Scrambling speed of random quantum circuits

Winton Brown* Omar Fawzi†

October 25, 2012

Abstract

Random transformations are typically good at "scrambling" information. Specifically, in the quantum setting, scrambling usually refers to the process of mapping most initial pure product states under a unitary transformation to states which are macroscopically entangled, in the sense of being close to completely mixed on most subsystems containing a fraction fn of all n particles for some constant f. While the term scrambling is used in the context of the black hole information paradox, scrambling is related to problems involving decoupling in general, and to the question of how large isolated many-body systems reach local thermal equilibrium under their own unitary dynamics.

Here, we study the *speed* at which various notions of scrambling/decoupling occur in a simplified but natural model of random two-particle interactions: random quantum circuits. For a circuit representing the dynamics generated by a local Hamiltonian, the depth of the circuit corresponds to time. Thus, we consider the depth of these circuits and we are typically interested in what can be done in a depth that is sublinear or even logarithmic in the size of the system. We resolve an outstanding conjecture raised in the context of the black hole information paradox with respect to the depth at which a typical quantum circuit generates an entanglement assisted encoding against the erasure channel. In addition, we prove that typical quantum circuits of $\mathrm{poly}(\log n)$ depth satisfy a stronger notion of scrambling and can be used to encode αn qubits into n qubits so that up to βn errors can be corrected, for some constants $\alpha, \beta > 0$.

1 Introduction

Random quantum circuits of polynomial size are meant to be efficient implementations that inherent many useful properties of "uniformly" chosen unitary transformations which are typically very inefficient. A lot of work was done in analyzing convergence properties of the distribution defined by random quantum circuits to the Haar measure on the full unitary group acting on n qubits [EWS+03, ELL05, ODP07, Žni08, Oli09, HL09, Low10, BV10, BHH12]. Here, instead of trying to study the convergence of these circuits to some limit, we study the information-theoretic property of interest directly. This property can be intuitively pictured as "scrambling" or spreading some structured or localized information over the global system of n qubits. The term scrambling is used in the context of the black hole information paradox [HP07, SS08, LSH+11], but such a property can also be understood in terms of decoupling, a central notion in the study of quantum communication [HOW05, HOW06, HHYW08, Dup10, ADHW09, DBWR10]. On a more technical level, a typical property of a scrambler can be seen when we decompose the input and output states in the Pauli basis (which can be seen as a Fourier basis): a scrambler tends to reduce the mass of the low-weight Pauli operators. In fact, all of our arguments prove a statement of that form.

1.1 Strong scrambling, quantum error correction and decoupling

An important example of a scrambler is an encoding circuit for a quantum error correcting code. In particular a k-qubit, distance d, non-degenerate error correcting code maps all initial states localized on k qubits to states which are completely mixed on all subsystems of size less than d, which can be considered a strong form of scrambling when the distance is a constant fraction of n. Another way of defining a good quantum error correcting code is

*Département de Physique, Université de Sherbrooke
†Institute for Theoretical Physics, ETH Zürich

that it decouples a purification of the encoded qubits from any subsystems of size smaller than d. Proving coding theorems by proving a decoupling statement has been quite successful culminating in a very general decoupling theorem [HOW05, HOW06, HHYW08, Dup10, ADHW09, DBWR10]. Our objective can also be seen as trying to determine how fast decoupling occurs.

We prove the following results:

- We give a random quantum circuit model of depth $O(\log^3 n)$ that satisfies a strong notion of scrambling. That is, for any initial state, on average over the circuit, all subsystems of size at most fn are close to completely mixed.
- This result can also be considered as giving decoupling unitaries that are more efficient than standard (approximate) two-designs. Relying on the fact that random quantum circuits are approximate two-designs [HL09], it was shown by [SDTR11] that random circuits of size $O(n^2)$ are decouplers in a quite general setting. Here we prove that in some particular cases, we can obtain much faster decoupling with circuits of depth $O(\log^3 n)$.
- As another application, we prove the existence of stabilizer codes with an encoding circuit of depth $O(\log^3 n)$ that have a constant encoding rate and a minimum distance that grows linearly with n.

It would be interesting to prove that scrambling occurs in depth $O(\log n)$ instead. Our second set of results proves a weaker notion of scrambling in depth $O(\log n)$. This notion of scrambling is particularly relevant in the study of the black hole information paradox question.

1.2 Black holes and the fast scrambling conjecture

It was noted in [Pag93], that by collecting the Hawking radiation from a black hole, an arbitrary message dropped into the black hole could be recovered after half the black hole had evaporated if the dynamics of the black hole could be approximated as a random unitary transformation. This approach was tightened significantly in [HP07], where it was shown that at any time after the black hole has evaporated past its half way point, an m-qubit quantum state that was dropped into the black hole could be recovered with high fidelity from an amount of Hawking radiation containing slightly more than m qubits of quantum information, as long as the dynamics of the black hole approximates a unitary two-design sufficiently. A random quantum circuit model analyzed in [DCEL09] was invoked which could, for the purposes of recovering an initial state of constant size, scramble the degrees of freedom by a local circuit of depth $O(\log n)$. This random quantum circuit model, though highly contrived, could be performed by two-qubit gates between nearest neighbours on a 2-dimensional lattice, in a depth of $O(\sqrt{n}\log n)$. This amount of time is just enough to avoid a violation of the quantum no cloning principle assuming complementarity at the event horizon. This motivated interest in the scrambling properties of more natural models of random quantum circuits that may better represent a naturally arising Hamiltonian. It was conjectured in [SS08] that this was possible in time $O(n^{1/d})$ and $O(\log n)$ for a local Hamiltonian in d-dimensions and infinite dimensions respectively, with k-body interactions. Since the signaling bound precludes faster scrambling, such unitary transformations are referred to as "fast scramblers".

- Here we resolve the fast scrambling conjecture for random quantum circuits in the case of constant message size. We show that typical random quantum circuits on d-dimensional lattices and the complete graph, of depth $O(n^{1/d}\log^2 n)$ and $O(\log n)$ respectively, scramble a message of constant size m such that it may be recovered with high fidelity using only $m + c$ randomly selected qubits, for some constant c. Since a straightforward lower bound of $\Omega(n^{1/d})$ and $\Omega(\log n)$ can be shown, our results are nearly optimal.

1.3 Proof technique

The first step of the proof is to relate the property of interest, which is most naturally stated in terms of the trace-norm, to the two norm, whose behavior under under the random quantum circuit can be completely described by its second-order moment operator. For the random quantum circuits we consider, this moment operator, when evaluated in the Pauli basis, can be seen as a Markov chain on the set of Pauli basis elements (also called Pauli strings). This means that the properties of interest can be seen as properties of this Markov chain.

Most previous studies of random quantum circuits bounded the convergence using the spectral gap of the moment operators. However, as we show, the spectral gap only weakly depends on the underlying interaction

graph of the circuit. This means that any result that uses the spectral gap of the second moment operator will give bounds on the scrambling time that would also apply to circuits where the gates are applied between neighbouring cells on a one dimensional line. In particular, as the diameter of the interaction graph plays a crucial role in determining the scrambling speed, it is necessary in our proofs to go beyond placing bounds on the spectral gap and to make use directly of the Markov chain (or a Markov chain obtained from lumping certain states), which heavily depends on the interaction graph.

2 Preliminaries

2.1 Generalities

The state of a pure quantum system is represented by a unit vector in a Hilbert space. Quantum systems are denoted $A, B, C \ldots$ and are identified with their corresponding Hilbert spaces. The Hilbert spaces we consider here will be mostly n-qubits spaces of the form $(\mathbb{C}^2)^{\otimes n}$. To describe a distribution $\{p_1, \ldots, p_r\}$ over quantum states $\{|\psi_1\rangle, \ldots, |\psi_r\rangle\}$ (also called a mixed state), we use a density operator $\rho = \sum_{i=1}^{r} p_i |\psi_i\rangle\langle\psi_i|$, where $|\psi\rangle\langle\psi|$ refers to the projector on the line defined by $|\psi\rangle$. A density operator is a Hermitian positive semidefinite operator with unit trace. The density operator associated with a pure state is abbreviated by omitting the ket and bra $\psi \stackrel{\text{def}}{=} |\psi\rangle\langle\psi|$. $\mathcal{S}(A)$ is the set of density operators acting on A. The Hilbert space on which a density operator $\rho \in \mathcal{S}(A)$ acts is sometimes denoted by a subscript, as in ρ_A. This notation is also used for pure states $|\psi\rangle_A \in A$.

In order to describe the joint state of a system AB, we use the tensor product Hilbert space $A \otimes B$, which is sometimes simply denoted AB. If ρ_{AB} describes the joint state on AB, the state on the system A is described by the partial trace $\rho_A \stackrel{\text{def}}{=} \operatorname{tr}_B \rho_{AB}$. If U is a unitary acting on A, and $|\psi\rangle$ a state in $A \otimes B$, we sometimes use $U|\psi\rangle$ to denote the state $(U \otimes \mathbb{1}_B)|\psi\rangle$, where the symbol $\mathbb{1}_B$ is reserved for the identity map on B. For an introduction to quantum information, we refer the reader to [NC00].

Throughout the paper, we use the Pauli basis, which is an orthogonal basis for 2×2 matrices:

$$\sigma_0 = \mathbb{1} \qquad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

For a string $\nu \in \{0,1,2,3\}^n$, we define $\sigma_\nu = \sigma_{\nu_1} \otimes \cdots \otimes \sigma_{\nu_n}$. The support $\operatorname{supp}(\nu)$ of ν is simply the subset $\{i \in [n] : \nu_i \neq 0\}$ and the weight $w(\nu) = |\operatorname{supp}(\nu)|$.

We now introduce some various notation. The notation $\operatorname{poly}(n)$ refers to a term that could be chosen to be any polynomial and the power of the polynomial can be made larger by appropriately choosing the related constants. As we are going to deal with binomial coefficients, the binary entropy function $h(x) = -x \log x - (1-x)\log(1-x)$ is going to be used. We also use the shorthand $[n] \stackrel{\text{def}}{=} \{1, \ldots, n\}$.

2.2 Random quantum circuits

We consider two related models for random quantum circuits. In a sequential random quantum circuit a random two-qubit gate is applied to a randomly chosen pair of qubits in each time step. For a general interaction graph, instead of choosing a pair at random from all the possible pairs, we choose a random edge in the graph. Here the random two-qubit gate is going to be a random Clifford gate or a gate uniformly chosen from the Haar measure on the unitary group acting on two qubits. In fact, as the second-order moment operator is the same for these two models, our results apply equally well to them. However, the result of Theorem 3.6 proving the existence of stabilizer codes with efficient encoding makes explicit use of the model with random Clifford gate. Since we are interested in the speed at which scrambling occurs rather than the gate complexity, we ask into how many layers of gates can the sequence be decomposed so that no two gates act on the same qubit. To construct the parallelized circuit, one keeps adding gates to the current level until there is a gate that shares a qubit with a previously added gate in that level, in which case you create a new level and continue. We show that parallelizing a size n random circuit results in a depth of $O(\log n)$ with high probability.

In order to avoid this overhead, we also consider a circuit model which is parallelized by construction. In this second model, a random maximum matching of on the complete graph is chosen and a random two-qubit gate is applied to qubits that are joined by an edge. We will also be interested in partially parallelized construction when in each time step a random edge is drawn from each of a set of coarse grained cells on a d-dimensional lattice.

A model of random circuits of a certain size defines a measure over unitary transformations on n qubits that we call p_{circ}. We will sometimes compare the behaviour of the circuit to a unitary transformation chosen from the Haar measure p_{haar} over the full unitary group on n qubits.

As mentioned earlier, the second-order moment operator will play an important role in all our proofs. The second-order moment operator is a super-operator acting on two copies of the space of operators acting on the ambient Hilbert space, which is an n-qubit space in our setting. For a measure p over the unitary group, we can define the second moment operator M as

$$M[X \otimes Y] = \mathop{\mathbf{E}}_{U \sim p} \left\{ UXU^\dagger \otimes UYU^\dagger \right\}.$$

In particular $M_{\text{haar}} = \mathop{\mathbf{E}}_{U \sim p_{\text{haar}}} \left\{ UXU^\dagger \otimes UYU^\dagger \right\}$. Any distribution for which $M = M_{\text{haar}}$ is referred to as a two-design. We will be using the following properties of the second moment operator.

- For a circuit composed of t gates chosen independently, the second moment operator is $M_{\text{circ}} = M^t$ where M is the second-order moment operator corresponding to the measure obtained when applying one gate.
- For all the measures p we consider here, the second moment operator is Hermitian. In addition, all eigenvalues of the moment operator are bounded in absolute value by 1 and for the measures we consider, 1 is the only eigenvalue of magnitude 1.
- The eigenspace $\mathcal{V}$ for the eigenvalue 1 can be shown to be the space of operators X acting on $2n$ qubits such that for all unitary transformations $U \otimes UXU^\dagger \otimes U^\dagger = X$ (for the distributions p we consider here). It follows that this space is the span of the identity operator and the swap operator. The moment operator of the Haar measure, M_{haar}, is the projector onto $\mathcal{V}$.

These properties imply that if M is the second-order moment operator associated to the random quantum circuits we consider here, M^t converges to M_{haar} as $t \to \infty$ at an asymptotic rate determined by the second largest λ_2 (in absolute value) eigenvalue of M. The gap of the moment operator is defined by $\Delta = 1 - \lambda_2$, and the larger the gap, the faster the moment operator converges to M_{haar}. In order to study random quantum circuits when the interaction graph is a d-dimensional lattice, we will need a lower bound on the gap of these random quantum circuits. In order to obtain that, we proceed as in [Žni08, BHH12] seeing the second-order moment operator as a local Hamiltonian. In fact, for a sequential random quantum circuit, we can write the second-order moment operator as follows:

$$M_{\text{circ}} = \sum_{i<j} q_{ij} m_{ij},$$

where $m_{ij} = \mathop{\mathbf{E}}_{U \sim \tilde{p}} \left\{ UXU^\dagger \otimes UYU^\dagger \right\}$ and $\tilde{p}$ is the normalized measure over gates that act only on qubits i and j and q_{ij} is the total probability over such gates. Then, one can use a result on the gap of local frustration free Hamiltonians [Nac96]. The property of being frustration free in this context simply mean that if X is invariant for M_{circ}, then it is also invariant for the terms m_{ij}, which follows easily from the properties mentioned above.

As mentioned earlier, the Pauli basis will play an important role in our analysis. Consider a representation of the moment operator M in the basis defined by $\sigma_\nu \otimes \sigma_\mu$, with $\nu, \mu \in \{0,1,2,3\}^n$. This defines a matrix $\{Q((\mu,\mu'),(\nu,\nu'))\}_{\mu,\mu',\nu,\nu'}$ of size $16^n \times 16^n$.

First, it can be shown that for the random quantum circuits we consider here, we have

$$Q((\mu,\mu'),(\nu,\nu')) = \frac{1}{4^n} \mathop{\mathbf{E}}_{U \sim p} \left\{ \text{tr}[(\sigma_\nu \otimes \sigma_{\nu'})(U\sigma_\mu U^\dagger \otimes U\sigma_{\mu'}U^\dagger)] \right\} = 0,$$

unless $\nu = \nu'$ and $\mu = \mu'$. As a result, we will simply write $Q(\mu, \nu)$. Note that for any μ, we have

$$\begin{aligned}\sum_{\nu\in\{0,1,2,3\}^n} Q(\mu,\nu) &= \sum_{\nu\in\{0,1,2,3\}^n} \frac{1}{4^n} \mathop{\mathbf{E}}_{U\sim p} \left\{\mathrm{tr}[(\sigma_\nu \otimes \sigma_\nu)(U\sigma_\mu U^\dagger \otimes U\sigma_\mu U^\dagger)]\right\} \\ &= \frac{1}{4^n} \mathop{\mathbf{E}}_{U\sim p} \left\{ \sum_{\nu\in\{0,1,2,3\}^n} \mathrm{tr}[\sigma_\nu U\sigma_\mu U^\dagger]^2 \right\} \\ &= \frac{1}{2^n} \mathop{\mathbf{E}}_{U\sim p} \left\{ \mathrm{tr}\left[\left(U\sigma_\mu U^\dagger\right)^2\right] \right\} \\ &= \frac{1}{2^n} \mathrm{tr}[\sigma_\mu^2] \\ &= 1.\end{aligned}$$

This proves that $\{Q(\mu,\nu)\}_{\mu,\nu}$ can be seen as the transition matrix of a Markov chain on the set of Pauli strings $\{0,1,2,3\}^n$. This Markov chain is going to play an important role throughout the paper and the information theoretic properties we are interested in are going to be expressed in terms of its properties. More precisely, we are going to consider the chain obtained by removing the state 0^n (which is isolated from the rest of the chain).

Given a set of interacting pairs, to which Haar random (or Clifford) two-qubit gates are applied, the Markov chain for the sequential random quantum circuit is constructed in the following way. Consider only gates acting on qubits i and j. If the Pauli string is $\sigma_0 \otimes \sigma_0$ on i and j, then there are no transitions induced as the gate acts as the identity on the string. If the value of the string is $\sigma_a \otimes \sigma_b$ with $(a,b) \neq (0,0)$ on qubits i and j, then following from invariance of the Haar measure under unitary transformations, the value of the string on i and j transitions to each $\sigma_c \otimes \sigma_d$ where (c,d) is chosen uniformly from $\{0,1,2,3\}^2 - \{(0,0)\}$. An average over all such transition for each interacting pair allowed by the interaction graph results in the Markov chain for this circuit.

3 Strong scrambling, decoupling and quantum error correction

In this section, we prove that a random circuit of size $O(n\log^2 n)$ scrambles (on average over the choice of circuit) any initial state in the sense that all subsets of size at most fn for some constant f are very close to maximally mixed. In order to prove such a result, we consider the total mass of the coefficients corresponding to the Pauli strings of weight at most fn, and prove that it is small with very high probability. The common thing between proving strong scrambling and obtaining error correcting codes with large minimum distance is that the probability bounds should be close to optimal. More precisely, we prove that a Pauli string of weight ℓ is mapped by the random circuit to a Pauli string of weight at least fn with probability at least roughly $1 - \frac{1}{\binom{n}{\ell}}$. The following Section 3.2 then says that this circuit can with high probability be parallelized so that it has depth $O(\log^3 n)$. In the following sections, we see how we can interpret our upper bounds on the total mass on low-weight Pauli strings to prove results on decoupling and quantum error correcting codes for low-depth random quantum circuits.

3.1 Sequential random circuit

Theorem 3.1. *Let $\rho(0)$ be an initial arbitrary mixed state on n qubits and $\rho(t)$ be the corresponding state after the application of t random two-qubit gates (the sequential circuit model). Then provided f is such that $f\log 3 + h(f) - \frac{\log 3}{2} < 0$ and $t > cn\log^2 n$ (for some large enough constant c), we have for all subsets S of size at most fn,*

$$\mathbf{E}\left\{\mathrm{tr}[\rho_S(t)^2]\right\} \leq \frac{1}{2^{|S|}} + \frac{1}{2^{|S|}\,\mathrm{poly}(n)}, \tag{1}$$

where the expectation is taken over the random circuit. This implies that

$$\mathbf{E}\left\{\left\|\rho_S(t) - \frac{\mathbb{1}}{2^{|S|}}\right\|_1^2\right\} \leq \frac{1}{\mathrm{poly}(n)}. \tag{2}$$

Proof First, observe that (1) easily implies (2) using the Cauchy-Schwarz inequality:

$$\begin{aligned}\left\|\rho_S(t) - \frac{\mathbb{1}}{2^{|S|}}\right\|_1^2 &\leq 2^{|S|}\left\|\rho_S(t) - \frac{\mathbb{1}}{2^{|S|}}\right\|_2^2 \\ &= 2^{|S|}\left(\operatorname{tr}[\rho_S(t)^2] - 2\frac{\operatorname{tr}[\rho_S(t)]}{2^{|S|}} + \frac{\operatorname{tr}[\mathbb{1}]}{2^{2|S|}}\right) \\ &= 2^{|S|}\operatorname{tr}[\rho_S(t)^2] - 1.\end{aligned}$$

To compute $\operatorname{tr}[\rho_S(t)^2]$, we decompose $\rho_S(t)$ in the Pauli basis:

$$\rho_S(t) = \sum_{\nu\in\{0,1,2,3\}^S} 2^{-|S|}\operatorname{tr}[\sigma_\nu\rho(t)]\sigma_\nu.$$

As a result, we have

$$\begin{aligned}\operatorname{tr}[\rho_S^2(t)] &= \sum_{\nu\in\{0,1,2,3\}^S}\frac{\operatorname{tr}[\sigma_\nu\rho_S(t)]^2}{2^{|S|}} \\ &= \sum_{\nu\in\{0,1,2,3\}^S}\frac{\operatorname{tr}[\sigma_\nu\otimes\mathbb{1}_{S^c}\operatorname{tr}_{S^c}[\rho(t)]]^2}{2^{|S|}} \\ &= \frac{1}{2^{|S|}} + \sum_{\nu\in\{0,1,2,3\}^S,\nu\neq 0}\frac{\operatorname{tr}[\sigma_\nu\otimes\mathbb{1}_{S^c}\rho(t)]^2}{2^{|S|}} \\ &\leq \frac{1}{2^{|S|}} + \sum_{\nu:1\leq w(\nu)\leq|S|}\frac{\operatorname{tr}[\sigma_\nu\rho(t)]^2}{2^{|S|}}.\end{aligned}$$

Our objective now is to study the evolution of the quantity $\mathbf{E}\left\{\sum_{\nu:1\leq w(\nu)\leq|S|}\operatorname{tr}[\sigma_\nu\rho(t)]^2\right\}$ as a function of t. As we described in the preliminaries, applying a random two-qubit gate has a simple effect on the decomposition into the Pauli basis: an identity on two qubits always gets mapped to an identity and a non-identity Pauli string on two qubits gets mapped to a uniformly chosen non-identity Pauli string (of which there are 15).

Our focus will be to study the Markov chain that describes the evolution of the distribution of the weight of the different levels $\sum_{\nu:w(\nu)=k}\mathbf{E}\left\{\operatorname{tr}[\sigma_\nu\rho(t)]^2\right\}$. More precisely, we can write for any $k\in\{1,\ldots,n\}$,

$$\begin{aligned}\sum_{\nu:w(\nu)=k}\mathbf{E}\left\{\operatorname{tr}[\sigma_\nu\rho(t)]^2\right\} &= \sum_{\nu:w(\nu)=k,\mu}\mathbf{E}\left\{\operatorname{tr}[\sigma_\mu\rho(t-1)]^2\right\}\mathbf{E}\left\{\operatorname{tr}[\sigma_\nu U_t\sigma_\mu U_t^\dagger]^2\right\} \\ &= P(k-1,k)\sum_{\mu:w(\mu)=k-1}\mathbf{E}\left\{\operatorname{tr}[\sigma_\mu\rho(t-1)]^2\right\} \\ &+ P(k,k)\sum_{\mu:w(\mu)=k}\mathbf{E}\left\{\operatorname{tr}[\sigma_\mu\rho(t-1)]^2\right\} \\ &+ P(k+1,k)\sum_{\mu:w(\mu)=k+1}\mathbf{E}\left\{\operatorname{tr}[\sigma_\mu\rho(t-1)]^2\right\},\end{aligned}$$

where the matrix $P\in\mathbb{R}^{n\times n}$ is defined by

$$P(x,y) = \begin{cases} 1-\frac{2x(3n-2x-1)}{5n(n-1)} & \text{if } y=x \\ \frac{2x(x-1)}{5n(n-1)} & \text{if } y=x-1 \\ \frac{6x(n-x)}{5n(n-1)} & \text{if } y=x+1 \\ 0 & \text{otherwise.}\end{cases}$$

We refer the reader to [HL09] for more details on how to derive the parameters of this Markov chain. In fact, [HL09] study the mixing time of this Markov chain. Here, we need to analyze a slightly different property: starting at some point ℓ, what is the probability that after t steps the random walk ends up in a point $\leq fn$? One can obtain bounds on this probability using the mixing time but these bounds only give something useful for our setting if

$t = \Omega(n^2)$. So we will need to improve the analysis of [HL09] and go directly for computing the desired probability instead of going through the mixing time. More precisely, by defining the Markov chain $\{X_s(\ell)\}_{s\geq 0}$ that starts at ℓ and has transition probabilities given by P, we have

$$\sum_{k=1}^{fn} \sum_{\nu:w(\nu)=k} \mathbf{E}\left\{\mathrm{tr}[\sigma_\nu \rho(t)]^2\right\} = \sum_{\ell=1}^{n} \sum_{\nu:w(\nu)=\ell} \mathrm{tr}[\sigma_\nu \rho(0)]^2 \mathbf{P}\left\{X_t(\ell) \leq fn\right\}. \tag{3}$$

If the initial state $\rho(0)$ is a pure product state, then one can verify that

$$\sum_{\nu:w(\nu)=\ell} \mathrm{tr}[\sigma_\nu \rho(0)]^2 = \binom{n}{\ell}.$$

In general, we have

$$\begin{aligned}\sum_{\nu:w(\nu)=\ell} \mathrm{tr}[\sigma_\nu \rho(0)]^2 &\leq \sum_{S:|S|=\ell} 2^{|S|} \,\mathrm{tr}[\rho_S(0)^2] \\ &\leq 2^\ell \binom{n}{\ell}.\end{aligned}$$

The main technical result in this proof is in Theorem A.1 (which we defer to the appendix), where we obtain a bound on $\mathbf{P}\left\{X_t(\ell) \leq fn)\right\} \leq \frac{2^{f\log 3 + h(f)}}{\binom{n}{n/2} 3^{n/2}} + \frac{1}{2^\ell \binom{n}{\ell} \mathrm{poly}(n)}$. Plugging this into (3), we obtain

$$\begin{aligned}\sum_{k=1}^{fn} \sum_{\nu:w(\nu)=k} \mathbf{E}\left\{\mathrm{tr}[\sigma_\nu \rho(t)]^2\right\} &\leq \sum_{\ell=1}^{n} \sum_{\nu:w(\nu)=\ell} \mathbf{E}\left\{\mathrm{tr}[\sigma_\nu \rho(0)]^2\right\} \cdot \frac{1}{2^\ell \binom{n}{\ell} \mathrm{poly}(n)} + \frac{2^{f\log 3 + h(f)}}{\binom{n}{n/2} 3^{n/2}} \sum_{\ell=1}^{n} \sum_{\nu:w(\nu)=k} \mathbf{E}\left\{\mathrm{tr}[\sigma_\nu \rho(0)]^2\right\} \\ &\leq \sum_{\ell=1}^{n} 2^\ell \binom{n}{\ell} \cdot \frac{1}{2^\ell \binom{n}{\ell} \mathrm{poly}(n)} + 2^n \cdot \mathrm{tr}[\rho(0)^2] \cdot \frac{2^{f\log 3 + h(f)}}{\binom{n}{n/2} 3^{n/2}} \\ &\leq \frac{1}{\mathrm{poly}(n)},\end{aligned}$$

provided f is such that $f \log 3 + h(f) - \frac{\log 3}{2} < 0$. □

3.2 Parallelizing the circuit

Recall that we are interested in the depth of random circuits. A priori, the circuit studied in the previous section has a depth that is as large as the number of gates which is nearly linear. But in general in such a circuit there are many successive gates that are applied on disjoint qubits so they could be actually performed in parallel. More precisely, we look at the gates one by one in the order they are applied. For the purpose of this section, the gates can simply be labelled by the two qubits the gate acts upon. To construct the parallelized circuit, one keeps adding gates to the current level until there is a gate that shares a qubit with a previously added gate in that level, in which case you create a new level and continue. In the following proposition, we prove that by parallelizing a random circuit on n qubits having t gates we obtain with high probability a circuit of depth $O(\frac{t}{n}\log n)$.

Proposition 3.2. *Consider a random sequential circuit composed of t gates where t is a polynomial in n. Then parallelize the circuit as described above. Except with probability $1/\mathrm{poly}(n)$, you end up with a circuit of depth at most $O\left(\frac{t}{n}\log n\right)$.*

In order to prove this lemma, we use the following calculation:

Lemma 3.3. *Let $G_1, \ldots, G_k$ be a sequence of independent and random gates $G_i \in \binom{n}{2}$, then the probability that $G_1, \ldots, G_k$ form a circuit of depth k is at most $\left(\frac{2}{n}\right)^{k-1} \cdot k!$*

Proof We prove this by induction on k. For $k = 2$, we may assume $G_1 = (1, 2)$, in which case $\mathbf{P}\left\{G_2 \cap \{1, 2\} \neq \emptyset\right\} \leq 4/n$. Now the probability that $G_1, \ldots, G_{k+1}$ form a circuit of depth $k + 1$ can be bounded by

$$\mathbf{P}\left\{G_1, \ldots, G_k \text{ form a circuit of depth } k\right\} \cdot \mathbf{P}\left\{G_{k+1} \cap (G_1 \cup \cdot \cup G_k) \neq \emptyset | G_1, \ldots, G_k \text{ form a circuit of depth } k\right\}.$$

Now it suffices to see that, conditioned on $[G_1, \ldots, G_k$ form a circuit of depth $k]$, the number of nodes occupied by $G_1, \ldots, G_k$ is at most $k+1$. Thus, using this fact and the induction hypothesis, we obtain a bound of

$$\left(\frac{2}{n}\right)^{k-1} k! \cdot 2 \cdot \frac{k+1}{n} = \left(\frac{2}{n}\right)^{k} (k+1)!\,,$$

which conclude the proof. □

Proof [of Proposition 3.2] Suppose we apply m gates for some m to be chosen later.

$$\begin{aligned}\mathbf{P}\left\{G_1, \ldots, G_m \text{ form a circuit of depth at least } d\right\} &= \mathbf{P}\left\{\exists (i_1, \ldots, i_d) \in [m]^d : G_{i_1}, \cdots, G_{i_d} \text{ form a circuit of depth d}\right\} \\ &\leq \binom{m}{d}\left(\frac{2}{n}\right)^{d-1} \cdot d! \\ &\leq m^d \cdot \left(\frac{2}{n}\right)^{d-1}.\end{aligned}$$

Now we can fix $m = n/4$ and $d = c\log n + 1$ for some constant c to be chosen depending on the desired probability bound, then we have

$$\mathbf{P}\left\{G_1, \ldots, G_m \text{ form a circuit of depth at least } d\right\} \leq m \cdot \left(\frac{2m}{n}\right)^{d-1} \leq n^{-c+1}.$$

This proves that every set of $n/4$ gates generate a circuit of depth at most $c\log n + 1$ with probability at least $1 - 1/n^{-c+1}$, and so if we have $4t/n$ such sets, we get depth at most $4t/n(c\log n + 1)$ with probability at least $1 - 4t/n^c$. □

The next corollary follows directly from Theorem 3.1 and Proposition 3.2.

Corollary 3.4. *In the parallelized random quantum circuit model with depth $O(\log^3 n)$, we have*

$$\mathbf{E}\left\{\left\|\rho_S(t) - \frac{\mathbb{1}}{2^{|S|}}\right\|_1^2\right\} \leq \frac{1}{\mathrm{poly}(n)} \tag{4}$$

for all subsets S of size at most fn with f such that $f\log 3 + h(f) - \frac{\log 3}{2} < 0$.

3.3 Decoupling and quantum error correcting codes

Scrambling is related to the notion of decoupling. The idea of decoupling plays an important role in quantum information theory and many coding theorems amount to proving a decoupling theorem [HOW05, HOW06, HHYW08, ADHW09, Dup10, DBWR10].

Consider the setting described in Figure 1. Let $|\Phi\rangle_{MM'}$ and $|\psi\rangle_{AA'}$ be pure states on MM' and AA' respectively. Then apply some unitary transformation to the system $M'A'$ (which for us is going to be a random quantum circuit) and map it to a system that we call B. Let us denote by $|\rho\rangle_{BMA}$ the output state. Assume now that the reduced state ρ_{MS} on M together with some subset S of the qubits of B is a product state: $\rho_{MS} = \rho_M \otimes \rho_S$ (the subsystem S is *decoupled* from the reference M). Then by Uhlmann's theorem (or the unitary equivalence of purifications), there exists an isometry acting on AS^c that recovers a purification of the system M. If for example $|\Phi\rangle_{MM'}$ is a maximally entangled state, then the previous argument shows that if we input quantum information into the M' system, it can be recovered from the systems AS^c alone with no need for the system S.

In the following for simplicity, we focus on the case

$$\Phi_{MM'} = \frac{1}{2^{2m}} \sum_{\nu \in \{0,1,2,3\}^m} \sigma_\nu \otimes \sigma_\nu,$$

where the systems M and M' consist of m qubits. For the AA' system, we will focus on two important cases: First where A' is already in a pure state $|\psi\rangle_{A'} = |0\rangle_{A'}$, so that it can be written in the Pauli basis as

$$\psi_{A'} = \frac{1}{2^{n-m}} \sum_{\nu\{0,3\}^{n-m}} \sigma_\nu\,, \tag{5}$$

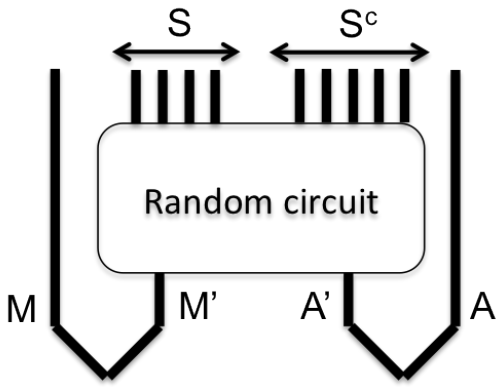

Figure 1: Illustration of decoupling for random quantum circuits

and second the case where $\psi_{AA'}$ is maximally entangled so that

$$\psi_{AA'} = \frac{1}{2^{2(n-m)}} \sum_{\nu \in \{0,1,2,3\}^{n-m}} \sigma_\nu \otimes \sigma_\nu, \tag{6}$$

which corresponds to entanglement assisted communication. Of course, one could obtain a statement for general states and this would depend on the decomposition of the states $\Phi_{MM'}$ and $\psi_{AA'}$ in the basis of Pauli strings and more precisely on the weight distribution of this decomposition.

A decoupling statement is very similar in spirit to scrambling and the analysis is almost the same except that we use a specific state for the input. Using Theorem A.1 which was the main technical ingredient in the proof of Theorem 3.1, we can get the following decoupling and coding results.

Theorem 3.5. *In the setting of equation* (5)*, we have if $m < \beta n$ and any S of size $|S| \leq fn$ with $\beta < 1/9$ and $\beta < \log 3/2 - f \log 3 - h(f)$, and ρ is the state obtained after applying parallelized random quantum circuit of depth $O(\log^3 n)$, we have*

$$\left\| \rho_{MS} - \frac{\mathbb{1}}{2^m} \otimes \frac{\mathbb{1}}{2^{|S|}} \right\|_1 \leq \frac{1}{\mathrm{poly}(n)}$$

with probability $1 - 1/\mathrm{poly}(n)$ over the choice of the circuit.

In the setting of equation (6) *(entanglement assisted coding), we have for $\beta < 2/3$ and $\beta < \frac{1+\log 3/2 - f \log 3 - h(f)}{2}$,*

$$\left\| \rho_{MS} - \frac{\mathbb{1}}{2^m} \otimes \frac{\mathbb{1}}{2^{|S|}} \right\|_1 \leq \frac{1}{\mathrm{poly}(n)}$$

with probability $1 - 1/\mathrm{poly}(n)$ over the choice of the circuit.

Remark. The rates we obtain here are not optimal, but we prove that it is possible to code at constant rate with a constant fraction of errors using a circuit of polylogarithmic depth. It would be interesting to determine whether it is possible to achieve the capacity of the erasure channel using such shallow circuits. For example, it would be interesting to improve the bound in the entanglement assisted case to $\beta < \frac{2 - f \log 3 - h(f)}{2}$. This is the bound one would get for a random unitary distributed according to the Haar measure on the unitary group acting on n qubits and is reminiscent of the entanglement assisted capacity of the depolarizing channel. □

Proof We start with the entanglement assisted case, for which the calculation is a bit simpler. We apply a random quantum circuit to the system $M'A'$. We can write the initial state on $MM'A'$ as

$$\rho(0) = \frac{1}{2^{2m}} \sum_{\nu \in \{0,1,2,3\}^m} \sigma_\nu \otimes \sigma_\nu \otimes \frac{\mathbb{1}}{2^{n-m}}.$$

We study the evolution of the state $\rho(t)$ when we apply t random gates, more precisely we study the behaviour of the reduced state when a subset size $k = (1-f)n$ qubits are discarded (from the $M'A'$ system), the objective is to

show that the remaining state is close to maximally mixed. We have

$$\begin{aligned}\mathbf{E}\left\{\left\|\rho(t)_{MS}-\frac{\mathbb{1}}{2^{m+fn}}\right\|_1^2\right\} &\leq \sum_{\nu\in\{0,1,2,3\}^m,\mu\in\{0,1,2,3\}^S,(\nu,\mu)\neq 0}\mathbf{E}\left\{\mathrm{tr}[\sigma_\nu\otimes\sigma_\mu\otimes\mathbb{1}\rho(t)]^2\right\}\\ &\leq \sum_{\nu\in\{0,1,2,3\}^m,w(\mu)\leq fn,(\nu,\mu)\neq 0}\mathbf{E}\left\{\mathrm{tr}[\sigma_\nu\otimes\sigma_\mu\rho(t)]^2\right\}\\ &= \sum_{\nu\in\{0,1,2,3\}^m,\nu\neq 0}\mathrm{tr}[\sigma_\nu\otimes\sigma_\nu\otimes\mathbb{1}\rho(0)]^2\mathbf{P}\left\{X_t(w(\nu))\leq fn\right\}\\ &= \sum_{\ell=1}^m\binom{m}{\ell}3^\ell\mathbf{P}\left\{X_t(\ell)\leq fn\right\},\end{aligned}$$

where we used the same notation as in the proof of Theorem 3.1: $X_t(\ell)$ is the random variable denoting the weight of the Pauli string obtained after applying t random gates to the operator σ_μ for some μ of weight ℓ.

If $t > cn\log^2 n$, we can apply Theorem A.1, and obtain

$$\begin{aligned}\mathbf{E}\left\{\left\|\rho(t)_{MS}-\frac{\mathbb{1}}{2^{m+fn}}\right\|_1^2\right\} &\leq \frac{1}{\mathrm{poly}(n)}\cdot\sum_{\ell=1}^m\binom{m}{\ell}3^\ell\frac{1}{2^\ell\binom{n}{\ell}}+(4^m-1)\cdot\frac{2^{f\log 3+h(f)}}{\binom{n}{n/2}3^{n/2}}\\ &= \frac{1}{\mathrm{poly}(n)}\cdot\sum_{\ell=1}^m\frac{m(m-1)\cdots(m-\ell+1)}{n(n-1)\cdots(n-\ell+1)}\left(\frac{3}{2}\right)^\ell+(4^m-1)\cdot\frac{2^{f\log 3+h(f)}}{\binom{n}{n/2}3^{n/2}}.\end{aligned}$$

This means that provided m and f are small enough, most of the random circuits with $t = O(n\log^2 n)$ gates are good encoders that allow the (approximate) correction of any fn erasure when entanglement assistance is available. In other words, if we write $m = \beta n$, then as long as $\beta < 2/3$ and $2\beta + f\log 3 + h(f) - 1 - \log 3/2 < 0$, the reference system M is decoupled from any subset of at most fn qubits of the output.

We now move to the case where the state $|\psi\rangle_{A'}$ is pure. In this case the initial state on $MM'A'$ can be written as

$$\rho(0) = \frac{1}{2^{2m}}\cdot\frac{1}{2^{n-m}}\sum_{\nu\in\{0,1,2,3\}^m,\mu\in\{0,3\}^{n-m}}\sigma_\nu\otimes\sigma_\nu\otimes\sigma_\mu.$$

Then, the analysis is the same

$$\begin{aligned}\mathbf{E}\left\{\left\|\rho(t)_{MS}-\frac{\mathbb{1}}{2^{m+fn}}\right\|_1^2\right\} &= \sum_{\nu\in\{0,1,2,3\}^m,\mu\in\{0,3\}^{n-m},,(\nu,\mu)\neq 0}\mathrm{tr}[\sigma_\nu\otimes\sigma_\mu\rho(0)_{M'A'}]^2\mathbf{P}\left\{X_t(w(\nu)+w(\mu))\leq fn\right\}\\ &= \sum_{\ell=1}^n\left(\sum_{p=0}^{\min(\ell,m)}\binom{m}{p}3^\ell\binom{n-m}{\ell-p}\right)\mathbf{P}\left\{X_t(\ell)\leq fn\right\} \qquad (7)\\ &\leq \sum_{\ell=1}^n\left(\sum_{p=0}^{\min(\ell,m)}\binom{m}{p}3^\ell\binom{n-m}{\ell-p}\right)\frac{1}{\binom{n}{\ell}2^\ell\,\mathrm{poly}(n)}+2^{m+n}\frac{2^{f\log 3+h(f)}}{\binom{n}{n/2}3^{n/2}}. \qquad (8)\end{aligned}$$

If $m = \beta n$, the second term vanishes provided $\beta + f\log 3 + h(f) - \log 3/2 < 0$. For the first term, we need to analyze carefully the number of Pauli strings $\sigma_\nu\otimes\sigma_\mu$ with a given weight ℓ which does not have an expression that is as simple as in the entanglement assisted case. Our objective is to prove that

$$\sum_{p=0}^{\min(\ell,m)}\binom{m}{p}3^\ell\binom{n-m}{\ell-p}\leq\ell\binom{n}{\ell}2^\ell.$$

using the fact that m is not too large, so that we get a vanishing bound on the trace distance. We bound the terms for $p\leq\ell/2$ and $p\geq\ell/2$ separately. We have

$$\begin{aligned}\sum_{p=0}^{\ell/2}\binom{m}{p}3^\ell\binom{n-m}{\ell-p} &\leq 3^{\ell/2}\sum_{p=0}^{\ell/2}\binom{m}{p}\binom{n-m}{\ell-p}\\ &\leq 2^\ell\binom{n}{\ell}.\end{aligned}$$

For $p \geq k/2$, this needs a bit more work. We have

$$\binom{m}{p}\binom{n-m}{\ell-p} = \frac{m(m-1)\cdots(m-p+1)\cdot(n-m)(n-m-1)\cdots(n-m-(\ell-p)-1)}{p!(\ell-p)!}.$$

First we have $p!(\ell-p)! \geq ((\ell/2)!)^2 \geq \frac{\ell!}{2^\ell}$. We also have

$$m(m-1)\cdot(m-p+1) \leq \left(\frac{m}{n}\right)^p n(n-1)\cdots(n-p+1).$$

Moreover, as $p \leq m$, we have

$$(n-m)\cdots(n-m-\ell+p+1) \leq (n-p)\cdots(n-p-\ell+p+1).$$

As a result, for $p \geq \ell/2$,

$$3^p\binom{m}{p}\binom{n-m}{\ell-p} \leq 3^\ell \cdot 2^\ell \left(\frac{m}{n}\right)^{\ell/2} \cdot \binom{n}{\ell}.$$

By choosing $m/n \leq 1/9$, we can now bound the number of Pauli strings of weight ℓ:

$$\sum_{p=0}^{\min(\ell,m)} \binom{m}{p} 3^\ell \binom{n-m}{\ell-p} \leq \ell 2^\ell \binom{n}{\ell}.$$

Returning to (8), provided $m/n < 1/9$ we get

$$\mathbf{E}\left\{\left\|\rho(t)_{MS} - \frac{\mathbb{1}}{2^{m+fn}}\right\|_1^2\right\} \leq \sum_{\ell=1}^{n} \frac{\ell}{\mathrm{poly}(n)} + 2^{n+m}\cdot\frac{2^{f\log 3 + h(f)}}{\binom{n}{n/2}3^{n/2}}.$$

We can then obtain the claimed results by parallelizing these sequential circuits (Proposition 3.2). □

Another way of interpreting the analysis above is that random quantum circuits define good stabilizer codes, i.e., codes with a positive rate and linear distance. Such a result can be seen as a step towards understanding the complexity of encoding into good quantum error correcting codes. There are many results on various classical versions of this problem; see e.g., [GHK+12] for a recent result in this spirit.

Theorem 3.6 (Good codes from low-depth circuits)**.** *There exists non-degenerate stabilizer codes with encoding circuits of depth $O(\log^3 n)$ encoding βn qubits into n qubits and having a minimum distance αn for some constants $\alpha, \beta > 0$.*

Proof

It is not hard to see that a circuit composed of Clifford gates that maps all Pauli strings of the form $\sigma_\nu \otimes \sigma_\mu$ with $\nu \in \{0,1,2,3\}^m$ and $\mu \in \{0,3\}^{n-m}$ into Pauli strings of weight at least fn defines a stabilizer code encoding m qubits and having distance fn. That is exactly what the analysis in the proof of Theorem 3.5 shows. □

The results in this section involve random quantum circuits of depth $O(\log^3 n)$. It would be interesting to improve these scrambling times to $O(\log n)$ instead. Our second set of results presented in the following section proves a weaker notion of scrambling in depth $O(\log n)$. This notion of scrambling is particularly relevant in the study of the black hole information paradox question. In addition, we also consider this notion of scrambling when the interaction graph is a d-dimensional lattice.

4 Scrambling and the black hole information paradox

Here we show that there are natural random quantum circuit models that perform good entanglement assisted codes for the erasure channel. Note that this is a weaker notion of scrambling since we require that only a constant number of initial low weight Pauli strings are brought to linear weight strings with high probability. In this section we will consider a different model of random quantum circuit where gates are selected from among sets of matchings between neighbors on lattices of fixed dimension and from the complete graph. For the d-dimensional models, to aid in our proofs we introduce an additional set of coarse grained blocks of size $O(\log n)$ between which disallow gates to be performed for coarse time steps of $O(\log^2 n)$. We show that a constant size message for

typical random quantum circuits of depth $n^{1/d}\log^2 n$ and $\log n$ for random quantum circuits with gates that act on a bounded number of qubits between neighbors on d-dimensional lattices and two-qubits gates on the complete (infinite dimensional) graph. For quantum circuits consisting of gates of fixed weight a straightforward upper bound is given by the radius of the interactions graphs of $n^{1/d}$ and $\log n$, so that our results are essentially optimal. Assuming the random quantum circuit models accurately capture the scaling behavior of typical Hamiltonians with the same interaction graph, as argued in [HP07] these time scales determine the time at which a quantum state which falls into a black hole sometime after half the black hole has evaporated will be accessible from an observer who knows the dynamics of the black hole and has been collecting all off the Hawking radiation.

4.1 Parallel circuit model on the complete graph

Recall that in the parallel circuit model, a random maximum matching of the qubits is chosen and a random gate is applied on each edge of the matching. Consider figure 1 with the systems M and M' having a constant size m (think of M' as the message), and A and A' are in a maximally entangled state. Clearly, if we have access to the whole output we can recover the message, i.e., a purification of M. The following theorem proves that if we apply a parallel random circuit of depth $O(\log n)$, a sufficiently large constant number of randomly chosen qubits of the output together with the system A are sufficient for approximately recovering the message. As mentioned earlier, this is equivalent to proving that the system M is decoupled from a subset of the qubits of size $n-c$ for some constant c.

Theorem 4.1. *Let $\epsilon > 0$ and m be a constant. In the setting described above, we have on average over a randomly chosen T of size $n-c$ for some sufficiently large c (depending on ϵ and m) such that*

$$\left\| \rho_{MT} - \frac{\mathbb{1}}{2^m} \otimes \frac{\mathbb{1}}{2^{|T|}} \right\|_1 \leq \epsilon$$

with probability $1 - O(\log^3 n/n)$ over the choice of the circuit. Here, ρ_{MT} refers to the state you obtain by applying a random parallel circuit of depth $O(\log n)$.

Proof We can write the initial state on $MM'A'$ as

$$\rho(0) = \frac{1}{2^{2m}} \sum_{\nu \in \{0,1,2,3\}^m} \sigma_\nu \otimes \sigma_\nu \otimes \frac{\mathbb{1}}{2^{n-m}}.$$

Then we have, as in the proof of Theorem 3.1,

$$\begin{aligned}
\mathbf{E}\left\{ \left\| \rho(t)_{MT} - \frac{\mathbb{1}}{2^{m+n-c}} \right\|_1^2 \right\} &\leq \sum_{\nu \in \{0,1,2,3\}^m, \mu \in \{0,1,2,3\}^T, (\nu,\mu) \neq 0} \mathbf{E}\left\{ \mathrm{tr}[\sigma_\nu \otimes \sigma_\mu \otimes \mathbb{1}\rho(t)]^2 \right\} \\
&= \sum_{\nu \in \{0,1,2,3\}^m, (\nu,\mu) \neq 0} \mathbf{E}\left\{ \mathrm{tr}[\sigma_\nu \otimes \sigma_\mu \rho(t)]^2 \right\} \\
&= \sum_{\nu \in \{0,1,2,3\}^m, \nu \neq 0} \mathrm{tr}[\sigma_\nu \otimes \sigma_\nu \otimes \mathbb{1}\rho(0)]^2 \mathbf{P}\left\{ S_t(\mathrm{supp}(\nu)) \subseteq T \right\} \\
&= O(\mathbf{P}\left\{ S_t(\{1\}) \subseteq T \right\})
\end{aligned}$$

where we defined the Markov chain $\{S_t\}$ whose state space is the set of subsets $[n]$ which corresponds to the set of non-zero Pauli operators. The transition probabilities of the Markov chain are defined as follows. We start by choosing a random maximum matching of the nodes. For each edge $\{i,j\}$ of the matching, we do the following: if neither i nor j are in S_t, they are still not in S_{t+1}, but if one of the nodes $\{i,j\}$ is in S_t, then with probability $9/15$, i and j are in S_{t+1} and with probability $3/15$, $i \in S_{t+1}$ and $j \notin S_{t+1}$ and with probability $3/15$, $j \in S_{t+1}$ and $i \notin S_{t+1}$. As before, we use the notation $S_t(A)$ when the Markov chain starts in the state A. Here our Markov chain is assumed to start in the state $\{1\}$, so we will drop the $(\{1\})$ from now on.

Lemma 4.2. *For a sufficiently large constant c and $t \geq c\log n$ and sufficiently small constant f, we have*

$$\mathbf{P}\left\{ |S_t| \leq fn \right\} \leq O\left(\frac{\log^3 n}{n} \right).$$

Before proving the lemma, we just note that it is sufficient to prove the desired result. In fact, for a randomly chosen T of size $n-c$, we have

$$\begin{aligned}\mathbf{P}\{S_t \subseteq T\} &\leq \mathbf{P}\{\forall x \in [n]-T, x \notin S_t, |S_t| > fn\} + \mathbf{P}\{|S_t| \leq fn\} \\ &\leq \mathbf{P}\{x \notin S_t, |S_t| > fn\}^c + O\left(\frac{\log^3 n}{n}\right) \\ &\leq (1-f)^c + O\left(\frac{\log^3 n}{n}\right)\end{aligned}$$

where x is uniformly distributed on $[n]$. This proves the theorem. □

Proof [of Lemma 4.2] The analysis has two steps. The first part of the proof deals with the case where $S = O(\log n)$ and the second part with the case $S = \Omega(\log n)$.

Define $T_1 = \min\{t : |S_t| \geq 10\log n\}$. The pre-factor 10 is chosen only for concreteness and can of course be chosen to be any constant and the statement remains unchanged. We start by proving that

$$\mathbf{P}\{T_1 \geq c_1 \log n\} = O\left(\frac{\log^3 n}{n}\right).$$

Let E be the event that for all $s \leq c_1 \log n$, nodes $i, j \in S_s$ never get matched. We have

$$\mathbf{P}\{T_1 \geq c_1 \log n\} = \mathbf{P}\{T_1 \geq c_1 \log n, \mathsf{E}\} + \mathbf{P}\{T_1 \geq c_1 \log n, \mathsf{E}^c\}. \tag{9}$$

Let us analyze the second term first. For this, we denote the matching by $\{(M_k^1, M_k^2)\}_{1 \leq k \leq n/2}$.

$$\begin{aligned}\mathbf{P}\{T_1 \geq c_1 \log n, \mathsf{E}^c\} &= \mathbf{P}\left\{\exists s \in [c_1 \log n], k \in [n/2] : M_k^1, M_k^2 \in S_s, T_1 \geq c_1 \log n\right\} \\ &\leq \sum_{s=1}^{c_1 \log n} \mathbf{P}\left\{\exists k \in [n/2] : M_k^1, M_k^2 \in S_s, S_s \leq 10\log n\right\} \\ &\leq c_1 \log n \cdot \frac{10\log n \cdot (10\log n - 1)}{2n} \\ &\leq \frac{100 c_1 \log^3 n}{n}.\end{aligned}$$

We now focus on the first term in (9). Because E holds, we know that $|S_{s+1}| \geq |S_s|$ for all $s \in [c_1 \log n]$. More precisely, if $|S_s| = k$, we have $|S_{s+1}|$ is distributed as $k + \mathrm{Bin}(k, 9/15)$. But using a Chernoff-Hoeffding bound, we have

$$\mathbf{P}\{\mathrm{Bin}(k, 9/15) \leq k/2\} \leq e^{-\frac{1}{2\cdot 3/5}k(3/5-1/2)^2} \leq e^{-k/200}.$$

We can now define the times $T(i) = \min\{t : |S_t| \geq 2^i\}$. With this notation, and letting $m = \log(10\log n)$, we have $T_1 = T(m)$. We can write

$$\begin{aligned}\mathbf{P}\{T_1 \geq c_1 \log n, \mathsf{E}\} &\leq \mathbf{P}\left\{T(1) \geq c_1' 2^{m-1}\right\} + \mathbf{P}\left\{T(1) < c_1' 2^{m-1}, T(2) \geq c_1'(2^{m-1} + 2^{m-2})\right\} + \\ &\qquad \cdots + \mathbf{P}\left\{T(1) + \cdots + T(m-1) < c_1'(2^{m-1} + \ldots 2^1), T(m) \geq c_1'(2^{m-1} + \cdots + 1)\right\} \\ &\leq e^{-1/200 \cdot c_1' 2^{m-1}} + e^{-2/200 \cdot c_1' 2^{m-2}} + \cdots + e^{-2^{m-1}/200 \cdot c_1'} \\ &= m \cdot e^{-c_1' 2^{m-1}/200}.\end{aligned}$$

where c_1' is chosen so that $c_1'(2^m - 1) = c_1 \log n$. For c_1 (or equivalently c_1') large enough, this expression is at most $1/n$.

For the second part, we consider a large enough subset $S \subseteq [n]$ and we prove that in one step of the random circuit, the size will increase by a constant fraction with high probability. Now it is not possible to assume that we do not have any gates within S itself. But the fact that S is large, we can have better concentration. First given an S, let N_S be the number of gates that are between two nodes of S. It is easy to see that the expected number of such gates is $\mathbf{E}\{N_S\} = \frac{|S|(|S|-1)}{2} \cdot \frac{1}{n-1}$. Actually what we want is to bound $\mathbf{P}\{N_S > \beta|S|\}$ where β is some small constant to be fixed later. We could use a straight Markov inequality

$$\mathbf{P}\{N_S > \beta|S|\} \leq \frac{\mathbf{E}\{N_S\}}{\beta|S|} = \frac{(|S|-1)}{2\beta(n-1)},$$

which is good enough for $|S| = o(n)$ but does not give a good bound for linear $|S|$. That's why we compute the second moment of N_S.

$$\begin{aligned}\mathbf{E}\left\{N_S^2\right\} &= \sum_{i<j,k<l} \mathbf{E}\left\{\mathbb{1}_{(i,j)}\mathbb{1}_{(k,l)}\right\} \\ &= \sum_{i<j} \mathbf{E}\left\{\mathbb{1}_{(i,j)}\right\} + \sum_{i<j,k<l,k,l\notin\{i,j\}} \mathbf{E}\left\{\mathbb{1}_{(i,j)}\mathbb{1}_{(k,l)}\right\} \\ &= \frac{|S|(|S|-1)}{2}\frac{1}{n-1} + \frac{|S|(|S|-1)}{2}\cdot\frac{(|S|-2)(|S|-3)}{2}\cdot\frac{1}{(n-1)(n-3)}.\end{aligned}$$

where $\mathbb{1}_{(i,j)}$ is one if there is a gate applied between nodes i and j, which are both in S. Thus the variance is equal to

$$\begin{aligned}\mathbf{E}\left\{N_S^2\right\} - \mathbf{E}\left\{N_S\right\}^2 &= \frac{|S|(|S|-1)}{2}\frac{1}{n-1} + \frac{|S|(|S|-1)}{2}\cdot\frac{(|S|-2)(|S|-3)}{2}\cdot\frac{1}{(n-1)(n-3)} - \frac{|S|^2(|S|-1)^2}{4}\cdot\frac{1}{(n-1)^2} \\ &= \frac{|S|(|S|-1)}{2}\frac{1}{n-1}\left(1 + \frac{(|S|-2)(|S|-3)}{2}\cdot\frac{1}{n-3} - \frac{|S|(|S|-1)}{2}\frac{1}{n-1}\right).\end{aligned}$$

But $(|S|-2)(|S|-3)(n-1) = (|S|^2-5|S|+6)(n-1) = (|S|^2-|S|)(n-1) - 2(2|S|-3)(n-1)$ and we compare that to $(|S|^2-|S|)(n-3) = (|S|^2-|S|)(n-1) - 2(|S|(|S|-1))$. The first term is smaller than the second one provided $|S| \geq 3$ and $|S| \leq n$ (which is the case). Thus we can bound the variance by the expected value:

$$\mathbf{E}\left\{N_S^2\right\} - \mathbf{E}\left\{N_S\right\}^2 \leq \mathbf{E}\left\{N_S\right\}.$$

By applying Chebyshev's inequality, we have for any $\gamma > 0$,

$$\mathbf{P}\left\{N_S > (1+\gamma)\mathbf{E}\left\{N_S\right\}\right\} \leq \frac{1}{\gamma^2\mathbf{E}\left\{N_S\right\}}.$$

Now if β is such that $|S| < 2\beta n$, we define γ so that $(1+\gamma) = \beta|S|/\mathbf{E}\left\{N_S\right\} = 2\beta\frac{n-1}{|S|-1}$. As a result,

$$\begin{aligned}\mathbf{P}\left\{N_S > \beta|S|\right\} &\leq \frac{1}{(\beta\frac{|S|}{\mathbf{E}\{N_S\}} - 1)^2\mathbf{E}\left\{N_S\right\}} \\ &= \frac{2(n-1)}{(\beta\frac{2(n-1)}{|S|-1} - 1)^2|S|(|S|-1)} \\ &\leq \frac{2(n-1)}{(2\beta(n-1) - (|S|-1))^2} \\ &= O\left(\frac{1}{n}\right),\end{aligned}$$

provided for example $|S| - 1 \leq \beta(n-1)$.

We proved that we can assume that the number of gates within S is small. For the gates that associate a node in S with a node outside S, we need to prove that many of these gates lead to a Pauli operator of weight two so that we obtain an overall increase in the size of S. In fact, provided $N_S < \beta|S|$, the number of non-zero Pauli operator is distributed at least as $(1-\beta)|S| + \mathrm{Bin}((1-2\beta)|S|, 3/5)$. Now we can bound using a standard Chernoff bound

$$\begin{aligned}\mathbf{P}\left\{\mathrm{Bin}((1-2\beta)|S|, 3/5) < (\beta+1/4)\cdot|S|\right\} &\leq \exp\left(-\frac{5}{6}\cdot\frac{(3/5(1-2\beta)|S| - (\beta+1/4)|S|)^2}{(1-2\beta)|S|}\right) \\ &= \exp\left(-\frac{5|S|}{6}\cdot\frac{(7/20 - (6/5+1)\beta)^2}{(1-2\beta)}\right).\end{aligned}$$

For sufficiently small β and sufficiently large $|S| = \Omega(\log n)$, this probability is $O(1/n)$.

This proves that provided $c\log n \leq |S_s| \leq \beta n$ for a sufficiently large c and sufficiently small β, then we have $|S_{s+1}| \geq 5/4\cdot|S_s|$. Together with the first part of the proof, we obtain that after $O(\log n)$ steps of the random circuit we have $|S_t| \geq \beta n$ with probability $1 - O(\log^3 n/n)$. □

4.2 Random circuit on a d-dimensional lattice

We now turn to examining the depth of a random quantum circuit restricted to nearest neighbours on a d-dimensional square lattice, required to scramble a constant number of initial low weight Pauli strings. As was shown in the previous section, this determines the depth at which we obtain entanglement assisted codes.

We analyze scrambling in a model for which a partial parallelization has been performed of a sequential random quantum circuit on on a d-dimensional lattice. By a d-dimensional sequential random quantum circuit we mean one for which a random two-qubit gate selected according to the Haar measure or uniformly from the Clifford group is applied to a pair of qubits selected uniformly from among nearest neighbors on a square d-dimensional lattice with open boundary conditions. The specific model consists of partitioning the lattice into coarse grained cells and in each time step performing a random two-qubit gate between a randomly selected pair of nearest neighbour within each cell. We consider two equivalent coarse grainings of the lattice into square cells of size $O(\log n)$, such that the midpoints of the cells of the first set are the corners of the cells of the second set. We then have cells of type 1 corresponding to the first coarse graining and cells of type 2 corresponding to the second coarse graining. In alternating coarse time steps, gates are applied within each of the the cells of one set at a time. In each coarse time step a total of $O(\log^2 n)$ gates will be performed.

The following theorem proves that after $O((\frac{n}{\log n})^{1/d})$ coarse grained time steps, a sufficiently large constant number of randomly chosen qubits of the output together with the system A are sufficient for approximately recovering the message (see Figure 1). As mentioned earlier, this is equivalent to proving that the system M is decoupled from a subset of the qubits of size $n - c$ for some constant c.

Theorem 4.3. *Let $\epsilon > 0$ and m be a constant. In the setting described above, we have on average over a randomly chosen T of size $n - c$ for some sufficiently large c (depending on ϵ) such that*

$$\left\| \rho_{MT} - \frac{\mathbb{1}}{2^m} \otimes \frac{\mathbb{1}}{2^{|T|}} \right\|_1 \leq \epsilon$$

with probability $1 - O(1/n)$ over the choice of the circuit. Here, ρ_{MT} refers to the state you obtain by applying a random quantum circuit of depth $O(n^{1/d} \log^2(n))$ as described above.

Proof As in the proof of Theorem 4.1, we only need to show that a Pauli string of weight one becomes a Pauli string of linear weight within $O(n^{1/d} \log^2 n)$ time steps with high probability. We start by proving (in Theorem 4.4 below) a lower bound of $\Omega(1/n)$ on the gap of the second moment operator of a sequential random quantum circuit on n qubits with a d-dimensional lattice interaction graph (or equivalently on the corresponding the Markov chain described in Section 2.2). Then by a standard argument, one can obtain an upper bound on the mixing time of the Markov chain; see e.g., [MT06]. That is after $t = O(\frac{1}{\Delta}(n + \log(1/\delta)))$, the distribution on Pauli strings is δ-close to the the stationary distribution of the Markov chain which is the uniform distribution over all non-zero Pauli strings.

We then apply this result to a cell which is a d-dimensional lattice with $O(\log n)$ nodes. Then partition the cell into 2^d d-dimensional sub-cells with half the length of the original cell. Then, if we choose δ to be inverse polynomial with a sufficiently large power, and applying $O(\log n(\log n + \log(1/\delta))) = O(\log^2 n)$ gates, a non-zero Pauli string in the cell gets mapped to a Pauli string whose support contains at least on element in each one of these sub-cells with probability $1 - 1/\operatorname{poly}(n)$. We can summarize this as follows: in each successful coarse time step every cell that contains a non-zero Pauli string gets mapped to a Pauli string that has support in each one of its sub-cells. By choosing the constants appropriately, we can make the probability of success of a coarse step to be $1 - 1/\operatorname{poly}(n)$.

Consider now the following coarse time step that uses the alternate coarse graining. What the success of the previous coarse step is saying is that a cell with non-zero Pauli weight has contaminated all the cells of type 2 that overlap with it. By the same argument as in the previous paragraph, we see that in the coarse step involving cells of type 2, each one of these contaminated cells of type 2 will in turn contaminate the cells of type 1 that overlap with it. Thus, by repeating these alternate steps a number of times corresponding to the diameter of the graph of cells $O\left(\left(\frac{n}{\log n}\right)^{1/d}\right)$, we reach a Pauli string of linear weight. □

Theorem 4.4. *The spectral gap, Δ_{ds}, of the second-order moment operator for the d-dimensional sequential random quantum circuit described above is bounded from below by $\Delta_{ds} \geq \frac{a}{n}$ for a constant a.*

Proof The second-order moment operator of a sequential one-dimensional random quantum circuit is of the form,

$$M_{1s} = \frac{1}{n-1} \sum_{i=1}^{n-1} m_{ii+1}.$$

We will use the fact that the gap, Δ_{1s}, of the second-order moment operator of a sequential $1D$ random quantum circuit was shown in [Žni08, BĆH+12] to be lower bounded by $\Delta_{1s} \geq \frac{a}{n}$, to show a similar lower bound on the spectral gap, Δ_q, of the second-order moment operator for a non-uniform sequential $1D$ random quantum circuit, for which the probability of applying a gate to qubits i and $i+1$ is $\frac{q_{ii+1}}{n-1}$ with $q_{ii+1} > 0$. We next show how to write the second-order moment operator of the d-dimensional sequential random quantum circuit as a convex sum of such non-uniform $1D$ sequential random quantum circuits, which we will lead to the desired bound by using the following lemma on the convexity of the spectral gap.

Lemma 4.5. *For two random quantum circuits whose gate distributions are universal and are invariant under Hermitian conjugation, with second-order moment operators M_1 and M_2 with spectral gaps of Δ_1 and Δ_2 respectively, the second-order moment operator describing any convex combination of the two random quantum circuits, $M = p_1 M_1 + p_2 M_2$, has a gap Δ that is lower bounded by $\Delta \geq p_1 \Delta_1 + p_2 \Delta_2$.*

Proof We use the fact that every second-order moment operator of a random quantum circuit over a universal gate set has the same space of fixed points, $\mathcal{V}$, onto which the second-order moment operator of the Haar measure, M_{haar}, is the projector. Consequently, we may define the following operators $\tilde{M} = M - M_{\text{haar}}$, $\tilde{M_1} = M_1 - M_{\text{haar}}$, $\tilde{M_2} = M_2 - M_{\text{haar}}$ which sets the eigenvalue of this eigenspace to 0. Now by the triangle inequality it follows that,

$$\|\tilde{M}\|_{sp} \leq p_1 \|\tilde{M_1}\|_{sp} + p_2 \|\tilde{M_2}\|_{sp},$$

where $\| \ \|_{sp}$ is the spectral norm. Since invariance under Hermitian conjugation of the gate distribution implies that the moment operators are Hermitian, it follows that $\lambda \leq p_1 \lambda_1 + p_2 \lambda_2$, where λ, λ_1 and λ_2 are the subdominant eigenvalues M, M_1 and M_2, respectively. Since $\Delta = 1 - \lambda$, the lemma follows. □

The second-order moment operator for a non-uniform $1D$ sequential random quantum circuit is given by,

$$M_q = \frac{1}{n-1} \sum_{i=1}^{n-1} q_{ii+1} m_{ii+1}.$$

Observe that because m_{ii+1} are positive semidefinite, we have

$$M_q \geq \min_i q_{ii+1} \cdot \frac{1}{n-1} \sum_{i=1}^{n-1} m_{ii+1}.$$

This implies that the spectral gaps satisfy $\Delta_q \geq \min_i q_{ii+1} \Delta_{1s} = \min_i q_{ii+1} \cdot \Omega(1/n)$.

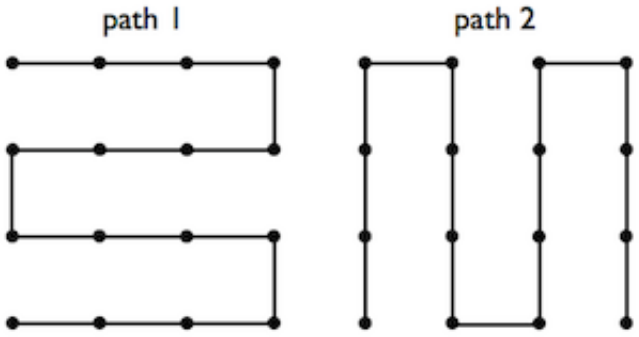

Figure 2: A set of non-intersecting 1D-paths on a 2D 4x4 square lattice

The goal now is to write the second moment operator for the d-dimensional circuit as a convex combination of second moment operators for one-dimensional circuits. For this, we find a set of paths on the d-dimensional

lattice such that each path includes every vertex and each edge is included in at least one path. Such a set may be constructed using d paths where the i-th path consists of every edge oriented in the i-th direction plus some perpendicular edges on the surface of the lattice. An illustration of such a set is given in Figure 2 for $d = 2$. For such a set of paths every internal edge is traversed by only one path, while an external edge may be traversed by as many as d paths. Thus, one may write the d-dimensional sequential random quantum circuit as an average of d one-dimensional non-uniform sequential random quantum circuits where the pair with the lowest probability is $1/d$ of that of the largest. Lemma 4.5 now implies that the gap, Δ_{ds}, of the d-dimensional sequential random quantum circuit is bounded by $\Delta_{ds} \geq \frac{1}{d}\Delta_{1s} \geq \frac{a}{n}$ for a constant a. □

4.3 Lower bound on the scrambling time

For a circuit of depth t consisting of gates, each of which act on at most k-qubits, an initial Pauli operator of weight 1 can only have support on a qubit distance kt away. Thus, on a d-dimensional lattice it may have support on at most $(kt)^d$ qubits, implying a depth of at least $\Omega(n^{1/d})$ for the weigh to be linear in n. For a random quantum circuit on the complete graph, the weight of a Pauli operator may increase by at most a factor of k in each time step, yielding an lower bound on the depth required for scrambling of $\Omega(\log n)$. Thus, the time at which most random quantum circuit scramble is within a constant or $O(\log^2 n)$ factor of the fastest possible circuit on the complete graph and a d-dimensional graph respectively. We think that the $O(\log^2 n)$ factor in the case of the d-dimensional lattice is an artifact of our proof technique.

5 Outlook

An interesting question is whether decoupling occurs in the general setting of [Dup10, DBWR10] with random circuits of depth $O(\mathrm{poly}(\log n))$. This would imply that random encoding circuits of $O(\mathrm{poly}(\log n))$ depth generate codes that are close to achieving the capacity of the erasure channel.

Since the scrambling condition in d-dimensional random quantum circuits utilizes a weak bound on the success probability of filling a sufficient number of cells, we think it may be possible to tighten our result to show strong scrambling, and thus stronger decoupling results for circuits of depth $O(n^{1/d} \log n)$ on d-dimensional lattices. It would be interesting to see if the course graining technique employed here can be used to show that the depth at which random quantum circuits are ϵ-approximate k-designs also scales with the radius of the interaction graph as conjectured in [BHH12].

It is known that the unitary generated by an arbitrary local Hamiltonian at time, t, which is a polynomial in the size, n, of the system can be approximated by a circuit consisting of single and two-qubit gates whose depth is polynomial in t [PQSV11]. Thus, whether our results imply that Hamiltonian evolution scrambles quickly depends on whether typical time independent Hamiltonians explore sufficiently uniformly the measure accessible to them at times sublinear in n. This question appears to be linked with the approach of random matrix theory [Sre94, RDO08] to understand thermalization under dynamics generated by strongly-nonintegrable, time independent Hamiltonians, whereby the eigenstates of the Hamiltonian resemble those drawn uniformly from an appropriate matrix ensemble. It would be interesting to further explore the connection between quantum chaos, properties of random quantum circuits and quantum aspects of thermalization such as scrambling and decoupling.

Acknowledgements

We would like to thank Patrick Hayden and David Poulin for helpful discussions. The research of WB is supported by the Centre de Recherches Mathématiques at the University of Montreal, Mprime, and the Lockheed Martin Corporation. The research of OF is supported by the European Research Council grant No. 258932, and was started while he was affiliated with McGill University.

A Analysis of the Markov chain

Theorem A.1. *Let $X_t(\ell)$ be the random variable representing the position of the random walk starting at ℓ after t steps. There is a constant c such that for any $f < 1/2$ and $t \geq cn\log^2 n$ and all $\ell \in \{1, \ldots, n\}$,*

$$\mathbf{P}\left\{X_t(\ell) \leq fn\right\} \leq \frac{2^{f\log 3 + h(f)}}{\binom{n}{n/2}3^{n/2}} + \frac{1}{2^\ell\binom{n}{\ell}}\frac{1}{\mathrm{poly}(n)}.$$

Proof The general strategy of the proof is as follows. First, we pick a reference point r (which is going to be $n/2$) for which we can bound $\mathbf{P}\left\{X_t(r) \leq fn\right\}$ easily. Then we will prove that for any $\ell < r$, starting at ℓ, we will reach r within t steps with high probability.

The stationary distribution of the chain is given by $\pi(k) = \frac{3^k\binom{n}{k}}{4^n-1}$ (see [HL09, Lemma 5.3]). As a result, we have for any $t \geq 1$,

$$\begin{aligned}\frac{1}{4^n-1}\sum_{\ell=1}^{n} 3^\ell\binom{n}{\ell}\mathbf{P}\left\{X_t(\ell) \leq fn\right\} &= \frac{1}{4^n-1}\sum_{\ell=1}^{fn} 3^\ell\binom{n}{\ell} \\ &\leq \frac{3^{fn}2^{h(f)n}}{4^n-1} \\ &= \frac{2^{(f\log 3+h(f))n}}{4^n-1}.\end{aligned}$$

This allows us to bound the probability of the event $[X_t(n/2) \leq fn]$. In fact, for any t,

$$\begin{aligned}\mathbf{P}\left\{X_t(n/2) \leq fn\right\} &= \frac{4^n-1}{\binom{n}{n/2}3^{n/2}}\cdot\frac{\binom{n}{n/2}3^{n/2}}{4^n-1}\mathbf{P}\left\{X_t(n/2) \leq fn\right\} \\ &\leq \frac{4^n-1}{\binom{n}{n/2}3^{n/2}}\cdot\frac{1}{4^n-1}\sum_{\ell=1}^{n} 3^\ell\binom{n}{\ell}\mathbf{P}\left\{X_t(\ell) \leq fn\right\} \\ &\leq \frac{2^{f\log 3+h(f)}}{\binom{n}{n/2}3^{n/2}}.\end{aligned} \tag{10}$$

Moreover, note that for $\ell \geq n/2$, we have

$$\begin{aligned}\mathbf{P}\left\{X_t(\ell) \leq fn\right\} &\leq \max_{1\leq s\leq t}\mathbf{P}\left\{X_s(n/2) \leq fn\right\} \\ &\leq \frac{2^{f\log 3+h(f)}}{\binom{n}{n/2}3^{n/2}}\end{aligned}$$

The remaining case is then $\ell \leq n/2$. In this case, the objective is to show that

$$\mathbf{P}\left\{X_t(\ell) \leq fn\right\} \leq \frac{2^{f\log 3+h(f)}}{\binom{n}{n/2}3^{n/2}} + \frac{1}{2^\ell\binom{n}{\ell}}\frac{1}{\mathrm{poly}(n)}.$$

Define $T = \min\{t \geq 1 : X_t(\ell) \geq n/2\}$. Note that we have for any t

$$\begin{aligned}\mathbf{P}\left\{X_t(\ell) \leq fn\right\} &\leq \mathbf{P}\left\{T < t, X_t(\ell) \leq fn\right\} + \mathbf{P}\left\{T \geq t\right\} \\ &= \mathbf{P}\left\{T < t, X_{t-T}(n/2) \leq fn\right\} + \mathbf{P}\left\{T \geq t\right\} \\ &\leq \max_{1\leq s\leq t}\mathbf{P}\left\{X_s(n/2) \leq fn\right\} + \mathbf{P}\left\{T \geq t\right\}.\end{aligned}$$

Using (10), we can bound the first term. The objective of the remainder of the proof is to bound the probability $\mathbf{P}\left\{T \geq t\right\}$ when $t = cn\log^2 n$. This is done in Lemma A.2 below. Once we have that, the result follows. □

Lemma A.2. *For a large enough constant c,*

$$\mathbf{P}\left\{T > cn\log^2 n\right\} \leq 2^{-2n} + \frac{1}{2^\ell\binom{n}{\ell}}\cdot\frac{1}{\mathrm{poly}(n)}.$$

Proof To prove this result, we start by defining an accelerated walk $\{Y_i\}$ as in [HL09] and the corresponding stopping time $S = \min\{s : Y_s \geq n/2\}$. More formally, let $N_0 = 0$ and $N_{i+1} = \min\{k \geq N_i : X_k \neq X_{N_i}\}$ and then $Y_i = X_{N_i}$. It is not hard to see that $\{Y_i\}$ is a Markov chain and the transition probabilities are given by the transition probabilities for $\{X_k\}$ conditioned on moving.

We also define the waiting time $W_i = N_{i+1} - N_i - 1$ to be the number of times the self-loop edge is taken. Conditioned on Y_i, W_i has a geometric distribution with parameter $\frac{2Y_i(3n-2Y_i-1)}{5n(n-1)}$. Notice that this distribution is stochastically dominated by a geometric distribution with parameter $\frac{2Y_i}{5n}$, which we will use instead (we are only interested in upper bounds on the waiting times).

Getting back to T, notice that $T = S + W_1 + W_2 + \cdots + W_S$. So we have for all s

$$\mathbf{P}\{T > t + s\} \leq \mathbf{P}\{S > s\} + \mathbf{P}\{S \leq s, W_1 + \cdots + W_S > t\}. \tag{11}$$

We will choose s later so that both terms are small.

Lemma A.3. *For any $s \geq 2n$, we have*

$$\mathbf{P}\{S > s\} \leq \exp(-s/8).$$

Proof For this we just use a concentration bound on the position of a random walk relative to its expectation. First we define a random walk Y_i' with $Y_0' = 0$ and it moves to the right with probability $3/4$ and to the left with probability $1/4$. Observe that the probability of moving right is at most $3/4$ for Y_i provided $Y_i \leq n/2$. For this reason, before S, we can assume that $Y_i' \leq Y_i$. In other words, we have $S' \geq S$ where $S' = \min\{i : Y_i' \geq n/2\}$. Thus,

$$\begin{aligned}
\mathbf{P}\{S > s\} &\leq \mathbf{P}\{S' > s\} \\
&\leq \mathbf{P}\{Y_s' < n/2\} \\
&= \mathbf{P}\{Y_s' < \ell + s/2 - (s/2 + \ell - n/2)\} \\
&\leq \exp\left(-\frac{(s/2 + \ell - n)^2}{2s}\right) \\
&\leq \exp(-s/8)
\end{aligned}$$

where we used the fact that $\mathbf{E}\{Y_s'\} = \ell + s/2$ and a Chernoff-type bound, see for example [HL09, Lemma A.4]. □

We now move to the second step of the proof where we analyze the waiting times $W_1 + \cdots + W_S$. Recall this is the total waiting time before the node $r = n/2$ is reached.

Lemma A.4. *We have*

$$\mathbf{P}\left\{S \leq s, W_1 + \cdots + W_S > cn\log^2 n\right\} \leq \frac{1}{2^\ell \binom{n}{\ell}} \cdot \frac{1}{\mathrm{poly}(n)}$$

Proof The techniques we use are similar to the techniques in [HL09], but we need to improve the analysis in several places. We try to use similar notation as [HL09] as much as possible.

As in the proof of [HL09, Lemma A.11], we start by defining the good event

$$\mathsf{H} = \bigcap_{x=1}^{n} \left[\sum_{k=1}^{S} \mathbb{1}(Y_k \leq x) \leq \gamma x/\mu\right],$$

where $\mu = 1/2$.[1] The parameter γ is going to be chosen later. This event is saying that states with small labels are not visited too many times. Later in the proof, we will show that the $\mathbf{P}\{\mathsf{H}^c\}$ is small. Define the random variable

[1] We use this notation to apply [HL09, Lemma A.5] later. μ corresponds to the probability of going forward minus the probability of going backward for a simplified walk that moves forward at most as fast as Y_k. In our case, we have $\mu = 1/2$ because we stop after reaching state $r = n/2$, and the probability of moving forward at $n/2$ is $3/4$.

$M = \min_{1 \le i \le S} Y_i$. We have

$$\begin{aligned}
\mathbf{P}\left\{W_1 + \ldots W_S > t, S \le s, \mathsf{H}\right\} &= \sum_{m=1}^{\ell} \mathbf{P}\left\{M = m, S \le s, W_1 + \ldots W_S > t, \mathsf{H}\right\} \\
&= \sum_{m=1}^{\ell} \mathbf{P}\left\{M = m\right\} \mathbf{P}\left\{S \le s, W_1 + \ldots W_S > t, \mathsf{H} | M = m\right\} \\
&\le \sum_{m=1}^{\ell} \mathbf{P}\left\{M \le m\right\} \max_{\{y_i\} \text{ satisfying } M=m \text{ and } \mathsf{H} \text{ and } S \le s} \mathbf{P}\left\{W(y_1) + \cdots + W(y_s) \ge t\right\},
\end{aligned} \tag{12}$$

where the maximum is taken over all sequences $y_1, \ldots, y_s$ of possible walks and $W(y)$ is the waiting time at state y.

We will bound $\mathbf{P}\{M \le m\}$ using Lemma B.1. Our random walk starts at position ℓ so that, in the notation of Lemma B.1, $p_- = \frac{6\ell(n-\ell)}{6\ell(n-\ell)+2\ell(\ell-1)}$ and for $k \ge \ell+1$, $p_+(k) = \frac{6k(n-k)}{6k(n-k)+2(\ell+1)\ell}$. As a result, we have

$$\alpha_- = \frac{6\ell(n-\ell)}{2\ell(\ell-1)} = 3 \cdot \frac{n-\ell}{\ell-1}.$$

As we stop after reaching the reference point $r = n/2$, we can bound $p_+ \ge 3/4$. As a result, we have

$$\begin{aligned}
\mathbf{P}\left\{M \le \ell - 1\right\} &\le \frac{1}{1 + 3 \cdot \frac{n-\ell}{\ell-1}(1 - 1/3)} \\
&= \frac{1}{1 + 2 \cdot \frac{n-\ell}{\ell-1}} \\
&\le \frac{1}{2} \cdot \frac{\ell-1}{n-\ell}.
\end{aligned}$$

Reaching $\ell - 2$ before r means reaching $\ell - 1$ before r starting at ℓ and reaching $\ell - 2$ before r starting at $\ell - 1$, and these parts of the walk are independent. As a result, by induction, we can then see that

$$\begin{aligned}
\mathbf{P}\left\{M \le m\right\} &\le \frac{1}{2^{\ell-m}} \cdot \frac{(\ell-1)(\ell-2)\cdots m}{(n-\ell)(n-\ell+1)\cdots(n-m-1)} \\
&= \frac{1}{2^\ell} \cdot \frac{\ell!}{n(n-1)\cdots(n-\ell+1)} \cdot \frac{2^m}{\ell(n-\ell)} \cdot \frac{n(n-1)\cdot(n-m)}{(m-1)!} \\
&\le \frac{1}{2^\ell \binom{n}{\ell}} \cdot (2n)^m.
\end{aligned} \tag{13}$$

We now look at the term $\max_{\{y_i\} \text{ satisfying } M=m \text{ and } \mathsf{H} \text{ and } S \le s} \mathbf{P}\{W(y_1) + \cdots + W(y_s) \ge t\}$. As argued in the proof of [HL09, Lemma A.11], the maximum is achieved when we make the walk visit as many times as possible the states with smaller labels. This means state m is visited $\gamma m/\mu$ times, and all $i > m$ are visited γ/μ times. So we can write

$$W(y_1) + \cdots + W(y_s) \le \sum_{i=1}^{\gamma m/\mu} G_{m,i} + \sum_{i=1}^{\gamma/\mu} \sum_{k=m+1}^{n/2} G_{k,i},$$

where $G_{k,i}$ has a geometric distribution with parameter $2k/5n$ and the random variables $\{G_{k,i}\}$ are independent. We are going to give upper tail bounds on the right hand side by computing the moment generating function. For any $\lambda \ge 0$, we have, using the moment generating function of a geometric distribution and the independence of the random variables:

$$\mathbf{E}\left\{\exp\left(\lambda\left(\sum_{i=1}^{\gamma m/\mu} G_{m,i} + \sum_{i=1}^{\gamma/\mu}\sum_{k=m+1}^{n/2} G_{k,i}\right)\right)\right\} = \left(\frac{2m/5n}{e^{-\lambda} - 1 + 2m/5n}\right)^{\gamma m/2} \prod_{k=m+1}^{n/2} \left(\frac{2k/5n}{e^{-\lambda} - 1 + 2k/5n}\right)^{\gamma/\mu}.$$

Now take λ so that $e^\lambda = \frac{1}{1-m/(5n)}$. This leads to

$$\begin{aligned}\mathbf{E}\left\{\exp\left(\lambda\left(\sum_{i=1}^{\gamma m/\mu} G_{m,i} + \sum_{i=1}^{\gamma/\mu}\sum_{k=m+1}^{n/2} G_{k,i}\right)\right)\right\} &= \left(\frac{2m}{2m-m}\right)^{\gamma m/\mu} \cdot \prod_{k=m+1}^{n/2}\left(\frac{2k}{2k-m}\right)^{\gamma/\mu} \\ &\le 2^{\gamma m/\mu}\left(\prod_{k=m+1}^{n/2} e^{\frac{m/2}{k-m/2}}\right)^{\gamma/\mu} \\ &\le 2^{\gamma m/\mu}\left(e^{m/2\cdot\ln n}\right)^{\gamma/\mu}.\end{aligned}$$

As a result, using Markov's inequality, we obtain

$$\begin{aligned}\mathbf{P}\left\{\sum_{i=1}^{\gamma m/\mu} G_{m,i} + \sum_{i=1}^{\gamma/\mu}\sum_{k=m+1}^{n/2} G_{k,i} > t\right\} &= \mathbf{P}\left\{\exp\left(\lambda\left(\sum_{i=1}^{\gamma m/\mu} G_{m,i} + \sum_{i=1}^{\gamma/\mu}\sum_{k=m+1}^{n/2} G_{k,i}\right)\right) > e^{\lambda t}\right\} \\ &\le 2^{\gamma m/\mu} e^{\gamma m/(2\mu)\cdot\ln n}\cdot(1-m/5n)^t \\ &\le 2^{\gamma m/\mu} e^{\gamma m/(2\mu)\cdot\ln n}\cdot e^{-tm/(5n)}.\end{aligned}$$

Getting back to equation (12), we have

$$\begin{aligned}\mathbf{P}\left\{W_1 + \ldots W_S > t, S \le s, \mathsf{H}\right\} &\le \frac{1}{2^\ell\binom{n}{\ell}}\sum_{m=1}^{\ell}(2n)^m 2^{\gamma m/\mu} e^{\gamma m/(2\mu)\cdot\ln n}\cdot e^{-tm/(5n)} \\ &= \frac{1}{2^\ell\binom{n}{\ell}}\sum_{m=1}^{\ell}\left(2n 2^{\gamma/\mu} e^{\gamma/(2\mu)\cdot\ln n}\cdot e^{-t/(5n)}\right)^m\end{aligned}$$

Thus, for $t > cn\log^2 n$ with sufficiently large c, this probability is bounded by $O\left(\frac{2^{-\ell}}{\binom{n}{\ell}\,\mathrm{poly}(n)}\right)$.

It now remains to bound $\mathbf{P}\left\{\mathsf{H}^c, S \le s\right\}$. Fix $x \in \{1,\ldots,n\}$, we have

$$\begin{aligned}\mathbf{P}\left\{\sum_{k=1}^{S}\mathbb{1}(Y_k \le x) > \gamma x/\mu, S \le s\right\} &\le \sum_{j=1}^{s}\mathbf{P}\left\{Y_j = x, [\forall i<j, Y_i > x], j < S, \sum_{k=1}^{S}\mathbb{1}(Y_k \le x) > \gamma x/\mu\right\} \\ &\le \sum_{j=1}^{s}\mathbf{P}\left\{Y_j = x, [\forall i<j, Y_i > x], j < S, \sum_{k=j+1}^{S_j}\mathbb{1}(Y_k \le x) \ge \gamma x/\mu\right\} \\ &\le \sum_{j=1}^{s}\mathbf{P}\left\{M \le x\right\}\cdot\mathbf{P}\left\{\sum_{k=j+1}^{S_j}\mathbb{1}(Y_k \le x) \ge \gamma x/\mu | Y_j = x, j < S\right\},\end{aligned}$$

where we defined $S_j = \min\{s \ge j+1 : Y_k \ge n/2\}$. To obtain the last inequality, we simply used the fact that $[Y_j = x, j < S] \subseteq [M \le x]$. Moreover, $[j < S]$ can be determined by looking at $Y_1, \ldots, Y_j$ and thus conditioned on $[Y_j = x]$, Y_k for $k \ge j+1$ and also S_j are independent of $[j < S]$. This means that we can drop $[j < S]$ from the conditioning.

To bound $\mathbf{P}\left\{M \le x\right\}$, we use (13). We can also bound Y_k by a simpler random walk Y'_k that moves forward with probability $3/4$, as we did in the proof of Lemma A.3. Thus, we obtain

$$\begin{aligned}\mathbf{P}\left\{\sum_{k=1}^{S}\mathbb{1}(Y_k \le x) > \gamma x/\mu, S \le s\right\} &\le \frac{1}{2^\ell\binom{n}{\ell}}(2n)^x\cdot s\cdot\mathbf{P}\left\{\sum_{k=1}^{\infty}\mathbb{1}(Y'_k \le x) \ge \gamma x/\mu | Y'_0 = 0\right\} \\ &\le \frac{1}{2^\ell\binom{n}{\ell}}(2n)^x\cdot s\cdot 2\exp\left(-\frac{\mu(\gamma-2)x}{2}\right),\end{aligned}$$

where we used [HL09, Lemma A.5]. As a result, by a union bound,

$$\begin{aligned}\mathbf{P}\left\{\mathsf{H}^c, S \le s\right\} &\le \frac{1}{2^\ell \binom{n}{\ell}} \cdot 2s \cdot \sum_{x=1}^{n} \exp\left(x\left(\log(2n) - \frac{\mu(\gamma-2)}{2}\right)\right) \\ &\le \frac{1}{2^\ell \binom{n}{\ell}} \cdot \frac{1}{\operatorname{poly} n},\end{aligned}$$

where to get the last inequality, we choose $\gamma = c' \log n$ for large enough c' and use the fact that s will be chosen linear in n. Continuing, we reach

$$\begin{aligned}\mathbf{P}\left\{W_1 + \ldots W_S > t, S \le s\right\} &\le \mathbf{P}\left\{W_1 + \ldots W_S > t, S \le s, \mathsf{H}\right\} + \mathbf{P}\left\{\mathsf{H}^c, S \le s\right\} \\ &\le \frac{1}{2^\ell \binom{n}{\ell}} \frac{1}{\operatorname{poly}(n)}.\end{aligned}$$

□

To complete the proof of Lemma A.2, we just plug the bounds obtained from Lemma A.3 with $s = 16n$ and from Lemma A.4 into equation (11). □

B An additional lemma

Consider a random walk on a line indexed from -1 to a. At positions $i > 0$, the probability of moving to the right is $p_+(i)$ (depending on i and for points $i \le 0$, the probability of moving to the right is p_-. The following lemma gives a bound on the probability of hitting the node -1 before hitting a when starting at position 0. In our setting, we are interested in the case where p_- and p_+ are (significantly) larger than $1/2$ so that the probability of hitting -1 before a is small.

Lemma B.1. *Assume $p_+(i), p_- > 1/2$. Then the probability of hitting -1 before a is exactly*

$$\frac{1}{1 + \alpha_- \cdot \frac{\prod_{j=1}^{a-1} \alpha_+(j)}{1 + \sum_{i=1}^{a-1} \prod_{j=i}^{a-1} \alpha_+(j)}},$$

where $\alpha_+(i) = \frac{p_+(i)}{1-p_+(i)}$ and $\alpha_- = \frac{p_-}{1-p_-}$. In particular, if $\alpha_+(i) = \alpha_+$ for all i, this probability becomes

$$\frac{1}{1 + \alpha_- \cdot \frac{\alpha_+^a - \alpha_+^{a-1}}{\alpha_+^a - 1}} \le \frac{1}{1 + \alpha_- \cdot (1 - 1/\alpha_+)}.$$

Proof Let P_i be the probability of first reaching -1 when starting at position i. We can write for any for $i \in [1, a-1]$, $P_i = p_+(i)P_{i+1} + (1 - p_+(i))P_{i-1}$, which can be re-written as

$$\frac{p_+(i)}{1 - p_+(i)} \left(P_i - P_{i+1}\right) = \left(P_{i-1} - P_i\right).$$

We now use the boundary condition at node a: $P_a = 0$. Thus, $(P_{a-2} - P_{a-1}) = \frac{p_+(a-1)}{1-p_+(a-1)} P_{a-1}$. Moreover, we see by induction that for any $i \ge 1$, $P_{i-1} - P_i = \left(\prod_{j=i}^{a-1} \frac{p_+(j)}{1-p_+(j)}\right) P_{a-1}$. We can now write a telescoping sum

$$P_0 - P_{a-1} = \sum_{i=1}^{a-1} P_{i-1} - P_i = \sum_{i=1}^{a-1} \prod_{j=i}^{a-1} \alpha_+(j) \cdot P_{a-1}.$$

As a result,

$$P_0 = P_{a-1} \left(1 + \sum_{i=1}^{a-1} \prod_{j=i}^{a-1} \alpha_+(j).\right).$$

We can then write $P_{-1} - P_0 = \frac{p_-}{1-p_-}(P_0 - P_1) = P_{a-1} \cdot \prod_{j=1}^{a-1} \alpha_+(j) \cdot \frac{p_-}{1-p_-}$.

Now, we use our second boundary condition $P_{-1} = 1$. We have

$$\begin{aligned} 1 = P_{-1} &= P_0 + P_{a-1} \cdot \alpha_- \prod_{j=1}^{a-1} \alpha_+(j) \\ &= P_0 \left(1 + \alpha_- \frac{\prod_{j=1}^{a-1} \alpha_+(j)}{\sum_{i=1}^{a-1} \prod_{j=i}^{a-1} \alpha_+(j)} \right), \end{aligned}$$

which leads to the desired result. □

References

[ADHW09] A. Abeyesinghe, I. Devetak, P. Hayden, and A. Winter. The mother of all protocols: Restructering quantum information's family tree. *Proceedings of Royal Society A*, 465:2537, 2009. arXiv:quant-ph/0606225v1.

[BĆH+12] F.G.S.L. Brandão, P. Ćwikliński, M. Horodecki, P. Horodecki, J.K. Korbicz, and M. Mozrzymas. Convergence to equilibrium under a random hamiltonian. *Phys. Rev. E*, 86(3):031101, 2012.

[BHH12] F.G.S.L Brandao, A.W. Harrow, and M. Horodecki. Local random quantum circuits are approximate polynomial-designs. 2012. arXiv:1208.0692.

[BV10] W.G. Brown and L. Viola. Convergence rates for arbitrary statistical moments of random quantum circuits. *Phys. Rev. Lett.*, 104:250501, 2010.

[DBWR10] F. Dupuis, M. Berta, J. Wullschleger, and R. Renner. The decoupling theorem. 2010. arXiv:1012.6044v1.

[DCEL09] C. Dankert, R. Cleve, J. Emerson, and E. Livine. Exact and approximate unitary 2-designs and their application to fidelity estimation. *Phys. Rev. A*, 80(1):12304, 2009. arXiv:quant-ph/0606161.

[Dup10] F. Dupuis. *A decoupling approach to quantum information theory*. PhD thesis, Université de Montreal, 2010. arXiv:1004.1641.

[ELL05] J. Emerson, E. Livine, and S. Lloyd. Convergence conditions for random quantum circuits. *Physical Review A*, 72(6):060302, 2005.

[EWS+03] J. Emerson, Y.S. Weinstein, M. Saraceno, S. Lloyd, and D.G. Cory. Pseudo-random unitary operators for quantum information processing. *Science*, 302(5653):2098–2100, 2003.

[GHK+12] A. Gál, K.A. Hansen, M. Koucký, P. Pudlák, and E. Viola. Tight bounds on computing error-correcting codes by bounded-depth circuits with arbitrary gates. In *Proceedings of the 44th symposium on Theory of Computing*, pages 479–494. ACM, 2012.

[HHYW08] P. Hayden, M. Horodecki, J. Yard, and A. Winter. A decoupling approach to the quantum capacity. *Open Systems and Information Dynamics*, 15:7–19, 2008. arXiv:quant-ph/0702005v1.

[HL09] A. Harrow and R. Low. Random quantum circuits are approximate 2-designs. *Communications in Mathematical Physics*, 291:257–302, 2009. arXiv:0802.1919v3.

[HOW05] M. Horodecki, J. Oppenheim, and A. Winter. Partial quantum information. *Nature*, 436:673–676, 2005. arXiv:quant-ph/0505062v1.

[HOW06] M. Horodecki, J. Oppenheim, and A. Winter. Quantum state merging and negative information. *Communications in Mathematical Physics*, 269:107, 2006. arXiv:quant-ph/0512247v1.

[HP07] P. Hayden and J. Preskill. Black holes as mirrors: quantum information in random subsystems. *Journal of High Energy Physics*, 2007(09):120, 2007.

[Low10] Richard A. Low. *Pseudo-randomness and Learning in Quantum Computation*. PhD thesis, Bristol, 2010. arXiv:1006.5227.

[LSH+11] N. Lashkari, D. Stanford, M. Hastings, T. Osborne, and P. Hayden. Towards the fast scrambling conjecture. *arXiv preprint arXiv:1111.6580*, 2011.

[MT06] R. Montenegro and P. Tetali. Mathematical aspects of mixing times in markov chains. *Found. Trend Theor. Comput. Sci.*, 3(1):237, 2006.

[Nac96] B. Nachtergaele. The spectral gap for some spin chains with discrete symmetry breaking. *Comm. Math. Phys.*, 175:565, 1996.

[NC00] M.A. Nielsen and I.L. Chuang. *Quantum computation and quantum information*. Cambridge Series on Information and the Natural Sciences. Cambridge University Press, 2000.

[ODP07] R. Oliveira, OCO Dahlsten, and MB Plenio. Generic entanglement can be generated efficiently. *Phys. Rev. Lett.*, 98(13):130502, 2007.

[Oli09] R.I. Oliveira. On the convergence to equilibrium of kac's random walk on matrices. *The Annals of Applied Probablility*, 19(3), 2009.

[Pag93] D.N. Page. Average entropy of a subsystem. *Phys. Rev. Lett.*, 71(9):1291–1294, 1993.

[PQSV11] D. Poulin, A. Qarry, R. Somma, and F. Verstraete. Quantum simulation of time-dependent hamiltonians and the convenient illusion of hilbert space. *Phys. Rev. Lett.*, 106(17):170501, 2011.

[RDO08] M. Rigol, V. Dunjko, and M. Olshanii. Thermalization and its mechanism for generic isolated quantum systems. *Nature*, 452(7189):854–858, 2008.

[SDTR11] O. Szehr, F. Dupuis, M. Tomamichel, and R. Renner. Decoupling with unitary almost two-designs. arXiv:1109.4348, 2011.

[Sre94] M. Srednicki. Chaos and quantum thermalization. *Phys. Rev. E*, 50(2):888, 1994.

[SS08] Y. Sekino and L. Susskind. Fast scramblers. *Journal of High Energy Physics*, 2008(10):065, 2008.

[Žni08] M. Žnidarič. Exact convergence times for generation of random bipartite entanglement. *Phys. Rev. A*, 78(3):032324, 2008.