Amenable groups with nearly exponential sofic profile, and quantum channels that need nearly linear memory

Seth Douglas Nidhal Mghirbi

November 2026 (version 1.2.0)

Abstract

How far from finite can an amenable group be? Cornulier's sofic profile counts the points needed to imitate a fixed finite piece by permutations at a given accuracy. We build a finitely presented elementary amenable group with nearly exponential profile, close to the ceiling of our counting method. Its lamps are symmetric groups on three letters placed on binary configurations of three rays. The group embeds in Brin's group 3V, which therefore also has a finite piece of nearly exponential profile. A sweep form of the counting criterion charges a whole ordered suffix at once. Quadratic sweep fillings improve the lower bound for the dyadic lamplighter to exponential in the square root of the accuracy parameter, for the same finite piece.

The configuration group gives an explicit quantum channel in dimension 873 with Choi rank 130. A device that applies it repeatedly, releasing each output before the next input arrives, needs memory proportional to the number of uses divided by its purity budget, up to polylogarithmic factors; exact devices attain this tradeoff. At the optimal exchange rate the least memory is of square-root order, up to subpower factors, attained with purity of the same order. The channel is factorizable and is a limit of channels with finite maximally mixed baths, but accurate such baths require nearly exponential dimension in inverse error. A pure environment of dimension 130 implements it exactly. The engine is a one-round theorem: after one use, the device's memory carries an approximate representation of the group.

Contents

1 Introduction **3**

2 Counting approximately commuting copies of S_3 **7**
2.1 Whole suffix sweeps 9

I The group (Theorem A) **10**

3 Configuration lamps **11**
3.1 A certificate of 25 relations 12
3.2 Theorem A 13

4 Local fillings **14**

5 Pair areas **16**

6 Hash models, and the proof of Theorem A **18**
6.1 Affine hashing and the upper bound 18
6.2 Proof of Theorem 3.3 20

7 More about Γ **21**
7.1 A finite presentation . . . 21
7.2 An embedding in a Brin–Thompson group . . . 23
7.3 The Følner function . . . 25

8 The ladder below Γ **25**
8.1 An amenable group with a square-root lower bound . . . 26
8.2 Thompson's group V . . . 28

II The channel (Theorem B) 31

9 Causal services and gadget channels **31**

10 The weighted one-round theorem **34**

11 The channel of the configuration lamps, and Theorem B **35**

12 Sofic models give devices, and streaming at the rank rate **37**
12.1 An upper bound from sofic models . . . 37
12.2 Streaming at the rank rate . . . 39

13 A purity budget at the rank rate **40**

14 Proof of Theorem B **41**

III Physics and open problems 44

15 Local physics **44**
15.1 Local reservoirs are cheap at every finite time . . . 44
15.2 Autonomous limits are self-adjoint . . . 45
15.3 An expensive autonomous limit: the lazy doubling . . . 46

16 Open problems **48**

Use of AI **49**

A Tools from Houghton's group **49**

B The counting criterion and sequential extraction **51**
B.1 Three facts about copies of S_3 . . . 52
B.2 Proof of the counting criterion . . . 54
B.3 Proof of sequential extraction . . . 54

C Proof of the weighted one-round theorem **55**
C.1 Weighted norms . . . 55
C.2 Extraction . . . 56
C.3 Charged area . . . 58
C.4 Exact S_3 rounding in place . . . 59
C.5 A weighted sector bound . . . 59
C.6 Completion . . . 61

D Proof of the streaming theorem **62**

E The lazy doubling in detail **64**

F Finer constraints on autonomous limits **69**

G Hyperlinearity transfers from F to $\mathcal{L}$ **70**

H The grid host **71**

I Results from the companion papers, and machine checks **73**
I.1 Machine checks 73

1 Introduction

Amenable groups are the infinite groups closest to finite ones. Every finite piece of an amenable group can be imitated by permutations of a finite set, with products correct on all but a small fraction of the points. How many points does such an imitation need? For a finite piece E of a group, a *chunk*, Cornulier's *sofic profile* $\sigma_E(r)$ is the least number of points on which E can be modelled with multiplication correct, and distinct elements separated, on all but a fraction $1/r$ of the points [25]. He asked whether a sofic group can have a profile that is unbounded and not linear in r [25, Problem 3.18].

Groups whose chunks all have finite profile are the sofic groups of Gromov and Weiss [34, 70]; amenable and residually finite groups are sofic. Whether every group is sofic was open until 2026, when OpenAI announced a proof that the unit group of the binary Leavitt algebra is not sofic [62]. Fournier-Facio then gave a finitely presented torsion-free example [29], and Kun and Thom showed that generalized wreath products, with lamps $\mathbb{Z}/2$ over the cosets of a subgroup, are not sofic for suitable pairs of residually finite groups with property (T) [54]. In a non-sofic group some chunk has infinite profile, so the question becomes how large a profile can be when it is finite. Our group Γ has the shape of those of Kun and Thom, lamps over a permutation action, but it is amenable, so all its profiles are finite. Slofstra and Vidick tied the hyperlinear profile, its analogue for unitary matrices, to the entanglement needed to play linear-system non-local games near-optimally [68]. Slofstra then gave a finitely presented group with a chunk whose profile is at least $\exp(r^c)$ for some $c > 0$, but his group is not known to be sofic, so this profile may be infinite [67].

In earlier work we found a profile of this size in an amenable group, where every profile is finite. Houghton's group H_3 is elementary amenable and finitely presented [14], with an explicit presentation due to Johnson [48], and one of its chunks has profile at least $\exp(\Omega(r^{0.129}))$ [26]. To our knowledge it was the first amenable group shown to have superpolynomial profile. This contradicts Cornulier's extension theorem [25, Theorem 3.15], which would bound that profile by $O(r^2)$, and the conclusion drawn from it in [25, Example 3.16], that every elementary amenable group has polynomial profile; Section 7 of [26] shows where the argument for that theorem breaks. The proof is a counting criterion (Theorem 2.1). If a few relations force many displaced copies of S_3 to commute cheaply, every model must be large, because nearly commuting copies of S_3 carry entropy. The criterion can never give more than $\exp(O(r))$. This paper shows that an amenable group can come close to that ceiling.

Theorem A. There is a finitely presented elementary amenable group Γ, generated by five elements, with a chunk E for which the following hold. Put $\kappa = 6 + \log 107 < 12.75$.

(a) There are absolute constants $c, C > 0$ such that for all $r \geq 2$

$$\exp\Big(\frac{c\,r}{(\log r)^{\kappa}}\Big) \;\leq\; \sigma_E(r) \;\leq\; \exp(C\,r\log r).$$

So $\sigma_E(r) = \exp(r^{1+o(1)})$. Every chunk E' of Γ satisfies $\sigma_{E'}(r) \leq \exp(C_{E'}\,r\log r)$ for all $r \geq 2$. The lower bound is certified by an explicit set of 25 relations of Γ.

(b) The same relations control unitary models. Two of the generators, a and c, give $b = ac^4$, and a, b generate a copy of S_3 in Γ. Let $0 < \Delta \leq 1$. There are $c_\Delta, e_\Delta > 0$ with the following property. Let ϕ assign unitaries in $U(D)$ to the five generators so that each

of the 25 relations is mapped within e of I in the normalized Hilbert–Schmidt norm, where $0 < e \le e_\Delta$, and put $\phi(b) = \phi(a)\phi(c)^4$. If $\|\phi(a)\phi(b) - I\|_2 \ge \Delta$, then

$$\log D \;\ge\; c_\Delta \, \frac{e^{-2}}{(\log(1/e))^{2\kappa}}.$$

Theorem 3.3 gives the precise statement, with the chunk E and the 25 relations of Table 1. The lower bound is within a polylogarithmic factor in the exponent of the most the criterion can give. The exponent κ records the price of bringing two lamps together: about N^2 local moves, each paid for by the Dehn function of H_3, which grows at most like $n^{4+\log 107}$ [26, Theorems 3.2 and 3.3]. That polynomial bound is the one input of Theorem A whose proof we do not reproduce; with Lee's exponential bound alone the method gives only $\exp(c\, r^{1/(1+B)})$ for a constant B that we cannot make explicit (Remark 6.4). The 107 is the factor by which a doubling endomorphism of H_3 multiplies area in [26, Section 3]; it is the only constant in κ that is not structural, and a quadratic Dehn function for H_3 would give $\kappa = 4$. The upper bound comes from models that hash configurations by random affine maps. These are far smaller than the models that Følner sets give [25, Proposition 3.14], [4, Section 4.3], because the Følner function of Γ is doubly exponential (Proposition 7.3). The finite presentation (Proposition 7.1) comes from Cornulier's criterion for permutational wreath products [24, Theorem 1.1]. Finally, Γ embeds in Brin's higher-dimensional Thompson group $3V$ (Proposition 7.2). So $3V$, a finitely presented simple group whose soficity is open, has a chunk whose profile is finite and $\exp(r^{1+o(1)})$.

Configuration lamps. In Houghton's group the copies of S_3 sit on the points of a ray. Bringing two of them together costs a word as long as the distance k between them, and commuting them costs at least of order k^2 relations [26, Proposition 3.4]. A family of N copies spreads over distance N, so its commutators are expensive. We move the lamps off the points and onto the finitely supported binary configurations of the three rays of H_3. Houghton's group moves the points. Finitely supported affine maps act on the configurations: a translation flips one point, and a transvection adds one point into another. One copy of S_3, a lamp, sits on each configuration (Figure 1). The 2^N configurations supported on the first N points of a ray are reached by words of length $O(N^2)$, and two of them differ by a configuration that at most N transvections clear down to a single point. Each step touches at most three points, so its cost is governed by the Dehn function of H_3 at scale N, which is polynomial [26, Theorems 3.2 and 3.3]. Commuting two of the 2^N lamps therefore costs a polynomial in N, which is polylogarithmic in the number of lamps. That is exactly what the criterion needs.

A ladder. The configuration lamps are the last step of a ladder of transport mechanisms, each of which the criterion turns into a profile bound.

transport	group	$\log \sigma_E(r) \gtrsim$	profile finite
along a ray	H_3 [26]	$r^{0.129}$	yes
by one element	dyadic $\mathcal{G}$ (§8.1)	$r^{1/2}$	yes
exponential orbits	$\mathcal{L} \le V$ (§8.2)	$r/\log^2 r$	yes [72]
configurations	Γ (§§3–6)	$r/(\log r)^\kappa$	yes

The pairwise criterion for transport by the powers of one element gives exponent $1/(p+1)$ when far commutators have area of order k^p. Along a ray of H_3 the area is at least quadratic [26, Proposition 3.4], so the pairwise estimate stops at $1/3$. The sweep criterion

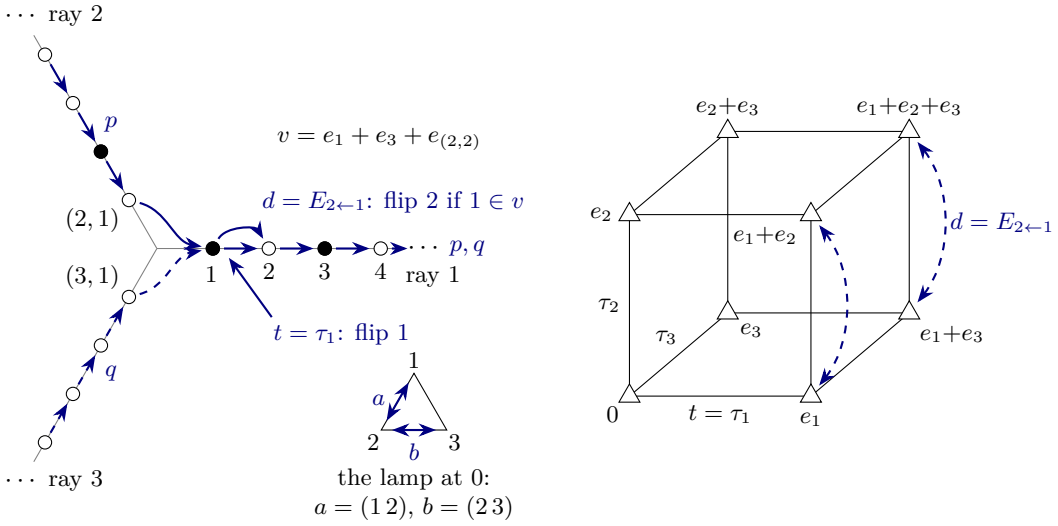


Figure 1: Left: the points of the three rays, a configuration v (filled points), the lamp at 0, and the generators as moves: p and q move points, and t and d move configurations. Right: the 2^3 configurations supported on the points $1, 2, 3$, each carrying a lamp, joined by the translations τ_1, τ_2, τ_3 and folded by the transvection $d = E_{2\leftarrow 1}$, which exchanges e_1 with $e_1 + e_2$ and $e_1 + e_3$ with $e_1 + e_2 + e_3$ and fixes the other four. N points carry 2^N lamps, and any two are brought together by moves that each involve at most three points.

instead charges a whole suffix at once; quadratic sweeps give the dyadic group $\mathcal{G}$ exponent $1/2$ with the same seventeen relations. Thompson's group F transports cheaply along exponentially growing orbits, and a finitely presented subgroup $\mathcal{L}$ of Thompson's group V gives $\log \sigma_E(r) \geq cr/\log^2 r$ for a fixed chunk. The recent preprint [72, Theorem 1.1] proves that V is sofic, so the embedding $\mathcal{L} \leq V$ gives finite profile for this chunk. Proposition 8.5 also proves directly that $\mathcal{L}$ is hyperlinear exactly when F is. The configuration lamps achieve a nearly linear lower exponent inside an elementary amenable group.

Quantum memory. The second half of the paper turns Γ into a quantum channel. Picture a device that must apply the same channel n times. Each input arrives, the device acts on it together with its memory, and the output leaves before the next input comes in. The device may swap memory qubits for fresh ones and may start with some purity, but it never resets, and every qubit it keeps counts. How much memory does it need? This question goes back to Kretschmann and Werner, who showed that every causal quantum process is a concatenation of memory channels [52], and to Rybár and Ziman, who asked when a device that is never reset applies the same channel to each of its inputs. They showed that random unitary channels can be applied this way with a fixed finite memory, while nonunital channels cannot [65]. We let the memory grow with the number of uses and ask how fast it must grow. We set up the resource accounting in [27]. The *exchange* counts the qubits the device swaps out, and the *purity* counts the qubits it starts with or imports, minus their entropy (an entropy deficit, not $\operatorname{Tr}\rho^2$). The error is measured against observers who choose their inputs adaptively and keep quantum references, as for quantum strategies, combs and process tensors [37, 21, 63]. A finite set of relations of a group defines a channel, its *gadget channel*, whose memory cost is governed by how well the group can be imitated: small memory would give small approximate representations. For a gadget channel in the closure of channels with finite tracial factorizations, the optimal exchange is $\frac{1}{2}\log r_*$ qubits per use, where r_* is the Choi rank [27, Propositions 3.1 and 9.1]. We call it the rank rate.

The quantum reverse Shannon theorem simulates n uses of a channel as one block, acting

on all inputs at once with free shared entanglement, and charges classical communication [5, 6]; our device sees one input at a time and pays in memory. Storage and retrieval keeps an unknown unitary in a quantum memory for later use [66]; our channel is known, and memory is needed because it is not unitary. Memory costs are known for classical simulations of sequences of quantum measurements [50, 15] and for quantum models of stochastic processes and of adaptive agents [35, 28]; their targets have temporal correlations, while n independent uses of one channel have none. Our devices have the form of collision models [23]: the memory plays the system, and the inputs and imported qubits are units that each meet it once. Collision models usually fix the collisions and derive the dynamics of the system; we fix the channel that the inputs see and ask how large the system must be. Purity, counted as above, is the asymptotic measure of the resource theory of purity, or informational nonequilibrium [44, 33]: maximally mixed qubits are free, and its free operations from a system to itself, the noisy operations, are the channels with a finite tracial factorization that appear in Theorem C.

One use reveals nothing: the channel below lives on $\mathbb{C}^{873}$, and a single round is easy. The dimension 873 is what the gadget construction gives for the 29 words that define the channel; we made no attempt to minimize it. The cost shows up only over many rounds. From H_3 we obtained a channel that needs memory $(n/\log n)^{0.069}$ at logarithmic purity [27]. The configuration lamps give nearly linear memory, and a trade-off between memory and purity that is sharp up to polylogarithmic factors. All these bounds are asymptotic: the polylogarithmic factors have exponent up to $2\kappa + 1 < 26.5$, and the constants, which come from the Dehn function of H_3, are not evaluated, so the lower bounds take effect only for very large n.

Theorem B. The 25 relations of Theorem A, with two new letters for $b = ac^4$ and $j = ab$ and four short words that relate them (29 words in seven letters), define an explicit quantum channel Φ on $\mathbb{C}^{873}$. It lies in the closure of channels with finite tracial factorizations, and its Choi rank is $r_* = 130$. With κ as in Theorem A, there are constants $c, c_0, K > 0$ and n_0 such that the following hold for causal services of n uses of Φ with error at most $1/16$.

(a) *Memory against purity.* If $n \geq n_0$ and $\log n \leq B \leq c_0 n$, every service with purity at most B has memory at least $c\,n/(B(\log n)^{2\kappa})$, at any exchange. For every $n \geq 1$ and $0 < B \leq n$ there is an exact service, with error 0, purity less than B and exchange $(n-1)Q$, whose memory is $Q \leq K(n/B)\log(2n/B)$. In particular, services with purity $O(\log n)$ need memory $n^{1-o(1)}$, while memory $Kn\log n$ suffices at zero purity.

(b) *The rank rate.* If $n \geq n_0$, every service with exchange at most $\frac{n}{2}\log r_*$ has memory at least $c\sqrt{n}/(\log n)^{\kappa}$, whatever its purity. For every $n \geq 2$ there is a service with exchange at most $\frac{n}{2}\log r_*$ whose memory and purity are both at most $K\sqrt{n\log n}$.

So memory and purity trade at a fixed rate: in the range of (a), the least memory at purity B, times B, is $n^{1+o(1)}$. With exchange at most the rank rate the least memory is $n^{1/2+o(1)}$, and it is attained with purity of the same order.

Theorem 11.1 gives the precise statement, and Figure 4 draws it. The engine is a one-round theorem (Theorem 10.1) for every gadget channel whose relations contain those of S_3 and a name for ab: a single round of any device reveals an approximate representation of the group on the device's memory, with relator defect of order the square root of the leakage and entropy production of that round. In [27, Theorem 5.5 and Appendix B] this was proved for one gadget, built from Houghton's group. The upper bounds come from sofic models, which give devices directly (Proposition 12.1). At most points of a model the pure bath state reproduces the channel exactly, so a memory that is maximally mixed on those points serves the channel with no error at all. At the rank rate a device cannot buy memory with purity, because the purity it can spend is at most linear in its memory (Proposition 13.1).

Finite environments. The same bounds say how far from finite the channel itself is. A channel is factorizable if it factors through a finite von Neumann algebra with a faithful trace [2, 38], and it has a finite tracial factorization if it is $X \mapsto (\mathrm{id} \otimes \tau_M)[U(X \otimes I)U^*]$ for a unitary U on $\mathbb{C}^d \otimes \mathbb{C}^M$, that is, if it can be implemented with a finite bath in its maximally mixed state. By a theorem of Haagerup and Musat, the closure of these channels consists of the channels that factor through an ultrapower of the hyperfinite II_1 factor, and whether it contains every factorizable channel, in every dimension, is equivalent to the Connes embedding problem [39, Theorems 3.6 and 3.7]. That problem has a negative answer [47], so some factorizable channels lie outside the closure. Inside it, Musat and Rørdam gave channels that need an infinite-dimensional ancilla [60]. Our channel measures how expensive the finite approximations can be.

Theorem C. The channel Φ of Theorem B is factorizable and lies in the closure of the channels with finite tracial factorizations, and its minimal Stinespring dilation implements it exactly with a pure environment of dimension 130. Yet for small u, every channel with a finite tracial factorization within accuracy u of Φ, in Choi or in diamond distance, and every finite mixture of such channels, needs a bath of dimension at least $\exp(c\,u^{-1}/(\log(1/u))^{2\kappa})$, while dimension $\exp(C\,u^{-1}\log(1/u))$ suffices.

Corollary 14.1 gives the precise statement. Channels in the closure with no finite tracial factorization were known [60], and for the Houghton channel of [27, Theorem D.1] the exponent of the profile lies between 0.069 and 1/3. What is new here is the quantitative profile: for Φ the exponent is $1 - o(1)$, close to the most the method can certify.

Physics. A local lattice system in ν dimensions whose bath starts maximally mixed produces, in time T, only channels that can be served with memory polynomial in T and $\log n$ (Proposition 15.1). So it needs time at least about $\delta^{-1/\nu}$ to produce Φ to accuracy δ (Corollary 15.2). Autonomous dynamics with a bath in its tracial state have long-time limits that are self-adjoint (Proposition 15.3; [49, Remark 10.3], [38, Remark 5.2]), yet such a limit can still be expensive: a bounded time-independent Hamiltonian, not local, produces in the long run a channel that trades memory and purity at the same rate as Φ (Proposition 15.6). Whether local autonomous dynamics can produce an expensive channel is open.

How to read this paper. Part I (Sections 3–8) proves Theorem A. A group theorist can read Sections 2–6 and stop: Sections 7 and 8 add the finite presentation, the embedding in $3V$ and the ladder; the bounds in parts (a) and (b) of Theorem A do not use them. Part II (Sections 9–14) proves Theorem B and needs from Part I only the definitions of Section 2, the statement of Theorem A and the pair-area theorem. Part III treats local physics and lists open problems. Long proofs are in the appendices. Appendix B reproduces the proofs of the two results from the companion papers on which the lower bounds of Theorems A and B stand, and Table 2 in Appendix I lists everything else we use from them. The certificate of 25 relations is in the accompanying files, with programs that check it. Logarithms are to base two unless written ln.

2 Counting approximately commuting copies of S_3

Every lower bound in this paper comes from one principle, proved in [26]. If finitely many relations force many displaced copies of S_3 to commute cheaply, every model of a fixed chunk must be large, because approximately commuting copies of S_3 carry entropy. We fix the conventions used in both halves of the paper, state the principle in the form used here, with a, b in place of the letters u, v of [26], and note the ceiling $\exp(O(r))$ that it can

never pass. The scheme, many cheaply transported copies of a finite group, relations whose cost is controlled by area, and a dimension bound, follows Slofstra [67, Sections 2–3]. The pairwise criterion needs only approximate commutation of displaced copies of S_3; Theorem 2.3 strengthens its permutation form using whole suffixes.

Conventions. Functions are composed from right to left, groups act on the left, $[x, y] = xyx^{-1}y^{-1}$, and log is to base two. For a finite set S and a finite set R of words in $S^{\pm 1}$, $\mathrm{Area}_R(w)$ is the least number of conjugates of words of $R^{\pm 1}$ whose product is w in the free group on S, and ∞ if there is none. A *chunk* of a group is a finite subset $E \ni 1$ with the partial multiplication $xy = z$ whenever $x, y, z \in E$ and $xy = z$ in the group. Its *sofic profile* $\sigma_E(r)$ is the least n for which some $f : E \to \mathrm{Sym}_n$ with $f(1) = \mathrm{id}$ multiplies and separates correctly up to $1/r$: $d(f(x)f(y), f(xy)) \le 1/r$ whenever xy is defined in E, and $d(f(x), f(y)) \ge 1 - 1/r$ for distinct $x, y \in E$ [25]. Here d is the normalized Hamming distance, the fraction of points on which two permutations differ. The normalized Hilbert–Schmidt norm of a $D \times D$ matrix is $\|X\|_2 = (\mathrm{Tr}(X^*X)/D)^{1/2}$.

The criterion is Theorem 8.1 of [26], and Appendix B reproduces its proof.

Theorem 2.1 (Counting criterion)**.** *Let G be a group, $S \subset G$ a finite set containing a, b, and R a finite set of words in $S^{\pm 1}$, each representing 1 in G, with $a^2, b^2, (ab)^3 \in R$ and $ab \neq 1$ in G. Let l be the largest length of a word in R, and let E be the chunk consisting of 1, $S^{\pm 1}$, the prefixes of the words of R and of ab, and the inverses of these elements. Let $w_1, \dots, w_N$ be words in $S^{\pm 1}$ and $\Lambda \ge 1$ such that*

$$\mathrm{Area}_R([w_i x w_i^{-1}, w_j y w_j^{-1}]) \le \Lambda \qquad (i \neq j,\ x, y \in \{a, b\}).$$

Then $\log \sigma_E(r) \ge N/16$ for every $r \ge 10^8 N\Lambda(2l-1)$. Moreover, for $0 < \Delta \le 1$ and $K_\Delta = \sqrt{80000\,(1 + \log(1/\Delta))}/\Delta$, every assignment ϕ of unitaries in $U(D)$ to S under which every word of R lies within e of I in the normalized Hilbert–Schmidt norm and $\|\phi(ab) - I\|_2 \ge \Delta$ satisfies $\log D \ge N\Delta^2/12$ whenever $K_\Delta \sqrt{N}\,\Lambda e \le 1$.

The proof in [26] has four steps, and only the last is specific to S_3. Conjugating the images of a and b by the image of w_i gives N copies of S_3, one for each word. Each copy satisfies the relations of S_3 up to the defect of a relator, because those relations are conjugates of a^2, b^2 and $(ab)^3$, so it can be rounded to an exact representation of S_3. Its three-cycle is a conjugate of the image of ab, so it is displaced from the identity as much as that image is. Two copies commute up to Λ times the relator defect, because each of their commutators is a product of at most Λ conjugates of relators. Finally, exact copies of S_3 that nearly commute and have displaced three-cycles cannot share a small space: they force a fixed amount of entropy per copy, at least $\Delta^2/12$ bits for unitaries and $1/16$ bit for permutations [26, Section 4]. The chunk E does not depend on N or on the words w_i; only the areas matter.

The theorem has a ceiling. Its hypotheses force $N \le r/10^8$ and $N \le (K_\Delta e)^{-2}$, so it never gives $\log \sigma_E(r)$ beyond order r, nor $\log D$ beyond order e^{-2}. Below the ceiling, everything depends on how cheaply the group carries copies of S_3 to many places. Transport by the powers of one element, with area of order k^p at distance k, gives exponent $1/(p+1)$; this is Corollary 8.2 of [26].

Corollary 2.2 (Transport by the powers of one element)**.** *Let G, S, R, l and E be as in Theorem 2.1, let $t \in S$, and suppose that $\mathrm{Area}_R([x, t^k y t^{-k}]) \le Ck^p$ for $x, y \in \{a, b\}$ and $k \ge 1$, with $C \ge 1$ and $p \ge 0$. Then*

$$\log \sigma_E(r) \ \ge\ \frac{1}{16}\left\lfloor \left(\frac{r}{10^8 C(2l-1)}\right)^{1/(p+1)} \right\rfloor \qquad (r > 1),$$

and every unitary assignment as in Theorem 2.1 has $\log D \ge \frac{\Delta^2}{12}\left((CK_\Delta e)^{-2/(2p+1)} - 1\right)$.

In Houghton's group H_3 the copies of S_3 along a ray have commutators of at least quadratic area [26, Proposition 3.4], so the pairwise estimate of Corollary 2.2 never passes exponent $1/3$ along that ray. Whole suffix sweeps can improve this conclusion, as Theorem 2.3 shows.

One more fact of Cornulier's is used twice, in Sections 7 and 8. The profile of a chunk is bounded exactly when the chunk embeds in a finite group, and otherwise it is at least r [25, Fact 3.5], indeed at least $2r$ [26, Lemma 2.1]. A chunk with unbounded profile therefore certifies that its group is not locally embeddable into finite groups, not LEF [25, Fact 3.8].

2.1 Whole suffix sweeps

Pairwise commutator estimates can overcount the cost of moving one copy past an ordered suffix. The following form charges one filling for the whole suffix. It uses permutation rounding in Hamming distance before passing to the entropy estimate.

Theorem 2.3 (Sweep counting criterion)**.** *Use the group, relators, chunk and transport words of Theorem 2.1, and suppose $N \geq 1$. For each of the six fixed normal forms*

$$\mathcal{N} = \{1, a, b, ab, ba, aba\}, \qquad W_i(g) = w_i g w_i^{-1},$$

assume that, for every i, every $h \in \mathcal{N}$ and every choice $g_j \in \mathcal{N}$ for $j > i$,

$$\mathrm{Area}_R\left([W_i(h), W_{i+1}(g_{i+1}) \cdots W_N(g_N)]\right) \leq \Lambda_N.$$

Then

$$(\Lambda_N + 40N)(2l-1) \leq r/20000 \quad \Longrightarrow \quad \log \sigma_E(r) \geq N/16. \tag{1}$$

If only the sweeps with $h = a, b$ are bounded by Λ_N, the same conclusion holds with $r/60000$ in the hypothesis in place of $r/20000$. In both versions every later factor ranges over all six normal forms.

Proof. Relator errors. Given a chunk model $f : E \to \mathrm{Sym}_n$, evaluate generator words by $\phi(s) = f(s)$ and $\phi(s^{-1}) = f(s)^{-1}$. Merging successive prefixes of a relator of length m costs at most $(m-1)/r$; replacing inverse letters costs at most m/r, since $s, s^{-1}, 1 \in E$. Thus each relator has Hamming defect at most $\epsilon = (2l-1)/r$. Bi-invariance implies that a word of area A has defect at most $A\epsilon$. The prefix of ab and expansiveness give $d(\phi(ab), 1) \geq 1 - 2/r$. All transported pairs have the same square and braid errors and three-cycle displacement: conjugation is an exact Hamming isometry, regardless of the transport word's length.

Native rounding. Let A, B be any transported pair. Replace cycles of length at least three in each permutation by fixed points, obtaining involutions P, Q with $d(P, A), d(Q, B) \leq \epsilon$. Put $R = PQ$. Then $d(R^3, 1) \leq 7\epsilon$. The identity $QRQ = R^{-1}$ implies that $\mathrm{Fix}(R^3)$ is invariant under Q, R, and hence under $P = RQ$. Keep P, Q on this set and make them both trivial on its invariant complement. They now give an exact permutation representation $u_i : S_3 \to \mathrm{Sym}_n$. Every normal form has length at most three: its first replacement costs at most 3ϵ, and deleting bad orbits costs at most 7ϵ. Consequently

$$d(u_i(g), \phi(W_i(g))) \leq 10\epsilon \qquad (g \in \mathcal{N}).$$

This construction allows trivial and sign orbits. In particular, the three-cycle loses at most 10ϵ of its Hamming displacement.

One ordered error. Put $U_{>i} = u_{i+1}(g_{i+1}) \cdots u_N(g_N)$ and let $V_{>i}$ be its raw-word counterpart. Replacing both occurrences of the first factor and of the suffix in the commutator gives

$$d([u_i(h), U_{>i}], 1) \leq (\Lambda_N + 20(N - i + 1))\epsilon \leq (\Lambda_N + 40N)\epsilon.$$

Let $F(g) = u_1(g_1)\cdots u_N(g_N)$, for uniform $g \in S_3^N$. Exact same-copy multiplication shows that the error $F(gh) - F(g)u_i(h)$ is exactly the error in commuting $u_i(h)$ past its suffix, up to unitary factors. For permutation matrices $\|P - Q\|_2^2 = 2d(P, Q)$; hence

$$T^2 := \max_{i,h} \mathbb{E}_g \|F(gh) - F(g)u_i(h)\|_2^2 \le 2(\Lambda_N + 40N)\epsilon.$$

Here h lies in factor i. There is no same-copy multiplication defect after rounding.

Entropy from this error. We give the sector argument to specify exactly what is used. Suppose $\|u_i(c) - I\|_2 \ge \Delta$ for a three-cycle c, $T \le \Delta$ and $T^2 \le 5/6$. On $\ell^2(S_3^N) \otimes \mathbb{C}^n$ use the regular action R and the isometry

$$Wv = |S_3^N|^{-1/2} \sum_g |g\rangle \otimes F(g)v, \qquad \Pi = WW^*, \quad \rho = \Pi/n.$$

The ordered-error identity gives $\|\mathsf{R}(h_i)W - Wu_i(h)\|_{HS} \le \sqrt{n}\,T$. Thus, for $Z_i = 2I - \mathsf{R}(c_i) - \mathsf{R}(c_i)^*$, $\mathrm{Tr}(\rho Z_i) \ge (\Delta - T)^2$. Let $\mathcal{E}_i$ average conjugation by the six regular operators in factor i. Its leakage outside Π is at most T^2. Pinching into Π and $I - \Pi$ gives ranks at most n and $5n$, because one of the six states being averaged is ρ. Therefore $S(\mathcal{E}_i\rho) \le \log n + h_2(T^2) + T^2 \log 5$. The $\mathcal{E}_i$ commute. Monotonicity of relative entropy, together with $S(\mathcal{E}\sigma) - S(\sigma) = D(\sigma\|\mathcal{E}\sigma)$, bounds each successive entropy increment by its increment on ρ. For $\bar\rho = \mathcal{E}_N \cdots \mathcal{E}_1\rho$ this yields

$$S(\bar\rho) \le \log n + N\{h_2(T^2) + T^2 \log 5\}.$$

On each irreducible sector of S_3^N, Z_i is $3I$ if factor i is standard and zero otherwise. If a sector has k standard factors, its dimension is 2^k. Since $\bar\rho$ commutes with the regular action, its entropy is at least the mean of k. The twirls preserve the expectation of each central Z_i, so $N(\Delta - T)^2 \le 3S(\bar\rho)$. Combining gives

$$\log n \ge N\{(\Delta - T)^2/3 - h_2(T^2) - T^2 \log 5\}.$$

No pairwise estimate entered this argument.

Constants. The stated hypothesis gives $T \le 1/100$ and $\epsilon \le 1/(800000N)$. The preceding displacement estimate gives $\Delta \ge 49/100$ (a deliberately conservative Hilbert–Schmidt lower bound). Using $\log_2(10000) < 14$ and $1/\ln 2 < 3/2$, one has

$$h_2(1/10000) \le 31/20000, \qquad (1/10000)\log 5 \le 6/20000.$$

Thus the entropy per copy is at least $(48/100)^2/3 - 37/20000 = 1499/20000 > 1/16$. For elementary-letter sweeps, split h into at most three letters and use the free commutator product identity. The full sweep area is at most $3\Lambda_N$, and $3\Lambda_N + 40N \le 3(\Lambda_N + 40N)$, proving the $1/60000$ version. □

The $40N$ term retains the counting method's linear scale in r even if sweep areas are very small. This is a restriction on lower bounds produced by the criterion, not an upper bound on any group's actual profile. The transported finite-group and area-to-defect strategy follows Slofstra and Slofstra–Vidick [67, 68]; metric approximation growth with growing balls, as studied by Arzhantseva–Cherix and Bradford [4, 71], is distinct from the fixed-chunk accuracy profile used here.

Part I
The group (Theorem A)

3 Configuration lamps

This section defines the group Γ of Theorem A, its five generators p, q, t, c, a and the 25 relations that certify its profile, and states Theorem A in its precise form. The idea is visible in Figure 1: the lamps, copies of S_3, sit on configurations rather than on points, so N points carry 2^N lamps while every generator touches at most three points. Two of the 2^N lamps are compared by about N^2 local moves, each filled with the Dehn function of H_3 at scale N, so the area is polylogarithmic in the number of lamps. That is what the criterion needs. Theorem 2.1 turns M copies of S_3 whose pairwise commutators have area at most Λ into $\log \sigma_E(r) \geq M/16$ at r of order $M\Lambda$, and it approaches its ceiling $\exp(O(r))$ only when Λ is tiny compared with M.

Put differently, the criterion needs three things at once: exponentially many sites reachable by short words, pairwise commutators of polylogarithmic area, and an amenable ambient group. The two groups of the ladder in Section 8 each meet two of the three. The dyadic group $\mathcal{G}$ is amenable and its orbit of 0 grows exponentially, but we cannot commute two copies at points of opposite parity cheaply. In the subgroup $\mathcal{L}$ of Thompson's group V everything works except amenability, which is the question whether F is amenable. The configuration lamps meet all three by changing what a site is. The sites are the finitely supported binary configurations of the points of Houghton's three rays, not the points themselves. Affine maps that touch a bounded number of coordinates at a time carry the copies between sites. And the ambient group is locally finite-by-$\mathbb{Z}^2$.

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 $H = H_3$ consists of the 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. The map $g \mapsto (-m_2(g), -m_3(g))$ is a surjective homomorphism $H \to \mathbb{Z}^2$ whose kernel is the group $\mathrm{FSym}(Y)$ of finitary permutations [45]. Let p be the permutation with $p(1, j) = (1, j + 1)$, $p(2, 1) = (1, 1)$ and $p(2, j) = (2, j - 1)$ for $j \geq 2$, fixing ray 3, and let q be defined in the same way with rays 2 and 3 exchanged. They generate H (Appendix A), and their images in $\mathbb{Z}^2$ are $(1, 0)$ and $(0, 1)$.

The group. Let $V = \mathbb{F}_2^{(Y)}$ be the space of finitely supported functions $Y \to \mathbb{F}_2$, the *configurations*,[1] with basis $(e_y)_{y \in Y}$. The group H acts 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 $\mathrm{GL}(\mathbb{F}_2^J)$ over finite $J \subset Y$, each extended by the identity on the other basis vectors. It is normalized by H, and $H \cap \mathrm{GL}_{\mathrm{fin}}(Y) = \mathrm{FSym}(Y)$. Let $L = \mathrm{GL}_{\mathrm{fin}}(Y)\, H \leq \mathrm{GL}(V)$. Write τ_u for the translation $v \mapsto v + u$ of V, and let $A = V \rtimes L$ be the group of affine maps $v \mapsto g(v) + u$ with $g \in L$ and $u \in V$. Our group is

$$\Gamma = \Big(\bigoplus_{v \in V} S_3 \Big) \rtimes A,$$

where A permutes the summands through its action on V. The copy of S_3 at v is the *lamp* at v; writing λ_v for the copy at v of $\lambda \in S_3$, we have $g \lambda_v g^{-1} = \lambda_{g(v)}$ for $g \in A$. Thus Γ

[1] This V is unrelated to Thompson's group V of Sections 7.2 and 8.2.

is a lamplighter group. The lamps are copies of S_3, one at each finite configuration of lit points of the three rays. The lighter is the affine group A, which moves configurations by permuting points, toggling a point, or adding one point into another. Equivalently, Γ is the group of permutations of $V \times S_3$ generated by the maps $(v, s) \mapsto (g(v), s)$ for $g \in A$ and $(v, s) \mapsto (v, \lambda(v)s)$ for finitely supported $\lambda : V \to S_3$. This action is faithful. Groups obtained from a permutation action by adjoining the finitary linear or symmetric groups of the permuted set, and lamps over it, appear as enrichments [8, 1] and as halo products [32].

For distinct points $y, y' \in Y$ let $E_{y \leftarrow y'}$ be the transvection that sends $e_{y'}$ to $e_{y'} + e_y$ and fixes the other basis vectors. For $g \in H$ and a linear map M of V,

$$g\tau_{e_y}g^{-1} = \tau_{e_{g(y)}}, \qquad gE_{y \leftarrow y'}g^{-1} = E_{g(y) \leftarrow g(y')}, \qquad M\tau_u M^{-1} = \tau_{M(u)}. \tag{2}$$

Linear maps fix 0, so they commute with the lamp at 0. On the first ray we abbreviate: for $x \geq 1$ we write x for the point $(1, x)$, e_x for $e_{(1,x)}$ and τ_x for τ_{e_x}.

Let a and b be the transpositions $(1\,2)$ and $(2\,3)$ in the lamp at 0, so that ab has order three, and put

$$t = \tau_1, \qquad d = E_{2 \leftarrow 1}, \qquad c = d\,ab;$$

thus t adds the point 1 to a configuration, and d adds the point 2 to every configuration that contains the point 1.

Lemma 3.1. *(a) The element c has order six, $c^3 = d$ and $c^4 = ab$; hence $b = ac^4$.*

(b) Γ is generated by the five elements p, q, t, c, a.

(c) Let $\pi : \Gamma \to \mathbb{Z}^2$ be the composite of $\Gamma \to A \to L \to L/\operatorname{GL}_{\mathrm{fin}}(Y) \cong H/\operatorname{FSym}(Y) \cong \mathbb{Z}^2$. Its kernel $(\bigoplus_V S_3) \rtimes (V \rtimes \operatorname{GL}_{\mathrm{fin}}(Y))$ is locally finite. Hence Γ is locally finite-by-$\mathbb{Z}^2$, and in particular elementary amenable.

Proof. (a) The linear map d fixes 0, so it commutes with ab; since d has order two and ab order three, c has order six, $c^3 = d^3(ab)^3 = d$ and $c^4 = d^4(ab)^4 = ab$.

(b) Let Γ' be the subgroup generated by the five elements. It contains H, because $H = \langle p, q\rangle$. Since $\operatorname{FSym}(Y)$ acts 2-transitively on Y, (2) shows that Γ' contains every τ_{e_y}, hence V, and, since $d = c^3$, every transvection $E_{y \leftarrow y'}$. Transvections generate every $\operatorname{GL}(\mathbb{F}_2^J)$, so $\operatorname{GL}_{\mathrm{fin}}(Y) \leq \Gamma'$, and $A \leq \Gamma'$. Finally $b = ac^4 \in \Gamma'$, and the lamp at $u \in V$ is $\tau_u\langle a, b\rangle\tau_u^{-1}$.

(c) Since $H \cap \operatorname{GL}_{\mathrm{fin}}(Y) = \operatorname{FSym}(Y)$, we have $L/\operatorname{GL}_{\mathrm{fin}}(Y) \cong H/\operatorname{FSym}(Y)$, and the kernel of π is as stated. It is an extension of $\bigoplus_V S_3$ by $V \rtimes \operatorname{GL}_{\mathrm{fin}}(Y)$, the directed union of the finite groups $\mathbb{F}_2^J \rtimes \operatorname{GL}(\mathbb{F}_2^J)$; both are 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 finitely generated, and is finite. Locally finite groups are directed unions of finite groups, and elementary amenable groups are closed under directed unions and extensions [22]. □

3.1 A certificate of 25 relations

All fillings in Γ are built from one finite set R of words in the five generators of Lemma 3.1, each representing 1 in Γ. The set R does not present Γ (Section 7.1). Areas are always taken with respect to R.

Abbreviations. Put

$$s = [p, q], \qquad B = psp^{-1}, \qquad t_m = p^{m-1}\,t\,p^{1-m} \quad (m = 1, 2, 3), \qquad d = c^3, \qquad b = ac^4.$$

By Lemma A.1, s and B are the transpositions (1 2) and (2 3), and t_m represents τ_m, since $p^{m-1}(1,1) = (1,m)$. By Lemma A.2, with $\zeta_1 = s$ and $\zeta_2 = Bs$, the pairs

$$K_1 = (sp,\ sq), \qquad K_2 = (Bsp,\ Bsq)$$

generate the pointwise stabilizers $H^{(1)}$ of the point 1 and $H^{(2)}$ of the points $1, 2$. The elements t and d have the *prototypes* $P_t = \{1\}$ and $P_d = \{1,2\}$, the points on which they depend: by (2), an element of H that fixes P_f pointwise commutes with f.

	relations	says	number	length
H	s^2, $(sps\,p^{-1})^3$, $[s, p^2sp^{-2}]$, $psp^{-1}qs^{-1}q^{-1}$	the presentation (13) of Houghton's group	4	64
Q	t^2	a translation is an involution	1	2
S	$[t,k]$ for $k \in K_1$; $[d,k]$ for $k \in K_2$	the stabilizer of a prototype commutes with it	4	56
Z	$[g,x]$ for $g \in \{p,q\}$, $x \in \{a,b\}$; $[d,a]$	linear maps fix the lamp at 0	5	40
C	$[t_1,t_2]$	translations commute	1	8
A	$d\,t_1d^{-1}(t_1t_2)^{-1}$, $[d,t_2]$, $[d,t_3]$	how d moves the translations at $1,2,3$	3	39
L	a^2, b^2, $(ab)^3$	the lamp at 0 is S_3	3	30
P	$[x, tyt^{-1}]$ for $x,y \in \{a,b\}$	the lamps at 0 and at e_1 commute	4	64
	total		25	303

Table 1: The certificate R: eight families of relations in the five generators p, q, t, c, a, written with the abbreviations s, B, t_m, d and b. Each relation is taken to be the freely reduced word obtained by expanding the abbreviations; lengths are those of these words, and the longest has length 24.

Lemma 3.2. *Every word of Table 1 represents* 1 *in* Γ.

Proof. H: the words of (13) hold in H. Q: V has exponent two. S: the elements of $H^{(1)}$ fix the point 1, so they commute with $t = \tau_1$; those of $H^{(2)}$ fix the points 1 and 2, so they commute with $d = E_{2\leftarrow 1}$ (Lemma 3.1(a) and (2)). Z: p, q and d are linear, so they commute with a and with $b = ac^4 = a \cdot ab$, which lie in the lamp at 0. C: translations commute. A: $dt_md^{-1} = \tau_{d(e_m)}$, and $d(e_1) = e_1 + e_2$, while d fixes e_2 and e_3. L: these are relations of S_3. P: tyt^{-1} lies in the lamp at e_1, which commutes with the lamp at 0. □

A program in the accompanying files checks the 25 words independently of Lemma 3.2, in the faithful action of Γ on $V \times S_3$; Appendix I says what it checks.

3.2 Theorem A

Theorem 3.3 (Theorem A)**.** *Let $E \subset \Gamma$ be the chunk consisting of* 1*, the five generators p, q, t, c, a and their inverses, the elements represented by the prefixes of the* 25 *freely reduced words of Table 1, and the inverses of these elements. Let $p_H = \log 107 + 4$ and*

$$\kappa = 2 + p_H = 6 + \log 107 < 12.75.$$

There are absolute constants $c, C > 0$ such that for every $r \geq 2$:

(a) $\log \sigma_E(r) \geq c\,r/(\log r)^{\kappa}$*;*

(b) $\log \sigma_E(r) \leq C\,r \log r$*; more generally, every chunk E' of Γ satisfies* $\log \sigma_{E'}(r) \leq C_{E'}\,r \log r$.

In particular $\sigma_E(r) = \exp(r^{1+o(1)})$. *Moreover:*

(c) For every $0 < \Delta \le 1$ *there are* $c_\Delta, e_\Delta > 0$ *such that the following holds for* $0 < e \le e_\Delta$. *Let* ϕ *assign unitaries in* $U(D)$ *to the five generators so that every word of Table 1 is mapped within* e *of* I *in the normalized Hilbert–Schmidt norm, and put* $\phi(b) = \phi(a)\phi(c)^4$. *If* $\|\phi(a)\phi(b) - I\|_2 \ge \Delta$, *then*

$$\log D \;\ge\; c_\Delta \, \frac{e^{-2}}{(\log(1/e))^{2\kappa}}, \qquad 2\kappa < 25.49.$$

(d) Γ *is finitely presented and elementary amenable.*

The exponent $\kappa = 2 + p_H$ has two parts. Two of the 2^N lamps are brought together by about N^2 local moves (Theorem 5.1), each priced by the Dehn function of H_3 at scale N, which is at most CN^{p_H} (Proposition A.3(b)). The 107 is the factor by which the doubling endomorphism of H_3 in [26, Section 3] multiplies a reweighted area at each step, so that far commutators at distance M have area $O(M^{\log 107})$ [26, Theorem 3.2]. The 4 comes from reducing the Dehn function to far commutators [26, Theorem 3.3]. The 107 is the only constant in κ that is not structural, and a quadratic Dehn function for H_3 would give $\kappa = 4$.

The proof occupies Sections 4–7.

(i) Section 4: every local identity among transported translations, transvections and lamps is filled at a cost $L(n)$ governed by δ_H at scale n.

(ii) Section 5: two of the 2^N lamps on the first N points commute at the cost of about N^2 local fillings (Theorem 5.1).

(iii) Section 6: feeding these pair areas into Theorem 2.1 gives the lower bounds (a) and (c).

(iv) Section 6: models that hash configurations by random affine maps give the upper bound (b).

(v) Section 7: finite presentation, part (d), by Cornulier's criterion for permutational wreath products.

4 Local fillings

To compare lamps we carry the two prototypes, the translation t and the transvection d, around the first n points of ray 1 by short words in p and q. This section shows that any two short words that agree on the prototype give the same conjugate, at the cost of two fillings in H_3 at scale n. The output is one number, the $L(n)$ of Lemma 4.2: the price of every local identity used in Section 5.

Three facts about Houghton's group. All transport takes place among the first points of ray 1. For $n \ge 1$ let $\Sigma_n \le H$ be the group of permutations of Y supported in $\{(1,1), \dots, (1,n)\}$, identified with $\mathrm{Sym}(n)$, and write $(x\ y)$ for the transposition of the points x and y of the first ray. Let δ_H be the Dehn function of the presentation P_H of H in (13), whose four relators are the family H. That is, $\delta_H(n)$ is the largest area, with respect to these relators, of a word of length at most n in $p^{\pm 1}, q^{\pm 1}$ that represents 1 in H. Appendix A proves the three facts used here.

(H1) *Short words for finite permutations* (Lemma A.1). The word $\sigma_k = p^k s p^{-k}$ represents the transposition $(k{+}1\ \ k{+}2)$, and for $1 \le x < y$ the transposition $(x\ y)$ is represented by a word $\omega_{x,y}$ of length less than $10y$ in p, q. For distinct $x_1, \dots, x_l \in \{1, \dots, n\}$ there is a word $\pi[x_1, \dots, x_l]$ of length less than $10ln$ that sends m to x_m for $m \le l$; it is a product of at most l transpositions of points in $\{1, \dots, n\}$.

(H2) *Point stabilizers* (Lemma A.2). Let $H^{(k)}$ be the pointwise stabilizer in H of the points $1,\dots,k$ of ray 1, and put $\zeta_k = \sigma_{k-1}\cdots\sigma_1\sigma_0$. Every element of $\Sigma_n \cap H^{(k)}$ that moves $m \geq 2$ points is represented by a word of length less than $10(m-1)n$ in the two letters of $K_k = (\zeta_k p, \zeta_k q)$. For $k = 1, 2$ these are the pairs of Section 3.1, whose letters have length at most 5 and 11 in p, q.

(H3) *The Dehn function* (Proposition A.3). There are absolute constants K and C with $\delta_H(n) \leq e^{Kn}$, by Lee [55, Theorem D], and $\delta_H(n) \leq Cn^{p_H}$ with p_H as in Theorem 3.3, by [26, Theorems 3.2 and 3.3].

Let $n \geq 3$. Three elementary properties of area are also used. The area is invariant under free equality, inversion and conjugation, and subadditive under products. If N represents 1, then $\mu N \nu = (\mu N \mu^{-1})\mu\nu$ for all words μ, ν, so replacing a subword U by U' changes the area by at most $\mathrm{Area}_R(UU'^{-1})$. And $[w, z_1 z_2] = [w, z_1] \cdot z_1 [w, z_2] z_1^{-1}$ and $[w, z^{-1}] = z^{-1}[w, z]^{-1} z$, so $\mathrm{Area}_R([w, z])$ is at most the sum of the areas $\mathrm{Area}_R([w, z_l])$ over the letters z_l of any factorization $z = z_1^{\pm 1} \cdots z_m^{\pm 1}$.

Canonical transporters. For points $i \neq j$ in $\{1,\dots,n\}$ put

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

with the words of Lemma A.1(c). We use τ_i both for this word and for the translation by e_i that it represents, since $U_i(1) = i$; and X_{ij} represents $E_{i \leftarrow j}$, since U_{ij} maps 1 to j and 2 to i. By Lemma A.1, $|U_i| < 10n$ and $|U_{ij}| < 20n$, and U_i is a product of at most one transposition and U_{ij} of at most two.

Lemma 4.1 says that it does not matter which short word transports a prototype: two such words differ by an element of a stabilizer, which commutes with the transported element by the family S. We allow the transporting word to end with a power of p, because the relations of Table 1 use $t_2 = ptp^{-1}$ and $t_3 = p^2 t p^{-2}$.

Lemma 4.1 (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, and let $W = W_0 p^j$, where $j = 0$ if $f = d$ and $j \in \{0, 1, 2\}$ if $f = t$. If W and U agree on P_f, then*

$$\mathrm{Area}_R(WfW^{-1} \cdot Uf^{-1}U^{-1}) \;\leq\; \mathcal{E}(n) := 2\,\delta_H(1500n) + 130n.$$

Proof. 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 A.2(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, since g and sp do. By Lemma A.2(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.

Let $N = W^{-1}UZ^{-1}$. It represents 1 in H and has length less than $30n + 2 + 30n + 1342n \leq 1500n$, so by the family H its area is at most $\delta_H(1500n)$. By the family S, $\mathrm{Area}_R([f, Z]) < 122n$. Since $U = WNZ$ freely,

$$WfW^{-1} \cdot Uf^{-1}U^{-1} = W\,(fNf^{-1})\,[f, Z]\,N^{-1}\,W^{-1},$$

whose area is at most $2\delta_H(1500n) + 122n \leq \mathcal{E}(n)$. □

Lemma 4.2 (Local fillings)**.** *Let $n \geq 3$ and*

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

Let $i, j, k \in \{1, \dots, n\}$ be points with $i \neq j$, and $x, y \in \{a, b\}$, where b stands for the word ac^4. Each of the following words represents 1 and has area at most $L(n)$ with respect to R.

(a) τ_i^2*, of area* 1*, and* $[\tau_i, \tau_j]$*.*

(b) $X_{ij}\tau_k X_{ij}^{-1}\tau_k^{-1}$ *if* $k \neq j$*, and* $X_{ij}\tau_j X_{ij}^{-1}\tau_i^{-1}\tau_j^{-1}$*.*

(c) $[X_{ij}, x]$*, of area less than* $40n + 1$*.*

(d) $[x, \tau_i y \tau_i^{-1}]$*, of area less than* $40n + 1$*.*

Proof. All the words represent 1 by (2). Call replacing a subword UfU^{-1} by WfW^{-1} as in Lemma 4.1 a *change of transporter*; it costs at most $\mathcal{E}(n)$, and $5\mathcal{E}(n) + 1 \leq L(n)$. In each part, W_0 is a product of at most three transpositions, of length less than $30n$, 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) $\tau_i^2 = U_i t^2 U_i^{-1}$ freely, a conjugate of the relation Q. For $[\tau_i, \tau_j]$ let $W_0 = \pi[i, j]$. Replace both occurrences of $\tau_i^{\pm 1}$ by $(W_0 t W_0^{-1})^{\pm 1}$, and both occurrences 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 both map 1 to i, and $W_0 p$ and U_j both map 1 to j. The result $W_0[t_1, t_2]W_0^{-1}$ is a conjugate of the relation C, so the area is at most $4\mathcal{E}(n) + 1$.

(b) Let $W_0 = \pi[j, i, k]$, or $\pi[j, i]$ if $k \in \{i, j\}$, so that W_0 maps 1 to j, 2 to i, and 3 to k when $k \notin \{i, j\}$; let $m \in \{1, 2, 3\}$ be the point with $W_0(m) = k$, so that $m = 1$ exactly when $k = j$. 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 $(W_0 t_m W_0^{-1})^{\pm 1} = (W_0 p^{m-1}\, t\, p^{1-m} W_0^{-1})^{\pm 1}$, and $\tau_i^{\pm 1}$ by $(W_0 t_2 W_0^{-1})^{\pm 1}$, again changes of transporter. If $k \neq j$, four of them turn the word into $W_0(dt_m d^{-1} t_m^{-1})W_0^{-1}$ with $m \in \{2, 3\}$, a conjugate of a relation A. If $k = j$, five of them turn it into $W_0(dt_1 d^{-1} t_2^{-1} t_1^{-1})W_0^{-1}$, again a conjugate of a relation A. So the area is at most $5\mathcal{E}(n) + 1$.

(c) Moving x across a letter of U_{ij} costs one relation Z, and moving it across $d = c^3$ costs one relation $[d, a]$, since $[c^3, ac^4] = [c^3, a]$ freely. Hence $\mathrm{Area}_R([X_{ij}, x]) \leq 2|U_{ij}| + 1 < 40n + 1$.

(d) Freely the word is $U_i[U_i^{-1} x U_i,\ t(U_i^{-1} y U_i)t^{-1}]U_i^{-1}$. Replacing the two occurrences of $U_i^{-1} x^{\pm 1} U_i$ and the two of $U_i^{-1} y^{\pm 1} U_i$ by $x^{\pm 1}$ and $y^{\pm 1}$ costs at most $4|U_i|$ relations Z and leaves $U_i[x, tyt^{-1}]U_i^{-1}$, a conjugate of a relation P; so the area is at most $4|U_i| + 1 < 40n + 1$. □

5 Pair areas

We now bring together two of the 2^N lamps supported on the first N points of ray 1. Their transport words are products of the τ_i, and the two configurations differ by a configuration that at most N transvections clear down to a single point, where the commutator of the two lamps is a local filling. Clearing naively would let the translation word grow; we keep it normalized after every step, so that it never has more than N factors, and each step costs $O(N)$ local fillings.

A worked example. Let $N = 3$ and take the lamps at $u = \{1, 3\}$ and $w = \{2\}$, with transport words $T_u = \tau_1\tau_3$ and $T_w = \tau_2$. Figure 2 draws the clearing. Translations commute, so the commutator of the two lamps is $T_u[x, ZyZ^{-1}]T_u^{-1}$ with $Z = T_u^{-1}T_w = \tau_3^{-1}\tau_1^{-1}\tau_2$, and two relations Q and two swaps turn Z into $T_{u_0} = \tau_1\tau_2\tau_3$, where $u_0 = u + w = \{1, 2, 3\}$. Choose the point 3 of u_0 as the *pivot*, into which the other points are cleared. The transvection X_{13} adds the value at 3 into 1. Conjugating T_{u_0} by it leaves τ_1 and τ_2 alone and turns τ_3 into $\tau_1\tau_3$;

one swap brings the new τ_1 next to the old one, and one relation Q cancels them, leaving $\tau_2\tau_3$. Then X_{23} turns $\tau_2\tau_3$ into τ_3 in the same way. Conjugating by $G = X_{23}X_{13}$ therefore carries the lamp at u_0 to the lamp at e_3, while G is linear and commutes with the lamp at 0. What is left is $[x, \tau_3 y \tau_3^{-1}]$, a local filling. Every identity used involves at most three points, so Lemma 4.2 prices each of them at $L(3)$.

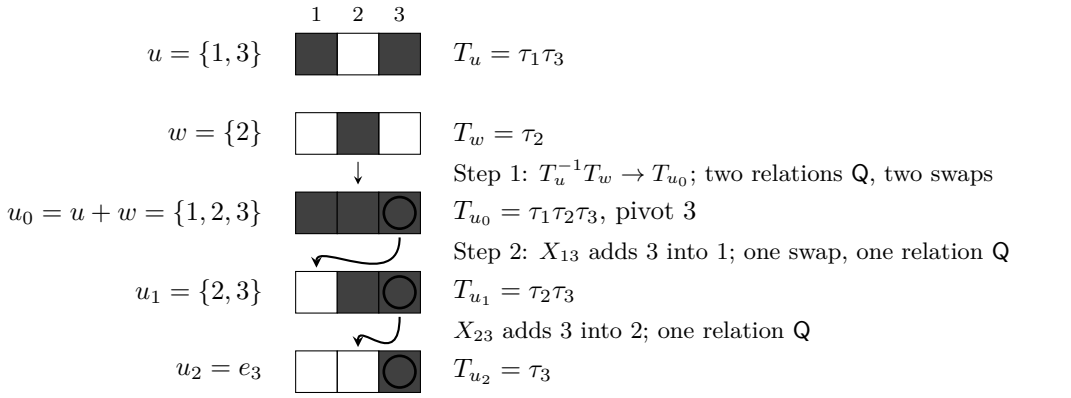


Figure 2: The clearing of the worked example, $N = 3$: the lit cells are the support of a configuration on the points $1, 2, 3$ of ray 1, the ringed cell is the pivot, and each arrow is a transvection that adds the value at the pivot into the cell it points to. After X_{13} and X_{23} the configuration $u_0 = \{1, 2, 3\}$ has become e_3, and the translation word has been renormalized after each step.

Theorem 5.1 (Pair areas)**.** *Let $N \geq 1$ and $n = \max(3, N)$. For a configuration u supported in the first N points of ray 1 let $T_u = \tau_{i_1} \cdots \tau_{i_m}$, where $i_1 < \cdots < i_m$ are the points of the support of u. Then T_u represents τ_u, so $T_u a T_u^{-1}$ and $T_u b T_u^{-1}$ generate the copy of S_3 at u; $|T_u| \leq 21Nn$; and for distinct such u, w and $x, y \in \{a, b\}$, with b the word ac^4,*

$$\mathrm{Area}_R([T_u x T_u^{-1},\, T_w y T_w^{-1}]) \;\leq\; \Lambda_N := 20N^2 L(n) \;\leq\; C_0 N^2 (\delta_H(C_0 N) + N)$$

with $C_0 = 4.5 \cdot 10^4$. Consequently there are absolute constants K' and C' with

$$\Lambda_N \leq e^{K'N} \qquad \text{and} \qquad \Lambda_N \leq C' N^{2+p_H} \qquad (N \geq 1), \qquad p_H = \log 107 + 4.$$

Proof. Translations commute, so T_u represents τ_u, and $|T_u| \leq N(2 \cdot 10n + 1) \leq 21Nn$. Write $L = L(n)$; by Lemma 4.2, each of its four local identities has area at most L.

Step 1: one translation word. Freely, $[T_u x T_u^{-1}, T_w y T_w^{-1}] = T_u [x, ZyZ^{-1}] T_u^{-1}$ 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 one relation Q each. Sort the resulting product of at most $2N$ factors τ_i into increasing order; this takes at most $\binom{2N}{2}$ swaps of adjacent distinct factors, each a local filling of Lemma 4.2(a). Cancel the at most N adjacent equal pairs, again at one relation Q each. The result is T_{u_0} with $u_0 = u + w \neq 0$, and $\mathrm{Area}_R(Z T_{u_0}^{-1}) \leq 2N + N(2N-1)L \leq 2N^2 L$. Since $Z^{\pm 1}$ occurs four times in $[x, ZyZ^{-1}]$,

$$\mathrm{Area}_R([T_u x T_u^{-1}, T_w y T_w^{-1}]) \leq 8N^2 L + \mathrm{Area}_R([x, T_{u_0} y T_{u_0}^{-1}]).$$

Step 2: normalized clearing. Let $s \leq N$ be the size of the support of u_0, choose a point j in it as the pivot, and list the other points of the support as $i_1, \ldots, i_{s-1}$. Put $G_l = X_{i_l j} \cdots X_{i_1 j}$ and $u_l = u_0 + e_{i_1} + \cdots + e_{i_l}$, whose support is that of u_0 with $i_1, \ldots, i_l$ removed; it still contains j, and $u_{s-1} = e_j$. We claim

$$\mathrm{Area}_R(G_l T_{u_0} G_l^{-1} T_{u_l}^{-1}) \leq 2lsL \qquad (0 \leq l \leq s-1).$$

For $l = 0$ there is nothing to prove. Given the claim for l, write $X = X_{i_{l+1}j}$ and $i = i_{l+1}$. Freely $G_{l+1}T_{u_0}G_{l+1}^{-1} = X(G_lT_{u_0}G_l^{-1})X^{-1}$, which differs from $XT_{u_l}X^{-1}$ by a conjugate of a word of area at most $2lsL$. Next, $XT_{u_l}X^{-1}$ is freely the product of the words $X\tau_{i'}X^{-1}$ over the support of u_l, in order. By Lemma 4.2(b), at a cost of at most sL, each of these 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}, because i lies in the support of u_l; at most $s-1$ swaps (Lemma 4.2(a)) make the two adjacent, and one relation Q cancels them. What remains is $T_{u_{l+1}}$. This step costs at most $sL + (s-1)L + 1 \le 2sL$, which proves the claim. The translation word is renormalized after every transvection and never has more than $s \le N$ factors.

Step 3: a single point. Let $G = G_{s-1}$ and let $M = GT_{u_0}G^{-1}\tau_j^{-1}$, which has area at most $2s(s-1)L \le 2N^2L$ by Step 2. Freely $T_{u_0} = G^{-1}M\tau_jG$. Substituting this into the four occurrences of $T_{u_0}^{\pm1}$ in $[x, T_{u_0}yT_{u_0}^{-1}]$ and deleting the four occurrences of $M^{\pm1}$ costs at most $8N^2L$. By Lemma 4.2(c), $\mathrm{Area}_R([G,z]) \le (s-1)(40n+1)$ for $z \in \{a,b\}$, and the same bound holds for $[G, z^{-1}]$ and $[z, G^{-1}]$. Moving $y^{\pm1}$ across G and G^{-1}, and then $x^{\pm1}$, leaves $G^{-1}[x, \tau_jy\tau_j^{-1}]G$, at a cost of at most $4(s-1)(40n+1)$; and $[x, \tau_jy\tau_j^{-1}]$ has area at most $40n+1$ by Lemma 4.2(d). Altogether, since $L \ge 700n$,

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

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

Growth. Since $n \le 3N$, $L(n) \le 10\delta_H(4500N) + 2100N$, which gives the first bound on Λ_N. With Proposition A.3(a) it is at most $e^{K'N}$ for a suitable K'. With Proposition A.3(b) it is at most $C_0N^2(C(C_0N)^{p_H} + N) \le C'N^{2+p_H}$, since $p_H > 1$. □

The factor N^2 counts local moves: at most N transvections, each followed by at most N commutations. The Dehn function of H enters through the local fillings, at the scale N of the transport. Theorem 5.1 is the only transport input of the paper: the lower bounds of Theorem A (Section 6) and of Theorem B (Section 11) feed it into the criterion and into the one-round theorem.

6 Hash models, and the proof of Theorem A

The lower bound of Theorem A follows by feeding 2^N lamps of pair area polynomial in N (Theorem 5.1) into the criterion. The upper bound needs models far smaller than Følner sets, which we build by hashing configurations through random affine maps, a universal hash family that the affine group moves covariantly. The two halves meet in the proof of Theorem A at the end of the section.

6.1 Affine hashing and the upper bound

The Følner sets of Γ are doubly exponential (Proposition 7.3), but its chunks have far smaller models. A Følner set gives models of every chunk, by Cornulier's comparison of the sofic profile with the Følner function [25, Proposition 3.14], but it records the lamp at every configuration of a large finite part of V. A model need only record the lamps that are lit, and it can store them through a random affine hash of the configurations, which the affine group moves covariantly. Random $\mathbb{F}_2$-linear maps form a universal hash family of Carter and Wegman [20], and a random translation does not change collision probabilities.

Let W be a finite-dimensional vector space over $\mathbb{F}_2$, and let $K_W = S_3^W \rtimes \mathrm{AGL}(W)$, where S_3^W is the group of functions $W \to S_3$ under pointwise multiplication, and $g \in \mathrm{AGL}(W)$ acts by $(g\lambda)(w) = \lambda(g^{-1}w)$. Thus $(\lambda, g)(\mu, h) = (\lambda \cdot g\mu,\, gh)$. The support of λ is the set of w

with $\lambda(w) \neq 1$. Fix $k \geq 1$ and a total order on W, let $\mathrm{Aff}(W, \mathbb{F}_2^k)$ be the set of affine maps $W \to \mathbb{F}_2^k$, and let

$$\Omega = \mathrm{Aff}(W, \mathbb{F}_2^k) \times S_3^{\mathbb{F}_2^k}, \qquad |\Omega| = 2^{k(\dim W+1)}\, 6^{2^k}.$$

For $\eta \in \mathrm{Aff}(W, \mathbb{F}_2^k)$ and $\lambda \in S_3^W$ let $\lambda^\eta : \mathbb{F}_2^k \to S_3$ send z to the product, in the order of W, of the $\lambda(w)$ with $\eta(w) = z$. For $\gamma = (\lambda, g) \in K_W$ define a permutation Θ_γ of Ω by

$$\Theta_\gamma(\eta, u) = (\eta \circ g^{-1},\ \lambda^{\eta \circ g^{-1}} \cdot u);$$

its inverse sends (η', u') to $(\eta' \circ g, (\lambda^{\eta'})^{-1} u')$, and Θ_1 is the identity.

Lemma 6.1 (Affine hashing)**.** *Let $b \geq 1$, let $\gamma_1 = (\lambda, g)$ and $\gamma_2 = (\mu, h)$ be elements of K_W whose lamp parts λ, μ have supports of size at most b, and let ω be uniformly distributed on Ω. Then*

(i) $\Theta_{\gamma_1}\Theta_{\gamma_2}\omega = \Theta_{\gamma_1\gamma_2}\omega$ except with probability at most $\binom{2b}{2} 2^{-k}$;

(ii) if $\gamma_1 \neq \gamma_2$, then $\Theta_{\gamma_1}\omega \neq \Theta_{\gamma_2}\omega$ except with probability at most $\binom{2b}{2} 2^{-k}$.

Proof. Write $\omega = (\eta, u)$ and $\eta = \eta_0 + z_0$ with η_0 linear; then η_0 and z_0 are independent and uniform. For $w \neq w'$ in W, $\eta(w) - \eta(w') = \eta_0(w - w')$ is uniform on $\mathbb{F}_2^k$, because $\eta_0 \mapsto \eta_0(w - w')$ is a surjective linear map; so $\eta(w) = \eta(w')$ with probability exactly 2^{-k}. Precomposition with a fixed affine bijection of W permutes $\mathrm{Aff}(W, \mathbb{F}_2^k)$, so $\eta \circ f$ is again uniform for every $f \in \mathrm{AGL}(W)$. Hence, for a set $S \subset W$ of size at most $2b$, $\eta \circ f$ is injective on S except with probability at most $\binom{2b}{2} 2^{-k}$. If η' is injective on a set S containing the support of λ, then $\lambda^{\eta'}(\eta'(w)) = \lambda(w)$ for $w \in S$, and $\lambda^{\eta'} = 1$ off $\eta'(S)$.

(i) Put $\eta' = \eta \circ (gh)^{-1}$. Then $\Theta_{\gamma_1}\Theta_{\gamma_2}(\eta, u) = (\eta', \lambda^{\eta'} \mu^{\eta \circ h^{-1}} u)$ and $\Theta_{\gamma_1\gamma_2}(\eta, u) = (\eta', (\lambda \cdot g\mu)^{\eta'} u)$. Let S be the union of the support of λ and the image under g of the support of μ, and suppose that η' is injective on S. Since $\eta \circ h^{-1}(v) = \eta'(gv)$, the map $\eta \circ h^{-1}$ is injective on the support of μ, and $\mu^{\eta \circ h^{-1}}(\eta'(w)) = \mu(g^{-1}w)$ for $w \in S$. The support of $\lambda \cdot g\mu$ lies in S. So all three functions are trivial off $\eta'(S)$, and at $\eta'(w)$, $w \in S$, both sides equal $\lambda(w)\mu(g^{-1}w)$.

(ii) If $g \neq h$, choose w_0 with $h^{-1}g(w_0) \neq w_0$. If $\eta \circ g^{-1} = \eta \circ h^{-1}$, then $\eta = \eta \circ h^{-1}g$, so $\eta(h^{-1}g(w_0)) = \eta(w_0)$, which has probability 2^{-k}. If $g = h$, then $\lambda \neq \mu$; put $\eta' = \eta \circ g^{-1}$ and let S be the union of the supports. If η' is injective on S, then $\lambda^{\eta'}$ and $\mu^{\eta'}$ take the different values $\lambda(w)$ and $\mu(w)$ at $\eta'(w)$ for some $w \in S$. □

To use the lemma in Γ we need finite subgroups that contain the relevant cocycles. The quotient $\mathbb{Z}^2$ of Lemma 3.1(c) serves as a clock. A model is a box in $\mathbb{Z}^2$ times the hash space Ω; a generator advances the clock and applies a cocycle, which lies in a finite subgroup of the locally finite kernel of π. For $z \in \mathbb{Z}^2$ put $g_z = p^{z_1} q^{z_2}$, so that $\pi(g_z) = z$, and for $m \geq 1$ let $B_m = \{-m, \ldots, m\}^2$. For $R \geq 1$ let $Y_R \subset Y$ be the set of points of height at most R, and let $W_R = \mathbb{F}_2^{Y_R} \subset V$. Let $K_R \leq \ker \pi$ consist of the elements whose lamps lie at configurations in W_R and whose affine part lies in $W_R \rtimes \mathrm{GL}(W_R)$, with $\mathrm{GL}(W_R)$ extended by the identity on the other basis vectors. Restricting to W_R identifies K_R with K_{W_R}.

Lemma 6.2 (The clock)**.** *Let $m \geq 1$, $R = 2m + 3$, $z \in B_m$ and $s \in \{p, q, t, c, a\}^{\pm 1}$. Then $g_{z+\pi(s)}^{-1} s g_z$ and $g_z s g_{z+\pi(s)}^{-1}$ lie in K_R, and their lamps sit at no more than one configuration.*

Proof. Each of $p^{\pm 1}, q^{\pm 1}$ changes heights by at most one and acts on the points of height at least 2 as a translation along the rays, and these translations commute. Hence for $(y_1, y_2) \in \mathbb{Z}^2$ the commutator $[p^{y_1}, q^{y_2}]$ fixes every point of height greater than $|y_1| + |y_2| + 1$: while it is applied, such a point stays at height at least 2. For $s = p^{\pm 1}$ we have $\pi(s) = (\pm 1, 0)$, $g_{z+\pi(s)}^{-1} s g_z = 1$ and $g_z s g_{z+\pi(s)}^{-1} = p^{z_1}[q^{z_2}, p^{\pm 1}]p^{-z_1}$, which is supported in heights at most $|z_2| + 2 + |z_1| \leq 2m + 2$.

For $s = q^{\pm 1}$ we have $\pi(s) = (0, \pm 1)$, $g_z s g_{z+\pi(s)}^{-1} = 1$ and $g_{z+\pi(s)}^{-1} s g_z = q^{-z_2 \mp 1}[p^{-z_1}, q^{\pm 1}]q^{z_2 \pm 1}$, supported in heights at most $|z_1| + 2 + |z_2| + 1 \le 2m + 3$. These are permutations of Y_R, hence elements of $\mathrm{GL}(W_R)$. For $s \in \{t, c, a\}^{\pm 1}$ we have $\pi(s) = 0$, and $g_z^{\pm 1}$ moves the points 1 and 2 to height at most $2m + 2$; so the conjugates of t and d by $g_z^{\pm 1}$ are a translation and a transvection of W_R, while those of a and ab are a and ab, at the configuration 0. □

Proposition 6.3 (Hash models)**.** *Let E' be a chunk of Γ each of whose elements is a product of at most $\ell \ge 1$ of the generators p, q, t, c, a and their inverses. Then*

$$\log \sigma_{E'}(r) \;\le\; 500\,\ell^2\, r \log(2 + r) \qquad (r > 1).$$

Proof. Let $m = \lceil 8\ell r \rceil$, $R = 2m + 3$, $D = |Y_R| = 6m + 9$ and $k = \lceil \log(4\ell^2 r) \rceil$, and let Ω be the hash space for $W = W_R$. For $g \in \Gamma$ and $z \in \mathbb{Z}^2$ put $\psi(g, z) = g_{z+\pi(g)}^{-1} g g_z \in \ker \pi$; then $\psi(gh, z) = \psi(g, z + \pi(h))\, \psi(h, z)$. If g is a product of at most ℓ generators and inverses and $z \in B_{m-\ell}$, then $\psi(g, z)$ is the product of the $\psi(s, z')$ for the letters s of the word, at points $z' \in B_m$, so by Lemma 6.2 it lies in K_R with lamps at no more than ℓ configurations.

Let $X = B_m \times \Omega$. For $g \in E'$ let F_g be a permutation of X that extends the injective partial map $(z, \omega) \mapsto (z + \pi(g), \Theta_{\psi(g,z)}\omega)$, defined for $z \in B_{m-\ell}$, with $F_1 = \mathrm{id}$. Let $g, h, gh \in E'$ and $z \in B_{m-2\ell}$. Then z and $z + \pi(h)$ lie in $B_{m-\ell}$, so $F_g F_h(z, \omega) = (z + \pi(gh), \Theta_{\psi(g, z+\pi(h))}\Theta_{\psi(h,z)}\omega)$ and $F_{gh}(z, \omega) = (z + \pi(gh), \Theta_{\psi(g, z+\pi(h))\psi(h,z)}\omega)$; by Lemma 6.1(i) these agree for all but a fraction $\binom{2\ell}{2} 2^{-k}$ of the ω. Since $|B_{m-2\ell}| \ge (1 - 8\ell/(2m+1))|B_m|$,

$$d(F_g F_h, F_{gh}) \;\le\; \varepsilon := \frac{8\ell}{2m+1} + \binom{2\ell}{2} 2^{-k} < \frac{1}{2r} + \frac{1}{2r} = \frac{1}{r}.$$

If $g \ne h$ in E' and $z \in B_{m-\ell}$, then either $\pi(g) \ne \pi(h)$ and $F_g(z, \omega) \ne F_h(z, \omega)$, or $\psi(g, z) \ne \psi(h, z)$ and Lemma 6.1(ii) applies; so $d(F_g, F_h) \ge 1 - \varepsilon$. Hence $\sigma_{E'}(r) \le |X|$.

Finally $\log |X| = 2 \log(2m + 1) + k(D + 1) + 2^k \log 6$. Here $2^k < 8\ell^2 r$, $k < \log(8\ell^2 r) \le 5\ell \log(2 + r)$ and $D + 1 \le 64\ell r$, so $k(D + 1) \le 320\ell^2 r \log(2 + r)$; also $2^k \log 6 \le 21\ell^2 r \le 14\ell^2 r \log(2+r)$ and $2 \log(2m+1) \le 10\ell^2 r \log(2+r)$. The sum is less than $500\ell^2 r \log(2+r)$. □

6.2 Proof of Theorem 3.3

We now feed the pair areas of Theorem 5.1 into the criterion and the hash models of Proposition 6.3 into the upper bound.

Proof of Theorem 3.3. Setting up the criterion. Theorem 2.1 asks for a and b among the generators, while in R the letter b abbreviates ac^4. So let S^+ consist of the five generators and $b = ac^4 \in \Gamma$, and let R^+ consist of R and the three words $b^{-1}ac^4$, b^2 and $(ab)^3$ in the letters of S^+. These words represent 1, $a^2, b^2, (ab)^3 \in R^+$, $ab \ne 1$, and the longest word of R^+ has length $l = 24$. The chunk E^+ that Theorem 2.1 associates with S^+ and R^+ is contained in E. The reason, in one line, is that every new element is a power of c or is represented, using $a^2 = 1$, by a prefix of the expanded words b^2 and $(ab)^3$. In detail, besides the elements of E, the chunk E^+ contains $b^{\pm 1}$, the prefixes $b^{-1}, c^{-4}, c^{-3}, c^{-2}, c^{-1}$ of $b^{-1}ac^4$, and the prefixes b, $ab = c^4$, $aba = c^4 a$, $abab = c^2$ and $ababa = c^2 a$ of b^2, $(ab)^3$ and ab. These lie in E, because ac^4, a^2c^4, a^2c^4a and $a^2c^4a^2c^4a$ are prefixes of b^2 and $(ab)^3$ in Table 1, c^2 and c^3 are prefixes of the relations $[d, k]$ of the family S, $a^2 = 1$, and $c^6 = 1$. Restricting a model of E to E^+ gives a model of E^+, so $\sigma_E \ge \sigma_{E^+}$.

For $N \ge 1$ take the $M = 2^N$ words T_u of Theorem 5.1, one for each configuration u supported in the first N points of ray 1. A commutator $[T_u x T_u^{-1}, T_w y T_w^{-1}]$ with $x, y \in \{a, b\}$ contains the letter $b^{\pm 1}$ at most four times, and replacing each occurrence by $(ac^4)^{\pm 1}$ costs one

relation $b^{-1}ac^4$. So its area with respect to R^+ is at most $\Lambda_N^+ = \Lambda_N + 4$, and Theorem 2.1 gives

$$\log \sigma_E(r) \geq 2^N/16 \qquad \text{whenever} \qquad r \geq 4.7 \cdot 10^9 \cdot 2^N \Lambda_N^+. \tag{3}$$

Since $a \neq 1$, a $(1-1/r)$-expansive map needs at least two points, so $\log \sigma_E(r) \geq 1$; for r below any fixed bound r_0 claim (a) therefore holds once c is small enough, and we may assume r large.

(a) By Theorem 5.1, $\Lambda_N^+ \leq C_1 N^\kappa$ with C_1 absolute. Let $r' = r/(4.7 \cdot 10^9 C_1)$ and $N = \lfloor \log r' - \kappa \log\log r' \rfloor$, which is at least 1 for large r. Then $2^N \Lambda_N^+ \leq C_1 r' (\log r')^{-\kappa} (\log r')^\kappa = r/(4.7 \cdot 10^9)$, so (3) applies, and $\log \sigma_E(r) \geq 2^N/16 \geq r'/(32(\log r')^\kappa)$. Since $\log r' \leq \log r$, this is (a).

(b) Every element of E is a generator, the inverse of one, or represented by a prefix of a word of length at most 24 or by its inverse. Proposition 6.3 with $\ell = 24$ gives $\log \sigma_E(r) \leq 500 \cdot 24^2 r \log(2+r)$, and $\log(2+r) \leq 2 \log r$ for $r \geq 2$. The same proposition applies to every chunk.

(c) Extend ϕ to S^+ by $\phi(b) = \phi(a)\phi(c)^4$. Then $b^{-1}ac^4$ is mapped to I exactly, and b^2 and $(ab)^3$ are mapped to the images of the words b^2 and $(ab)^3$ of Table 1; so every word of R^+ is mapped within e of I. With $\Lambda_N^+ \leq C_1 N^\kappa$ as in (a), the condition $K_\Delta \sqrt{M}\, \Lambda_N^+ e \leq 1$ of Theorem 2.1 holds when $2^N N^{2\kappa} \leq Y := (C_1 K_\Delta e)^{-2}$, which is the case for $N = \lfloor \log Y - 2\kappa \log\log Y \rfloor$; this N is at least 1 for e small in terms of Δ. Then $2^N \geq Y/(2(\log Y)^{2\kappa})$, and since $C_1 K_\Delta \geq 1$ we have $\log Y \leq 2\log(1/e)$. Hence

$$\log D \;\geq\; \frac{\Delta^2}{12} 2^N \;\geq\; \frac{\Delta^2}{24 \cdot 4^\kappa C_1^2 K_\Delta^2} \cdot \frac{e^{-2}}{(\log(1/e))^{2\kappa}}.$$

(d) Proposition 7.1 gives finite presentation and Lemma 3.1 elementary amenability. □

Remark 6.4. Part (a) uses the polynomial Dehn bound of Proposition A.3(b). With Lee's exponential bound alone, Theorem 5.1 gives $\Lambda_N^+ \leq 2^{BN}$ for a constant B that depends on the constant in Lee's inequality, and $N = \lfloor \log(r/(4.7 \cdot 10^9))/(1+B) \rfloor$ in (3) gives only $\log \sigma_E(r) \geq c\, r^{1/(1+B)}$: a superpolynomial profile, with an exponent we cannot make explicit. Transport at scale N is too expensive for an exponential Dehn bound. The grid host transports at scale $\sqrt{N}$ and reaches $\log \sigma \geq c\, r\, e^{-C\sqrt{\log r}}$ with Lee's bound alone (conditionally, Remark H.1).

7 More about Γ

Theorem A asserts that Γ is finitely presented, and the introduction claims that Γ embeds in Brin's group $3V$. This section proves both, and then shows that the Følner function of Γ is doubly exponential, so that the hash models of Section 6 are exponentially smaller than any model built from a Følner set. Nothing here is used in the profile bounds of Theorem 3.3, parts (a) to (c), or in any proof of Part II, where the embedding in $3V$ serves only as a second check of the Choi rank; part (d) of that theorem, finite presentation, is Proposition 7.1. A reader heading for Theorem B can skip this section and Section 8.

7.1 A finite presentation

The proof that Γ is finitely presented has three steps. A presentation lemma for groups of the form $F(\mathcal{X}) \rtimes H$ reduces everything to finitely many H-orbits of relations. Steinberg's relations for the transvections, with Johnson's presentation of H and the relations of the translations, then present the affine group A, and omitting the translations presents L. Cornulier's criterion

for permutational wreath products presents Γ, because A acts doubly transitively on the configurations with finitely generated stabilizers.

The certificate of Section 3.1 is a finite set of relations, but it does not present Γ. The group it presents has abelianization $\mathbb{Z}^2 \times \mathbb{Z}/2 \times \mathbb{Z}/4$, in which c has order four; this is a computation with exponent sums, for instance a^2, $b^2 = (ac^4)^2$ and $(ab)^3$ give $2a = 8c = 12c = 0$. In Γ, by contrast, $c = d\,ab$ is a product of commutators, since $d = [E_{2\leftarrow 3}, E_{3\leftarrow 1}]$ and ab is a three-cycle.

Proposition 7.1 (Finite presentation)**.** *The groups L, A and Γ are finitely presented.*

Proof. Stabilizers. For a finite set $D \subset Y$ the pointwise stabilizer $H_{(D)}$ of D in H is isomorphic to H. Renumbering the points of each ray outside D in increasing order turns $Y \setminus D$ into three rays, and a permutation fixing D is eventually a translation on the old rays exactly when it is on the new ones. So $H_{(D)}$ is finitely generated. Since H contains $\mathrm{FSym}(Y)$, it acts transitively on ordered k-tuples of distinct points for every k.

A presentation lemma. Let a finitely presented group H act on a set $\mathcal{X}$ with finitely many orbits and finitely generated stabilizers, and let $F(\mathcal{X})$ be the free group on $\mathcal{X}$. Then $F(\mathcal{X}) \rtimes H$, with H permuting the free generators, is finitely presented. Take a presentation of H, adjoin one representative x_i of each orbit, and add the relations $[x_i, k] = 1$ for k in a finite generating set of the stabilizer of x_i. In the group so presented, hx_ih^{-1} depends only on $h \cdot x_i$, and the group is isomorphic to $F(\mathcal{X}) \rtimes H$: both are the iterated amalgam of H with the groups $\mathrm{Stab}(x_i) \times \langle x_i \rangle$ over the stabilizers $\mathrm{Stab}(x_i)$. Imposing in addition a set of relations that is a finite union of H-orbits keeps the group finitely presented, since one representative of each orbit suffices.

A presentation of A. For distinct $i, j \in Y$ let x_{ij} be the transvection $e_j \mapsto e_j + e_i$, let $t_i = \tau_{e_i}$, and let $\varsigma_{ij} \in \mathrm{FSym}(Y) \leq H$ be the transposition of i and j. Let $\widetilde{A}$ be the quotient of $F(\mathcal{X}) \rtimes H$, where $\mathcal{X} = \{x_{ij}\} \sqcup \{t_i\}$ and H relabels indices, by the relations

$$x_{ij}^2 = 1, \qquad [x_{ij}, x_{kl}] = 1 \ (i \neq l,\ j \neq k), \qquad [x_{ij}, x_{jk}] = x_{ik} \ (i, j, k \text{ distinct}),$$
$$\varsigma_{ij} = x_{ij}x_{ji}x_{ij}, \qquad t_i^2 = 1, \qquad [t_i, t_j] = 1,$$
$$x_{ij}t_jx_{ij}^{-1} = t_it_j, \qquad x_{ij}t_kx_{ij}^{-1} = t_k \ (k \neq j).$$

They hold in A, so there is a surjection $\widetilde{A} \to A$. Each relation involves at most four points, so by transitivity the relations form finitely many H-orbits; the stabilizers of x_{ij} and t_i are $H_{(\{i,j\})}$ and $H_{(\{i\})}$. Since H is finitely presented [14], for instance by Johnson's presentation (12) [48], $\widetilde{A}$ is finitely presented, and it remains to show that $\widetilde{A} \to A$ is injective. The linear relations are Steinberg's [58]; we give the short direct argument.

Fix a finite set $D \subset Y$, number its points $1, \dots, m$ in an order unrelated to heights, and let $U_D \leq \widetilde{A}$ be generated by the x_{ij} with $i < j$ in D. By the first three relations, $x_{1m}, \dots, x_{m-1,m}$ are commuting involutions, and they generate a subgroup normalized by the x_{ij} with $i < j < m$. By induction $|U_D| \leq 2^{m-1}2^{(m-1)(m-2)/2} = 2^{m(m-1)/2}$, the order of the upper unitriangular group $\mathrm{UT}_m(\mathbb{F}_2)$, onto which U_D maps. So $U_D \cong \mathrm{UT}_m(\mathbb{F}_2)$. Let $W_D = \mathrm{Sym}(D) \leq H$ and $\varsigma_i = \varsigma_{i,i+1}$. By the relation $\varsigma_{ij} = x_{ij}x_{ji}x_{ij}$ and relabeling, the subgroup G_D generated by all x_{ij} with $i, j \in D$ contains W_D and is generated by U_D and the ς_i. We claim $G_D = U_DW_DU_D$; since this set contains 1, it suffices that it is closed under right multiplication by U_D and by each ς_i. Every $b \in U_D$ is $x_{i,i+1}^{\epsilon}b_0$ with b_0 in the subgroup generated by the other positive roots, the kernel of the entry $(i, i+1)$, which ς_i normalizes. So it suffices to treat $wx_{i,i+1}\varsigma_i$ with $w \in W_D$; put $k = w(i)$ and $l = w(i+1)$. If $k < l$, then $wx_{i,i+1}\varsigma_i = x_{kl}w\varsigma_i \in U_DW_D$. If $k > l$, then $w\varsigma_i = \varsigma_{kl}w$ and $x_{kl}\varsigma_{kl} = x_{lk}x_{kl}$, so $wx_{i,i+1}\varsigma_i = x_{kl}\varsigma_{kl}w = x_{lk}wx_{i,i+1} \in U_DW_DU_D$. The map $G_D \to \mathrm{GL}(\mathbb{F}_2^D)$ is onto, since transvections generate $\mathrm{GL}(\mathbb{F}_2^D)$. If b_1wb_2 maps to the identity of $\mathrm{GL}(\mathbb{F}_2^D)$, the permutation

matrix of w is upper unitriangular, so $w = 1$ in H, and then $b_1 b_2 = 1$ in U_D. Hence $G_D \cong \mathrm{GL}(\mathbb{F}_2^D)$. Every element of the subgroup $\widetilde{J}$ generated by all x_{ij} lies in some G_D, so $\widetilde{J} \cong \mathrm{GL}_{\mathrm{fin}}(Y)$. By relabeling, every element of the subgroup generated by $\widetilde{J}$ and H has the form jh. If it maps to 1, then $h \in \mathrm{GL}_{\mathrm{fin}}(Y) \cap H = \mathrm{FSym}(Y)$, so h is a product of transpositions, each equal in $\widetilde{A}$ to a word in the x_{ij}, and $jh \in \widetilde{J}$ maps to 1, hence is 1. Finally the t_i generate a normal elementary abelian subgroup. It is a quotient of $\mathbb{F}_2^{(Y)}$ and maps onto V by $t_i \mapsto e_i$, hence isomorphically, and the linear part acts on it as on V. Hence every element of $\widetilde{A}$ is $t\ell$ with t in that subgroup. If its image is the identity affine map, evaluating at 0 gives $t = 1$, and then $\ell = 1$. So $\widetilde{A} \cong A$. Omitting the t_i gives L.

The lamps. The group A acts on V doubly transitively: a translation carries v to 0, and a finitary linear map carries any nonzero vector to e_1. The stabilizer of 0 is L, which is generated by p, q and d, because the H-conjugates of d are all the transvections $E_{y \leftarrow y'}$. So A is finitely presented, its point stabilizers are finitely generated, and it has two orbits on $V \times V$. By Cornulier's criterion for permutational wreath products [24, Theorem 1.1], $\Gamma = S_3 \wr_V A$ is finitely presented. Explicitly, let $\mathcal{R}_A$ present A on generators that include p, q, d, t. Add the generators a, b and the relations a^2, b^2, $(ab)^3$, $[a, \ell]$ and $[b, \ell]$ for $\ell \in \{p, q, d\}$, and $[u, tvt^{-1}]$ for $u, v \in \{a, b\}$. The result presents Γ. The lamp at $v \in V$ is well defined as a conjugate of $\langle a, b \rangle$, because L centralizes $\langle a, b \rangle$; and double transitivity carries the commutation of the lamps at 0 and at e_1 to every pair of lamps. □

7.2 An embedding in a Brin–Thompson group

This subsection places Γ inside a finitely presented simple group whose soficity is open, so that this group has a chunk with finite, nearly exponential profile. Let $\mathfrak{C} = \{0, 1\}^{\mathbb{N}}$ be Cantor space. For $n \geq 1$, Brin's group nV [11] consists of the homeomorphisms h of $\mathfrak{C}^n$ given by a finite table of prefix replacements. Such a table is a pair of partitions of $\mathfrak{C}^n$ into the same number of rectangles $u_1\mathfrak{C} \times \cdots \times u_n\mathfrak{C}$, with $u_1, \ldots, u_n$ finite binary words, matched in pairs. The map h sends each rectangle of the first partition onto its partner by replacing the n prefixes and keeping the n suffixes. Thus $1V$ is Thompson's group V, and nV embeds in $n'V$ for $n \leq n'$ by acting trivially on the additional coordinates. The groups nV are finitely presented, by Brin for $2V$ [12] and by Hennig and Matucci for every n [42], and simple [13]. For $n \geq 2$, it is not known whether nV is sofic. The letter V in these names has nothing to do with our space of configurations, which is also called V.

Proposition 7.2 (Embedding in $3V$)**.** *Γ embeds in $3V$, and hence in nV for every $n \geq 3$. Consequently $3V$ has a chunk E_{3V} with*

$$\exp\Big(\frac{c\,r}{(\log r)^{12.75}}\Big) \;\leq\; \sigma_{E_{3V}}(r) \;\leq\; \exp(C\,r\log r) \qquad (r \geq 2).$$

Proof. Coding configurations. For a finite set $S = \{j_1 < \cdots < j_m\}$ of positive integers, with $j_0 = 0$, let

$$e(S) = \Big(\prod_{k=1}^{m} 1\,0^{\,j_k - j_{k-1} - 1}\,1\Big)\,0,$$

so that $e(\emptyset) = 0$, $e(\{1\}) = 110$ and $e(\{2\}) = 1010$. Read a word from the left: at a record boundary the letter 0 ends the word and the letter 1 opens a record 10^k1, which ends at its next letter 1. So the words $e(S)$ form a prefix code. A finite word whose reading has not ended can be extended to one that begins with some $e(S)$, by appending 0 at a record boundary or 10 inside a record. Hence the cylinders $e(S)\mathfrak{C}$ are disjoint and their union is dense in $\mathfrak{C}$. For $v \in V$ and $i = 1, 2, 3$ let $S_i(v) = \{j : v(i, j) = 1\}$, let

$$Z_v = e(S_1(v))\mathfrak{C} \times e(S_2(v))\mathfrak{C} \times e(S_3(v))\mathfrak{C},$$

and let $\iota_v : \mathfrak{C}^3 \to Z_v$ prefix the three codewords. The rectangles Z_v are disjoint, and their union is dense in $\mathfrak{C}^3$.

Prepending a site. Define $s_0, s_1 : \mathfrak{C} \to \mathfrak{C}$ by $s_0(0\xi) = 0\xi$, $s_0(1\xi) = 10\xi$ and $s_1(\xi) = 11\xi$. Their images $0\mathfrak{C} \cup 10\mathfrak{C}$ and $11\mathfrak{C}$ partition $\mathfrak{C}$, and, with $S+1 = \{j+1 : j \in S\}$,

$$s_0\big(e(S)\xi\big) = e(S+1)\,\xi, \qquad s_1\big(e(S)\xi\big) = e\big(\{1\} \cup (S+1)\big)\,\xi : \tag{4}$$

an empty site in front lengthens the first gap by one, and an occupied site in front adds the record 11.

The affine group. For $b \in \{0,1\}$ put

$$\hat{P}(x, s_b(y), z) = (s_b(x), y, z), \qquad \hat{Q}(x, y, s_b(z)) = (s_b(x), y, z),$$
$$\hat{T}(s_b(x), y, z) = (s_{1-b}(x), y, z),$$

and let $\hat{D}$ fix $s_0(\mathfrak{C}) \times \mathfrak{C}^2$ pointwise and send $(11x, y, z)$ to $(11x', y, z)$, where $\hat{T}(x, y, z) = (x', y, z)$. These maps lie in $3V$. In prefix tables, with ϵ the empty word, $\hat{P}$ replaces the prefixes $(\epsilon, 11, \epsilon)$, $(0, 0, \epsilon)$, $(1, 0, \epsilon)$, $(0, 10, \epsilon)$ and $(1, 10, \epsilon)$ by $(11, \epsilon, \epsilon)$, $(0, 0, \epsilon)$, $(10, 0, \epsilon)$, $(0, 1, \epsilon)$ and $(10, 1, \epsilon)$. $\hat{Q}$ does the same with the second and third coordinates exchanged. $\hat{T}$ exchanges the first-coordinate prefixes $0 \leftrightarrow 110$ and $10 \leftrightarrow 111$. $\hat{D}$ exchanges $110 \leftrightarrow 11110$ and $1110 \leftrightarrow 11111$ and keeps 0 and 10. The permutation p moves the first point of ray 2 to the first point of ray 1 and shifts the other points of these rays, and q does the same with ray 3. The element t changes the value of a configuration at the point 1, and d adds the value at 1 to the value at 2. So (4) gives, for every $v \in V$,

$$\hat{P}\iota_v = \iota_{p(v)}, \qquad \hat{Q}\iota_v = \iota_{q(v)}, \qquad \hat{T}\iota_v = \iota_{t(v)}, \qquad \hat{D}\iota_v = \iota_{d(v)}.$$

By the proof of Lemma 3.1(b), p, q, t, d generate A. If a word in them represents $g \in A$, the same word in $\hat{P}, \hat{Q}, \hat{T}, \hat{D}$ maps $\iota_v(\xi)$ to $\iota_{g(v)}(\xi)$ for all v and ξ. If $g = 1$, it is the identity on the dense set $\bigcup_v Z_v$, hence the identity. So $p, q, t, d \mapsto \hat{P}, \hat{Q}, \hat{T}, \hat{D}$ defines a homomorphism $\rho : A \to 3V$ with

$$\rho(g)\,\iota_v = \iota_{g(v)} \qquad (g \in A,\ v \in V), \tag{5}$$

and ρ is injective: if $g \neq 1$, then $g(v) \neq v$ for some v, and $\rho(g)$ maps Z_v onto the disjoint rectangle $Z_{g(v)}$.

The lamps. Let $\beta_1 = 0$, $\beta_2 = 10$ and $\beta_3 = 11$, and let S_3 act faithfully on $\mathfrak{C}^3$ by the prefix replacements $\theta(\lambda)(\beta_i\xi, y, z) = (\beta_{\lambda(i)}\xi, y, z)$. For $\lambda \in S_3$ and $v \in V$ let $\rho(\lambda_v)$ be $\iota_v\theta(\lambda)\iota_v^{-1}$ on Z_v and the identity off Z_v; it lies in $3V$, since the complement of a rectangle is a finite union of rectangles. Copies at different configurations have disjoint supports and commute, each copy is faithful, and by (5) $\rho(g)\rho(\lambda_v)\rho(g)^{-1} = \rho(\lambda_{g(v)})$ for $g \in A$. These are the defining relations of Γ as a semidirect product, so ρ extends to a homomorphism $\rho : \Gamma \to 3V$. Suppose that $\rho(\gamma) = 1$, where γ has lamp part $\lambda : V \to S_3$ and affine part g. The lamps preserve every Z_v, so $\rho(g)$ does too, and $g(v) = v$ for all v; so $g = 1$. Then $\rho(\gamma)$ restricts on Z_v to $\iota_v\theta(\lambda(v))\iota_v^{-1}$, so $\lambda(v) = 1$ for all v. Hence ρ is injective. Explicitly, $\rho(a)$ exchanges the rectangles $00\mathfrak{C} \times 0\mathfrak{C} \times 0\mathfrak{C}$ and $010\mathfrak{C} \times 0\mathfrak{C} \times 0\mathfrak{C}$. Also $\rho(c) = \hat{D}\rho(ab)$, where $\rho(ab)$ permutes the rectangles with prefixes $(00, 0, 0)$, $(010, 0, 0)$ and $(011, 0, 0)$ cyclically in this order and $\hat{D}$ is the identity on Z_0.

The chunk. The sofic profile of a chunk depends only on its identity element and its partial multiplication (Section 2), and an injective homomorphism preserves both. So the chunk $E_{3V} = \rho(E)$, for the chunk E of Theorem 3.3, has $\sigma_{E_{3V}} = \sigma_E$, and parts (a) and (b) of that theorem give the bounds. □

By Section 8.2, Thompson's group V has chunks with finite superpolynomial profile, from its Houghton subgroups, and a chunk with profile at least $\exp(c\,r/\log^2 r)$, which is finite using the recent soficity result for V [72, Theorem 1.1]. Proposition 7.2 gives finite, nearly exponential profile in $3V$ with no assumption. Whether Γ embeds in V or in $2V$ is open (Section 16).

7.3 The Følner function

The hash models of Proposition 6.3 are far smaller than any Følner set of Γ, and this subsection measures the gap. For $r \geq 1$ let $\mathrm{Føl}(r)$ be the least size of a finite nonempty set $F \subset \Gamma$ with $|Fs \setminus F| \leq |F|/r$ for $s \in \{p, q, t, c, a\}$.

Proposition 7.3 (The Følner function)**.** *For $r \geq 12$,*

$$2^{\,2^{\lfloor r/12 \rfloor - 1}} \;\leq\; \mathrm{Føl}(r) \;\leq\; 2^{\,2^{3r+14}}.$$

Thus the Følner function of Γ is doubly exponential, while by Theorem 3.3(b) the sofic profile of every chunk is $\exp(O(r \log r))$.

Proof. Lower bound. Let F be such a set and $l = \lfloor r/12 \rfloor \geq 1$. For $v \in \{0,1\}^l$ the word $T_v = t^{v_1} p\, t^{v_2} p \cdots p\, t^{v_l} p^{1-l}$ is freely equal to $\prod_{j=1}^{l} p^{j-1} t^{v_j} p^{1-j}$, so it represents the translation by $\sum_j v_j e_j$, and it has length less than $3l$. The elements $a_v = T_v a T_v^{-1}$ are the copies of a at 2^l distinct configurations, so they are commuting involutions and form a basis of an elementary abelian group Q of order 2^M, $M = 2^l$. Each a_v is a product of fewer than $6l$ generators and inverses. Since $|Fxy \setminus F| \leq |Fx \setminus F| + |Fy \setminus F|$ and $|Fx^{-1} \setminus F| = |Fx \setminus F|$, we get $|Fa_v \setminus F| \leq 6l|F|/r \leq |F|/2$, and so $\sum_v |Fa_v \setminus F| \leq M|F|/2$.

On the other hand, for every finite $S \subset Q$,

$$\sum_v |Sa_v \setminus S| \;\geq\; |S|\,(M - \log|S|). \tag{6}$$

Indeed, the left side is $M|S| - 2e(S)$, where $e(S)$ is the number of pairs $\{x, xa_v\} \subset S$, and $e(S) \leq \frac{1}{2}|S| \log|S|$ by the edge-isoperimetric inequality of the cube [40]. We recall the proof, by induction on M. Split S by the last coordinate into S_0, S_1 of sizes $s_0 \leq s_1$. Then $e(S) \leq e(S_0) + e(S_1) + s_0 \leq \frac{1}{2}(s_0 \log s_0 + s_1 \log s_1) + s_0$, and this is at most $\frac{1}{2} s \log s$ with $s = s_0 + s_1$, because the binary entropy function h satisfies $h(x) \geq 2x$ for $0 \leq x \leq \frac{1}{2}$. The right multiplications by the a_v preserve each coset xQ; applying (6) to the sets $x^{-1}(F \cap xQ)$ and using $|F \cap xQ| \leq |F|$ gives $\sum_v |Fa_v \setminus F| \geq |F|(M - \log|F|)$. Hence $\log|F| \geq M/2 = 2^{l-1}$.

Upper bound. Let $m = \lceil (r-1)/2 \rceil$, $R = 2m + 3$ and $F = \bigcup_{z \in B_m} K_R g_z$, a disjoint union because $\pi(g_z) = z$. By Lemma 6.2, $g_z s \in K_R g_{z+\pi(s)}$ for $z \in B_m$ and each generator s, so $|Fs \setminus F| \leq |K_R|(2m+1) = |F|/(2m+1) \leq |F|/r$. With $D = |Y_R| = 6m + 9$ we have $|K_R| = 6^{2^D} \cdot 2^D \cdot |\mathrm{GL}_D(\mathbb{F}_2)|$, so $\log|F| \leq 2^D \log 6 + D + D^2 + 2\log(2m+1) \leq 2^{D+2} \leq 2^{3r+14}$. □

The group Γ is not LEF. This follows from $H_3 \leq \Gamma$ in the standard way. H_3 is finitely presented and not residually finite, because the finitary alternating group of an infinite set is simple. A finitely presented LEF group is residually finite [69], and subgroups of LEF groups are LEF. The chunk E shows it too. Its profile is unbounded by Theorem 3.3(a), and a chunk with unbounded profile embeds in no finite group (Section 2).

8 The ladder below Γ

Before Γ we found two groups that each supply part of what the criterion needs. We record them here for comparison, and because each leaves an open problem; nothing in this section is

used in the proof of Theorem A or in Part II. The dyadic group $\mathcal{G}$ of Section 8.1 is elementary amenable and transports copies of S_3 by the powers of one element, with far commutators of area at most quadratic. Whole suffix fillings and Theorem 2.3 give exponent $1/2$, improving the pairwise exponent $1/3$ (Section 2). The subgroup $\mathcal{L}$ of Thompson's group V in Section 8.2 transports along the exponentially growing orbits of Thompson's group F: words of length $O(\log N)$ reach N sites, and the quadratic Dehn function of F fills all pairwise commutators cheaply. This comes within a factor $\log^2 r$ in the exponent of the ceiling of Theorem 2.1. The recent preprint [72, Theorem 1.1 and Corollary 1.2] proves soficity of V and F; using its result and the embedding $\mathcal{L} \leq V$, the chunk below has finite profile. The proof that $\mathcal{L}$ is hyperlinear exactly when F is uses tools that nothing else in the paper needs, and it is in Appendix G.

8.1 An amenable group with a square-root lower bound

The group $\mathcal{G}$ is a lamplighter over the dyadic rationals $\mathbb{Z}[\frac{1}{2}]$: the lamps are copies of S_3, and the lighter is generated by the shift of the integers and by doubling. In this section a group acting on a set X acts on $\bigoplus_{x\in X} S_3$ by carrying the copy of S_3 at x to the copy at $g(x)$ through the identity of S_3.

Let $X = \mathbb{Z}[\frac{1}{2}]$, and let t and u be the permutations of X given by $t(x) = x + 1$ for $x \in \mathbb{Z}$, $t(x) = x$ for $x \notin \mathbb{Z}$, and $u(x) = 2x$. Let $\mathcal{B} = \langle t, u\rangle$ and $\mathcal{G} = (\bigoplus_{x\in X} S_3) \rtimes \mathcal{B}$, and let a, b be generating involutions of the copy of S_3 at 0. Put $s = utu^{-1}$ and $v = tst^{-1}$. Then s adds 2 to the even integers and v adds 2 to the odd integers, each fixing every other point, so $t^2 = sv$ and $[s, v] = 1$. Moreover u and v fix 0, and s fixes $1 = t(0)$. Since $\mathcal{B}$ acts transitively on X, the group $\mathcal{G}$ is generated by a, b, t and u. By these identities the following seventeen words in a, b, t, u, s, v represent 1 in $\mathcal{G}$; they form the set $R_{\mathcal{G}}$:

$$a^2,\ b^2,\ (ab)^3,\quad s^{-1}utu^{-1},\ v^{-1}tst^{-1},\ t^2v^{-1}s^{-1},\ [s,v],\quad [u,x],\ [v,x],\ [s,txt^{-1}],\ [x,tyt^{-1}],$$

where $x, y \in \{a, b\}$. Theorem 8.1 also records that $\mathcal{G}$ is not solvable, so the square-root lower bound holds outside the solvable class.

Theorem 8.1 (The dyadic group)**.** *The group $\mathcal{G}$ is elementary amenable and not solvable, and every commutator $[x, t^k y t^{-k}]$ with $x, y \in \{a, b\}$ and $k \geq 1$ has area at most $5k^2$ with respect to $R_{\mathcal{G}}$. Consequently the chunk $E_{\mathcal{G}}$ that Theorem 2.1 associates with $R_{\mathcal{G}}$ has finite sofic profile with*

$$\log \sigma_{E_{\mathcal{G}}}(r) \ \geq\ \frac{1}{16}\Big\lfloor\sqrt{\frac{r}{20000000}}\Big\rfloor \qquad \Big(\Big\lfloor\sqrt{\frac{r}{20000000}}\Big\rfloor \geq 8\Big),$$

every unitary assignment ϕ as in Theorem 2.1 has $\log D \geq \frac{\Delta^2}{12}((5K_\Delta e)^{-2/5} - 1)$, and $\mathcal{G}$ is not LEF.

Proof. Amenability. For $j \in \mathbb{Z}$ let $t_j = u^j t u^{-j}$; it adds 2^j to the points of $2^j\mathbb{Z}$ and fixes the others. The group $\mathcal{K}$ generated by all t_j is normalized by u and contains t, so it is normal in $\mathcal{B}$ and $\mathcal{B}/\mathcal{K}$ is cyclic. For $n \geq 1$ the elements t_j with $|j| \leq n$ move only points of $2^{-n}\mathbb{Z}$, and they commute with the translation $x \mapsto x + 2^n$ of $2^{-n}\mathbb{Z}$, which has 4^n orbits. A permutation of $2^{-n}\mathbb{Z}$ that commutes with this translation maps each orbit onto an orbit by a translation, so these t_j generate a subgroup of $\mathbb{Z}^{4^n} \rtimes \mathrm{Sym}_{4^n}$, which is virtually abelian. Hence $\mathcal{K}$ is a directed union of virtually abelian groups, $\mathcal{B}$ is elementary amenable, and so is $\mathcal{G}$, because $\bigoplus_X S_3$ is locally finite.

Not solvable. For $n \geq 1$ the elements $t_0, \dots, t_{n-1}$ preserve $\mathbb{Z}$ and commute with $x \mapsto x+2^n$, so they act on $\mathbb{Z}/2^n$, where they permute the classes modulo 2^i for every $i \leq n$. So they generate a subgroup Γ_n of the iterated wreath product W_n of n copies of $\mathbb{Z}/2$, which has order

2^{2^n-1}. On the even classes, identified with $\mathbb{Z}/2^{n-1}$ by halving, $t_1, \dots, t_{n-1}$ act as $t_0, \dots, t_{n-2}$ act on $\mathbb{Z}/2^{n-1}$, and they fix the odd classes. Conjugation by t_0 carries this action to the odd classes, and t_0 exchanges the two halves. So $|\Gamma_n| \geq 2|\Gamma_{n-1}|^2$, and $\Gamma_n = \mathsf{W}_n$ by induction from $\Gamma_1 = \mathsf{W}_1 = \mathbb{Z}/2$. The derived subgroup of W_n lies in $\mathsf{W}_{n-1} \times \mathsf{W}_{n-1}$ and projects onto each factor, since the commutator of $(g, 1)$ with the exchange is (g^{-1}, g). By induction the $(n-1)$st derived subgroup of W_n is nontrivial. Since Γ_n is a quotient of the subgroup $\langle t_0, \dots, t_{n-1} \rangle$ of $\mathcal{B}$, the derived length of $\mathcal{B}$ is at least n for every n, so neither $\mathcal{B}$ nor $\mathcal{G}$ is solvable.

Area. Let $A_{\mathcal{G}}(k)$ be the largest area of $[x, t^k y t^{-k}]$ over $x, y \in \{a, b\}$, so that $A_{\mathcal{G}}(1) = 1$. Five kinds of move cost one cell each.

- Replacing t^2 by sv.
- Exchanging adjacent letters $s^{\pm 1}$ and $v^{\pm 1}$.
- Replacing s by utu^{-1}.
- Moving a letter $x^{\pm 1}$, $x \in \{a, b\}$, across $u^{\pm 1}$ or $v^{\pm 1}$.
- Moving txt^{-1} or its inverse across $s^{\pm 1}$.

The commutator contains four powers $t^{\pm k}$.

For $k = 2m$, write each $t^{\pm 2m}$ as $(s^m v^m)^{\pm 1}$, with m replacements and $m(m-1)/2$ exchanges each, $2m^2 + 2m$ cells in all. Moving $y^{\pm 1}$ across $v^{\pm m}$ ($2m$ cells) leaves $[x, s^m y s^{-m}]$. Replacing its $4m$ letters $s^{\pm 1}$ gives $[x, ut^m u^{-1} y u t^{-m} u^{-1}]$, and moving $y^{\pm 1}$ and $x^{\pm 1}$ across u (four cells) gives $u[x, t^m y t^{-m}]u^{-1}$. Hence $A_{\mathcal{G}}(2m) \leq A_{\mathcal{G}}(m) + 2m^2 + 8m + 4$.

For $k = 2m + 1$, put $z = tyt^{-1}$, so that the commutator is freely $[x, t^{2m} z t^{-2m}]$. Write each $t^{\pm 2m}$ as $(v^m s^m)^{\pm 1}$, now with $m(m+1)/2$ exchanges each, $2m^2 + 6m$ cells in all. Moving $z^{\pm 1}$ across $s^{\pm m}$ ($2m$ cells) leaves $[x, v^m z v^{-m}]$, moving $x^{\pm 1}$ across $v^{\pm m}$ ($2m$ cells) leaves $v^m[x, z]v^{-m}$, and $[x, z]$ is a relation. Hence $A_{\mathcal{G}}(2m+1) \leq 2m^2 + 10m + 1$.

By induction $A_{\mathcal{G}}(k) \leq 5k^2$, since $A_{\mathcal{G}}(2) \leq 15$, $A_{\mathcal{G}}(m) + 2m^2 + 8m + 4 \leq 7m^2 + 8m + 4 \leq 20m^2$ for $m \geq 1$, and $2m^2 + 10m + 1 \leq 5(2m+1)^2$.

Profile. The words of $R_{\mathcal{G}}$ have length at most 8, so Corollary 2.2 with $C = 5$, $p = 2$ and $l = 8$ gives the unitary bound. The stronger permutation bound follows from Theorem 8.2 below. Since $\mathcal{G}$ is amenable, it is sofic and every chunk has finite profile. The profile of $E_{\mathcal{G}}$ is unbounded, so $E_{\mathcal{G}}$ embeds in no finite group, and $\mathcal{G}$ is not LEF (Section 2). □

The contrast with H_3 is instructive. There the doubling endomorphism of [26, Section 3], which differs by a conjugation from the doubling map of Burillo, Cleary, Martino and Röver [16, Section 7], multiplies a reweighted area by 107 at each step; hence the exponent $\log 107$. In $\mathcal{G}$ the dilation u relates distance $2m$ to distance m exactly, and the two commuting shifts cost $O(m^2)$ cells. The orbit of 0 under $\mathcal{B}$ grows exponentially, as the orbits of Thompson's group F do. Whether $\mathcal{G}$ also supports the transport of many copies that we use for F in Section 8.2 is open (Section 16).

Theorem 8.2 (Quadratic dyadic sweeps)**.** *Let $x \in \{a, b\}$ and*

$$P = t^{p_1} c_1 t^{p_2 - p_1} c_2 \cdots t^{p_m - p_{m-1}} c_m t^{-p_m}, \quad 1 \leq p_1 \leq \cdots \leq p_m \leq K, \quad c_j \in \{a, b\}.$$

If at most μ letters occur at any position, then

$$\mathrm{Area}_{R_{\mathcal{G}}}([x, P]) \leq \frac{8}{3}(1 + \mu/2)^2 K^2 + (10 + 4\mu)K + 2\log(K+1) + 2. \tag{7}$$

The empty block has area zero. The multiplicity hypothesis, and hence the constant's dependence on μ, is part of the assertion.

Proof. Let $F(K, L)$ denote the cost of the construction for the displayed lamp list L. Put $I = \lfloor p_m/2 \rfloor$, and let m_e, m_o be its even- and odd-position letter counts. Regard $z_c = tct^{-1}$ as one odd lamp unit. Free regrouping expresses P as a walk with paired t-steps, with even

lamps c and odd lamps z_c attached to that walk. In one occurrence of P, replace outbound and return t-pairs using $t^2 = sv$, at cost $2I$ cells. Call $s^{\pm1}, a, b$ even units, and $v^{\pm1}, z_a, z_b$ odd units. Each even unit commutes with each odd unit by one signed conjugate of one of the listed relations $[s, v], [v, c], [s, tct^{-1}], [c, tdt^{-1}]$. This includes inverse transport letters, and keeps each z_c intact. Sorting the two alphabets costs at most

$$(2I + m_e)(2I + m_o) \le (K + m/2)^2$$

per occurrence. No same-parity lamps are interchanged. The result is $P_E P_O$; mirror the derivation in the second occurrence in $[x, P]$. These pair replacements and two sorts cost at most $2K + 2(K + m/2)^2$.

Move x^{-1} left across the odd block, at cost at most $K + m_o$, and freely cancel that block to obtain $[x, P_E]$. In each of the two even blocks replace $s^{\pm1}$ by $ut^{\pm1}u^{-1}$ and move u^{-1} past its lamps, at cost at most $K + m_e$ per occurrence. It becomes $uP'u^{-1}$, where P' is the surviving even lamp list with every position halved. Two final $[u, x]$ moves leave $u[x, P']u^{-1}$. Thus

$$F(K, L) \le F(\lfloor K/2 \rfloor, L_e/2) + 2(K + m/2)^2 + 5K + 2m + 2.$$

If there are no even lamps, stop after removing the odd block. Surviving positions stay positive and retain multiplicity at most μ. At level j, let $K_j = \lfloor K/2^j \rfloor$ and let m_j count the actual original letters whose positions are divisible by 2^j, now placed at their halved positions. Sum the recurrence only over nonempty recursive levels, in particular $K_j \ge 1$:

$$F(K, L) \le \sum_{j \text{ active}} \{2(K_j + m_j/2)^2 + 5K_j + 2m_j + 2\}.$$

Since $m_j \le \mu K_j$, $\sum K_j^2 \le 4K^2/3$, $\sum K_j \le 2K$, and there are at most $\log(K + 1) + 1$ levels, this finite sum gives (7). □

Square-root profile bound in Theorem 8.1. Take copies transported by t^i. Conjugating a sweep by t^{-i} gives $[h, P]$ with P supported at positions $1, \dots, N - i$ and at most three letters at each position. Split h into at most three letters. Theorem 8.2, with $\mu = 3$, gives

$$\Lambda_N \le 50N^2 + 66N + 6\log(N + 1) + 6 \le 60N^2 \qquad (N \ge 8).$$

Apply Theorem 2.3 with $l = 8$. For $N = \lfloor \sqrt{r/20000000} \rfloor \ge 8$, $20000 \cdot 15(60N^2 + 40N) \le 20000000N^2 \le r$. Hence $\log \sigma_{E_G}(r) \ge N/16$. This uses exactly the existing chunk E_G; no transport words are added to it. Amenability gives finiteness, independently of this lower bound. □

Finite certified computations. The distributed witnesses record every substitution as a signed defining-relator conjugate or a free change. A separate checker reconstructs the boundary from the lamp list, checks each signed cell against the seventeen relations, and freely reduces the product of all cell conjugates to that boundary. Seven certificates passed this test: full three-letter blocks at $K = 16, 64, 128$ used $2646, 36282, 140860$ cells; a random 68-letter block reaching 47 used 9168 cells; and three random full blocks at $K = 128$ each used 140860 cells. These are finite certified computations of construction costs, not optimal-area claims. The files and reproduction command are described in `verification/README.md`.

8.2 Thompson's group V

Röver proved that every Houghton group embeds in Thompson's group V [64, Proposition 2.6], and Genevois extended the embedding to generalised Houghton groups [31, Corollary 5.11].

Since H_3 is amenable, Thompson's group V therefore has a chunk whose profile is finite at every r and superpolynomial by [26, Theorem 1.1]. A different subgroup of V gives a much larger bound.

Let $\mathbb{D}$ be the set of dyadic rationals in $(0,1)$, and let Thompson's group F act on $[0,1]$ by its piecewise-linear homeomorphisms. Let $\mathcal{L} = (\bigoplus_{x\in\mathbb{D}} S_3) \rtimes F$. To embed $\mathcal{L}$ in V we double every 1 of a binary word, so that the pattern 10 never occurs inside a coded word. The first 10 then marks a gap, the gaps are indexed by the dyadic rationals, and F acts on the gaps as it acts on $\mathbb{D}$.

Lemma 8.3 (Gap coding)**.** *$\mathcal{L}$ embeds in V.*

Proof. Let c be the substitution $0 \mapsto 0$, $1 \mapsto 11$ on finite binary words. An infinite binary word either is a concatenation of blocks 0 and 11, or it has a first occurrence of 10 at a block boundary. So Cantor space is the disjoint union of the set $\mathcal{C}$ of block sequences and the cylinders $Z_w = \mathsf{c}(w)10\{0,1\}^{\mathbb{N}}$, one for each finite word w. Index Z_w by the dyadic $x_w = 0.w1 \in \mathbb{D}$ in binary; this is a bijection. Let $f \in F$ have a tree-pair diagram with leaves $w_1,\dots,w_n$ and $w'_1,\dots,w'_n$, so that $f(0.w_k z) = 0.w'_k z$. Define $\hat{f}$ by the prefix replacements $\mathsf{c}(w_k) \mapsto \mathsf{c}(w'_k)$ and $\mathsf{c}(v)10 \mapsto \mathsf{c}(v')10$, where the interior vertices v of the first tree are matched in left-to-right order, the k-th with the k-th, with the interior vertices v' of the second. Both families of prefixes are complete prefix codes, because splitting a leaf w replaces $\mathsf{c}(w)$ by $\mathsf{c}(w0)$, $\mathsf{c}(w)10$ and $\mathsf{c}(w1)$. Expanding a matched pair of leaves adds a matched pair of gaps and changes nothing else, so $\hat{f}$ depends only on f. It maps Z_w onto $Z_{w'}$ with $x_{w'} = f(x_w)$: below a leaf this is the formula for f, and interior vertices correspond to the breakpoints between leaves. On $\mathcal{C}$ it acts as f acts on binary sequences, so $f \mapsto \hat{f}$ is an injective homomorphism $F \to V$. Inside each Z_w let S_3 permute the three cylinders $\mathsf{c}(w)10y\{0,1\}^{\mathbb{N}}$, $y \in \{0,10,11\}$. Copies with different w have disjoint supports, and $\hat{f}$ carries the copy at x to the copy at $f(x)$. An element of the group they generate that acts trivially acts trivially on $\mathcal{C}$, so its F-part is trivial, and then each of its S_3 coordinates is trivial. □

Let a, b generate the copy of S_3 at $\frac{1}{2}$, let $t \in F$ be

$$t(x) = \begin{cases} 2x, & 0 \le x \le \frac{1}{4}, \\ x + \frac{1}{4}, & \frac{1}{4} \le x \le \frac{1}{2}, \\ \frac{x+1}{2}, & \frac{1}{2} \le x \le 1, \end{cases}$$

so that $t(\frac{1}{2}) = \frac{3}{4}$, and for $m \ge 1$ let $\mathbb{D}_m$ be the set of the $2^m - 1$ points of $\mathbb{D}$ with denominator at most 2^m.

The group $\mathcal{L}$ is finitely presented. This is Cornulier's criterion for permutational wreath products [24, Theorem 1.1], applied to the action of F on $\mathbb{D}$; he checks its hypotheses in [24, Example 3.4]. Thompson's group F is finitely presented, and it acts transitively on increasing pairs of points of $\mathbb{D}$ [18, Lemma 4.2], so it has finitely many orbits on $\mathbb{D} \times \mathbb{D}$. The stabilizer $F_{1/2}$ of $\frac{1}{2}$ is the product of the elements supported in $[0, \frac{1}{2}]$ and those supported in $[\frac{1}{2}, 1]$, two copies of F, so it is finitely generated. Nothing in Proposition 8.4 uses finite presentation.

Proposition 8.4 (Thompson transport)**.** *The group $\mathcal{L}$ contains F and contains no Houghton group H_m, so the bounds below do not come from the Houghton subgroups of V. There are a finite set R of relations of $\mathcal{L}$, containing $a^2, b^2, (ab)^3$, and a constant C with the following property: for every $m \ge 1$ there are words w_p, $p \in \mathbb{D}_m$, with $w_p(\frac{1}{2}) = p$ and*

$$\mathrm{Area}_R([w_p x w_p^{-1}, w_q y w_q^{-1}]) \le C m^2 \qquad (p \ne q \in \mathbb{D}_m,\ x, y \in \{a, b\}).$$

Consequently a chunk E of $\mathcal{L}$, and hence of V, has $\log \sigma_E(r) \ge c\, r / \log^2 r$ for all large r, and for $0 < \Delta \le 1$ every unitary assignment as in Theorem 2.1 has $\log D \ge c_\Delta\, e^{-2} / \log^4(1/e)$ for all small e, where $c > 0$ is a constant and $c_\Delta > 0$ depends only on Δ.

Proof. Transport. A standard dyadic interval is one of the form $[k2^{-j}, (k+1)2^{-j}]$. Let two partitions of $[0,1]$ into standard dyadic intervals have the same number n of intervals. The map sending the ith interval of the first affinely onto the ith interval of the second lies in F and has a tree-pair diagram with $n-1$ carets. Every standard dyadic interval can be split into any given number of standard dyadic intervals. For $p < q$ in $\mathbb{D}_m$, the binary digits of p split $[0,p]$ and $[p,1]$ into at most m standard dyadic intervals each. The interval $[p,q]$, cut at its point c with the shortest binary expansion, splits into at most m standard dyadic intervals on each side of c. Hence there are $w_p \in F$ with $w_p(\frac{1}{2}) = p$ and fewer than $2m$ carets, and $g_{pq} \in F$ with $g_{pq}(\frac{1}{2}) = p$, $g_{pq}(\frac{3}{4}) = q$ and fewer than $4m$ carets. The word length in F is comparable to the number of carets of the reduced tree-pair diagram [17]. Hence these elements are represented by words of length $O(m)$ in the generating set of the finite presentation of F that the set R of relations uses. The Dehn functions of two finite presentations of F are equivalent, so Guba's quadratic bound holds for it. We write w_p also for such a word. An element fixing $\frac{1}{2}$ maps $[0, \frac{1}{2}]$ to itself, so the left and right subtrees of its reduced diagram are diagrams for its two restrictions, which are elements of F after affine rescaling, with fewer carets in total. Hence $F_{1/2} \cong F \times F$ is undistorted in F: its elements of length n in F are words of length $O(n)$ in a fixed finite set of generators h_j of $F_{1/2}$.

Relations. Let R consist of a finite presentation of F on a generating set that includes t and the h_j, the words $a^2, b^2, (ab)^3$, the commutators $[x, h_j]$, and the commutators $[x, tyt^{-1}]$, for $x, y \in \{a, b\}$. They hold in $\mathcal{L}$: each h_j fixes $\frac{1}{2}$, and the copies at $\frac{1}{2}$ and at $t(\frac{1}{2}) = \frac{3}{4}$ commute.

Area. Let $p < q$ in $\mathbb{D}_m$ and $g = g_{pq}$. The elements $g^{-1}w_p$ and $t^{-1}g^{-1}w_q$ fix $\frac{1}{2}$ and have length $O(m)$, so they are represented by words ω, ω' of length $O(m)$ in the h_j. The words $w_p^{-1}g\omega$ and $w_q^{-1}gt\omega'$ are null in F and have length $O(m)$, so Guba's quadratic Dehn function [36] fills each with $O(m^2)$ cells. Replacing the four occurrences of $w_p^{\pm 1}$ in $[w_p x w_p^{-1}, w_q y w_q^{-1}]$ by $(g\omega)^{\pm 1}$, and the four of $w_q^{\pm 1}$ by $(gt\omega')^{\pm 1}$, therefore costs $O(m^2)$ cells. Moving $x^{\pm 1}$ across ω and $y^{\pm 1}$ across ω' with the relations $[x, h_j]$ costs $O(m)$ cells and leaves $g[x, tyt^{-1}]g^{-1}$, a conjugate of a relation. For $p > q$ the commutator is the inverse of the one with the two copies exchanged. So every area is $O(m^2)$.

Profile. Apply Theorem 2.1 with $N = 2^m - 1$ and $\Lambda = Cm^2$, where $C \geq 1$, and let l be the largest length of a word of R. Given r, let m be the largest integer with $2^m m^2 \leq r/(10^8 C(2l-1))$; then 2^m is at least of order $r/\log^2 r$, and $\log \sigma_E(r) \geq (2^m - 1)/16$. Given e, let m be the largest integer with $2^{m/2} m^2 \leq 1/(CK_\Delta e)$; then 2^m is at least of order $(CK_\Delta)^{-2} e^{-2}/\log^4(1/e)$, and $\log D \geq (2^m - 1)\Delta^2/12$.

No Houghton subgroup. A torsion element of $\mathcal{L}$ maps to a torsion element of the torsion-free group F, so it lies in $\bigoplus_{\mathbb{D}} S_3$, which has exponent six. Every H_m contains a five-cycle. Finally, $\mathcal{L}$ embeds in V by Lemma 8.3. □

By the recent soficity result for V [72, Theorem 1.1] and Lemma 8.3, $\mathcal{L}$ is sofic. Thus the chunk in Proposition 8.4 has finite profile and satisfies $\log \sigma_E(r) \geq cr/\log^2 r$ for all large r. The corresponding unitary models also exist, since sofic groups are hyperlinear; Proposition 8.5 records the separate transfer argument.

Slofstra's bounds are of the same kind: they bound models that may not exist, since his group is not known to be hyperlinear. Any unitary models of it need $\log D \gtrsim e^{-c}$ for every $c < \frac{1}{2}$ [67, Theorem 1.2 and the remark after it], and the proof of [67, Corollary 1.3] shows that any permutation models of a suitable chunk need $\log \sigma \gtrsim r^c$ for every $c < \frac{1}{4}$. Slofstra's profiles may be infinite. The group $\mathcal{L}$ gives e^{-2} and r up to logarithms.

Since $\mathbb{D}$ is the coset space $F/F_{1/2}$, the group $\mathcal{L} = (\bigoplus_{F/F_{1/2}} S_3) \rtimes F$ has the shape $(\bigoplus_{G/H} A) \rtimes G$ of the generalised wreath products of Kun and Thom [54, Theorem A]. They prove such a product non-sofic for $A = \mathbb{Z}/2$ when $H < G$ is infranormal and not normal, that

is, the elements g with $gHg^{-1} \subseteq H$ generate G, and both H and G have property (T). Our input is different: cheap transport of non-abelian copies of S_3. Soficity and hyperlinearity pass to wreath products whose acting group is sofic [41]. They pass to permutational wreath products when the acting group is sofic (respectively hyperlinear) and the action is sofic [30, Theorems 3.6 and 3.8], [1]. Since F acts on $\mathbb{D}$ transitively and faithfully, a sofic action of F on $\mathbb{D}$ would already make F sofic [30, Theorem 2.12], so these results give nothing here. Proposition 8.5 assumes only that F is hyperlinear.

Proposition 8.5 (Hyperlinearity transfer)**.** *The group $\mathcal{L}$ is hyperlinear if and only if F is.*

The proof, in Appendix G, needs a copy of S_3 in the ultrapower $L(F)^\omega$ at $\frac{1}{2}$ that commutes with the stabilizer $F_{1/2}$. Near $\frac{1}{2}$ the stabilizer acts by dilations. The proof therefore places a copy of S_3 in each of a sequence of intervals shrinking to $\frac{1}{2}$, and selects among them by a random index whose law is invariant under dilation in the limit.

Part II
The channel (Theorem B)

9 Causal services and gadget channels

Part II can be read on its own by a reader who comes from quantum information. From Part I it needs three definitions, all in Section 2. A *chunk* of a group is a finite subset containing 1, with the partial multiplication it inherits. A *sofic model* of the chunk at accuracy $1/r$ assigns a permutation of a finite set to each element, with products correct and distinct elements separated on all but a fraction $1/r$ of the points. The *profile* $\sigma_E(r)$ is the least number of points of such a model. The configuration lamps enter mainly through two facts, quoted where they are used: their pair areas (Theorem 5.1) and their profile, Theorem A in the precise form of Theorem 3.3. Concretely, the lower bounds use one black box, the transport of Section 11: for each N there are 2^N words of length at most $63N^2$ whose copies of S_3 commute pairwise at area at most $C_1 N^\kappa$, with $\kappa = 6 + \log 107$ (the introduction says where 107 comes from). The upper bounds use another, the hash models behind Theorem A(b), which give sofic models with $\exp(O(r \log r))$ points.

A finite set of relations of a group becomes a quantum channel of fixed dimension. Each relation contributes a few small unitary blocks; the channel evaluates the blocks in the regular representation of the group and traces out the group von Neumann algebra. Every channel on $\mathbb{C}^d$ has a Stinespring dilation, a unitary on the input and a pure environment of dimension at most d^2 that is discarded after use. Applied n times with a fresh environment each time, the channel needs almost no memory, but the discarded environments carry away entropy that fresh pure qubits paid for. We charge for that purity, and the memory is never reset. The question is how much memory a device needs when its purity is far below what the Stinespring device spends.

In [27] we showed that the memory needed to apply such a channel repeatedly, releasing each output before the next input arrives, is bounded below by the size of approximate representations of its relations. Part II sharpens that link, so that it loses a single square root, and applies it to the configuration lamps. The outcome is Theorem B, stated precisely as Theorem 11.1: an explicit channel in the closure of finite tracial channels whose repeated use with logarithmic purity needs memory $n^{1-o(1)}$, while memory $O(n \log n)$ suffices. The lower bounds come from the sharpened link, the weighted one-round theorem of Section 10. The upper bounds come from the sofic models of Γ, which give devices directly (Proposition 12.1) and, at the rank rate, through Theorem 12.2.

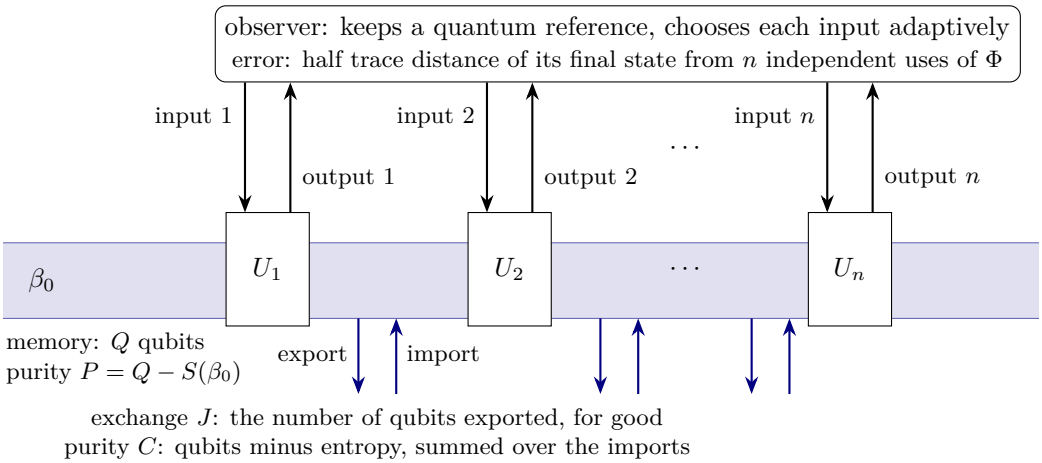


Figure 3: A causal service of Φ for n rounds. In each round the input enters, a fixed unitary U_t acts on it and the memory, and the output is released before the next input arrives; after each round but the last, a prescribed set of memory qubits is exported for good and replaced by the same number of fresh qubits in a fixed state. The resources are the memory Q, the exchange J, the purity $P + C$, and the error, the largest half trace distance between the observer's final state and its final state after n independent uses of Φ.

We recall the framework of [27] and keep its notation.

Channels. Channels act on $M_d = M_d(\mathbb{C})$. The normalized Choi state of a channel Ψ is $\mathcal{J}(\Psi) = (\Psi \otimes \mathrm{id})(|\Omega\rangle\langle\Omega|)$, with $|\Omega\rangle = d^{-1/2}\sum_i |ii\rangle$, and $d_\diamond(\Psi, \Psi') = \frac{1}{2}\|\Psi - \Psi'\|_\diamond$ is the diamond distance. We write τ_M for the normalized trace on M_M. A channel Ψ has a *finite tracial factorization* with bath dimension M if $\Psi(X) = (\mathrm{id} \otimes \tau_M)[U(X \otimes I_M)U^*]$ for a unitary U on $\mathbb{C}^d \otimes \mathbb{C}^M$, and $\mathcal{K}_d$ denotes the closure of the set of channels on M_d with a finite tracial factorization of some bath dimension. It is also the closure of the convex hull of that set [59, Proposition 3.8(ii)], and it consists of unital channels. Channels with a finite tracial factorization are exactly the noisy operations of [44] from M_d to M_d, and $\mathcal{K}_d$ is their closure.

Causal services. A *causal service* of a channel Φ on M_d for n rounds specifies a *device*, its physical realization, which serves one use of the channel in each round. The device has a memory register of Q qubits, started in a fixed state β_0. In each round t it applies a unitary U_t, fixed in advance, to the input and the memory and then releases the d-dimensional output. After each round but the last, a prescribed set of memory qubits is exported for good and replaced by the same number of imported qubits in a fixed state. That state may be correlated within one step but is independent of everything else. An observer chooses the inputs adaptively, possibly entangled with references of its own. The *error* of the service is the largest half trace distance between the observer's final state in this experiment and in the ideal experiment with n independent uses of Φ. Figure 3 shows the device and its rounds.

The resources are the memory Q, the exchange J and the purity $P + C$. The *exchange* is the total number of exported qubits; exporting is the only way the device can discard a qubit without handing it to the observer. The *purity* is an entropy deficit in bits, not $\mathrm{Tr}\,\rho^2$. Here $P = Q - S(\beta_0)$ is the deficit of the initial memory, and C is the sum over the imports of their number of qubits minus their entropy. A pure qubit is a resource, since it costs work to prepare, while a maximally mixed one is free; the Stinespring device of a channel with Choi rank r_* spends about $\log r_*$ bits of purity in every round. Every retained qubit counts; there is no free reset, measurement record or postselection.

Gadget channels. Let G be a group generated by a finite set S, and let R be a finite set of words in $S^{\pm 1}$, each of length at least two and each representing 1 in G. Write $r_i = \sigma_{i,1} \cdots \sigma_{i,\ell_i}$ $(1 \le i \le r_0)$ for the words of R, and put $s = |S|$ and $\Lambda = \sum_i \ell_i$. The *symbols* are the $2s$ letters $x^{\pm 1}$, $x \in S$, and the proper prefixes $p_{i,l} = \sigma_{i,1} \cdots \sigma_{i,l}$, $2 \le l \le \ell_i - 1$, one for each pair (i, l); we also put $p_{i,1} = \sigma_{i,1}$ and let p_{i,ℓ_i} be the empty word. The *triangles* are the triples $(p_{i,l-1}, \sigma_{i,l}, p_{i,l})$, $2 \le l \le \ell_i$, and $(x, x^{-1}, \varnothing)$, $x \in S$. For unitaries $\lambda(w)$ assigned to the symbols, with $\lambda(\varnothing) = I$, let

$$U_\lambda = \operatorname{diag}\big(I, (\lambda(w))_w\big) \;\oplus \bigoplus_{(x,y,z)} \frac{1}{\sqrt{2}} \begin{pmatrix} I & \lambda(y) \\ \lambda(x) & -\lambda(x)\lambda(y) \end{pmatrix},$$

where w runs over the symbols and (x, y, z) over the triangles. Each triangle block equals $\operatorname{diag}(I, \lambda(x))\,(H \otimes I)\operatorname{diag}(I, \lambda(y))$ with H the Hadamard matrix, so U_λ is unitary for all unitary λ. A triangle is a Hadamard test of one step $xy = z$ of one relation. Its block is unitary whatever is put in it, and its lower right entry $-\lambda(x)\lambda(y)$ is compared, through the Choi support defined next, with the entry that carries the label of z. A device that reproduces the channel must pass every test with the operators it holds in memory. There are $2s + \Lambda - 2r_0$ symbols and $\Lambda - r_0 + s$ triangles, so the system dimension is

$$d = 1 + (2s + \Lambda - 2r_0) + 2(\Lambda - r_0 + s) = 1 + 4s + 3\Lambda - 4r_0. \tag{8}$$

The *gadget channel* Φ_G on M_d evaluates each symbol w in the left regular representation, $\lambda(w) = \lambda([w])$ with $[w] \in G$ the element that w represents, and traces out the group von Neumann algebra $L(G)$ with its canonical trace τ: $\Phi_G(X) = (\mathrm{id} \otimes \tau)[U_\lambda(X \otimes 1)U_\lambda^*]$. In particular Φ_G is factorizable, through $(L(G), \tau)$. Since the relations hold, the lower right entry of the block of a triangle (x, y, z) is $-\lambda([z])$.

We write the gadget as $U_\lambda = \sum_{a \in \mathcal{T}} c_a E_a \otimes \lambda([w_a])$, where $\mathcal{T}$ is the set of *slots*, the nonzero entries, E_a is the matrix unit of slot a, and $c_a \in \{1, \pm 2^{-1/2}\}$. The word w_a is empty, a symbol, or the concatenation xy in the lower right corner of a triangle. The elements $[w_a]$ are the *labels*. With $A_g = \sum_{[w_a]=g} c_a E_a$, the matrices A_g have disjoint supports, $\Phi_G(X) = \sum_g A_g X A_g^*$ and $\mathcal{J}(\Phi_G) = \frac{1}{d}\sum_g |A_g\rangle\!\rangle\langle\!\langle A_g|$, because $\tau(\lambda(g)\lambda(h)^*) = \delta_{gh}$. The Choi rank r_* of Φ_G is therefore the number of distinct labels. Every label is carried by an *anchor*, a diagonal slot with coefficient 1 whose row and column contain no other slot. The anchor of the label 1 is the first diagonal entry, and every other label has a symbol entry as anchor, since the lower right entry of a triangle carries the label of a prefix symbol or 1. Feeding one half of a maximally entangled state on the r_* anchor positions gives a pure probe whose output is flat of rank r_*, and $\log r_*$ is the largest entropy of a complementary output of Φ_G [27, Proposition 9.1].

The rank rate. The probe fixes the least exchange that bounded memory allows. Over n rounds the ideal experiment hands the observer $n \log r_*$ bits of entropy. The observer's final state is a marginal of a pure state whose other half consists of the final memory, the exported qubits and the purifiers of the initial memory and of the imports. Those are at most $2(Q + J)$ qubits in all, so a device hands over at most $2(Q + J)$ bits. With memory sublinear in n it must export about $\frac{1}{2}\log r_*$ qubits per round at least. This rate is the optimal exchange rate of a gadget channel in $\mathcal{K}_d$ [27, Propositions 3.1 and 9.1], and we call it the *rank rate*: a service is at the rank rate if $J \le \frac{n}{2}\log r_*$. For the lamp channel, Theorem 12.2 attains it with memory $n^{1/2+o(1)}$.

One round. Let U be a unitary on $\mathbb{C}^d \otimes \mathbb{C}^M$ and ρ a state on $\mathbb{C}^M$, and put

$$\Psi(X) = \operatorname{Tr}_M U(X \otimes \rho)U^*, \qquad T(\rho) = \operatorname{Tr}_d U(\tau_d \otimes \rho)U^*.$$

Relative to a channel Φ with Choi support projection Π_Φ, the *support leakage* of Ψ is $\ell = \mathrm{Tr}[(I - \Pi_\Phi)\mathcal{J}(\Psi)]$. The *entropy production* of the round is $\Delta S = S(T(\rho)) - S(\rho)$, which is nonnegative because T is unital. A service produces such rounds, and the next result, from [27, Theorem 5.2], finds one in which both quantities are small. Appendix B reproduces its proof.

Theorem 9.1 (Sequential extraction)**.** *Let a causal service of a channel Φ on M_d for n rounds have memory Q, error $\epsilon < 1/2$ and $P + C \le B$, and put*

$$t_n = \frac{B + \epsilon + h_2(\epsilon) + \epsilon \log n}{n(1-\epsilon)}.$$

Then some round t, with its unitary U_t and the memory state ρ_t entering it, defines a channel $\Psi_t(X) = \mathrm{Tr}_M\, U_t(X \otimes \rho_t)U_t^*$ with bath dimension $M = 2^Q$, $d_\diamond(\Psi_t, \Phi) \le \epsilon/(1-\epsilon)$, and support leakage ℓ_t and entropy production ΔS_t satisfying $\ell_t + \Delta S_t \le t_n$.

The bath of this round is the memory itself, of dimension 2^Q, so $S(\rho_t) \le Q$. Every lower bound on memory in Part II is a lower bound on the entropy of a memory state that realizes one good round.

A worked example: the gadget of S_3. Take $G = S_3$ with $S = \{a, b\}$ and $R = \{a^2,\, b^2,\, (ab)^3\}$, so that $s = 2$, $r_0 = 3$ and $\Lambda = 2 + 2 + 6 = 10$. The symbols are the four letters $a^{\pm 1}, b^{\pm 1}$ and the four proper prefixes ab, aba, $abab$, $ababa$ of $(ab)^3$; the words a^2 and b^2 have no proper prefix of length two or more. The triangles are the nine triples $(a, a, \varnothing)$, $(b, b, \varnothing)$, (a, b, ab), (ab, a, aba), $(aba, b, abab)$, $(abab, a, ababa)$, $(ababa, b, \varnothing)$, $(a, a^{-1}, \varnothing)$ and $(b, b^{-1}, \varnothing)$. Formula (8) gives $d = 1 + 8 + 2 \cdot 9 = 27$. The labels are the elements of S_3 that the slots carry, namely 1, a, b, ab, aba, $abab = ba$ and $ababa = b$; these are all six elements, so $r_* = 6$. Since S_3 is finite, its group von Neumann algebra is the regular representation on $\mathbb{C}^6$ with the normalized trace. The channel Φ_{S_3} therefore has a finite tracial factorization with bath dimension 6 by definition, and one round with a flat bath on six points reproduces it exactly. Serving n rounds costs little more. The regular action of S_3 on itself is a model with no bad points, and Theorem 12.2 with $k = k' = 6$ gives a service at the rank rate with purity at most 2 and memory $O(\log n)$. The lamp channel of Section 11 adds five letters and twenty-six words to these three. What those words cost a device is what Theorem B measures.

10 The weighted one-round theorem

The engine of Part II reads an approximate representation of the relations directly off the memory state of one round, weighted by that state, and turns many cheaply commuting copies of S_3 into entropy of the memory. We state it here with its amplified form; the proof is in Appendix C. Transport words $u_1, \dots, u_N$ over $S^{\pm 1}$ have *length at most L* and *pairwise areas at most A* if $|u_i| \le L$ for all i and

$$\mathrm{Area}_R([u_i x u_i^{-1},\, u_k y u_k^{-1}]) \le A \qquad (i \ne k,\ x, y \in \{a, b\}),$$

where Area_R is the area in the free group on S, as in Section 2.

Theorem 10.1 (Weighted one-round theorem)**.** *Let G, S and R be as in Section 9, and suppose that S contains a, b and j, that a^2, b^2, $(ab)^3$ and $j^{-1}ab$ lie in R, and that $j \ne 1$ in G. Put $m = \max\{6, \ell_1, \dots, \ell_{r_0}\}$. Let $u_1, \dots, u_N$ be transport words of length at most L and pairwise areas at most A. Let $\Psi(X) = \mathrm{Tr}_M\, U(X \otimes \rho)U^*$ be one round with $d_\diamond(\Psi, \Phi_G) \le 1/8$, support leakage ℓ relative to Φ_G and entropy production ΔS, and put $\nu = \ell + \Delta S$. If*

$$2^{25}\, m^2 \sqrt{d}\, \sqrt{N}\, (A + L + 1)\sqrt{\nu} \;\le\; 1, \tag{9}$$

then $S(\rho) \geq N/16$.

The theorem reads the round at relator defect of order $\sqrt{\nu}$, a single square root of the quantity that Theorem 9.1 controls; a service with purity $O(\log n)$ is thus read at defect about $n^{-1/2}$. It does so without tracialization, working with the actual memory state ρ, which need not be flat. The general correspondence of [27] instead replaces the round by a nearby finite tracial mixture, which costs a second square root and reads the profile at defect about $n^{-1/4}$. That paper carried the weighted route out for a single gadget, built from Houghton's group. Theorem 10.1 holds for every relation set that contains the relations of S_3 and a name for ab.

One round reveals a representation because the Choi support of Φ_G records which slots carry the same label. A round with small leakage must therefore put nearly the same bath operator at every slot of a given label; the anchors make these operators unitaries, and the triangle tests make them satisfy the relations. Small entropy production makes them nearly commute with $\sqrt{\rho}$, and that is what lets every error be measured in the norm weighted by the memory state rather than in the flat one. Nearly commuting copies of S_3 then cost entropy. Measure the N three-cycles one after another on the memory. Each outcome is nearly a fair coin, because the involution of its copy flips it, yet a purifying reference nearly knows it, because the copy nearly commutes with $\sqrt{\rho}$. Each copy thus contributes about one bit of mutual information with the reference, and that mutual information is at most $S(\rho)$.

Applied to a family of transport words that grows with a parameter, Theorem 10.1 gives a lower bound on the entropy of one round in terms of ν alone.

Corollary 10.2 (Amplification)**.** *Keep the assumptions of Theorem 10.1 on G, S, R and the round, and let $K \geq 1$ and $\kappa \geq 0$.*

(a) Suppose that for every integer $k \geq 1$ there are 2^k transport words of length at most L_k and pairwise areas at most A_k, with $A_k + L_k + 1 \leq K k^{\kappa}$. Put $C_R = 2^{50} m^4 d K^2$. There is y_0 depending only on κ such that

$$S(\rho) \;\geq\; \frac{y}{32\,(\log y)^{2\kappa}} \qquad \text{for every } y \geq y_0 \text{ with } C_R\,\nu\,y \leq 1.$$

(b) Suppose instead that $A_k + L_k + 1 \leq 2^{K\sqrt{k}}$ for every $k \geq 1$, and put $C'_R = 2^{50} m^4 d$. There is y_1 depending only on K such that

$$S(\rho) \;\geq\; \frac{y}{32}\, 2^{-2K\sqrt{\log y}} \qquad \text{for every } y \geq y_1 \text{ with } C'_R\,\nu\,y \leq 1.$$

Proof. (a) Let $y_0 \geq 4$ be such that $k = \lfloor \log y - 2\kappa \log\log y \rfloor \geq 1$ for $y \geq y_0$. Then $k \leq \log y$ and $2^k \leq y(\log y)^{-2\kappa}$, so $2^k k^{2\kappa} \leq y$. With $N = 2^k$, the left side of (9) is at most $2^{25} m^2 \sqrt{d}\, K\, (2^k k^{2\kappa})^{1/2} \sqrt{\nu} \leq (C_R \nu y)^{1/2} \leq 1$. So $S(\rho) \geq 2^k/16$, and $2^k > y(\log y)^{-2\kappa}/2$.

(b) Let y_1 be such that $k = \lfloor \log y - 2K\sqrt{\log y} \rfloor \geq 1$ for $y \geq y_1$. Then $K\sqrt{k} \leq K\sqrt{\log y}$, so $2^k (A_k + L_k + 1)^2 \leq 2^{k + 2K\sqrt{\log y}} \leq y$, and (9) holds with $N = 2^k$ as in part (a). Hence $S(\rho) \geq 2^k/16 > y\, 2^{-2K\sqrt{\log y}}/32$. □

11 The channel of the configuration lamps, and Theorem B

This section builds the channel of Theorem B from the certificate of Theorem A and states the theorem in full. Theorem 10.1 needs a, b and a name for ab among the generators. Let R be the certificate of Table 1, in the five generators p, q, t, c, a, with b standing for the word ac^4. Let $S_\Phi = \{p, q, t, c, a, b, j\}$, where the new letters represent $b = ac^4$ and $j = ab$ in Γ, and let R_Φ consist of the 25 words of R and the four words

$$b^{-1} a c^4, \qquad b^2, \qquad (ab)^3, \qquad j^{-1} a b$$

in the letters of S_Φ. They represent 1 in Γ, $a^2 \in R$, and $j \neq 1$ because ab has order three (Lemma 3.1). The channel of the configuration lamps is the gadget channel $\Phi = \Phi_\Gamma$ of R_Φ.

Dimension and rank. The channel by the numbers:

letters	words	total length	symbols	triangles	longest word	dimension	Choi rank
$s = 7$	$r_0 = 29$	$\Lambda = 320$	276	298	$m = 24$	$d = 873$	$r_* = 130$

The total length is $\Lambda = 303 + 6 + 2 + 6 + 3$, and (8) gives $d = 1 + 4 \cdot 7 + 3 \cdot 320 - 4 \cdot 29 = 873$. The Choi rank r_* is the number of distinct labels. Of the 277 candidates, the identity and the 276 symbols, exactly 130 are distinct elements of Γ, so $r_* = 130$. Most coincidences are shared prefixes: pq and pqp^{-1}, for instance, begin several words. The others are relations such as $a^2 = 1$ and $j = ab = c^4$. The script `verification/choi_rank_check.py` computes this partition with two independent engines, the faithful action of Γ on configurations and the embedding in $3V$ (Appendix I.1). The rank rate is therefore $\frac{1}{2}\log 130 \approx 3.51$ qubits per use.

Transport. For $N \geq 1$, the proof of Theorem 3.3 provides 2^N words T_u, one for each configuration u supported in the first N points of ray 1, with $|T_u| \leq 21N \max(3, N) \leq 63N^2$ (Theorem 5.1) and $\mathrm{Area}_{R_\Phi}([T_u x T_u^{-1}, T_w y T_w^{-1}]) \leq \Lambda_N + 4 \leq C_1 N^\kappa$ for $u \neq w$ and $x, y \in \{a, b\}$, where

$$p_H = \log 107 + 4, \qquad \kappa = 2 + p_H < 12.75,$$

and C_1 is absolute; the letter b is converted to ac^4 by the word $b^{-1}ac^4$, at most four times, and areas with respect to $R_\Phi \supseteq R$ are no larger than with respect to R. So the hypothesis of Corollary 10.2(a) holds with this κ and $K = C_1 + 64$.

Theorem 11.1 (Theorem B)**.** *The channel Φ on M_{873} lies in $\mathcal{K}_{873}$, and its Choi rank is $r_* = 130$. Let $\kappa = 2 + p_H = 6 + \log 107$. There are constants $c, c_0, K > 0$ and n_0 such that the following hold; all services have error at most $1/16$.*

(a) Memory against purity. *For $n \geq n_0$ and $\log n \leq B \leq c_0 n$, every causal service of Φ for n rounds with purity $P + C \leq B$ has, at any exchange,*

$$Q \;\geq\; c\,\frac{n}{B\,(\log n)^{2\kappa}}.$$

For every $n \geq 1$ and $0 < B \leq n$ there is an exact causal service, with error 0, purity $P + C < B$, exchange $J = (n-1)Q$ and memory $Q \leq K(n/B)\log(2n/B)$. In particular, for every $A_0 \geq 0$ there is $c' > 0$ such that services with $P + C \leq A_0 \log n$ have $Q \geq c' n/(\log n)^{2\kappa+1}$ for all large n, where $2\kappa + 1 = 2p_H + 5 < 26.49$, and there are services with $P = C = 0$ and $Q \leq Kn \log n$.

(b) The rank rate. *For $n \geq n_0$, every causal service of Φ for n rounds at the rank rate, $J \leq \frac{n}{2}\log r_*$, has, whatever its purity,*

$$Q \;\geq\; c\,\frac{\sqrt{n}}{(\log n)^{\kappa}}.$$

For every $n \geq 2$ there is a causal service at the rank rate with $C = 0$ and $P, Q \leq K\sqrt{n \log n}$.

So memory and purity trade at a fixed rate: for $n \geq n_0$ and $\log n \leq B \leq c_0 n$ the least memory of a service with purity at most B lies between $c\,n/(B(\log n)^{2\kappa})$ and $K(n/B)\log(2n/B)$, and the product of memory and purity is $n^{1+o(1)}$. At the rank rate the least memory is $n^{1/2+o(1)}$, and it is attained with purity of the same order: the rank rate sits where memory and purity are equal.

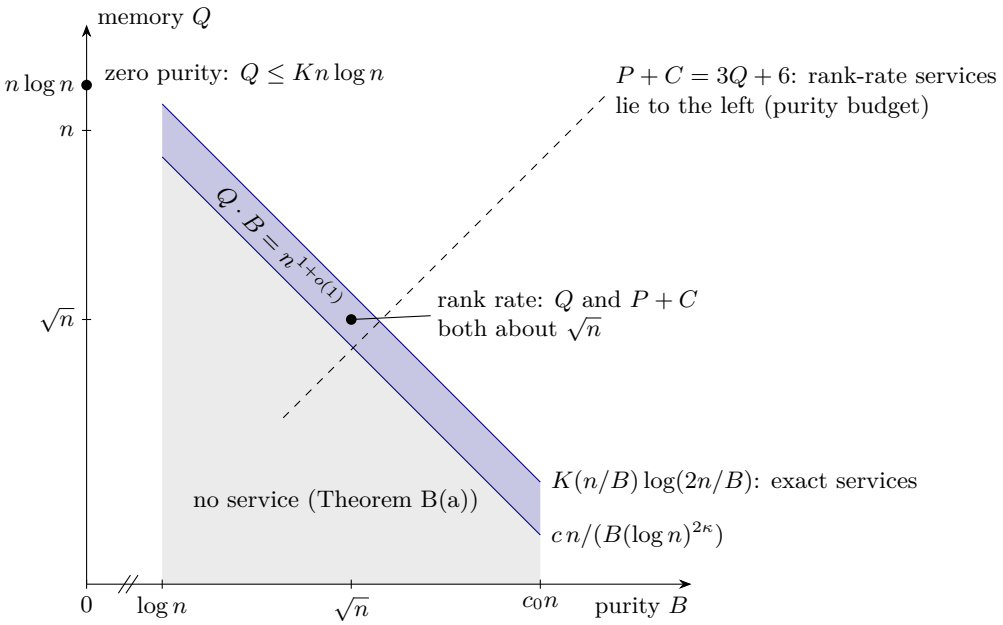


Figure 4: Memory against purity for the lamp channel, on logarithmic axes. For $\log n \le B \le c_0 n$ the least memory of a service with purity at most B lies in the band $Q \cdot B = n^{1+o(1)}$, between $c\,n/(B(\log n)^{2\kappa})$ and $K(n/B)\log(2n/B)$ (Theorem 11.1(a)), and at zero purity memory $Kn \log n$ suffices. At the rank rate the purity is less than $3Q + 6$ (Proposition 13.1), so the band forces memory about $\sqrt{n}$, and streaming services attain this with purity of the same order (Theorem 11.1(b)).

The exponent 2κ in (a) is the unitary exponent of Theorem 3.3(c): a service with purity B is read at relator defect about $(B/n)^{1/2}$. The lower bound in (b) needs no assumption on purity, because at the rank rate purity is paid for with memory (Proposition 13.1): there $P + C < 3Q + 6$, and the argument of (a) with B of order Q gives $Q^2 \gtrsim n$ up to logarithms. The services in (a) and (b) are the two devices of Section 12. We prove Theorem 11.1 in Section 14, after the purity budget and the upper bound, and read it as a statement about finite tracial approximations in Corollary 14.1.

12 Sofic models give devices, and streaming at the rank rate

12.1 An upper bound from sofic models

Sofic models of the labels give devices, for two reasons. The triangle blocks are exactly unitary for arbitrary permutations. And at most points of a model the pure bath vector reproduces the target channel exactly, not approximately, so a memory that is maximally mixed on the good points of a model serves the channel exactly. This gives every upper bound in Theorem B. Recall from Section 2 that $\sigma_E(r)$ is the least M for which some $f : E \to \mathrm{Sym}(\Omega)$, $|\Omega| = M$, with $f(1) = \mathrm{id}$, is a $(1 - r^{-1})$-*expansive* r^{-1}-*morphism*. This means that $f(x)f(y)$ and $f(xy)$ differ on at most M/r points whenever $x, y, xy \in E$, and that $f(x)$ and $f(y)$ agree on at most M/r points whenever $x \ne y$ in E.

Proposition 12.1 (Sofic models give devices)**.** *Let G, S and R be as in Section 9, with $t_\triangle$ triangles and r_* labels, let E be a chunk of G containing all labels, and put $K_G = t_\triangle + \binom{r_*}{2}$. Every $(1 - r^{-1})$-expansive r^{-1}-morphism $f : E \to \mathrm{Sym}(\Omega)$, $|\Omega| = M$, gives a unitary U_f on*

$\mathbb{C}^d \otimes \mathbb{C}^\Omega$ *whose tracial channel* $\Phi_f(X) = (\mathrm{id} \otimes \tau_M)[U_f(X \otimes I)U_f^*]$ *satisfies*

$$\Phi_f = (1-\vartheta)\Phi_G + \vartheta\,\Xi$$

for some channel Ξ *and some* $0 \le \vartheta \le K_G/r$. *Consequently, for all* $n \ge 1$ *and* $0 < \epsilon \le 1$ *with* $\sigma_E(2K_Gn/\epsilon) < \infty$, Φ_G *has*

(a) a causal service with error at most ϵ, $P = C = 0$, $J = (n-1)Q$ *and* $Q = \lceil \log \sigma_E(2K_Gn/\epsilon) + \log(2n/\epsilon) \rceil$;

(b) a causal service with error at most ϵ, $J = P = C = 0$ *and* $Q = \lceil n \log \sigma_E(2K_Gn/\epsilon) + \log(2/\epsilon) \rceil$;

and, for all $n \ge 1$ *and* $0 < B \le n$ *with* $\sigma_E(4K_Gn/B) < \infty$,

(c) an exact causal service, with error 0, *purity* $P + C < B$, $J = (n-1)Q$ *and* $Q = \lceil \log \sigma_E(4K_Gn/B) + \log(4n/B) \rceil$.

Proof. The permutation gadget is unitary. For a symbol w let P_w be the permutation matrix of $f([w])$ on $\mathbb{C}^\Omega$, so that symbols with the same label get the same matrix, and let U_f be the gadget U_λ with $\lambda(w) = P_w$. Each triangle block is $\frac{1}{\sqrt{2}}\left(\begin{smallmatrix} I & P_y \\ P_x & -P_xP_y \end{smallmatrix}\right) = \mathrm{diag}(I, P_x)(H \otimes I)\mathrm{diag}(I, P_y)$, so U_f is unitary although no relation holds among the P_w. Writing π_a for the permutation of Ω in slot a, namely $f([w_a])$ for every slot except the lower right corner of a triangle (x, y, z), where it is $f([x])f([y])$, we have $U_f = \sum_a c_a E_a \otimes P_{\pi_a}$.

Good points reproduce Φ_G. Since τ_M is the average of the vector states of the points of Ω, $\Phi_f = \frac{1}{M}\sum_{\omega\in\Omega} \Phi_\omega$ with $\Phi_\omega(X) = \mathrm{Tr}_\Omega\, U_f(X \otimes |\omega\rangle\langle\omega|)U_f^*$. The isometry $|\psi\rangle \mapsto U_f(|\psi\rangle \otimes |\omega\rangle)$ is $\sum_a c_a E_a \otimes |\pi_a(\omega)\rangle$, so

$$\Phi_\omega(X) = \sum_{a,b\in\mathcal{T}} c_a\overline{c_b}\ \mathbf{1}[\pi_a(\omega) = \pi_b(\omega)]\ E_aXE_b^*,$$
$$\Phi_G(X) = \sum_{a,b\in\mathcal{T}} c_a\overline{c_b}\ \mathbf{1}[[w_a] = [w_b]]\ E_aXE_b^*,$$

the second because $\tau(\lambda(g)\lambda(h)^*) = \delta_{gh}$. Call ω *good* if $f([x])f([y])(\omega) = f([z])(\omega)$ for every triangle (x, y, z) and $f(g)(\omega) \ne f(h)(\omega)$ for all distinct labels g, h. At a good point $\pi_a(\omega) = f([w_a])(\omega)$ for every slot, since $[w_a] = [z]$ at the lower right corner of a triangle (x, y, z); so $\pi_a(\omega) = \pi_b(\omega)$ exactly when $[w_a] = [w_b]$, and $\Phi_\omega = \Phi_G$.

Bad points are few. The labels of a triangle satisfy $[x][y] = [z]$ with all three in E, so $f([x])f([y])$ and $f([z])$ differ on at most M/r points; and distinct labels $g, h \in E$ have $f(g)(\omega) = f(h)(\omega)$ for at most M/r points. So the fraction ϑ of bad points is at most K_G/r, and $\Phi_f = (1-\vartheta)\Phi_G + \vartheta\,\Xi$, where Ξ is the average of Φ_ω over the bad points. If there are no bad points, Ξ may be any channel.

Refreshing the memory gives part (a). Let $r = 2K_Gn/\epsilon$, let f have degree $M = \sigma_E(r)$, and let Q be as in part (a), so that $M \le 2^Q\epsilon/(2n)$. Split the Q-qubit memory as $(\mathbb{C}^\Omega \otimes \mathbb{C}^{k_0}) \oplus \mathbb{C}^{g_0}$ with $k_0 = \lfloor 2^Q/M \rfloor$ and $g_0 = 2^Q - k_0M < M$. In every round apply $U_f \otimes I_{k_0}$ on the first summand and the identity on the second. After every round but the last, export all Q memory qubits and import Q fresh qubits in the maximally mixed state; the initial state is maximally mixed as well, so $P = C = 0$. Every round then meets a fresh flat memory, and the observer faces n independent uses of the channel $(1 - g_0/2^Q)\Phi_f + (g_0/2^Q)\mathrm{id}$, which is within $\vartheta + g_0/2^Q \le \epsilon/(2n) + \epsilon/(2n)$ of Φ_G in diamond distance, by convexity. A hybrid argument over the rounds bounds the error by ϵ.

A memory flat on the good points gives part (c). Let $0 < B \le n$, put $\eta = B/(4n) \le \frac{1}{4}$ and $r = K_G/\eta$, let f have degree $M = \sigma_E(r)$, and let Q be as in part (c), so that $M \le \eta 2^Q$.

Let $\Omega_0 \subseteq \Omega$ be the set of good points, so that $|\Omega_0| \geq (1-\eta)M$. Split the memory as $(\mathbb{C}^{\Omega} \otimes \mathbb{C}^{k_0}) \oplus \mathbb{C}^{g_0}$ as in part (a), with $k_0 = \lfloor 2^Q/M \rfloor$ and $g_0 < M \leq \eta 2^Q$, and let β be the maximally mixed state on $\operatorname{span}\{|\omega\rangle : \omega \in \Omega_0\} \otimes \mathbb{C}^{k_0}$. Since $k_0 M = 2^Q - g_0$, its deficit is

$$Q - \log(|\Omega_0| k_0) = \log \frac{2^Q}{2^Q - g_0} + \log \frac{M}{|\Omega_0|} \leq 2 \log \frac{1}{1-\eta} \leq \frac{8\eta}{3 \ln 2} < \frac{B}{n}.$$

Start the memory in β and apply $U_f \otimes I_{k_0}$ on the first summand in every round. After every round but the last, export all Q memory qubits and import Q fresh qubits in the state β. The n registers used have total deficit $P + C < B$. In every round the memory is a fresh copy of β, a uniform mixture of vector states $|\omega\rangle \otimes |j\rangle$ with ω good, and $\Phi_\omega = \Phi_G$ at good points. So every round applies exactly Φ_G, independently of the other rounds, and the error is 0.

One factor per round gives part (b). Let f be the model of part (a) and Q as in part (b), so that $M^n \leq 2^Q \epsilon/2$. Split the memory as $(\mathbb{C}^{\Omega^n} \otimes \mathbb{C}^{k_0}) \oplus \mathbb{C}^{g_0}$ with $k_0 = \lfloor 2^Q/M^n \rfloor$ and $g_0 < M^n$, start it maximally mixed, and in round s apply U_f to the input and the s-th factor $\mathbb{C}^{\Omega}$ of the first summand, and the identity on the second. With probability $1 - g_0/2^Q \geq 1 - \epsilon/2$ the memory starts in the first summand, uniformly distributed, and then the n factors are independent flat baths, each used once. The observer then faces n independent uses of Φ_f, which is within $n\vartheta \leq \epsilon/2$ of the ideal experiment. The remaining branch has weight at most $\epsilon/2$, so the error is at most ϵ. □

In particular, if $\sigma_E(r)$ is finite for every r, as it is when G is sofic, then $\Phi_G \in \mathcal{K}_d$, since $d_\diamond(\Phi_f, \Phi_G) \leq K_G/r$ for models at every accuracy. This is a quantitative form, for sofic groups, of the fact that gadget channels of hyperlinear groups lie in $\mathcal{K}_d$ [27].

12.2 Streaming at the rank rate

The services of Proposition 12.1 refresh the whole memory in every round, so their exchange is far above the rank rate. At the rank rate the memory has to be kept, and a sofic model cannot simply be reused, since its bad points would then accumulate error. Theorem 12.2 keeps one sofic model for all rounds and pays for its bad points with purity. It uses the good points of the model as an exact *rectangular* dilation, whose input bath is smaller than its output bath, and lets the part of the memory in use grow slightly in every round. The mixing and export steps are those of the rolling reservoir of [27, Appendix A], and the proof is in Appendix D.

Theorem 12.2 (Streaming at the rank rate)**.** *Let Φ be a channel on M_d with linearly independent Kraus operators $A_1, \dots, A_r$, $r \geq 2$, and let $1 \leq k \leq k'$. Suppose that $W = \sum_i A_i \otimes B_i$, with $B_i : \mathbb{C}^k \to \mathbb{C}^{k'}$, is an isometry from $\mathbb{C}^d \otimes \mathbb{C}^k$ to $\mathbb{C}^d \otimes \mathbb{C}^{k'}$ and that*

$$\operatorname{Tr}(B_i^* B_j) = k\,\delta_{ij}, \qquad \sum_i B_i^* B_i = r\, I_k. \tag{10}$$

Then for all $n \geq 2$ and $0 < \epsilon < 1$ there is a causal service of Φ for n rounds with error at most ϵ, $C = 0$, $J = \lfloor \frac{n-1}{2} \log r \rfloor$ and

$$P \leq n \log \frac{k'}{k} + 2, \qquad Q \leq \log k + n \log \frac{k'}{k} + O_r\Big(\log \frac{n}{\epsilon} + \log(2 + \log k)\Big).$$

The first identity in (10) says that W with a flat input bath realizes Φ exactly: $\Phi(X) = \operatorname{Tr}_{k'}[W(X \otimes I_k/k)W^*] = \sum_i A_i X A_i^*$. The second says that every Kraus history costs the same. For a channel with a flat probe of rank r it follows from $W^*W = I$ alone: apply $\operatorname{Tr}(\rho_* \cdot)$ to the first factor of $\sum_{i,j} A_i^* A_j \otimes B_i^* B_j = I$, where ρ_* is the input marginal of the probe, for which $\operatorname{Tr}(\rho_* A_i^* A_j) = \delta_{ij}/r$.

Corollary 12.3 (The lamp channel at the rank rate)**.** *There is K such that for every $n \geq 2$ the channel Φ of Theorem 11.1 has a causal service for n rounds with error at most $1/16$, $C = 0$, $J = \lfloor \frac{n-1}{2} \log r_* \rfloor$ and $P, Q \leq K\sqrt{n \log n}$.*

Proof. Let E be the chunk of labels of Proposition 12.1, let $s \geq 2K_G$, and let $f : E \to \mathrm{Sym}(\Omega)$ be a $(1 - s^{-1})$-expansive s^{-1}-morphism of degree $M = \sigma_E(s)$. Let Ω_0 be its set of good points, so that $k = |\Omega_0| \geq (1 - K_G/s)M$, and for each label g let $B_g : \mathbb{C}^{\Omega_0} \to \mathbb{C}^{\Omega}$ send $|\omega\rangle$ to $|f(g)\omega\rangle$. At a good point ω every slot of U_f acts through the permutation of its label, so $U_f(\xi \otimes |\omega\rangle) = \sum_g A_g \xi \otimes |f(g)\omega\rangle$, and $W = \sum_g A_g \otimes B_g$ is the restriction of the unitary U_f to $\mathbb{C}^d \otimes \mathbb{C}^{\Omega_0}$, an isometry. Each $B_g^* B_g$ is I_k, and $\mathrm{Tr}(B_g^* B_h) = 0$ for $g \neq h$ because distinct labels do not collide at good points; the A_g are linearly independent, having disjoint supports. So Theorem 12.2 applies with $r = r_*$ and $k' = M$. Here $\log(M/k) \leq -\log(1 - K_G/s) \leq 2K_G/(s \ln 2)$, and $\log k \leq \log \sigma_E(s) \leq C_E s \log s$ by Theorem 3.3(b). With $\epsilon = 1/16$ the theorem gives $Q \leq C_E s \log s + 2K_G n/(s \ln 2) + O(\log n)$ and $P \leq 2K_G n/(s \ln 2) + 2$; take $s = \max(2K_G, \lceil \sqrt{n/\log n} \rceil)$. □

13 A purity budget at the rank rate

At the rank rate a device cannot buy memory with purity: the purity it can spend is linear in its memory. This is the last tool for the lower bound at the rank rate in Theorem B. We sharpen the budget $14Q + 3$ of [27] by combining its rank count with a one-sided second-moment bound.

Proposition 13.1 (Purity budget)**.** *Let Φ be a channel on M_d, and suppose that some pure state ψ_* of the input and a reference has output $\omega = (\Phi \otimes \mathrm{id})(\psi_*)$ flat of rank r. Every causal service of Φ for n rounds with error $\epsilon \leq 1/16$ and exchange $J \leq \frac{n}{2} \log r$ satisfies*

$$P + C \;<\; 3Q + 6.$$

Every gadget channel has such a probe with $r = r_*$ (Section 9). The proof uses the variance of the surprisal of a state with small deficit, which is [27, Lemma 5.9]; we include the short argument.

Lemma 13.2 (Surprisal variance)**.** *Let β be a state on $q \geq 1$ qubits with spectrum (β_i) and deficit $D = q - S(\beta)$. The surprisal $-\log \beta_i$, drawn with probability β_i, has variance at most $(q + 2/\ln 2)\, D$.*

Proof. Put $\varphi(x) = x \ln x - x + 1 \geq 0$. We claim $x(\ln x)^2 \leq (2 + \ln_+ x)\varphi(x)$ for $x \geq 0$. For $x = e^t$ with $t \geq 0$ the difference of the two sides is $f(t) = e^t(t-2) + t + 2$, with $f(0) = f'(0) = 0$ and $f''(t) = te^t \geq 0$. For $t \leq 0$ it is $g(t) = 2(te^t - e^t + 1) - t^2 e^t$, with $g(0) = 0$ and $g'(t) = -t^2 e^t \leq 0$; and $x = 0$ is clear. Now let $Q_0 = 2^q$ and $x_i = Q_0 \beta_i \leq Q_0$, the index running over all Q_0 eigenvalues. Then $\frac{1}{Q_0} \sum_i x_i = 1$ and $D \ln 2 = \sum_i \beta_i \ln(Q_0 \beta_i) = \frac{1}{Q_0} \sum_i \varphi(x_i)$. The variance is at most

$$\sum_i \beta_i \big(\log(Q_0 \beta_i)\big)^2 = \frac{1}{Q_0 (\ln 2)^2} \sum_i x_i (\ln x_i)^2 \leq \frac{2 + q \ln 2}{(\ln 2)^2}\, D \ln 2 = \Big(q + \frac{2}{\ln 2}\Big) D.$$

□

Proof of Proposition 13.1. Put $b = P + C$ and suppose, for a contradiction, that $b \geq 3Q + 6$.

Preloading. The imports are fixed states, independent of everything else, and the exported qubits are never used again. So the observer's final state does not change if every import is prepared at the start in a register of its own, and each replacement swaps it with the exported

qubits, which then stay in the device. The device becomes a register of $L = Q + J$ qubits in the product state $\sigma_0 = \beta_0 \otimes \bigotimes_f \omega_f$ of the initial memory state and the imported states, and it applies a fixed unitary in each round.

The surprisal. Let (σ_i) be the spectrum of σ_0 and let $Z = L + \log \sigma_i$, with i drawn with probability σ_i. Then $\mathbb{E}Z = L - S(\sigma_0) = b$. Also Z is the sum of the independent variables $Z_f = q_f + \log \beta_i^{(f)}$ attached to the factors of σ_0, where q_f is the number of qubits of the factor and $\beta^{(f)}$ its spectrum. Every factor has at most Q qubits, since an import replaces exported memory qubits, and the deficits $b_f = \mathbb{E}Z_f$ of the factors add up to b. By Lemma 13.2,

$$\operatorname{Var} Z = \sum_f \operatorname{Var} Z_f \le \Big(Q + \frac{2}{\ln 2}\Big) b.$$

A tail bound from the rank. Let the observer feed the input halves of n copies of ψ_* and keep the references. Its final state is $\tau_O = \sum_i \sigma_i \tau_i$, where τ_i is its final state when the register starts in the eigenvector of σ_i. Each τ_i is a marginal of a pure state of the observer's system and the L-qubit register, so $\operatorname{rank} \tau_i \le 2^L$. Fix a real number k and let I_k be the set of indices with $\sigma_i \ge 2^{k-L}$, so that $|I_k| \le 2^{L-k}$. Let E be the support projection of $\sum_{i \in I_k} \tau_i$. Then $\operatorname{rank} E \le 2^{2L-k}$ and $\operatorname{Tr}(E\tau_O) \ge \sum_{i \in I_k} \sigma_i = \Pr(Z \ge k)$. The ideal final state $\omega^{\otimes n}$ is flat of rank r^n, so $\operatorname{Tr}(E\omega^{\otimes n}) \le 2^{2L-k} r^{-n} \le 2^{2Q-k}$, using $2J \le n \log r$. Since the error is at most ϵ,

$$\Pr(Z \ge k) \;\le\; \epsilon + 2^{2Q-k} \qquad (k \in \mathbb{R}). \tag{11}$$

The contradiction. Put $u = b - 2Q - 3$, so that $u \ge b/3 + 1 > 0$. By the one-sided Chebyshev inequality, $\Pr(Z \ge b - u) \ge u^2/(\operatorname{Var} Z + u^2)$, while (11) with $k = b - u = 2Q + 3$ gives $\Pr(Z \ge b - u) \le \frac{1}{16} + \frac{1}{8} = \frac{3}{16}$. Hence $13u^2 \le 3 \operatorname{Var} Z$, and with $u > b/3$,

$$\frac{13}{9} b^2 < 3\Big(Q + \frac{2}{\ln 2}\Big) b, \qquad \text{that is} \qquad b < \frac{27}{13}\Big(Q + \frac{2}{\ln 2}\Big) < 2.08\, Q + 5.993,$$

which contradicts $b \ge 3Q + 6$. □

14 Proof of Theorem B

Theorem B, in the form of Theorem 11.1, is assembled from sequential extraction, the weighted one-round theorem, the purity budget and the sofic-model devices. Two remarks after the proof say what it shows.

Sofic models give the closure and the services. By Theorem 3.3(b) the chunk E of Γ that consists of 1 and the labels of the gadget of R_Φ has $\log \sigma_E(r) \le C_E r \log r$. Proposition 12.1 therefore gives finite tracial channels within K_G/r of Φ for every r, so $\Phi \in \mathcal{K}_{873}$. This also follows from [27], since Γ is amenable (Lemma 3.1) and hence hyperlinear. The exact services of part (a) are those of Proposition 12.1(c), and the zero-purity services are those of Proposition 12.1(a) with $\epsilon = 1/16$. Their memory bounds follow from $\log \sigma_E(R) \le C_E R \log R$ for $R \ge 2$ and from $\log(4K_G n/B) \le (1 + \log 2K_G) \log(2n/B)$, which holds because $\log(2n/B) \ge 1$. Corollary 12.3 gives the services of part (b). The bounds on r_* were proved in Section 11.

Part (a). Let a service have error $\epsilon \le 1/16$ and $P + C \le B$, where $\log n \le B$. Theorem 9.1 gives a round with bath dimension 2^Q, $d_\diamond(\Psi_t, \Phi) \le \epsilon/(1-\epsilon) \le 1/15$ and $\nu_t = \ell_t + \Delta S_t \le t_n$. Since $\epsilon + h_2(\epsilon) < 1$, $\epsilon \log n \le \log n$ and $1 \le \log n \le B$,

$$t_n \le \frac{16}{15} \cdot \frac{B + 1 + \log n}{n} \le \frac{16B}{5n} \qquad (n \ge 2).$$

Let $C_R \geq 1$ be the constant of Corollary 10.2(a) for R_Φ and the transport of Section 11, and put $y_n = 5n/(16C_RB)$, so that $C_R\nu_t y_n \leq 1$ and $y_n \leq n$. Put $c_0 = 5/(16C_Ry_0)$. If $B \leq c_0 n$, then $y_n \geq y_0$, and the corollary gives

$$Q \;\geq\; S(\rho_t) \;\geq\; \frac{y_n}{32(\log y_n)^{2\kappa}} \;\geq\; \frac{5}{512\,C_R}\cdot\frac{n}{B(\log n)^{2\kappa}}.$$

For the second statement take $B = \max(1, A_0)\log n$, which lies between $\log n$ and $c_0 n$ for large n.

Part (b). Proposition 13.1, with the flat probe of rank r_*, gives $P + C \leq B := 3Q + 6$. Theorem 9.1 now gives a round with $\nu_t \leq t_n \leq \frac{16}{15}(3Q{+}7{+}\log n)/n$. Put $y_n = 15n/(16C_R(3Q{+}7 + \log n)) \leq n$. If $y_n < y_0$, then $3Q + 7 + \log n > 15n/(16C_Ry_0)$, so $Q \geq c'n$ for large n. Otherwise Corollary 10.2(a) gives $Q \geq y_n/(32(\log n)^{2\kappa})$, that is,

$$Q\,(3Q + 7 + \log n) \;\geq\; \frac{15}{512\,C_R}\cdot\frac{n}{(\log n)^{2\kappa}}.$$

If $Q \leq \log n + 7$ the left side is at most $4(\log n + 7)^2$, which is smaller than the right side for large n. So $Q > \log n + 7$, the left side is at most $4Q^2$, and $Q \geq c\sqrt{n}/(\log n)^\kappa$. □

Finite tracial approximations. The same argument has a static reading. A finite tracial factorization of bath dimension M is a round with the flat bath state τ_M: its entropy production vanishes, and its support leakage is at most its Choi distance to Φ. So Theorem 10.1 also reads the static approximation profile of Φ, and the hash models bound it from above. For channels Ψ, Ψ' on M_d let $d_J(\Psi, \Psi') = \frac{1}{2}\|\mathcal{J}(\Psi) - \mathcal{J}(\Psi')\|_1$ be the distance of their normalized Choi states, so that $d_J \leq d_\diamond \leq d\,d_J$ [27, Section 2]. For $u \geq 0$ let $D_\Phi(u)$ be the least M for which some channel Ψ with a finite tracial factorization of bath dimension M has $d_J(\Psi, \Phi) \leq u$; this is the Choi profile of [27, Definition 2.1]. Let $D^\diamond_\Phi(u)$ be the same with $d_\diamond$ in place of d_J, and $D^{\mathrm{mix}}_\Phi(u)$ the same with Ψ a finite convex combination of channels with finite tracial factorizations of bath dimensions at most M. Each is $+\infty$ if no such Ψ exists, each is nonincreasing in u, and $D^{\mathrm{mix}}_\Phi \leq D_\Phi \leq D^\diamond_\Phi \leq D_\Phi(\,\cdot\,/d)$.

Corollary 14.1 (Finite tracial approximations of the lamp channel)**.** *Let Φ be the channel of Theorem 11.1 on M_{873}, let $\kappa = 6 + \log 107$, let E be the chunk of Γ consisting of 1 and the labels of Φ, and let $K_G = 298 + \binom{r_*}{2} = 8683$ be its constant in Proposition 12.1. There are constants $c, C > 0$ and $u_0 > 0$ such that for $0 < u \leq u_0$*

$$c\,\frac{u^{-1}}{(\log(1/u))^{2\kappa}} \leq \log D^{\mathrm{mix}}_\Phi(u) \leq \log D_\Phi(u) \leq \log D^\diamond_\Phi(u) \leq \log \sigma_E(K_G/u) \leq C\,u^{-1}\log(1/u),$$

where $2\kappa < 25.49$. So the least bath dimension of a finite tracial approximation of Φ to accuracy u, in Choi or in diamond distance, is $\exp(u^{-1+o(1)})$, and $D^{\mathrm{mix}}_\Phi(0) = \infty$: Φ is a limit of channels with finite tracial factorizations but is not a finite convex combination of such channels, although it has a Stinespring dilation with a pure environment of dimension $r_* = 130$.

Proof. Upper bound. Let $0 < u \leq \frac{1}{2}$ and $r = K_G/u \geq 2$. Since $\sigma_E(r)$ is finite by Theorem 3.3(b), there is a $(1 - r^{-1})$-expansive r^{-1}-morphism $f : E \to \mathrm{Sym}(\Omega)$ with $|\Omega| = \sigma_E(r)$. Proposition 12.1 gives a channel $\Phi_f = (1 - \vartheta)\Phi + \vartheta\,\Xi$ with a finite tracial factorization of bath dimension $\sigma_E(r)$ and $0 \leq \vartheta \leq K_G/r = u$, so $d_\diamond(\Phi_f, \Phi) = \vartheta\, d_\diamond(\Xi, \Phi) \leq u$ and $D^\diamond_\Phi(u) \leq \sigma_E(K_G/u)$. By Theorem 3.3(b), $\log \sigma_E(r) \leq C_E r \log r$, and $\log(K_G/u) \leq (1{+}\log K_G)\log(1/u)$ since $\log(1/u) \geq 1$; so $C = C_E K_G(1 + \log K_G)$ serves. The inequalities between the three profiles hold because $d_J \leq d_\diamond$ and a single factorization is a convex combination with one term.

A mixture is a flat round. Let $\Psi = \sum_{i=1}^{k} p_i \Psi_i$ with $d_J(\Psi, \Phi) \le u$, where $\Psi_i(X) = (\mathrm{id} \otimes \tau_{m_i})[U_i(X \otimes I)U_i^*]$ and $m_i \le M$. By Carathéodory's theorem we may take $k \le d^4$, since the channels on M_d lie in a real affine space of dimension less than d^4. On $\mathbb{C}^d \otimes \bigoplus_i \mathbb{C}^{m_i}$ put $U = \bigoplus_i U_i$ and $\rho = \bigoplus_i p_i \tau_{m_i}$. Then the partial trace of $U(X \otimes \rho)U^*$ over $\bigoplus_i \mathbb{C}^{m_i}$ is $\Psi(X)$, so Ψ is a round in the sense of Section 9, with $S(\rho) = H(p) + \sum_i p_i \log m_i \le 4 \log d + \log M$. Its entropy production vanishes: $T(\rho) = \bigoplus_i p_i \operatorname{Tr}_d U_i(\tau_d \otimes \tau_{m_i})U_i^* = \rho$, because $\tau_d \otimes \tau_{m_i}$ is the normalized identity and U_i is unitary. Its support leakage is $\ell = \operatorname{Tr}[(I - \Pi_\Phi)(\mathcal{J}(\Psi) - \mathcal{J}(\Phi))] \le \frac{1}{2}\|\mathcal{J}(\Psi) - \mathcal{J}(\Phi)\|_1 \le u$, because $\mathcal{J}(\Phi)$ is supported in Π_Φ and $\operatorname{Tr}[PX] \le \frac{1}{2}\|X\|_1$ for a projection P and a traceless Hermitian X. So $\nu = \ell + \Delta S \le u$, and $d_\diamond(\Psi, \Phi) \le d\,u \le \frac{1}{8}$ when $u \le 1/(8d)$: the round satisfies the hypotheses of Theorem 10.1.

Lower bound. Section 11 verifies the hypothesis of Corollary 10.2(a) for R_Φ with the present κ and $K = C_1 + 64$. Let $C_R = 2^{50} m^4 d K^2 \ge 1$ and y_0 be the constants of that corollary, and put $y = 1/(C_R u)$. Then $C_R \nu y \le 1$, and $y \ge y_0$ when $u \le 1/(C_R y_0)$. For $u \le \min\{1/(8d), 1/(C_R y_0)\}$ the corollary gives, since $1 < y \le 1/u$,

$$4 \log d + \log M \;\ge\; S(\rho) \;\ge\; \frac{y}{32(\log y)^{2\kappa}} \;\ge\; \frac{1}{32 C_R} \cdot \frac{u^{-1}}{(\log(1/u))^{2\kappa}}.$$

Let u_0 be so small that the right side is at least $8 \log d$ for $u \le u_0$. Then $\log M \ge c\, u^{-1} (\log(1/u))^{-2\kappa}$ with $c = 1/(64 C_R)$, and with $c = 1/(32 C_R)$ for a single factorization, where $H(p) = 0$. Finally $D_\Phi^{\mathrm{mix}}(0) \ge D_\Phi^{\mathrm{mix}}(u)$ for every $u > 0$, and the lower bound is unbounded as $u \to 0$. The pure environment is the Stinespring dilation of Remark C.8. □

By a theorem of Haagerup and Musat [39, Theorem 3.6], a channel on M_d lies in $\mathcal{K}_d$ exactly when it factors through an ultrapower of the hyperfinite II_1 factor, and Musat and Rørdam showed that such a channel may have no finite tracial factorization at all [60]. The Houghton channel of [27, Theorem D.1] is an explicit channel of this kind, with profile exponent between 0.069 and 1/3. For Φ the exponent is $1 - o(1)$, nearly the most that the one-round method can certify (Remark 14.3), while channels in the relative interior of $\mathcal{K}_d$ have profiles at most linear in $1/u$ [27, Section 10].

Remark 14.2 (Memory between two readings of the profile). Take a gadget channel of a sofic group that satisfies the hypotheses of Theorem 10.1. The least memory of its services with error 1/16 and purity $O(\log n)$ lies between two readings of the group at scale n. The lower reading is the entropy forced by as many cheaply commuting copies of S_3 as Corollary 10.2 fits at relator defect about $\sqrt{\log n / n}$. The upper reading is $\log \sigma_E(Cn) + O(\log n)$, by Proposition 12.1; it is attained already at zero purity, with exchange $(n-1)Q$. For the configuration lamps both readings are $n^{1+o(1)}$. The upper reading is the sofic profile, not the Følner function. The latter is doubly exponential for Γ (Proposition 7.3), so models built from Følner sets [25, Proposition 3.14] give memory exponential in n, while the hash models behind Theorem 3.3(b) give $O(n \log n)$. The grid host gives a second pair of readings, conditionally on the sketches of Appendix H (Remark H.2).

Remark 14.3 (The rank rate is the diagonal of the trade-off). At the rank rate a device with error at most 1/16 has purity less than $3Q + 6$ (Proposition 13.1). Theorem 11.1(a) with B of order Q therefore forces $Q \gtrsim \sqrt{n}$, and the streaming services attain this with purity of the same order. In general, for every channel with a flat probe, padding a rank-rate service with error at most 1/16 by $2Q + 6$ maximally mixed qubits turns it into a service whose purity is at most its memory. The least memory at purity at most memory is therefore at most three times the least memory at the rank rate, plus six. For Φ both are $n^{1/2+o(1)}$. The one-round method cannot give more than $\sqrt{n}$ at the rank rate, since condition (9) certifies at most about $1/\nu$ copies; Theorem 12.2 shows that for Φ nothing more is true.

Part III
Physics and open problems

15 Local physics

Theorem B concerns an abstract channel. This section asks which physical systems could produce the lamp channel Φ, and what a system left to itself can produce in the long run. The answers are given at the scale of Theorem B. A channel is *cheap* if for all n and ϵ it has causal services for n rounds with error at most ϵ, $P = C = 0$ and memory polynomial in $\log(n/\epsilon)$. It is *expensive* if for some $A, c > 0$ every service for n rounds with error at most $1/16$ and $P + C \le A \log(n+1)$ needs memory at least n^c for all large n. Neither notion constrains the exchange.

Local dynamics with a maximally mixed bath are cheap at every finite time, so they need time about $\delta^{-1/\nu}$ to come within δ of Φ (Section 15.1). An autonomous dynamics with a tracial bath has only self-adjoint, positive channels as long-time limits (Section 15.2). Those two properties do not force cheapness: a lazy doubling of Φ has both, is the exact value of an autonomous dynamics at every time $t \ge 1$, and is still expensive (Section 15.3). Finer constraints on autonomous limits, which exclude some exact finite factors, are in Appendix F.

15.1 Local reservoirs are cheap at every finite time

Local reservoirs. A *local reservoir* is a finite box of the lattice $\mathbb{Z}^\nu$ with a system of dimension at most q at each site, the probe $\mathbb{C}^d$ attached to the origin, and a Hamiltonian $H(t) = \sum_X h_X(t)$, possibly time dependent. Its terms are supported in sets X of diameter at most R_0 and satisfy $\|h_X(t)\| \le J_0$. The reservoir starts in its tracial state, and the probe is coupled to it for a time T. The resulting channel on the probe is $\Phi_T(X) = (\mathrm{id} \otimes \tau)[W_T(X \otimes I)W_T^*]$, where W_T is the time-ordered evolution and τ the normalized trace of the reservoir. We call ν, q, R_0, J_0 and d the *parameters* of the reservoir; the size of the box is arbitrary.

Proposition 15.1 (Local reservoirs are cheap at every finite time)**.** *There is K, depending only on the parameters, such that for every local reservoir, every $T \ge 0$, every $n \ge 1$ and every $0 < \epsilon \le 1$, the channel Φ_T has a causal service for n rounds with error at most ϵ, $P = C = 0$, $J = (n-1)Q$ and*

$$Q \;\le\; K\,[1 + T + \log(n/\epsilon)]^{\nu}.$$

Proof. For $r \ge 0$ let $\Phi_T^{(r)}$ be the channel produced by the terms h_X with X inside the ball B_r of radius r about the origin. The Lieb–Robinson bound [56, 9, 61] gives constants $C_1, v, \mu > 0$, depending only on the parameters, such that the Heisenberg evolutions of an operator A at the origin under the full and the truncated dynamics differ in norm by at most $C_1\|A\|e^{vT-\mu r}$. The bound does not depend on the dimension of the site at the origin, so it holds for operators on the probe tensored with a reference system that the dynamics does not touch. Hence the Heisenberg-picture maps of Φ_T and $\Phi_T^{(r)}$ differ in completely bounded norm by at most $C_1 e^{vT-\mu r}$, and by duality $d_\diamond(\Phi_T, \Phi_T^{(r)}) \le C_1 e^{vT-\mu r}$. The truncated evolution acts only on the probe and the sites of B_r, and the other sites start in their tracial state, so $\Phi_T^{(r)}$ has a finite tracial factorization with bath dimension $m \le q^{(2r+1)^\nu}$.

Choose $r = \lceil (vT + \ln(2C_1 n/\epsilon))/\mu \rceil$, so that $d_\diamond(\Phi_T, \Phi_T^{(r)}) \le \epsilon/(2n)$, and let $Q = \lceil \log m + \log(2n/\epsilon) \rceil$. The exchange construction in the proof of Proposition 12.1, with U_f replaced by the unitary of this factorization and $\vartheta = 0$, gives a service of $\Phi_T^{(r)}$ in which every round

is within $\epsilon/(2n)$ of $\Phi_T^{(r)}$, hence within ϵ/n of Φ_T. A hybrid argument bounds the error by ϵ. The memory is $Q \le (2r+1)^\nu \log q + \log(2n/\epsilon) + 1 \le K[1 + T + \log(n/\epsilon)]^\nu$. □

With Theorem B this says that local physics cannot produce the lamp channel quickly.

Corollary 15.2 (Preparation time)**.** *Let Φ be the channel of Theorem 11.1, and fix the parameters of a local reservoir with $d = 873$. There are $c, \delta_0 > 0$ such that, whenever such a reservoir produces at time T a channel within $\delta \le \delta_0$ of Φ in diamond distance,*

$$T \;\ge\; c\,\delta^{-1/\nu}\big(\log(1/\delta)\big)^{-(2\kappa+1)/\nu}, \qquad 2\kappa + 1 < 26.49.$$

Proof. Let $n = \lfloor 1/(32\delta) \rfloor$. Proposition 15.1 gives a service of Φ_T with error at most $1/32$, zero purity and $Q \le K[1 + T + \log(32n)]^\nu$. Replacing the n uses of Φ_T by uses of Φ one at a time changes the observer's final state by at most $n\delta \le 1/32$, so the same device serves Φ with error at most $1/16$ and zero purity. For δ_0 small, n is large, and Theorem 11.1(a) with $A_0 = 0$ gives $Q \ge c' n/(\log n)^{2\kappa+1}$. Hence $1 + T + \log(32n) \ge \big(c' n/(K(\log n)^{2\kappa+1})\big)^{1/\nu}$, and since $n \ge 1/(64\delta)$ this gives the claim for δ_0 small. □

The maximally mixed start of the reservoir is essential: Φ has a Stinespring dilation with a pure environment of dimension r_*, so from a pure state a bounded number of sites produce Φ exactly in bounded time. The same argument applies to a circuit of depth T of gates of bounded range, whose light cone is exact, and so bounds the depth of any local circuit that prepares Φ. The corollary does not exclude that local dynamics produce Φ in the limit of long times; it says how slowly. What the limit of long times can be, for a system left to itself, is the question of the rest of the section.

15.2 Autonomous limits are self-adjoint

Autonomous dynamics. Let $\mathcal{B}$ be a von Neumann algebra with a faithful normal tracial state $\tau_{\mathcal{B}}$. Put $\mathcal{M} = M_d \otimes \mathcal{B}$ with the trace $\tau = \mathrm{tr}_d \otimes \tau_{\mathcal{B}}$, where tr_d is the normalized trace, and let $\iota(a) = a \otimes 1$ and $E = \mathrm{id} \otimes \tau_{\mathcal{B}}$. An *autonomous dynamics* is a trace-preserving automorphism α of $\mathcal{M}$, and the channel after $t \in \mathbb{N}$ steps is $\Phi_t = E \circ \alpha^t \circ \iota$. For $\alpha = \mathrm{Ad}\,U$ with $\mathcal{B} = M_m$ this is the tracial factorization $\Phi_t(X) = (\mathrm{id} \otimes \tau_m)[U^t(X \otimes I)U^{-t}]$. Let $\mathcal{N}$ be the fixed-point algebra of α and $E_{\mathcal{N}}$ the τ-preserving conditional expectation onto it.

The next proposition is standard. It combines the mean ergodic theorem with an observation of Junge, Le Merdy and Xu [49, Remark 10.3], and of Haagerup and Musat [38, Remark 5.2]: compressions $E \circ E_{\mathcal{N}} \circ \iota$ of trace-preserving conditional expectations are self-adjoint and positive.

Proposition 15.3 (Autonomous limits are self-adjoint)**.** *The averages $\frac{1}{T}\sum_{t=0}^{T-1} \Phi_t$ converge in diamond norm to $\overline{\Phi} = E \circ E_{\mathcal{N}} \circ \iota$, and if Φ_t converges, its limit is $\overline{\Phi}$. The channel $\overline{\Phi}$ is self-adjoint and positive semidefinite for the Hilbert–Schmidt inner product on M_d. In particular, a channel Ψ with $\Psi^* \ne \Psi$ is not the long-time limit of any autonomous dynamics with a tracial bath, local or not.*

Proof. The automorphism extends to a unitary V on $L^2(\mathcal{M}, \tau)$. By the mean ergodic theorem [51], $\frac{1}{T}\sum_{t<T} V^t$ converges strongly to the orthogonal projection P onto the invariant vectors. For $x \in \mathcal{M}$ the averages of x lie in the ball of radius $\|x\|$ of $\mathcal{M}$, which is closed in L^2, so $Px \in \mathcal{M}$, and Px is invariant; hence $P(\mathcal{M}) \subseteq \mathcal{N}$. Since $\mathcal{N}$ consists of invariant vectors and $\mathcal{M}$ is dense in $L^2(\mathcal{M})$, P is the orthogonal projection onto $L^2(\mathcal{N})$, whose restriction to $\mathcal{M}$ is $E_{\mathcal{N}}$. The map ι is an isometry from (M_d, tr_d) into $L^2(\mathcal{M}, \tau)$ with adjoint E, so $\Phi_t = \iota^* V^t \iota$, and the averages converge to $\iota^* P \iota = \overline{\Phi}$ on every $a \in M_d$, hence in diamond norm, since M_d is finite dimensional. An ordinary limit equals the limit of the averages. And $\mathrm{tr}_d(b^* \overline{\Phi}(a)) = \langle P\iota(b), P\iota(a) \rangle$ is Hermitian in (a, b) and nonnegative for $a = b$. □

Conversely, every channel of the form $\Psi\Psi^*$ with Ψ in the tracial closure is a compression of a conditional expectation: apply [38, Theorem 5.9] to Ψ^*, which is again factorizable. It is in fact an autonomous long-time limit. The dynamics that produces it adapts Kümmerer's tensor dilations with a Bernoulli shift [53], and nothing in the construction makes it local.

Proposition 15.4 (Autonomy without locality)**.** *Let $\mathcal{N}_0$ be a von Neumann algebra with a faithful normal tracial state τ_0, let U be a unitary in $M_d \otimes \mathcal{N}_0$, and put $\Psi(X) = (\mathrm{id} \otimes \tau_0)[U(X \otimes 1)U^*]$. Let $\mathcal{B} = \overline{\bigotimes}_{k\in\mathbb{Z}}(\mathcal{N}_0, \tau_0)$, with U acting on the factor $k = 0$, and let β be the Bernoulli shift of $\mathcal{B}$. Then the autonomous dynamics $\alpha = \mathrm{Ad}\, U \circ (\mathrm{id} \otimes \beta) \circ \mathrm{Ad}\, U^*$ has $\Phi_t \to \Psi\Psi^*$, where $\Psi^*(a) = (\mathrm{id} \otimes \tau_0)[U^*(a \otimes 1)U]$ is the Hilbert–Schmidt adjoint of Ψ.*

Proof. We have $\alpha^t = \mathrm{Ad}\, U \circ (\mathrm{id} \otimes \beta^t) \circ \mathrm{Ad}\, U^*$. The shift is mixing: $\tau_{\mathcal{B}}(w\, \beta^t(z)) \to \tau_{\mathcal{B}}(w)\tau_{\mathcal{B}}(z)$ for $w, z \in \mathcal{B}$. Equality holds as soon as w and $\beta^t(z)$ are elementary tensors with disjoint supports, and by Kaplansky's density theorem such tensors of norm at most $\|w\|$ and $\|z\|$ approximate w and z in 2-norm. Let $a \in M_d$ and $y = U^*(a\otimes 1)U$, and write $y = \sum_{i,j} e_{ij} \otimes y_{ij}$. For every $x = \sum_{k,l} e_{kl} \otimes x_{kl}$ in $M_d \otimes \mathcal{B}$,

$$\tau\big(x\,(\mathrm{id} \otimes \beta^t)(y)\big) = \frac{1}{d}\sum_{k,l} \tau_{\mathcal{B}}(x_{kl}\beta^t(y_{lk})) \;\longrightarrow\; \tau\big(x\,(E(y) \otimes 1)\big).$$

Taking $x = U^*(e_{ji} \otimes 1)U$ shows that every matrix entry of $\Phi_t(a) = E[U(\mathrm{id} \otimes \beta^t)(y)U^*]$ converges to the corresponding entry of $E[U(E(y) \otimes 1)U^*] = \Psi(\Psi^*(a))$. Convergence on a basis of M_d gives convergence in diamond norm. □

By a theorem of Haagerup and Musat [39, Theorem 3.6], every channel in $\mathcal{K}_d$ has the form Ψ of Proposition 15.4 for some $\mathcal{N}_0$, and the lamp channel has it with $\mathcal{N}_0 = L(\Gamma)$. Nothing in Proposition 15.4 makes the dynamics local. The unitary U is arbitrary in $M_d \otimes \mathcal{N}_0$, and whether a dynamics of this kind can be given by a local rule is open (Section 16).

Applied to the lamp channel, Proposition 15.4 makes $\Phi\Phi^*$ an autonomous limit. That limit is cheap, since $\Phi\Phi^*$ has an exact finite factorization with a bounded bath (Proposition F.3). An expensive limit needs a construction that keeps more of the group, and the next subsection gives one.

15.3 An expensive autonomous limit: the lazy doubling

The limit we build is a self-adjoint, positive and expensive channel on a doubled system, reached in two steps. The first step doubles Φ to a self-adjoint channel Θ that keeps the lower bounds of Theorem B. The second averages Θ with the identity, which makes it positive and leaves it expensive.

Remark 15.5 (Self-adjoint expensive channels). A self-adjoint channel can still be expensive. Let z be the unitary of order three in $L(\mathbb{Z}_3)$, so that $\tau(z) = \tau(z^2) = 0$, let U be the gadget unitary of the channel Φ of Theorem B, and let $V = \left(\begin{smallmatrix} 0 & U\otimes z \\ U^*\otimes z^* & 0 \end{smallmatrix}\right)$, a self-adjoint unitary over $\mathbb{C}^2 \otimes \mathbb{C}^{873}$ and $L(\Gamma \times \mathbb{Z}_3)$. The channel it defines on M_{1746} is $\Theta\left(\begin{smallmatrix} x & y \\ u & w \end{smallmatrix}\right) = \mathrm{diag}(\Phi(w), \Phi^*(x))$, which is self-adjoint. It keeps the lower bounds of Theorem B. The memory-against-purity bound transfers: a service of Θ becomes a service of Φ with one more memory qubit and one more bit of purity, at the same exchange. Feed each input together with a retained tag qubit that starts in $|1\rangle$ and is flipped after every round; this works because Θ maps $|1\rangle\langle 1| \otimes \rho$ to $|0\rangle\langle 0| \otimes \Phi(\rho)$. The bound at the rank rate transfers as well. Θ has a flat probe of rank $2r_*$, with anchors in both blocks, so Proposition 13.1 gives $P + C < 3Q + 6$ at its own rank rate $J \le \frac{n}{2}\log(2r_*)$; the tag transfer and Theorem 11.1(a) then give $Q \ge c\sqrt{n}/(\log n)^{\kappa}$. But Θ maps $\mathrm{diag}(I, -I)$ to $-\mathrm{diag}(I, -I)$, so it is not positive and not an autonomous limit. Its square $\Theta^2 = \mathrm{diag}(\Phi\Phi^*, \Phi^*\Phi)$ is an autonomous limit, by Proposition 15.4, and it is cheap (Proposition F.3).

A lazy doubling. The doubled channel Θ is expensive but not positive, and its square is cheap. Averaging Θ with the identity restores positivity and keeps it expensive. Put

$$\Theta_+ = \tfrac{1}{2}(\mathrm{id} + \Theta), \qquad \Theta_+ \begin{pmatrix} x & y \\ u & w \end{pmatrix} = \frac{1}{2} \begin{pmatrix} x + \Phi(w) & y \\ u & w + \Phi^*(x) \end{pmatrix} \qquad \text{on } M_{1746}.$$

Remark 15.5 uses only $zz^* = 1$ and $\tau(z^2) = 0$, so z may also be the unitary $\mathrm{diag}(1, i)$ of a qubit with its normalized trace. Since Θ is a unital channel that is self-adjoint for the Hilbert–Schmidt inner product, its spectrum lies in $[-1, 1]$; unital channels are contractions for the Hilbert–Schmidt norm. Hence Θ_+ is self-adjoint and positive semidefinite, as Proposition 15.3 requires of an autonomous limit.

Proposition 15.6 (An expensive autonomous limit)**.** *Let Φ, V and Θ be as in Remark 15.5, and $\kappa = 6 + \log 107$.*

(a) Autonomy. *Let Z be the coordinate of $L^\infty[-1, 1]$ with the normalized Lebesgue measure, put $\mathcal{B} = L(\Gamma \times \mathbb{Z}_3) \mathbin{\overline{\otimes}} L^\infty[-1, 1]$ with the product trace, and $W = \exp(i\frac{\pi}{2} V \otimes Z) = \cos(\frac{\pi}{2} Z) + i \sin(\frac{\pi}{2} Z)\, V \in M_{1746} \otimes \mathcal{B}$. The autonomous dynamics $\alpha = \mathrm{Ad}\, W$ has $\Phi_t = \Theta_+$ for every $t \geq 1$. More generally, $\exp(isV \otimes Z)$ gives the channel $\Xi_s = \frac{1+\mathrm{sinc}\, 2s}{2}\, \mathrm{id} + \frac{1-\mathrm{sinc}\, 2s}{2}\, \Theta$, with $\mathrm{sinc}\, x = \sin x / x$, and $d_\diamond(\Xi_s, \Theta_+) = |\sin 2s|/(4s)$.*

(b) Memory against purity, lower bound. *There are $c, c_0 > 0$ and n_0 such that for $n \geq n_0$ and $\log n \leq B \leq c_0 n$ every causal service of Θ_+ for n rounds with error at most $1/16$ and purity $P + C \leq B$ has, at any exchange,*

$$Q \;\geq\; c\, \frac{n}{B\, (\log n)^{2\kappa}}.$$

In particular, for every $A_0 \geq 0$ there is $c' > 0$ such that services with $P + C \leq A_0 \log n$ have $Q \geq c' n/(\log n)^{2\kappa+1}$ for all large n.

(c) Upper bound. *There is K such that for every $n \geq 1$ and $0 < B \leq n$, Θ_+ has an exact causal service with purity $P + C < B$, $J = (n-1)Q$ and $Q \leq K(n/B) \log(2n/B)$. Moreover $\Theta_+ \in \mathcal{K}_{1746}$.*

Thus Θ_+ is the long-time limit, indeed the exact value at every time $t \geq 1$, of an autonomous dynamics with a tracial bath. It is expensive in the sense of this section, and it trades memory and purity at the same rate as Φ (Theorem 11.1(a)). Like the dynamics of Proposition 15.4, the construction is not local. Part (a) is a computation, and we give it here. Appendix E proves parts (b) and (c).

Part (b) loses nothing against Theorem 11.1(a). A service of Θ_+ yields, by Theorem 9.1, one round (U, ρ) on the bath $\mathbb{C}^M$, $M = 2^Q$, with small leakage and entropy production. Write U in 2×2 blocks with respect to the qubit factor of the doubled system. The upper right block U_{01} is where Θ_+ applies Φ. It is nearly a round of Φ on the same bath, except that it is not unitary: in the norm weighted by ρ, $U_{01}^* U_{01}$ is close to $1 \otimes R$ for an operator $0 \leq R \leq 1$ on the bath. Where R is bounded away from 0, the operator $U_{01} R^{-1/2}$ is nearly an isometry. A smooth cutoff restricts to that part of the bath, and a polar decomposition produces a unitary U' on $\mathbb{C}^{873} \otimes \mathbb{C}^M$ and a state ρ', the state ρ reweighted by the cutoff (Lemma E.3). The leakage of the new round and its commutator $\|[U', 1 \otimes \sqrt{\rho'}\,]\|_F^2$ are at most a constant times the leakage and the commutator $\|[U, 1 \otimes \sqrt{\rho}\,]\|_F^2$ of the old round, a linear loss. The one-round theorem accepts the commutator in place of entropy production, because its proof uses entropy production only to bound that commutator. It also accepts, in place of a small diamond distance to Φ, a bound on one coherence between two anchors, which the block extraction gives as well (Lemma E.1). Corollary 10.2 then yields the bound of Theorem 11.1(a) for Θ_+, with the same exponent.

For part (c), two maximally mixed qubits join the bath of a sofic model of the labels: a coin that chooses between the identity and Θ, and a phase qubit carrying $\mathrm{diag}(1, i)$ in place of z. The model then serves Θ_+ exactly at its good points. The exact service of Proposition 12.1(c) then goes through unchanged, and the same unitary with a flat bath puts Θ_+ in $\mathcal{K}_{1746}$.

Proof of (a). Since $V^2 = 1$ and Z commutes with V, $\exp(isV \otimes Z) = \cos(sZ) + i \sin(sZ)V$, so that $W^t = \exp(it\frac{\pi}{2} V \otimes Z)$ and

$$W^t(X \otimes 1)W^{-t} = \cos^2(\tfrac{t\pi}{2} Z)\, X + \sin^2(\tfrac{t\pi}{2} Z)\, V(X \otimes 1)V + \tfrac{i}{2} \sin(t\pi Z)(V(X \otimes 1) - (X \otimes 1)V).$$

The trace of $\mathcal{B}$ is the product of the trace of $L(\Gamma \times \mathbb{Z}_3)$ and the integral over Z. The last term integrates to 0 because $\sin(t\pi Z)$ is odd, $\int \cos^2(\frac{t\pi}{2} Z) = \frac{1}{2}(1 + \operatorname{sinc} t\pi) = \frac{1}{2}$ for $t \geq 1$, and $(\mathrm{id} \otimes \tau)[V(X \otimes 1)V] = \Theta(X)$. So $\Phi_t = \frac{1}{2}(\mathrm{id} + \Theta)$. The same computation with s in place of $\frac{t\pi}{2}$ gives Ξ_s. Finally $\Xi_s - \Theta_+ = \frac{1}{2}\operatorname{sinc}(2s)(\mathrm{id} - \Theta)$, and $\|\mathrm{id} - \Theta\|_\diamond = 2$: the input $|0\rangle\langle 0| \otimes \sigma$ is mapped by Θ into the block $|1\rangle\langle 1|$. The automorphism $\mathrm{Ad}\, W$ is inner, hence trace preserving. □

The same channel is reached, for every $t \geq 1$, by a Bernoulli-shift construction like that of Proposition 15.4. Take $\mathcal{B} = L(\Gamma \times \mathbb{Z}_3) \mathbin{\overline{\otimes}} \overline{\bigotimes}_{k \in \mathbb{Z}} (M_2, \tau_2)$, let β be the shift of the chain, and let p_0 be the projection $|1\rangle\langle 1|$ of the qubit at site 0. The automorphism $\mathrm{Ad}((1-p_0)+p_0 V) \circ (\mathrm{id} \otimes \beta)$ raises V to the parity of t fair bits.

Autonomous dynamics with a tracial bath can therefore produce expensive channels. The reduced dynamics of the probe matters. If the channels Φ_t themselves form a semigroup of unital channels, they are mixed unitary for all large t [7], hence cheap; here $\Theta_+^2 \neq \Theta_+$. Locality constrains how Θ_+ can be approached. Proposition 15.1 and Proposition 15.6(b) combine as in the proof of Corollary 15.2. Together they show that a local reservoir whose probe is the doubled system, and which produces a channel within $\delta \leq \delta_0$ of Θ_+ at time T, has $T \geq c\, \delta^{-1/\nu} (\log(1/\delta))^{-(2\kappa+1)/\nu}$, with constants $c, \delta_0 > 0$ depending only on the parameters of the reservoir. No local reservoir, then, produces Θ_+ exactly at a finite time or approaches it exponentially fast. Whether a local dynamics can produce an expensive channel at all is open (Section 16). Self-adjointness and positivity, which Θ_+ has, are not all that an autonomous limit must satisfy: Appendix F gives a sharper necessary condition and excludes an exact finite factor by it.

16 Open problems

Groups.

(1) *The exact profile.* Theorem A gives $\exp(r^{1+o(1)})$, with polylogarithmic factors in the exponent between $c\, r/(\log r)^\kappa$, $\kappa = 6 + \log 107$, and $C\, r \log r$. The exponent κ comes from the Dehn function of H_3, known to be polynomial of degree at most $\log 107 + 4$ and at least 2 [26]; a quadratic Dehn function would give $\kappa = 4$. What is the correct polylogarithmic factor; is $\log \sigma_E(r) = \Theta(r \log r)$?

(2) *Beyond the counting ceiling.* Does some amenable group have a chunk with profile $\exp(c\, r)$? Does any sofic group have a chunk with profile $\exp(r^{1+c})$ for some $c > 0$? The counting criterion cannot give either; a different lower-bound mechanism would be needed. For the second question, grids of fixed dimension should not help: the argument sketched in Remark H.1 would bound the profile of their chunks by $\exp(O(r \log r))$. By Remark 14.2, a gadget channel of a sofic group can need superlinear memory at zero purity only if its chunk profile exceeds $\exp(O(r))$. Every fixed chunk of a sofic group

has a computable profile [46, Lemma 1] (search the degrees $1, 2, \dots$ and the finitely many maps at each), so no chunk outgrows every computable function. How fast the profile of a chunk of a finitely presented amenable group can grow is open.

(3) *The dyadic group.* Its profile is finite and at least $\exp(c\,r^{1/2})$. In $\mathcal{G}$ the words $w_0 = 1$, $w_{2n} = uw_n$ and $w_{2n+1} = tuw_n$ reach the integer n with length $O(\log n)$, and commutators of copies at points of equal parity reduce freely to smaller indices. What is the area of the commutators at points of opposite parity? More precisely, can the whole suffix sweeps for these binary transports be filled with area $O(N(\log N)^d)$ for some fixed d? Theorem 2.3 would then give $\log \sigma_{E_{\mathcal{G}}}(r) \geq cr/(\log r)^d$, reaching the counting ceiling up to logarithms. This is an open target, not an upper bound on the profile.

(4) *Thompson's groups.* The embedding $\mathcal{L} \leq V$ and the recent result [72, Theorem 1.1] give soficity of $\mathcal{L}$. This does not by itself show that the action of F on $\mathbb{D}$ is sofic in the sense of [30]. Is that action sofic? Does Γ embed in V or in $2V$? The embedding of Proposition 7.2 uses one coordinate for each ray.

Channels and physics.

(5) *The optimal exchange rate in general.* For Φ the least memory at the rank rate and the least memory with purity at most the memory are both $n^{1/2+o(1)}$ (Remark 14.3). For every channel with a flat probe the second is at most three times the first, plus six. Is the converse true up to polylogarithmic factors for every channel in $\mathcal{K}_d$ with a flat probe? Theorem 12.2 bounds the rank-rate memory by $\log k + n \log(k'/k) + O(\log n)$, so the converse holds whenever some exact rectangular dilation makes this at most a polylogarithmic factor times the least memory at purity at most memory.

(6) *Local autonomous physics.* Proposition 15.6 gives an expensive long-time limit of an autonomous dynamics with a tracial bath, but the dynamics is not local. Can an expensive channel be the long-time limit of a fixed translation-invariant local rule, such as a quantum cellular automaton or a time-independent finite-range Hamiltonian, with the bath in its tracial state? By Proposition 15.1 such a limit would be approached slowly; for Θ_+ the distance at time t is at least of order $t^{-\nu}$ up to logarithmic factors (Section 15). By [7] the reduced dynamics of the probe could not be a semigroup.

Use of AI

This work was carried out with extensive use of generative AI systems. They proposed constructions and proof strategies, wrote and checked proofs, carried out computations, searched the literature, wrote the verification code, and drafted and edited the text. The authors directed the research, assessed and selected the results, and take full responsibility for the content. The proofs are given in full in the text; repeated adversarial checking and the accompanying programs support them but do not replace them. The verification notes state which statements are checked by machine. New sessions developed the whole-suffix sweep argument and its applications, and an independent blind check tested the proofs and finite certificates before this revision.

A Tools from Houghton's group

Section 4 uses three facts about $H = H_3$, and this appendix proves them. Lemma A.1 gives short words for finite permutations, Lemma A.2 gives generators of point stabilizers

with linear word length, and Proposition A.3 bounds the Dehn function. A comparison of conventions with [26] comes first, because permutations act on the right there. All transport in Sections 4 and 5 takes place among the first points of ray 1, so we need words only for permutations supported there. For $n \geq 1$ let $\Sigma_n \leq H$ be the group of permutations of Y supported in $\{(1,1),\dots,(1,n)\}$, identified with $\mathrm{Sym}(n)$; we write $(x\ y)$ for the transposition of the points x and y of the first ray.

Conventions. Put $\alpha = [p,q]$; a direct check shows $\alpha = (1\ 2)$. In [26] permutations act on the right. Johnson's presentation of H [48], also in [55, Theorem 2.8], is written there in generators a, b, α that are the inverses, as maps, of our p, q, α. The map $g \mapsto g^{-1}$ is an isomorphism from the permutations of Y with the product of [26] to the permutations of Y under composition, and it sends a, b, α to p, q, α. A word in a, b, α therefore represents 1 in the conventions of [26] if and only if the same word in p, q, α represents 1 here. Hence the relators of [26, Section 2], read in our letters, give the presentation

$$P_J = \langle p, q, \alpha \mid\ \alpha^2,\ (\alpha p \alpha p^{-1})^3,\ [\alpha, p^2 \alpha p^{-2}],\ [p,q]\alpha^{-1},\ p\alpha p^{-1} q \alpha^{-1} q^{-1} \rangle \tag{12}$$

of H; in particular $H = \langle p, q \rangle$. Renaming the letters is an isomorphism of free groups that carries the relators of [26] to those of P_J, so every area computed in [26] for Johnson's presentation holds verbatim for P_J.

Eliminating α by the relator $[p,q]\alpha^{-1}$, a Tietze transformation, gives the presentation

$$P_H = \langle p, q \mid\ s^2,\ (sps\,p^{-1})^3,\ [s, p^2 s p^{-2}],\ psp^{-1}qs^{-1}q^{-1} \rangle, \qquad s = [p,q], \tag{13}$$

of H. Let δ_H be its Dehn function: $\delta_H(n)$ is the largest area, with respect to the four relators of P_H, of a word of length at most n in $p^{\pm 1}, q^{\pm 1}$ that represents 1 in H. It is nondecreasing.

Lemma A.1 (Finite permutations)**.** *Put $\sigma_k = p^k \alpha p^{-k}$ for $k \geq 0$.*

(a) $\sigma_k = (k{+}1\ \ k{+}2)$, and for $1 \leq x < y$ the word $\omega_{x,y} = p^{x-1}(\alpha p)^{y-x-1}(\alpha p^{-1})^{y-x-1}\alpha\, p^{1-x}$ represents $(x\ y)$. With $\alpha = pqp^{-1}q^{-1}$ it has length at most $10y - 8x - 8 < 10y$ in p, q.

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

(c) Let $x_1, \dots, x_l \in \{1, \dots, n\}$ be distinct. Put $\pi_0 = 1$ and, for $m = 1, \dots, l$, $w_m = \pi_{m-1}^{-1}(x_m)$ and $\pi_m = \pi_{m-1}(m\ w_m)$, where $(m\ m) = 1$. Then $\pi[x_1, \dots, x_l] := \pi_l$ sends m to x_m for $m \leq l$, it is a product of at most l transpositions of points in $\{1, \dots, n\}$, and the product of the corresponding words ω represents it with length less than $10ln$.

Proof. (a) Since p^k maps $(1,1), (1,2)$ to $(1,k+1), (1,k+2)$, σ_k is the stated transposition. Hence $(x\ y) = \sigma_{x-1}\sigma_x \cdots \sigma_{y-3}\,\sigma_{y-2}\,\sigma_{y-3} \cdots \sigma_{x-1}$, by induction on $y - x$ from $(x\ y) = \sigma_{x-1}(x{+}1\ \ y)\sigma_{x-1}$. In the product the powers of p telescope: $\sigma_k\sigma_{k+1} = p^k \alpha p \alpha p^{-k-1}$ and $\sigma_{k+1}\sigma_k = p^{k+1}\alpha p^{-1}\alpha p^{-k}$, which gives $\omega_{x,y}$ after free reduction. Its length is $2(x-1) + 10(y - x - 1) + 4$.

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

(c) Since $\pi_{m-1}(j) = x_j \neq x_m$ for $j < m$, we have $w_m \geq m$, so $(m\ w_m)$ fixes $1, \dots, m-1$ and $\pi_m(j) = x_j$ for $j \leq m$. Each factor has length less than $10n$ by (a). □

That a point stabilizer of a Houghton group is isomorphic to the group is known [24, Example 3.6]; part (a) of Lemma A.2 gives such isomorphisms explicitly on generators.

Lemma A.2 (Point stabilizers)**.** *For $k \geq 1$ let $H^{(k)}$ be the pointwise stabilizer in H of $(1,1), \dots, (1,k)$, 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, α to $\zeta_k p, \zeta_k q, \sigma_k$ and the transposition $(x\ y)$ to $(x{+}k\ \ y{+}k)$. In particular $\zeta_k p$ and $\zeta_k q$ generate $H^{(k)}$.*

(b) *Every element of $\Sigma_n \cap H^{(k)}$ that moves $m \geq 2$ points is represented by a word of length less than $10(m-1)n$ in $\zeta_k p$ and $\zeta_k q$.*

In p, q the words $\zeta_1 p = [p, q]\, p$ and $\zeta_2 p = p[p, q]p^{-1}[p, q]\, p$ have length at most 5 and 11, and likewise for q.

Proof. (a) Let $\theta_k : Y \to Y \setminus \{(1,1), \dots, (1,k)\}$ send $(1, j)$ 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 $(1,1), \dots, (1,k)$. Then $g^{[k]}$ is eventually a translation on each ray, and $g \mapsto g^{[k]}$ is an isomorphism $H \to H^{(k)}$, since an element of $H^{(k)}$ restricts to a permutation of the image of θ_k. Clearly $(x\ y)^{[k]} = (x{+}k\ \ y{+}k)$, so $\alpha^{[k]} = \sigma_k$. The element ζ_k is the cycle $(1,1) \mapsto (1,k+1) \mapsto (1,k) \mapsto \cdots \mapsto (1,2) \mapsto (1,1)$. Hence $\zeta_k p$ fixes $(1, j)$ for $j \leq k$, maps $(1, j)$ to $(1, j+1)$ for $j > k$, maps $(2,1)$ to $(1, k+1)$, and agrees with p elsewhere; that is, $\zeta_k p = p^{[k]}$. Likewise $\zeta_k q = q^{[k]}$.

(b) An element $g \in \Sigma_n \cap H^{(k)}$ moving m points is $g'^{[k]}$ for some $g' \in \Sigma_{n-k}$ moving m points. Write g' as in Lemma A.1(b) and substitute $\zeta_k p, \zeta_k q$ for p, q. □

Part (a) of the next proposition is Lee's theorem [55, Theorem D], and part (b) is [26, Theorems 3.2 and 3.3]. Both pass to P_H by a Tietze transformation.

Proposition A.3 (Dehn function)**.** *Let δ_J be the Dehn function of P_J. Then $\delta_H \leq \delta_J$, and:*

(a) *there is an absolute constant K with $\delta_H(n) \leq e^{Kn}$ for $n \geq 1$;*

(b) *there is an absolute constant C with $\delta_H(n) \leq C n^{p_H}$ for $n \geq 1$, where $p_H = \log 107 + 4 \approx 10.74$.*

Proof. Let w be a word in p, q of length at most n representing 1. In the free group on p, q, α it is a product of at most $\delta_J(n)$ conjugates of relators of P_J and their inverses. Apply the homomorphism to the free group on p, q that fixes p, q and sends α to $[p, q]$. It fixes w, sends $[p, q]\alpha^{-1}$ to the trivial word, and sends each of the other four relators to a relator of P_H. Hence $\mathrm{Area}_{P_H}(w) \leq \delta_J(n)$.

(a) Lee proved that H_3 satisfies an exponential isoperimetric inequality [55, Theorem D]. The Dehn functions δ, δ' of two finite presentations of the same group are equivalent: $\delta(n) \leq C'\delta'(C'n) + C'n + C'$ for a constant C' [10]. This bound preserves the class of functions bounded by e^{Kn}.

(b) By [26, Theorem 3.3], $\delta_J(n) \leq 400n^4(1 + A^*(4n+2))$, where the far-commutator area satisfies $A^*(M) \leq 135M^{\log 107}$ [26, Theorem 3.2]; so $\delta_J(n) = O(n^{\log 107 + 4})$. By the comparison of conventions at the start of this appendix, this holds for P_J in our letters, and $\delta_H \leq \delta_J$. □

Part (a) rests on Lee's published exponential bound, part (b) on our polynomial one. Both reach the lower bounds of Part I through a single quantity, δ_H at scale N, where N is the number of points that carry the configurations.

B The counting criterion and sequential extraction

The lower bounds of Theorems A and B stand on two results of the companion papers: the counting criterion (Theorem 2.1, which is [26, Theorem 8.1]) and sequential extraction (Theorem 9.1, which is [27, Theorem 5.2]). This appendix reproduces both proofs in the notation of this paper, so that a reader can check the main theorems without the companions. The remaining imports, chiefly the Dehn bound for H_3 (Proposition A.3), are listed in Table 2. Within this appendix, $\|uv - vu\|_2$ measures how far two unitaries are from commuting, and ln is the natural logarithm.

B.1 Three facts about copies of S_3

The first fact rounds an approximate representation of $S_3 = \langle s, t \mid s^2, t^2, (st)^3 \rangle$ to an exact one.

Lemma B.1 (Exact S_3 rounding)**.** *If $A, B \in U(D)$ satisfy $\|A^2 - I\|_2, \|B^2 - I\|_2, \|(AB)^3 - I\|_2 \le e$, there is a unitary representation u of S_3 on $\mathbb{C}^D$ with $\|u(s) - A\|_2 \le e$, $\|u(t) - B\|_2 \le 8e$ and $\|u(st) - AB\|_2 \le 9e$, and each normal form $1, s, t, st, ts, sts$ evaluated in u is within $10e$ of the same word in A, B.*

Proof. Round the spectra of A and B to the nearer of ± 1. For $|z| = 1$ this moves z by at most $|z^2 - 1|$, so the resulting Hermitian unitaries S, T satisfy $\|S - A\|_2, \|T - B\|_2 \le e$. Then $C = ST$ satisfies $SCS = C^*$ and $\|C^3 - I\|_2 \le 7e$. Round the spectrum of C to the nearest cube root of unity, choosing 1 at ties, with one exception: send -1 to 1, although it is not the nearest root there. At that point the displacement is $2 = |(-1)^3 - 1|$. This rule commutes with complex conjugation and moves z by at most $|z^3 - 1|$, so C_0 satisfies $C_0^3 = I$, $SC_0S = C_0^*$ and $\|C_0 - C\|_2 \le 7e$. Then $u(s) = S$ and $u(t) = SC_0$ are involutions with $u(s)u(t) = C_0$, and the bounds follow from the triangle inequality. □

The second fact is the entropy that nearly commuting exact copies force. Write h_2 for the binary entropy.

Lemma B.2 (Entropic sector bound)**.** *Let $u_1, \dots, u_N : S_3 \to U(D)$ be unitary representations, not assumed to commute, such that for a three-cycle c*

$$\|u_i(c) - I\|_2 \ge \Delta \quad (1 \le i \le N), \qquad \|u_i(g)u_j(h) - u_j(h)u_i(g)\|_2 \le \eta \quad (i \ne j,\ g, h \in S_3).$$

Put $\theta = \sqrt{2N}\,\eta$. If $\theta \le \Delta$ and $\theta^2 \le 5/6$, then

$$\log D \ \ge\ N\Big[\frac{(\Delta - \theta)^2}{3} - h_2(\theta^2) - \theta^2 \log 5\Big].$$

Proof. Let $\Sigma = S_3^N$ and $f(g) = u_1(g_1) \cdots u_N(g_N)$. For h in factor i, $\|f(gh) - f(g)u_i(h)\|_2 = \|Y^*XY - X\|_2$ with $X = u_i(h)$ and $Y = u_{i+1}(g_{i+1}) \cdots u_N(g_N)$, and $\|Y^*XY - X\|_2^2 = 2 - 2\,\mathrm{Re}\,\tau(X^*Y^*XY)$. The averages $T_j(Z) = \frac{1}{6}\sum_a u_j(a)^* Z u_j(a)$ are orthogonal projections on M_D with the normalized Hilbert–Schmidt inner product, the average of Y^*XY over g is $T_N \cdots T_{i+1}(X)$, and $\|(I - T_j)X\|_2^2 = \frac{1}{12}\sum_a \|[X, u_j(a)]\|_2^2 \le \eta^2/2$. Lemma C.4 therefore gives

$$\mathbb{E}_g \|f(gh) - f(g)u_i(h)\|_2^2 \le 2(N - i)\eta^2 \le \theta^2.$$

Let R be the right regular representation of Σ and $W : \mathbb{C}^D \to \ell^2(\Sigma) \otimes \mathbb{C}^D$ the isometry $Wv = |\Sigma|^{-1/2} \sum_g |g\rangle \otimes f(g)v$. Then $\|\mathsf{R}(h)W - Wu_i(h)\|_{HS} \le \sqrt{D}\,\theta$ for h in factor i. Put $\Pi = WW^*$ and $\rho = \Pi/D$, a state of entropy $\log D$. For the three-cycle c_i of factor i and $Z_i = 2I - \mathsf{R}(c_i) - \mathsf{R}(c_i)^*$,

$$\mathrm{Tr}(\rho Z_i) = D^{-1}\|(\mathsf{R}(c_i) - I)W\|_{HS}^2 \ge (\Delta - \theta)^2.$$

Let $\mathcal{E}_i(X) = \frac{1}{6}\sum_h \mathsf{R}(h_i) X \mathsf{R}(h_i)^*$, with h_i the element h placed in factor i. These conditional expectations commute. Since $(I - \Pi)W = 0$, $\mathrm{Tr}[(I - \Pi)\mathcal{E}_i(\rho)] \le \theta^2$. The state $\mathcal{E}_i(\rho)$ is an average of six states of rank D, one of them ρ, so after pinching by Π and $I - \Pi$ its two blocks have rank at most D and $5D$. Pinching does not decrease entropy, and $x \mapsto h_2(x) + x \log 5$ increases on $[0, 5/6]$, so $S(\mathcal{E}_i(\rho)) \le \log D + h_2(\theta^2) + \theta^2 \log 5$. For a conditional expectation $S(\mathcal{E}(\sigma)) - S(\sigma) = D(\sigma \| \mathcal{E}(\sigma))$. Since the $\mathcal{E}_j$ commute, monotonicity of relative entropy under $\mathcal{E}_{i-1} \cdots \mathcal{E}_1$ bounds the entropy increment of $\mathcal{E}_i$, applied after $\mathcal{E}_{i-1} \cdots \mathcal{E}_1$, by $D(\rho \| \mathcal{E}_i \rho)$. Hence $\bar{\rho} = \mathcal{E}_N \cdots \mathcal{E}_1(\rho)$ satisfies $S(\bar{\rho}) \le \log D + N[h_2(\theta^2) + \theta^2 \log 5]$.

The state $\bar\rho$ commutes with $\mathsf{R}(\Sigma)$, so $\bar\rho = \bigoplus_\lambda p_\lambda (I_{d_\lambda}/d_\lambda \otimes \sigma_\lambda)$ over the irreducible representations λ of S_3^N, and $S(\bar\rho) \geq \sum_\lambda p_\lambda \log d_\lambda = \sum_\lambda p_\lambda k_\lambda$, where k_λ is the number of factors of λ that carry the two-dimensional representation of S_3. The operator Z_i is central in the group algebra, so $\mathrm{Tr}(\bar\rho Z_i) = \mathrm{Tr}(\rho Z_i)$, and it acts as $3I$ on the sectors whose factor i is two-dimensional and as 0 on the others. Hence $N(\Delta-\theta)^2 \leq \sum_i \mathrm{Tr}(\bar\rho Z_i) = 3\sum_\lambda p_\lambda k_\lambda \leq 3S(\bar\rho)$. □

The third fact runs the same count on the three-cycles alone. It needs no relation between the transpositions of different copies. Put $\chi(x) = h_2(\min\{\frac{2}{3}x, \frac{2}{3}\}) + \min\{\frac{2}{3}x, \frac{2}{3}\}$, a concave function.

Theorem B.3 (Phase-robust sector bound)**.** *Let $c_i, x_i \in U(D)$, $1 \leq i \leq N$, satisfy $c_i^3 = x_i^2 = I$ and $x_i c_i x_i = c_i^{-1}$, with no condition on how x_i and x_j commute. Put $\Delta_i = \|c_i - I\|_2$ and $f(g) = c_1^{g_1} \cdots c_N^{g_N}$ for $g \in \mathbb{Z}_3^N$, let $g^{(i)}$ be g with its i-th coordinate negated, and, for uniform g, let*

$$\hat a_i^2 = \mathbb{E}_g \|f(g+e_i) - f(g)c_i\|_2^2, \qquad \hat b_i^2 = \mathbb{E}_g \|x_i f(g^{(i)}) - f(g)x_i\|_2^2.$$

Then $\hat a_i^2 \leq \frac{4}{3}\sum_{j>i} \|c_ic_j - c_jc_i\|_2^2$, $\hat b_i^2 \leq \frac{4}{3}\sum_{j\neq i} \|x_ic_j - c_jx_i\|_2^2$, and

$$\log D \;\geq\; \sum_{i=1}^N \Big[\frac{(\Delta_i^2 - \hat a_i^2)_+}{3} - \frac{\hat b_i^2}{2\ln 2} - \chi(\hat a_i^2)\Big].$$

Proof. The ordered-product errors. For a unitary h, $\|(I - \mathcal{A}_j)h\|_2^2 = \frac{1}{3}\|hc_j - c_jh\|_2^2$ for the average $\mathcal{A}_j$ of conjugation by the powers of c_j. Moving c_i through the suffix $c_{i+1}^{g_{i+1}} \cdots c_N^{g_N}$ and averaging over its exponents gives $\hat a_i^2 = 2 - 2\,\mathrm{Re}\langle c_i, \mathcal{A}_N \cdots \mathcal{A}_{i+1} c_i\rangle$, and Lemma C.4 gives the first bound. For the second, the maps $\mathcal{B}_j(Z) = \frac{1}{3}\sum_k c_j^{-k} Z c_j^{\epsilon_j k}$, with $\epsilon_i = -1$ and $\epsilon_j = 1$ for $j \neq i$, are orthogonal projections. The local relation gives $\mathcal{B}_i(x_i) = x_i$, and $\|(I - \mathcal{B}_j)x_i\|_2^2 = \frac{1}{3}\|x_ic_j - c_jx_i\|_2^2$ for $j \neq i$. Moreover $\hat b_i^2 = 2 - 2\,\mathrm{Re}\langle x_i, \mathcal{B}_N \cdots \mathcal{B}_1 x_i\rangle$, and Lemma C.4 gives the second bound.

Embedding. Let $W\xi = 3^{-N/2}\sum_g f(g)\xi \otimes |g\rangle$, let C_i shift the i-th coordinate of $\ell^2(\mathbb{Z}_3^N)$ by -1, let F_i negate it, and put $\mathsf{X}_i = x_i \otimes \mathsf{F}_i$. The C_i commute, $\mathsf{X}_i\mathsf{C}_i\mathsf{X}_i = \mathsf{C}_i^{-1}$ and $\mathsf{X}_i\mathsf{C}_j\mathsf{X}_i = \mathsf{C}_j$ for $j \neq i$. Moreover $D^{-1/2}\|\mathsf{C}_iW - Wc_i\|_{HS} = \hat a_i$ and $D^{-1/2}\|\mathsf{X}_iW - Wx_i\|_{HS} = \hat b_i$.

Entropy. Let $\rho = WW^*/D$, and let $\mathcal{T}_i$ average conjugation by the powers of C_i. The twirl moves weight at most $\min\{\frac{2}{3}\hat a_i^2, \frac{2}{3}\}$ outside the range of WW^*, into a block of rank at most $2D$, so $S(\mathcal{T}_i\rho) \leq \log D + \chi(\hat a_i^2)$. As in Lemma B.2 the twirls commute, and the increment of each twirl, the Holevo quantity of the conjugates of the current state, can only decrease under the earlier twirls by data processing for the Holevo quantity. So $\bar\rho = \mathcal{T}_1 \cdots \mathcal{T}_N\rho$ has $S(\bar\rho) \leq \log D + \sum_i \chi(\hat a_i^2)$.

Sectors. Let p be the joint distribution of the eigenvalue labels $Z_i \in \{1, \omega, \omega^2\}$ of the C_i, the same for ρ and $\bar\rho$. Then $S(\bar\rho) \geq H(p) \geq \sum_i H(Z_i \mid Z_{-i})$. Let Q_α and P_β be the spectral projections of c_i and C_i. The numbers $r_{\alpha\beta} = D^{-1}\|P_\beta W Q_\alpha\|_{HS}^2$ form a joint distribution whose marginals are the spectral distribution of c_i in the normalized trace and that of C_i in ρ. Distinct cube roots of unity are at squared distance 3, so $\hat a_i^2 = 3\Pr(\alpha \neq \beta)$ and $\Delta_i^2 = 3\Pr(\alpha \neq 1)$, whence $s_i := \Pr(Z_i \neq 1) \geq (\Delta_i^2 - \hat a_i^2)_+/3$. Next, the normalized vectorizations of W and of $\mathsf{X}_iWx_i^*$ are at squared distance $\hat b_i^2$, so their real overlap is $1 - \hat b_i^2/2$. Measuring the joint spectral projections of the C_j gives p from the first and its flip $p^{(i)}$ in coordinate i from the second, because X_i exchanges ω and ω^2 in coordinate i and fixes the other labels. By the Cauchy–Schwarz inequality, outcome by outcome, $\sum_z \sqrt{p(z)p(z^{(i)})} \geq 1 - \hat b_i^2/2$. For $0 \leq x \leq 1$ one has $1 - h_2(x) \leq (1 - 2\sqrt{x(1-x)})/\ln 2$: with $y = 2x - 1$, the difference of $1 - \sqrt{1-y^2}$ and $\ln 2\,(1 - h_2(x))$ vanishes with its derivative at $y = 0$ and has second derivative $(1-y^2)^{-3/2} - (1-y^2)^{-1} \geq 0$. So, for fixed other labels with masses m, m' on the two nontrivial values of Z_i, the conditional entropy receives at least $m + m' - (m + m' - 2\sqrt{mm'})/\ln 2$. Summing gives $H(Z_i \mid Z_{-i}) \geq s_i - \hat b_i^2/(2\ln 2)$. □

B.2 Proof of the counting criterion

Proof of Theorem 2.1. Both parts use the copies $A_i = \phi(w_i)\phi(a)\phi(w_i)^{-1}$ and $B_i = \phi(w_i)\phi(b)\phi(w_i)^{-1}$, $1 \le i \le N$, of an assignment ϕ extended to the free group. They have three properties. The S_3 relations of each pair are conjugates of $a^2, b^2, (ab)^3$, so they have the defect of a relator. The product $A_iB_i = \phi(w_i)\phi(ab)\phi(w_i)^{-1}$ has the displacement of $\phi(ab)$. And a commutator of two letters from copies $i \neq j$, or of their inverses, is a conjugate of $[w_i x w_i^{-1}, w_j y w_j^{-1}]^{\pm 1}$ with $x, y \in \{a, b\}$, so its area is at most Λ. A word of area A is freely a product of A conjugates of relators, so under a unitary assignment it lies within A times the relator defect of I; for permutations the same holds for the normalized Hamming distance, which is bi-invariant.

Permutations. Let $f : E \to \mathrm{Sym}_n$ have $f(1) = \mathrm{id}$ and multiply and separate correctly up to $1/r$, with $r \ge 10^8 N\Lambda(2l-1)$, and put $\phi(s) = f(s)$ for $s \in S$. For a word $s_1 \cdots s_m$ of R with prefixes p_k, the chunk contains $p_k s_{k+1} = p_{k+1}$ and $p_m = 1$, and $s \cdot s^{-1} = 1$, so $m-1$ merges and at most m inverse replacements give $d(\phi(w), \mathrm{id}) \le \delta := (2l-1)/r$ for every $w \in R$. In the same way $d(\phi(ab), f(ab)) \le 1/r$, and $d(f(ab), \mathrm{id}) \ge 1 - 1/r$ because $ab \neq 1$; so $d(\phi(ab), \mathrm{id}) \ge 1 - 2/r \ge \frac{1}{2}$. Regard permutations as unitary matrices, with $\|P - I\|_2^2 = 2d(P, \mathrm{id})$. The S_3 relators of each pair then have defect at most $e = \sqrt{2\delta}$, and Lemma B.1 gives exact representations u_i within $10e$ of the normal forms. Commuting two normal forms from different pairs takes at most nine interchanges of letters, so their commutator has $d \le 9\Lambda\delta$ and $\|\cdot\|_2 \le 3\sqrt{2\Lambda\delta}$, and rounding adds at most $40e$; so $\eta \le 43\sqrt{2\Lambda\delta}$. The element $u_i(st)$ lies within $9e$ of A_iB_i, whose displacement in $\|\cdot\|_2$ is at least 1, so $\Delta \ge 1 - 9\sqrt{2\delta} \ge \frac{1}{2}$. Then $\theta^2 = 2N\eta^2 \le 7396 N\Lambda\delta < 10^{-4}$, and Lemma B.2 gives $\log n \ge N[(49/100)^2/3 - h_2(10^{-4}) - 10^{-4}\log 5] \ge N/16$. So $\log \sigma_E(r) \ge N/16$.

Unitaries. Let ϕ map every word of R within e of I, with $\|\phi(ab) - I\|_2 \ge \Delta$ and $K\sqrt{N}\,\Lambda e \le 1$, where $K = K_\Delta = \sqrt{80000\,(1 + \log(1/\Delta))}/\Delta$. Lemma B.1 gives exact pairs $x_i = u_i(s)$ and $c_i = u_i(st)$ with $\|x_i - A_i\|_2 \le e$ and $\|c_i - A_iB_i\|_2 \le 9e$, so $\Delta_i \ge \Delta - 9e$. Four and two interchanges of letters give $\|c_ic_j - c_jc_i\|_2 \le 4\Lambda e + 36e \le 40\Lambda e$ and $\|x_ic_j - c_jx_i\|_2 \le 2\Lambda e + 20e \le 22\Lambda e$ for $i \neq j$. By Theorem B.3 and the concavity of χ,

$$\log D \;\ge\; N\Big[\frac{(\Delta - 9/K)^2 - \bar{a}}{3} - \frac{\bar{b}}{2\ln 2} - \chi(\bar{a})\Big], \qquad \bar{a} \le \frac{3200}{3K^2}, \quad \bar{b} \le \frac{1936}{3K^2},$$

where $\bar{a}$ and $\bar{b}$ are the averages of $\hat{a}_i^2$ and $\hat{b}_i^2$, since $e \le 1/K$ and $N\Lambda^2 e^2 \le 1/K^2$. It remains to show that the bracket is at least $\Delta^2/12$. Since $K \ge 36/\Delta$, $(\Delta - 9/K)^2 \ge 9\Delta^2/16$, so it suffices that $\bar{a}/3 + \bar{b}/(2\ln 2) + \chi(\bar{a}) \le 5\Delta^2/48$. Here $\bar{a}/3 + \bar{b}/(2\ln 2) < 821.2/K^2$. With $q = 2\bar{a}/3 < 711.2/K^2$, the bound $(1-q)\log\frac{1}{1-q} \le q\log e$ gives $\chi(\bar{a}) = h_2(q) + q \le q\log(2e/q)$, and $x\log(2e/x)$ increases for $x < 2$. Put $L = \log(1/\Delta)$, so that $K^2 = 80000(1+L)/\Delta^2$. Multiplying by K^2 and using $\log(1+L) \le L + \frac{1}{10}$ for $L \ge 0$, the left side becomes at most

$$821.2 + 711.2\Big[\log\frac{2e \cdot 80000}{711.2} + \log(1+L) + 2L\Big] \;\le\; 7476 + 2134L \;\le\; \frac{5}{48}\,80000\,(1+L),$$

which is the required inequality. Hence $\log D \ge N\Delta^2/12$. □

B.3 Proof of sequential extraction

Proof of Theorem 9.1. A test with perfect completeness. In every round the observer submits half of a fresh maximally entangled pair, measures $\{\Pi_\Phi, I - \Pi_\Phi\}$ on the output and the reference, keeps the outcome, discards the pair and continues. Let p_t be the probability that the tests of rounds $1, \ldots, t-1$ all pass, ℓ_t the probability of failing round t given that, and $F \in \{1, \ldots, n, \infty\}$ the first failure, with ∞ for none. The ideal experiment never fails, so $f = \Pr(F \le n) \le \epsilon$. Hence $p_t \ge 1 - \epsilon$ and $\sum_t p_t\ell_t = f \le \epsilon$.

Entropy ledger. Let ρ_t be the memory state given that the first $t-1$ tests passed; for the bookkeeping, a branch that fails is frozen. Since the input marginal is τ_d, the round takes ρ_t to $T_t(\rho_t) = \mathrm{Tr}_d\, U_t(\tau_d \otimes \rho_t)U_t^*$, raising the entropy by ΔS_t. Recording the outcome lowers the average conditional entropy by at most $h_2(\ell_t)$. Exporting k qubits lowers the entropy by at most k, and importing k qubits in a state of entropy $k-c$ raises it by $k-c$, so an exchange lowers it by at most the imported deficit c. The average entropy starts at $Q-P$ and ends at most at Q, so

$$\sum_t p_t \Delta S_t \le P + C + \sum_t p_t h_2(\ell_t) \le B + H(F) \le B + h_2(\epsilon) + \epsilon \log n,$$

where $\sum_t p_t h_2(\ell_t) = H(F)$ by the chain rule and $H(F) \le h_2(f) + f \log n$. Therefore $\sum_t p_t(\ell_t + \Delta S_t) \le B + \epsilon + h_2(\epsilon) + \epsilon \log n$, and since $\sum_t p_t \ge n(1-\epsilon)$, some round has $\ell_t + \Delta S_t \le t_n$.

The extracted channel. Each round's unitary acts only on the input and the Q memory qubits, and the memory state ρ_t conditioned on the earlier passes does not depend on the current input, since the earlier test registers are discarded. So the round defines a channel Ψ_t with a bath of dimension 2^Q, and its support leakage is exactly ℓ_t. For an arbitrary current input with reference, the actual state is p_t times a normalized pass-block state σ plus a failure block, while the ideal state ω lies in the pass block. Closeness of the complete experiments gives $\frac{1}{2}\|p_t\sigma - \omega\|_1 + \frac{1-p_t}{2} \le \epsilon$, hence $p_t \frac{1}{2}\|\sigma - \omega\|_1 \le \epsilon$ and $d_\diamond(\Psi_t, \Phi) \le \epsilon/(1-\epsilon)$. □

The device performs no test. The tests and the conditioning on success belong to one particular observer and to the proof.

C Proof of the weighted one-round theorem

This appendix proves Theorem 10.1, the weighted one-round theorem. Theorem B uses it through Corollary 10.2 in Section 14, and Appendix E reuses its proof in a commutator form.

Four steps lead from the round to the entropy bound. A round with small leakage and entropy production yields unitaries on the bath under which every relation holds up to $O(\sqrt{\nu})$, with errors weighted by the memory state (Lemma C.1). Weighted errors add up over a van Kampen diagram, at a price paid once for each face and each vertex (Lemma C.2). Each transported copy of S_3 is rounded to an exact representation on the same space, so the state and its entropy do not change (Lemma C.3). Measuring the rounded three-cycles one after another on one half of a purification of the memory state produces about one bit per copy that the other half also knows (Theorem C.6).

The argument generalises [27, Appendix B], which proves the one-round theorem for the Houghton gadget [27, Theorem 5.5], to every gadget channel that satisfies the hypotheses of Theorem 10.1. The weighted norms are those of [27, Appendix B]. Lemmas C.1 and C.2 extend Lemmas B.4 and B.2 there from the Houghton gadget to every such gadget. Lemma C.3 is Lemma B.3 there. Theorem C.6 is a form of Theorem B.1 there, now proved by sequential measurement.

C.1 Weighted norms

Throughout, $\|\cdot\|_F$ is the unnormalized Hilbert–Schmidt norm on M_M, ρ is a state on $\mathbb{C}^M$ and $\xi = \sqrt{\rho}$, so $\|\xi\|_F = 1$. For operators W on $\mathbb{C}^M$ put

$$E_\rho(W) = \|(W - I)\xi\|_F, \qquad \chi_\rho(W) = \|W\xi - \xi W\|_F,$$

the *displacement* and the *centrality defect* of W at ρ. We drop the subscript ρ. For unitaries u, v and any operator F:

(W1) $E(uv) \le E(u) + E(v)$ and $E(u^*) = E(u)$;

(W2) $\chi(uv) \le \chi(u) + \chi(v)$ and $\chi(u^*) = \chi(u)$;

(W3) $\|Fv\xi\|_F \le \|F\xi\|_F + \|F\|\,\chi(v)$;

(W4) $|E(uWu^*) - E(W)| \le 2\chi(u)$ for every W with $\|W\| \le 1$;

(W5) $\|\xi F\|_F = \|F^*\xi\|_F$, and $\|\xi F\|_F = \|F\xi\|_F$ if F is normal.

For (W1) write $(uv - I)\xi = u(v - I)\xi + (u - I)\xi$ and $(u^* - I)\xi = -u^*(u - I)\xi$; for (W2), $uv\xi - \xi uv = u(v\xi - \xi v) + (u\xi - \xi u)v$ and $u^*\xi - \xi u^* = -u^*(u\xi - \xi u)u^*$. For (W3) write $Fv\xi = F\xi v + F(v\xi - \xi v)$. For (W4), $(uWu^* - I)\xi = u(W - I)u^*\xi$, to which (W3) applies with $\|W - I\| \le 2$, in both directions. For (W5), $\|F^*\xi\|_F^2 = \mathrm{Tr}(\xi F F^*\xi)$, which equals $\mathrm{Tr}(\xi F^* F\xi)$ when F is normal. A word w in $S^{\pm 1}$ is evaluated by an assignment ϕ of unitaries to S through $\phi(x^{-1}) = \phi(x)^*$; by (W2), $\chi(\phi(w)) \le |w| \max_{x \in S} \chi(\phi(x))$.

C.2 Extraction

Lemma C.1 (Weighted extraction)**.** *Let G, S and R be as in Section 9, with $a, b, j \in S$, $j^{-1}ab \in R$ and $j \ne 1$ in G, and let m be as in Theorem 10.1. Let $\Psi(X) = \mathrm{Tr}_M U(X \otimes \rho)U^*$ be a round with $u = d_\diamond(\Psi, \Phi_G) \le 1/8$, leakage ℓ relative to Φ_G and entropy production ΔS, and put $\nu = \ell + \Delta S$, $\eta = \sqrt{d\nu}$ and $e = 256\, m\, \eta$. There are unitaries $\phi(x)$, $x \in S$, on $\mathbb{C}^M$ such that*

(i) $E(\phi(r)) \le e$ *for every* $r \in R$;

(ii) $\chi(\phi(x)) \le e$ *for every* $x \in S$;

(iii) $E(\phi(a)\phi(b)) \ge 1 - e$.

Proof. Normalize the output basis. Let (ζ_i) and (ζ_i') be orthonormal eigenbases of ρ and $T(\rho)$ with decreasing eigenvalues λ_i and μ_i. Replacing U by $(I \otimes W)U$, where the unitary W maps ζ_i' to ζ_i, changes neither Ψ nor its leakage and conjugates $T(\rho)$ by W. So we may assume $T(\rho) = \sum_i \mu_i |\zeta_i\rangle\langle\zeta_i|$. Input and output bath are now the same space, and $\xi = \sum_i \sqrt{\lambda_i}|\zeta_i\rangle\langle\zeta_i|$ acts on both; this does not assert $T(\rho) = \rho$.

Entropy production controls commutators. Write $U_{ji} = (I \otimes \langle\zeta_j|)U(I \otimes |\zeta_i\rangle) \in M_d$ and $t_{ji} = \frac{1}{d}\|U_{ji}\|_F^2$. Since the columns and the rows of U are orthonormal, t is doubly stochastic, and the diagonal of $T(\rho)$ gives $\mu_j = \sum_i t_{ji}\lambda_i$. The distributions $p_{ji} = t_{ji}\lambda_i$ and $q_{ji} = t_{ji}\mu_j$ satisfy

$$D(p\|q) = \sum_{j,i} t_{ji}\lambda_i \log \frac{\lambda_i}{\mu_j} = -S(\rho) + S(T(\rho)) = \Delta S.$$

By Jensen's inequality $-2\ln \sum \sqrt{p_{ji}q_{ji}} \le D(p\|q)\ln 2$, and $-2\ln x \ge 2(1-x)$, so $\sum_{j,i} t_{ji}(\sqrt{\lambda_i} - \sqrt{\mu_j})^2 \le \Delta S \ln 2$. The matrix t is a convex combination of permutation matrices, by Birkhoff's theorem, and both sequences decrease, so the rearrangement inequality gives $\sum_{j,i} t_{ji}\sqrt{\lambda_i\mu_j} \le \sum_j \sqrt{\lambda_j\mu_j}$. Expanding the squares and using double stochasticity, $\sum_j(\sqrt{\lambda_j} - \sqrt{\mu_j})^2 \le \sum_{j,i} t_{ji}(\sqrt{\lambda_i} - \sqrt{\mu_j})^2$. The bath block (j,i) of $U(I \otimes \xi) - (I \otimes \xi)U$ is $(\sqrt{\lambda_i} - \sqrt{\lambda_j})U_{ji}$, and $(\sqrt{\lambda_i} - \sqrt{\lambda_j})^2 \le 2(\sqrt{\lambda_i} - \sqrt{\mu_j})^2 + 2(\sqrt{\mu_j} - \sqrt{\lambda_j})^2$. Therefore, with $U = \sum_{o,o'} |o\rangle\langle o'| \otimes A_{oo'}$ in system blocks $A_{oo'} \in M_M$,

$$\sum_{o,o'} \|[A_{oo'}, \xi]\|_F^2 = \|[U, I \otimes \xi]\|_F^2 \le 4d\,\Delta S \ln 2 =: \epsilon_S^2. \tag{14}$$

Leakage controls residuals. The Choi state $\mathcal{J}(\Psi)$ is the marginal on $\mathbb{C}^d \otimes \mathbb{C}^d$ of the unit vector $d^{-1/2}\sum_{o,o'} |oo'\rangle \otimes A_{oo'}\xi$ in $\mathbb{C}^d \otimes \mathbb{C}^d \otimes \mathcal{H}$, where $\mathcal{H}$ is M_M with the Hilbert–Schmidt inner

product. Indeed $\langle A_{o_2o_2'}\xi, A_{o_1o_1'}\xi\rangle = \mathrm{Tr}(A_{o_1o_1'}\rho A^*_{o_2o_2'}) = \langle o_1|\Psi(|o_1'\rangle\langle o_2'|)|o_2\rangle$. The support of $\mathcal{J}(\Phi_G)$ is spanned by the orthogonal vectors $|A_g\rangle\!\rangle$, of squared norms $\mu_g = \|A_g\|_F^2$. Projecting onto it, put

$$X_g = \mu_g^{-1} \sum_{[w_a]=g} \overline{c_a} A_a, \qquad R_{oo'} = A_{oo'} - \sum_g (A_g)_{oo'} X_g,$$

where A_a is the block at slot a. Then $\sum_{o,o'} \|R_{oo'}\xi\|_F^2 = d\ell =: \epsilon_\ell^2$. At a slot $A_a = c_a X_{[w_a]} + R_a$, and at every other position $A_{oo'} = R_{oo'}$. We have $\epsilon_\ell \le \eta$, $\epsilon_S \le 2\sqrt{\ln 2}\,\eta$, and, by Cauchy–Schwarz, $\epsilon_\ell + \epsilon_S \le \sqrt{1+4\ln 2}\,\eta < 2\eta$ and $2\epsilon_\ell^2 + 2\epsilon_S^2 \le 8\ln 2\,\eta^2$.

Anchors are nearly unitary. Every symbol w, and the empty word, occupies a diagonal position $z = z(w)$ whose row and column contain no other slot; there the coefficient is 1, so $B_w := A_{zz} = X_{[w]} + R_{zz}$. Column unitarity gives $\mathrm{Tr}\,\rho(I - B_w^* B_w) = \sum_{o\neq z} \|R_{oz}\xi\|_F^2 \le \epsilon_\ell^2$. Row unitarity gives $\mathrm{Tr}\,\rho(I - B_w B_w^*) = \sum_{o'\neq z} \|\xi A_{zo'}\|_F^2 \le \sum_{o'\neq z} (\|R_{zo'}\xi\|_F + \|[A_{zo'},\xi]\|_F)^2 \le 2\epsilon_\ell^2 + 2\epsilon_S^2$. Let $B_w = V_w|B_w| = |B_w^*|V_w$ with V_w unitary. Since $(1-x)^2 \le 1 - x^2$ on $[0,1]$,

$$\|(V_w - B_w)\xi\|_F \le \epsilon_\ell, \qquad \|\xi(V_w - B_w)\|_F = \|\xi(I - |B_w^*|)\|_F \le \sqrt{8\ln 2}\,\eta,$$

and so $\chi(V_w) \le \|(V_w - B_w)\xi\|_F + \|[B_w,\xi]\|_F + \|\xi(B_w - V_w)\|_F \le \epsilon_\ell + \epsilon_S + \sqrt{8\ln 2}\,\eta < 4.3\eta$, since $\epsilon_\ell + \epsilon_S \le \sqrt{1+4\ln 2}\,\eta$.

Triangles enforce the products. In the block of a triangle (x,y,z), compare each slot a with the anchor of the symbol $w'(a)$ that carries the same label. These symbols are the empty word at the upper left, y at the upper right, x at the lower left and z at the lower right. Put $F_a = A_a - c_a V_{w'(a)}$. Then $F_a = R_a - c_a R_{z'z'} + c_a(B_{w'(a)} - V_{w'(a)})$ with $z' = z(w'(a))$, and since a and (z',z') are different positions, Cauchy–Schwarz gives

$$\|F_a\xi\|_F \le \sqrt{3/2}\,\epsilon_\ell + 2^{-1/2}\epsilon_\ell < 2\eta, \qquad \|F_a\| < 2,$$
$$\|\xi F_a\|_F \le \|F_a\xi\|_F + \|[A_a,\xi]\|_F + 2^{-1/2}\chi(V_{w'(a)}) < 7\eta.$$

For two slots a_1, a_2 in the same row of the block, with $A_{a_i} = c_i V_i + F_i$,

$$A_{a_1}^* A_{a_2} - \overline{c_1} c_2 V_1^* V_2 = A_{a_1}^* F_2 + c_2 F_1^* V_2,$$

whose right-weighted norm is at most $\|F_2\xi\|_F + 2^{-1/2}(\|\xi F_1\|_F + \|F_1\|\chi(V_2)) < 2\eta + 2^{-1/2}\cdot 17\eta < 15\eta$ by (W3) and (W5). The two columns of the block are orthogonal columns of U, so the products for the two rows of the block add up to $-C_1^* C_2$, where C_1, C_2 are the parts of the two columns in the other rows. These are residuals, so $\|C_1^* C_2\xi\|_F \le \|C_2\xi\|_F \le \epsilon_\ell$. The coefficients are $\overline{c_1}c_2 = \frac{1}{2}$ for the pair (upper left, upper right) and $-\frac{1}{2}$ for the pair (lower left, lower right), so

$$\tfrac{1}{2}\|(V_\varnothing^* V_y - V_x^* V_z)\xi\|_F < 15\eta + 15\eta + \eta.$$

Put $W_w = V_\varnothing^* V_w$ for every symbol w, and $W_\varnothing = I$. Since $W_x W_y - W_z = V_\varnothing^* V_x (V_\varnothing^* V_y - V_x^* V_z)$,

$$\|(W_x W_y - W_z)\xi\|_F \le 64\eta \quad \text{for every triangle } (x,y,z), \qquad \chi(W_w) < 10\eta. \tag{15}$$

The relations nearly hold. Let $\phi(x) = W_x$ for $x \in S$. For a letter $x \in S$, the triangle $(x, x^{-1}, \varnothing)$ gives $\|(W_{x^{-1}} - \phi(x^{-1}))\xi\|_F = \|W_x^*(W_x W_{x^{-1}} - I)\xi\|_F \le 64\eta$. Fix a word $r = \sigma_1\cdots\sigma_\ell$ of R and let $D_l = \|(\phi(\sigma_1\cdots\sigma_l) - W_{p_l})\xi\|_F$, with p_l its prefix symbols, $p_1 = \sigma_1$ and $W_{p_\ell} = I$. Then $D_1 \le 64\eta$, and

$$\phi(\sigma_1\cdots\sigma_l) - W_{p_l} = (\phi(\sigma_1\cdots\sigma_{l-1}) - W_{p_{l-1}})\phi(\sigma_l) + W_{p_{l-1}}(\phi(\sigma_l) - W_{\sigma_l}) + (W_{p_{l-1}} W_{\sigma_l} - W_{p_l}).$$

By (W3), (15) and the triangle (p_{l-1}, σ_l, p_l), $D_l \le D_{l-1} + 2\cdot 10\eta + 64\eta + 64\eta$. So $E(\phi(r)) = D_\ell \le 148\,\ell\,\eta \le e$, which is (i); and (ii) holds because $\chi(\phi(x)) < 10\eta \le e$.

The copy of ab is displaced. Let $z = z(j)$ and let 0 be the position of the empty word. The labels there are $[j] \neq 1$ and 1, so $\langle z|\Phi_G(|z\rangle\langle 0|)|0\rangle = \tau(\lambda([j])) = 0$, while $\langle z|\Psi(|z\rangle\langle 0|)|0\rangle = \operatorname{Tr}(B_j \rho B_\varnothing^*)$. Feed $(|zz\rangle + |00\rangle)/\sqrt{2}$ into $\Psi \otimes \mathrm{id}$ and $\Phi_G \otimes \mathrm{id}$. The difference Δ of the outputs is Hermitian and traceless with $\|\Delta\|_1 \leq 2u$, so each of its entries has modulus at most $\|\Delta\| \leq \|\Delta\|_1/2 \leq u$. The entry at $(zz, 00)$ is half the difference of $\langle z|\Psi(|z\rangle\langle 0|)|0\rangle$ and $\langle z|\Phi_G(|z\rangle\langle 0|)|0\rangle$, so $|\operatorname{Tr}(\rho B_\varnothing^* B_j)| \leq 2u$. By the anchor bounds, $|\operatorname{Tr}(\rho V_\varnothing^* V_j) - \operatorname{Tr}(\rho B_\varnothing^* B_j)| \leq 2\epsilon_\ell$, and so

$$E(\phi(j))^2 = \|(V_j - V_\varnothing)\xi\|_F^2 = 2 - 2\operatorname{Re}\operatorname{Tr}(\rho V_\varnothing^* V_j) \geq 2 - 4u - 4\epsilon_\ell \geq \tfrac{3}{2} - 4\eta.$$

By (i) for the word $j^{-1}ab$, $\|(\phi(a)\phi(b) - \phi(j))\xi\|_F = E(\phi(j)^*\phi(a)\phi(b)) \leq e$, so $E(\phi(a)\phi(b)) \geq \sqrt{3/2 - 4\eta} - e \geq 1 - e$ if $\eta \leq 1/8$. If $\eta > 1/8$ then $e > 1$ and (iii) is trivial. □

C.3 Charged area

Relator errors add up over a van Kampen diagram, but in the weighted norms every conjugation costs twice the centrality defect of the conjugator. Writing a null word naively as a product of conjugates of relations would charge each conjugating word once for every relation it encloses. Peeling the diagram one boundary face at a time charges instead each face a bounded amount and each vertex once.

Lemma C.2 (Charged area)**.** *Let S and R be finite, with every word of R of length at most m, and let ϕ assign unitaries to S with $E(\phi(r)) \leq e$ for $r \in R$ and $\chi(\phi(x)) \leq e$ for $x \in S$. Every word w of finite area satisfies*

$$E(\phi(w)) \;\leq\; \left(3m \operatorname{Area}_R(w) + |w|\right) e \;\leq\; \left((4m+1)\operatorname{Area}_R(w) + 2|w|\right) e.$$

Proof. Since ϕ is a homomorphism on the free group, $E(\phi(w))$ depends only on the free reduction of w, which is not longer and has the same area; so let w be freely reduced. By van Kampen's lemma [57, Chapter V], w has a van Kampen diagram with at most $\operatorname{Area}_R(w)$ faces. Such a diagram is a finite, connected and simply connected planar 2-complex with edges labelled by letters of S. Its boundary cycle, read from a base vertex, is w, and the boundary cycle of every face, read from a suitable vertex in a suitable direction, is a word of R. We show, by induction on the number of edges, that every such diagram D with V vertices and A faces satisfies

$$E(\phi(w_D)) \leq \big((2m-1)A + 2(V-1)\big)e, \tag{16}$$

where w_D is its boundary word. If D is a single vertex, w_D is empty. Otherwise let ε be the first edge of the boundary cycle, from the base vertex v_0 to v_1, with label $x^{\pm 1}$.

If ε lies on no face, the boundary cycle traverses it twice, and removing it leaves diagrams $D_0 \ni v_0$ and $D_1 \ni v_1$ with $w_D = x^{\pm 1} w_{D_1} x^{\mp 1} w_{D_0}$, the words read from v_1 and v_0. By (W1) and (W4), $E(\phi(w_D)) \leq 2e + E(\phi(w_{D_1})) + E(\phi(w_{D_0}))$, and the vertices and faces of D_0 and D_1 add up to those of D.

Otherwise ε lies on the boundary of a face f. Let c be the boundary cycle of f read from v_0 and beginning with ε, so that $c = \varepsilon\pi$ with π a path in ∂f from v_1 to v_0. Removing f and ε merges f with the outer region and leaves a van Kampen diagram D' with the same vertices and one face fewer, whose boundary word read from v_0 is $w_{D'} = \pi^{-1}\varrho$, where $w_D = \varepsilon\varrho$. Hence $w_D = c\, w_{D'}$ freely. The word c is a cyclic permutation of some $r^{\pm 1}$, $r \in R$, that is $c = y^{-1} r^{\pm 1} y$ freely with $|y| \leq m - 1$, so $E(\phi(c)) \leq e + 2(m-1)e$ by (W4) and (W2), and $E(\phi(w_D)) \leq (2m-1)e + E(\phi(w_{D'}))$ by (W1).

In both cases (16) follows from the induction hypothesis. To finish, $V - 1$ is at most the number of edges of D, since D is connected. Counting the two sides of every edge gives twice the number of edges as at most $mA + |w|$, the sum of the lengths of the face boundaries and of the boundary cycle. So (16) gives $E(\phi(w)) \leq ((2m-1)A + mA + |w|)e$. □

C.4 Exact S_3 rounding in place

An approximate copy of S_3 is rounded to an exact one by functional calculus on the same space. The state is untouched, so no entropy is gained or lost. The lemma is [27, Lemma B.3].

Lemma C.3 (Exact rounding)**.** *Let X and Y be unitaries on $\mathbb{C}^M$, and suppose that the displacements of X^2, Y^2 and $(XY)^3$ and the centrality defects of X and Y are at most δ. There are unitaries x, c on $\mathbb{C}^M$ with $x^2 = c^3 = I$ and $xcx = c^{-1}$, such that*

$$\|(x - X)\xi\|_F \le \delta, \qquad \chi(x) \le 3\delta, \qquad \|(c - XY)\xi\|_F \le 29\delta, \qquad \chi(c) \le 56\delta.$$

Proof. Let $\operatorname{sgn} t = 1$ for $t \ge 0$ and -1 for $t < 0$, and let $x = \operatorname{sgn}(\operatorname{Re} X)$ and $y = \operatorname{sgn}(\operatorname{Re} Y)$ by functional calculus; they are self-adjoint unitaries. On the unit circle $|z - \operatorname{sgn}\operatorname{Re} z| \le |z^2 - 1|$: for $z = e^{i\vartheta}$ with $|\vartheta| \le \pi/2$ this reads $2|\sin(\vartheta/2)| \le 2|\sin\vartheta|$, which holds as $\cos(\vartheta/2) \ge \frac{1}{2}$, and the case $\operatorname{Re} z < 0$ follows by symmetry. Hence $|x - X|^2 \le |X^2 - I|^2$ as functions of the normal operator X, and $\|(x - X)\xi\|_F \le E(X^2) \le \delta$; by (W5) also $\|\xi(x - X)\|_F \le \delta$, so $\chi(x) \le \chi(X) + 2\delta \le 3\delta$. The same holds for y.

Put $Z = XY$ and $w = xy$, so that $xwx = yx = w^*$. Then $\chi(Z) \le 2\delta$, $\chi(w) \le 6\delta$, and, by (W3), $\|(w - Z)\xi\|_F \le \|x(y - Y)\xi\|_F + \|(x - X)Y\xi\|_F \le \delta + 3\delta = 4\delta$. From $w^3 - Z^3 = w^2(w - Z) + w(w - Z)Z + (w - Z)Z^2$ and (W3),

$$E(w^3) \le E(Z^3) + 4\delta + (4\delta + 2 \cdot 2\delta) + (4\delta + 2 \cdot 4\delta) \le 25\delta.$$

Let f send each point of the unit circle to a nearest cube root of unity, with the ties $e^{\pm i\pi/3}$ and -1 sent to 1. Then $f(\bar z) = \overline{f(z)}$, and $|z - f(z)| \le |z^3 - 1|$: for $z = -1$ both sides equal 2, and otherwise $z = f(z)e^{i\vartheta}$ with $|\vartheta| \le \pi/3$, and the inequality reads $2|\sin(\vartheta/2)| \le 2|\sin(3\vartheta/2)|$. Let $c = f(w)$. Then $c^3 = I$, and $xcx = f(xwx) = f(w^*) = c^* = c^{-1}$. As before $\|(c - w)\xi\|_F \le E(w^3) \le 25\delta$ and $\|\xi(c - w)\|_F \le 25\delta$. Hence $\|(c - Z)\xi\|_F \le 29\delta$ and $\chi(c) \le \chi(w) + \|(c - w)\xi\|_F + \|\xi(c - w)\|_F \le 56\delta$. □

C.5 A weighted sector bound

Rounded copies of S_3 that nearly commute in the weighted sense force entropy. We work in the Hilbert space $\mathcal{H}$ of operators on $\mathbb{C}^M$ with the Hilbert–Schmidt inner product $\langle Y, Z\rangle = \operatorname{Tr}(Y^*Z)$, identified with $\mathbb{C}^M \otimes \mathbb{C}^M$. Left multiplications act on the first factor, which we call the system, and right multiplications on the second, the reference; they commute. The unit vector ξ purifies ρ on the system, and its reference marginal has the same nonzero spectrum, so it has entropy $S(\rho)$. The first lemma is [26, Lemma 4.1], with its short proof.

Lemma C.4 (Projection products)**.** *Let $P_1, \dots, P_k$ be orthogonal projections on a Hilbert space. Every vector y satisfies $\|y\|^2 - \operatorname{Re}\langle y, P_k \cdots P_1 y\rangle \le 2\sum_l \|(I - P_l)y\|^2$.*

Proof. By homogeneity let $\|y\| = 1$. Put $y_0 = y$, $y_l = P_l y_{l-1}$ and $s_l = y_{l-1} - y_l$, which lies in the range of $I - P_l$. Then $\sum_l \|s_l\|^2 = 1 - \|y_k\|^2 =: \beta$, and $\alpha := 1 - \operatorname{Re}\langle y, y_k\rangle = \operatorname{Re}\sum_l \langle (I - P_l)y, s_l\rangle \le \sqrt{E\beta}$ with $E = \sum_l \|(I - P_l)y\|^2$. Since $0 \le \|y - y_k\|^2 = 2\alpha - \beta$, we get $\alpha \le \sqrt{2E\alpha}$, that is $\alpha \le 2E$. □

Let $c_1, \dots, c_N$ be unitaries on $\mathbb{C}^M$ with $c_i^3 = I$, and write $c_i = \sum_{k \in \mathbb{Z}/3} \omega^k P_i^k$ with $\omega = e^{2\pi i/3}$ and spectral projections P_i^k. For every $Y \in \mathcal{H}$,

$$\|(c_i - I)Y\|_F^2 = 3\sum_{k \ne 0} \|P_i^k Y\|_F^2, \qquad \|[c_i, Y]\|_F^2 = 3\sum_{k \ne l} \|P_i^k Y P_i^l\|_F^2, \tag{17}$$

because $|\omega^k - \omega^l|^2 = 3$ for $k \ne l$ and the pieces $P_i^k Y P_i^l$ are orthogonal. For $\mathbf{a} = (a_1, \dots, a_{i-1}) \in (\mathbb{Z}/3)^{i-1}$ put $Q_{\mathbf{a}} = P_{i-1}^{a_{i-1}} \cdots P_1^{a_1}$. Summing over the coordinates one at a time gives $\sum_{\mathbf{a}} Q_{\mathbf{a}}^* Q_{\mathbf{a}} = \sum_{\mathbf{a}} Q_{\mathbf{a}} Q_{\mathbf{a}}^* = I$.

Lemma C.5 (Transfer to the reference)**.** *For every $Y \in \mathcal{H}$ and $1 \le i \le N$,*

$$\sum_{\mathbf{a}} \|Q_{\mathbf{a}} Y - Y Q_{\mathbf{a}}\|_F^2 \;\le\; \frac{4}{3} \sum_{k<i} \|[c_k, Y]\|_F^2 .$$

Proof. The pinchings $\mathcal{E}_k(Y) = \sum_l P_k^l Y P_k^l$ are orthogonal projections on $\mathcal{H}$, and by (17) $\|(I - \mathcal{E}_k)Y\|_F^2 = \sum_{l \neq l'} \|P_k^l Y P_k^{l'}\|_F^2 = \frac{1}{3}\|[c_k, Y]\|_F^2$. Both $\sum_{\mathbf{a}} \|Q_{\mathbf{a}} Y\|_F^2$ and $\sum_{\mathbf{a}} \|Y Q_{\mathbf{a}}\|_F^2$ equal $\|Y\|_F^2$, and summing over the coordinates one at a time, $\sum_{\mathbf{a}} \langle Y Q_{\mathbf{a}}, Q_{\mathbf{a}} Y\rangle = \sum_{\mathbf{a}} \operatorname{Tr}(Q_{\mathbf{a}}^* Y^* Q_{\mathbf{a}} Y) = \langle \mathcal{E}_1 \cdots \mathcal{E}_{i-1} Y, Y\rangle$. So the left side equals $2\|Y\|_F^2 - 2\operatorname{Re}\langle Y, \mathcal{E}_1 \cdots \mathcal{E}_{i-1} Y\rangle$, and Lemma C.4 bounds it by $4 \sum_{k<i} \|(I - \mathcal{E}_k)Y\|_F^2$. □

A three-valued distribution that is nearly symmetric under $k \mapsto -k$ carries about one bit for each unit of weight off 0. Let h_2 be the binary entropy. With $t = (1+y)/2$,

$$\ln 2 \,(1 - h_2(t)) = \sum_{k \ge 1} \frac{y^{2k}}{2k(2k-1)} \;\le\; \sum_{k \ge 1} \binom{2k}{k} \frac{y^{2k}}{(2k-1)4^k} = 1 - \sqrt{1-y^2} = 1 - 2\sqrt{t(1-t)},$$

by comparing coefficients: $4^k \le 2k\binom{2k}{k}$ holds for $k = 1$, and the right side grows by the factor $2(2k+1)/k \ge 4$ from k to $k+1$. For a distribution (r_0, r_1, r_2) with $s = r_1 + r_2 > 0$ this gives, with $t = r_1/s$,

$$H(r_0, r_1, r_2) \;\ge\; s\, h_2(r_1/s) \;\ge\; s - \frac{(\sqrt{r_1} - \sqrt{r_2})^2}{\ln 2}, \tag{18}$$

which also holds when $s = 0$.

Theorem C.6 (Weighted sector bound)**.** *Let c_i, x_i, $1 \le i \le N$, be unitaries on $\mathbb{C}^M$ with $c_i^3 = x_i^2 = I$ and $x_i c_i x_i = c_i^{-1}$, with no condition relating different indices. For a state ρ with $\xi = \sqrt{\rho}$ put $\Delta_i = E(c_i)$, $\theta_i = \chi(c_i)$, $\theta_i' = \chi(x_i)$, and*

$$T_i = \Big(\sum_{k<i} \theta_k^2\Big)^{1/2}, \qquad K_i = \Big(\sum_{k<i} \|[c_i, c_k]\xi\|_F^2\Big)^{1/2}, \qquad K_i' = \Big(\sum_{k<i} \|[x_i, c_k]\xi\|_F^2\Big)^{1/2},$$

$$w_i = \frac{1}{3}\Big(\Delta_i - \frac{4}{\sqrt{3}} T_i\Big)_+^2, \qquad \beta_i = \theta_i' + \frac{2}{\sqrt{3}}(2T_i + K_i'), \qquad \pi_i = \frac{1}{3}\Big(\theta_i + \frac{2}{\sqrt{3}}(2T_i + K_i)\Big)^2 .$$

If $\pi_i \le 2/3$ for all i, then

$$S(\rho) \;\ge\; \sum_{i=1}^{N} \Big[w_i - \frac{\beta_i^2}{2\ln 2} - h_2(\pi_i) - \pi_i \Big].$$

Proof. Measure the copies in turn. Measure $c_1, \dots, c_N$ in turn on the system, with the projections P_i^k, and let $A_i \in \mathbb{Z}/3$ be the outcomes. Then $\Pr(A_{<i} = \mathbf{a}, A_i = k) = \|P_i^k Q_{\mathbf{a}} \xi\|_F^2 =: p(\mathbf{a}, k)$. The measurements do not touch the reference, whose entropy is $S(\rho)$, so

$$S(\rho) \;\ge\; I(A_1 \cdots A_N : \mathrm{Ref}) = \sum_{i=1}^{N} I(A_i : \mathrm{Ref} \mid A_{<i}),$$

since the conditional entropy of the reference given the classical outcomes is nonnegative. Fix i.

The copy stays displaced. By (17), $\Pr(A_i \neq 0) = \frac{1}{3}\sum_{\mathbf{a}} \|(c_i - I) Q_{\mathbf{a}} \xi\|_F^2$. By Lemma C.5 with $Y = \xi$, the vectors $Q_{\mathbf{a}}\xi$ and $\xi Q_{\mathbf{a}}$ differ by at most $\frac{2}{\sqrt{3}} T_i$ in ℓ^2 over $\mathbf{a}$, and $\sum_{\mathbf{a}} \|(c_i - I)\xi Q_{\mathbf{a}}\|_F^2 = \Delta_i^2$ because $\sum_{\mathbf{a}} Q_{\mathbf{a}} Q_{\mathbf{a}}^* = I$. Since $\|c_i - I\| \le 2$, the triangle inequality in ℓ^2 gives $\Pr(A_i \neq 0) \ge w_i$.

The sign is a fair coin. Since $x_i P_i^k x_i = P_i^{-k}$ and right multiplication by x_i is unitary, $\sqrt{p(\mathbf{a},-k)} = \|P_i^k x_i Q_{\mathbf{a}}\xi\|_F$ and $\sqrt{p(\mathbf{a},k)} = \|P_i^k Q_{\mathbf{a}}\xi x_i\|_F$, so $\sum_{\mathbf{a},k}(\sqrt{p(\mathbf{a},k)} - \sqrt{p(\mathbf{a},-k)})^2 \le \sum_{\mathbf{a}} \|x_i Q_{\mathbf{a}}\xi - Q_{\mathbf{a}}\xi x_i\|_F^2$. Write

$$x_i Q_{\mathbf{a}}\xi - Q_{\mathbf{a}}\xi x_i = x_i(Q_{\mathbf{a}}\xi - \xi Q_{\mathbf{a}}) + ((x_i\xi)Q_{\mathbf{a}} - Q_{\mathbf{a}}(x_i\xi)) + Q_{\mathbf{a}}(x_i\xi - \xi x_i).$$

In ℓ^2 over $\mathbf{a}$ the first term is at most $\frac{2}{\sqrt{3}}T_i$ and the third is θ_i'. For the second apply Lemma C.5 with $Y = x_i\xi$: since $[c_k, x_i\xi] = [c_k, x_i]\xi + x_i[c_k,\xi]$, it is at most $\frac{2}{\sqrt{3}}(K_i' + T_i)$. So the sum is at most β_i^2, and since the terms $k=1$ and $k=2$ contribute equally, (18) gives

$$H(A_i \mid A_{<i}) \;\ge\; \Pr(A_i \neq 0) - \frac{1}{\ln 2}\sum_{\mathbf{a}}\big(\sqrt{p(\mathbf{a},1)} - \sqrt{p(\mathbf{a},2)}\big)^2 \;\ge\; w_i - \frac{\beta_i^2}{2\ln 2}.$$

The reference knows the outcome. Let the reference measure c_i from the right, with the projections $Y \mapsto YP_i^l$, and call the outcome B_i. Left and right multiplications commute, so $\Pr(A_{<i} = \mathbf{a}, A_i = k, B_i = l) = \|P_i^k Q_{\mathbf{a}}\xi P_i^l\|_F^2$, and by (17) $\Pr(A_i \neq B_i) = \frac{1}{3}\sum_{\mathbf{a}} \|[c_i, Q_{\mathbf{a}}\xi]\|_F^2$. The same three-term splitting gives

$$[c_i, Q_{\mathbf{a}}\xi] = c_i(Q_{\mathbf{a}}\xi - \xi Q_{\mathbf{a}}) + ((c_i\xi)Q_{\mathbf{a}} - Q_{\mathbf{a}}(c_i\xi)) + Q_{\mathbf{a}}(c_i\xi - \xi c_i),$$

and Lemma C.5 with $Y = \xi$ and $Y = c_i\xi$ gives $\Pr(A_i \neq B_i) \le \pi_i$. By data processing and Fano's inequality for three values,

$$I(A_i : \mathrm{Ref} \mid A_{<i}) \;\ge\; I(A_i : B_i \mid A_{<i}) \;\ge\; H(A_i \mid A_{<i}) - h_2(\pi_i) - \pi_i,$$

where we used that $h_2(\pi) + \pi$ increases on $[0, 2/3]$. Summing over i proves the theorem. □

Corollary C.7. *In Theorem C.6, suppose $\Delta_i \ge 1/2$ for all i, and that θ_i, θ_i', $\|[c_i, c_k]\xi\|_F$ and $\|[x_i, c_k]\xi\|_F$ are at most b for all $i \neq k$, where $\sqrt{N}\,b \le 10^{-3}$. Then $S(\rho) \ge N/16$.*

Proof. Now $T_i, K_i, K_i' \le \sqrt{N}\,b \le 10^{-3}$ and $\theta_i, \theta_i' \le 10^{-3}$. So $w_i \ge \frac{1}{3}(\frac{1}{2} - \frac{4}{\sqrt{3}}10^{-3})^2 > 0.0825$, and β_i and $\sqrt{3\pi_i}$ are at most $10^{-3} + 2\sqrt{3}\cdot 10^{-3} < 4.5\cdot 10^{-3}$. Hence $\beta_i^2/(2\ln 2) < 1.5\cdot 10^{-5}$ and $\pi_i < 6.8\cdot 10^{-6}$, and with $h_2(\pi) \le \pi\log(e/\pi)$ we get $h_2(\pi_i) < 1.3\cdot 10^{-4}$. Each summand exceeds $0.0823 > 1/16$. □

C.6 Completion

Proof of Theorem 10.1. Let ϕ be given by Lemma C.1, with $e = 256\,m\sqrt{d\nu}$, and put

$$b_* = 100\,m\,(A + L + 1)\,e = 25600\,m^2(A + L + 1)\sqrt{d\nu}, \qquad \delta_L = (2L+1)e.$$

Condition (9) gives $\sqrt{N}\,b_* \le 25600\cdot 2^{-25} < 10^{-3}$.

Round the copies. Let $\psi_i = \phi(u_i)$, so $\chi(\psi_i) \le Le$, and put $X_i = \psi_i\phi(a)\psi_i^*$, $Y_i = \psi_i\phi(b)\psi_i^*$ and $Z_i = X_iY_i$. Since a^2, b^2 and $(ab)^3$ lie in R, (W4) and Lemma C.1 give $E(X_i^2), E(Y_i^2), E(Z_i^3) \le e + 2Le = \delta_L$, and (W2) gives $\chi(X_i), \chi(Y_i) \le \delta_L$. Lemma C.3 gives unitaries x_i, c_i with $x_i^2 = c_i^3 = I$, $x_ic_ix_i = c_i^{-1}$ and

$$\|(x_i - X_i)\xi\|_F \le \delta_L, \qquad \chi(x_i) \le 3\delta_L, \qquad \|(c_i - Z_i)\xi\|_F \le 29\delta_L, \qquad \chi(c_i) \le 56\delta_L.$$

By (W4) and Lemma C.1(iii), $E(Z_i) \ge E(\phi(a)\phi(b)) - 2Le \ge 1 - \delta_L$, so $\Delta_i = E(c_i) \ge 1 - 30\delta_L$.

The raw copies nearly commute. Let $i \neq k$ and $x, y \in \{a, b\}$, and let $P = \psi_i\phi(x)\psi_i^*$ and $P' = \psi_k\phi(y)\psi_k^*$. The group commutator $g = [u_ixu_i^{-1}, u_kyu_k^{-1}]$ is a word of length at

most $8L+4$ and area at most A, so Lemma C.2 gives $E(\phi(g)) \le (3mA+8L+4)e$. Since $PP' - P'P = (\phi(g) - I)P'P$ and $\chi(P'P) \le 2\delta_L$, (W3) gives

$$\|(PP' - P'P)\xi\|_F \le \gamma := (3mA+8L+4)e + 4\delta_L = (3mA+16L+8)e.$$

Expanding $[X_i, X_kY_k] = [X_i, X_k]Y_k + X_k[X_i, Y_k]$ and $[X_iY_i, Z_k] = X_i[Y_i, Z_k] + [X_i, Z_k]Y_i$ and using (W3), $\|[X_i, Z_k]\xi\|_F, \|[Y_i, Z_k]\xi\|_F \le 2\gamma + 2\delta_L$ and $\|[Z_i, Z_k]\xi\|_F \le 4\gamma + 6\delta_L$.

The rounded copies nearly commute too. For unitaries U, V, U_0, V_0,

$$[U, V] - [U_0, V_0] = U(V - V_0) + (U - U_0)V_0 - V(U - U_0) - (V - V_0)U_0,$$

so by (W3) the weighted norms of the two commutators differ by at most $2\|(U - U_0)\xi\|_F + 2\|(V - V_0)\xi\|_F + 2\chi(U_0) + 2\chi(V_0)$. With $(U, U_0, V, V_0) = (x_i, X_i, c_k, Z_k)$ and (c_i, Z_i, c_k, Z_k) this gives

$$\|[x_i, c_k]\xi\|_F \le 2\gamma + 68\delta_L = (6mA + 168L + 84)e,$$
$$\|[c_i, c_k]\xi\|_F \le 4\gamma + 130\delta_L = (12mA + 324L + 162)e.$$

Since $m \ge 6$, these and $\chi(c_i) \le (112L+56)e$ and $\chi(x_i) \le (6L+3)e$ are at most b_*, and $\Delta_i \ge 1 - (60L+30)e \ge 1 - b_* > 1/2$. Corollary C.7 gives $S(\rho) \ge N/16$. □

The constant 2^{25} is not optimized; the proof uses only $25600 \cdot 2^{-25} < 10^{-3}$, and Corollary C.7 has room to spare.

Remark C.8 (The constant term). The term $+1$ in (9) cannot be dropped. Take $N = 1$, the empty transport word, so that $L = 0$, and $A = 0$, which is allowed since there are no pairs. The gadget channel has a Stinespring dilation with a pure environment of dimension r_*, so there is a round with $M = r_*$, ρ pure and $\Psi = \Phi_G$ exactly. Then $d_\diamond(\Psi, \Phi_G) = 0$ and the condition with $A + L$ in place of $A + L + 1$ holds, since its left side vanishes, but $S(\rho) = 0 < 1/16$. With the $+1$, the condition forces $\nu \le 2^{-50}m^{-4}d^{-1}$, while this round has $\ell = 0$ and $\Delta S = S(T(\rho))$, the entropy of the complementary output at the maximally mixed input. That output is diagonal with entries $\|A_g\|_F^2/d \ge 1/d$, and there are at least two labels, since $ab \ne 1$; so $\Delta S \ge h_2(1/d)$, far above the threshold.

D Proof of the streaming theorem

This appendix proves Theorem 12.2, the streaming service at the rank rate. Theorem B uses it through Corollary 12.3, which gives the services at the rank rate in Section 14.

The device keeps one rectangular dilation W for all rounds and lets the part of the memory in use, the *active space*, grow by the factor k'/k in every round; that growth is the whole purity bill. Three things are then controlled. The Kraus histories must stay nearly orthogonal, and a Haar-random unitary of the active space before each round ensures this, through a concentration bound on the Gram matrix of the histories. The concentration bound needs the state of everything outside the memory to stay spread out, and exporting about $\frac{1}{2}\log r$ qubits per round ensures that. And the error follows from the Gram bound alone, because the final vector of the service differs from the ideal one by the polar part of a frame. The proof fixes the schedule, chooses the unitaries round by round, and then bounds the error and the memory.

Proof of Theorem 12.2. We use the frames of [27, Appendix A]. A *frame* is a linear map $F : \mathbb{C}^M \to \mathbb{C}^D \otimes \mathbb{C}^S$; its columns are the vectors, in the memory $\mathbb{C}^D$ and an inaccessible exterior $\mathbb{C}^S$, attached to the Kraus histories. The exterior holds the purifier of the initial memory, the exported qubits and the purifiers of the imported ones. Put $g(F) = \|F^*F\|$ and $\alpha(F) = \|\mathrm{Tr}_{\mathbb{C}^D} FF^*\|$. Two lemmas from there carry the argument.

- *Concentration* [27, Lemma A.1]. For Hermitian X on $\mathbb{C}^D$ and Haar-random $V \in U(D)$, the matrix $Z = F^*((V^*XV) \otimes I)F - \tau_D(X)F^*F$ satisfies $\Pr\{\|Z\| \geq s\} \leq 2M \exp[-Ds^2/(4\alpha(F)g(F)\|X\|^2)]$.
- *Export* [27, Lemma A.4]. If $\mathbb{C}^D = \mathbb{C}^{D/e} \otimes \mathbb{C}^e$, there is a unitary of $\mathbb{C}^D$ after which exporting the factor $\mathbb{C}^e$ and importing a maximally mixed e-dimensional register give a frame F' with $F'^*F' = F^*F$ and, for every $\eta > 0$, $\alpha(F') \leq (1+\eta)\alpha(F)/e^2 + 4e^2 g(F)\ln(8Se^2)/(\eta D)$.

Fix the schedule and the dimensions. Put $h = \log r$, $E_p = 2^{\lceil h/2 \rceil}$, $J_t = \lfloor th/2 \rfloor$, $j_t = J_t - J_{t-1} \leq \lceil h/2 \rceil$ and $e_t = 2^{j_t}$, a divisor of E_p. Choose D_0, a multiple of kE_p, with $D_0 \geq nkE_p$ and

$$D_0^2 \;\geq\; \frac{32\, A_*\, r^4 n^2 \ln(8r^{n+2})}{\epsilon^2}, \qquad \text{where} \quad A_* = 12(1+8E_p^2 n^2 L_*), \quad L_* = \ln(8D_0 r^n E_p^2). \tag{19}$$

Put $q_0 = k'/k$. For $t \geq 1$ let $\widetilde{D}_t = q_0 D_{t-1}$, a multiple of $k'E_p$, and let D_t be the least multiple of kE_p with $D_t \geq \widetilde{D}_t$. Then $D_t \leq q_0 D_{t-1} + kE_p$, so $D_n \leq q_0^n(D_0 + nkE_p) \leq 2q_0^n D_0$. The memory has $Q = \lceil \log D_n \rceil$ qubits, of which the last $\lceil h/2 \rceil$ form a *port*. A space whose dimension D' is a multiple of E_p is realized as a subspace of the first $Q - \lceil h/2 \rceil$ qubits of dimension D'/E_p, tensored with the port. The active space of round t is such a space of dimension D_{t-1}, identified with $\mathbb{C}^k \otimes \mathbb{C}^{D_{t-1}/k}$.

The device never leaves the active spaces. The memory starts maximally mixed on the active space of round 1, of dimension D_0. In round t the device applies a unitary V_t of the active space and then $W \otimes I_{D_{t-1}/k}$, which maps the input and the active space isometrically onto the output and a space of dimension $\widetilde{D}_t$. It releases the output. It then applies a unitary V_t' of that space; since V_t' acts on the memory alone, it could equally act before the release. If $t < n$, the device exports j_t port qubits and imports j_t qubits in the maximally mixed state in their place. The resulting space of dimension $\widetilde{D}_t$ lies in the active space of round $t+1$. These maps are completed arbitrarily to unitaries of the input and the whole memory; the completion never acts, since the memory never leaves the active spaces. So $C = 0$, $J = J_{n-1}$, and $P = Q - \log D_0 \leq n \log q_0 + 2$.

Frames track the histories. Purify the initial memory: its frame F_0 is one unit vector, with $\alpha(F_0) = 1/D_0$. Let F be the frame before round t, with one column F_y for each history $y \in [r]^{t-1}$. After the call the frame has the columns $(B_i \otimes I)(V_t \otimes I)F_y$, indexed by the histories (y,i). Its Gram matrix has the blocks $F^*((V_t^*(B_i^*B_j \otimes I)V_t) \otimes I)F$, whose Haar means are $\tau(B_i^*B_j \otimes I)F^*F = \delta_{ij}F^*F$ by the first identity in (10). By the second identity and cyclicity of the partial trace over the memory, the exterior marginal after the call is exactly r times the one before, whatever V_t is; so the call multiplies α by r.

Choose the unitaries round by round. We choose $V_1, V_1', V_2, \ldots$ in turn so that the frame F_t after round t satisfies

$$\|F_t^*F_t - I\| \leq t\epsilon/n, \qquad \alpha(F_t) \leq A_*/D_0. \tag{20}$$

This holds for $t = 0$. Suppose it holds for $t-1$; then $g(F_{t-1}) \leq 2$. Since $B_i^*B_i \leq rI$, the Hermitian and anti-Hermitian parts of the $B_i^*B_j$ have norm at most r. Apply the concentration lemma to these $2r^2$ operators, tensored with $I_{D_{t-1}/k}$, with $D = D_{t-1} \geq D_0$, at most r^n columns and level $s = \epsilon/(2rn)$: by (19) the failure probabilities add up to at most $4r^{n+2}\exp[-D_0^2\epsilon^2/(32A_*r^4n^2)] \leq \frac{1}{2}$. So some V_t makes every block of the new Gram matrix deviate from its mean by at most $\epsilon/(rn)$, hence the $r \times r$ block matrix by at most ϵ/n, and the Gram matrix after the call is within $t\epsilon/n$ of I. Exports and imports preserve it. Choose V_t' by the export lemma with $e = e_t$, $\eta = 1/n$ and $D = \widetilde{D}_t \geq D_0$. The exterior has dimension $S \leq D_0\, 2^{2J_{t-1}} \leq D_0 r^n$, so $\ln(8Se_t^2) \leq L_*$, and

$$\alpha(F_t) \;\leq\; \Big(1 + \frac{1}{n}\Big)\frac{r\,\alpha(F_{t-1})}{e_t^2} + \frac{8E_p^2 n L_*}{D_0}.$$

Put $c_t = 2^{2J_t} r^{-t}$, so that $c_0 = 1$, $\frac{1}{4} < c_t \le 1$ and $c_t r / e_t^2 = c_{t-1}$. Multiplying by c_t and unrolling gives $c_t \alpha(F_t) \le (1 + \frac{1}{n})^t (1 + 8E_p^2 n^2 L_*)/D_0$, hence $\alpha(F_t) \le 4e(1 + 8E_p^2 n^2 L_*)/D_0 \le A_*/D_0$. In round n there is no export, and only the Gram bound is needed.

The error is at most ϵ. Purify the observer. In the ideal experiment, with the Stinespring isometry $\xi \mapsto \sum_i A_i \xi \otimes |i\rangle$ in each round, its final vector is $\sum_y \psi_y \otimes |y\rangle$. In the service it is $\sum_y \psi_y \otimes F_n |y\rangle$ with the same vectors ψ_y, because the device acts on the input and the memory through $W = \sum_i A_i \otimes B_i$ and unitaries of the memory, and the observer never touches the memory or the exterior. The polar part $T = F_n (F_n^* F_n)^{-1/2}$ is an isometry with $\|F_n - T\| = \|(F_n^* F_n)^{1/2} - I\| \le \epsilon$. So the actual final vector is within ϵ of the ideal one mapped by $I \otimes T$, and after tracing out the memory and the exterior the observer's final states are within ϵ in trace distance.

The memory is as claimed. We have $Q \le \log D_n + 1 \le \log D_0 + n \log q_0 + 2$. The conditions on D_0 hold for $D_0 = kE_p \lceil K_r (n/\epsilon)^4 (1 + \ln k) \rceil$ with K_r large enough in terms of r, and then $\log D_0 = \log k + O_r(\log(n/\epsilon) + \log(2 + \log k))$. □

E The lazy doubling in detail

This appendix proves parts (b) and (c) of Proposition 15.6, used in Section 15.3, and Proposition F.3, used in Appendix F; it also records two remarks on Θ_+. Part (b) needs the one-round theorem in a commutator form (Lemma E.1), a weighted intertwining lemma (Lemma E.2) and a block extraction with a smooth cutoff (Lemma E.3).

Throughout, Φ is a general unital channel on M_d, with doubling $\Theta\left(\begin{smallmatrix} x & y \\ u & w \end{smallmatrix}\right) = \operatorname{diag}(\Phi(w), \Phi^*(x))$ and $\Theta_+ = \frac{1}{2}(\mathrm{id} + \Theta)$ on M_{2d}. Let $\mathcal{S}_\Phi \subseteq M_d$ be the span of the Kraus operators of Φ, so that the Choi support of Φ is $\{|A\rangle\rangle : A \in \mathcal{S}_\Phi\}$, and let Π_Φ also denote the Hilbert–Schmidt projection of M_d onto $\mathcal{S}_\Phi$. Since $\Theta(X) = \sum_g (e_{01} \otimes A_g) X (e_{01} \otimes A_g)^* + (e_{10} \otimes A_g^*) X (e_{10} \otimes A_g^*)^*$ for Kraus operators A_g of Φ, the Choi support of Θ_+ corresponds to the span of the three pairwise orthogonal subspaces

$$\mathbb{C}\, 1_{2d}, \qquad e_{01} \otimes \mathcal{S}_\Phi, \qquad e_{10} \otimes \mathcal{S}_\Phi^* \tag{21}$$

of $M_{2d} = M_2 \otimes M_d$. As in Appendix C, $\|\cdot\|_F$ is the Hilbert–Schmidt norm and $\|\cdot\|$ the operator norm. An operator X on the bath $\mathbb{C}^M$ is identified with $1 \otimes X$ when it acts on $\mathbb{C}^d \otimes \mathbb{C}^M$, so that $\|1 \otimes X\|_F = \sqrt{d}\, \|X\|_F$.

Lemma E.1 (Commutator form of the one-round theorem)**.** *Lemma C.1, Theorem 10.1 and Corollary 10.2 remain true if, in the definition $\nu = \ell + \Delta S$, the entropy production ΔS of the round $\Psi(X) = \operatorname{Tr}_M U(X \otimes \rho) U^*$ is replaced by any $\tilde{s} \ge 0$ with $\|[U, 1 \otimes \sqrt{\rho}]\|_F^2 \le 4d\, \tilde{s} \ln 2$, and if the hypothesis $d_\diamond(\Psi, \Phi_G) \le 1/8$ is replaced by the weaker hypothesis*

$$|\langle z | \Psi(|z\rangle\langle 0|) |0\rangle| \le \tfrac{3}{4}, \tag{22}$$

where 0 and z are the anchor positions of the empty word and of j; in Lemma C.1(iii) the bound then reads $E(\phi(a)\phi(b)) \ge 0.66 - e$. Conversely, every round can be rewritten, without changing its channel or its leakage, as a round (U', ρ) with $U' = (1 \otimes W)U$ for which $\tilde{s} = \Delta S$ is admissible.

Proof. The rewriting is the normalization step of the proof of Lemma C.1, after which (14) is the inequality $\|[U', 1 \otimes \sqrt{\rho}]\|_F^2 \le 4d\, \Delta S \ln 2$; its proof uses only that U' is a unitary on $\mathbb{C}^d \otimes \mathbb{C}^M$. In the proof of Lemma C.1, ΔS enters only through (14). Afterwards only the bounds $\epsilon_S \le 2\sqrt{\ln 2}\, \eta$, $\epsilon_\ell + \epsilon_S \le \sqrt{1 + 4 \ln 2}\, \eta$ and $2\epsilon_\ell^2 + 2\epsilon_S^2 \le 8 \ln 2\, \eta^2$ with $\eta = \sqrt{d\nu}$ are used, together with the unitarity of U, its leakage and $d_\diamond(\Psi, \Phi_G)$. These bounds hold with $\epsilon_S^2 \le 4d\tilde{s} \ln 2$ and $\nu = \ell + \tilde{s}$. The diamond hypothesis enters only in the

displacement step, through $|\operatorname{Tr}(\rho B_\varnothing^* B_j)| = |\langle z|\Psi(|z\rangle\langle 0|)|0\rangle| \le 2u \le \frac14$. Under (22) instead, $E(\phi(j))^2 \ge 2 - \frac32 - 4\epsilon_\ell \ge \frac12 - 4\eta$, so $E(\phi(a)\phi(b)) \ge \sqrt{1/2 - 4\eta} - e \ge 0.66 - e$ if $\eta \le \frac{1}{64}$, while $e > 1$ otherwise. In the proof of Theorem 10.1 this gives $E(Z_i) \ge 0.66 - \delta_L$ and $\Delta_i \ge 0.66 - 30\delta_L \ge 0.66 - b_* > \frac12$ in place of $\Delta_i \ge 1 - b_*$, and $\Delta_i \ge \frac12$ is all that Corollary C.7 needs. Theorem 10.1 and Corollary 10.2 use the round only through Lemma C.1. □

Lemma E.2 (Weighted intertwining)**.** *Let B be an operator on $\mathbb{C}^d \otimes \mathbb{C}^M$ with $\|B\| \le 1$, let R, R' be self-adjoint operators on $\mathbb{C}^M$, ξ an operator on $\mathbb{C}^M$, and $\Sigma = BR - R'B$.*

(i) If h is Lipschitz with constant L_h on an interval containing the spectrum of R, then $\|[h(R), \xi]\|_F \le L_h \, \|[R, \xi]\|_F$.

(ii) If $g(x) = \int \hat g(t) e^{itx} dt$ with $M_1 = \int |t\,\hat g(t)|\, dt$ and $M_2 = \int t^2 |\hat g(t)|\, dt$ finite, then

$$\big\|(Bg(R) - g(R')B)\xi\big\|_F \;\le\; M_1 \, \|\Sigma\xi\|_F + \tfrac12 M_2 \, \|\Sigma\| \, \sqrt{d} \, \|[R, \xi]\|_F .$$

Proof. (i) In an eigenbasis of R with eigenvalues r_k, the entries of $[h(R), \xi]$ and $[R, \xi]$ are $(h(r_k) - h(r_l))\xi_{kl}$ and $(r_k - r_l)\xi_{kl}$. (ii) Differentiating $\theta \mapsto e^{i\theta t R'} B e^{i(1-\theta)tR}$ gives $Be^{itR} - e^{itR'}B = it \int_0^1 e^{i\theta t R'} \, \Sigma \, e^{i(1-\theta)tR} \, d\theta$. Write $e^{i(1-\theta)tR}\xi = \xi e^{i(1-\theta)tR} + [e^{i(1-\theta)tR}, \xi]$ and use (i) with $h(x) = e^{i(1-\theta)tx}$; unitaries do not change Hilbert–Schmidt norms, so $\|(Be^{itR} - e^{itR'}B)\xi\|_F \le |t| \, \|\Sigma\xi\|_F + \frac{t^2}{2} \|\Sigma\| \sqrt{d} \, \|[R, \xi]\|_F$. Integrate against $|\hat g(t)|\, dt$. □

We fix the cutoff once and for all. Let $S : \mathbb{R} \to [0, 1]$ be smooth with $S = 0$ on $(-\infty, 0]$ and $S = 1$ on $[1, \infty)$, let $0 < s_0 \le \frac14$, and put

$$g(x) = \sin\big(\tfrac{\pi}{2} S(2x/s_0 - 1)\big) \, S(3 - 2x), \qquad g_2(x) = \cos\big(\tfrac{\pi}{2} S(2x/s_0 - 1)\big), \qquad f(x) = \sqrt{x}\, g(x).$$

Then g is smooth with compact support, so $M_g := \int (|t| + t^2)|\hat g(t)|\, dt < \infty$; on $[0, 1]$ we have $0 \le g \le 1$, $g = 0$ on $[0, s_0/2]$, $g = 1$ and $g_2 = 0$ on $[s_0, 1]$, and $g^2 + g_2^2 = 1$. Let L_g be a common Lipschitz constant of g and g_2 on $[0, 1]$, put $\sigma = \sqrt{2/s_0}$ and

$$K_a = 11 M_g + 1 + 1.5 L_g + 2\sigma(4 + 1.5 L_g).$$

These depend only on s_0.

Lemma E.3 (Block extraction)**.** *Let Φ be a unital channel on M_d, and let $\Psi(X) = \operatorname{Tr}_M U(X \otimes \rho)U^*$ be a round of Θ_+, with U unitary on $\mathbb{C}^2 \otimes \mathbb{C}^d \otimes \mathbb{C}^M$, $u = d_\diamond(\Psi, \Theta_+)$ and leakage ℓ relative to Θ_+. Put $\epsilon_\ell = \sqrt{2d\ell}$, $\epsilon_S = \|[U, 1 \otimes \sqrt{\rho}\,]\|_F$, $\gamma = \epsilon_\ell + \epsilon_S$ and $\mathcal{E} = 2u + s_0 + 4K_a\sqrt{d}\,\gamma$. If $\mathcal{E} < \frac12$, there are a unitary U' on $\mathbb{C}^d \otimes \mathbb{C}^M$ and a state ρ' on $\mathbb{C}^M$ such that the round $\Psi'(X) = \operatorname{Tr}_M U'(X \otimes \rho')U'^*$ of Φ, with leakage $\ell_\Phi(\Psi')$ relative to Φ, satisfies*

$$d_\diamond(\Psi', \Phi) \le \frac{\mathcal{E}}{\frac12 - \mathcal{E}}, \qquad \ell_\Phi(\Psi') \le \frac{4K_a^2\gamma^2}{d(\frac12 - \mathcal{E})}, \qquad \big\|[U', 1 \otimes \sqrt{\rho'}\,]\big\|_F^2 \le \frac{8K_a^2\gamma^2}{\frac12 - \mathcal{E}},$$

and $|\langle x|\Psi'(|x\rangle\langle y|)|y\rangle| \le \mathcal{E}/(\frac12 - \mathcal{E})$ for all unit vectors x, y with $\langle x|\Phi(|x\rangle\langle y|)|y\rangle = 0$.

Proof. Put $\xi = \sqrt{\rho}$ and write $U = \sum_{a,b\in\{0,1\}} e_{ab} \otimes U_{ab}$ with $D_0 = U_{00}$, $B = U_{01}$, $D_1 = U_{11}$, operators on $\mathbb{C}^d \otimes \mathbb{C}^M$.

Leakage bounds the residuals. As in the proof of Lemma C.1, $\mathcal{J}(\Psi)$ is the marginal of the unit vector $(2d)^{-1/2} U\xi \in M_{2d} \otimes M_M$, so $\ell = (2d)^{-1}\|(\Pi^\perp \otimes \mathrm{id})(U\xi)\|_F^2$, with Π the projection of M_{2d} onto the span of (21). Put $\alpha = (2d)^{-1}(\operatorname{Tr} \otimes \mathrm{id})(U)$, the average of the $2d$ diagonal entries of U in M_M, and $E_a = D_a - \alpha$, $F = (\Pi_\Phi^\perp \otimes \mathrm{id})(B)$. By (21), $(\Pi^\perp \otimes \mathrm{id})(U)$ has the diagonal blocks E_0, E_1 and the upper right block F; since $\Pi^\perp \otimes \mathrm{id}$ commutes with right multiplication by ξ, the numbers $e_a = \|E_a\xi\|_F$ and $e_F = \|F\xi\|_F$ are at most ϵ_ℓ. Also $\|\alpha\| \le 1$ and $\|E_a\| \le 2$.

The blocks nearly commute with ξ. Put $k_X = \|[X,\xi]\|_F$ for $X \in \{D_0, B, D_1\}$; these are at most ϵ_S. The diagonal entries of U are entries of D_0 and D_1, so by Cauchy–Schwarz $k_\alpha := \sqrt{d}\,\|[\alpha,\xi]\|_F \le (k_{D_0}^2 + k_{D_1}^2)^{1/2}/\sqrt{2} \le \epsilon_S/\sqrt{2}$. With $R = 1-\alpha^*\alpha$ and $R' = 1-\alpha\alpha^*$, both between 0 and 1, $k_R := \sqrt{d}\,\|[R,\xi]\|_F \le 2k_\alpha \le \sqrt{2}\,\epsilon_S$, and likewise $k_{R'} \le \sqrt{2}\,\epsilon_S$.

The unitarity defects are small. Unitarity gives $B^*B = 1 - D_1^*D_1$ and $BB^* = 1 - D_0D_0^*$, hence

$$\Delta := R - B^*B = \alpha^*E_1 + E_1^*D_1, \qquad \Delta' := R' - BB^* = \alpha E_0^* + E_0D_0^*,$$
$$\Sigma := BR - R'B = B\Delta - \Delta'B,$$

with $\|\Delta\|, \|\Delta'\| \le 1$ and $\|\Sigma\| \le 2$. From $\xi E_1 = E_1\xi - [D_1,\xi] + [\alpha,\xi]$ and $D_1\xi = \xi D_1 + [D_1,\xi]$, $\|\Delta\xi\|_F \le e_1 + \|\xi E_1\|_F + 2k_{D_1} \le 2e_1 + 3k_{D_1} + k_\alpha \le 4\gamma$. Symmetrically $\|\Delta'\xi\|_F \le 2e_0 + 3k_{D_0} + k_\alpha \le 4\gamma$, and $\Sigma\xi = B\Delta\xi - \Delta'\xi B - \Delta'[B,\xi]$ gives $\|\Sigma\xi\|_F \le 9\gamma$.

Z_0 factors in two ways. Let P and P' be the spectral projections of R and R' for $(s_0/2,\infty)$. Since g vanishes on $[0,s_0/2]$, $g(R) = R^{-1/2}Pf(R)$ with $\|R^{-1/2}P\| \le \sigma$, and likewise for R'. The operators $B_1 = BR^{-1/2}P$ and $B_1' = R'^{-1/2}P'B$ satisfy

$$B_1^*B_1 = P - \tilde\Delta, \quad \tilde\Delta = PR^{-1/2}\Delta R^{-1/2}P, \qquad B_1'B_1'^* = P' - \tilde\Delta', \quad \tilde\Delta' = P'R'^{-1/2}\Delta' R'^{-1/2}P'.$$

Take polar decompositions $B_1 = V_1|B_1|$ and $B_1' = |B_1'^*|V_1'$ with V_1, V_1' unitary. On the range of $1\otimes P$ we have $|B_1| = 1 + \varphi(\tilde\Delta)$ with $\varphi(x) = \sqrt{1-x} - 1$ and $\tilde\Delta \le 1$, and $\varphi(x)^2 \le x^2$; so $\|(|B_1| - P)Y\|_F \le \|\tilde\Delta Y\|_F$ for every Y with range in that subspace, and similarly for B_1'. Put $Z_0 = Bg(R)\xi = B_1f(R)\xi$. Then $Z_0 - V_1f(R)\xi = V_1(|B_1| - P)f(R)\xi$, and $\tilde\Delta f(R)\xi = PR^{-1/2}\Delta g(R)\xi$, while $\|\Delta g(R)\xi\|_F \le \|\Delta\xi\|_F + \sqrt{d}\,\|[g(R),\xi]\|_F \le \|\Delta\xi\|_F + L_g\,k_R$ by Lemma E.2(i). Hence

$$a_1 := \|Z_0 - V_1f(R)\xi\|_F \le \sigma(4 + 1.5L_g)\gamma.$$

On the other side, $Z_0 = (Bg(R) - g(R')B)\xi + g(R')[B,\xi] + [g(R'),\xi]B + \xi f(R')B_1'$ and $\xi f(R')B_1' = \xi f(R')V_1' + \xi f(R')(|B_1'^*| - P')V_1'$, where $\|\xi f(R')(|B_1'^*| - P')\|_F \le \|\tilde\Delta' f(R')\xi\|_F \le \sigma(\|\Delta'\xi\|_F + L_g\,k_{R'})$. By Lemma E.2(ii), $\|(Bg(R) - g(R')B)\xi\|_F \le M_g(\|\Sigma\xi\|_F + k_R) \le 11M_g\gamma$. Hence

$$a_2 := \|Z_0 - \xi f(R')V_1'\|_F \le 11M_g\gamma + (1 + 1.5L_g)\gamma + \sigma(4 + 1.5L_g)\gamma, \qquad a_1 + a_2 \le K_a\gamma.$$

The polar part of Z_0 nearly intertwines ζ_1 and ζ_2. Put $\zeta_1 = |f(R)\xi| = (\xi f(R)^2\xi)^{1/2}$ and $\zeta_2 = |(\xi f(R'))^*| = (\xi f(R')^2\xi)^{1/2}$. Then $|V_1f(R)\xi| = \zeta_1$ and $|(\xi f(R')V_1')^*| = \zeta_2$, as operators $1\otimes\zeta_i$. By the Araki–Yamagami inequality $\||X| - |Y|\|_F \le \sqrt{2}\,\|X - Y\|_F$ [3], applied to Z_0 and to Z_0^*, $\||Z_0| - \zeta_1\|_F \le \sqrt{2}\,a_1$ and $\||Z_0^*| - \zeta_2\|_F \le \sqrt{2}\,a_2$. Let $Z_0 = O|Z_0| = |Z_0^*|O$ with O unitary. Then

$$\|O\zeta_1 - \zeta_2O\|_F \le \|O(\zeta_1 - |Z_0|)\|_F + \|(|Z_0^*| - \zeta_2)O\|_F \le \sqrt{2}(a_1 + a_2) =: \delta.$$

Build the new round. The Hermitian operators $O(1\otimes\zeta_1)O^*$ and $1\otimes\zeta_2$ are within δ in $\|\cdot\|_F$, and their spectra are those of ζ_1 and ζ_2, each repeated d times. By the Hoffman–Wielandt inequality [43], the decreasing eigenvalue lists (x_k) of ζ_1 and (y_k) of ζ_2 satisfy $d\sum_k(x_k - y_k)^2 \le \delta^2$; so the unitary W_h of $\mathbb{C}^M$ that maps an eigenbasis of ζ_1 to one of ζ_2 in the same order has $\|1\otimes(W_h\zeta_1W_h^* - \zeta_2)\|_F \le \delta$. Put $U' = (1\otimes W_h^*)O$, $c_1 = \operatorname{Tr}\zeta_1^2 = \operatorname{Tr}\rho f(R)^2$ and $\rho' = \zeta_1^2/c_1$. Then $\|[U', 1\otimes\zeta_1]\|_F \le \|O\zeta_1 - \zeta_2O\|_F + \|(\zeta_2 - W_h\zeta_1W_h^*)O\|_F \le 2\delta$, which is the third bound once we know $c_1 \ge \frac12 - \mathcal{E}$. The channel of (U',ρ') is $\Psi' = \Phi_{O\zeta_1}/c_1$, where $\Phi_X(w) = \operatorname{Tr}_M[X(w\otimes 1)X^*]$. Its leakage is $(dc_1)^{-1}\|(\Pi_\Phi^\perp\otimes\mathrm{id})(O\zeta_1)\|_F^2$, and

$$\|(\Pi_\Phi^\perp\otimes\mathrm{id})(O\zeta_1)\|_F \le \|(\Pi_\Phi^\perp\otimes\mathrm{id})(Z_0)\|_F + \sqrt{2}\,a_1 \le e_F + L_g\,k_R + \sqrt{2}\,a_1 \le 2K_a\gamma,$$

since $(\Pi_\Phi^\perp \otimes \mathrm{id})(Bg(R)\xi) = F\xi\, g(R) + (\Pi_\Phi^\perp \otimes \mathrm{id})(B[g(R),\xi])$. This is the second bound.

The new round is close to Φ. For the completely positive maps Φ_X just defined, $\|\Phi_X\|_\diamond = \|\mathrm{Tr}_M(X^*X)\|$, and $\|\Phi_X - \Phi_Y\|_\diamond \le \|X-Y\|_F(\|\mathrm{Tr}_M X^*X\|^{1/2} + \|\mathrm{Tr}_M Y^*Y\|^{1/2})$, where Tr_M traces out the input bath. Let $Y = B\xi$. The upper left block of $\Psi(e_{11}\otimes w)$ is $\Phi_Y(w)$ and that of $\Theta_+(e_{11}\otimes w)$ is $\frac12\Phi(w)$; compressing to the block, $\|\Phi_Y - \frac12\Phi\|_\diamond \le 2u$. Since $g^2+g_2^2 = 1$ on $[0,1]$, $\Phi_Y - \Phi_{Z_0} - \Phi_{Bg_2(R)\xi}$ is $w \mapsto \mathrm{Tr}_M[B(w\otimes X)B^*]$ with $X = \sum_{h\in\{g,g_2\}} h(R)[h(R),\rho]$, and $\|X\|_1 \le 4L_g\|[R,\xi]\|_F$, so its diamond norm is at most $4L_g\, k_R/\sqrt{d}$, by splitting X into positive and negative parts and using $B^*B \le 1$. Next, $\|\Phi_{Bg_2(R)\xi}\|_\diamond \le \mathrm{Tr}\,\rho\, g_2(R)Rg_2(R) + \|\xi g_2(R)\Delta g_2(R)\xi\|_1 \le s_0 + \sqrt{d}(\|\Delta\xi\|_F + L_g\, k_R)$, because $xg_2(x)^2 \le s_0$ on $[0,1]$. And $\mathrm{Tr}_M\zeta_1^2 = c_1 \le 1$ and $\mathrm{Tr}_M Z_0^*Z_0 \le 1$ give $\|\Phi_{O\zeta_1} - \Phi_{Z_0}\|_\diamond \le 2\sqrt{2}\, a_1$. Adding up,

$$\|\Phi_{O\zeta_1} - \tfrac12\Phi\|_\diamond \le 2u + s_0 + 2\sqrt{2}\, a_1 + \sqrt{d}(4 + 1.5L_g)\gamma + 6L_g\gamma/\sqrt{d} \le \mathcal{E}.$$

Evaluating at τ_d gives $|c_1 - \frac12| \le \mathcal{E}$, so $c_1 \ge \frac12 - \mathcal{E}$, and $\|\Psi' - \Phi\|_\diamond \le \mathcal{E}/c_1 + |1/(2c_1) - 1| \le 2\mathcal{E}/c_1$. If $\langle x|\Phi(|x\rangle\langle y|)|y\rangle = 0$, then $|\langle x|\Phi_{O\zeta_1}(|x\rangle\langle y|)|y\rangle| \le \|\Phi_{O\zeta_1} - \frac12\Phi\|_\diamond \le \mathcal{E}$, and dividing by c_1 gives the last bound. □

Proof of Proposition 15.6(b). Let a service of Θ_+ for n rounds have error $\epsilon \le \frac{1}{16}$ and $P + C \le B$ with $\log n \le B$. Theorem 9.1, applied to Θ_+ on M_{1746}, gives a round with bath $\mathbb{C}^M$, $M = 2^Q$, $u = d_\diamond(\Psi_t, \Theta_+) \le \frac{1}{15}$ and $\nu_t = \ell_t + \Delta S_t \le t_n \le \frac{16}{15}\cdot\frac{B+1+\log n}{n} \le \frac{3.2\,B}{n}$, as in the proof of Theorem 11.1. The normalization step of the proof of Lemma C.1, which uses only that U_t is a unitary on $\mathbb{C}^{2d}\otimes\mathbb{C}^M$, replaces U_t by $(1\otimes W)U_t$ without changing Ψ_t or its leakage, after which $\epsilon_S^2 = \|[U_t, 1\otimes\sqrt{\rho_t}\,]\|_F^2 \le 4(2d)\Delta S_t \ln 2$; so $\gamma^2 \le 2(2d\ell_t + 8d\ln 2\,\Delta S_t) \le 12d\,\nu_t$. Apply Lemma E.3 with $s_0 = \frac{1}{100}$ and $d = 873$: $\mathcal{E} \le \frac{2}{15} + \frac{1}{100} + 14K_a d\sqrt{\nu_t} \le \frac{3}{14}$ as soon as $14K_a d\sqrt{3.2B/n} \le 0.07$. Then $c_1 \ge \frac12 - \mathcal{E} \ge \frac27$, and the anchors 0 and $z = z(j)$ satisfy $\langle z|\Phi(|z\rangle\langle 0|)|0\rangle = \tau(\lambda([j])) = 0$, so the round Ψ' satisfies (22). By Lemma E.1 it may be used in Corollary 10.2 with

$$\nu' = \ell_\Phi(\Psi') + \frac{\|[U', 1\otimes\sqrt{\rho'}]\|_F^2}{4d\ln 2} \le \Big(4 + \frac{2}{\ln 2}\Big)\frac{K_a^2\gamma^2}{d\,c_1} \le 290K_a^2\,\nu_t \le 928K_a^2\,\frac{B}{n}.$$

Let C_R and y_0 be the constants of Corollary 10.2(a) for R_Φ and the transport of Section 11, and put $y_n = n/(928K_a^2C_RB)$, so that $C_R\nu' y_n \le 1$ and $y_n \le n$. With $c_0 = \min\{1/(928K_a^2C_Ry_0),\ (0.07/(14K_ad))^2/3.2\}$ and $B \le c_0 n$, the corollary gives

$$Q = \log M \ \ge\ S(\rho') \ \ge\ \frac{y_n}{32(\log y_n)^{2\kappa}} \ \ge\ \frac{1}{29696\,K_a^2C_R}\cdot\frac{n}{B(\log n)^{2\kappa}}.$$

For the second statement take $B = \max(1, A_0)\log n$. □

Proof of Proposition 15.6(c). Let E, K_G and a $(1-r^{-1})$-expansive r^{-1}-morphism $f : E \to \mathrm{Sym}(\Omega)$ be as in Proposition 12.1, with its unitary $U_f = \sum_a c_aE_a\otimes P_{\pi_a}$ and its set Ω_0 of good points, at which $\mathrm{Tr}_\Omega U_f(w\otimes|\omega\rangle\langle\omega|)U_f^* = \Phi(w)$. Let Ω_1 be the set of ω with $(f([x])f([y]))^{-1}(\omega) = f([z])^{-1}(\omega)$ for every triangle (x,y,z) and $f(g)^{-1}(\omega) \ne f(h)^{-1}(\omega)$ for distinct labels g,h. As $U_f^*(\psi\otimes|\omega\rangle) = \sum_a \overline{c_a}E_a^*\psi\otimes|\pi_a^{-1}(\omega)\rangle$, the computation of Proposition 12.1 gives $\mathrm{Tr}_\Omega U_f^*(x\otimes|\omega\rangle\langle\omega|)U_f = \Phi^*(x)$ for $\omega\in\Omega_1$. A point outside Ω_1 is the image under the bijection $f([z])$ of a point where $f([x])f([y])$ and $f([z])$ differ, or under $f(g)$ of a point where $f(g)$ and $f(h)$ agree; so $|\Omega\setminus\Omega_1| \le K_GM/r$, and $\Omega_* = \Omega_0\cap\Omega_1$ has $|\Omega_*| \ge (1-2K_G/r)M$.

Let the bath be a coin qubit, a phase qubit carrying $z = \mathrm{diag}(1,i)$, and $\mathbb{C}^\Omega$, and put $W_f = |0\rangle\langle 0|\otimes 1 + |1\rangle\langle 1|\otimes V_f$ with $V_f = \left(\begin{smallmatrix} 0 & U_f\otimes z \\ U_f^*\otimes z^* & 0\end{smallmatrix}\right)$. For every state β that is a mixture of the $|\omega\rangle\langle\omega|$, $\omega\in\Omega_*$, the bath state $\tau_2\otimes\tau_2\otimes\beta$ makes W_f apply exactly Θ_+. The coin splits

the channel into $\frac{1}{2}\mathrm{id}$ and $\frac{1}{2}$ the channel of V_f, the phase qubit removes the off-diagonal blocks since $\tau(z^2) = 0$, and the diagonal blocks are $\Phi(w)$ and $\Phi^*(x)$ at the points of Ω_*. Now repeat the exact service of Proposition 12.1(c) with K_G replaced by $2K_G$ and two more qubits, which are imported maximally mixed. With $\eta = B/(4n)$ and $r = 2K_G/\eta$ the imported register is maximally mixed on $\mathrm{span}\{|\omega\rangle : \omega \in \Omega_*\} \otimes \mathbb{C}^{k_0}$ and on the two qubits, its deficit is below B/n, every round applies exactly Θ_+, and $Q = \lceil \log \sigma_E(8K_G n/B) + \log(4n/B) \rceil + 2$. The bound $\log \sigma_E(R) \le C_E R \log R$ of Theorem 3.3(b) gives $Q \le K(n/B)\log(2n/B)$. With β maximally mixed on all of Ω, the same unitary gives a finite tracial factorization of $(1-\vartheta)\Theta_+ + \vartheta\,\Upsilon$ for some channel Υ and some $\vartheta \le 2K_G/r$, for every r; hence $\Theta_+ \in \mathcal{K}_{1746}$. □

Remark E.4 (Exact finite factorizations of Θ_+). Θ_+ is an exact autonomous value but not an exact finite factor. Indeed, if $\Theta_+(X) = (\mathrm{id} \otimes \tau_m)[W(X \otimes 1)W^*]$ for a unitary W on $\mathbb{C}^{2d} \otimes \mathbb{C}^m$, the proof of Lemma E.3 with $\rho = \tau_m$ has $\ell = 0$ and $\epsilon_S = 0$: the diagonal blocks of W are both $1 \otimes \alpha$, its upper right block B lies in $\mathcal{S}_\Phi \otimes M_m$, $B^*B = 1 \otimes R$ and $BB^* = 1 \otimes R'$. Then $B(1 \otimes R) = (1 \otimes R')B$, so for every eigenvalue $\mu > 0$ of R the operator $\mu^{-1/2}B$ maps $\mathbb{C}^d \otimes \ker(R - \mu)$ unitarily onto $\mathbb{C}^d \otimes \ker(R' - \mu)$, and the upper left block of $\Theta_+(e_{11} \otimes w)$ writes Φ as the convex combination $\sum_\mu \frac{2\mu m_\mu}{m}\,\Phi_\mu$ of exact factors with baths of dimension $m_\mu = \dim\ker(R - \mu) \le m$. For the channel of Theorem 11.1 this is impossible: such a combination is a single round with $\ell = \Delta S = 0$, a direct sum of flat baths, and Theorem 10.1 with $\nu = 0$ would give it entropy at least $N/16$ for every N.

Remark E.5 (A lower exchange). At an anchor z of the gadget, $\Phi(|z\rangle\langle z|) = |z\rangle\langle z|$, so Θ_+ maps $|1\rangle\langle 1| \otimes |z\rangle\langle z|$ to the flat state $\frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|) \otimes |z\rangle\langle z|$ of rank two. Proposition 13.1 with $r = 2$ therefore gives $P + C < 3Q + 6$ for services of Θ_+ with error at most $1/16$ and exchange $J \le n/2$. Proposition 15.6(b), used as in the proof of Theorem 11.1(b) with $B = \max(3Q + 6, \log n)$, then gives $Q \ge c\sqrt{n}/(\log n)^\kappa$ for such services, whatever their purity. If $B > c_0 n$, then already $Q \ge (c_0 n - 6)/3$.

Proof of Proposition F.3. Let $\mathcal{F}$ be the set of labels, and write $A_g = \sum_{[w_a]=g} c_a E_a$, so that $\Phi_G\Phi_G^*(X) = \sum_{g,h} A_g A_h^* X A_h A_g^*$. A product $A_g A_h^*$ is a sum of terms $c_a\overline{c_b}E_a E_b^*$ over slots a, b in a common column, so it vanishes unless g and h label two slots of one column. In a triangle block with labels $1, y$ in the first row and x, z in the second, where $z = xy$, the columns carry the labels $\{1, x\}$ and $\{y, z\}$; so $A_g A_h^* \ne 0$ forces $hg^{-1} \in K = \{1\} \cup \{x^{\pm 1}\}$, the union over the triangles. For $k \in K$ let π_k be a permutation of $\mathcal{F}$ that extends the partial map $g \mapsto kg$, defined when $g, kg \in \mathcal{F}$, with $\pi_1 = \mathrm{id}$ and $\pi_{k^{-1}} = \pi_k^{-1}$. For $k \ne k^{-1}$ extend one of the two maps and invert it; for an involution k the partial map is an involution on its domain, which we extend by the identity.

The bath is $r_* + |K| - 1 \le 3r_* - 2$ qubits. On the first r_*, indexed by $\mathcal{F}$, let V_g be the Pauli matrix Z on the qubit g, so that $\tau(V_g V_h) = \delta_{gh}$, and let P_k permute these qubits by π_k, so that $P_k V_g P_k^* = V_{kg}$ whenever $g, kg \in \mathcal{F}$. On the other qubits choose diagonal unitaries χ_k with $\chi_1 = I$, $\chi_{k^{-1}} = \chi_k^*$ and $\tau(\chi_k^*\chi_l) = \delta_{kl}$: for an involution, Z on one qubit, and for a pair $k \ne k^{-1}$, $\mathrm{diag}(1, i, -1, -i)$ and its adjoint on two qubits, different pairs on different qubits. Put $R_k = P_k \otimes \chi_k$, so that $R_{k^{-1}} = R_k^*$ and $R_k V_g R_k^* = V_{kg}$ for $g, kg \in \mathcal{F}$, and put $C_{g,h} = V_g R^*_{hg^{-1}}$. These are orthonormal for the normalized trace: if $hg^{-1} \ne h'g'^{-1}$ the factor χ has trace zero, and otherwise $\tau(C_{g,h}^* C_{g',h'}) = \tau(R_k V_g V_{g'} R_k^*) = \delta_{gg'}$, which then forces $h = h'$. In the same way $\tau(C_{g,h} C_{g',h'}^*) = \tau(V_g V_{g'}) = \delta_{gg'}$ when $hg^{-1} = h'g'^{-1}$, and it vanishes otherwise.

Let $W = \sum A_g A_h^* \otimes C_{g,h}$, summed over the pairs with $hg^{-1} \in K$; the other products $A_g A_h^*$ vanish. On the block of a symbol with label g, including the first block, where $g = 1$, it is V_g. On a triangle block, with H the Hadamard matrix and using $R_x V_1 R_x^* = V_x$ and

$R_x V_y R_x^* = V_z$,

$$W = \frac{1}{2}\begin{pmatrix} V_1 + V_y & (V_1 - V_y)R_x^* \\ R_x(V_1 - V_y) & V_x + V_z \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & R_x \end{pmatrix}(H \otimes I)\begin{pmatrix} V_1 & 0 \\ 0 & V_y \end{pmatrix}(H \otimes I)\begin{pmatrix} I & 0 \\ 0 & R_x^* \end{pmatrix},$$

a self-adjoint unitary; triangles with coinciding labels are covered by the same computation. So W is unitary, and by orthonormality of the $C_{g,h}$, $(\mathrm{id}\otimes\tau)[W(X\otimes I)W^*] = \sum_{g,h} A_g A_h^* X A_h A_g^* = \Phi_G\Phi_G^*(X)$. For $\Phi_G^*\Phi_G$ use right translations $g \mapsto gk$ with $k \in \{1\} \cup \{y^{\pm 1}\}$, the operators $C_{g,h} = V_g R_{g^{-1}h}^*$, and rows instead of columns; on a triangle block the rows carry $\{1, y\}$ and $\{x, z\}$, and the same computation applies with x and y exchanged. For Θ^2, one further maximally mixed qubit dephases the two blocks, and controlled on the block the two factorizations act on a common bath. Refreshing the whole bath in every round gives the services. □

F Finer constraints on autonomous limits

This appendix complements Section 15.2: a return inequality that autonomous limits obey, an exact finite factor that it excludes, and the cheapness of adjoint compositions of gadget channels.

The limit $\overline{\Phi}$ of Proposition 15.3 is more than self-adjoint and positive. It is a compression of a conditional expectation, and this ties the decay of a coherence to the return probability of a state.

Proposition F.1 (A return inequality)**.** *Let $T = E \circ E_{\mathcal{N}} \circ \iota$ be as in Proposition 15.3, on M_D. If $T(e_{00}) = e_{00}$ and $T(e_{0j}) = a\, e_{0j}$ for some $j \neq 0$ and $0 \leq a \leq 1$, then*

$$\langle j|T(e_{jj})|j\rangle \;\geq\; a^2 + \frac{(1-a)^2}{D-1}.$$

Proof. Write $P = E_{\mathcal{N}}$, the orthogonal projection onto $L^2(\mathcal{N})$, and identify $x \in M_D$ with $\iota(x)$; then $\mathrm{tr}_D(x^*T(y)) = \langle x, Py\rangle$ with $\langle x, y\rangle = \tau(x^*y)$. Put $p_0 = e_{00}$, $p_j = e_{jj}$ and $x = e_{0j}$. Since $\|Pp_0\|_2^2 = \langle p_0, T(p_0)\rangle = \|p_0\|_2^2$, we have $Pp_0 = p_0$, so $p_0 \in \mathcal{N}$. Let $y = P(x)$, $X = P(p_j)$ and $B = y^*y$. As P is $\mathcal{N}$-bimodular, $p_0 y = P(p_0 x) = y$, $yp_0 = P(xp_0) = 0$ and $p_0 X = P(p_0 p_j) = 0$; so yy^* is supported in p_0, and B and X in $1 - p_0$. The Schwarz inequality for conditional expectations gives $B \leq P(x^*x) = X$. Moreover $\tau(X) = \tau(p_j) = 1/D$ and $\tau(B) = \|Px\|_2^2 = \langle x, T(x)\rangle = a/D$, so $R = X - B \geq 0$ has $\tau(R) = (1-a)/D$. By the Cauchy–Schwarz inequality, $\tau(B^2) = \tau((yy^*)^2) \geq \tau(yy^*)^2/\tau(p_0) = a^2/D$ and $\tau(R^2) \geq \tau(R)^2/\tau(1-p_0) = (1-a)^2/(D(D-1))$, and $\tau(BR) \geq 0$. Hence

$$\frac{1}{D}\,\langle j|T(e_{jj})|j\rangle = \langle p_j, Pp_j\rangle = \tau(X^2) = \tau(B^2) + 2\tau(BR) + \tau(R^2) \;\geq\; \frac{a^2}{D} + \frac{(1-a)^2}{D(D-1)}. \quad \square$$

Remark F.2 (Self-adjoint and positive is not enough). Haagerup and Musat give self-adjoint positive maps that are not such compressions because they are not factorizable [38, Theorem 5.4]. One exact finite factor is excluded all the same. Let $a = \sqrt{3}/2$ and $b = \frac{1}{2}$, and let Λ on M_3 have the Kraus operators $\mathrm{diag}(1, a, a)$, $b\, e_{12}$ and $-b\, e_{21}$. With $\omega = e^{2\pi i/3}$ and $W_k = 1 \oplus \left(\begin{smallmatrix} a & b\omega^k \\ -b\omega^{-k} & a \end{smallmatrix}\right)$ one has $\Lambda = \frac{1}{3}\sum_{k=0}^{2} \mathrm{Ad}\, W_k$, so Λ has a finite tracial factorization with bath dimension three. It is self-adjoint, and positive for the Hilbert–Schmidt inner product, with eigenvalues between $\frac{1}{2}$ and 1. But $\Lambda(e_{00}) = e_{00}$, $\Lambda(e_{01}) = a\, e_{01}$ and $\langle 1|\Lambda(e_{11})|1\rangle = a^2$, which violates Proposition F.1. Hence Λ is not the long-time limit of any autonomous dynamics with a tracial bath.

Composing a gadget channel with its adjoint erases the multiplication of the group. In a triangle block the unitarity of the gadget forces the coefficient of the product label to be the product of the coefficients, and this is what the one-round theorem reads. In $\Phi\Phi^*$ a triangle only requires a permutation π_x of the labels that carries 1 to x and y to xy, and no relation between the permutations π_x for different x is needed.

Proposition F.3 (Adjoint compositions of gadgets are exact finite factors)**.** *Let Φ_G be a gadget channel with r_* labels. Then $\Phi_G\Phi_G^*$ and $\Phi_G^*\Phi_G$ have finite tracial factorizations with baths of at most $3r_* - 2$ qubits. Consequently, for every n, they have exact causal services with $Q \le 3r_* - 2$, $P = C = 0$ and $J = (n-1)Q$, and so does Θ^2 in Remark 15.5, with one more qubit.*

The proof is in Appendix E. The exchange $(n-1)Q$ of these services is far above the rank rate, as for the services of Proposition 12.1. At its own rank rate $\Phi\Phi^*$ is cheap too: as an exact finite factor with a bath of at most $3r_* - 2$ qubits, it has services with $P = C = 0$ and memory $O(r_* + \log(n/\epsilon))$ by [27, Theorem 8.6]. An autonomous dynamics with a tracial bath therefore cannot be expensive through $\Phi\Phi^*$. It can be through a lazy version of the doubling (Section 15.3).

G Hyperlinearity transfers from F to $\mathcal{L}$

Proposition 8.5 states that $\mathcal{L} = (\bigoplus_{\mathbb{D}} S_3) \rtimes F$ is hyperlinear exactly when F is. This appendix proves it by embedding $L(\mathcal{L})$ into an ultrapower $L(F)^\omega$: a copy of S_3 is built from a localized sequence of II_1 subfactors that $F_{1/2}$ asymptotically fixes. Nothing else in the paper depends on this appendix.

Proof of Proposition 8.5. Subgroups of hyperlinear groups are hyperlinear, so suppose that F is hyperlinear. A countable group Γ is hyperlinear if and only if its von Neumann algebra $L(\Gamma)$ has a trace-preserving embedding into an ultrapower $\mathcal{R}^\omega$ of the hyperfinite II_1 factor [19]. Then every separable von Neumann subalgebra of $L(\Gamma)^\omega$ embeds in $\mathcal{R}^\omega$ as well, by a diagonal argument. It therefore suffices to embed $L(\mathcal{L})$ into $L(F)^\omega$ preserving the trace, for a free ultrafilter ω on $\mathbb{N}$. We write τ for the canonical traces and F_J for the subgroup of elements supported in an interval J. If J, J' are disjoint, $L(F_J)$ and $L(F_{J'})$ commute and $\tau(xy) = \tau(x)\tau(y)$ for $x \in L(F_J)$ and $y \in L(F_{J'})$, since F_J and $F_{J'}$ generate their direct product. If $f, f' \in F$ agree on J, then $\lambda(f)x\lambda(f)^{-1} = \lambda(f')x\lambda(f')^{-1}$ for $x \in L(F_J)$, because $f'^{-1}f$ fixes J pointwise and therefore commutes with F_J.

A localized copy of S_3. Choose $\sigma \in F$ supported in $[\frac{1}{2}, \frac{3}{4}]$, with $\sigma(x) < x$ on $(\frac{1}{2}, \frac{3}{4})$ and $\sigma(x) = \frac{1}{2} + \frac{1}{2}(x - \frac{1}{2})$ near $\frac{1}{2}$, and a dyadic $x_0 \in (\frac{1}{2}, \frac{3}{4})$. The intervals $J_k = \sigma^k((\sigma(x_0), x_0))$, $k \in \mathbb{Z}$, are disjoint and accumulate at $\frac{1}{2}$ as $k \to \infty$. For a nonempty open interval J with dyadic endpoints, $F_J \cong F$ has infinite conjugacy classes, since a nontrivial element has conjugates with infinitely many different supports, so $L(F_J)$ is a II_1 factor. In $L(F_{J_0})$ choose a unital copy of M_6 and, in its relative commutant, which is again a II_1 factor, a self-adjoint Z_0 with uniform distribution on $[0, 1]$. Let $v_0 : S_3 \to U(M_6)$ be the regular representation, so that $\tau(v_0(g)) = 0$ for $g \neq 1$, and put $v_k = \lambda(\sigma)^k v_0 \lambda(\sigma)^{-k}$ and $Z_k = \lambda(\sigma)^k Z_0 \lambda(\sigma)^{-k}$, which lie in $L(F_{J_k})$. By the factorization of traces, the Z_k are independent and uniform. For a finite $K \subset \mathbb{Z}$ and $k \in K$, let p_k^K be the spectral projection of the event that Z_k is the largest of the Z_j, $j \in K$. Ties have trace zero, so these projections are orthogonal and sum to 1, and they commute with every $v_j(g)$. Hence

$$V_K(g) = \sum_{k \in K} p_k^K v_k(g)$$

is a representation of S_3. Since $v_k(g)$ lies in a copy of M_6 whose relative commutant contains Z_k, and the other Z_j lie in algebras with factorized trace, $\tau(p_k^K v_k(g)) = \tau(p_k^K)\tau(v_0(g))$; so $\tau(V_K(g)) = 0$ for $g \neq 1$. If K is an interval of n integers and $|q| \leq n$, the two selected indices agree unless the largest Z_j over $K \cup (K+q)$ has its index outside $K \cap (K+q)$, an event of trace $2|q|/(n+|q|)$. On its complement $V_K(g) = V_{K+q}(g)$. Hence $\|V_{K+q}(g) - V_K(g)\|_2^2 \leq 8|q|/(n+|q|)$.

Invariance. Put $V_n = V_{\{n,\dots,2n-1\}}$, and let $h \in F_{1/2}$. Near $\frac{1}{2}$ on the right, $h(x) = \frac{1}{2} + 2^m(x - \frac{1}{2})$ for an integer m. Since σ has slope $\frac{1}{2}$ on the right of $\frac{1}{2}$, h agrees with σ^{-m} on J_k for all large k. Conjugation by $\lambda(h)$ therefore carries v_k and Z_k to v_{k-m} and Z_{k-m} for large k, and V_n to $V_{\{n,\dots,2n-1\}-m}$ for large n, so $\|\lambda(h)V_n(g)\lambda(h)^{-1} - V_n(g)\|_2 \to 0$. Thus $V(g) = (V_n(g))_\omega$ is a representation of S_3 in $L(F)^\omega$ with $\tau(V(g)) = 0$ for $g \neq 1$, and it commutes with $\lambda(F_{1/2})$.

The embedding. For $x \in \mathbb{D}$ choose $f_x \in F$ with $f_x(\frac{1}{2}) = x$ and put $V_x = \lambda(f_x)V\lambda(f_x)^{-1}$. By the invariance of V under $\lambda(F_{1/2})$, V_x does not depend on the choice of f_x, and $\lambda(f)V_x\lambda(f)^{-1} = V_{f(x)}$ for $f \in F$. The representatives $V_n(g)$ lie in $L(F_{I_n})$, where $I_n = \bigcup_{n \leq k < 2n} J_k$ shrinks to $\frac{1}{2}$ from the right, so those of V_x lie in $L(F_{f_x(I_n)})$. For $x \neq y$ the intervals $f_x(I_n)$ and $f_y(I_n)$ are disjoint for large n. So copies at distinct points commute, and the V_x together with λ define a homomorphism $\pi : \mathcal{L} \to U(L(F)^\omega)$. Let $\gamma = (\prod_{x \in X} \ell_x)f$, with X finite, ℓ_x in the copy of S_3 at x and $f \in F$. If $f = 1$ and some $\ell_x \neq 1$, the factors lie in commuting algebras with factorized trace for large n, so $\tau(\pi(\gamma)) = \prod_x \tau(V(\ell_x)) = 0$. If $f \neq 1$, choose a point moved by f outside X. For large n it lies outside the closure of every $f_x(I_n)$, so f is not in the group K_n generated by the $F_{f_x(I_n)}$. Then $\tau(a\lambda(f)) = 0$ for every $a \in L(K_n)$, in particular for the representatives of $\prod_x \pi(\ell_x)$. Hence $\tau(\pi(\gamma)) = 0$ whenever $\gamma \neq 1$, and π extends to a trace-preserving embedding of $L(\mathcal{L})$ into $L(F)^\omega$. □

H The grid host

This appendix is a sketch: its conclusions are conditional on proofs that we only outline, and they are stated in the indicative with that understanding. Nothing elsewhere in the paper depends on them. Placing the lamps on configurations of the cells of a grid, rather than of the points of Y, gives a stronger lower bound with the same upper bound. It also gives a nearly linear exponent $r^{1-o(1)}$ from Lee's exponential Dehn bound alone.

Remark H.1 (The grid host). Let $I = Y \times Y$, let $T = H \times H$ act on I coordinatewise, and let

$$\Gamma_2 = \Big(\bigoplus_{v \in V_2} S_3\Big) \rtimes (V_2 \rtimes (\mathrm{GL}_{\mathrm{fin}}(I) \rtimes T)), \qquad V_2 = \mathbb{F}_2^{(I)}.$$

It is locally finite-by-$\mathbb{Z}^4$ and generated by seven elements: the generators p_1, q_1, p_2, q_2 of the two copies of H, the translation t by the cell $o = ((1,1),(1,1))$, the element $c = d\,ab$, where d is the transvection that adds e_o to $e_{((1,2),(1,2))}$, and a. Transvections within a row or a column are not needed as generators, since commutators $[E_{x \leftarrow y}, E_{y \leftarrow z}] = E_{x \leftarrow z}$ of transvections of the third kind produce them. A certificate of 77 relations, of total length 5445 and maximum length 224, is listed in `verification/relators.txt` and checked by `verification/relators_check.py`. It has the families of Table 1, with two copies of the Houghton relators and their commutation, and with stabilizer and action relations for three kinds of transvection: within a row, within a column, and neither.

Transport now happens at scale $\sqrt{N}$. N cells fit in a square of side $\lceil \sqrt{N} \rceil$, so the transporters have length $O(\sqrt{N})$, and the local fillings cost $C(\delta_H(C\sqrt{N}) + N)$; the term N counts the swaps of letters from the two copies of H in a change of transporter. The clearing argument of Theorem 5.1 then bounds the pair areas by $CN^2(\delta_H(C\sqrt{N}) + N)$, which is at

most $CN^{2+p_H/2}$ by Proposition A.3(b) and at most $e^{C\sqrt{N}}$ by Proposition A.3(a). As in the proof of Theorem 3.3, the chunk E_2 of this certificate therefore has

$$\log \sigma_{E_2}(r) \;\geq\; c\,\frac{r}{(\log r)^{2+p_H/2}}, \qquad 2+\tfrac{p_H}{2} < 7.38,$$

and, using only Lee's bound, $\log \sigma_{E_2}(r) \geq c\,r\,e^{-C\sqrt{\log r}}$; unitary models need $\log D \geq c_\Delta e^{-2}/(\log(1/e))^{4+p_H}$. In the other direction, hashing the configurations directly, as in Proposition 6.3, needs a configuration space of dimension of order r^2. Hashing the cells first avoids this. A cell (y_1, y_2) gets the label $a_1(f_1^{-1}y_1) + a_2(f_2^{-1}y_2) \in \mathbb{F}_q$. Here the f_j are positions in finite Følner charts of H of size $e^{O(r\log r)}$, inverses of right Følner sets because the charts move by left multiplication, and the a_j are random tables on $O(r)$ points. The label is covariant under the charts, and it separates the finitely many cells of a chunk with probability $1 - O(1/q)$. The configurations are then compressed linearly to $\mathbb{F}_2^{\mathbb{F}_q}$ and hashed affinely as in Proposition 6.3. With q of order $\ell^2 r$, for a chunk of words of length at most ℓ, and the other parameters of order r, this gives $\log \sigma_{E'}(r) \leq C_{E'} r \log(2+r)$ for every chunk E' of Γ_2. The same holds for the d-dimensional grid host, built in the same way on the cells of Y^d with H^d acting, for every fixed d. Hence $\sigma_{E_2}(r) = \exp(r^{1+o(1)})$ as well, and no fixed-dimensional grid host exceeds $\exp(r^{1+o(1)})$. The proofs follow those of Sections 4–6 with the changes indicated.

Remark H.2 (The gadget channel of the grid host). Adding b and j and the four words of Section 11 to the certificate of Γ_2 gives $s = 9$, $r_0 = 81$ and $\Lambda = 5462$, so the gadget channel Φ_2 of Γ_2 acts on M_d with

$$d = 1 + 4\cdot 9 + 3\cdot 5462 - 4\cdot 81 = 16099,$$

and has $m = 224$. Of its 81 words, 78 have length at least three, and the next-to-last prefix of each of them represents the inverse of its last letter; $a = a^{-1}$, $b = b^{-1}$ and $j = ab$ give three more coincidences. Its Choi rank, the number of distinct labels (Section 11), is therefore at most $1 + 18 + 5300 - 78 - 3 = 5238$. With the transport of Remark H.1, words of length $O(N^{3/2})$ and pair areas $O(N^2(\delta_H(C\sqrt{N}) + N))$, the hypothesis of Corollary 10.2(a) holds with the exponent $2 + p_H/2$, and part (b) of that corollary holds with Lee's exponential Dehn bound [55] alone. The proof of Theorem 11.1 then gives three bounds for Φ_2. At purity $O(\log n)$ the memory is at least $c\,n/(\log n)^{p_H+5}$, where $p_H + 5 < 15.75$, and at least $n\,2^{-C\sqrt{\log n}}$ using only Lee's bound. At the rank rate it is at least $c\sqrt{n}/(\log n)^{2+p_H/2}$, whatever the purity. By the two-stage hashing of Remark H.1, $\log \sigma(r) = O(r \log r)$ for Γ_2 as well, which gives an upper bound of order $n \log n$. For Γ itself, Lee's bound alone gives pair areas $2^{O(N)}$, and the argument gives only memory n^c for some $c > 0$.

I Results from the companion papers, and machine checks

Table 2 lists every result of the companion papers [26, 27] on which an argument here depends, and where it is used. Results of the companions mentioned only for comparison are cited in the text.

result	source	used in
counting criterion (proof in App. B)	[26, Thm. 8.1]	Thm. 2.1, every lower bound
transport by the powers of one element	[26, Cor. 8.2]	Section 2
Johnson's presentation of H_3 and its conventions	[26, Section 2]	Appendix A
far-commutator area and polynomial Dehn bound for H_3	[26, Thms. 3.2, 3.3]	Prop. A.3
quadratic lower bound along a ray	[26, Prop. 3.4]	Sections 1, 2, 8
bounded or at least $2r$ (sharpening Cornulier's Fact 3.5); embeddings in finite groups	[26, Lemma 2.1]	Section 2
superpolynomial profile of H_3	[26, Thm. 1.1]	Section 8.2
projection products, exact S_3 rounding, sector bounds	[26, Section 4]	Lemma C.4, App. B
causal services, exchange, purity and error	[27, Def. 2.3, Section 3]	Section 9
gadget channels: Choi rank r_*, flat probe, entropy $\log r_*$	[27, Prop. 9.1]	Section 9
the rank rate is the optimal exchange rate of gadget channels in $\mathcal{K}_d$	[27, Props. 3.1, 9.1]	Thm. B(b)
sequential extraction (proof in App. B)	[27, Thm. 5.2]	Thm. 9.1
purity budget $14Q + 3$ at the rank rate (sharpened here)	[27, Thm. 5.8]	Prop. 13.1
entropy variance	[27, Lemma 5.9]	Lemma 13.2
rolling-reservoir compiler	[27, Appendix A, Lemmas A.1, A.4]	Thm. 12.2
channel profiles are group profiles; hyperlinear groups give channels in $\mathcal{K}_d$	[27, Thm. 9.2]	Section 9
distances between channels and the Choi profile	[27, Def. 2.1, Section 2]	Cor. 14.1
exact factors are cheap at their rank rate	[27, Thm. 8.6]	Appendix F
one-round theorem for the Houghton gadget (generalised here)	[27, Thm. 5.5, Appendix B]	Thm. 10.1, Appendix C
weighted norms; charged transport and weighted extraction for the Houghton gadget (generalised here)	[27, Appendix B, Lemmas B.2, B.4]	Lemmas C.1, C.2
weighted S_3 rounding	[27, Lemma B.3]	Lemma C.3
weighted sector bound (proved here by sequential measurement)	[27, Thm. B.1]	Thm. C.6

Table 2: Results of the companion papers used in this paper (numbering of their version 1.1.2).

I.1 Machine checks

The 25 freely reduced words are listed in the file `verification/relators1d.txt`. The script `verification/relators1d_check.py` rebuilds each word from the abbreviations and checks it, independently of Lemma 3.2, in the faithful action of Γ on $V \times S_3$. For a word w of length l it checks that the exponent sums of p and q vanish; that w fixes the empty configuration

and every configuration $\{y\}$ with y of height at most $l+2$ or far out on a ray; and that it fixes (v, s) for every $s \in S_3$ at every configuration v at which one of its lamp letters acts. These checks are exhaustive: while w is applied, a point of height greater than $l+2$ never reaches the points 1 and 2 and is moved only by translations along its ray, so the checks determine the affine part of w, and then the lamps of w sit at the configurations just listed. The script also checks random configurations, that the abbreviations act as stated, and, as a negative control, that deleting any single letter from a sample of eleven relations produces a word that acts nontrivially.

The script `verification/embedding3v_check.py` composes the prefix tables exactly and checks the identities of the proof of Proposition 7.2 and that the images of p, q, t, c, a satisfy the 25 relations of Table 1.

The script `verification/choi_rank_check.py` computes the Choi rank of the lamp channel. It sorts the identity and the 276 symbols of the gadget of R_Φ into classes of equal elements of Γ twice: once with the faithful action on configurations and the triviality test just described, and once with the exact images in $3V$. The two partitions agree and have 130 classes, so $r_* = 130$.

The fillings in the proof of Theorem 8.1 have been generated and checked by free reduction for $k \le 64$; the largest ratio of cell count to k^2 among them is $15/4$, at $k = 2$.

References

[1] V. Alekseev and H. Bradford, Sofic actions, halo products, and metric approximations of groups, arXiv:2601.18742 (2026).

[2] C. Anantharaman-Delaroche, On ergodic theorems for free group actions on noncommutative spaces, Probab. Theory Related Fields **135** (2006), 520–546.

[3] H. Araki and S. Yamagami, An inequality for Hilbert–Schmidt norm, Comm. Math. Phys. **81** (1981), 89–96.

[4] G. Arzhantseva and P.-A. Cherix, Quantifying metric approximations of discrete groups, Ann. Fac. Sci. Toulouse Math. **33** (2024), 831–877; doi:10.5802/afst.1788.

[5] C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor and A. Winter, The quantum reverse Shannon theorem and resource tradeoffs for simulating quantum channels, IEEE Trans. Inform. Theory **60** (2014), 2926–2959; arXiv:0912.5537.

[6] M. Berta, M. Christandl and R. Renner, The quantum reverse Shannon theorem based on one-shot information theory, Comm. Math. Phys. **306** (2011), 579–615; arXiv:0912.3805.

[7] B. V. R. Bhat and R. Devendra, On regions of mixed unitarity for semigroups of unital quantum channels, arXiv:2512.23598 (2025).

[8] H. Bradford, Controlling LEF growth in some group extensions, arXiv:2201.05052 (2022).

[9] S. Bravyi, M. B. Hastings and F. Verstraete, Lieb–Robinson bounds and the generation of correlations and topological quantum order, Phys. Rev. Lett. **97** (2006), 050401.

[10] M. R. Bridson, The geometry of the word problem, in: M. R. Bridson and S. M. Salamon (eds.), Invitations to Geometry and Topology, Oxf. Grad. Texts Math. **7**, Oxford University Press, Oxford, 2002, 29–91.

[11] M. G. Brin, Higher dimensional Thompson groups, Geom. Dedicata **108** (2004), 163–192.

[12] M. G. Brin, Presentations of higher dimensional Thompson groups, J. Algebra **284** (2005), 520–558; arXiv:math/0501082.

[13] M. G. Brin, On the baker's map and the simplicity of the higher dimensional Thompson groups nV, Publ. Mat. **54** (2010), 433–439; arXiv:0904.2624.

[14] K. S. Brown, Finiteness properties of groups, J. Pure Appl. Algebra **44** (1987), 45–75.

[15] C. Budroni, G. Fagundes and M. Kleinmann, Memory cost of temporal correlations, New J. Phys. **21** (2019), 093018; arXiv:1902.06517.

[16] J. Burillo, S. Cleary, A. Martino and C. E. Röver, Commensurations and metric properties of Houghton's groups, Pacific J. Math. **285** (2016), 289–301; arXiv:1403.0026.

[17] J. Burillo, S. Cleary and M. I. Stein, Metrics and embeddings of generalizations of Thompson's group F, Trans. Amer. Math. Soc. **353** (2001), 1677–1689; arXiv:math/9809185.

[18] J. W. Cannon, W. J. Floyd and W. R. Parry, Introductory notes on Richard Thompson's groups, Enseign. Math. (2) **42** (1996), 215–256.

[19] V. Capraro and M. Lupini, Introduction to Sofic and Hyperlinear Groups and Connes' Embedding Conjecture, Lecture Notes in Math. **2136**, Springer, Cham, 2015.

[20] J. L. Carter and M. N. Wegman, Universal classes of hash functions, J. Comput. System Sci. **18** (1979), 143–154.

[21] G. Chiribella, G. M. D'Ariano and P. Perinotti, Theoretical framework for quantum networks, Phys. Rev. A **80** (2009), 022339; arXiv:0904.4483.

[22] C. Chou, Elementary amenable groups, Illinois J. Math. **24** (1980), 396–407.

[23] F. Ciccarello, S. Lorenzo, V. Giovannetti and G. M. Palma, Quantum collision models: open system dynamics from repeated interactions, Phys. Rep. **954** (2022), 1–70; arXiv:2106.11974.

[24] Y. de Cornulier, Finitely presented wreath products and double coset decompositions, Geom. Dedicata **122** (2006), 89–108; arXiv:math/0509090.

[25] Y. Cornulier, Sofic profile and computability of Cremona groups, Michigan Math. J. **62**(4) (2013), 823–841; doi:10.1307/mmj/1387226167. Numbered references use arXiv:1305.0993v1.

[26] S. Douglas and N. Mghirbi, Houghton's group H_3 has superpolynomial sofic profile, version 1.1.2, Zenodo (2026); doi:10.5281/zenodo.23070869. Section numbers refer to version 1.1.2.

[27] S. Douglas and N. Mghirbi, Causal quantum-channel simulation: memory beyond entropy, version 1.1.2, Zenodo (2026); doi:10.5281/zenodo.23070870. Section numbers refer to version 1.1.2.

[28] T. J. Elliott, M. Gu, A. J. P. Garner and J. Thompson, Quantum adaptive agents with efficient long-term memories, Phys. Rev. X **12** (2022), 011007; arXiv:2108.10876.

[29] F. Fournier-Facio, A torsion-free non-sofic group, arXiv:2608.02025 (2026).

[30] D. Gao, S. Kunnawalkam Elayavalli and G. Patchell, Soficity for group actions on sets and applications, Res. Math. Sci. **12** (2025), 48; arXiv:2401.04945.

[31] A. Genevois, Embeddings into Thompson's groups from quasi-median geometry, Groups Geom. Dyn. **13** (2019), 1457–1510; arXiv:1709.03888v2. Numbered references use this arXiv version.

[32] A. Genevois and R. Tessera, Lamplighter-like geometry of groups, arXiv:2401.13520 (2024).

[33] G. Gour, M. P. Müller, V. Narasimhachar, R. W. Spekkens and N. Yunger Halpern, The resource theory of informational nonequilibrium in thermodynamics, Phys. Rep. **583** (2015), 1–58; arXiv:1309.6586.

[34] M. Gromov, Endomorphisms of symbolic algebraic varieties, J. Eur. Math. Soc. **1** (1999), 109–197; doi:10.1007/PL00011162.

[35] M. Gu, K. Wiesner, E. Rieper and V. Vedral, Quantum mechanics can reduce the complexity of classical models, Nat. Commun. **3** (2012), 762; arXiv:1102.1994.

[36] V. S. Guba, The Dehn function of Richard Thompson's group F is quadratic, Invent. Math. **163** (2006), 313–342.

[37] G. Gutoski and J. Watrous, Toward a general theory of quantum games, Proceedings of the 39th Annual ACM Symposium on Theory of Computing (STOC 2007), 565–574; arXiv:quant-ph/0611234.

[38] U. Haagerup and M. Musat, Factorization and dilation problems for completely positive maps on von Neumann algebras, Comm. Math. Phys. **303** (2011), 555–594; arXiv:1009.0778.

[39] U. Haagerup and M. Musat, An asymptotic property of factorizable completely positive maps and the Connes embedding problem, Comm. Math. Phys. **338** (2015), 721–752; arXiv:1408.6476.

[40] L. H. Harper, Optimal assignments of numbers to vertices, J. Soc. Indust. Appl. Math. **12** (1964), 131–135.

[41] B. Hayes and A. W. Sale, Metric approximations of wreath products, Ann. Inst. Fourier **68** (2018), 423–455; arXiv:1608.02610.

[42] J. Hennig and F. Matucci, Presentations for the higher-dimensional Thompson groups nV, Pacific J. Math. **257** (2012), 53–74; arXiv:1105.3714.

[43] A. J. Hoffman and H. W. Wielandt, The variation of the spectrum of a normal matrix, Duke Math. J. **20** (1953), 37–39.

[44] M. Horodecki, P. Horodecki and J. Oppenheim, Reversible transformations from pure to mixed states and the unique measure of information, Phys. Rev. A **67** (2003), 062104.

[45] C. H. Houghton, The first cohomology of a group with permutation module coefficients, Arch. Math. **31** (1978), 254–258.

[46] A. Ivanov, Sofic profiles of $S(\omega)$ and computability, Arch. Math. Logic **60** (2021), 477–494; arXiv:2001.02648.

[47] Z. Ji, A. Natarajan, T. Vidick, J. Wright and H. Yuen, MIP* = RE, arXiv:2001.04383 (2020).

[48] D. L. Johnson, Embedding some recursively presented groups, in: Groups St Andrews 1997 in Bath, II, London Math. Soc. Lecture Note Ser. **261**, Cambridge Univ. Press, Cambridge, 1999, 410–416.

[49] M. Junge, C. Le Merdy and Q. Xu, H^∞ functional calculus and square functions on noncommutative L^p-spaces, Astérisque **305** (2006), vi+138 pp.

[50] M. Kleinmann, O. Gühne, J. R. Portillo, J.-Å. Larsson and A. Cabello, Memory cost of quantum contextuality, New J. Phys. **13** (2011), 113011; arXiv:1007.3650.

[51] U. Krengel, Ergodic Theorems, de Gruyter Stud. Math. **6**, Walter de Gruyter, Berlin, 1985.

[52] D. Kretschmann and R. F. Werner, Quantum channels with memory, Phys. Rev. A **72** (2005), 062323; arXiv:quant-ph/0502106.

[53] B. Kümmerer, Markov dilations on W*-algebras, J. Funct. Anal. **63** (1985), 139–177.

[54] G. Kun and A. Thom, Nonsofic wreath products of residually finite groups, arXiv:2608.06222 (2026).

[55] S. R. Lee, Geometry of Houghton's groups, PhD thesis, University of Oklahoma (2012); arXiv:1212.0257v1. Numbered references use this arXiv version.

[56] E. H. Lieb and D. W. Robinson, The finite group velocity of quantum spin systems, Comm. Math. Phys. **28** (1972), 251–257.

[57] R. C. Lyndon and P. E. Schupp, Combinatorial Group Theory, Ergeb. Math. Grenzgeb. **89**, Springer, Berlin, 1977.

[58] J. Milnor, Introduction to Algebraic K-Theory, Ann. of Math. Stud. **72**, Princeton Univ. Press, Princeton, 1971.

[59] M. Musat, The Connes–Kirchberg problem and infinite-dimensional phenomena in quantum information theory, arXiv:2512.00432 (2025).

[60] M. Musat and M. Rørdam, Non-closure of quantum correlation matrices and factorizable channels that require infinite dimensional ancilla (with an appendix by N. Ozawa), Comm. Math. Phys. **375** (2020), 1761–1776; arXiv:1806.10242.

[61] B. Nachtergaele, R. Sims and A. Young, Quasi-locality bounds for quantum lattice systems. Part I. Lieb–Robinson bounds, quasi-local maps, and spectral flow automorphisms, J. Math. Phys. **60** (2019), 061101; arXiv:1810.02428.

[62] OpenAI, Ten advances in mathematics and theoretical computer science, Chapter 3: Nonsofic groups exist (2026); cdn.openai.com/pdf/ten-proofs-oai.pdf.

[63] F. A. Pollock, C. Rodríguez-Rosario, T. Frauenheim, M. Paternostro and K. Modi, Non-Markovian quantum processes: complete framework and efficient characterization, Phys. Rev. A **97** (2018), 012127; arXiv:1512.00589.

[64] C. E. Röver, Subgroups of finitely presented simple groups, PhD thesis, University of Oxford (1999).

[65] T. Rybár and M. Ziman, Repeatable quantum memory channels, Phys. Rev. A **78** (2008), 052114; arXiv:0808.3851.

[66] M. Sedlák, A. Bisio and M. Ziman, Optimal probabilistic storage and retrieval of unitary channels, Phys. Rev. Lett. **122** (2019), 170502; arXiv:1809.04552.

[67] W. Slofstra, A group with at least subexponential hyperlinear profile, arXiv:1806.05267v1 (2018).

[68] W. Slofstra and T. Vidick, Entanglement in non-local games and the hyperlinear profile of groups, Ann. Henri Poincaré **19** (2018), 2979–3005; arXiv:1711.10676.

[69] A. M. Vershik and E. I. Gordon, Groups that are locally embeddable in the class of finite groups, Algebra i Analiz **9**(1) (1997), 71–97; English transl. St. Petersburg Math. J. **9**(1) (1998), 49–67.

[70] B. Weiss, Sofic groups and dynamical systems, Sankhyā Ser. A **62** (2000), 350–359.

[71] H. Bradford, Quantifying local embeddings into finite groups, arXiv:2104.07111v2 (2021).

[72] M. Chaudkhari, K. Juschenko and K. Muliarchyk, Thompson's groups are sofic, arXiv:2610.25415v1 (2026).