ of words, which we call the *block words* of r. Hence, if every block word of r is trivial in the group of γ with area at most k_0, the defect of r is at most k_0 times the largest defect of a relator under γ (Lemma 2.3). We write

$$\Omega = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}, \qquad U \oplus V = \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix}, \qquad \operatorname{adiag}(U,V) = \begin{pmatrix} 0 & U \\ V & 0 \end{pmatrix},$$

so that $\operatorname{adiag}(U,V)\operatorname{adiag}(U',V') = UV' \oplus VU'$ and $\operatorname{adiag}(U,V)^{-1} = \operatorname{adiag}(V^{-1},U^{-1})$.

B.2 The construction

Lemma B.1 (Slofstra's embedding, with $w \mapsto -I$; cf. [Slo19, §4], [SV18, §6])**.** *Let $\mathcal{P} = \langle S_0 \sqcup S_1 \mid R\rangle$ be an $E\Gamma_0$ presentation with $|S_1| \le 1$ and nonempty linear sets of at most three elements, and let $w \in S_0$. Assume that the product of each linear set of $\mathcal{P}$ is taken in an order that is a cyclic rotation of the order in which S_0 is listed. The construction below gives a binary linear system (A,b) in variables $x_1,\dots,x_n$, with at most three variables in each equation; a homomorphism $\Psi : F(S_0 \sqcup S_1) \to F(x_1,\dots,x_n,J)$ that sends each $g \in S_0$ to a variable x_g and the generator of S_1, if there is one, to a product $x_f x_v$ of two variables; a variable x_{j_0}; and a constant $A_0 \ge 1$, such that, with $\Gamma = \Gamma(A,b)$ presented as in Definition 2.7:*

(G1) for every $r \in R$, the word $\Psi(r)$ is trivial in Γ and has area at most A_0;

(G2) the word $[x_w, x_{j_0}]J$ is trivial in Γ and has area at most A_0;

(G3) every ε-representation ϕ of $\mathcal{P}$ on $\mathbb{C}^d$ with $\phi(w) = -I$ yields a 3ε-representation of Γ on $\mathbb{C}^{8d}$ with $J \mapsto -I$.

If $\mathcal{P}$ has n_0 involutive generators, ℓ_0 linear sets, c_0 triples in C_0 and c_1 in C_1, put $N = n_0 + c_1 + 4$ and $c = c_0 + 2c_1 + 1$ if $S_1 \neq \emptyset$, and $N = n_0 + 2$, $c = c_0 + 1$ otherwise. Then (A,b) has

$$n = 11N + 16c + 1 \quad \text{variables and} \quad m = \ell_0 + 1 + 8N + 14c \quad \text{equations},$$

all with right-hand side 0 except one, $x_w + x_u = 1$.

The hypothesis on the order of the products costs nothing: reordering the product of a linear set changes neither the group nor, up to a constant factor, the defects (Lemma 2.4), and the presentation of Proposition 10.3 satisfies it (Appendix C.3).

Proof. Let ϕ be an ε-representation of $\mathcal{P}$ on $\mathbb{C}^d$ with $\phi(w) = -I$. We pass through presentations G', $\widehat{G}$ and a nice $\widehat{G}_1$, each of which receives the relators of the previous one as trivial words and an approximate representation built from the previous one. The dimension doubles in Steps 1, 3 and 4, and the defect grows by the factor 3 in Step 4; before that, only the commutators inside the linear sets of $\mathcal{P}$, which are not relators of $\mathcal{P}$, reach defect $\sqrt{2}\varepsilon$.

Step 1: removing the non-involutive generator [Slo19, Prop. 4.8]. If $S_1 = \emptyset$, let $G' = \mathcal{P}$ and put $\gamma(x) = \phi(x) \oplus I$ on $\mathbb{C}^d \oplus \mathbb{C}^d$. Otherwise let $S_1 = \{y\}$. Add involutions f, v, with $y = fv$ in mind, and for each triple $(y,b,c) \in C_1$ an involution Z_{bc} with the two conjugacy relations $vbv = Z_{bc}$ and $fZ_{bc}f = c$. These imply the old relation, since $fvbvf = (fv)b(fv)^{-1}$ once $f^2 = v^2 = e$. With them imposed, y occurs only in $y = fv$ and can be deleted. The result G' has only involutive generators, the linear sets of $\mathcal{P}$, and conjugacy relations. On $\mathbb{C}^d \oplus \mathbb{C}^d$ put

$$\gamma(x) = \phi(x) \oplus I \quad (x \in S_0), \qquad \gamma(f) = \Omega,$$
$$\gamma(v) = \operatorname{adiag}\big(\phi(y)^*, \phi(y)\big), \qquad \gamma(Z_{bc}) = I \oplus \phi(y)\phi(b)\phi(y)^*.$$

Then $\gamma(f)\gamma(v) = \phi(y) \oplus \phi(y)^*$. The relations $f^2 = v^2 = e$ and $vbv = Z_{bc}$ hold exactly; Z_{bc}^2 has block words e and yb^2y^{-1}, so its defect is at most that of b^2; and $\gamma(f)\gamma(Z_{bc})\gamma(f)\gamma(c)^{-1} = \phi(yby^{-1}c^{-1}) \oplus I$ has defect at most ε. The relations among S_0 keep defect at most ε, because $\gamma(x) = \phi(x) \oplus I$ and the assignment $x \mapsto I$ satisfies every homogeneous linear relation and every conjugacy relation. So in both cases γ is an ε-representation of G' on $\mathbb{C}^{2d}$ with $\gamma(w) = -I_d \oplus I_d$. The commutators $[x, x']$ of two elements of a linear set of $\mathcal{P}$ are not relators of $\mathcal{P}$, but they are relators of the linear-plus-conjugacy presentations below. Since (x, x', x') or (x', x, x) lies in C_0, the word $[x, x']$ is the relator $xx'xx'^{-1}$, or the inverse of $x'xx'x^{-1}$, with one letter inverted, so it has area at most 2 in $\mathcal{P}$ and $\|\phi([x, x']) - I\|_f \le 2\varepsilon$ (Lemma 2.3); hence $\gamma([x, x']) = \phi([x, x']) \oplus I$ has defect at most $\sqrt{2}\varepsilon$.

Step 2: making J a commutator [Slo19, Lemma 5.2]. Let $\widehat{G}$ be presented by the relations of G' together with new involutions t^*, J, u and the relations $J^2 = t^{*2} = u^2 = e$, $[x, J] = e$ for every generator x, the linear set $\{w, u\}$ with $b_V = 1$ (that is, the relators $[w, u]$ and Jwu), and the conjugacy relation $t^*wt^* = u$. In $\widehat{G}$ these give $[w, t^*] = w\,(t^*wt^*) = wu = J$, so the word $[w, t^*]J$ is trivial. Extend γ by $\gamma(J) = -I_{2d}$, $\gamma(u) = -\gamma(w) = I_d \oplus -I_d$ and $\gamma(t^*) = \Omega$. Then $\gamma(t^*)\gamma(w)\gamma(t^*) = \gamma(u)$, and all new relations hold exactly. So γ is an ε-representation of $\widehat{G}$ on $\mathbb{C}^{2d}$ with $J \mapsto -I$, except that the commutators inside the linear sets of $\mathcal{P}$ have defect at most $\sqrt{2}\varepsilon$. The presentation of $\widehat{G}$ is a linear-plus-conjugacy presentation; its conjugacy relations are those of C_0 (among them the commutations, listed as triples (x, y, y)), the $2c_1$ relations of Step 1, and $t^*wt^* = u$.

Step 3: making the conjugacy relations nice [Slo19, Lemma 4.4]. Let T be the set of generators of $\widehat{G}$ other than J, and $\mathcal{C}$ its set of conjugacy triples. The presentation of $\widehat{G}_1$ has the generators of $\widehat{G}$, one more involution g^*, and for each $a \in T$ three involutions $a', a'', \bar{a}$. Its relators are those of $\widehat{G}$ except the conjugacy relators, together with: the linear sets $\{a, a', a''\}$ and $\{a, g^*, \bar{a}\}$ for $a \in T$, with $b_V = 0$; the conjugacy relators $g^*a'g^*a''^{-1}$ for $a \in T$; and, for each $(a, b, c) \in \mathcal{C}$, the conjugacy relators $\bar{a}b'\bar{a}c''^{-1}$ and $\bar{a}b''\bar{a}c'^{-1}$ together with the linear sets $\{b', c'', k_{abc}\}$ and $\{c', b'', k'_{abc}\}$ in new involutions k_{abc}, k'_{abc}. This presentation is nice: a', a'' lie in $\{a, a', a''\}$, b', c'' in $\{b', c'', k_{abc}\}$, and c', b'' in $\{c', b'', k'_{abc}\}$. The old conjugacy relators follow. Indeed $a = g^*\bar{a} = \bar{a}g^*$ and $b = b'b''$, so

$$aba = g^*\,(\bar{a}b'\bar{a})(\bar{a}b''\bar{a})\,g^* = g^*c''c'g^* = (g^*c''g^*)(g^*c'g^*) = c'c'' = c,$$

using $g^*c'g^* = c''$, hence $g^*c''g^* = c'$, and the commutation of c' and c''. So the identity on the generators of $\widehat{G}$ sends its relators to trivial words of $\widehat{G}_1$.

Let $D = 2d$, and on $\mathbb{C}^D \oplus \mathbb{C}^D$ put

$$\gamma_1(x) = \gamma(x) \oplus \gamma(x) \quad (x \in T \cup \{J\}), \qquad \gamma_1(g^*) = \Omega, \qquad \gamma_1(\bar{a}) = \operatorname{adiag}\big(\gamma(a), \gamma(a)\big),$$
$$\gamma_1(a') = \gamma(a) \oplus I, \quad \gamma_1(a'') = I \oplus \gamma(a), \quad \gamma_1(k_{abc}) = \gamma(b) \oplus \gamma(c), \quad \gamma_1(k'_{abc}) = \gamma(c) \oplus \gamma(b).$$

Write $U_a = \gamma(a)$, $U_b = \gamma(b)$, $U_c = \gamma(c)$. The relators of $\widehat{G}$ that were kept have block words r, r, so their defects are unchanged, and $\gamma_1(J) = -I$ commutes with everything. The commutators inside the new linear sets and the relators $g^*a'g^*a''^{-1}$ are mapped exactly to I, as one sees by multiplying the matrices; for instance $\gamma_1(\bar{a})\gamma_1(g^*) = U_a \oplus U_a$ and $\Omega(U_a \oplus I)\Omega = I \oplus U_a$. The squares of the new generators and the product relators of the new linear sets have block words among aa, bb, cc and the empty word, in any order of the factors; each has area at most 1. Finally

$$\gamma_1(\bar{a})\gamma_1(b')\gamma_1(\bar{a})\gamma_1(c'')^{-1} = U_a^2 \oplus U_aU_bU_aU_c^{-1},$$
$$\gamma_1(\bar{a})\gamma_1(b'')\gamma_1(\bar{a})\gamma_1(c')^{-1} = U_aU_bU_aU_c^{-1} \oplus U_a^2,$$

with block words of area 1. So γ_1 is an ε-representation of $\widehat{G}_1$ on $\mathbb{C}^{2D}$ with $J \mapsto -I$, with the same exception.

Step 4: linear gadgets for nice conjugacy relations [Slo19, Prop. 4.2]. Let $\Gamma(A,b)$ be the solution group obtained from $\widehat{G}_1$ by deleting each conjugacy relator $abac^{-1}$ (from now on a,b,c denote generators of $\widehat{G}_1$) and adding seven new variables $y_1,\dots,y_7$, depending on the triple, and the six equations

$$ay_1y_2 = by_2y_3 = y_3y_4y_5 = ay_5y_6 = cy_6y_7 = y_1y_4y_7 = e. \tag{24}$$

Every linear set of $\widehat{G}_1$ becomes one equation. The six equations force the relation: using them in turn,

$$aba = (y_1y_2)(y_2y_3)(y_5y_6) = y_1y_3y_5y_6 = y_1(y_4y_5)y_5y_6 = y_1y_4y_6 = y_1(y_1y_7)y_6 = y_7y_6 = c.$$

So the identity on the old generators sends the relators of $\widehat{G}_1$ to trivial words of $\Gamma(A,b)$.

Ordering. Number the variables as follows: the generators of $\widehat{G}$ other than J (those of $\mathcal{P}$ in their listed order, then f, v, the Z_{bc}, t^*, u), then all a', all a'', g^*, all $\bar{a}$, all k_{abc} and k'_{abc}, and then the gadget variables $y_1 < \cdots < y_7$ of each triple. The products in (24) and in the linear sets of Steps 2 and 3, in the order written, are then increasing. The products of the linear sets of $\mathcal{P}$ are cyclic rotations of increasing ones, by hypothesis, so the relator of Definition 2.7 is a conjugate of the one written. Replacing a relator by a conjugate changes neither flat defects nor areas, so we may work with the orders written.

The representation. On $\mathbb{C}^{2D} \oplus \mathbb{C}^{2D}$ let $\gamma_2(x) = \gamma_1(x) \oplus \gamma_1(x)$ for the generators of $\widehat{G}_1$, so the kept relators have unchanged defects and $\gamma_2(J) = -I$. For a triple (a,b,c) write now $U_a = \gamma_1(a)$, $U_b = \gamma_1(b)$, $U_c = \gamma_1(c)$, and put

$$\begin{gathered}\gamma_2(y_1) = \operatorname{adiag}(U_a,U_a), \qquad \gamma_2(y_2) = \Omega, \qquad \gamma_2(y_3) = \operatorname{adiag}(U_b,U_b),\\ \gamma_2(y_4) = \operatorname{adiag}(U_bU_a,U_aU_b), \qquad \gamma_2(y_5) = U_bU_aU_b \oplus U_a,\\ \gamma_2(y_6) = U_bU_c \oplus I, \qquad \gamma_2(y_7) = U_b \oplus U_c.\end{gathered}$$

If γ_1 were an exact representation, γ_2 would be one. Multiplying out the matrices, the image of every relator of $\Gamma(A,b)$ that involves the gadget is block-diagonal; Table 3 lists its block words, which are words in a,b,c. We bound their areas with respect to the relators a^2, b^2, c^2, $abac^{-1}$ and $bcb^{-1}c^{-1}$ of $\widehat{G}_1$; the last one is a relator, or freely trivial, because $\widehat{G}_1$ is nice. All five have defect at most ε under γ_1: for every conjugacy triple of $\widehat{G}_1$ the letters b,c lie in one of the linear sets added in Step 3, whose commutators are mapped exactly to I. A *cell* replaces a subword of a cyclic permutation of the word by the inverse of the complementary subword of a cyclic permutation of one of these relators or its inverse; for instance $x^{-1} \to x$, $xx \to e$, $aba \to c$, $c^{-1}a \to a^{-1}b^{-1}$, $bc \to cb$. Free reductions and cyclic permutations cost nothing. Every block word reaches the empty word in at most 3 cells (Table 3), so it has area at most 3 in $\widehat{G}_1$.

Reductions. Every arrow is one cell, $\sim$ marks a cyclic permutation, and free reductions are made silently.

- $a^2b^2 \to a^2 \to e$; $b^2 \cdot ab^{-2}a^{-1} \to ab^{-2}a^{-1} \to e$, and likewise $a^2 \cdot ba^{-2}b^{-1}$.
- $abac \to cc \to e$; $aca^{-1}b^{-1} \sim a^{-1}b^{-1}ac \to a^{-1}b^{-1}a^{-1}c \to e$; $aba^{-1}c^{-1} \sim a^{-1}c^{-1}ab \to a^{-2} \to e$.
- $bab^2ab \to ba^2b \to b^2 \to e$; $abab^2c \to abac \to cc \to e$.
- $bcbc \to b^{-1}cbc \to b^{-1}c^{-1}bc \to e$, the last step because $b^{-1}c^{-1}bc$ is a cyclic permutation of $[b,c]$.
- $abab\,a^{-1}b^{-1}a^{-1}b^{-1} \to cb\,a^{-1}b^{-1}a^{-1}b^{-1} \to cbc^{-1}b^{-1} \to e$.
- $abca^{-1}c^{-1}b^{-1} \sim a^{-1}c^{-1}b^{-1}abc \to a^{-1}b^{-1}c^{-1}abc \to a^{-1}b^{-1}a^{-1}c \to e$, using $c^{-1}b^{-1} \to b^{-1}c^{-1}$ and then $c^{-1}a \to a^{-1}b^{-1}$.

block word	where it occurs	cells
a^2, b^2, c^2	squares of y_1, y_3, y_5, y_7; products $1, 2, 4, 5$; $[y_3, y_5]$	1
$bab^{-2}a^{-1}b^{-1}$, ab^2a^{-1} (conjugates of $b^{\mp 2}$)	$[y_3, y_5]$, $[y_4, y_5]$	1
$cbc^{-1}b^{-1}$, $bcbc^{-1}b^{-2}$ (conjugates of $[c, b]$)	$[c, y_6]$, $[c, y_7]$, $[y_6, y_7]$	1
a^2b^2	square of y_4; products $3, 6$	2
$b^2ab^{-2}a^{-1}$, $a^2ba^{-2}b^{-1}$	$[y_3, y_4]$, $[y_1, y_4]$	2
$abac$	product 6	2
$aca^{-1}b^{-1}$, $aba^{-1}c^{-1}$	$[y_1, y_7]$, $[y_4, y_7]$	2
bab^2ab	square of y_5; product 3	3
$bcbc$	square of y_6; product 5	3
$abab^2c$	product 4	3
$abab\,a^{-1}b^{-1}a^{-1}b^{-1}$	$[a, y_5]$	3
$abca^{-1}c^{-1}b^{-1}$	$[a, y_6]$	3
$bab^2cb^{-1}a^{-1}b^{-1}c^{-1}b^{-1}$	$[y_5, y_6]$	3

Table 3: The block words of the gadget relators in Step 4 of the proof of Lemma B.1, up to cyclic permutation and inversion, with the number of cells of the reductions listed after it. "Product i" is the product relator of the i-th equation in (24). All other block words are freely trivial, for instance both blocks of the square of y_2 and of every commutator within the first two equations; the relators $[y_i, J]$ are mapped exactly to I.

- $bab^2cb^{-1}a^{-1}b^{-1}c^{-1}b^{-1} \sim a^{-1}b^{-1}c^{-1}ab^2cb^{-1} \to a^{-1}b^{-1}a^{-1}bcb^{-1} \to a^{-1}b^{-1}a^{-1}c \to e$, using $c^{-1}a \to a^{-1}b^{-1}$ and then $bc \to cb$.

The kept relators have defect at most $\sqrt{2}\varepsilon < 3\varepsilon$, so γ_2 is a 3ε-representation of $\Gamma(A, b)$ on $\mathbb{C}^{4D} = \mathbb{C}^{8d}$ with $\gamma_2(J) = -I$. This proves (G3).

(G1) and (G2). Let Ψ be the composite of the substitutions of the four steps: $y \mapsto x_f x_v$ for $y \in S_1$, and $g \mapsto x_g$ for $g \in S_0$. Every relator of $\mathcal{P}$ is sent to a word that is trivial in G' (Step 1), hence in $\widehat{G}$, $\widehat{G}_1$ and $\Gamma(A, b)$, since each step sends relators to trivial words. Put $x_{j_0} = x_{t^*}$. The word $[x_w, x_{t^*}]J$ is trivial in $\widehat{G}$ (Step 2), hence in $\Gamma(A, b)$. The areas of these finitely many trivial words depend only on $\mathcal{P}$; let $A_0 \geq 1$ bound them. For the presentation of Proposition 10.3, $A_0 = 42$ (Remark B.2).

Accounting. The generators of $\widehat{G}$ other than J are the n_0 of $\mathcal{P}$, then f, v and the c_1 involutions Z_{bc} if $S_1 \neq \emptyset$, and t^*, u: N in all. Its conjugacy relations are the c_0 of $\mathcal{P}$, the $2c_1$ of Step 1 and $t^*wt^* = u$: c in all. Step 3 adds $3N + 1 + 2c$ involutions and $2N + 2c$ linear sets, and turns the c conjugacy relations into $N + 2c$ nice ones; Step 4 adds $7(N + 2c)$ variables and $6(N+2c)$ equations. So $n = N+(3N+1+2c)+7(N+2c) = 11N+16c+1$, and the linear sets of $\mathcal{P}$, the set $\{w, u\}$, those of Step 3 and the gadget equations give $m = \ell_0+1+(2N+2c)+6(N+2c)$ equations. All have at most three variables, and only $x_w + x_u = 1$ has right-hand side 1. □

Remark B.2 (The constant A_0 for $\mathcal{P}$). For the presentation $\mathcal{P}$ of Proposition 10.3 we may take $A_0 = 42$, and then $C_\Psi = A_0(8A_0 + 17) = 14826$ in Theorem B.3. For every relator r of $\mathcal{P}$, and for the word $[x_w, x_{t^*}]J$, we have written down a product of conjugates of relators of $\Gamma(A, b)$, in the presentation of Definition 2.7, that is freely equal to the word; the number of factors bounds its area.

The basic piece is the gadget of a nice relation $aba = c$. The six equations (24), together with the commutators $[a, y_6]$ and $[b, y_3]$, give the word $aca^{-1}b^{-1}$ area at most 8. A conjugacy relation $aba = c$ of $\widehat{G}$ then gives the word $aba^{-1}c^{-1}$ area at most 22:

- $a = \bar{a}^{-1}g^{*-1}$ by the linear set $\{a, g^*, \bar{a}\}$;
- g^* commutes with b because $\{b, g^*, \bar{b}\}$ is an equation, so $[b, g^*]$ is a relator of $\Gamma(A, b)$;
- $b = b''^{-1}b'^{-1}$;

- the gadgets of $\bar{a}b''\bar{a} = c'$ and $\bar{a}b'\bar{a} = c''$ turn $\bar{a}^{-1}b\bar{a}$ into $c'^{-1}c''^{-1}$;
- one commutation and the linear set $\{c, c', c''\}$ give c.

This uses the relators $[a, g^*]$ of $\Gamma(A, b)$ directly, rather than the nice relations $g^*a'g^* = a''$ of Step 3, which serve (G3).

The resulting areas are as follows:

- the squares and the linear relations of $\mathcal{P}$ have area 1;
- the commutations inside a linear set have area 2;
- the other relations of C_0 have area 23;
- the relations of C_1 have area 42, by four gadgets through $vbv = Z_{bc}$ and $fZ_{bc}f = c$;
- the word of (G2) has area 24.

The longest conjugator that occurs has 59 letters, within the bound $8 + 4A_0$ used in the proof of Theorem B.3. The program `a0/a0_check.py` in the ancillary files rebuilds every certificate from `egamma0_P.json` and `game_Ab.json`. It checks that each factor is a conjugate of a relator of $\Gamma(A, b)$ and that the product reduces freely to the word. It also rejects corrupted certificates and corrupted systems (Appendix C).

B.3 The embedding theorem

Theorem B.3 (Slofstra's embedding with $w \mapsto -I$; cf. [SV18, Prop. 6.2])**.** *Let $\mathcal{P}$ and w be as in Lemma B.1, let (A, b), Ψ, x_{j_0} and A_0 be as there, $\Gamma = \Gamma(A, b)$, and put $C_\Psi = A_0(8A_0 + 17)$.*

(a) (From $\mathcal{P}$ to Γ.) Every ε-representation of $\mathcal{P}$ on $\mathbb{C}^d$ with $w \mapsto -I$ yields a 3ε-representation of Γ on $\mathbb{C}^{8d}$ with $J \mapsto -I$. Hence $\mathrm{hlp}_\Gamma(J; 2, 3\varepsilon) \le 8\ \mathrm{hlp}_\mathcal{P}(w; 2, \varepsilon)$ for every $\varepsilon > 0$.

(b) (From Γ to $\mathcal{P}$, in a state.) Let (ϕ, ρ) be a state-weighted model of Γ at precision e on $\mathbb{C}^d$ (Definition 2.5) with $\phi(J) = -I$, and $\xi = \rho^{1/2}$. Then $(\phi \circ \Psi, \rho)$ is a state-weighted model of $\mathcal{P}$ at precision $C_\Psi e$, and

$$\|(\phi(x_w) - I)\xi\|_2 \ \ge\ 1 - C_\Psi e.$$

In (b) the involution w is *visible*: in the state, $\phi(\Psi(w)) = \phi(x_w)$ stays at distance about 1 from I. Nothing in (b) is flattened, and w need not be central.

Proof. (a) is (G3), and the inequality follows by (1), by which defect 2 means $\phi(w) = -I$, resp. $\phi(J) = -I$.

(b) The relators of Γ have length at most 4, since its equations have at most three variables, and those of $\mathcal{P}$ have length at most 4 (Definition 10.1). By (S2) for Γ, every generator of Γ has state commutator at most e.

(S2) for $\mathcal{P}$. Each $\Psi(g)$ has at most two letters, so $\|[\phi(\Psi(g)), \xi]\|_2 \le 2e$ by Lemma 3.3(a).

(S1) for $\mathcal{P}$. Let $r \in R$. The word $\Psi(r)$ has at most 8 letters and, by (G1), area at most A_0 in Γ. By Lemma 3.4 it is freely equal to a product of at most A_0 conjugates of relators of Γ and their inverses, with conjugators of length at most $8 + 4A_0$. By Lemma 3.3(d),

$$\|(\phi(\Psi(r)) - I)\xi\|_2 \ \le\ A_0\big(1 + 2(8 + 4A_0)\big)e = A_0(8A_0 + 17)\, e = C_\Psi e.$$

Visibility. Let $W = \phi(x_w)$ and $V = \phi(x_{j_0})$, and $X = WVW^*V^*$. The word $[x_w, x_{j_0}]J$ has 5 letters and, by (G2), area at most A_0; as above, $\|(\phi([x_w, x_{j_0}]J) - I)\xi\|_2 \le A_0(1 + 2(5 + 4A_0))e = A_0(8A_0 + 11)e$. Since $\phi(J) = -I$, this says $\|(X + I)\xi\|_2 \le A_0(8A_0 + 11)e$. On the other hand

$X-I = W(VW^*V^*-I)+W(I-W^*)$, so by Lemma 3.3(c), applied to the conjugate VW^*V^*, and $\|(W^*-I)\xi\|_2 = \|(W-I)\xi\|_2$,

$$\|(X-I)\xi\|_2 \;\le\; 2\|(W-I)\xi\|_2 + 2\|[V,\xi]\|_2 \;\le\; 2\|(W-I)\xi\|_2 + 2e.$$

Since $\|\xi\|_2 = 1$, $2 = \|2\xi\|_2 \le \|(X+I)\xi\|_2 + \|(X-I)\xi\|_2$. Hence $\|(W-I)\xi\|_2 \ge 1 - \big(\frac{1}{2}A_0(8A_0+11)+1\big)e \ge 1 - C_\Psi e$, as $A_0 \ge 1$. □

Remark B.4 (Sources). Steps 1, 3 and 4 of Lemma B.1 follow [Slo19, Prop. 4.8, Lemma 4.4, Prop. 4.2]; Step 3 introduces two nice relations for each conjugacy relation, so the counts of Lemma B.1 are those of this construction and not those of [Slo19, Remark 4.5]. Step 2 is the commutator device of [Slo19, Lemma 5.2], where an involution t with $tat = Ja$ is adjoined; here $t^*wt^* = u$ and $wu = J$. By [Slo19, Remark 4.5], Lemma 4.4 and Proposition 4.2 there turn a d-dimensional ε-representation into a $4d$-dimensional $C\varepsilon$-representation with $C \le 75$; here the block words of Step 3 have area at most 1 and those of Step 4 at most 3, which gives the factor 3, and Step 1 doubles the dimension once more. Theorem B.3 is the case $|S_1| \le 1$ of [SV18, Prop. 6.2] (for two or more non-involutive generators see the discussion before Lemma 11.13), which is stated there for every extended $E\Gamma_0$ presentation, with unspecified constants, for a central involution w that is nontrivial in approximate representations, and whose lower bound passes through an amplification of w to $-I$ [SV18, Prop. 2.7]. Part (b) replaces that amplification by the visibility estimate, which keeps the state; the inequality $2 \le \|(X+I)\xi\|_2 + \|(X-I)\xi\|_2$ is the state-weighted form of [SV18, Lemma 6.3]. Linear system games and their solution groups go back to Cleve and Mittal [CM14] and Cleve, Liu and Slofstra [CLS17].

C Machine checks

Four finite statements of this paper are verified by computer, and no proof depends on a computation except these:

(C1) the 27 words of R_{P} represent 1 in G_{P} (Lemma 7.2);

(C2) the presentation $\mathcal{P}$: its relations hold in G_{P}^{+} under the defining words, and the 29 derivations behind Proposition 10.3(d), with the area bound $A_* = 60$;

(C3) the linear system (A, b) of the game is the output of Lemma B.1 for $\mathcal{P}$ (Remark 10.7);

(C4) the area certificates of Remark B.2, which give $A_0 = 42$ in Lemma B.1.

Each is checked by at least two programs that share no code. The words of (C1) are evaluated by three programs. In (C2) the relations of $\mathcal{P}$ are checked in G_{P}^{+} by three programs and the 29 derivations are replayed by four; the area bound 60 is established by two of these, `area_check.py`, through a fresh search with an exact count of cells, and `area_xverify.py`, through explicit products of conjugates of relators checked by free reduction, while `independent_check.py` gives the weaker bound 335 and the count kept by `egamma0_certificate.py` is not a cell count. For (C3), `game_Ab_generate.py` builds the system from $\mathcal{P}$ and does not validate its input; `game_Ab_check.py` re-derives the system on its own and compares. For (C4), `a0/a0_check.py` and `a0/a0_xverify.py` both replay every certificate and both require at most 42 factors. Every checking program runs negative controls (Appendix C.3). Every other statement has a written proof in the paper that does not depend on a computation: in particular the pair areas (Theorem 7.8), the Dehn bound of Appendix A, the qubit count, the transfer, the sector theorem, the hash models, Lemma B.1, Theorem B.3 and Lemma 11.13. The value of A_0 enters only the constants c and ε_0 of Theorem A. This appendix describes the programs, which are in the ancillary directory `anc/`, and lists $\mathcal{P}$ in full (Appendix C.5).

C.1 The files

The programs need Python 3, and `numpy` and `scipy` only for the numerical sanity tests, which no proof uses. Each program is run inside `anc/` as `python3 -I <program> <arguments>`, takes a few seconds (the numerical attacks a few minutes), and exits with status 0 exactly when all its checks pass. In the checks of (C1)–(C3) all arithmetic is exact: words are expanded, freely reduced and evaluated in a faithful model of G_{P}^{+} with integer and $\mathbb{F}_2$ arithmetic. Table 4 lists the files.

file	content
`relators_pauli.txt`	the 27 words of R_{P}, in the abbreviations of Definition 7.1, with $J = xzxz$
`relators_pauli_CORRUPTED.txt`	a negative control: the word $(xz)^4$ replaced by $(xz)^2$
`check_pauli_lamps.py [file]`	the model of G_{P} and the check of (C1); log `run.log`, and `negative_control.log` on the corrupted file
`egamma0_P.json`	$\mathcal{P}$: the defining words of the 26 involutive generators, C_0, C_1, the linear sets, and the 29 derivations
`egamma0_certificate.py`	modes `verify` (the default), `build` and `negctl`; imports the model of `check_pauli_lamps.py` and extends it by ι; logs `egamma0_verify.log`, `egamma0_build.log`, `egamma0_negctl.log`
`independent_check.py <dir>`	an independent checker of (C1) and (C2); log `independent_check.log`
`area_check.py <dir>`	a second independent checker of (C2), with an exact area count; log `area_check.log`
`area_xverify.py <dir>`	a third independent checker of the derivations of (C2): explicit certificates and the cap 60, with eight negative controls (`--negctl`); log `area_xverify.log`
`game_Ab_generate.py <dir>`	builds (A, b) from `egamma0_P.json` by the four steps of Lemma B.1; writes `game_Ab.json`, log `game_Ab_generate.log`
`game_Ab.json`	the game: 1993 variables, each with its role in Lemma B.1, and 1687 equations with their right-hand sides
`game_Ab_check.py <dir>`	an independent checker of (C3); log `game_Ab_check.log`
`pauli_sector_numeric.py`	a numerical sanity test of a flat qubit count; no log
`numerics/analytic_checks.py`	numerical attacks on Theorem 4.1, Lemma 3.6, Lemma 5.3 and Theorem 5.4 (on the magic square game, with a negative control) and Theorem 11.9; log `numerics/analytic_checks.log`
`sv_balanced_lift.py`	a numerical check of Lemma 11.13, with the obstruction for the literal lifts as a negative control; log `sv_balanced_lift.log`
`a0/a0_check.py <dir>`	rebuilds and checks the certificates of (C4) from `egamma0_P.json` and `game_Ab.json`, with nine negative controls (`--negctl`); writes `a0/a0_certificates.json` with `--dump`; log `a0/a0_check.log`
`a0/a0_xverify.py <dir> <file>`	an independent checker of the dumped certificates of (C4), with eight negative controls (`--negctl`); log `a0/a0_xverify.log`
`README.md`	how to run everything

Table 4: The files in `anc/`.

For (C1) and (C2) the programs form four independent code bases:

- `check_pauli_lamps.py` and `egamma0_certificate.py`, which share one model of G_{P}^{+};

- `independent_check.py`, with its own model of G_{P}^{+} and its own reading of R_{P} and of $\mathcal{P}$;
- `area_check.py`, likewise;
- `area_xverify.py`, which checks the derivations only, with its own transcription of R_{P}.

The last three import none of the other programs. For (C3), `game_Ab_check.py` imports nothing and re-derives the system on its own.

C.2 The model

An element of G_{P}^{+} is stored as a pair: a signed Pauli string $J^c X^a Z^b$, with $c \in \mathbb{F}_2$ and a, b finite sets of sites, which is a normal form in E_V; and an affine bijection of V, given by its action on configurations. The letters p, q, ι act by the permutations of Y of Definition 6.1 and Lemma 6.9, t by the translation τ_1, and d by the transvection $E_{(1,2)\leftarrow(1,1)}$; x, z are the Pauli pair at the site 0. A lamp letter read after an affine prefix is moved to the image of its site. A word u is trivial exactly when its Pauli part is the identity and its affine part α is the identity.

The programs check $\alpha(0) = 0$, and $\alpha(e_y) = e_y$ for every point y of height at most $|u| + 3$ on each ray (or a little higher). This suffices. Once $\alpha(0) = 0$, α is linear, so it is the composite of the linear parts of its letters: permutations of Y for p, q, ι, the identity for t, and for d a transvection that moves only configurations containing $(1, 1)$. A permutation changes heights by at most 1, so a point of height at least $|u| + 3$ stays at height at least 2 while the word is applied; there d acts trivially and each permutation acts as a translation along a ray or a permutation of rays. Hence $\alpha(e_{(r,h)}) = e_{(r',h+k)}$ for all $h \geq |u| + 3$, with r' and k independent of h, and the check at $h = |u| + 3$ forces $r' = r$ and $k = 0$. `game_Ab_check.py` uses a different representation: the linear part of an affine map is stored by the ray permutation and translations that it eventually induces, and by its values on the basis vectors below the least height from which on it acts so. This form is canonical, so equality of elements is decided exactly.

C.3 What each program checks

(C1). `check_pauli_lamps.py` evaluates the 27 words of R_{P}; checks that $J \neq 1$ commutes with all six generators, that $(x_u z_u)^2 = J$ at the eight sites u of a test ladder, and that the 224 commutators of lamps at distinct sites of the ladder are trivial; and rejects four built-in corrupted words. On `relators_pauli_CORRUPTED.txt` it reports the word $xzxz$ as failed and exits with status 1. `independent_check.py` evaluates R_{P} in its own model, both as written and with $q \mapsto \iota p \iota$, checks $q = \iota p \iota$ and $[J, \iota] = 1$, and rejects twelve test words that are nontrivial in G_{P}^{+}. `area_check.py` evaluates R_{P} a third time and finds that its longest freely reduced word has 22 letters.

(C2): the relations. `egamma0_certificate.py verify` checks that every defining word is an involution in G_{P}^{+}; that every relation of C_0, C_1 and of the linear sets holds in G_{P}^{+}; that every pair inside a linear set is a listed commutation; and that $w \mapsto J$, with the 25 triples $(g, w, w) \in C_0$ and $pwp^{-1} = w \in C_1$ present. `independent_check.py` and `area_check.py` re-check every relation in their own models, together with $w \mapsto J$, the centrality relations and the format of Definition 10.1. `area_check.py` also checks the counts of Proposition 10.3(a) and the composition $76 = 32 + 25 + 19$ of C_0, and finds that the 26 generators represent 26 distinct elements, that the free reduction of $\vartheta(r)$ has at most $\ell_E = 44$ letters for $r \in R$ (46 before free reduction), and that $\iota_1 \iota_2 \iota_1 = \sigma\iota$ (equation (19)).

(C2): the derivations. `egamma0_certificate.py verify` replays the 29 derivations against the lists of C_0, C_1 and linear sets alone, without the model, in the three stages of the proof of Proposition 10.3: the decomposition into p-conjugates by free reduction, the resolution through C_1, and the recorded moves. Its own count, one unit per move and per step of C_1, does not charge the uses of (E1) and is not used; nor does it check the labels `area_bound` stored with the derivations, which no program uses. `independent_check.py` recomputes the resolved sequence of every derivation from R_{P} with its own front end, and re-justifies each move as a rewriting along a cyclic rotation of a relator; with a generous count of every relator application, including the uses of $x^2 = e$ hidden in a swap and the s-lemma at each occurrence of s, it finds area at most 335 per word. `area_check.py` searches afresh, by a shortest-path search over sequences of involutive generators that uses only listed relations, for a derivation of each of the 29 words, and counts cells exactly in the free group on $S_0 \sqcup S_1$, charging every relator and every use of (E1). Each word has area at most 60, attained by $(sps\,p^{-1})^3$; the s-lemma costs 9 cells. It also replays the recorded derivations under the same count, with the same maximum 60. `area_xverify.py` turns each recorded derivation into an explicit certificate: a list of conjugates $ur^{\pm1}u^{-1}$ of relators r of $\mathcal{P}$, one or more per stage and move. Each literal $[p, \iota p \iota]$ is replaced by σ through the certificate of the s-lemma, each conjugation by p costs one relator of C_1 and each inverted letter one relator g^2, and a move costs one relator, or two or three when a linear set is used out of its listed order. The program then checks, separately from this construction, that every factor names a relator of $\mathcal{P}$ with sign exactly $+1$ or -1 and that the product reduces freely to the target word, which it builds from its own transcription of Table 2. Word by word its counts equal those of the replay in `area_check.py`: 9 for the s-lemma and at most 60, attained by $(sps\,p^{-1})^3$. It requires at most 60 factors for every word, and reports the replay and this cap separately. The recorded derivations use 32 triples of C_0, none among the 25 triples (g, w, w) or the 19 added commutations, and both searching checkers confirm that derivations exist without the added commutations.

(C2): negative controls. `egamma0_certificate.py negctl` deletes a triple of C_0 used by a derivation, falsifies a triple of C_1, redefines w as xz, and deletes a commutation of a linear set; each corrupted file is rejected. `independent_check.py` rejects the false relation $xtx = txt$, a derivation after deletion of the triple (σ, g_4, g_5), a falsified triple of C_1, an illegal cancellation, and a claimed derivation of the nontrivial word xz. `area_check.py` rejects the false relation $xtx = txt$, finds no derivation of the nontrivial words xz and $(xzx)(ztz)$, and none of the s-lemma once the triple (g_1, g_2, g_3) is deleted. `area_xverify.py --negctl` rejects a certificate with a factor dropped, a valid certificate padded to 62 factors by a factor and its inverse, which only the cap rejects, a factor with sign 2, the presentation without the triple (σ, g_4, g_5), a derivation with one recorded move falsified, one with its resolved sequence falsified, and two claimed certificates of the nontrivial word xz.

(C3). `game_Ab_generate.py` applies Steps 1–4 of Lemma B.1 to $\mathcal{P}$, with the numbering of their proof. It only transforms its input: it does not check that the relations of $\mathcal{P}$ hold, and it produces a system from a corrupted presentation as well. It finds 40 generators besides J after Step 2, 97 conjugacy relations, 234 nice ones after Step 3, and $283 + 6 \cdot 234 = 1687$ equations in $26 + 12 + 2 + 315 + 1638 = 1993$ variables. One equation, $x_w + x_u = 1$, has two variables and right-hand side 1; the others have three variables and right-hand side 0. Every product is increasing in the numbering, including those of the eight linear sets of $\mathcal{P}$ (Appendix C.5), so Lemma B.1 applies. `game_Ab_check.py` checks the format, the SHA-256 digest of `egamma0_P.json`, and these counts, and that the equations, read as sets of variable roles with their right-hand sides, are exactly those of its own derivation of Steps 1–4 from `egamma0_P.json`. In its own model of G_{P}^{+} it re-checks every relation of $\mathcal{P}$, and forms the

representations $\gamma, \gamma_1, \gamma_2$ of the proof of Lemma B.1 from the defining words, with the central involution J of G_{P}^{+} in place of -1: every variable becomes an 8×8 block-monomial matrix over G_{P}^{+}. The 1993 squares, the 1993 commutators with J, the 5059 commutators within equations and the 1687 product relators of $\Gamma(A, b)$ all evaluate to the identity, while J goes to $J\, I_8 \neq I$; so this is a homomorphism $\Gamma(A, b) \to G_{\mathrm{P}}^{+} \wr \mathrm{Sym}(8)$ in which $J \neq 1$. It also checks by Gaussian elimination that $Ax = b$ has no solution over $\mathbb{F}_2$ (A has rank 1512). It rejects a gadget equation with the wrong right-hand side or with a wrong variable, a deleted equation, and the system with $b = 0$ in the equation $x_w + x_u = 1$, which becomes solvable.

(C4). `a0/a0_check.py` finds every variable of the game through the role recorded for it in `game_Ab.json`, so it does not depend on the numbering. For each of the 121 words of (G1) and (G2), namely the 120 relators of $\mathcal{P}$ and the word $[x_w, x_{t^*}]J$, it instantiates the template of Remark B.2 for the type of the word and records each cell as a conjugate of a relator. It then checks the certificate. Every factor must be a square, a commutator of two variables that share an equation, or the product relator of an equation of the file, and the product of the conjugates must reduce freely to $\Psi(r)$, which is computed directly from `egamma0_P.json`. The largest number of factors is 42, attained by the relations of C_1. A valid certificate stays valid when a factor and its inverse are appended, so the replay alone does not bound the area; the program therefore also requires every certificate to have at most 42 factors, and reports the replay and this cap separately. It rejects nine corruptions: a factor inverted or dropped, a commutator of variables in no common equation, a wrong variable in a gadget equation, a deleted equation, the system with $b = 0$ in $x_w + x_u = 1$, a false triple of C_0, a true certificate checked against a corrupted game, and a true certificate of 42 factors padded to 44 by a factor and its inverse, which only the cap rejects. `a0/a0_xverify.py` shares no code with it. Before it builds any word, it validates every factor of the dumped certificates against the game: the key of the relator must have its exact form and length, with integer indices that name a variable, an equation, or two distinct variables of one equation; the sign must be exactly $+1$ or -1; and every letter of the conjugator must be one of $x_1^{\pm 1}, \ldots, x_n^{\pm 1}, J^{\pm 1}$. It then rebuilds every factor from `game_Ab.json`, forms the target words from `egamma0_P.json` on its own, repeats the comparison, and requires the same cap of 42 factors. With `--negctl` it rejects a dump padded in the same way, factors with signs 2 and 0, a dropped factor, and four dumps that only the validation rejects: in each, a factor and its inverse, which cancel freely, are appended to a certificate of one factor, and they name the square of a variable that the game does not have, the relator $[x_j, J]$ for j one past the last variable, the equation of index -1, or a conjugate by a letter outside this alphabet.

Numerical tests. Part 2 of `check_pauli_lamps.py` and `pauli_sector_numeric.py` test, on random near-Pauli instances and on exact Pauli instances, a flat qubit count of the form $\log_2 D \geq \sum_i \big[1 - b_i^2/(2 \ln 2) - h(\min(a_i^2/2, \tfrac{1}{2}))\big]$, a variant of Corollary 4.3, and a scalar entropy inequality on a grid. They are sanity tests, and no proof uses them. `sv_balanced_lift.py` checks Lemma 11.13 for $m \leq 8$. It confirms that the model of Example 11.11 has the single inexact relator stated there; that the literal lifts of [Slo19, Prop. 4.8] are approximate representations whose marked involution W has the eigenvalues -1 and $+1$ with multiplicities d and $3d$, and that $\|TWT^* + W\|_f^2 \geq 2$ for the exchange of the two halves, for all 24 block permutations and for random unitaries, with equality for a block permutation, so that the relator $tctu^{-1}$ keeps defect $\sqrt{2}$; that τ, f' and v' have the stated properties, that Steps 1–3 keep every defect at most e_m, and that the marker is balanced; and that the solution group built in Step 4, here with 262 variables and 206 equations, gets an $8d$-dimensional lift with $J \mapsto -I$ in which all 2373 relators have defect at most $e_m/\sqrt{2}$. No proof uses it.

C.4 Rerunning and regenerating

The following runs, made inside `anc/` with `python3 -I`, reproduce the shipped logs. In `run.log`, `negative_control.log`, `independent_check.log` and the three `egamma0` logs, the last line `exit` N records the exit status and was appended by the shell.

arguments of `python3 -I`	log
`check_pauli_lamps.py`	`run.log`
`check_pauli_lamps.py relators_pauli_CORRUPTED.txt`	`negative_control.log` (status 1)
`egamma0_certificate.py verify`	`egamma0_verify.log`
`egamma0_certificate.py build`	`egamma0_build.log`
`egamma0_certificate.py negctl`	`egamma0_negctl.log`
`independent_check.py .`	`independent_check.log`
`area_check.py .`	`area_check.log`
`area_xverify.py .`, and with `--negctl`	`area_xverify.log`
`game_Ab_generate.py .`	`game_Ab_generate.log`
`game_Ab_check.py .`	`game_Ab_check.log`
`a0_check.py ..`, and with `--negctl`; inside `a0/`	`a0/a0_check.log`
`a0_xverify.py ..  a0_certificates.json`; inside `a0/`	`a0/a0_xverify.log`
the same with `--negctl`; inside `a0/`	`a0/a0_xverify.log`
`numerics/analytic_checks.py all`	`numerics/analytic_checks.log`
`sv_balanced_lift.py`	`sv_balanced_lift.log`

Both data files are regenerated byte for byte. `egamma0_certificate.py build` reconstructs the presentation and its derivations from `relators_pauli.txt` into a temporary file, reports that it is byte-identical to `egamma0_P.json`, deletes it, and verifies the shipped file. Its search meets more involutions than the derivations use; it keeps those that occur on the returned derivations, 26 in all, and numbers them in the order in which it met them. `game_Ab_generate.py` overwrites `game_Ab.json` with a file identical to the shipped one. The SHA-256 digests are

```
egamma0_P.json   ad0fd91d9378bc956a78e65c813b94fbb93a0e7090b3068b4d76ac6cbc1590f0
game_Ab.json     c82d4cbbd72b0683ba2a7fceba8c92c346a4c3565e3785ddd2a56415f569c889
```

and `game_Ab_check.py` checks the first of them. In `game_Ab.json`, the product relator of each equation is J^{b_i} times the product of its variables in the listed order.

The comments and messages of the programs cite statements of this paper by their numbers here, with their labels in the LaTeX source in brackets, such as Lemma B.1 [`lem:gadget`]. The references to Definition 4.7 and Section 6 of the papers of Slofstra and of Slofstra and Vidick are to [Slo19] and [SV18].

C.5 The presentation $\mathcal{P}$ in full

The 26 involutive generators, in their listed order, are $\iota, t, d, x, z, \sigma, g_1, \ldots, g_{20}$. Table 5 gives their defining words $\vartheta(g)$ in p, ι, t, d, x, z; negative exponents are inverse letters, and ι, t, d, x, z are their own defining words. The non-involutive generator is p, and $w = g_{14}$.

σ	$p\iota p\iota p^{-1}\iota^{-1}p^{-1}\iota^{-1}$	g_1	$p\iota p^{-1}$
g_2	$p^2\iota p^{-2}$	g_3	$p\iota p\iota p^{-1}\iota p^{-1}$
g_4	$p^2\iota p\iota p^{-1}\iota^{-1}p^{-1}\iota^{-1}p^{-1}$	g_5	$p\iota p\iota p^{-1}\iota^{-1}p^{-1}\iota^{-1}p^2\iota p\iota p^{-2}\iota^{-1}p^{-1}\iota^{-1}$
g_6	$p^3\iota p\iota p^{-1}\iota^{-1}p^{-1}\iota^{-1}p^{-2}$	g_7	ptp^{-1}
g_8	pdp^{-1}	g_9	$p\iota p\iota p^{-1}\iota^{-1}p^{-1}\iota^{-1}pd\iota p\iota p^{-1}\iota^{-1}p^{-1}\iota^{-1}$
g_{10}	$tptp^{-1}$	g_{11}	p^2tp^{-2}
g_{12}	xzx	g_{13}	zxz
$g_{14} = w$	$xzxz$	g_{15}	xtx
g_{16}	txt	g_{17}	$xtxt$
g_{18}	tzt	g_{19}	ztz
g_{20}	$ztzt$		

Table 5: The defining words of the involutive generators of $\mathcal{P}$ other than ι, t, d, x, z.

C_1**.** The ten relations $pbp^{-1} = c$ of type (E4) are, as pairs (b, c): (d, g_8), (g_1, g_2), (g_{14}, g_{14}), (g_4, g_6), (g_7, g_{11}), (ι, g_1), (σ, g_4), (t, g_7), (x, x), (z, z).

Linear sets. The eight linear sets are $\{\iota, \sigma, g_3\}$, $\{t, g_7, g_{10}\}$, $\{t, g_{15}, g_{17}\}$, $\{t, g_{19}, g_{20}\}$, $\{x, g_{13}, g_{14}\}$, $\{x, g_{16}, g_{17}\}$, $\{z, g_{12}, g_{14}\}$, $\{z, g_{18}, g_{20}\}$; each product is taken in the order listed, which is the order of the generators.

C_0**.** The 76 triples $(a, b, c) \in C_0$, meaning $aba = c$, are: the 25 centrality triples (g, w, w), $g \neq w$; the 32 triples used by the derivations, (d, t, g_{10}), (g_1, g_2, g_3), (g_1, g_4, g_4), (g_1, g_7, g_7), (g_1, g_8, g_8), (g_1, x, x), (g_1, z, z), (g_{11}, d, d), (g_{12}, g_{19}, t), (g_4, g_9, d), (g_4, σ, g_5), (g_7, d, d), (ι, g_4, g_4), (ι, g_7, g_7), (ι, g_8, g_8), (σ, g_4, g_5), (σ, g_6, g_6), (σ, g_7, t), (σ, g_8, g_9), (t, g_7, g_7), (t, x, g_{16}), (t, z, g_{18}), (x, d, d), (x, g_{18}, g_{18}), (x, ι, ι), (x, t, g_{15}), (x, z, g_{12}), (z, d, d), (z, g_{16}, g_{16}), (z, ι, ι), (z, t, g_{19}), (z, x, g_{13}); and the 19 commutations inside linear sets not already listed, (g_{10}, t, t), (g_{12}, z, z), (g_{13}, x, x), (g_{15}, g_{17}, g_{17}), (g_{15}, t, t), (g_{16}, g_{17}, g_{17}), (g_{16}, x, x), (g_{17}, t, t), (g_{17}, x, x), (g_{18}, g_{20}, g_{20}), (g_{18}, z, z), (g_{19}, g_{20}, g_{20}), (g_{19}, t, t), (g_{20}, t, t), (g_{20}, z, z), (g_3, ι, ι), (g_3, σ, σ), (g_7, g_{10}, g_{10}), (ι, σ, σ).

The 29 derivations. Table 6 gives, for each target word, its resolved sequence of involutive generators and the moves that reduce it to the empty word: c cancels aa; $\kappa(a, b)$ replaces aba by the generator c with $(a, b, c) \in C_0$; m replaces two letters of a linear set by the third; λ deletes the three letters of a linear set. The words of R_{P} are written with the family letters of Table 2 and the abbreviations of Definition 7.1 ($s = [p, q]$, $B = psp^{-1}$, $t_m = p^{m-1}tp^{1-m}$, $J = xzxz$); a capital letter is the inverse of the lower-case one; and q stands for $\iota p\iota$.

target	resolved sequence	moves
$[p, \iota p \iota]\sigma^{-1}$	$g_1\, g_2\, g_1\, \iota\, \sigma$	$\kappa(g_1, g_2)$, λ
H: `s s`	$\sigma\, \sigma$	c
H: `s p s P s p s P s p s P`	$\sigma\, g_4\, \sigma\, g_4\, \sigma\, g_4$	$\kappa(\sigma, g_4)$, $\kappa(g_4, \sigma)$, c
H: `s p p s P P s^-1 p p s^-1 P P`	$\sigma\, g_6\, \sigma\, g_6$	$\kappa(\sigma, g_6)$, c
H: `p s P q s^-1 Q`	$g_4\, \iota\, g_1\, g_4\, g_1\, \iota$	$\kappa(g_1, g_4)$, $\kappa(\iota, g_4)$, c
Q: `t t`	$t\, t$	c
S: `t s p T P s^-1`	$t\, \sigma\, g_7\, \sigma$	$\kappa(\sigma, g_7)$, c
S: `t s q T Q s^-1`	$t\, \sigma\, \iota\, g_1\, g_7\, g_1\, \iota\, \sigma$	$\kappa(g_1, g_7)$, $\kappa(\iota, g_7)$, $\kappa(\sigma, g_7)$, c
S: `d B s p D P s^-1 B^-1`	$d\, g_4\, \sigma\, g_8\, \sigma\, g_4$	$\kappa(\sigma, g_8)$, $\kappa(g_4, g_9)$, c
S: `d B s q D Q s^-1 B^-1`	$d\, g_4\, \sigma\, \iota\, g_1\, g_8\, g_1\, \iota\, \sigma\, g_4$	$\kappa(g_1, g_8)$, $\kappa(\iota, g_8)$, $\kappa(\sigma, g_8)$, $\kappa(g_4, g_9)$, c
C: `t1 t2 t1^-1 t2^-1`	$t\, g_7\, t\, g_7$	$\kappa(t, g_7)$, c
A: `d t1 D t2^-1 t1^-1`	$d\, t\, d\, g_7\, t$	$\kappa(d, t)$, λ
A: `d t2 D t2^-1`	$d\, g_7\, d\, g_7$	$\kappa(g_7, d)$, c
A: `d t3 D t3^-1`	$d\, g_{11}\, d\, g_{11}$	$\kappa(g_{11}, d)$, c
Z: `p x P X`	$x\, x$	c
Z: `p z P Z`	$z\, z$	c
Z: `q x Q X`	$\iota\, g_1\, x\, g_1\, \iota\, x$	$\kappa(g_1, x)$, $\kappa(x, \iota)$, c
Z: `q z Q Z`	$\iota\, g_1\, z\, g_1\, \iota\, z$	$\kappa(g_1, z)$, $\kappa(z, \iota)$, c
Z: `d x D X`	$d\, x\, d\, x$	$\kappa(x, d)$, c
Z: `d z D Z`	$d\, z\, d\, z$	$\kappa(z, d)$, c
L: `x x`	$x\, x$	c
L: `z z`	$z\, z$	c
L: `x z x z x z x z`	$x\, z\, x\, z\, x\, z\, x\, z$	$\kappa(x, z)$, $\kappa(z, x)$, m, λ
P: `x t x T X t X T`	$x\, t\, x\, t\, x\, t\, x\, t$	$\kappa(x, t)$, $\kappa(t, x)$, m, λ
P: `x t z T X t Z T`	$x\, t\, z\, t\, x\, t\, z\, t$	$\kappa(t, z)$, $\kappa(x, g_{18})$, $\kappa(t, z)$, c
P: `z t x T Z t X T`	$z\, t\, x\, t\, z\, t\, x\, t$	$\kappa(t, x)$, $\kappa(z, g_{16})$, $\kappa(t, x)$, c
P: `z t z T Z t Z T`	$z\, t\, z\, t\, z\, t\, z\, t$	$\kappa(z, t)$, $\kappa(t, z)$, m, λ
K: `J t J^-1 T`	$x\, z\, x\, z\, t\, z\, x\, z\, x\, t$	$\kappa(x, z)$, $\kappa(z, t)$, $\kappa(x, z)$, $\kappa(g_{12}, g_{19})$, c
$xzxz\, w^{-1}$	$x\, z\, x\, z\, g_{14}$	$\kappa(x, z)$, λ

Table 6: The 29 derivations of Proposition 10.3(d), as recorded in `egamma0_P.json`.

D Formal verification and its scope

The Lean sources are in the `lean/` directory of the accompanying repository. The exact types, dependency pins, source hashes, third-party notices and commands for a separate build and kernel replay are given in `lean/README.md`, the dependency record and the source manifest in that directory. The frozen file `Challenge.lean` contains historical admitted declarations; these are statement markers and are not the proofs of the certified roots. The final `Audit.lean` compares its six analytic statement types with the proved declarations and prints the axioms of every listed root. The A/E/F aggregate propositions are defined in `Blueprint.lean`, beyond the original frozen statement file.

Paper statement	Lean declaration (qualified by namespace)
Theorem A	`Blueprint.theoremA`
Lemma 3.4	`Blueprint.DiagramProof.conjugatorLengths`
A second Theorem A wrapper	`Blueprint.DiagramProof.theoremA`
Pauli hash models used in the upper bound	`Blueprint.Published.pauliHashModels`
Theorem 4.1	`Companion.Analytic.CountProof.count`
Theorem C and Corollary 4.2	`Companion.Analytic.CountProof.uniformCount`
Theorem 5.1	`Companion.Analytic.TransferProof.transfer`
Theorem 5.4, including the specified measurement and zero error	`Companion.Analytic.ReturnProof.returnTheorem`
Positive-error existence from Theorem 5.4	`Companion.Analytic.ReturnProof.returnExistence`
Theorem D and the three conclusions of Corollary 5.5	`Companion.Analytic.ProfileProof.twoSided`
Theorem E	`Blueprint.theoremE`
Theorem F	`Blueprint.theoremF`
Theorem 8.1, entropy and rank inequalities	`Blueprint.ASector.source_sector`
Proposition 9.4	`Blueprint.AHash.general_hash_models`
Literal boundary merge in the conjugator proof	`Blueprint.DiagramProof.merge_matching`

Exact scope. The type of `Blueprint.theoremA` is `Published.TensorPerfectFinite` → `AStatement`, in namespace `Blueprint`. The latter includes the concrete counts, arity three, quantum value one, exclusion of finite perfect strategies, pure and mixed lower bounds, the maximally entangled upper bound, and dimension-profile exponents. It additionally includes exclusion of perfect spatial strategies; that conclusion uses the published parameter described above. The pure lower bound also bounds entropy by log rank. The entropy asymptotic is supplied by the lower and upper bounds rather than a separate entropy-limit conjunct. The uniform count includes the rank conclusion. The analytic statements are finite-dimensional; linear-system rows are nonempty, sorted, duplicate-free and bounded, with an admissible question distribution. The return theorem fixes Bob's rounded observables and Alice's sequential measurement with parity correction. The sector and hash-model adapter declarations retain, respectively, a conjugator-length and a Pauli-hash-model parameter; the Theorem A assembly supplies both proofs internally. The sector root records the entropy and rank inequalities, not all intermediate sector-state conclusions as separate conjuncts. The literal boundary merge is an internal combinatorial lemma, not an additional paper theorem. The E/F propositions use the definitions of profiles in `Blueprint/RatesDefinitions.lean` and

`Blueprint/FSVGame.lean`. The K profile has the explicit range $0 < \delta \leq 2$.

What is not formalized. The external perfect-tensor-strategy implication itself is a parameter rather than a Lean proof. There is no standalone certified root for Theorem B, no claim that every proposition, discussion, literature result or appendix of the paper has been formalized, and no formal certification of the ancillary numerical experiments, the prose, or its identification with the definitions. Historical comments in the unchanged statement and proof sources may describe an earlier state of completion; the proved declarations and their audited types specify the present claims. In particular, the remaining point of trust is that the definitions in `Challenge.lean` mean what the paper means.

References

[BDSW96] Charles H. Bennett, David P. DiVincenzo, John A. Smolin, and William K. Wootters. Mixed-state entanglement and quantum error correction. *Phys. Rev. A*, 54:3824–3851, 1996. arXiv:quant-ph/9604024.

[Bri02] Martin R. Bridson. The geometry of the word problem. In Martin R. Bridson and Simon M. Salamon, editors, *Invitations to Geometry and Topology*, volume 7 of *Oxford Graduate Texts in Mathematics*, pages 29–91. Oxford University Press, 2002.

[Cho80] Ching Chou. Elementary amenable groups. *Illinois J. Math.*, 24:396–407, 1980.

[Cir80] Boris S. Cirel'son. Quantum generalizations of Bell's inequality. *Lett. Math. Phys.*, 4:93–100, 1980.

[CLS17] Richard Cleve, Li Liu, and William Slofstra. Perfect commuting-operator strategies for linear system games. *J. Math. Phys.*, 58:012202, 2017. arXiv:1606.02278.

[CM14] Richard Cleve and Rajat Mittal. Characterization of binary constraint system games. In *Automata, Languages, and Programming (ICALP 2014)*, volume 8572 of *Lecture Notes in Comput. Sci.*, pages 320–331. Springer, 2014. arXiv:1209.2729.

[Col20] Andrea Coladangelo. A two-player dimension witness based on embezzlement, and an elementary proof of the non-closure of the set of quantum correlations. *Quantum*, 4:282, 2020. doi:10.22331/q-2020-06-18-282; arXiv:1904.02350.

[Con76] Alain Connes. Classification of injective factors. Cases II_1, II_∞, III_λ, $\lambda \neq 1$. *Ann. of Math. (2)*, 104:73–115, 1976.

[CRSV17] Rui Chao, Ben W. Reichardt, Chris Sutherland, and Thomas Vidick. Overlapping qubits. In *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*, volume 67 of *LIPIcs*, pages 48:1–48:21, 2017.

[dC06] Yves de Cornulier. Finitely presented wreath products and double coset decompositions. *Geom. Dedicata*, 122:89–108, 2006.

[dlS25] Mikael de la Salle. Spectral gap and stability for groups and non-local games. *J. Éc. polytech. Math.*, 12:1417–1444, 2025. arXiv:2204.07084.

[DM26a] Seth Douglas and Nidhal Mghirbi. Amenable groups with nearly exponential sofic profile, and quantum channels that need nearly linear memory. arXiv:2610.04271v1, identical to Zenodo version 1.1.2, doi:10.5281/zenodo.23111927. Unrefereed preprint; result numbers refer to this version, 2026.

[DM26b] Seth Douglas and Nidhal Mghirbi. Houghton's group H_3 has superpolynomial sofic profile. Version 1.1.2, Zenodo, 2026. doi:10.5281/zenodo.23070869. Unrefereed preprint. Result numbers refer to version 1.1.2.

[DPP19] Ken Dykema, Vern I. Paulsen, and Jitendra Prakash. Non-closure of the set of quantum correlations via graphs. *Comm. Math. Phys.*, 365:1125–1142, 2019. arXiv:1709.05032.

[ES05] Gábor Elek and Endre Szabó. Hyperlinearity, essentially free actions and L^2-invariants. The sofic property. *Math. Ann.*, 332:421–441, 2005.

[Gao15] Jingliang Gao. Quantum union bounds for sequential projective measurements. *Phys. Rev. A*, 92(5):052331, 2015.

[GH17] W. T. Gowers and O. Hatami. Inverse and stability theorems for approximate representations of finite groups. *Sb. Math.*, 208:1784–1817, 2017. arXiv:1510.04085.

[Gro99] Misha Gromov. Endomorphisms of symbolic algebraic varieties. *J. Eur. Math. Soc.*, 1:109–197, 1999.

[GYZ26] Minbo Gao, Tianshi Yu, and Lihong Zhi. An explicit counterexample to Tsirelson's problem via a linear system game, 2026. arXiv:2610.10248.

[Hou78] C. H. Houghton. The first cohomology of a group with permutation module coefficients. *Archiv der Mathematik*, 31:254–258, 1978.

[JLV20] Zhengfeng Ji, Debbie Leung, and Thomas Vidick. A three-player coherent state embezzlement game. *Quantum*, 4:349, 2020. arXiv:1802.04926.

[Joh99] D. L. Johnson. Embedding some recursively presented groups. In *Groups St Andrews 1997 in Bath, II*, volume 261 of *London Mathematical Society Lecture Note Series*, pages 410–416. Cambridge University Press, 1999.

[KTW14] Jedrzej Kaniewski, Marco Tomamichel, and Stephanie Wehner. Entropic uncertainty from effective anticommutators. *Phys. Rev. A*, 90:012332, 2014. arXiv:1402.5722.

[Lee12] Sang Rae Lee. *Geometry of Houghton's groups*. PhD thesis, University of Oklahoma, 2012. arXiv:1212.0257v1; result numbers refer to that version.

[Lie73] Elliott H. Lieb. Convex trace functions and the Wigner–Yanase–Dyson conjecture. *Adv. Math.*, 11:267–288, 1973.

[Lin74] Göran Lindblad. Expectations and entropy inequalities for finite quantum systems. *Comm. Math. Phys.*, 39:111–119, 1974.

[Lin75] Göran Lindblad. Completely positive maps and entropy inequalities. *Comm. Math. Phys.*, 40:147–151, 1975.

[Lin23] Junqiao Lin. Tracially embeddable strategies: lifting MIP* tricks to MIP^{co}, 2023. arXiv:2304.01940.

[LS77] Roger C. Lyndon and Paul E. Schupp. *Combinatorial Group Theory*, volume 89 of *Ergebnisse der Mathematik und ihrer Grenzgebiete*. Springer-Verlag, Berlin, 1977.

[MdlS23] Amine Marrakchi and Mikael de la Salle. Almost synchronous correlations and Tomita–Takesaki theory, 2023. arXiv:2307.08129.

[MNY22] Hamoon Mousavi, Seyed Sajjad Nezhadi, and Henry Yuen. Nonlocal games, compression theorems, and the arithmetical hierarchy. In *Proceedings of the 54th Annual ACM Symposium on Theory of Computing (STOC 2022)*, pages 1–11, 2022. doi:10.1145/3519935.3519949; arXiv:2110.04651.

[MV15] Laura Mančinska and Thomas Vidick. Unbounded entanglement in nonlocal games. *Quantum Inf. Comput.*, 15:1317–1332, 2015. Conference version, *Unbounded entanglement can be needed to achieve the optimal success probability*, in ICALP 2014, doi:10.1007/978-3-662-43948-7_69; arXiv:1402.4145.

[NV17] Anand Natarajan and Thomas Vidick. A quantum linearity test for robustly verifying entanglement. In *Proceedings of the 49th Annual ACM Symposium on Theory of Computing (STOC 2017)*, pages 1003–1015, 2017. arXiv:1610.03574.

[Ost16] Dimiter Ostrev. The structure of nearly-optimal quantum strategies for the CHSH(n) XOR games. *Quantum Inf. Comput.*, 16:1191–1211, 2016. arXiv:1506.00607, as *The structure of optimal and nearly-optimal quantum strategies for non-local XOR games.*

[OV18] Dimiter Ostrev and Thomas Vidick. Entanglement of approximate quantum strategies in XOR games. *Quantum Inf. Comput.*, 18:617–631, 2018. arXiv:1609.01652.

[Pad22] Connor Paddock. Rounding near-optimal quantum strategies for nonlocal games to strategies using maximally entangled states, 2022. arXiv:2203.02525. Result numbers refer to version 5 (September 2025).

[Pes08] Vladimir G. Pestov. Hyperlinear and sofic groups: a brief guide. *Bull. Symbolic Logic*, 14:449–480, 2008. arXiv:0804.3968.

[PS70] Robert T. Powers and Erling Størmer. Free states of the canonical anticommutation relations. *Comm. Math. Phys.*, 16:1–33, 1970.

[Slo11] William Slofstra. Lower bounds on the entanglement needed to play XOR non-local games. *J. Math. Phys.*, 52:102202, 2011. arXiv:1007.2248.

[Slo18] William Slofstra. A group with at least subexponential hyperlinear profile, 2018. arXiv:1806.05267v1.

[Slo19] William Slofstra. The set of quantum correlations is not closed. *Forum Math. Pi*, 7:e1, 2019. arXiv:1703.08618. Result numbers refer to arXiv:1703.08618v2.

[Slo20] William Slofstra. Tsirelson's problem and an embedding theorem for groups arising from non-local games. *J. Amer. Math. Soc.*, 33:1–56, 2020. arXiv:1606.03140.

[SV18] William Slofstra and Thomas Vidick. Entanglement in non-local games and the hyperlinear profile of groups. *Ann. Henri Poincaré*, 19:2979–3005, 2018. doi:10.1007/s00023-018-0718-y; arXiv:1711.10676. Result numbers refer to arXiv:1711.10676v2.

[Uhl77] Armin Uhlmann. Relative entropy and the Wigner–Yanase–Dyson–Lieb concavity in an interpolation theory. *Comm. Math. Phys.*, 54:21–32, 1977.

[Vid22] Thomas Vidick. Almost synchronous quantum correlations. *J. Math. Phys.*, 63:022201, 2022. arXiv:2103.02468.

[Vil09] Cédric Villani. *Optimal Transport: Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, 2009.

[Wei00] Benjamin Weiss. Sofic groups and dynamical systems. *Sankhyā Ser. A*, 62:350–359, 2000.