Decoupling with random quantum circuits

Winton Brown* Omar Fawzi†

July 3, 2013

Abstract

Decoupling has become a central concept in quantum information theory with applications including proving coding theorems, randomness extraction and the study of conditions for reaching thermal equilibrium. However, our understanding of the dynamics that lead to decoupling is limited. In fact, the only families of transformations that are known to lead to decoupling are (approximate) unitary two-designs, i.e., measures over the unitary group which behave like the Haar measure as far as the first two moments are concerned. Such families include for example random quantum circuits with $O(n^2)$ gates, where n is the number of qubits in the system under consideration. In fact, all known constructions of decoupling circuits use $\Omega(n^2)$ gates.

Here, we prove that random quantum circuits with $O(n \log^2 n)$ gates satisfy an essentially optimal decoupling theorem. In addition, these circuits can be implemented in depth $O(\log^3 n)$. This proves that decoupling can happen in a time that scales polylogarithmically in the number of particles in the system, provided all the particles are allowed to interact. Our proof does not proceed by showing that such circuits are approximate two-designs in the usual sense, but rather we directly analyze the decoupling property.

1 Introduction

Consider an observer E that holds some information about a large system A, modeled by a joint state ρ_{AE}. In many settings, one wants this information to be mapped to global properties of the system A. This allows the information not to be affected by transformations (such as noise) provided they act on a small enough subsystem B. Such a condition is described formally by saying that the systems B and E are *decoupled*, i.e., $\rho_{BE} = \rho_B \otimes \rho_E$. In other words, this describes the absence of correlations between B and E. This condition naturally arises in the context of quantum error correcting codes, where information about which state was encoded must be unavailable on any corrupted subsystem, and in the notion of topological order, where information becomes stored in a topological degree of freedom and is inaccessible to measurements on a topologically trivial region.

A decoupling statement generally has the following form: applying a typical unitary transform chosen from some specified set to the system A leads to a state $\rho_{BE} \approx \rho_B \otimes \rho_E$, provided B is small enough compared to the initial correlations between A and E. A statement of this form is essential in proving a coding theorem for many information processing tasks. But taking the point of view of decoupling for proving coding theorems is especially useful in quantum information, mainly because of the notion of purification. Decoupling appears now as the most successful technique for analyzing quantum information processing tasks. Such an approach was used to study very general quantum information processing tasks like state merging [HOW05, HOW06, Ber09] and fully quantum Slepian-Wolf [ADHW09], but also in many other settings [Dup10]. For each of these task, a specific decoupling statement was proved but recently Dupuis et al. [DBWR10] proved a very general essentially tight decoupling theorem from which the previously mentioned results can be derived.

The notion of decoupling when A is classical is also studied under the name of privacy amplification. The maps that are applied in order to obtain decoupling are known as randomness extractors, a combinatorial object that is heavily studied in the context of complexity theory and cryptography; see [Vad] for a survey on this topic. Quantum uncertainty relations can also be viewed as decoupling statements [BFW12].

Ideas from quantum information related to decoupling have also been used in the context of thermodynamics. For example, del Rio et al. [dRÅR+11] used the decoupling theorem of [DBWR10] to study the work cost of

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

an erasure in a fully quantum context. Also, general conditions under which thermal equilibrium is reached are analyzed in [HW13, Hut11, dRHRW13]. In a different area, Hayden and Preskill [HP07] proved that an m-qubit quantum state that was dropped into a 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 well. The speed at which decoupling occurs is particularly important for this question and it motivated the study of fast scramblers [SS08, LSH$^+$13].

1.1 Decoupling with random quantum circuits

In this paper, we are interested in understanding the dynamics that lead to decoupling. For example, in a system with n particles with only pairwise interactions, how long does it take for the correlations with some observer E to become global? The time required by the dynamics generated by such a Hamiltonian is roughly equivalent to the depth of a corresponding quantum circuit. Thus, in terms of computational complexity, we want to determine what is the minimum size, and particularly, depth for a family of quantum circuits that leads to a decoupled state?

We consider the simple but natural model of random quantum circuits, in which t random gates are applied to randomly chosen pairs of qubits. Random quantum circuits of polynomial size are efficient implementations that are meant to inherit many properties of completely random unitary transformations, which typically require a circuit decomposition which is exponentially large in system size. An important property of interest is that a random unitary maps product states into highly entangled states [HLW06]. As Haar random states are not physical in the sense of computational complexity, it is interesting to determine whether such generic entanglement can be achieved by efficient random quantum circuits. A lot of work has been 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, HL09, Low10, BV10, BHH12, TGR07, HSZ12] especially properties related to the second moment. Specifically, Harrow and Low [HL09] proved that random quantum circuits are approximate two-designs with $O(n^2)$ gates. Using the result of [SDTR13], it follows that such random circuits satisfy a decoupling theorem provided the number of gates is $\Omega(n^2)$. Such a circuit has at least depth $\Omega(n)$, which is much larger than the simple signaling lower bound of $\Omega(\log n)$. Another, arguably less natural random circuit model defined in [DCEL09] was shown to decouple a constant size observer E from any macroscopic size subsystem in depth $O(\log n)$. However, it requires a depth proportional to the size of E in general, and thus requires a circuit with depth that is linear in the system size in general. This can be shown using the exact solution to the convergence properties of this model given in [BV13].

1.2 Results

We prove that random quantum circuits with $t = O(n \log^2 n)$ gates achieve essentially optimal decoupling, improving on the results of [HL09] combined with [SDTR13], which proved this result for $t = O(n^2)$. Then, by applying gates that act on disjoint qubits in parallel, we show that this circuit runs in time $O(\log^3 n)$.

1.2.1 Proof technique

The first step of the proof is to relate the property of interest to the second moment operator of the random quantum circuit. For the random quantum circuits we consider, this moment operator, when evaluated in the Pauli basis, can be seen as the transition matrix of a Markov chain on the Pauli basis elements. The property of decoupling can be formulated in terms of this Markov chain. The convergence times of such Markov chains arising from the second order moments have been previously studied in [ODP07, Žni08, HL09]. However, these convergence times are not sufficient to prove the result we are aiming for and can only give useful bounds when $\Omega(n^2)$ gates are applied. Instead, we analyze the Markov chain in a finer way by bounding the probabilities of going from an initial Pauli string of weight ℓ to a Pauli string of weight k within $O(n \log^2 n)$ steps. This is proved by building on the techniques used in [HL09].

Another reason to see why the methods of [Žni08, HL09, BHH12] cannot lead to the results we obtain here is that they use the spectral gap of the moment operator. It can be shown that this spectral gap only weakly depends on the underlying interaction graph of the circuit [BV13]. But clearly, if the interaction graph is for example a one-dimensional line, Lieb-Robinson bounds give a lower bound of $\Omega(n)$ on the circuit depth at which decoupling can happen. Recalling that our aim is to prove decoupling after a polylogarithmic number of steps, the method we use cannot rely only on the spectral gap of the moment operator.

1.3 Applications

Our results show that many information processing tasks in the quantum setting can have very efficient encoding circuits with almost linear size and polylogarithmic depth in the system size. In particular, we can achieve the quantum capacity of the erasure channel to within an arbitrary error using such an encoding circuit. The measurements for optimal quantum state merging can also be implemented using such circuits. Our main technical result can also be used to show that almost-linear sized random quantum circuits define codes with distances that achieve the quantum Gilbert-Varshamov bound; see [BF13] for details. To our knowledge capacity achieving codes of such short depth are only known for the quantum polar codes [RDR12, WR12, SRDR13], which for some special channels can even be efficiently decoded. We note that though inefficient to decode, a code defined by a short depth random quantum circuit is insensitive to which qubits the information to be encoded is initially located.

From a thermodynamics viewpoint, decoupling can be seen as a strong form of thermalization. We refer the reader to recent works that used decoupling theorems in order to derive general conditions under which thermal equilibrium is achieved [HW13, Hut11, dRHRW13]. As such, we believe that our results shed light on the speed at which thermal equilibrium is reached for generic two-body dynamics. A simple lower bound for the speed at which global thermal equilibrium can be reached for a closed quantum system is given by Lieb-Robinson bounds on the speed at which a signal can travel under such dynamics. For pairwise interactions on a complete graph this is given by time $\log n$. Our results show that this lower bound is almost achieved by a family of time dependent two-body Hamiltonians, specifically those that generate the random quantum circuit model we study.

Whether or not decoupling can be accomplished at the time scale of the signaling speed is relevant to the study of fast scramblers, [SS08], which was motivated by questions pertaining to quantum information processing in a black hole [HP07]. Specifically, it was estimated that if the dynamics of the black hole can encode a message dropped into it in a time $O(\sqrt{n}\log n)$, which is just larger than the lower bound from signaling on the two-dimensional "stretched horizon" of the black hole, then a violation of the quantum no cloning principle assuming complementarity at the event horizon could occur. Our results imply that "infinite" dimensional random quantum circuits are (pretty) fast scramblers in a strong sense, i.e., scramble a message of any size in $O(\log^3 n)$ time.

1.4 Organization

Section 2 introduces some basic notation and the model of random quantum circuits we consider here. In Section 3 we state our main result on decoupling with random circuits and reduce the problem to the study of a Markov chain Q. This Markov chain Q is studied in Section 4, which is the main technical result of this paper. The fact that the circuits can be parallelized is proved in Section 3.1. The appendix contains various technical results that are used in the proofs, such as a generalization of the gambler's ruin lemma and simple estimates for binomial coefficients.

2 Preliminaries

2.1 Generalities

A quantum state for a system A is described by a density operator $\rho \in \mathcal{S}(A)$ acting on the Hilbert space A associated with the system A. A density operator on A lives in the set $\mathcal{S}(A)$ of positive semidefinite operator with unit trace. If ρ_{AE} describes the joint state on AE, the state on the system A is described by the partial trace $\rho_A \stackrel{\text{def}}{=} \operatorname{tr}_E \rho_{AE}$. A pure state is a state of rank 1 and is denoted by $\rho_A = |\rho\rangle\langle\rho|_A$ where $|\rho\rangle \in A$. A quantum operation with input system A and output system C is given by a completely positive map $\mathcal{T}$ that maps operators on A to operators on C. A map $\mathcal{T}$ is said to be completely positive if for any system B and $X \in \mathcal{S}(A \otimes B)$ we have $(\mathcal{T} \otimes \mathrm{id})(X) \geqslant 0$. The system A in this paper is always composed on n qubits, and we denote by $\Phi_{AA'} = \frac{1}{2^n}\sum_{a,a' \in \{0,1\}^n} |a\rangle\langle a'|_A \otimes |a\rangle\langle a'|_{A'}$ a maximally entangled state between A and A'. Here $\{|a\rangle\}$ is the standard basis for A.

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

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

For a string $\nu \in \{0,1,2,3\}^n$, we define $\sigma_\nu = \sigma_{\nu_1} \otimes \cdots \otimes \sigma_{\nu_n}$. Observe that $\mathrm{tr}[\sigma_\nu \sigma_{\nu'}] = 2^n$ if $\nu = \nu'$ and 0 otherwise. The support $\mathrm{supp}(\nu)$ of ν is simply the subset $\{i \in [n] : \nu_i \neq 0\}$ and the weight $|\nu| = |\mathrm{supp}(\nu)|$. We also need to introduce an entropic quantity to quantify the decoupling accuracy. In particular, for a state ρ_{AE}, define

$$\mathrm{H}_2(A|E)_\rho = -\log_2 \left[\mathrm{tr} \left[\left(\rho_E^{-1/4} \rho_{AE} \rho_E^{-1/4} \right)^2 \right] \right]. \tag{1}$$

In order to simplify the statement of the results we use the notation $\mathrm{poly}(n)$ for a number that could be chosen as any polynomial in n and the power of the polynomial can be made large by appropriately choosing the related constants. The set of permutations of $\{1, \ldots, n\}$ is denoted by $\mathfrak{S}_n$.

2.2 Random quantum circuits

In a sequential random quantum circuit $\mathrm{RQC}(t)$, t random two-qubit gates are applied to randomly chosen pairs of qubits sequentially. Here the random two-qubit gate is chosen from the Haar measure on the unitary group acting on two qubits. In fact, our results apply equally well to any gate set whose second-order moment operator is the same as the one for the Haar measure on two qubits. This means that our results would also work if the gates are Clifford unitaries on two qubits. The number of gates of the circuit is one complexity measure but we are also interested in the depth. In this setting, multiple gates can be applied in the same time step as long as they act on disjoint qubits.

We construct a parallelized version of the sequential model in a natural way. Gates are sequentially added to the current level until it is not possible, i.e., there is a gate that shares a qubit with a previously added gate in that level. In this case, a new level is created and the process continues. We then define the parallelized model $\mathrm{RQC}(t,d)$ as follows. Choose a random $\mathrm{RQC}(t)$ circuit then parallelize it using the method describe above. If the circuit has depth at most d, then we return this circuit, otherwise the circuit is discarded and we restart the procedure.

A model of random circuits of a certain size defines a measure over unitary transformations on n qubits that we call p_{circ}. 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_p as

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

In particular $\mathrm{M}_{\mathrm{haar}} = \mathop{\mathbf{E}}_{U \sim p_{\mathrm{haar}}} \left\{ UXU^\dagger \otimes UYU^\dagger \right\}$. Any distribution for which $\mathrm{M} = \mathrm{M}_{\mathrm{haar}}$ is referred to as a two-design. We denote by $\mathrm{M}_{\mathrm{circ}}$ the moment operator for the distribution obtained by applying one step of the random circuit. For the case of a random unitary distributed according to the Haar measure applied to a randomly chosen pair i, j of qubits, see e.g., [HL09, Section 3.2]. We have

$$\mathrm{M}_{\mathrm{circ}} = \frac{1}{n(n-1)} \sum_{i \neq j} \mathrm{m}_{ij},$$

where m_{ij} only acts on qubits i and j and is defined by

$$\mathrm{m}_{ij}[\sigma_\mu \otimes \sigma_{\mu'}] = \begin{cases} 0 & \text{if } \mu \neq \mu' \\ \sigma_0 \otimes \sigma_0 & \text{if } \mu = \mu' = 0 \\ \frac{1}{15} \sum_{\nu \in \{0,1,2,3\}^2, \nu \neq 0} \sigma_\nu \otimes \sigma_\nu & \text{if } \mu = \mu' \neq 0 \end{cases}$$

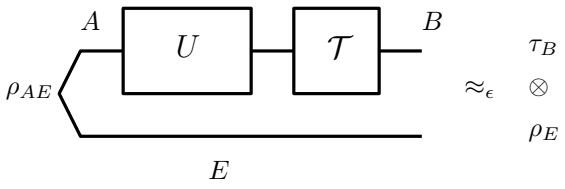
for all $\mu, \mu' \in \{0,1,2,3\}^2$. We can thus represent the operator $\mathrm{M}_{\mathrm{circ}}$ in the Pauli basis using the following $4^n \times 4^n$ matrix

$$Q(\mu, \nu) = \frac{1}{4^n} \mathrm{tr} \left[\sigma_\nu \otimes \sigma_\nu \, \mathrm{M}_{\mathrm{circ}}[\sigma_\mu \otimes \sigma_\mu] \right]. \tag{2}$$

In fact, it is simple to verify that $\sum_{\nu \in \{0,1,2,3\}^n} Q(\mu, \nu) = 1$ for all μ and so Q can be seen as a transition matrix for a Markov chain over the Pauli strings $\{0,1,2,3\}^n$ of length n.

Now for a random circuit with t independent random gates applied sequentially, the second moment operator is simply $\mathrm{M}_{\mathrm{circ}}^t$ and the corresponding matrix in the Pauli basis is also the t-th power of Q. The properties we are interested in can be expressed as quadratic functions of the entries of the unitary transformation defined by the circuit and thus can be computed from the second moment operator. This means that these properties can be completely reduced to studying the evolution of the Markov chain defined by Q.

Figure 1: A unitary U (which is going to be a random circuit in this paper) is applied to system A followed by a map $\mathcal{T}$.



3 Decoupling with random quantum circuits

We start by describing the setting for the general decoupling theorem of [DBWR10]. Consider a state ρ_{AE} on AE and a quantum channel, i.e., a completely positive trace preserving map $\mathcal{T} : \mathcal{S}(A) \to \mathcal{S}(B)$. For example, $\mathcal{T}$ might be the partial trace map keeping only the qubits in some subsystem B. See Figure 1 for an illustration. The theorem gives a sufficient condition on entropic quantities on the state ρ_{AE} and the state $\tau_{A'B} = \mathcal{T} \otimes \mathrm{id}_{A'}(\Phi_{AA'})$ where $\Phi_{AA'} = \frac{1}{2^n} \sum_{a,a'} |a\rangle\langle a'|_A \otimes |a\rangle\langle a'|_{A'}$ is a maximally entangled state on AA'. The definition of the entropy H_2 is given in (1).

Theorem 3.1 (General one-shot decoupling [DBWR10])**.** *With the notation above,*

$$\mathop{\mathbf{E}}_{U} \left\{ \|\mathcal{T}(U\rho_{AE}U^\dagger) - \tau_B \otimes \rho_E\|_1 \right\} \leq 2^{-\frac{1}{2}(\mathrm{H}_2(A|E)_\rho + \mathrm{H}_2(A|B)_\tau)}, \tag{3}$$

where U is distributed according to the Haar measure over unitaries acting on A.

In this section, we prove the main result of this paper which is a result analogous to Theorem 3.1 but where U is a unitary defined by applying a random circuit with $t = O(n \log^2 n)$ gates. Before proving the theorem, we provide a brief overview of the proof. Consider for simplicity that $\mathcal{T}$ is a partial trace map. We start by relating the trace distance of (3) to the purity $\mathrm{tr}[\mathcal{T}(U\tilde{\rho}_{AE}U^\dagger)^2]$ of the operator $\tilde{\rho}_{AE} = \rho_E^{-1/4}\rho_{AE}\rho_E^{-1/4}$. This step is standard and used in basically all decoupling theorems. Decomposing $\tilde{\rho}_{AE}$ using the Pauli basis on A, we can write

$$\tilde{\rho}_{AE} = \frac{1}{2^n} \sum_{\nu \in \{0,1,2,3\}^n} \sigma_\nu \otimes \mathrm{tr}_A[\sigma_\nu \tilde{\rho}_{AE}] \quad \text{and} \quad \mathrm{tr}[\tilde{\rho}_{AE}^2] = \frac{1}{2^n} \sum_{\nu \in \{0,1,2,3\}^n} \mathrm{tr}\left[\mathrm{tr}_A[\sigma_\nu \tilde{\rho}_{AE}]^2\right]. \tag{4}$$

Note that $\mathrm{tr}[\tilde{\rho}_{AE}^2]$ does not change when a unitary is applied on the system A. However, if we apply a unitary U and then keep a subset S of the qubits of A, the purity of the reduced state $\mathrm{tr}[\mathrm{tr}_{S^c}[U\tilde{\rho}_{AE}U^\dagger]^2] = \frac{1}{2^{|S|}} \sum_{\nu \in \{0,1,2,3\}^{|S|}} \mathrm{tr}\left[\mathrm{tr}_A[\sigma_\nu U\tilde{\rho}_{AE}U^\dagger]^2\right]$ in general depends on U. Observe for example that we only have terms $\mathrm{tr}\left[\mathrm{tr}_A[\sigma_\nu U\tilde{\rho}_{AE}U^\dagger]^2\right]$ where the weight of ν is at most $|S|$. It then becomes clear that in order to prove that $\mathrm{tr}[\mathrm{tr}_{S^c}[U\tilde{\rho}_{AE}U^\dagger]^2]$ is small when the subsystem S is sufficiently small, we should obtain bounds on $\mathrm{tr}\left[\mathrm{tr}_A[\sigma_\nu U\tilde{\rho}_{AE}U]^2\right]$ when ν is small. In particular, if U is a random quantum circuit with t gates, $\mathbf{E}\left\{\mathrm{tr}\left[\mathrm{tr}_A[\sigma_\nu U\tilde{\rho}_{AE}U]^2\right]\right\}$ can be written as a function of $Q^t(.,\nu)$ where Q is the transition matrix of the Markov chain introduced in (2) and using the decomposition of the initial state $\tilde{\rho}_{AE}$. The stationary distribution is given by the uniform distribution over all Pauli strings excluding the identity, $p_Q(\nu) = \frac{1}{4^n-1}$. The main technical result is then to prove that starting at a Pauli string, σ_μ of weight ℓ, we have that $\sum_\nu |Q^t(\mu,\nu) - p(\nu)| \leq \frac{1}{3^\ell \binom{n}{\ell}}$ where $p(\nu) \lesssim p_Q(\nu)$ provided $t > cn\log^2 n$. Note that when computing a mixing time, the worst case over all μ is considered. Note that the claimed bound on the distance improves with the weight $\ell = |\mu|$. For the result we aim to prove, obtaining this explicit dependence on ℓ is crucial.

Theorem 3.2. *Let $\rho_{AE} \in \mathcal{S}(AE)$ be an initial arbitrary mixed state and $U_t\rho_{AE}U_t^\dagger$ be the corresponding state after the application of t random two-qubit gates on the A system, which is composed of n qubits. Let $\mathcal{T} : \mathcal{S}(A) \to \mathcal{S}(B)$ be a completely positive trace preserving map. Define $\tau_{A'B} = \mathcal{T} \otimes \mathrm{id}_{A'}(\Phi_{AA'})$, where $|\Phi\rangle_{AA'} = \frac{1}{2^{n/2}} \sum_{a \in \{0,1\}^n} |a\rangle_A |a\rangle_{A'}$.*

Then we have for any $\delta > 0$, there exists a constant c such that for all n and all $t \geq cn\log^2 n$

$$\mathop{\mathbf{E}}_{U_t} \left\{ \left\| \mathcal{T}(U_t\rho_{AE}U_t^\dagger) - \tau_B \otimes \rho_E \right\|_1 \right\} \leq \sqrt{\frac{1}{\mathrm{poly}(n)} + 4^{\delta n} \cdot 2^{-\mathrm{H}_2(A|B)_\tau} \cdot 2^{-\mathrm{H}_2(A|E)_\rho}}, \tag{5}$$

where the expectation is take over the choice of random circuit of size t.

Proof As in [SDTR13], we use the following Hölder-type inequality for operators $\|\alpha\beta\gamma\|_1 \leq \||\alpha|^4\|_1^{1/4}\||\beta|^2\|_1^{1/2}\||\gamma|^4\|_1^{1/4}$, see e.g., [Bha97, Corollary IV.2.6].

$$\left\|\mathcal{T}(U_t\rho_{AE}U_t^\dagger) - \tau_B \otimes \rho_E\right\|_1^2 \leq \|(\tau_B^{1/4} \otimes \rho_E^{1/4})^4\|_1 \cdot \operatorname{tr}\left[\left(\tau_B^{-1/4} \otimes \rho_E^{-1/4}\left(\mathcal{T}(\rho_{AE}(t)) - \tau_B \otimes \rho_E\right)\tau_B^{-1/4} \otimes \rho_E^{-1/4}\right)^2\right].$$

Taking the expectation, we have

$$\begin{aligned}\mathbf{E}\left\{\left\|\mathcal{T}(U_t\rho_{AE}U_t^\dagger) - \tau_A \otimes \rho_E\right\|_1^2\right\} &\leq \mathbf{E}\left\{\operatorname{tr}[\tilde{\mathcal{T}}(U_t\tilde{\rho}_{AE}U_t^\dagger)^2]\right\} - 2\mathbf{E}\left\{\operatorname{tr}[\tilde{\mathcal{T}}(U_t\tilde{\rho}_{AE}U_t^\dagger) \cdot \tilde{\tau}_B \otimes \tilde{\rho}_E]\right\} + \operatorname{tr}[(\tilde{\tau}_B \otimes \tilde{\rho}_E)^2] \\ &\leq \mathbf{E}\left\{\operatorname{tr}\left[\tilde{\mathcal{T}}(U_t\tilde{\rho}_{AE}U_t^\dagger)^2\right]\right\} - \operatorname{tr}[\tilde{\tau}_B^2]\operatorname{tr}[\tilde{\rho}_E^2] + \frac{1}{\operatorname{poly}(n)}, \qquad (6)\end{aligned}$$

where we defined $\tilde{\rho}_{AE} = \rho_E^{-1/4}\rho_{AE}\rho_E^{-1/4}$ and $\tilde{\mathcal{T}}(.) = \tau_B^{-1/4}\mathcal{T}(.)\tau_B^{-1/4}$. If the map $\mathcal{T}$ is such that $\mathcal{T}(\mathrm{id})$ is a multiple of the identity then the last line follows directly without using any properties of U_t. If this is not the case, we explicitly bound the expectation and obtain the additional $1/\operatorname{poly}(n)$ term, which captures the fact that $\{U_t\}$ form an approximate 1-design; see Appendix B for a proof of this fact. We also use the fact that $\operatorname{tr}[\tilde{\tau}_B^2] = \operatorname{tr}[\tilde{\rho}_E^2] = 1$. To avoid complicating the expressions, we drop the $1/\operatorname{poly}(n)$ term in the remainder of the proof, as it is taken into account in the final desired statement.

Note that by definition $\operatorname{tr}[\tilde{\rho}_{AE}^2] = 2^{-\mathrm{H}_2(A|E)_\rho}$. Moreover, since $\Phi_{AA'} = \frac{1}{4^n}\sum_\nu \sigma_\nu \otimes \sigma_\nu$, we have $2^{-\mathrm{H}_2(A|B)_\tau} = \frac{1}{8^n}\sum_\nu \operatorname{tr}[\tilde{\mathcal{T}}(\sigma_\nu)^2]$. To compute $\operatorname{tr}[\tilde{\mathcal{T}}(U_t\tilde{\rho}_{AE}U_t^\dagger)^2]$, we decompose $U_t\tilde{\rho}_{AE}U_t^\dagger$ in the Pauli basis on A as follows:

$$U_t\tilde{\rho}_{AE}U_t^\dagger = \frac{1}{2^n}\sum_{\nu\in\{0,1,2,3\}^n} \sigma_\nu \otimes \operatorname{tr}_A[\sigma_\nu U_t\tilde{\rho}_{AE}U_t^\dagger]. \qquad (7)$$

Applying $\tilde{\mathcal{T}}$, we get

$$\begin{aligned}\tilde{\mathcal{T}}(U_t\tilde{\rho}_{AE}U_t^\dagger) &= \frac{1}{2^n}\sum_{\nu\in\{0,1,2,3\}^n} \tilde{\mathcal{T}}(\sigma_\nu) \otimes \operatorname{tr}_A[\sigma_\nu U_t\tilde{\rho}_{AE}U_t^\dagger] \\ &= \frac{1}{4^n}\sum_{\nu,\xi\in\{0,1,2,3\}^n} \operatorname{tr}[\sigma_\xi\tilde{\mathcal{T}}(\sigma_\nu)]\sigma_\xi \otimes \operatorname{tr}_A[\sigma_\nu U_t\tilde{\rho}_{AE}U_t^\dagger].\end{aligned}$$

As a result, we have

$$\begin{aligned}\operatorname{tr}[\tilde{\mathcal{T}}(U_t\tilde{\rho}_{AE}U_t^\dagger)^2] &= \frac{1}{2^n}\sum_{\xi\in\{0,1,2,3\}^n} \operatorname{tr}\left[\left(\frac{1}{2^n}\sum_\nu \operatorname{tr}[\sigma_\xi\tilde{\mathcal{T}}(\sigma_\nu)]\operatorname{tr}_A[\sigma_\nu U_t\tilde{\rho}_{AE}U_t^\dagger]\right)^2\right] \\ &= \frac{1}{2^n}\sum_{\xi\in\{0,1,2,3\}^n} \frac{1}{4^n}\operatorname{tr}[\sigma_\xi\tilde{\mathcal{T}}(\mathrm{id}_A)]^2\operatorname{tr}[\tilde{\rho}_E^2] \\ &+ \frac{1}{8^n}\sum_{\xi,\nu,\nu'\in\{0,1,2,3\}^n,\nu \text{ or } \nu'\neq 0} \operatorname{tr}[\sigma_\xi\tilde{\mathcal{T}}(\sigma_\nu)]\operatorname{tr}[\sigma_\xi\tilde{\mathcal{T}}(\sigma_{\nu'})] \cdot \operatorname{tr}\left[\operatorname{tr}_A[\sigma_\nu U_t\tilde{\rho}_{AE}U_t^\dagger]\operatorname{tr}_A[\sigma_{\nu'}U_t\tilde{\rho}_{AE}U_t^\dagger]\right] \\ &= \operatorname{tr}[\tilde{\tau}_B^2]\operatorname{tr}[\tilde{\rho}_E^2] + \frac{1}{8^n}\sum_{\nu,\nu'\in\{0,1,2,3\}^n,\nu \text{ or } \nu'\neq 0} T_{\nu,\nu'} \cdot \operatorname{tr}\left[\operatorname{tr}_A[\sigma_\nu U_t\tilde{\rho}_{AE}U_t^\dagger]\operatorname{tr}_A[\sigma_{\nu'}U_t\tilde{\rho}_{AE}U_t^\dagger]\right],\end{aligned}$$

where we defined $T_{\nu,\nu'} = \sum_\xi \operatorname{tr}[\sigma_\xi\tilde{\mathcal{T}}(\sigma_\nu)]\operatorname{tr}[\sigma_\xi\tilde{\mathcal{T}}(\sigma_{\nu'})]$. Getting back to equation (6) and using the concavity of the square root function, we have

$$\mathbf{E}\left\{\left\|\mathcal{T}(U_t\rho_{AE}U_t^\dagger) - \tau_B \otimes \rho_E\right\|_1\right\} \leq \sqrt{\mathbf{E}\left\{\frac{1}{8^n}\sum_{\nu,\nu'\in\{0,1,2,3\}^n,\nu \text{ or } \nu'\neq 0} T_{\nu,\nu'} \cdot \operatorname{tr}\left[\operatorname{tr}_A[\sigma_\nu U_t\tilde{\rho}_{AE}U_t^\dagger]\right]\operatorname{tr}\left[\operatorname{tr}_A[\sigma_{\nu'}U_t\tilde{\rho}_{AE}U_t^\dagger]\right]\right\}}. \qquad (8)$$

Observe that this term is a quadratic function of U_t and thus only depends on the second moment operator M of our distribution over unitary transformations on A. Recall that the second moment operator is a super operator acting on operators acting on two copies of A. For a random quantum circuit with t gates, the second moment operator is $\mathrm{M}_{\text{circ}}^t$. We have for any ν, ν',

$$\begin{aligned}\mathbf{E}\left\{\operatorname{tr}\left[\operatorname{tr}_A[\sigma_\nu U_t\tilde{\rho}_{AE}U_t^\dagger]\operatorname{tr}_A[\sigma_{\nu'}U_t\tilde{\rho}_{AE}U_t^\dagger]\right]\right\} &= \mathbf{E}\left\{\operatorname{tr}\left[\operatorname{tr}_A[\sigma_\nu U_t\tilde{\rho}_{AE}U_t^\dagger]\otimes\operatorname{tr}_{A'}[\sigma_{\nu'}U_t\tilde{\rho}_{A'E'}U_t^\dagger]F_{EE'}\right]\right\}\\ &= \operatorname{tr}\left[\operatorname{tr}_{AA'}[\sigma_\nu\otimes\sigma_{\nu'}(\mathrm{M}_{\text{circ}}^t\otimes\mathrm{id}_{EE'})[\tilde{\rho}_{AE}\otimes\tilde{\rho}_{A'E'}]F_{EE'}\right], \qquad (9)\end{aligned}$$

where we used in the first equality the fact that $\operatorname{tr}[\omega_E\omega'_E] = \operatorname{tr}[\omega_E\otimes\omega'_{E'}F_{EE'}]$ with $F_{EE'}$ being the swap operator. By expanding the initial state $\tilde{\rho}_{AE}$ in the Pauli basis, we obtain

$$\begin{aligned}(\mathrm{M}_{\text{circ}}^t\otimes\mathrm{id}_{EE'})[\tilde{\rho}_{AE}\otimes\tilde{\rho}_{A'E'}] &= \frac{1}{4^n}\sum_{\mu,\mu'\in\{0,1,2,3\}^n}(\mathrm{M}_{\text{circ}}^t\otimes\mathrm{id}_{EE'})\left[\sigma_\mu\otimes\operatorname{tr}_A[\sigma_\mu\tilde{\rho}_{AE}]\otimes\sigma_{\mu'}\otimes\operatorname{tr}_{A'}[\sigma_{\mu'}\tilde{\rho}_{A'E'}]\right]\\ &= \frac{1}{4^n}\sum_{\mu,\mu'\in\{0,1,2,3\}^n}\mathrm{M}_{\text{circ}}^t[\sigma_\mu\otimes\sigma_{\mu'}]\otimes\operatorname{tr}_A[\sigma_\mu\tilde{\rho}_{AE}]\otimes\operatorname{tr}_{A'}[\sigma_{\mu'}\tilde{\rho}_{A'E'}].\end{aligned}$$

Continuing, we get

$$\mathbf{E}\left\{\operatorname{tr}\left[\operatorname{tr}_A[\sigma_\nu\tilde{\rho}_{AE}(t)]\operatorname{tr}_A[\sigma_{\nu'}\tilde{\rho}_{AE}(t)]\right]\right\} = \frac{1}{4^n}\sum_{\mu,\mu'\in\{0,1,2,3\}^n}\operatorname{tr}\left[\sigma_\nu\otimes\sigma_{\nu'}\mathrm{M}_{\text{circ}}^t[\sigma_\mu\otimes\sigma_{\mu'}]\right]\otimes\operatorname{tr}\left[\operatorname{tr}_A[\sigma_\mu\tilde{\rho}_{AE}]\operatorname{tr}_A[\sigma_{\mu'}\tilde{\rho}_{AE}]\right].$$

Recall that $\frac{1}{4^n}\operatorname{tr}\left[\sigma_\nu\otimes\sigma_{\nu'}\mathrm{M}_{\text{circ}}^t[\sigma_\mu\otimes\sigma_{\mu'}]\right] = Q^t(\mu,\nu)$ if $\mu' = \mu$ and $\nu = \nu'$ and 0 otherwise. The expectation in equation (8) then becomes

$$\begin{aligned}&\frac{1}{8^n}\sum_{\nu\in\{0,1,2,3\}^n,\nu\neq 0}T_{\nu,\nu}\sum_{\mu\in\{0,1,2,3\}^n}Q^t(\mu,\nu)\operatorname{tr}[\operatorname{tr}_A[\sigma_\mu\tilde{\rho}_{AE}]^2]\\ &= \frac{1}{4^n}\sum_{\nu\in\{0,1,2,3\}^n,\nu\neq 0}\operatorname{tr}[\tilde{\mathcal{T}}(\sigma_\nu)^2]\sum_{\mu\in\{0,1,2,3\}^n,\mu\neq 0}Q^t(\mu,\nu)\operatorname{tr}[\operatorname{tr}_A[\sigma_\mu\tilde{\rho}_{AE}]^2]\\ &= \frac{1}{4^n}\sum_{\mu\in\{0,1,2,3\}^n,\mu\neq 0}\operatorname{tr}[\operatorname{tr}_A[\sigma_\mu\tilde{\rho}_{AE}]^2]\sum_{\nu\in\{0,1,2,3\}^n,\nu\neq 0}\operatorname{tr}[\tilde{\mathcal{T}}(\sigma_\nu)^2]Q^t(\mu,\nu). \qquad (10)\end{aligned}$$

The main technical result in this proof is in Theorem 4.1 (which we defer to Section 4), where we obtain a bound of

$$\sum_{\nu\in\{0,1,2,3\}^n,\nu\neq 0}\left|Q^t(\mu,\nu) - p_\delta(\nu)\right| \leq \frac{1}{(3-\eta)^\ell\binom{n}{\ell}\operatorname{poly}(n)}, \qquad (11)$$

where $p_\delta(\nu) \leq \frac{4^{\delta n}}{4^n-1}$ and $|\mu| = \ell$ and for any positive constants δ and η and $t \geq cn\log^2 n$ for some constant c depending on δ and η and the desired polynomial. We have by plugging equation (11) into (10), we obtain

$$\begin{aligned}&\mathbf{E}\left\{\frac{1}{8^n}\sum_{\nu,\nu'\in\{0,1,2,3\}^n,\,\nu\text{ or }\nu'\neq 0}T_{\nu,\nu'}\cdot\operatorname{tr}\left[\operatorname{tr}_A[\sigma_\nu U_t\tilde{\rho}_{AE}U_t^\dagger]\right]\operatorname{tr}\left[\operatorname{tr}_A[\sigma_{\nu'}U_t\tilde{\rho}_{AE}U_t^\dagger]\right]\right\}\\ &= \frac{1}{4^n}\sum_{\ell=1}^n\sum_{\mu:|\mu|=\ell}\operatorname{tr}[\operatorname{tr}_A[\sigma_\mu\tilde{\rho}_{AE}]^2]\sum_{\nu\in\{0,1,2,3\}^n,\nu\neq 0}\operatorname{tr}[\tilde{\mathcal{T}}(\sigma_\nu)^2]\left(p_\delta(\nu)+Q^t(\mu,\nu)-p_\delta(\nu)\right)\\ &\leq \frac{1}{4^n}\sum_{\mu\neq 0}\operatorname{tr}[\operatorname{tr}_A[\sigma_\mu\tilde{\rho}_{AE}]^2]\sum_{\nu\neq 0}\operatorname{tr}[\tilde{\mathcal{T}}(\sigma_\nu)^2]\frac{4^{\delta n}}{4^n-1}\\ &+ \frac{1}{4^n}\sum_{\ell=1}^n\sum_{\mu:|\mu|=\ell}\operatorname{tr}[\operatorname{tr}_A[\sigma_\mu\tilde{\rho}_{AE}]^2]\frac{1}{(3-\eta)^\ell\binom{n}{\ell}\operatorname{poly}(n)}\max_\nu\operatorname{tr}[\tilde{\mathcal{T}}(\sigma_\nu)^2]. \qquad (12)\end{aligned}$$

Let us start by considering the first term. Recall that $\sum_\mu\operatorname{tr}[\operatorname{tr}_A[\sigma_\mu\tilde{\rho}_{AE}]^2] = 2^n\operatorname{tr}[\tilde{\rho}_{AE}^2]$ and $\frac{1}{8^n}\sum_\nu\operatorname{tr}[\tilde{\mathcal{T}}(\sigma_\nu)^2] =$

$2^{-\mathrm{H}_2(A|B)_\tau}$. As a result,

$$\begin{aligned}\frac{1}{4^n}\sum_{\mu\neq 0}\mathrm{tr}[\mathrm{tr}_A[\sigma_\mu\tilde{\rho}_{AE}]^2]\sum_{\nu\neq 0}\mathrm{tr}[\tilde{\mathcal{T}}(\sigma_\nu)^2]\frac{4^{\delta n}}{4^n-1} &= 4^{\delta n}\frac{1}{4^n}\sum_{\nu\neq 0}\mathrm{tr}[\tilde{\mathcal{T}}(\sigma_\nu)^2]\frac{2^n\,\mathrm{tr}[\tilde{\rho}_{AE}^2]-\mathrm{tr}[\tilde{\rho}_E^2]}{4^n-1}\\ &\leq 4^{\delta n}\frac{1}{8^n}\sum_{\nu}\mathrm{tr}[\tilde{\mathcal{T}}(\sigma_\nu)^2]\frac{2^n\,\mathrm{tr}[\tilde{\rho}_{AE}^2]-\mathrm{tr}[\tilde{\rho}_E^2]}{2^n-1}\\ &\leq 4^{\delta n}2^{-\mathrm{H}_2(A|B)_\tau}2^{-\mathrm{H}_2(A|E)_\rho}.\end{aligned}$$

To prove that the second term can be bounded by an inverse polynomial, we use Lemma C.1 which is proven in the appendix. It states that

$$\sum_{\nu:|\nu|=\ell}\mathrm{tr}\left[\mathrm{tr}_A[\sigma_\nu\tilde{\rho}_{AE}]^2\right]\leq 12n^4\cdot(3-\eta)^\ell\binom{n}{\ell} \tag{13}$$

provided $\mathrm{tr}[\tilde{\rho}_{AE}^2]\leq 2^{(1-\delta)n}$. Also, we have for any $\nu\in\{0,1,2,3\}^n$,

$$\begin{aligned}\mathrm{tr}[\tilde{\mathcal{T}}(\sigma_\nu)^2] &= \mathrm{tr}\left[\mathcal{T}(\mathrm{id}/2^n)^{-1/2}\mathcal{T}(\sigma_\nu)\mathcal{T}(\mathrm{id}/2^n)^{-1/2}\mathcal{T}(\sigma_\nu)\right]\\ &\leq \mathrm{tr}[\mathrm{id}\sqrt{2^n}\sigma_\nu\mathrm{id}\sqrt{2^n}\sigma_\nu]\\ &= 4^n,\end{aligned}$$

using the monotonicity of the relative entropy of order 2; see e.g., [DFW13]. Plugging the value of η from (13) into the second term of (12), we obtain

$$\begin{aligned}\frac{1}{4^n}\sum_{\ell=1}^{n}\sum_{\mu:|\mu|=\ell}\mathrm{tr}[\mathrm{tr}_A[\sigma_\mu\tilde{\rho}_{AE}]^2]\frac{1}{(3-\eta)^\ell\binom{n}{\ell}\,\mathrm{poly}(n)}\max_\nu\mathrm{tr}[\tilde{\mathcal{T}}(\sigma_\nu)^2] &\leq \frac{1}{4^n}\max_\nu\mathrm{tr}[\tilde{\mathcal{T}}(\sigma_\nu)^2]\cdot\frac{12n^5}{\mathrm{poly}(n)}\\ &\leq \frac{1}{\mathrm{poly}(n)},\end{aligned}$$

by choosing a large enough c. Note that in the case where $\mathrm{tr}[\tilde{\rho}_{AE}^2]>2^{(1-\delta)n}$, the theorem clearly holds because the upper bound is greater than 2. □

An important example for the map $\mathcal{T}$ is the partial trace map.

Corollary 3.3. *Let ρ_{AE} be an initial arbitrary mixed state on n qubits and $U_t\rho_{AE}U_t^\dagger$ be the corresponding state after the application of t random two-qubit gates on the A system. Then let S be a random subset of the qubits $\{1,\ldots,n\}$ of size s.*

Then we have for any constant $\delta>0$, there exists a constant c such that $t\geq cn\log^2 n$, we have for subset S of size s:

$$\mathop{\mathbf{E}}_{U_t}\left\{\left\|\mathrm{tr}_{A_S{}^c}\left[U_t\rho_{AE}U_t^\dagger\right]-\frac{\mathrm{id}_{A_S}}{2^s}\otimes\rho_E\right\|_1\right\}\leq\sqrt{\frac{1}{\mathrm{poly}(n)}+4^{\delta n}\cdot 2^{2s-n}\cdot 2^{-\mathrm{H}_2(A|E)_\rho}}. \tag{14}$$

Proof It suffices to compute the entropic quantity for $\mathcal{T}$. If $\mathcal{T}$ is the erasure map for all but s qubits, we have $2^{-\mathrm{H}_2(A|B)_\tau}=2^{2s-n}$. □

3.1 Depth

We proved in the last section that decoupling can be accomplished using $O(n\log^2 n)$ gates. In this section, we study another complexity measure which is closely related to time: the depth. Gates acting on disjoint qubits are allowed to be executed in parallel. The depth of a circuit with t gates is at most t but it could be much smaller than t. In particular, for a random quantum circuit we expect many gates to act on disjoint qubits so that they can be implemented in a number of time steps that can be much smaller than t. As mentioned in the preliminaries, 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 this case, a new level is created and the process continues. 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 $O(\frac{t}{n}\log n)$. For the purpose of parallelization, the gates can simply be labelled by the two qubits the gate acts upon.

Proposition 3.4. *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 $\frac{1}{\mathrm{poly}(n)}$, the resulting circuit has depth at most $O\left(\frac{t}{n}\log n\right)$. In other words, in the model* RQC($cn\log^2 n$, $c'\log^3 n$)*, discarding a circuit only happens with probability $\frac{1}{\mathrm{poly}(n)}$ provided the constants c and c' are appropriately chosen.*

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

Lemma 3.5. *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 3.4] 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 generates 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 3.2 and Proposition 3.4.

Corollary 3.6. *In the setting of Theorem 3.2 and if U_t is the unitary computed by a random quantum circuit chosen according to the model* RQC($cn\log^2 n$,$c'\log^3 n$)*, then*

$$\mathbf{E}\left\{\left\|\mathcal{T}(U_t\rho_{AE}U_t^\dagger)-\tau_B\otimes\rho_E\right\|_1\right\}\leq\sqrt{\frac{1}{\mathrm{poly}(n)}+4^{\delta n}\cdot 2^{-\mathrm{H}_2(A|B)_\tau}\cdot 2^{-\mathrm{H}_2(A|E)_\rho}}. \tag{15}$$

Proof We write $\mathrm{depth}(U_t)$ for the depth of the circuit obtained by parallelizing the circuit defining U_t. Let $t=cn\log^2 n$ and $d=c'\log^3 n$. We have

$$\begin{aligned}\mathop{\mathbf{E}}_{\mathrm{RQC}(t,d)}\left\{\left\|\mathcal{T}(U_t\rho_{AE}U_t^\dagger)-\tau_B\otimes\rho_E\right\|_1\right\} &= \mathop{\mathbf{E}}_{\mathrm{RQC}(t)}\left\{\mathbb{1}_{\mathrm{depth}(U_t)\leq d}\left\|\mathcal{T}(U_t\rho_{AE}U_t^\dagger)-\tau_B\otimes\rho_E\right\|_1\right\}\\ &\leq \mathop{\mathbf{E}}_{\mathrm{RQC}(t)}\left\{\left\|\mathcal{T}(U_t\rho_{AE}U_t^\dagger)-\tau_B\otimes\rho_E\right\|_1\right\}+\mathbf{E}\left\{(1-\mathbb{1}_{\mathrm{depth}(U_t)\leq c'\log^3 n})\right\}\\ &\leq \mathop{\mathbf{E}}_{\mathrm{RQC}(t)}\left\{\left\|\mathcal{T}(U_t\rho_{AE}U_t^\dagger)-\tau_B\otimes\rho_E\right\|_1\right\}+\frac{1}{\mathrm{poly}(n)}.\end{aligned}$$

□

4 Analysis of the random walk over Pauli operators

This section is devoted to the analysis of the Markov chain Q over strings $\{0,1,2,3\}^n$ introduced in (2). The property we study is similar to the mixing time but differing in two ways. First, instead of considering the distance between the distribution $Q^t(\mu,.)$ obtained after t steps of the Markov chain and the stationary distribution p_Q, we can replace p_Q by any distribution that has the property $p \leq 2^{\delta n} p_Q$. In other words, we can compute the distance to any distribution p that has a small max-entropy relative to p_Q, i.e., $\mathrm{D}_{\max}(p, p_Q) \leq \delta n$. Second, the bound we obtain on the distance depends on the initial state μ.

Theorem 4.1. *Let Q be the Markov chain over Pauli strings defined in* (2)*. For any constants $\delta \in (0, 1/16), \eta \in (0,1)$, there exists a constant c such that for $t \geq cn \log^2 n$, and all Pauli strings σ_μ of weight ℓ, and large enough n, there exists a possible subnormalized distribution p_δ on strings $\{0,1,2,3\}^n$ such that for all ν,*

$$p_\delta(\nu) \leq \frac{16^{\delta n}}{4^n - 1}$$

and

$$\sum_{\nu \in \{0,1,2,3\}^n, \nu \neq 0} \left| Q^t(\mu,\nu) - p_\delta(\nu) \right| \leq \frac{1}{(3-\eta)^\ell \binom{n}{\ell}} \frac{1}{\mathrm{poly}(n)}.$$

We first prove a similar result for a Markov chain which acts only on the weights of the Pauli strings. More precisely, we define $P(\ell, k) = \sum_{\nu : |\nu| = k} Q(\mu, \nu)$ where μ is an arbitrary string with weight ℓ. Note that this definition is independent of the choice of μ. This follows from the fact that $Q(\pi(\mu), \pi(\nu)) = Q(\mu, \nu)$ for any permutation $\pi \in \mathfrak{S}_n$ of the n qubits, and also $Q(\gamma(\mu), \gamma(\nu)) = Q(\mu, \nu)$, where $\gamma \in \mathfrak{S}_3^{\times n}$ is a relabeling of the operators $\{1,2,3\}$. We have

$$P(\ell, k) = \begin{cases} 1 - \frac{2\ell(3n-2\ell-1)}{5n(n-1)} & \text{if } k = \ell \\ \frac{2\ell(\ell-1)}{5n(n-1)} & \text{if } k = \ell - 1 \\ \frac{6\ell(n-\ell)}{5n(n-1)} & \text{if } k = \ell + 1 \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

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 k? 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 need to improve the analysis of [HL09] and compute the desired probability directly.

Theorem 4.2. *Let P be the Markov chain transition matrix defined in* (16)*. For any constants $\delta \in (0, 1/16), \eta \in (0,1)$, there exists a constant c such that for $t \geq cn \log^2 n$ and all integers $1 \leq \ell \leq n$ and $1 \leq k \leq n$, we have for large enough n*

$$P^t(\ell, k) \leq 4^{\delta n} \cdot \frac{\binom{n}{k} 3^k}{4^n - 1} + \frac{1}{(3-\eta)^\ell \binom{n}{\ell}} \frac{1}{\mathrm{poly}(n)}.$$

Proof It is convenient for the proof to define variables $X_0, X_1, \ldots, X_t, \ldots$ for the Markov chain with transition probabilities P. We write $X_t(\ell)$ for a chain with $X_0 = \ell$. With this notation, $P^t(\ell, k) = \mathbf{P}\{X_t(\ell) = k\}$. The stationary distribution of P 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 $t \geq 1$,

$$\frac{1}{4^n - 1} \sum_{\ell=1}^{n} 3^\ell \binom{n}{\ell} \mathbf{P}\{X_t(\ell) = k\} = \frac{3^k \binom{n}{k}}{4^n - 1}. \tag{17}$$

The general strategy of the proof is as follows. First we choose two reference points r_- and r_+ with $r_- \leq 3n/4 \leq r_+$. The states r_- and r_+ are chosen for two properties: they should have a significant probability in the stationary distribution of P and moreover they should be bounded away from $3n/4$ so that the probability of reaching r_- starting below can be bounded and similarly for the probability of reaching r_+ starting above it. This divides the state space of the chain into three parts: $[1, r_-)$, $[r_-, r_+]$ and $(r_+, n]$. When $\ell \in [r_-, r_+]$, it is simple to prove the desired result. Whenever the starting point of the chain $\ell \in [1, r_-)$ or $\ell \in (r_+, n]$, we prove that the interval $[r_-, r_+]$ is reached with high probability (that depends on ℓ) if the chain is run for sufficiently long. We then conclude by using the first case. We note that most of the difficulty is in handling the case $\ell \in [1, r_-)$.

We start by picking specifically r_- and r_+. We choose $r_- = (3/4 - \delta)n$ and $r_+ = (3/4 + \delta)n$. They satisfy the following properties. The first one is

$$\binom{n}{r_-} 3^{r_-} \geq 4^{(1-\delta)n} \qquad \text{and} \qquad \binom{n}{r_+} 3^{r_+} \geq 4^{(1-\delta)n}, \tag{18}$$

for sufficiently large n. To see the second inequality, write

$$\binom{n}{r_+} = \frac{n(n-1)\cdots(n/4+1)\cdot n/4\cdots((1/4-\delta)n+1)}{(3/4n)!\cdot(3/4n+1)\cdots(3/4+\delta)n} \geq \binom{n}{3n/4}\left(\frac{1-4\delta}{3}\right)^{\delta n}.$$

The second property is that for all $x < r_-$ and $y > r_+$,

$$\frac{P(x,x+1)}{P(x,x-1)} = 3\cdot\frac{n-x}{x-1} \geq 1+2\delta \qquad \text{and} \qquad \frac{P(y,y-1)}{P(y,y+1)} \geq 1+2\delta. \tag{19}$$

We now start with the case $\ell \in [r_-, r_+]$. For this, we simply use (17). For any $t \geq 1$ and any $r \in [r_-, r_+]$,

$$\begin{aligned}
\mathbf{P}\left\{X_t(r) = k\right\} &= \frac{4^n-1}{\binom{n}{r}3^r}\cdot\frac{\binom{n}{r}3^r}{4^n-1}\mathbf{P}\left\{X_t(r)=k\right\} \\
&\leq \frac{4^n-1}{\binom{n}{r}3^r}\cdot\frac{1}{4^n-1}\sum_{\ell=1}^{n} 3^\ell\binom{n}{\ell}\mathbf{P}\left\{X_t(\ell)=k\right\} \\
&\leq \frac{4^n-1}{\binom{n}{r}3^r}\cdot\frac{\binom{n}{k}3^k}{4^n-1} \\
&\leq 4^{\delta n}\cdot\frac{\binom{n}{k}3^k}{4^n-1}.
\end{aligned} \tag{20}$$

In the last line, we used the inequalities in (18). This proves the case $\ell \in [r_-, r_+]$, and in fact for any $t \geq 1$.

We now handle the case $\ell \in [1, r_-)$. Introduce $T_{r_-}(\ell) = \min\{t \geq 1 : X_t(\ell) \geq r_-\}$. Note that we have for any t

$$\begin{aligned}
\mathbf{P}\left\{X_t(\ell)=k\right\} &\leq \mathbf{P}\left\{T_{r_-}(\ell) < t, X_t(\ell)=k\right\} + \mathbf{P}\left\{T_{r_-}(\ell)\geq t\right\} \\
&= \mathbf{P}\left\{T_{r_-}(\ell) < t, X_{t-T}(r)=k\right\} + \mathbf{P}\left\{T_{r_-}(\ell)\geq t\right\} \\
&\leq \max_{1\leq s\leq t}\mathbf{P}\left\{X_s(r)=k\right\} + \mathbf{P}\left\{T_{r_-}(\ell)\geq t\right\}.
\end{aligned} \tag{21}$$

Using (20), we can bound the first term. The objective of the remainder of the proof is to bound the probability $\mathbf{P}\left\{T_{r_-} \geq t\right\}$ when $t = cn\log^2 n$. This is done in Lemma 4.3 below and it concludes the case $\ell \in [1, r_-)$.

The case $\ell \in (r_+, n]$ is analogous, except that we use Lemma 4.6 instead, which has a similar proof but is significantly simpler. We note that in this case, it is possible to obtain a better probability bound without the dependence on the starting point ℓ. □

Lemma 4.3. *Let $\delta \in (0, 1/16)$ and $\eta \in (0,1)$ be constants and r_- satisfying condition* (19)*. Then for a large enough constant c (depending on δ and η) and large enough n, we have for all $\ell \leq r_-$,*

$$\mathbf{P}\left\{T_{r_-}(\ell) > cn\log^2 n\right\} \leq 2^{-2n} + \frac{1}{(3-\eta)^\ell\binom{n}{\ell}}\cdot\frac{1}{\mathrm{poly}(n)}.$$

Proof As this proof does not involve r_+, we write r instead of r_- to make the notation lighter. To prove this result, we start by defining an accelerated walk $\{Y_i\}$ as in [HL09] and the corresponding stopping time $S = \min\{s : Y_s \geq r\}$. 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 steps it takes the walk to change states. 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 sometimes use instead (we are only interested in upper bounds on the waiting times).

Getting back to T_r denoted simply T in the following, notice that $T = S + W_1 + W_2 + \cdots + W_S$. So we have for all s

$$\mathbf{P}\{T > t + s\} \leq \mathbf{P}\{S > s\} + \mathbf{P}\{S \leq s, W_1 + \cdots + W_S > t\}. \tag{22}$$

We will choose s later so that both terms are small. We start by bounding the first term, which can be done using a simple application of a Chernoff-type bound.

Lemma 4.4. *If $s > \frac{n}{3\delta}$, we have*

$$\mathbf{P}\{S > s\} \leq \exp\left(-\frac{\delta^2}{18} \cdot s\right).$$

Proof For this we just use a concentration bound on the position of a random walk relative to its expectation. Recall that the probability of moving forward when $Y_i = r$ is $\frac{6r(n-r)}{6r(n-r)+2r(r+1)}$. Then, using the property (19) the probability of moving forward is at most $1/2 + \delta/3$ for Y_i provided $Y_i \leq r$. Define a random walk Y_i' with $Y_0' = 0$ and it moves to the right with probability $1/2 + \delta/3$ and to the left with probability $1/2 - \delta/3$. For $i \leq S$, we can assume that $Y_i' \leq Y_i$. In other words, we have $S' \geq S$ where $S' = \min\{i : Y_i' \geq r\}$. Thus,

$$\begin{aligned}
\mathbf{P}\{S > s\} &\leq \mathbf{P}\{S' > s\} \\
&\leq \mathbf{P}\{Y_s' < r\} \\
&= \mathbf{P}\{Y_s' - \ell < 2 \cdot \delta/3 \cdot s - (2\delta/3 \cdot s + \ell - r)\} \\
&\leq \exp\left(-\frac{(2\delta/3s + \ell - r)^2}{2s}\right)
\end{aligned}$$

where we used the fact that $\mathbf{E}\{Y_s'\} = \ell + 2\delta/3s$ 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 $r = (3/4 - \delta)n$ is reached.

Lemma 4.5. *We have*

$$\mathbf{P}\left\{S \leq s, W_1 + \cdots + W_S > cn\log^2 n\right\} \leq \frac{1}{(3(1-8\eta))^\ell \binom{n}{\ell}} \cdot \frac{1}{\mathrm{poly}(n)}$$

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 notation of [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 $\mu = 2c'\delta$.[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 show that the $\mathbf{P}\{\mathsf{H}^c\}$ is small.

Define the random variable $M = \min_{1 \leq i \leq S} Y_i$. We have

$$\begin{aligned}
\mathbf{P}\{W_1 + \cdots + W_S > t, S \leq s, \mathsf{H}\} &= \sum_{m=1}^{\ell} \mathbf{P}\{M = m, S \leq s, W_1 + \cdots + W_S > t, \mathsf{H}\} \\
&= \sum_{m=1}^{\ell} \mathbf{P}\{M = m\}\, \mathbf{P}\{S \leq s, W_1 + \cdots + W_S > t, \mathsf{H} | M = m\} \\
&\leq \sum_{m=1}^{\ell} \mathbf{P}\{M \leq m\} \max_{\{y_i\} \text{ satisfying } M=m \text{ and } \mathsf{H} \text{ and } S \leq s} \mathbf{P}\{W(y_1) + \cdots + W(y_s) \geq t\},
\end{aligned} \tag{23}$$

[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 $\mu > 2\delta/3$ because we stop after reaching state $r = (3/4 - \delta)n$, and the probability of moving forward at r is at least $1/2 + \delta/3$.

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 bound $\mathbf{P}\{M \le m\}$ using Lemma A.1. Recall that the random walk we are considering has transition probabilities that depend on the state we are in. More precisely, the probabilities of going from state ℓ to state $\ell+1$ is a decreasing function of ℓ for $\ell \le r$. This makes it difficult to obtain a useful bound on $\mathbf{P}\{M \le m\}$ and so we are going to consider simplified walks for which $\mathbf{P}\{M \le m\}$ can only be greater. Note that at state r, the probability of moving to $r+1$ is $p_+(r) \ge 1/2 + \delta/3$ (see (19)).

Define $q \stackrel{\text{def}}{=} \left\lceil \frac{\log(n/\eta)}{\log(1+\delta)} \right\rceil$. We handle the cases $\ell < q/\eta + 1$ and $\ell \ge q/\eta + 1$ separately. We start with $\ell \ge q/\eta + 1$.

We consider the following chain: the probabilities of moving forward between $\ell+1$ and $\ell+q$ are all set to $p_+(\ell+q)$, the value of this probability at state $\ell+q$. Moreover, for all states larger than $\ell+q$, we assign an equal probability of moving forward and backward. This defines a new walk to which we can apply Lemma A.1. Assume for now that $\ell + q < r$. Using the same notation as in Lemma A.1, we write $\alpha_- = \frac{p_-}{1-p_-} = 3 \cdot \frac{n-\ell}{\ell-1}$, and $\alpha_q = \alpha_+(\ell+q) = \frac{p_+(\ell+q)}{1-p_+(\ell+q)}$, we obtain

$$\begin{aligned}\mathbf{P}\{M \le \ell - 1\} &\le \frac{1}{1 + \alpha_- \frac{\alpha_q^q}{1+\alpha_q^q+\cdots+\alpha_q+1+\cdots+1}} \\ &= \frac{1}{1 + \alpha_- \frac{\alpha_q^q}{\frac{\alpha_q^{q+1}-1}{\alpha_q-1}+(r-\ell-q-1)}}.\end{aligned}$$

We focus on the term involving α_q:

$$\begin{aligned}\frac{\alpha_q^q}{\frac{\alpha_q^{q+1}-1}{\alpha_q-1} + (r-\ell-q-1)} &= \frac{\alpha_q^q(\alpha_q-1)}{\alpha_q^{q+1} - 1 + (\alpha_q-1)(r-\ell-q-1)} \\ &\ge \frac{\alpha_q - 1}{\alpha_q} \cdot \frac{1}{1 + (\alpha_q-1)\frac{r-\ell-q-1}{\alpha_q^{q+1}}}.\end{aligned}$$

We know that $\alpha_q \ge \alpha_+(r) \ge 1 + 2\delta$ using property (19) and as a result the previous expression is lower bounded by $(1 - 1/\alpha_q)(1-\eta)$. Continuing, we get

$$\begin{aligned}\mathbf{P}\{M \le \ell - 1\} &\le \frac{1}{1 + (1-\eta)\cdot\alpha_-\cdot(1-\frac{1}{\alpha_q})} \\ &= \frac{1}{1 + (1-\eta)\alpha_- - (1-\eta)\frac{\alpha_-}{\alpha_q}}.\end{aligned}$$

We now bound the quotient α_-/α_q.

$$\begin{aligned}\frac{\alpha_-}{\alpha_q} &= \frac{n-\ell}{\ell-1}\frac{\ell+q-1}{n-(\ell+q)} \\ &= \left(1 + \frac{q}{\ell-1}\right)\left(1 + \frac{q}{n-\ell-q}\right).\end{aligned}$$

We have $\frac{q}{n-\ell-q} \le \frac{q}{n/4-q} \le \eta$ for large enough n. Moreover, by the assumption that $\ell \ge q/\eta + 1$, we have

$$\begin{aligned}\mathbf{P}\{M \le \ell - 1\} &\le \frac{1}{1 + (1-\eta)\alpha_- - (1-\eta)(1-\eta)^2} \\ &\le \frac{1}{(1-8\eta)\alpha_-}.\end{aligned}$$

This means that provided $q/\eta + 1 \le \ell < r - q$, we have

$$\mathbf{P}\{M \le \ell - 1\} \le \frac{1}{(1-8\eta)} \cdot \frac{1}{3}\frac{\ell-1}{n-\ell}.$$

Observe that if we have $\ell + q \geq r$, then we simply replace in the previous calculation $\ell + q$ with r and the previous bound still holds in this case. To obtain a bound on $\mathbf{P}\{M \leq m\}$ for $m < \ell - 1$, note that 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 have for $m \geq q/\eta + 1$,

$$\begin{aligned}\mathbf{P}\{M \leq m\} &\leq \frac{1}{(1-8\eta)^{\ell-m}3^{\ell-m}} \cdot \frac{(\ell-1)(\ell-2)\cdots m}{(n-\ell)(n-\ell+1)\cdots(n-m-1)} \\ &\leq \frac{1}{((1-8\eta)3)^{\ell}} \cdot \frac{\ell!}{n(n-1)\cdots(n-\ell+1)} \cdot \frac{3^m}{\ell(n-\ell)} \cdot \frac{n(n-1)\cdot(n-m)}{(m-1)!} \\ &\leq \frac{1}{(3(1-8\eta))^{\ell}\binom{n}{\ell}} \cdot (3n)^m. \end{aligned} \tag{24}$$

Note that whenever $m \leq q/\eta + 1$, we can use the bound

$$\mathbf{P}\{M \leq m\} \leq \mathbf{P}\{M \leq q/\eta + 1\} \leq \frac{1}{(3(1-8\eta))^{\ell}\binom{n}{\ell}} \cdot (3n)^{q/\eta+1}.$$

We now look at the term $\max_{\{y_i\} \text{ satisfying } M=m \text{ and H and } S\leq s} \mathbf{P}\{W(y_1) + \cdots + W(y_s) \geq t\}$. 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. To avoid making the notation heavy, we assume that γ/μ is an integer. So we can write

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

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 $\lambda \geq 0$, we have, using the moment generating function of a geometric distribution and the independence of the random variables:

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

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

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

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

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

Getting back to equation (23), we have using (24):

$$\begin{aligned}
\mathbf{P}\left\{W_1+\ldots W_S>t, S\le s, \mathsf{H}\right\} &\le \sum_{m=1}^{\ell}\mathbf{P}\left\{M\le m\right\}2^{\gamma m/\mu}e^{\gamma m/(2\mu)\cdot\ln n}\cdot e^{-tm/(5n)}\\
&\le \frac{1}{(1-8\eta)^{\ell}3^{\ell}\binom{n}{\ell}}\cdot(3n)^{q/\eta+1}\sum_{m=1}^{q/\eta}2^{\gamma m/\mu}e^{\gamma m/(2\mu)\cdot\ln n}\cdot e^{-tm/(5n)}\\
&+\frac{1}{(1-8\eta)^{\ell}3^{\ell}\binom{n}{\ell}}\cdot\sum_{m=q/\eta+1}^{\ell}(3n)^{m}2^{\gamma m/\mu}e^{\gamma m/(2\mu)\cdot\ln n}\cdot e^{-tm/(5n)}\\
&\le \frac{1}{(1-8\eta)^{\ell}3^{\ell}\binom{n}{\ell}}\cdot(1+(3n)^{q/\eta})\cdot\sum_{m=1}^{\ell}\left(3n2^{\gamma/\mu}e^{\gamma/(2\mu)\cdot\ln n}\cdot e^{-t/(5n)}\right)^{m}
\end{aligned}$$

Recall that $q=O(\log n)$ and thus if $t>cn\log^2 n$ with sufficiently large c, this probability is bounded by $O\left(\frac{1}{((1-8\eta)3)^{\ell}\binom{n}{\ell}}\cdot\frac{1}{\mathrm{poly}(n)}\right)$.

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 $S_j=\min\{s\ge j+1: Y_s\ge r\}$. 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 (24). We can also bound Y_k by a simpler random walk Y'_k that moves forward with probability $1/2+\delta/3$, as we did in the proof of Lemma 4.4. Thus, we obtain

$$\begin{aligned}
\mathbf{P}\left\{\sum_{k=1}^{S}\mathbb{1}(Y_k\le x)>\gamma x/\mu, S\le s\right\} &\le \frac{1}{((1-8\eta)3)^{\ell}\binom{n}{\ell}}\left((3n)^{x}+(3n)^{q/\eta}\right)\cdot s\cdot\mathbf{P}\left\{\sum_{k=1}^{\infty}\mathbb{1}(Y'_k\le x)\ge\gamma x/\mu | Y'_0=0\right\}\\
&\le \frac{1}{((1-8\eta)3)^{\ell}\binom{n}{\ell}}\left((3n)^{x}+(3n)^{q/\eta}\right)\cdot s\cdot 2\exp\left(-\frac{\mu(\gamma-2)x}{2}\right),
\end{aligned}$$

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

$$\begin{aligned}
\mathbf{P}\left\{\mathsf{H}^c, S\le s\right\} &\le \frac{1}{((1-8\eta)3)^{\ell}\binom{n}{\ell}}\cdot 2s\cdot\left(\sum_{x=1}^{n}\exp\left(x\left(\log(3n)-\frac{\mu(\gamma-2)}{2}\right)\right)+3n^{q/\eta}\sum_{x=1}^{n}\exp\left(-\frac{\mu(\gamma-2)x}{2}\right)\right)\\
&\le \frac{1}{((1-8\eta)3)^{\ell}\binom{n}{\ell}}\cdot\frac{1}{\mathrm{poly}(n)},
\end{aligned}$$

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

$$\begin{aligned}
\mathbf{P}\left\{W_1+\ldots W_S>t, S\le s\right\} &\le \mathbf{P}\left\{W_1+\ldots W_S>t, S\le s, \mathsf{H}\right\}+\mathbf{P}\left\{\mathsf{H}^c, S\le s\right\}\\
&\le \frac{1}{((1-8\eta)3)^{\ell}\binom{n}{\ell}}\frac{1}{\mathrm{poly}(n)}.
\end{aligned}$$

We proved the desired bound when $\ell \geq q/\eta + 1$. It remains to deal with the case $\ell < q/\eta + 1$. We need to bound $\mathbf{P}\{M \leq \ell - 1\}$ in a different way. For this we simply consider a walk that is even simpler than the one considered to obtain the bound in (24): let the probabilities of moving forward for all states above ℓ be equal to $p_+(r)$ which we know is at least $1/2 + \delta/3$. Applying Lemma A.1, we obtain

$$\mathbf{P}\{M \leq \ell - 1\} \leq \frac{1}{3\delta} \cdot \frac{\ell - 1}{n - \ell},$$

and then using the same argument as before

$$\mathbf{P}\{M \leq m\} \leq \frac{1}{(3\delta)^\ell \binom{n}{\ell}} \cdot (3\delta n)^m.$$

Then we apply the exact same argument to obtain a bound

$$\mathbf{P}\{W_1 + \ldots W_S > t, S \leq s\} \leq \frac{1}{(3\delta)^\ell \binom{n}{\ell}} \frac{1}{\mathrm{poly}(n)}.$$

Now recall that $\ell < q/\eta + 1 = O(\log n)$ and thus for large enough c, we can make the term $1/\mathrm{poly}(n)$ be small enough to obtain the desired bound. □

To complete the proof of Lemma 4.3, we just plug the bounds obtained from Lemma 4.4 with $s > 12n/\delta$ and from Lemma 4.5 into equation (22). □

Lemma 4.6. *Let $\delta \in (0, 1/16)$ and $\eta \in (0, 1)$ be constants and r_+ satisfying condition* (19)*. Then for a large enough constant c (depending on δ and η) and large enough n, we have for any $\ell \geq r_+$*

$$\mathbf{P}\left\{T_{r_+}(\ell) > cn \log^2 n\right\} \leq 2^{-2n}$$

Proof The proof is analogous to Lemma 4.3, except that it is much easier to bound the waiting time. In fact, when $x > r_+$ we have $P(x, x) \leq 4/5$. This means that the waiting times $W_1, \ldots, W_S$ can be assumed to have a geometric distribution with parameter $4/5$ and then proving a version of Lemma 4.5 becomes a simple application of a Chernoff-type bound, and in fact one can obtain a better bound that is independent of ℓ. □

Proof [of Theorem 4.1] Theorem 4.2 tells us that for $|\mu| = \ell$ and all $k \in \{1, \ldots, n\}$,

$$\sum_{|\nu|=k} Q^t(\mu, \nu) \leq 4^{\delta n} \frac{\binom{n}{k} 3^k}{4^n - 1} + \frac{1}{(3 - \eta)^\ell \binom{n}{\ell}} \frac{1}{\mathrm{poly}(n)}. \tag{25}$$

Recall that there are exactly $\binom{n}{k}3^k$ distinct strings $\nu \in \{0, 1, 2, 3\}^n$ such that $|\nu| = k$. We want to show that all these strings ν have basically the same value of $Q^t(\mu, \nu)$. For this, we view the chain Q as a mixture of a part $\tilde{R}$ that can only mix the sites of the string without increasing its weight and a part $\tilde{Q}$ that can change the weight of the string. We then use invariance properties of these chains with respect to permuting the qubits and relabeling of nonzero elements $\{1, 2, 3\}$ to get the desired conclusion.

More precisely, Let $Z_t(\mu) \in \{0, 1, 2, 3\}^n$ denote the state of the chain defined by Q at step t when started in the state μ. From inequality (25), we can find an event E_P (in the notation of the proof of of Theorem 4.2, $\mathsf{E}_P = \left[T_{r_-} < t\right]$, see equation (21)) such that

$$\mathbf{P}\{\mathsf{E}_P^c\} \leq \frac{1}{(3 - \eta)^\ell \binom{n}{\ell} \mathrm{poly}(n)} \qquad \text{and} \qquad \mathbf{P}\{|Z_t(\mu)| = k, \mathsf{E}_P\} \leq \frac{4^{\delta n} 3^k \binom{n}{k}}{4^n - 1},$$

where E^c denotes the complement of the event E. This gives a natural candidate for the desired p_δ, namely $p_\delta(\nu) = \mathbf{P}\{Z_t(\mu) = \nu, \mathsf{E}_P\}$. The distance condition on p_δ is clearly satisfied:

$$\sum_{\nu \in \{0,1,2,3\}^n - \{0\}} \mathbf{P}\{Z_t(\mu) = \nu\} - \mathbf{P}\{Z_t(\mu) = \nu, \mathsf{E}_P\} = \mathbf{P}\{\mathsf{E}_P^c\} \leq \frac{1}{(3 - \eta)^\ell \binom{n}{\ell} \mathrm{poly}(n)}.$$

The objective of the remainder of the proof is to show that we have $p_\delta(\nu) \leq \frac{4^{2\delta n}}{4^n - 1}$.

Define Q_{ij} to be the transition matrix of the Markov chain conditioned on having the gate act on qubits i, j. More precisely, $Q_{ij}(\mu,\nu) = 1$ if $\mu_i = \mu_j = 0$ and $\nu_i = \nu_j = 0$ and $Q_{ij}(\mu,\nu) = 1/15$ if $\mu_i\mu_j \neq 00$ and $\nu_i\nu_j \neq 00$, and all other entries of Q_{ij} are zero. Thus, $Q(\mu,\nu) = \frac{1}{n(n-1)}\sum_{i\neq j} Q_{ij}(\mu,\nu)$. We now define

$$R_{ij}(\mu,\nu) = \begin{cases} 1 & \text{if } |\mu_i\mu_j| = |\nu_i\nu_j| = 0 \\ 1/3 & \text{if } |\mu_i\mu_j| = |\nu_i\nu_j| = 1, \mu_i = \nu_i = 0 \\ 1/3 & \text{if } |\mu_i\mu_j| = |\nu_i\nu_j| = 1, \mu_j = \nu_j = 0 \\ 1/9 & \text{if } |\mu_i\mu_j| = |\nu_i\nu_j| = 2. \end{cases}$$

and $\tilde{R}_{ij} = \frac{1}{2}R_{ij} + \frac{1}{2}\Pi_{ij}R_{ij}$ where Π_{ij} simply swaps the coefficients at position i and j. Also define

$$\tilde{Q}_{ij}(\mu,\nu) = \begin{cases} 1 & \text{if } |\mu_i\mu_j| = |\nu_i\nu_j| = 0 \\ 1/9 & \text{if } |\mu_i\mu_j| = 1 \text{ and } |\nu_i\nu_j| = 2 \\ 2/3 \cdot 1/6 & \text{if } |\mu_i\mu_j| = 2 \text{ and } |\nu_i\nu_j| = 1 \\ 1/3 \cdot 1/9 & \text{if } |\mu_i\mu_j| = 2 \text{ and } |\nu_i\nu_j| = 2. \end{cases}$$

It is simple to see that $Q_{ij} = \frac{2}{5}\tilde{R}_{ij} + \frac{3}{5}\tilde{Q}_{ij}$. We can then define $\tilde{R} = \frac{1}{n(n-1)}\sum_{i\neq j}\tilde{R}_{ij}$ and $\tilde{Q} = \frac{1}{n(n-1)}\sum_{i\neq j}\tilde{Q}_{ij}$ so that

$$Q = \frac{2}{5}\tilde{R} + \frac{3}{5}\tilde{Q}.$$

Note that $\tilde{R}$ does not change the weight of any strings, but only performs swaps and locally randomizes 1, 2 and 3. An important observation that will allow us to study $\tilde{R}$ and $\tilde{Q}$ independently is that $\tilde{R}\tilde{Q} = \tilde{Q}\tilde{R}$. In order to see this, observe first that $\tilde{R}_{ij}\tilde{Q}_{ij} = \tilde{Q}_{ij} = \tilde{Q}_{ij}\tilde{R}_{ij}$. Also $\tilde{R}_{ij}$ and $\tilde{Q}_{i'j'}$ clearly commute if $\{i,j\} \cap \{i',j'\} = \emptyset$. Now for $j \neq j'$, we have $R_{ij}\tilde{Q}_{ij'} = \tilde{Q}_{ij'}R_{ij}$. However, $\Pi_{ij}R_{ij}$ does not commute with $\tilde{Q}_{ij'}$. But we can still write $\Pi_{ij}R_{ij}\tilde{Q}_{ij'} = R_{ij}\Pi_{ij}\tilde{Q}_{ij'} = R_{ij}\tilde{Q}_{jj'}\Pi_{ij} = \tilde{Q}_{jj'}\Pi_{ij}R_{ij}$. As a result,

$$\begin{aligned}
\tilde{R}\tilde{Q} &= \frac{1}{n^2(n-1)^2}\sum_{i\neq j, i'\neq j'} \tilde{R}_{ij}\tilde{Q}_{i'j'} \\
&= \frac{1}{n^2(n-1)^2}\left(\sum_{i\neq j, i'\neq j', |\{i,j\}\cap\{i',j'\}|\in\{0,2\}} \tilde{Q}_{i'j'}\tilde{R}_{ij} + 4\sum_{i\neq j, i'\neq j', j\neq j'} \frac{1}{2}R_{ij}\tilde{Q}_{ij'} + \frac{1}{2}\Pi_{ij}R_{ij}\tilde{Q}_{ij'}\right) \\
&= \frac{1}{n^2(n-1)^2}\left(\sum_{i\neq j, i'\neq j', |\{i,j\}\cap\{i',j'\}|\in\{0,2\}} \tilde{Q}_{ij}\tilde{R}_{i'j'} + 4\sum_{i\neq j, i\neq j', j\neq j'} \frac{1}{2}\tilde{Q}_{ij'}R_{ij} + \frac{1}{2}\tilde{Q}_{jj'}\Pi_{ij}R_{ij}\right) \\
&= \tilde{Q}\tilde{R}.
\end{aligned}$$

The factor 4 in the second line is to take into account the four possibilities $i = i'$, $i = j'$, $j = i'$ and $j = j'$. As a result, for any $t \geq 1$, we can write Q^t as

$$Q^t = \sum_{t_1+t_2=t} \left(\frac{3}{5}\right)^{t_1}\left(\frac{2}{5}\right)^{t_2}\binom{t}{t_1}\tilde{R}^{t_2}\tilde{Q}^{t_1}. \tag{26}$$

Using equation (26), we see that $Z_t(\mu)$ can be generated as follows. Choose T_1 according to a binomial distribution with parameters t and $3/5$ and run the chain $\tilde{Q}$ on μ for T_1 steps. Let $Z^w(\mu) \in \{0,1,2,3\}^n$ denote the state obtained at this time. Then, in the second phase, run the chain $\tilde{R}$ for $t - T_1$ steps obtaining the state $Z_t(\mu)$. Note that we have $|Z^w(\mu)| = |Z_t(\mu)|$.

We start with the case $k \leq \delta_0 n$ for some δ_0 to be chosen later. Using (33) and (36), we have $\binom{n}{k} \leq 2^{nh(k/n)} \leq 2^{2\sqrt{\delta_0}n}$ and thus

$$\begin{aligned}
\mathbf{P}\left\{Z_t(\mu) = \nu, \mathsf{E}_P\right\} \leq \mathbf{P}\left\{|Z_t(\mu)| = |\nu|, \mathsf{E}_P\right\} &\leq \frac{4^{\delta n}3^k\binom{n}{k}}{4^n - 1} \\
&\leq \frac{4^{\delta n}3^{\delta_0 n}2^{2\sqrt{\delta_0}n}}{4^n - 1}.
\end{aligned}$$

By choosing δ_0 appropriately small, we obtain the desired result.

Now we assume that $\delta_0 < k < (1-\delta_0)n$. We deal with the case $k \geq (1-\delta_0)n$ at the end of the proof. Note first that we have

$$\sum_{|\nu|=k} \mathbf{P}\left\{Z_t(\mu) = \nu, \mathsf{E}_P\right\} \leq \frac{4^{\delta n} 3^k \binom{n}{k}}{4^n - 1}.$$

Our objective is to show that this total probability is basically evenly spread among all the ν's of weight k. For this, we condition on the value of $Z^w(\mu)$.

$$\mathbf{P}\left\{Z_t(\mu) = \nu, \mathsf{E}_P\right\} = \sum_{|\nu^w|=k} \mathbf{P}\left\{Z^w(\mu) = \nu^w, \mathsf{E}_P\right\} \cdot \mathbf{P}\left\{Z_t(\mu) = \nu | Z^w(\mu) = \nu^w, \mathsf{E}_P\right\}. \tag{27}$$

Note that the event E_P only depends on the set of weights visited by the chain. As a result, by the Markov property for the second phase, the random variable $Z_t(\mu)$ is independent of E_P conditioned on $Z^w(\mu)$. In other words, $\mathbf{P}\left\{Z_t(\mu) = \nu | Z^w(\mu) = \nu^w, \mathsf{E}_P\right\} = \mathbf{P}\left\{Z_t(\mu) = \nu | Z^w(\mu) = \nu^w\right\}$. In order to evaluate this term, we study Markov chain for the second phase which is governed by the matrix $\tilde{R}$.

More precisely, we study the evolution of the support of $Z_s(\mu)$ for $s \geq T_1$ relative to the support of $Z^w(\mu)$. Define $I_s = |\text{supp}(Z_s(\mu)) \cap \text{supp}(Z^w(\mu))|$ for $s \geq T_1$. Recall that we have $|Z_s(\mu)| = |Z^w(\mu)| = k$ and thus the expected size for $\text{supp}(Z_s(\mu)) \cap \text{supp}(Z^w(\mu))$ if $\text{supp}(Z_s(\mu))$ were completely random is k^2/n.

It is simple to compute the transition probabilities of the chain $\{I_s\}_s$:

$$\begin{aligned}
\mathbf{P}\left\{I_{s+1} = I_s + 1\right\} &= \frac{(k - I_s)^2}{n(n-1)} \\
\mathbf{P}\left\{I_{s+1} = I_s - 1\right\} &= \frac{I_s(n - 2k + I_s)}{n(n-1)} \\
\mathbf{P}\left\{I_{s+1} = I_s\right\} &= 1 - \mathbf{P}\left\{I_{s+1} = I_s + 1\right\} - \mathbf{P}\left\{I_{s+1} = I_s - 1\right\}.
\end{aligned}$$

We can verify (for example by writing detailed balance equations) that the stationary distribution for this chain is given by $p_I(k') = \frac{\binom{k}{k'}\binom{n-k}{k-k'}}{\binom{n}{k}}$ for $k' \in \{0, \ldots, k\}$. This allows us to bound the probability of reaching the state k' when starting in a state r', as was done for the chain $\{X_t\}$ in (20). This bound gets closer to the stationary probability $p_I(k')$ as r' gets closer to $\frac{k^2}{n}$. More precisely, if $I_s(r')$ denotes the size of the intersection of the supports at step s given that the starting state has an intersection size of r', we have

$$\mathbf{P}\left\{I_s(r') = k'\right\} \leq \frac{\binom{n}{k}}{\binom{k}{r'}\binom{n-k}{k-r'}} \cdot \frac{\binom{k}{k'}\binom{n-k}{k-k'}}{\binom{n}{k}}.$$

We introduce the "good" event that for some $s \in [T_1, t]$, the walk I_s gets close to the state k^2/n: $\mathsf{E}_I = \left[\exists s \in [T_1, t] : \frac{k^2}{n} - \delta_2 n \leq I_s \leq \frac{k^2}{n} + \delta_2 n\right]$. Note that if $|r' - k^2/n| \leq \delta_2 n$, then using (33) and

$$\begin{aligned}
\binom{k}{r'}\binom{n-k}{k-r'} &\geq \frac{1}{n^2} 2^{k \cdot h(\frac{r'}{k}) + (n-k) \cdot h(\frac{k-r'}{n-k})} \\
&\geq \frac{1}{n^2} 2^{n \cdot h(\frac{k}{n}) - n h(\frac{\delta_1}{\delta_0})} \\
&\geq \frac{2^{-n h(\frac{\delta_1}{\delta_0})}}{n^2} \binom{n}{k}.
\end{aligned}$$

For the second line, we used inequality (35) which implies that $h(\frac{r'}{k}) \geq h(\frac{k}{n}) - h(\frac{\delta_2 n}{k}) \geq h(\frac{k}{n}) - h(\frac{\delta_2}{\delta_0})$, and similarly $h(\frac{k-r'}{n-k}) \geq h(\frac{k}{n}) - h(\frac{\delta_2 n}{n-k}) \geq h(\frac{k}{n}) - h(\frac{\delta_2}{\delta_0})$. This means that we have

$$\mathbf{P}\left\{I_s(r') = k'\right\} \leq n^2 2^{n h(\frac{\delta_2}{\delta_0})} \cdot \frac{\binom{k}{k'}\binom{n-k}{k-k'}}{\binom{n}{k}} \tag{28}$$

whenever $|r' - k^2/n| \leq \delta_2 n$. Getting back to equation (27), we can write

$$\mathbf{P}\left\{Z_t(\mu) = \nu | Z^w(\mu) = \nu^w\right\} = \mathbf{P}\left\{\mathsf{E}_I^c\right\} + \mathbf{P}\left\{Z_t(\mu) = \nu, \mathsf{E}_I | Z^w(\mu) = \nu^w\right\}.$$

We start by bounding $\mathbf{P}\{\mathsf{E}_I^c\}$. For this, observe that for $\mathbf{P}\{I_{s+1} = I_s + 1\} - \mathbf{P}\{I_{s+1} = I_s - 1\} = \frac{k^2 - I_s n}{n(n-1)}$. This means that if $I_s \geq \frac{k^2}{n} + \delta_2 n$, there is a δ_2 negative drift, and similarly there is a constant positive drift if $I_s \leq \frac{k^2}{n} - \delta_2 n$. Using standard methods, one can conclude that $\mathbf{P}\{\mathsf{E}_I^c | t - T_1 \geq n \log n\} \leq 2^{-10n}$. In addition for large enough t, $\mathbf{P}\{t - T_1 \geq n \log n\} \geq 1 - 2^{-10n}$. [2] Then one can directly conclude $\mathbf{P}\{\mathsf{E}_I^c\} \leq \mathbf{P}\{t - T_1 < n \log n\} + \mathbf{P}\{\mathsf{E}_I^c | t - T_1 \geq n \log n\} \leq 2 \cdot 2^{-10n}$.

We can write

$$\sum_{|\nu'|=k : |\mathrm{supp}(\nu') \cap \mathrm{supp}(\nu^w)| = k'} \mathbf{P}\{Z_t(\mu) = \nu', \mathsf{E}_I | Z^w(\mu) = \nu^w\} \leq \max_{|r' - k^2/n| \leq \delta_2 n} \max_s \mathbf{P}\{I_s(r') = k'\} \tag{29}$$

$$\leq n^2 2^{nh(\frac{\delta_2}{\delta_0})} \cdot \frac{\binom{k}{k'}\binom{n-k}{k-k'}}{\binom{n}{k}}. \tag{30}$$

Now it remains to say that many of the terms in this sum are actually the same. For this, we use invariance properties of $\tilde{R}$.

Under all permutations $\pi \in \mathfrak{S}_n$ of $\{1, \ldots, n\}$, and all functions $\gamma \in (\mathfrak{S}_3)^{\times n}$ that permute the Pauli operators $\{1, 2, 3\}$ on each qubit, we have

$$\tilde{R}((\pi \circ \gamma)(\mu), (\pi \circ \gamma)(\nu)) = \tilde{R}(\mu, \nu). \tag{31}$$

It follows that $\tilde{R}(\mu, (\pi_0 \circ \gamma_0)(\nu)) = \tilde{R}((\pi_0 \circ \gamma_0)(\mu), (\pi_0 \circ \gamma_0)(\nu)) = \tilde{R}(\mu, \nu)$ for any $\pi_0 \in \mathfrak{S}_n$ and $\gamma_0 \in (\mathfrak{S}_3)^n$ such that $\pi_0 \circ \gamma_0(\mu) = \mu$, e.g., if π_0 and γ_0 act outside the support of μ.

As a result, we have that $\mathbf{P}\{Z_t(\mu) = \nu | Z^w(\mu) = \nu^w\} = \mathbf{P}\{Z_t(\mu) = \nu' | Z^w(\mu) = \nu^w\}$ if ν' can be obtained from ν by a permutation and relabeling of the Pauli operators that act outside the support of ν^w. If $|\mathrm{supp}(\nu) \cap \mathrm{supp}(\nu^w)| = k'$, then there are $3^{k-k'}\binom{n-k}{k-k'}$ distinct ν' that can be obtained in this way.

Invariance of the transition probabilities under maps that act on the support of ν^w is slightly more complicated. For any permutation π of the support of ν^w, and any relabeling γ_π that satisfies $\gamma_\pi(\nu) = \pi^{-1}(\nu)$, $\pi \circ \gamma_\pi$ keeps ν^w unchanged. Note that for any π there is at least one such γ_π. This means that also $\nu' = \pi \circ \gamma_\pi(\nu)$ obtained in this way satisfy $\mathbf{P}\{Z_t(\mu) = \nu' | Z^w(\mu) = \nu^w\} = \mathbf{P}\{Z_t(\mu) = \nu | Z^w(\mu) = \nu^w\}$. By combining with invariants outside the support of ν^w, we obtain a total of $3^{k-k'}\binom{n-k}{k-k'} \cdot \binom{k}{k'}$ distinct ν' for which $\mathbf{P}\{Z_t(\mu) = \nu' | Z^w(\mu) = \nu^w\} = \mathbf{P}\{Z_t(\mu) = \nu | Z^w(\mu) = \nu^w\}$.

The total number of ν' such that $|\nu'| = k$ and $|\mathrm{supp}(\nu') \cap \mathrm{supp}(\nu^w)| = k'$ is $3^k\binom{n-k}{k-k'} \cdot \binom{k}{k'}$, so our objective is to prove that there are roughly $3^{k'}$ additional relabelings that keep the transition probability invariant. In particular, we want to show that relabelings acting on the support of ν^w keep this probability unchanged. For this we argue as in Appendix B, that with high probability, most of the sites are acted upon at least once in the second phase. More precisely introduce the event E_A that between times T_1 and t, a $(1 - \delta_1)$ fraction of the sites $\{1, \ldots, n\}$ are acted upon in at least one step. First, let us see that this event happens with high probability. In fact, by applying a union bound on all the subsets of size $\delta_1 n$, we directly get that for sufficiently large n, $\mathbf{P}\{\mathsf{E}_A^c | t - T_1 \geq n \log n\} \leq 2^{-10n}$ and thus $\mathbf{P}\{\mathsf{E}_A^c\} \leq 2 \cdot 2^{-10n}$.

As argued in Appendix B, we can condition on the set of all sites that are acted upon in some step between T_1 and t. Then any string that is obtained from ν by applying a relabeling γ that acts on these sites has the same probability as ν. If this set of sites has size at least $(1 - \delta_1)n$, i.e., the event E_A holds, there are at least $k' - \delta_1 n$ such sites that are in $\mathrm{supp}(\nu) \cap \mathrm{supp}(\nu^w)$. This means that under the event E_A, there are at least $3^{k' - \delta_1 n}$ strings ν' obtained from ν by applying a relabeling on some sites of $\mathrm{supp}(\nu) \cap \mathrm{supp}(\nu^w)$. As a result, using (30),

$$\begin{aligned}
&\mathbf{P}\{Z_t(\mu) = \nu | Z^w(\mu) = \nu^w\} \\
&\leq \frac{1}{3^{k' - \delta_1 n} 3^{k-k'}\binom{n-k}{k-k'} \cdot \binom{k}{k'}} \cdot \sum_{|\nu'|=k : |\mathrm{supp}(\nu') \cap \mathrm{supp}(\nu^w)| = k'} \mathbf{P}\{Z_t(\mu) = \nu', \mathsf{E}_A | Z^w(\mu) = \nu^w\} + \mathbf{P}\{\mathsf{E}_A^c\} \\
&\leq n^2 2^{nh(\frac{\delta_2}{\delta_0})} 3^{\delta_1 n} \cdot \frac{1}{3^k \binom{n}{k}} + 2 \cdot 2^{-10n}.
\end{aligned}$$

[2] Note that having $T_1 \geq c' n$ for some large enough constant c' depending on δ would be good enough; we choose $n \log n$ simply to avoid introducing additional constants.

Going back to (27), we obtain

$$\begin{aligned}
\mathbf{P}\left\{Z_t(\mu)=\nu,\mathsf{E}_P\right\} &\leq \frac{2n^2 2^{nh(\frac{\delta_2}{\delta_0})}3^{\delta_1 n}}{3^k\binom{n}{k}}\sum_{|\nu^w|=k}\mathbf{P}\left\{Z^w=\nu^w,\mathsf{E}_P\right\}\\
&= \frac{2n^2 2^{nh(\frac{\delta_2}{\delta_0})}3^{\delta_1 n}}{3^k\binom{n}{k}}\mathbf{P}\left\{|Z_t(\mu)|=k,\mathsf{E}_P\right\}\\
&\leq \frac{2n^2 2^{nh(\frac{\delta_2}{\delta_0})}3^{\delta_1 n}\cdot 4^{\delta n}}{4^n-1}\\
&\leq \frac{16^{\delta n}}{4^n-1},
\end{aligned}$$

for large enough n and where in the last step we choose $\delta_1>0$ and $\delta_2>0$ small enough constants.

Now it only remains to handle the case $k\geq(1-\delta_0)n$. In this case, the size of the intersection $k'=|\mathrm{supp}(\nu)\cap\mathrm{supp}(\nu^w)|\geq 2k-n$. We then observe that on the event E_A, we can obtain at least $3^{k'-\delta_1 n}$ distinct ν' such that $\mathbf{P}\left\{Z_t(\mu)=\nu'|Z^w(\mu)=\nu^w\right\}=\mathbf{P}\left\{Z_t(\mu)=\nu|Z^w(\mu)=\nu^w\right\}$. As a result

$$\begin{aligned}
\mathbf{P}\left\{Z_t(\mu)=\nu|Z^w(\mu)=\nu^w\right\} &\leq \frac{1}{3^{k'-\delta_1 n}}\sum_{|\nu'|=k:|\mathrm{supp}(\nu')\cap\mathrm{supp}(\nu^w)|=k'}\mathbf{P}\left\{Z_t(\mu)=\nu'|Z^w(\mu)=\nu^w\right\}\\
&\leq \frac{1}{3^{2k-n-\delta_1 n}}\sum_{|\nu'|=k}\mathbf{P}\left\{Z_t(\mu)=\nu'|Z^w(\mu)=\nu^w\right\}\\
&\leq \frac{1}{3^{k-\delta_0 n-\delta_1 n}}\frac{4^{\delta n}3^k\binom{n}{k}}{4^n-1}\\
&\leq \frac{3^{(\delta_1+\delta_0)n}2^{h(\delta_0)n}4^{\delta n}}{4^n-1}.
\end{aligned}$$

For small enough δ_0 and δ_1, this leads to the desired result. □

5 Conclusion

We proved that decoupling can be achieved using a number of two-qubit gates that is almost linear in the system size. This implies that information processing tasks that can be achieved via decoupling can be implemented with a circuit of almost linear size and polylogarithmic depth.

Our result also show that a class of random time dependent Hamiltonians self-thermalize at a speed that is close to the signaling bound. It is an interesting question if a similar result applies to the decoupling time for broader classes of two-body Hamiltonians on the complete graph, and whether decoupling can occur at a time scale close to the signaling bound of $O(n^{1/d})$ for interactions on d-dimensional lattices.

As far as optimality is concerned, it would be interesting to improve the depth to $O(\log n)$. For that, one would probably need to study directly a parallel random circuit model, as the parallelization step involves an additional $O(\log n)$ factor.

Acknowledgements

We would like to thank Fernando Brandao, Patrick Hayden, David Poulin, Renato Renner, Lidia del Rio, Marco Tomamichel and Stephanie Wehner for helpful discussions and Aram Harrow for his comments. 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.

A A generalisation of the gambler's ruin lemma

Consider a random walk on a line indexed from -1 to a. At positions $i > 0$, the probability of moving forward is $p_+(i)$ (depending on i) and for points $i \leq 0$, the probability of moving forward 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 A.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

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

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

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

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 $i \geq 1$, $P_{i-1} - P_i = \left(\prod_{j=i}^{a-1}\frac{p_+(j)}{1-p_+(j)}\right)P_{a-1}$. 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. □

B Sequential random quantum circuits are approximate 1-designs

The objective of this section is to show that we have for $t > cn\log n$,

$$\mathop{\mathbf{E}}_{U_t}\left\{\mathrm{tr}[\tilde{\mathcal{T}}(U_t\rho_{AE}U_t^\dagger)\tilde{\tau}_B \otimes \tilde{\rho}_E]\right\} \geq \left(1 - \frac{1}{\mathrm{poly}(n)}\right)\mathrm{tr}[\tilde{\tau}_B^2]\,\mathrm{tr}[\tilde{\rho}_E^2]. \tag{32}$$

Let us generate the circuit U_t by first choosing the pair of qubits $S = \{(i_1, j_1), \ldots, (i_t, j_t)\}$ on which each of the t gates act and then choosing the two-qubit unitaries $V_1, \ldots, V_t$ that are applied in each time step. We then write $U_t = V_t(i_t, j_t) \cdots V_1(i_1, j_1)$. Let G be the event that $\{i_1, j_1, i_2, j_2, \ldots, i_t, j_t\} = [n]$. It then follows that if we fix such an S and take the expectation over the choice of $V_1, \ldots, V_t$, we have for any S that satisfies G,

$$\mathop{\mathbf{E}}_{V_1,\ldots,V_t} \left\{ U_t \sigma_\mu U_t^\dagger \right\} = 0,$$

for all $\mu \neq 0$. As a result we have

$$\mathop{\mathbf{E}}_{V_1,\ldots,V_t} \left\{ \mathrm{tr}[\tilde{\mathcal{T}}(U_t \rho_{AE} U_t^\dagger)\tilde{\tau}_B \otimes \tilde{\rho}_E] \right\} = \mathrm{tr}\left[\tilde{\mathcal{T}}\left(\frac{\mathrm{id}}{2^n} \right) \otimes \tilde{\rho}_E \tilde{\tau}_B \otimes \tilde{\rho}_E \right] = \mathrm{tr}[\tilde{\tau}_B^2] \, \mathrm{tr}[\tilde{\rho}_E^2],$$

for any fixed S that satisfies G. Now it only remains to bound the probability of the event G^c, which is the complement of G. The probability that qubit 1 is not affected by any gate is $(1 - 2/n)^t$. Then, by a union bound, we have $\mathbf{P}\{\mathsf{G}^c\} \leq n(1 - 2/n)^t \leq n e^{2t/n} \leq \frac{1}{\mathrm{poly}(n)}$.

C Bounding the total mass of coefficients at a certain weight

Lemma C.1. *Let ρ_{AE} be such that $\mathrm{H}_2(A|E)_\rho \geq -(1-\epsilon)n$ with $\epsilon > 0$, i.e.,*

$$\mathrm{tr}[\tilde{\rho}_{AE}^2] \leq 2^{(1-\epsilon)n},$$

where $\tilde{\rho}_{AE} = \rho_E^{-1/4} \rho_{AE} \rho_E^{-1/4}$. Then, there exists $\eta > 0$ (depending only on ϵ) such that for all ℓ,

$$\sum_{\nu : |\nu| = \ell} \mathrm{tr}\left[\mathrm{tr}_A[\sigma_\nu \tilde{\rho}_{AE}]^2 \right] \leq 12 n^4 \cdot (3 - \eta)^\ell \binom{n}{\ell}$$

Proof Fix $m = \lceil 4\ell/3 \rceil$ and apply Theorem C.2, we obtain

$$\mathop{\mathbf{E}}_{|S|=m} \left\{ \mathrm{tr}[\tilde{\rho}_{A_S E}^2] \right\} \leq (n^2 + 1) \cdot 2^{(1-\delta)m}.$$

But we know that

$$\begin{aligned} \sum_{S : |S| = m} \mathrm{tr}[\tilde{\rho}_{A_S E}^2] &= \sum_{S : |S| = m} \frac{1}{2^m} \sum_{\nu \in \{0,1,2,3\}^S} \mathrm{tr}[\mathrm{tr}_A[\sigma_\nu \tilde{\rho}_{AE}]^2] \\ &\geq \frac{1}{2^m} \sum_{\nu \in \{0,1,2,3\}^n : |\nu| = \ell} \binom{n-\ell}{m-\ell} \mathrm{tr}[\mathrm{tr}_A[\sigma_\nu \tilde{\rho}_{AE}]^2], \end{aligned}$$

by simply forgetting the terms $\mathrm{tr}[\mathrm{tr}_A[\sigma_\nu \tilde{\rho}_{AE}]^2]$ for which $|\nu| \neq \ell$. Note that $\binom{n-\ell}{m-\ell}$ is the number of sets S of size m in which the support of ν is included. As a result, we have

$$\begin{aligned} \sum_{\nu : |\nu| = \ell} \mathrm{tr}[\mathrm{tr}_A[\sigma_\nu \tilde{\rho}_{AE}]^2] &\leq \frac{2^m}{\binom{n-\ell}{m-\ell}} \cdot \binom{n}{m} (n^2 + 1) 2^{(1-\delta)m} \\ &= (n^2 + 1) \binom{n}{\ell} \frac{4^m}{\binom{m}{\ell}} 2^{-\delta m}. \end{aligned}$$

To conclude, we note that $3^\ell \binom{m}{\ell} \geq 3^\ell \frac{2^{mh(3/4)}}{m(m+1)} \geq \frac{3^{3/4m-1} 2^{mh(3/4)}}{n(n+1)}$, where h is the binary entropy function. We conclude that

$$\begin{aligned} \sum_{\nu : |\nu| = \ell} \mathrm{tr}[\mathrm{tr}_A[\sigma_\nu \tilde{\rho}_{AE}]^2] &= 3(n^2 + 1) 3n(n+1) \binom{n}{\ell} 3^\ell 2^{-\delta m} \\ &\leq 12 n^4 \binom{n}{\ell} (3 - \eta)^\ell \end{aligned}$$

for an appropriate choice of constant $\eta > 0$. □

Theorem C.2 (Fully quantum entropy sampling [DFW13]). *Let ρ_{AE} be such that $\mathrm{H}_2(A|E)_\rho \geq -(1-\epsilon)n$ with $\epsilon > 0$, i.e.,*

$$\mathrm{tr}[\tilde{\rho}_{AE}^2] \leq 2^{(1-\epsilon)n},$$

where $\tilde{\rho}_{AE} = \rho_E^{-1/4}\rho_{AE}\rho_E^{-1/4}$. Then, there exists $\delta > 0$ (depending only on ϵ) such that for all m, when taking the average over all subsets S of size m,

$$\mathop{\mathbf{E}}_{|S|=m} \left\{\mathrm{tr}[\tilde{\rho}_{A_S E}^2]\right\} \leq (n^2+1)\cdot 2^{(1-\delta)m}.$$

D Properties of binomials

We use h to denote the binary entropy function $h(\alpha) = -\alpha\log(\alpha) - (1-\alpha)\log(1-\alpha)$. We use the following simple estimates for binomial coefficients (see [MU05, Lemma 9.2]). Let $\alpha \in [0,1]$ such that αn is an integer. Then

$$\sum_{k=0}^{\alpha n} \binom{n}{k} \leq 2^{nh(\alpha)}, \tag{33}$$

and

$$\frac{2^{nh(\alpha)}}{n+1} \leq \binom{n}{\alpha n}. \tag{34}$$

We also use

$$|h(\alpha+\delta) - h(\alpha)| \leq h(\delta), \tag{35}$$

for all $\alpha, \delta \geq 0$ with $\alpha + \delta \leq 1$. To prove this, we observe that $f : \alpha \mapsto h(\alpha+\delta) - h(\alpha)$ is a decreasing function of $\alpha \in [0, 1-\delta]$ and thus $|h(\alpha+\delta) - h(\alpha)| \leq \max(f(0), f(1-\delta)) = h(\delta)$. Moreover,

$$h(\alpha) \leq 2\sqrt{\alpha(1-\alpha)}. \tag{36}$$

References

[ADHW09] A. Abeyesinghe, I. Devetak, P. Hayden, and A. Winter. The mother of all protocols: Restructuring quantum information's family tree. *P. Roy. Soc. A - Math. Phy.*, 465:2537, 2009. arXiv:quant-ph/0606225.

[Ber09] M. Berta. Single-shot quantum state merging, 2009. arXiv:0912.4495.

[BF13] W. Brown and O. Fawzi. Short random circuits define good quantum error correcting codes. In *Proc. IEEE ISIT*, 2013.

[BFW12] M. Berta, O. Fawzi, and S. Wehner. Quantum to classical randomness extractors. In *Proc. CRYPTO*, volume 7417 of *LNCS*, pages 776–793, 2012. arXiv:1111.2026.

[Bha97] R. Bhatia. *Matrix Analysis*. Springer, 1997.

[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. Brown and L. Viola. Convergence rates for arbitrary statistical moments of random quantum circuits. *Phys. Rev. Lett.*, 104:250501, 2010. arXiv:0910.0913.

[BV13] W. Brown and L. Viola. Convergence rate of moments of random quantum circuits on arbitrary graphs. *in preparation*, 2013.

[DBWR10] F. Dupuis, M. Berta, J. Wullschleger, and R. Renner. One-shot decoupling. 2010. arXiv:1012.6044.

[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.

[DFW13] F. Dupuis, O. Fawzi, and S. Wehner. Entanglement sampling and applications. 2013. arXiv:1305.1316.

[dRÅR+11] L. del Rio, J. Åberg, R. Renner, O. Dahlsten, and V. Vedral. The thermodynamic meaning of negative entropy. *Nature*, 474(7349):61–63, 2011.

[dRHRW13] L. del Rio, A. Hutter, R. Renner, and S. Wehner. Relative thermalization. In preparation, 2013.

[Dup10] F. Dupuis. *The 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. *Phys. Rev. A*, 72(6):060302, 2005. arXiv:quant-ph/0503210.

[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.

[HL09] A. Harrow and R. Low. Random quantum circuits are approximate 2-designs. *Comm. Math. Phys.*, 291:257–302, 2009. arXiv:0802.1919.

[HLW06] P. Hayden, D. W. Leung, and A. Winter. Aspects of generic entanglement. *Comm. Math. Phys.*, 265(1):95–117, 2006. arXiv:quant-ph/0407049.

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

[HOW06] M. Horodecki, J. Oppenheim, and A. Winter. Quantum state merging and negative information. *Comm. Math. Phys.*, 269:107, 2006. arXiv:quant-ph/0512247.

[HP07] P. Hayden and J. Preskill. Black holes as mirrors: quantum information in random subsystems. *J. High Energy Phys.*, page 120, 2007. arXiv:0708.4025.

[HSZ12] A. Hamma, S. Santra, and P. Zanardi. Quantum entanglement in random physical states. *Phys. Rev. Lett.*, 109:040502, 2012. arXiv:1109.4391.

[Hut11] A. Hutter. Understanding Equipartition and Thermalization from Decoupling, 2011.

[HW13] A. Hutter and S. Wehner. Dependence of a quantum-mechanical system on its own initial state and the initial state of the environment it interacts with. *Phys. Rev. A*, 87:012121, Jan 2013. arXiv:1111.3080.

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

[LSH+13] N. Lashkari, D. Stanford, M. Hastings, T. Osborne, and P. Hayden. Towards the fast scrambling conjecture. *J. High Energy Phys.*, 2013(4):1–33, 2013. arXiv:1111.6580.

[MU05] M. Mitzenmacher and E. Upfal. *Probability and computing: Randomized algorithms and probabilistic analysis*. Cambridge Univ Press, 2005.

[ODP07] R. Oliveira, O.C.O. Dahlsten, and M.B. Plenio. Generic entanglement can be generated efficiently. *Phys. Rev. Lett.*, 98(13):130502, 2007. arXiv:quant-ph/0605126.

[RDR12] J. Renes, F. Dupuis, and R. Renner. Efficient polar coding of quantum information. *Phys. Rev. Lett.*, 109:050504, Aug 2012. arXiv:1109.3195.

[SDTR13] O. Szehr, F. Dupuis, M. Tomamichel, and R. Renner. Decoupling with unitary approximate two-designs. *New J. Phys.*, 15(5):053022, 2013. arXiv:1109.4348.

[SRDR13] D. Sutter, J. Renes, F. Dupuis, and R. Renner. Efficient quantum channel coding scheme requiring no preshared entanglement. In *Proc. IEEE ISIT*, 2013.

[SS08] Y. Sekino and L. Susskind. Fast scramblers. *J. High Energy Phys.*, 2008(10):065, 2008. arXiv:0808.2096.

[TGR07] G. Tóth and J.J. García-Ripoll. Efficient algorithm for multiqudit twirling for ensemble quantum computation. *Phys. Rev. A*, 75(4):042311, 2007. arXiv:quant-ph/0609052.

[Vad] S. Vadhan. Pseudorandomness.

[WR12] M. Wilde and J. Renes. Quantum polar codes for arbitrary channels. In *Proc. IEEE ISIT*, pages 334–338, 2012. arXiv:1201.2906.

[Žni08] M. Žnidarič. Exact convergence times for generation of random bipartite entanglement. *Phys. Rev. A*, 78(3):032324, 2008. arXiv:0809.0554.