The emergence of typical entanglement in two-party random processes

2007 J. Phys. A: Math. Theor. 40 8081

(http://iopscience.iop.org/1751-8121/40/28/S16)

IOP PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND THEORETICAL

J. Phys. A: Math. Theor. **40** (2007) 8081–8108 doi:10.1088/1751-8113/40/28/S16

The emergence of typical entanglement in two-party random processes

O C O Dahlsten[1,2], R Oliveira[3] and M B Plenio[1,2]

[1] The Institute for Mathematical Sciences, Imperial College London, 53 Prince's Gate, South Kensington London, SW7 2PG, UK
[2] QOLS, Blackett Laboratory, Imperial College London, London SW7 2BW, UK
[3] Instituto Nacional de Matemática Pura e Aplicada—IMPA Estrada Dona Castorina, 110 Jardim Botânico 22460-320, Rio de Janeiro, RJ, Brazil

E-mail: oscar.dahlsten@imperial.ac.uk, rimfo@impa.br and m.plenio@imperial.ac.uk

Received 13 November 2006, in final form 18 May 2007
Published 27 June 2007
Online at stacks.iop.org/JPhysA/40/8081

Abstract

We investigate the entanglement within a system undergoing a random, local process. We find that there is initially a phase of very fast generation and spread of entanglement. At the end of this phase the entanglement is typically maximal. In Oliveira *et al* (2007 *Phys. Rev. Lett.* **98** 130502) we proved that the maximal entanglement is reached to a fixed arbitrary accuracy within $O(N^3)$ steps, where N is the total number of qubits. Here we provide a detailed and more pedagogical proof. We demonstrate that one can use the so-called stabilizer gates to simulate this process efficiently on a classical computer. Furthermore, we discuss three ways of identifying the transition from the phase of rapid spread of entanglement to the stationary phase: (i) the time when saturation of the maximal entanglement is achieved, (ii) the cutoff moment, when the entanglement probability distribution is practically stationary, and (iii) the moment block entanglement exhibits volume scaling. We furthermore investigate the mixed state and multipartite setting. Numerically, we find that the mutual information appears to behave similarly to the quantum correlations and that there is a well-behaved phase-space flow of entanglement properties towards an equilibrium. We describe how the emergence of typical entanglement can be used to create a much simpler tripartite entanglement description. The results form a bridge between certain abstract results concerning typical (also known as generic) entanglement relative to an unbiased distribution on pure states and the more physical picture of distributions emerging from random local interactions.

PACS number: 03.67.Mn

M This article features online multimedia enhancements

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Entanglement is a key resource in quantum information tasks, and therefore the exploration of the structure of entanglement is of important concern in quantum information science [2]. Our quantitative understanding of this resource is very strong for bipartite entanglement; for reviews see [2–7] , or refer to [8] for an introduction to quantum information tasks. However multipartite entanglement [2, 9] is much less well understood. In particular, there appears to be a plethora of inequivalent classes of multipartite entanglement that are locally inequivalent [10–15]. There is hope that one can cut down on this plethora by considering which classes are typical (generic) relative to a certain measure on the set of states known as the unitarily invariant (Haar) measure [16]. In this measure, practically all pure states of large numbers of spins are maximally entangled [16–20]. This simplification suggests that the investigation of generic entanglement may hold some promise. However, an important question mark has existed as to whether the unitarily invariant measure is physical, in the sense that it can be approximated to arbitrary precision by two-particle interactions in a time that grows polynomially in the size of the system. The question has been raised in one form or another in [1, 21–24].

Our key objective is to determine whether this is the case. We find and prove that it is indeed possible to obtain generic entanglement properties in a polynomial number of steps in a two-party random process and give an explicit way of doing it. We also aim to gain a deeper understanding of the nature of the approach to the regime where entanglement displays generic behaviour. The paper expands on the results of [1], provides several new results and discusses future directions of this line of research.

The outline of the paper is as follows. We firstly consider the key process that is used throughout this work: random two-qubit interactions. These are modelled as two-qubit gates on a quantum computer picked at random. We then give the results of the present work. Firstly we prove that the generic entanglement as well as purity is achieved efficiently. We additionally prove that one can use the so-called stabilizer gates to simulate this process efficiently on a classical computer in the sense that the same averages will be achieved for the relevant quantities within the same time. We then discuss more in depth the observation that there is initially a phase of rapid spread of entanglement followed by a phase where the system is suffused with entanglement, and the average entanglement across any bipartite cut is practically maximal. We discuss three ways of identifying the transition between these two phases: (i) the moment of saturation of the average entanglement, (ii) the cutoff moment, and (iii) the moment the entanglement scales as the volume of the smaller of the two parties. We furthermore investigate the mixed state and multipartite setting. We find numerically that the classical correlations appear to behave similarly and that there is a well-behaved phase-space flow to the attracting equilibrium entanglement. We describe how the emergence of typical entanglement can be used to create a much simpler tripartite entanglement description. Finally, we give a conclusion as well as a discussion of the future of this line of enquiry.

The results relate certain abstract results concerning typical (also known as generic) entanglement relative to the unbiased distribution on pure states, on the one hand, to entanglement properties relative to distributions obtained by random local interactions on the other.

2. Random two-party interactions

Interactions in nature, such as the collisions of molecules in a gas, tend to be local two-party interactions. In the setting of qubits, this corresponds to two-qubit unitary gates. To have a concrete physical process in mind, consider a quantum computer which can perform

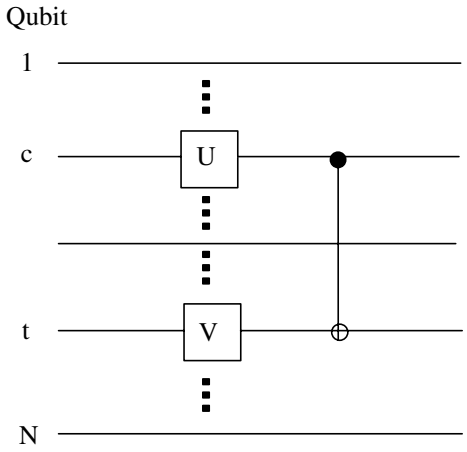

Figure 1. Two-qubit random interactions. This shows steps (iii) and (iv) of the random walk. For an explanation of such circuit diagrams see [8].

arbitrary single-qubit gates and CNOT gates[4] between any two qubits in the register. Let the randomization process be the application of random gates on randomly chosen qubits. This process leads to a distribution of pure states that will evolve over time gradually, and after a long time, approaches the flat distribution on pure states.

2.1. The random walk

We shall discuss the evolution of states $|\Psi\rangle_Q$ in an N-qubit Hilbert space $Q = Q_1 \otimes \cdots \otimes Q_N$ under a series of randomly chosen mappings. Each mapping is picked independently as in figure 1.

(i) Choose U and V independently from Haar measure on $U(2)$ (see appendix A for an introduction to the Haar measure).
(ii) Choose a pair of distinct qubits c and d, uniformly amongst all such pairs.
(iii) Apply $U \in U(2)$ to qubit c and $V \in U(2)$ to t.
(iv) Apply a CNOT with target qubit t and control qubit c.

2.2. Converges to uniform distribution

The Markov process described above converges to the uniform distribution on pure states (the distribution is described in appendix A). Each step in the process removes knowledge about our state, and asymptotically the state distribution becomes unbiased and invariant under the application of a new gate. Since CNOTs and single-qubit unitaries are universal, i.e. any unitary can be generated as a combination of them, this implies that in the asymptotic time limit the distribution is invariant under the application of any unitary. Thus the uniform distribution is induced in this limit.

It is important to note that the convergence rate to that final distribution is exponentially slow in the number of qubits, since approximating an arbitrary unitary to a fixed precision using a set of fixed-size gates requires a number of steps that grows exponentially in the number of qubits [8]. This leads one to question whether it is physically relevant to make statements relative to the unitarily invariant distribution. Interactions in nature tend to be

[4] CNOT: $|00\rangle \mapsto |00\rangle$, $|01\rangle \mapsto |01\rangle$, $|10\rangle \mapsto |11\rangle$, $|11\rangle \mapsto |10\rangle$ and is linear.

two-body interactions and should therefore not get close to the unitarily invariant distribution in a feasible time, i.e. a time that scales polynomially in the total number of qubits. We will return to that issue but firstly give some facts about the asymptotic entanglement probability distributions.

2.3. Asymptotic entanglement distribution

Consider the entanglement in the infinite time limit within a system undergoing the type of randomization described above.

When states are picked from the unitarily invariant measure there is an associated probability distribution of entanglement, whereby we mean the Von Neumann entropy of the reduced state of N_A (or N_B) particles in a system of $N = N_A + N_B$ total number of particles. Some of the first studies of this were [17, 18], and the explicit solution for the average entropy of entanglement (Page's conjecture) was conjectured in [19] and proven in [20]. It is given by

$$\mathbb{E}[S(\rho_A)] = \frac{1}{\ln 2}\left(\sum_{k=2^{N_B}+1}^{2^{N_A+N_B}} \frac{1}{k} - \frac{2^{N_A}-1}{2^{N_B+1}}\right) \tag{1}$$

with the convention that $N_A \leqslant N_B$.

This can be used to show that the average entanglement is very nearly maximal, i.e. close to $\min(N_A, N_B)$, for large quantum systems where $N \gg 1$. It is also interesting to note that there is a bound on the concentration of the distribution about this average. The probability that a randomly chosen state will have an entanglement E that deviates by more than δ from the mean value $\mathbb{E}[S(\rho_A)]$ decreases exponentially with δ^2 [16]. Therefore one is overwhelmingly likely to find a near-maximally entangled state if the system is large.

3. Theorem 1: maximal average is achieved efficiently

Despite the distribution on states requiring a number of steps that grows exponentially in the size of the system, we will now prove that achieving the entanglement distribution of generic quantum states to within any precision only requires a number of steps growing polynomially in the size of the system. An intuition why this is possible can be gained by noting that many states have the same entanglement, so attaching an entanglement value to all states results in a 'coarse-grained' state space which can be sampled in fewer steps. We now state our main theorem.

Theorem 1. *Suppose that $N_B - N_A = r \geqslant 0$ and some $\varepsilon \in (0, 1)$ is given. Then if the number of steps n satisfies*

$$n \geqslant \frac{9N(N-1)[(2\ln 2)N + \ln \varepsilon^{-1}]}{4},$$

we have

$$\mathbb{E}[S(\rho_{A,n})] \geqslant N_A - \frac{2^{-r}+\varepsilon}{\ln 2} \tag{2}$$

$$\mathbb{E}\left[\max_{|\Psi\rangle_{AB}\,\text{max entangled}} \langle\psi_n|\Psi\rangle\right] \geqslant 1 - \sqrt{\frac{4(2^{-r}+\varepsilon)}{\ln 2}}. \tag{3}$$

Note that the second statement of the theorem is most relevant for $r \gg 1$. This is because maximal entanglement N_A is not exactly achieved when $N_B - N_A = O(1)$. A similar problem is present in the asymptotic case analysis of [16].

3.1. Guide to the proof of theorem 1

Firstly, we simplify the problem by noting that the purity $\mathrm{Tr}\left(\rho_{A,n}^2\right)$ bounds the entanglement very tightly in the regime of almost maximal entanglement. Purity is more convenient to deal with, so we study the convergence rate of the purity to its asymptotic value. We expand the global density matrix in terms of elements of the Pauli group and track the time evolution of the coefficients. It will turn out that the computation of $\mathrm{Tr}\,\rho^2$ requires only the knowledge of the squares of some of these coefficients. It is then a key innovative step to realize that the relevant coefficients evolve according to a Markov chain on a small state space which we give and prove explicitly. We then use relatively recent Markov chain convergence rate analysis tools to determine how fast this chain converges to its stationary distribution. These arguments centre around the size of the 'spectral gap' of the stochastic matrix P defining the Markov chain, which simply means the difference between the second largest eigenvalue λ_1 and the eigenvalue 1. This is because $P^k = S\,\mathrm{diag}\left(1, \lambda_1^k, \lambda_2^k, \ldots\right)S^{-1}$ for some matrix S so the λ_1^k will define the most slowly decaying term and thus govern the distance to the stationary distribution.

3.2. Using purity to bound entanglement

Most of our mathematical work will be in estimating the quantity $\mathbb{E}\left[\mathrm{Tr}\left(\rho_{A,n}^2\right)\right]$. More precisely, theorem 1 follows from

Lemma 1. *For all n,*

$$\left|\mathbb{E}\left[\mathrm{Tr}\left(\rho_{A,n}^2\right)\right] - \frac{2^{N_A}+2^{N_B}}{2^N+1}\right| \leqslant 4^N \exp\left(-\frac{4n}{9N(N-1)}\right).$$

Before we prove lemma 1 we demonstrate that theorem 1 may be deduced from it. To see the implication, first note that the Von Neumann entropy $S(\rho_{A,n})$ is lower bounded by the *Rényi entropy*

$$S_2(\rho_{A,n}) = -\log_2\left(\mathrm{Tr}\left(\rho_{A,n}^2\right)\right).$$

By the concavity of $\log_2$, we have $\mathbb{E}[S_2(\rho_{A,n})] \geqslant -\log_2\left(\mathbb{E}\left[\mathrm{Tr}\left(\rho_{A,n}^2\right)\right]\right)$, and plugging in an n as suggested in the theorem

$$\mathbb{E}[S_2(\rho_{A,n})] \geqslant -\log_2\left(2^{-N_A}\frac{1+2^{-r}}{1+2^{-N}} + 2^{-N_A}\varepsilon\right) \geqslant N_A - \log_2(1+2^{-r}+\varepsilon) \geqslant N_A - \frac{2^{-r}+\varepsilon}{\ln 2}.$$

For the second statement, we note that by Uhlmann's theorem, the expectation on the LHS of (3) is given by a fidelity, Fid, of reduced density matrices. We can then use well-known relationships between distance measures on density matrices [8] to deduce

$$1 - \mathbb{E}\left[\max_{|\Psi\rangle_{AB}} \max_{\text{max. entangled}} \langle\psi_n|\Psi\rangle\right] = \mathbb{E}[1 - \mathrm{Fid}(I_A/2^{N_A}, \rho_{A,n})] \tag{4}$$

$$\leqslant \sqrt{\mathbb{E}[[1 - \mathrm{Fid}(I_A/2^{N_A}, \rho_{A,n})]^2]} \tag{5}$$

$$\leqslant \sqrt{\mathbb{E}[\|I_A/2^{N_A} - \rho_{A,n}\|_{\mathrm{tr}}]} \tag{6}$$

$$\leqslant \sqrt{2\mathbb{E}[D(\rho_{A,n}||I_A/2^{N_A})]}. \tag{7}$$

Here $D(\sigma||\rho) := \mathrm{Tr}[\sigma\log_2\sigma - \sigma\log_2\rho]$ is the relative entropy distance, which in this particular case reads

$$D(\rho_{A,n}||I_A/2^{N_A}) = N_A - S(\rho_{A,n}).$$

Using the same argument as above,

$$D(\rho_{A,n}||I_A/2^{N_A}) \leqslant N_A - \mathbb{E}[S_2(\rho_{A,n})] \leqslant N_A + \log_2\left(\mathbb{E}\left[\mathrm{Tr}\left[\rho_{A,n}^2\right]\right]\right)$$
$$\leqslant N_A + \log_2\left(\frac{2^{N_A}+2^{N_B}}{2^N+1} + 8^N \exp\left(-\frac{4n}{9N(N-1)}\right)\right),$$

and theorem 1 follows.

We are now proceeding with the proof of lemma 1. It takes an indirect route that requires a quick detour into a linear algebra. We therefore describe some basic linear algebra tools in the following subsubsection.

3.2.1. Linear algebra in the space of Hermitian operators. Let us make the following conventions: $V[c]$ represents an operator V acting on qubit c. Denote the Pauli operators (defined in appendix C) by σ_i for $\{i = 1, 2, 3\}$ and use $\sigma_0 = \mathbf{1}$. For a string $p = p_1, \ldots, p_N \in \{0, x, y, z\}^N$,

$$\Sigma(p) = 2^{-N/2} \bigotimes_{i=1}^{N} \sigma_{p_i}[i]$$

is a tensor product of Pauli operators, normalized so that $\mathrm{Tr}\left[\Sigma_p^2\right] = 1$. It is well known that the operators Σ_p form an orthonormal basis of the real vector space of Hermitian matrices over N qubits, with the inner product between A and B given by $\mathrm{Tr}(AB)$. Therefore, if we write $H = \sum_p h(p)\Sigma_p$ for a Hermitian operator H, we have

$$\mathrm{Tr}[H^2] = \sum_p h(p)^2.$$

Let us also note that if $B \subset [1, \ldots, N]$ is a non-empty subset of qubits and $A = [1, \ldots, N]\backslash B \neq \emptyset$, we can express the tracing out of B in the following form:

$$\mathrm{Tr}_B[H] = \sum_{p\in\{0,x,y,z\}^N} h(p)\,\mathrm{Tr}_B[\Sigma(p)]$$
$$= \sum_{p\in\{0,x,y,z\}^N} h(p)\left(\frac{\prod_{j\in B}\mathrm{Tr}\left[\sigma_{p_j}[j]\right]}{2^{|B|/2}}\right) \times \left(\frac{\bigotimes_{i\in A}\sigma_{p_i}[i]}{2^{|A|/2}}\right)$$
$$= 2^{|B|/2} \sum_{p\in\{0,x,y,z\}^N:\forall j\in B,\, p_j=0} h(p)\frac{\bigotimes_{i\in A}\sigma_{p_i}[i]}{2^{|A|/2}}.$$

3.2.2. Back to our problem. Let us now discuss how to apply the above to our problem. Assume that we write

$$|\psi_n\rangle\langle\psi_n| \equiv \sum_{p\in\{0,x,y,z\}^N} \xi_n(p)\Sigma(p),$$

where the $\xi_n(p)$'s are real coefficients. Then

$$\mathrm{Tr}\left(\rho_{A,n}^2\right) = \mathrm{Tr}\left[\left(2^{|B|/2} \sum_{p\in\{0,x,y,z\}^N:\forall j\notin A\, p_j=0} \xi_n(p)\frac{\bigotimes_{i\in A}\sigma_{p_i}[i]}{2^{|A|/2}}\right)^2\right]$$
$$= 2^{N_B} \sum_{p\in\{0,x,y,z\}^N:\forall j\notin A,\, p_j=0} \xi_n(p)^2. \tag{8}$$

Thus it suffices to determine the evolution of the positive coefficients $\xi_n(p)^2$.

3.3. Evolution of the coefficients

The main and final goal of this section is to map the evolution of the coefficients $\xi_n^2(p)$ under the random applications of quantum gates onto a random walk over states $p \in \{0, x, y, z\}^N$.

To achieve this end, several preliminary calculations are necessary. However, we will only use the following assumption about U and V.

Requirement 1. At each step of the process, U and V are independently chosen from a measure on $U(2)$ such that, if T is distributed according to the same measure, $F = F(A, B)$ is a bilinear function of A and B, and $a, b \in \{0, x, y, z\}$, then

$$\mathbb{E}[F(T\sigma_a T^\dagger, T\sigma_b T^\dagger)] = \begin{cases} F(I, I), & a = b = 0; \\ \frac{1}{3}\sum_{w\in\{x,y,z\}} F(\sigma_w, \sigma_w), & a = b \neq 0 \\ 0, & \text{otherwise.} \end{cases}$$

It is shown in appendix B that the Haar measure on $U(2)$ does have the above properties, and in section 4 that the flat distribution on the set of single-qubit 'stabilizer gates' also satisfies requirement 1.

3.3.1. Basic aspects. Now suppose we are given $|\psi_n\rangle_{AB}$ and choices of c, t, U and V. Then $|\psi_{n+1}\rangle = I_{[1,\ldots,N]\setminus\{c,t\}} \otimes W_n|\psi_n\rangle$, where

$$W_n = \mathrm{CNOT}[c, t](U[c] \otimes V[t]).$$

Therefore,

$$|\psi_{n+1}\rangle\langle\psi_{n+1}| = \sum_{q\in\{0,x,y,z\}^N} \xi_n(q)(I_{[1,\ldots,N]\setminus\{c,t\}} \otimes W_n)\Sigma(q)(I_{[1,\ldots,N]\setminus\{c,t\}} \otimes W_n)^\dagger$$

and for any $p \in \{0, x, y, z\}^N$ we find

$$\begin{aligned} \xi_{n+1}(p) &= \mathrm{tr}[\Sigma(p)|\psi_{n+1}\rangle\langle\psi_{n+1}|] \\ &= \frac{1}{4}\sum_{q:\forall i\notin\{c,t\} q_i=p_i} \xi_n(q)\,\mathrm{Tr}\left[\left(\sigma_{p_c}[c]\sigma_{p_t}[t]\right)W_n\left(\sigma_{q_c}[c]\sigma_{q_t}[t]\right)W_n^\dagger\right] \end{aligned} \tag{9}$$

and

$$\xi_{n+1}^2(p) = \frac{1}{16}\sum_{q,q':\forall i\notin\{c,t\} q_i=q_i'=p_i} \xi_n(q)\xi_n(q')G_n(p, q, q'), \tag{10}$$

where

$$G_n(p, q, q') \equiv \mathrm{Tr}\left[\left(\sigma_{p_c}[c]\sigma_{p_t}[t]\right)W_n\left(\sigma_{q_c}[c]\sigma_{q_t}[t]\right)W_n^\dagger\right]\mathrm{Tr}\left[\left(\sigma_{p_c}[c]\sigma_{p_t}[t]\right)W_n\left(\sigma_{q_c'}[c]\sigma_{q_t'}[t]\right)W_n^\dagger\right].$$

The above expression would appear to suggest that $\xi^2(p)$ depends on the non-positive ξ_n and ξ_n' which would prevent the formulation of a Markov process for ξ^2. However, we are only interested in averages and this will be the key for a further simplification. Let us consider the expectation of $G_n(p, q, q')$ conditioned on the values of c, t and ψ_n. Let us note first that $G_n(p, q, q')$ can be rewritten as

$$G_n(p, q, q') = \mathrm{Tr}\left[\sigma_{\hat{p}_c}U\sigma_{q_c}U^\dagger\right]\mathrm{Tr}\left[\sigma_{\hat{p}_t}V\sigma_{q_t}V^\dagger\right]\mathrm{Tr}\left[\sigma_{\hat{p}_c}U\sigma_{q_c'}U^\dagger\right]\mathrm{Tr}\left[\sigma_{\hat{p}_t}V\sigma_{q_t'}V^\dagger\right], \tag{11}$$

where $(\hat{p}_c, \hat{p}_t)$ is the unique pair in $\{0, x, y, z\}^2$ such that

$$\sigma_{\hat{p}_c}[c]\sigma_{\hat{p}_t}[t] = \pm\,\mathrm{CNOT}[c, t]\left(\sigma_{p_c}[c]\sigma_{p_t}[t]\right)\mathrm{CNOT}[c, t] \quad \text{(see table 1).} \tag{12}$$

Table 1. The map $p_c p_t \leftrightarrow \hat{p}_c \hat{p}_t$ as defined by equation (12).

00	$\leftrightarrow$	00	$x0$	$\leftrightarrow$	xx
$y0$	$\leftrightarrow$	yx	$z0$	$\leftrightarrow$	$z0$
$0x$	$\leftrightarrow$	$0x$	$0y$	$\leftrightarrow$	zy
$0z$	$\leftrightarrow$	zz	xz	$\leftrightarrow$	yy
yz	$\leftrightarrow$	xy	zx	$\leftrightarrow$	zx

Thus $G_n(p, q, q')$ is, for fixed V, a bilinear function of $U\sigma_{q_c}U^\dagger$ and $U\sigma_{q'_c}U^\dagger$, and for fixed U, a bilinear function of $V\sigma_{q_t}V^\dagger$ and $V\sigma_{q'_t}V^\dagger$. Because of requirement 1, we can deduce that $\mathbb{E}[G_n(p, q, q')|\psi_n, c, t] = 0$ *unless* $q_c = q'_c$ and $q_t = q'_t$, i.e. $q = q'$. Moreover, in this case we have

$$\mathbb{E}[G_n(p, q, q)|\psi_n, c, t] = \begin{cases} 1, & \hat{p}_c = \hat{p}_t = q_c = q_t = 0; \\ \frac{1}{3}, & \hat{p}_c = q_c = 0, \hat{p}_t \neq 0, q_t \neq 0; \\ \frac{1}{3}, & \hat{p}_c \neq 0, q_c \neq 0, \hat{p}_t = q_t = 0; \\ \frac{1}{9}, & \hat{p}_c \neq 0, q_c \neq 0, \hat{p}_t \neq 0, q_t \neq 0; \\ 0, & \text{otherwise.} \end{cases} \tag{13}$$

It follows that

$$\mathbb{E}\left[\xi^2_{n+1}(p)|\psi_n, c, t\right] = \begin{cases} \xi^2_n(p_{c\leftarrow 0, t\leftarrow 0}), & \hat{p}_c = \hat{p}_t = 0; \\ \frac{1}{3}\sum_{w\in\{x,y,z\}} \xi^2_n(p_{c\leftarrow 0, t\leftarrow w}), & \hat{p}_c = 0, \hat{p}_t \neq 0; \\ \frac{1}{3}\sum_{w\in\{x,y,z\}} \xi^2_n(p_{c\leftarrow w, t\leftarrow 0}), & \hat{p}_c \neq 0, \hat{p}_t = 0; \\ \frac{1}{9}\sum_{w,w'\in\{x,y,z\}} \xi^2_n(p_{c\leftarrow w, t\leftarrow w'}), & \hat{p}_c \neq 0, \hat{p}_t \neq 0, \end{cases} \tag{14}$$

where $p_{c\leftarrow w, t\leftarrow w'}$ is the string that equals p with p_c and p_t replaced by w and w', respectively. Thus we may now determine a Markov chain on the coefficients $\xi^2(p)$.

3.3.2. The mapping to a Markov chain. We now note the following key facts. First, $\left\{\xi^2_n(q)\right\}_{q\in\{0,x,y,z\}^N}$ is *a probability distribution* over $\{0, x, y, z\}^N$ because

$$\sum_q \xi^2_n(q) = \mathrm{Tr}[|\psi_n\rangle\langle\psi_n|^2] = 1.$$

Second, consider the following way to generate a random $p \in \{0, x, y, z\}^N$ from an element $q = q_1, \ldots, q_N \in \{0, x, y, z\}^N$:

(i) choose a pair $(c, t) \in [1, \ldots, N]^2$ of distinct elements in $[1, \ldots, N]$, uniformly amongst all such pairs;
(ii) if $q_c = 0$, set $w_c = 0$; else, choose $w_c \in \{x, y, z\}$ uniformly at random;
(iii) if $q_t = 0$, set $w_t = 0$; else, choose $w_t \in \{x, y, z\}$ uniformly at random;
(iv) set $p = q_{c\leftarrow \hat{w}_c, t\leftarrow \hat{w}_t}$ according to the mapping equation (12)

We *claim* that *if* q is distributed according to $\left\{\xi^2_n(q)\right\}_q$, the distribution of p is given by $\{\mathbb{E}[\xi_{n+1}(p)|\psi_n]\}_p$. In fact, fix c, t. The 'hat map' $w_c w_t \leftrightarrow \hat{w}_c\hat{w}_t$ is self-inverse, hence the above choice of p corresponds to setting $\hat{p}_c\hat{p}_t = w_c w_t$ and $p_i = q_i$ for all $i \in [1, \ldots, N]\backslash\{c, t\}$. Direct inspection reveals that the probability of obtaining p (given c, t) is given precisely by (14), which (by averaging over c, t) proves the claim.

We have shown the following.

Lemma 2. *Assume that P is a Markov chain on $\{0, x, y, z\}^N$ where the transitions from a state $q \in \{0, x, y, z\}^N$ are described by the above random choices. Start this chain from q_0 distributed according to $\{\xi_0^2(q)\}_q$. Then the distribution q_n of the state of the chain at time n satisfies*

$$\forall p \in \{0, x, y, z\}^N, \qquad \mathbb{P}(q_n = p) = \sum_{q,p} P^n(p, q)\xi_0^2(p) = \mathbb{E}[\xi_n(p)^2|\psi_0].$$

As a result, if $Z_A \equiv \{p \in \{0, x, y, z\}^N : \forall i \in [1, \ldots, N]\backslash A,\ p_i = 0\}$, then (by (8))

$$\mathbb{E}\big[\operatorname{Tr}\big[\rho_{A,n}^2\big]\big|\psi_0\big] = 2^{N_B}\mathbb{P}(q_n \in Z_A) = 2^{N_B} \sum_{p:\{N_n \subset A\}} \xi_n^2(p). \tag{15}$$

3.4. Analysis of the Markov chain and proof of the main result

Now we need to analyse the Markov chain that we have defined above with respect to its convergence properties. To this end we further simplify the problem by relating it to a simpler Markov chain using some standard techniques that will be outlined below.

3.4.1. Reduction of the chain. Define

$$\mathcal{X}(q) \equiv \{i \in [1, \ldots, N] : q_i \neq 0\}. \tag{16}$$

One can easily show the following.

Proposition 1. *For all n, all $E \subseteq [1, \ldots, N]$, all q with $\mathcal{X}(q_n) = E$, all $e \in E, d \notin E$, if $q_0, q_1, q_2, \ldots$ is an evolution of the Markov chain P,*

$$\mathbb{P}(\mathcal{X}(q_{n+1})) = E \cup \{d\}|q_n = q = \frac{2|E|}{3\binom{N}{2}},$$

$$\mathbb{P}(\mathcal{X}(q_{n+1}) = E\backslash\{e\}|q_n = q = \frac{2(|E| - 1)}{9\binom{N}{2}},$$

where $|E|$ is the cardinality of set E.

Proof. Assume that we are given $q_n = q, \mathcal{X}(q_n) = E$ and consider the random procedure for P described in the previous section. If $q_c = q_t = 0$, then $p = q$. If $q_c = 0$ but $q_t \neq 0$, then $w_c = 0$ and w_t is uniformly chosen from $\{x, y, z\}$; hence $p_c p_t = \hat{w}_c \hat{w}_t$ is uniform from $\{0x, zz, zy\}$, i.e. $\mathcal{X}(q)$ is replaced by $\mathcal{X}(q) \cup \{c\}$ with probability $2/3$ and remains the same with probability $1/3$. Similarly, if $q_c \neq 0$ but $q_t = 0$, t is added to $\mathcal{X}(q)$ with probability $2/3$ and stays the same otherwise. Finally, if we have $q_c \neq 0, q_t \neq 0, w_c w_t$ is chosen uniformly from $\{x, y, z\}^2$, hence $p_c p_t$ is uniform over

$$\{y0, 0z, yz, xy, x0, 0y, xz, yy, zx\}.$$

Thus in this case there are three possibilities: c (and c alone) is removed from $\mathcal{X}(q)$ (probability $2/9$); t (and t alone) is removed from $\mathcal{X}(q)$ (probability $2/9$); or nothing happens (probability $5/9$). We deduce from the above that an element $d \in [1, \ldots, N]\backslash E$ can be added to E only if it is one of $\{c, t\}$, and the remaining element in $\{c, t\}$ belongs to $\mathcal{X}(q) = E$. For each d there are $2|E|$ such pairs, out of all possible $N(N-1)$, and if such a pair is chosen, d is added with probability $2/3$. On the other hand, an element $e \in E$ can be removed only if it is one of $\{c, t\}$ and the remaining element is in E, in which case (corresponding to $2(|E| - 1)$ out of $N(N-1)$ pairs), e is actually removed with probability $2/9$. These assertions imply the proposition. □

By proposition 1, $\{\mathcal{X}_n \equiv \mathcal{X}(q_n)\}_n$ is a Markov chain. Since the only event we are interested in is $\{q_n \in Z_A\} = \{\mathcal{X}_n \subset A\}$ (see equation (15)), we may restrict our attention to this 'reduced' chain. For convenience, we state this as a proposition.

Proposition 2. *Let $\{\mathcal{X}_n\}_n$ be the Markov chain defined according to P and proposition 1 above, started from a $\mathcal{X}_0 = \mathcal{X}(q_0)$ with q_0 distributed according to the distribution $\{\xi_0^2(q)\}_q$ given by a state $|\psi_0\rangle$ (cf lemma 2). Then*

$$\mathbb{E}\left[\mathrm{Tr}\left[\rho_{A,n}^2\right]|\psi_0\right] = 2^{N_B}\mathbb{P}(\mathcal{X}_n \subset A).$$

From now on, we will only deal with the 'reduced' chain $\mathcal{X}_n$.

3.4.2. Dealing with the isolated state. It should be clear that $\mathcal{X}_n$ as defined above is *not* ergodic, as the state $\emptyset$ is isolated, i.e. there are no transitions to or from it from the rest of the state space. This corresponds to the fact that $q = 0\ldots0$ is an isolated state of the initial Markov chain.

However, this is not a problem, as we know that

$$\mathbb{P}(\mathcal{X}_n = \emptyset) = \mathbb{P}(\mathcal{X}_0 = \emptyset) = \mathbb{P}(q_0 = 0\ldots0) = \xi_0^2(0\ldots0) = \frac{\mathrm{Tr}[|\psi_0\rangle\langle\psi_0|]^2}{2^N} = \frac{1}{2^N}.$$

This means that we can neglect this state and restrict our calculations with $\mathcal{X}_n$ to the state space $\Omega = 2^{[1,\ldots,N]}\backslash\{\emptyset\}$, as we know the contribution of $\emptyset$ to the final result.

We now wish to show that the restricted chain is *ergodic*, i.e. that it has a unique stationary distribution $\mathcal{M}$ for which

$$\forall E,\ F \in \Omega,\ \lim_{n\to+\infty}\mathbb{P}(\mathcal{X}_n = F|\mathcal{X}_0 = E) = \mathcal{M}(F).$$

To prove this, it suffices [25] to show that the chain is *irreducible* and *aperiodic*. Irreducibility holds when there are sequences of valid transitions between any pair of states in Ω, which can be easily checked in our case. Aperiodicity means that there is no way to split $\Omega = \Omega_1 \cup \Omega_2$ so that all transitions happen from a state in one of Ω_1 or Ω_2 to a state in the other set. But this is implied by the fact that $\mathbb{P}(\mathcal{X}_{n+1} = E|\mathcal{X}_n = E) > 0$ for all $E \in \Omega$. This proves ergodicity, as desired.

3.4.3. Stationary distribution. We now prove that the $\mathcal{X}_n$ chain on Ω is *reversible* and determine the stationary distribution $\mathcal{M}$. *Reversibility* means that the stationary distribution $\mathcal{M}$ on Ω satisfies the *detailed balance condition*: for all distinct $C, D \in \Omega$,

$$\mathcal{M}(C)\mathbb{P}(\mathcal{X}_1 = D \mid \mathcal{X}_0 = C) = \mathcal{M}(D)\mathbb{P}(\mathcal{X}_1 = C \mid \mathcal{X}_0 = D). \tag{17}$$

Since we know that the chain is ergodic, the *existence* of a $\mathcal{M}$ satisfying the above equation implies that this $\mathcal{M}$ is the unique stationary distribution of the $\mathcal{X}_n$ process.

We make the ansatz that $\mathcal{M}(C) = f(|C|)$ is a function of $|C|$ only. In the above comparison of C and D, we can assume without loss of generality that $|C| = k$ and $|D| = k+1$ for some $1 \leqslant k \leqslant N-1$. Then the reversibility condition becomes

$$f(k)\frac{4k}{3N(N-1)} = f(k+1)\frac{4k}{9N(N-1)} \Rightarrow f(k+1) = 3f(k) \Rightarrow f(k) = \frac{3^k}{Z},$$

where Z is a normalizing factor determined by the condition $\sum_{C\in\Omega}\mathcal{M}(C) = 1$, i.e.

$$Z = \sum_{C\in\Omega} 3^{|C|} = \sum_{k=1}^{N}\binom{N}{k}3^k = 4^N - 1.$$

Thus

$$\mathcal{M}(C) = \frac{3^{|C|}}{4^N - 1}, \qquad C \in \Omega$$

is the unique stationary distribution of the $\{\mathcal{X}_n\}$ chain restricted to Ω.

3.4.4. Limits. Recall from lemma 2 that the quantity we wish to estimate is $\mathbb{E}\big[\mathrm{Tr}\left(\rho_{A,n}^2\right)\big|\psi_0\big] = 2^{N_B}\mathbb{P}(\mathcal{X}_n \subset A)$. Using ergodicity of the chain restricted to Ω, we know that this quantity converges as $n \to +\infty$ to

$$\begin{aligned}
\lim_{n\to+\infty} 2^{N_B}\mathbb{P}(\mathcal{X}_n \subset A) &= 2^{N_B}\left\{\frac{1}{2^N} + \left(\frac{2^N-1}{2^N}\right)\sum_{\emptyset\neq C\subset A}\mathcal{M}(C)\right\} \\
&= 2^{N_B-N}\left\{1 + \frac{2^N-1}{4^N-1}\left(\sum_{\emptyset\neq C\subset A} 3^{|C|}\right)\right\} \\
&= 2^{N_B-N}\left(1 + \frac{1}{2^N+1}\sum_{k=1}^{N_A}\binom{N_A}{k}3^k\right) \\
&= 2^{-N_A}\left(1 + \frac{4^{N_A}-1}{2^N+1}\right) \\
&= 2^{-N_A}\left(\frac{2^N+4^{N_A}}{2^N+1}\right) = \frac{2^{N_B}+2^{N_A}}{2^N+1}.
\end{aligned}$$

Thus (again using lemma 12)

$$\forall\psi_0,\ \lim_{n\to+\infty}\mathbb{E}\big[\mathrm{Tr}\left(\rho_{A,n}^2\right)\big|\psi_0\big] = \frac{2^{N_B}+2^{N_A}}{2^N+1}. \tag{18}$$

This result can also be deduced directly from the convergence of $|\psi_n\rangle$ to a uniformly random state as $n \to +\infty$, together with known formulae for the expected purity.

We now consider the rate at which this limit is approached during the random process.

3.4.5. Mixing of the reduced Markov chain. Our main goal in this section is to prove bounds on the mixing time of the Markov chain given by $\mathcal{X}_n$. We will take an indirect route to do so.

Lemma 3. *The Markov chain given by $\mathcal{X}_n$ has spectral gap $\geqslant 4/9N(N-1)$.*

Proof. Consider a chain B_n on Ω defined as follows. Assume $B_n = B$ and choose a $1 \leqslant j \leqslant N$ uniformly at random. If $j \in B$ and $|B| \geqslant 1$, set $B_{n+1} = B\backslash\{j\}$ with probability $1/3$ and $B_{n+1} = B$ with probability $2/3$. If $j \notin B$, set $B_{n+1} = B \cup \{j\}$. We *claim* the following.

Claim 1. *B_n is reversible, ergodic and has the same stationary distribution $\mathcal{M}$ as $\mathcal{X}_n$. Moreover, the spectral gap of B_n is at least $1/3N$.*

Before proving the claim, we show how it implies the lemma, i.e. the $\geqslant 4/9N(N-1)$ bound for the spectral gap of $\mathcal{X}_n$. This is possible via a *comparison* of Markov chains. By theorem 2.14 in [41], it suffices to show that for all distinct $C, D \in \Omega$,

$$\mathcal{M}(C)\mathbb{P}(\mathcal{X}_1 = D \mid \mathcal{X}_0 = C) \geqslant \frac{4\mathcal{M}(C)}{3(N-1)}\mathbb{P}(B_1 = D \mid B_0 = C). \tag{19}$$

Indeed, because both $\mathcal{X}_n$ and B_n are reversible chains, both the LHS and RHS are symmetric in C, D. Therefore, in proving equation (19) we can assume that $D = C \cup \{j\}$ for some $j \in [1, \ldots, N]\backslash C$. Then it is easy to see that

$$\mathcal{M}(C)\mathbb{P}(\mathcal{X}_1 = D \mid \mathcal{X}_0 = C) = \frac{2|C|\mathcal{M}(C)}{3\binom{N}{2}} \geqslant \frac{4\mathcal{M}(C)}{3(N-1)}\mathbb{P}(B_1 = D \mid B_0 = C).$$

To finish, we must prove claim 1.

Proof of claim 1. Reversibility of B_n follows from the fact that if $C \in \Omega$, $D = C \cup \{d\} \in \Omega$ with $d \notin C$ are given,

$$\mathcal{M}(C)\mathbb{P}(B_1 = D \mid B_0 = C) = \frac{\mathcal{M}(C)}{N} = \frac{\mathcal{M}(D)}{3N} = \mathcal{M}(D)\mathbb{P}(B_1 = C \mid B_0 = D).$$

We use the *path-coupling* technique of Bubler and Dyer (see e.g. [25]) to prove ergodicity and the desired spectral gap bound for B_n. For $B, B' \in \Omega$, let $d(B, B') = |B \Delta B'|$ be the Hamming distance between B and B'. Call B and B' *adjacent* if $d(B, B') = 1$. We will show that one can couple one-step evolutions B_1, B_1' started from adjacent B_0, B_0' so that

$$\mathbb{E}[d(B_1, B_1') - d(B_0, B_0')] \leqslant -\frac{1}{3N}. \tag{20}$$

This means (cf. reference [25]) that for *arbitrary* B_0, B_0', there is a coupling of $B_1, \ldots, B_n$ and $B_1', \ldots, B_n'$ such that for all $n \geqslant 0$,

$$\mathbb{E}[d(B_n, B_n')] \leqslant \left(1 - \frac{1}{3N}\right)^n d(B_0, B_0') \leqslant N\left(1 - \frac{1}{3N}\right)^n. \tag{21}$$

From this it follows via a standard argument that the statistical distance between B_n and B_n' decays at an exponential rate of $\leqslant(1 - 1/3N)$, which also implies the desired spectral gap bound.

The coupling in (20) is indeed very simple. Suppose we are given adjacent B_0 and B_0'. Without loss of generality, we can assume that $B_0 = \{1, \ldots, k\}$ and $B_0' = \{1, \ldots, k+1\}$ for some $1 \leqslant k \leqslant N-1$. We choose a $j \in [1, \ldots, N]$ uniformly at random and update B_0 and B_0' in the following way.

(i) If $1 \leqslant j \leqslant k$,
  (a) if $k \geqslant 2$, set $B_1' = B_0'\backslash\{j\}$ and $B_1 = B_0\backslash\{j\}$ with probability $1/3$ and do not change the states with probability $2/3$;
  (b) else if $k = 1$ do this only for B_1', leaving $B_1 = B_0$ always;

(ii) else if $j = k+1$,
  (a) set $B_1 = B_1' = B_0'$ with probability $2/3$
  (b) OR $B_1 = B_0'$, $B_1' = B_0$ with probability $1/3$;

(iii) else if $k+2 \leqslant j \leqslant N$, set $B_1 = B_0 \cup \{j\}$, $B_1' = B_0' \cup \{j\}$.

Clearly, B_1 and B_1' each have the right distribution. We have $d(B_1, B_1') = 2$ only if $j = 1$ (probability $1/N$), $k = 1$ (so that case $(i)b$ holds) and moreover $B_1 = B_0\backslash\{j\}$ (probability $1/3N$). Hence

$$\mathbb{P}(d(B_1, B_1') = 2) = \frac{1}{3N}.$$

On the other hand, $d(B_1, B_1') = 0$ when $j = k+1$ (probability $1/N$) and we are in case (ii)(a) (prob. $2/3$), hence

$$\mathbb{P}(d(B_1, B_1') = 0) = 2/3N.$$

One can check that $d(B_1, B_1') = d(B_0, B_0')$ in all other cases, so that

$$\mathbb{E}[d(B_1, B_1') - d(B_0, B_0')] = \frac{1}{3N} - \frac{2}{3N} = -\frac{1}{3N},$$

i.e. equations (20) and (21) follow. □

3.4.6. End of proof of the main theorem. We have shown in the introduction that the theorem follows from lemma 1. Moreover, proposition 2 shows that (omitting the initial state),

$$\mathbb{E}\big[\mathrm{Tr}\left(\rho_{A,n}^2\right)\big] = 2^{N_B}\mathbb{P}(\mathcal{X}_n \subset A),$$

where $\mathcal{X}_0$ has distribution given by $|\psi_0\rangle$ in the manner described above. Note that

$$\mathbb{P}(\mathcal{X}_n \subset A) = \mathbb{P}(\mathcal{X}_0 = \emptyset) + \mathbb{P}(\mathcal{X}_0 \neq \emptyset, \mathcal{X}_n \subset A) = \frac{1 + (2^N - 1)\mathbb{P}(\mathcal{X}_n \subset A | \mathcal{X}_n \neq \emptyset)}{2^N}.$$

Using the ergodicity of $\mathcal{X}_n$ restricted to Ω and the limit formula in section 3.4.4,

$$\left|2^{N_B}\mathbb{P}(\mathcal{X}_n \subset A) - \frac{2^{N_A} + 2^{N_B}}{2^N + 1}\right| \leqslant 2^{N_B}\|\mathcal{X}_n - \mathcal{M}\|_{sd},$$

where $\|\mathcal{X}_n - \mathcal{M}\|_{sd}$ is the statistical distance between $\mathcal{M}$ and the distribution of $\mathcal{X}_n$ restricted to Ω. However, we know that

(i) the restriction of $\mathcal{X}_n$ is ergodic and reversible;
(ii) its spectral gap is bounded below by $4/9N(N-1)$;
(iii) for any $E \in \Omega$, the probability that $\mathcal{X}_{n+1} = E$ given $\mathcal{X}_n = E$ is (cf. proposition 1)

$$\begin{aligned}\mathbb{P}(\mathcal{X}_{n+1} = E | \mathcal{X}_n = E) &= 1 - \sum_{e \in [1,\ldots,N]\setminus E} \frac{2|E|}{3\binom{N}{2}} - \sum_{e \in E} \frac{2(|E|-1)}{9\binom{N}{2}} \\ &= 1 - \frac{2|E|(N-|E|)}{3\binom{N}{2}} - \frac{2|E|(|E|-1)}{9\binom{N}{2}} \\ &= 1 - \frac{2\left((N-1)|E| - \frac{4}{3}|E|^2\right)}{3\binom{N}{2}}.\end{aligned}$$

A simple calculation shows that the numerator above is at most $3(N-1)^2/8 \leqslant 3/4\binom{N}{2}$ and that $\mathbb{P}(\mathcal{X}_{n+1} = E | \mathcal{X}_n = E) \geqslant 1/4$ always.

It follows from standard Markov chain theory (e.g. [42, corollary 1.15]) that all eigenvalues of the $\mathcal{X}_n$ chain that are different from 1 lie between $-3/4 + 1/4 = -1/2$ and $1 - 4/9N(N-1)$ and that for any initial distribution of $\mathcal{X}_0$:

$$\|\mathcal{X}_n - \mathcal{M}\|_{sd} \leqslant \frac{\left(1 - \frac{4}{9N(N-1)}\right)^n}{\sqrt{\min_{C \in \Omega} \mathcal{M}(C)}} \leqslant 2^N \exp\left(-\frac{4n}{9N(N-1)}\right).$$

Since $2^{N_B} \leqslant 2^N$, this implies that

$$\left|\mathbb{E}\big[\mathrm{Tr}\left(\rho_{A,n}^2\right)\big] - \frac{2^{N_A} + 2^{N_B}}{2^N + 1}\right| \leqslant 4^N \exp\left(-\frac{4n}{9N(N-1)}\right),$$

which finishes the proof of theorem 1.

4. Efficient simulation of process on a classical computer

Theorem 1 and lemma 1 concern the rate of convergence of the average entanglement and purity respectively during the random walk defined in section 2. This random walk can be simulated on a classical computer. However, since it is a random walk on a quantum state space, one is limited to comparatively small systems, as the quantum state space grows exponentially in N.

It will now be proven that the average entanglement and purity evolution during the random walk can be efficiently simulated on a classical computer. This is achieved in two stages: firstly it is noted that one can restrict the single-qubit unitaries allowed to be a particular finite subset whilst still retaining the validity of lemma 1 (and thereby theorem 1). Secondly it is shown that the restricted quantum state space can be simulated efficiently on a classical computer using so-called stabilizer states.

4.1. Restricted walk and theorem 1

Lemma 4. *Consider the random walk defined in section 2. Change step (i) to pick U and V independently from the uniform distribution of single-qubit stabilizer gates. Let steps (ii)–(iv) remain unchanged. Then lemma 1 (and thus theorem 1) still hold.*

Proof. In the notation defined in lemma 6 in appendix B, the six stabilizer states for one qubit are given by $(r_x, r_y, r_z) = (1, 0, 0), (-1, 0, 0), (0, 1, 0), (0, -1, 0), (0, 0, 1), (0, 0, -1)$. One can use the proof that Haar measure on $U(2)$ satisfies requirement 1 which is provided in appendix B, but replacing the Haar measure on $U(2)$ with the flat distribution on the six stabilizer states. Requirement 1 is the only requirement made about how we pick the single-qubit gates U and V in the proof of lemma 1 and as was previously shown, lemma 1 implies theorem 1. □

4.2. Efficient simulation of entanglement evolution

The restricted walk will remain within the state space of stabilizer states (see appendix C for an introduction to these states). It is well known that the gates used form a universal set of stabilizer gates, so the walk will be able to access all such states. It is furthermore well known that stabilizer states are efficiently simulable, in the sense that the number of parameters required to uniquely specify a single-stabilizer states grows only polynomially in N.

In order to efficiently track the evolution of the entanglement during a numerical simulation, we in addition require the ability of evaluating the entanglement of a stabilizer state with a $poly(N)$ effort. This is not trivial since we must evaluate the entanglement without converting the stabilizer state description to a standard quantum state description. To achieve this we use the tools in [32][5] to evaluate the entanglement after each step in the random walk. Examples of such numerical simulations of the restricted stabilizer walk can be seen in figures 2 and 3.

It is worth noting that the exact entanglement (E) probability distribution in the asymptotic limit, wherein one has an unbiased distribution on stabilizer states, was derived in [31] and is given by

$$P(E) = \frac{\prod_{i=1}^{N_A}(2^i+1)}{\prod_{k=N-N_A+1}^{N}(2^k+1)} \prod_{j=1}^{E} \frac{(2^{N-N_A+1-j}-1)(2^{N_A+j}-2^{2j-1})}{2^{2j}-1}, \tag{22}$$

[5] Note that the Matlab codes in this work can be downloaded from www.imperial.ac.uk/quantuminformation.

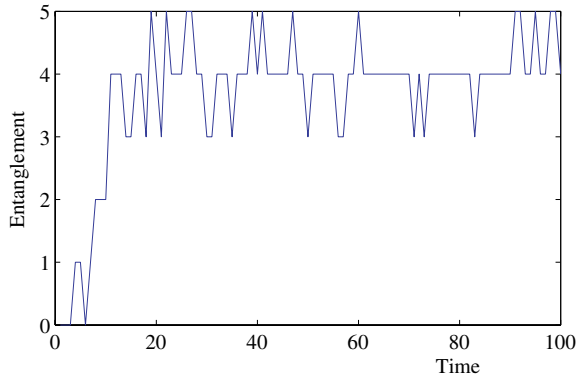

Figure 2. Entanglement evolution of stabilizer states where $N = 10$ and $N_A = 5$. Here a single realization only. The dramatic jumps reflect the fact that stabilizer entanglement only comes in integer values.

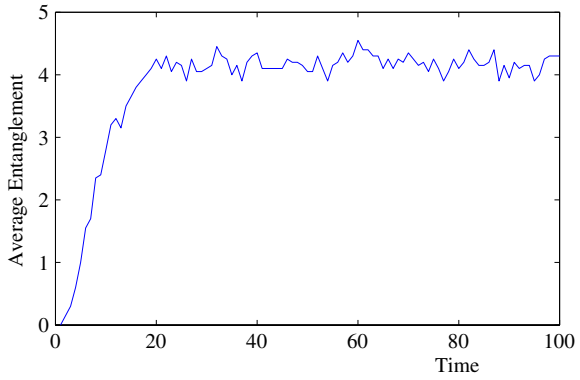

Figure 3. Here averaged over 20 realizations. One sees that the average behaviour is like that of general states .

where E is an integer since the entanglement of stabilizer states only comes in integers. The above expression can be used to show that the average is nearly maximal, and the distribution squeezes up around the average with increasing N [31]. The expected entanglement becomes equal to that of general states as $\frac{N_A}{N_B} \to 0$ and is approximately equal otherwise [31]. In turn, the expected purity of stabilizer states in the asymptotic time regime is identical to that of general states given by equation (18) [28][6].

It is hoped that the approach of simulating the entanglement evolution using stabilizer states will contribute to our ability of simulating highly entangled many-body systems. Stabilizer states, despite having various restrictions, possess a rich entanglement structure exhibiting multi-partite entanglement [29–34] and may be used in the approximate description of ground states of Hamiltonians [35].

In the following section we shall use the method described and justified above to simulate the entanglement evolution of 200 qubits undergoing the random walk.

[6] An alternative manner of explaining why the asymptotic time average of the purity is the same, whether using the stabilizer circuit or the general state circuit, is the concept of 2-designs; see [36, 37].

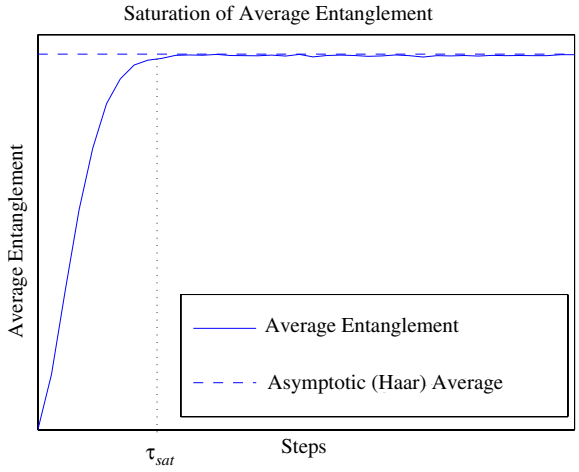

Figure 4. Illustrating τ_{sat}. We find that this behaviour is generic from numerical simulations.

5. Two phases in approach to typical entanglement

Theorem 1 tells us that systems undergoing this sort of random interactions will tend to become maximally entangled very fast. In the remaining section we would like to study this phenomenon in some more detail employing numerical tools both with the aim of providing a more intuitive understanding and to reveal several conjectures that may serve as motivation for future work.

These numerical studies reveal a finer structure in the approach, namely two phases, apparently separated by a well defined and short transition interval.

(i) A phase in which entanglement is rapidly increasing and spreading through the system.
(ii) A phase in which entanglement has spread through the entire system.

We identify three ways of defining the moment of time that separates these two phases that will be explained in the following: the saturation moment τ_{sat}, the cutoff moment τ_{cutoff} or the volume scaling moment τ_{vol}.

5.1. Saturation moment

Applying the random interaction we make the following.

Numerical observation. There is a moment of saturation of the average entanglement.

Before this moment the average block entanglement increases essentially linearly in the time n. After the transition moment it is however practically constant and nearly maximal. Therefore we term the transition time between the two regimes the 'saturation time' τ_{sat}. There is some degree of freedom in what exact value to assign to this time. One could for example specify $\tau_{\text{sat}}(\epsilon)$ as the moment that the gradient of the average entanglement curve is within some fixed accuracy ϵ to 1. Figure 4 shows how this moment is reached.

5.2. Cutoff moment

The numerical observation that there appears to be two time-scales involved here, first a rapid approach to the asymptotic value and then a slow one, led us to study the statistical mathematics

literature for tools that quantify this. In fact there has been extensive study of such problems, motivated by the fact that a randomization process, such as shuffling cards in a casino, is in practice performed only a finite number of times. The question is then, how many shuffles are necessary before one is certain, for all practical purposes, that the cards are shuffled. In the setting concerned here, that corresponds to asking when we are certain, for practical purposes, that the entanglement probability distribution has achieved its asymptotic form. The tool we will use here is the so-called 'cutoff effect', which is exhibited by many Markov chains [39].

The cutoff refers to an abrupt approach to the stationary distribution occurring at a certain number of steps taken in the chain. Say we have a Markov chain defined by its transition matrix P, and that it converges to a stationary distribution π. Initially the total variation distance $TV = \|P - \pi\| = \sup|P(E) - \pi(E)|$ between the corresponding probability distributions is given by $TV = 1$. After k steps this distance is given by $TV(k) = \|P^k - \pi\|$. A cutoff occurs, basically, if $TV(k) \simeq 1$ for $k = 0, 1, 2, \ldots, a$ and thereafter falls quickly such that after a few steps $TV(k) \simeq 0$. As we increase the size of the state space, the number of steps during which the abrupt approach takes place should decrease compared to a, the number of steps necessary to reach the cutoff. Then, for very large state spaces, we can say that the randomization occurs at a steps, some function of the size of the state space.

As a precise definition of a cutoff we use [39].

Definition. *Let P_n, π_n be Markov chains on sets χ_n. Let a_n, b_n be functions tending to infinity, with $\frac{b_n}{a_n}$ tending to zero. Say the chains satisfy an a_n, b_n cutoff if for some starting states x_n and all fixed real θ with $k_n = \lfloor a_n + \theta b_n \rfloor$, then*

$$\left\| P_n^{k_n} - \pi_n \right\| \longrightarrow c(\theta), \tag{23}$$

with $c(\theta)$ being a function tending to zero for θ tending to infinity and to 1 for theta tending to minus infinity.

With this definition it is now possible to study whether there is a sharp cutoff time associated with the entanglement probability distribution, a functional on our specific Markov process. Indeed we make the following observations.

Numerical observation. We observe an apparent cutoff effect in the entanglement probability distribution under the two-particle interaction random process described in this work.

Figure 5 shows an example of a cutoff for general states. The state space has been discretized by rounding off entanglement values to the nearest integer. We observe that $TV \simeq 1$ for a while and then falls. Finally, there is a stage where $TV \simeq 0$. The effect becomes more pronounced with increasing N.

Now consider the analogous situation for stabilizer states. From lemma 4 we would expect this behaviour to be representative of that of general states. Since stabilizer states are efficiently parametrized, using them here will allow us to scale further with N.

Numerical observation. We observe an apparent cutoff effect in the entanglement probability distribution under the stabilizer two-particle interaction random process described in this work.

How this squeezes up is showed for individual runs, averaged over 1000 realizations, in figure 6.

As stated, the cutoff effect is common in classical Markov chains such as card shuffling [39]. It is a testament to the universality of mathematics that applying quantum gates at random to a quantum register apparently exhibits the same features, in this regard, as applying shuffles to a deck of cards.

We term this the cutoff moment τ_{cutoff}. Using this moment to separate the first and second phases has the advantages that it unambiguously gives one moment, and this moment

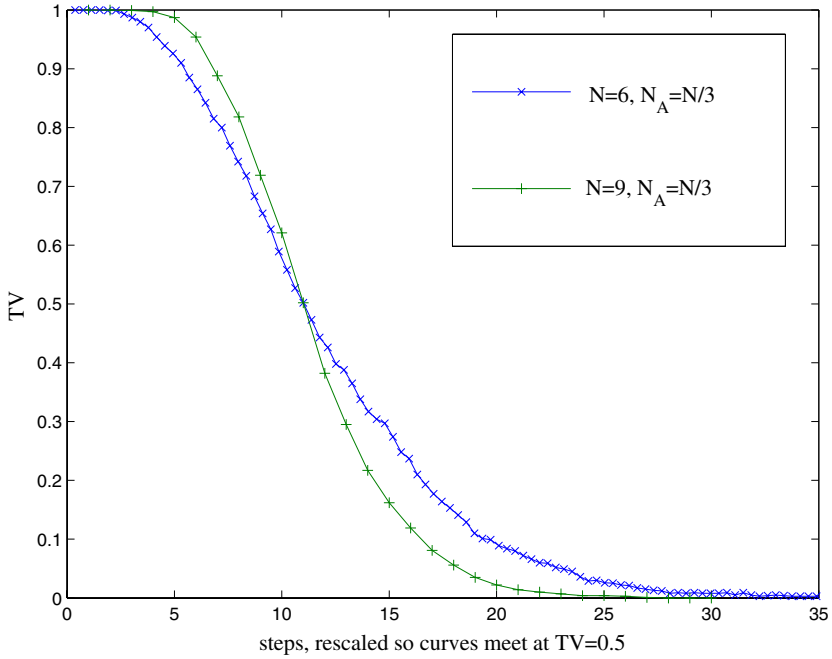

Figure 5. Observe cutoff effect for general states. The total variation distance to the asymptotic entanglement probability distribution, $TV \simeq 1$ for some finite time, before falling to 0. The fall becomes more dramatic with increasing system size. Note that, as is the customary way of representing this effect, we have rescaled the curves to meet where $TV \simeq 0.5$.

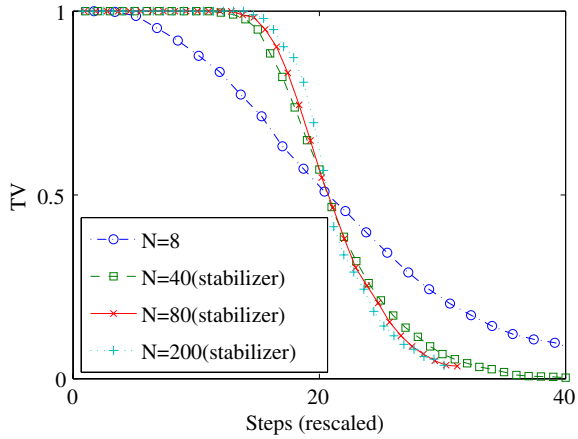

Figure 6. The cutoff effect in the total variation distance (TV) is here studied for larger N by employing stabilizer states and tools to efficiently evaluate their entanglement [32]. One sees that the effect becomes increasingly pronounced with increasing N. We conjecture that the function is a step function in the limit of $N \to \infty$, consistent with the normal behaviour of the cutoff effect. The curves have been resized so that they meet at $TV = 0.5$. This is a customary method of showing a cutoff. To see how the picture relates to the precise definition, pick any two allowed values for TV, e.g. 0.5 and 0.3. Let a_n be the time at which $TV = 0.5$ and $a_n + b_n\theta$ be the time $TV = 0.3$. There is then a cutoff if $b_n\theta/a_n$ vanishes when increasing the number of qubits. The figure indicates that this is indeed the case, since the resized curves approach a step-function in that limit.

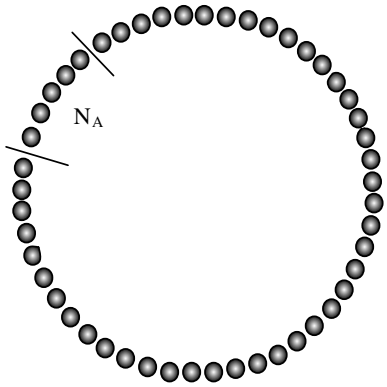

Figure 7. *N* qubits with nearest neighbour interactions. The stabilizer gate random walk is applied with the only difference that the qubits c and t picked at random must be nearest neighbours.

corresponds to the point when the entanglement distribution equals, for practical purposes, the asymptotic one.

5.3. The area to volume scaling transition

We now consider briefly the two-party random process restricted to nearest neighbours interactions, as in figure 7. This helps us to relate two pictures of typical entanglement. One is the notion that following a randomizing process entanglement appears all pervasive in the system. In this case one may expect that the entanglement between two blocks scales roughly as the size of the smaller part, i.e. we observe a volume scaling. While this is the behaviour for generic states it is in strong contrast to the behaviour of pure states that are often appearing in nature such as the ground states of Hamilton operators describing systems with short range interactions. Indeed, in such systems the block entanglement in ground states has been proven to scale as the boundary surface area between the two blocks for a wide variety of systems [43–46]. This indicates that ground states of short-range Hamiltonians tend to explore only a tiny fraction of the entire Hilbert space as their entanglement properties are indeed rather atypical.

Numerical Observation. The entanglement scales as the area at small times and as the volume of the smaller block at large times.
The result could have been expected since the nearest neighbour structure is only respected at small times as at longer times the two-qubit unitaries have combined to form global unitaries.

The above observation motivates the introduction of the transition time from area-to-volume scaling τ_{vol} and is shown in figure 8. To assign an exact value to this moment one can specify $\tau_{\mathrm{vol}}(\epsilon)$ as the moment $\max|\mathbb{E}[S(\rho_A)] - \mathbb{E}[S(\rho_{A,\tau})]| \leqslant \epsilon$ where the maximization is over all partitions.

6. Multipartite and mixed states

So far we have considered pure bipartite states. We will now discuss how the results relate to the mixed state and multipartite setting, making certain numerical observations as well as a conjecture.

It is known that the unitarily invariant measure is not uniquely defined for mixed states however. This can be visualized already in the case of one qubit. On one qubit a unitary

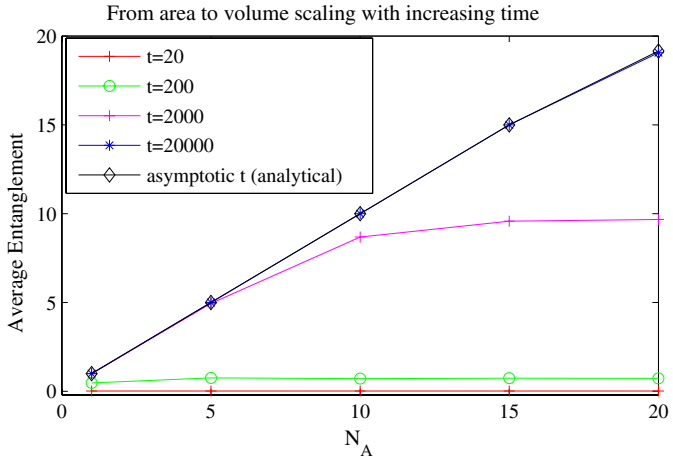

Figure 8. $N = 40$ qubits with nearest neighbour interactions. We vary N_A at different times and ask what the typical entanglement is between N_A and the rest of the system. Sufficiently high statistics were taken that the statistical error was not visible on this plot. The asymptotic curve was calculated using equation (22).

time evolution corresponds to a rotation of the Bloch sphere. Pure states lie on the surface of the sphere, and the unitarily invariant measure is a uniform distribution on the surface of the sphere. However, mixed states can lie anywhere in the ball inside the sphere, and there is no longer a unique measure as unitary invariance does not provide any restrictions on the radial distribution. Mixed measures may nevertheless be induced by considering environments C that allow for purifications of mixed states on AB. Then the measure will be obtained by using the uniform measure on ABC and employing the associated measure obtained on AB by partial tracing of C. This leads to AB being in a mixed state.

To quantify the entanglement in the mixed state we used the logarithmic negativity, E_N, defined in [47–49] and proven to be an entanglement monotone in [48]:

$$E_N(\rho) = \log_2 \|\rho^{T_A}\|_1. \tag{24}$$

Hence here the entanglement is quantified as $E_N(tr_C(\rho_{ABC}))$.

With the increasing size of the environment C the state of AB tends to become more mixed and the entanglement between A and B becomes negligible and ultimately disappears. This is interesting from the perspective of studies of emerging classicality in systems allowed to interact with environments.

The usefulness of the entanglement as a possible macroscopic parameter is highlighted in figure 9 which shows how the equilibrium values is an attractor point regardless of the initial state and that there is a flow towards this point. The latter observation hints that the average entanglement is a good parameter also in the approach to the attracting equilibrium.

One can also note that for mixed states there are classical correlations. The mutual information, $S(\rho_A) + S(\rho_B) - S(\rho_{AB})$, can be interpreted as the combination of classical and quantum correlations [50], and it is interesting to note that this quantity shows the same behaviour as the quantum correlations alone.

Numerical observation. The mutual information behaves very similarly to the quantum correlations (entanglement) in the two-particle interaction random walk simulations, as in figure 10.

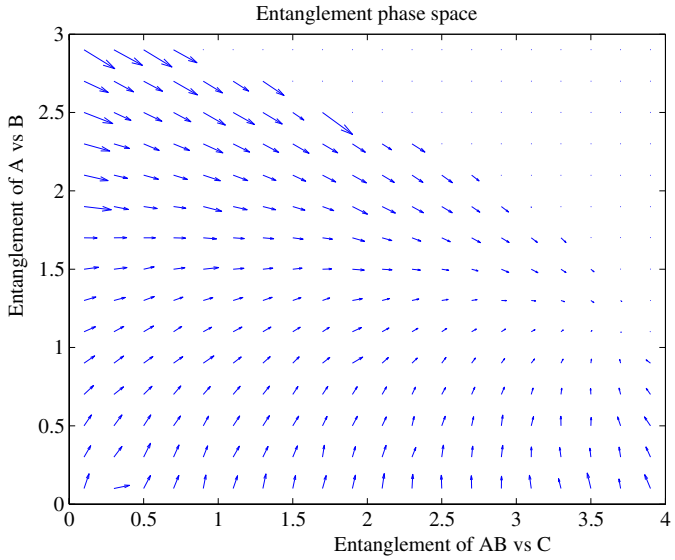

Figure 9. Entanglement 'Phase space' flow diagram. The arrows point in the average flow direction and the length indicates the speed. The entanglement between A and B is quantified as $E_N(\mathrm{tr}_C(\rho_{ABC}))$ and the entanglement between AB and C using $S(\rho_C)$ We see how different initial states will all tend to the attractor point. The horizontal coordinate of this point is given by equation 1. The vertical coordinate is given by $\mathbb{E}[E_N(tr_C(\rho_{ABC}))]$. Here $N = 10$, $N_A = 3$, $N_B = 3$ and $N_C = 4$. The region in the top right corner is empty as such states are not physically realizable.

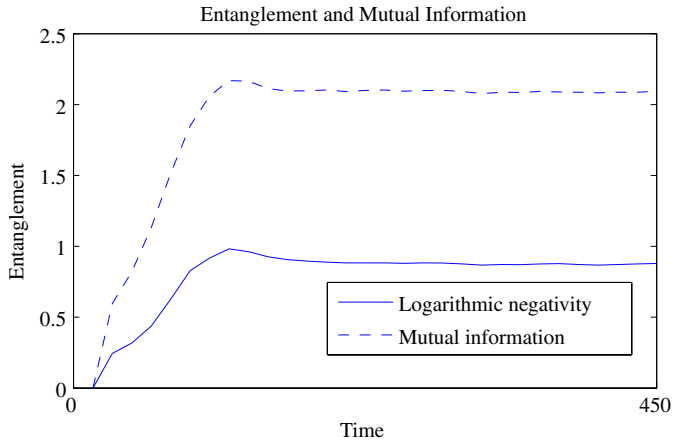

Figure 10. An example of a numerical simulation showing that the average classical and quantum correlations behave similarly. The quantum correlations are quantified using the logarithmic negativity. The combination of classical and quantum correlations is quantified using the mutual information. Here $N = 10$, $N_A = 3$, $N_B = 3$ and $N_C = 4$.

6.1. Simplified multipartite description

Considering generic entanglement only will hopefully provide simplifications in various settings. Here we suggest a way in which it simplifies the multipartite setting. We consider

the tripartite setting. Let N qubits be shared between three parties of size N_A, N_B and N_C, s.t. $N_A + N_B + N_C = N$ and $N_A \leqslant N_B \leqslant N_C$. We can then consider many types of bipartite cuts. Let $E(A|B)$ signify the (unitarily invariant) entanglement average of the entanglement across the $A|B$ cut after C has been traced out, and let $E(A+B|C)$ signify the corresponding quantity across the $A + B|C$ cut. The set Ω of all possible bipartite entanglement averages is then $\Omega = \{E(A|B), E(A|C), E(B|C), E(A+B|C), E(A+C|B), E(B+C|A)\}$. Let these values specify 'the tripartite entanglement'. Now for an arbitrary state, it is not the case that Ω is uniquely determined by $E(A|B)$ together with $E(A+B|C)$. However we make the following conjecture concerning the typical case.

Conjecture. Ω *is uniquely determined by* $E(A|B)$ *together with* $E(A+B|C)$

Support for conjecture. Firstly, note that if N_A, N_B and N_C are specified then Ω is uniquely determined. Then we need the following two statements to be true to prove the claim.

(i) $N_A + N_B$ and N_C are uniquely specified by $E(A+B \mid C)$. This is true if the Von Neumann entropy of entanglement increases monotonically when a qubit is given from the larger of the two parties to the smaller. We are concerned with the limit of large N, where equation (1) implies that $N_C = \lceil E(A+B \mid C) \rceil$ assuming $N_C \leqslant N_A + N_B$.

(ii) For the given N_C and N, N_A and N_B are uniquely determined by $E(A|B)$. Again one should be able to use a monotonicity argument here, although it will be more difficult due to the lack of a neat closed form of $E(A|B) = \mathbb{E}[E_N(\mathrm{tr}_C(\rho_{ABC}))]$. It seems reasonable to expect it to be monotonous when N_C is fixed.

This idea could presumably be extended to more than three parties and may lead to a description that is highly useful by virtue of requiring very few parameters.

7. Discussion and conclusion

The entanglement evolution during random two-qubit interactions was studied in some depth, and we have provided a detailed proof of the result presented in [1] that the average entanglement approaches the unitarily invariant value within $O(N^3)$ steps. The essential idea of the proof is a map from the evolution in state space onto one on a much smaller space that tracks the evolution of the purity of the reduced subsystem. We then prove that the process can be simulated efficiently on a classical computer using stabilizer states.

We found through numerical studies that there are two phases in the approach: first a phase during which entanglement is rapidly spreading through the system, and then a phase where the entanglement is suffused throughout the system. Three moments of time that could be used to define the partition between these phases were introduced and discussed. Firstly the saturation of the average τ_{sat} was considered, followed by the cut off moment τ_{cutoff}. Then we noted that if one restricts the interactions to be between nearest neighbours, the entanglement initially scales as the area of the smaller region and in the second phase as that of the volume. This led us to introduce τ_{vol}, the moment the entanglement is typically volume scaling. Of these perhaps the cutoff moment τ_{cutoff} is the most attractive choice, since it gives an unambiguous single time that corresponds to the moment that the entanglement probability distribution is equal to, for practical purposes, that of the unbiased distribution of pure states.

The results support the relevance of generic entanglement as we show it can be generated efficiently from two-qubit gates. Therefore protocols relying on typical/generic entanglement, like [16, 26, 27, 42] gain relevance [1].

The above results may be extended in various directions that will be described briefly here.

Multi-particle entanglement measures. It would be interesting to extend the above results to further investigate properties of typical multi-partite entanglement. For entanglement measures based on average purities [2] this is possible with the results established here. Other measures such as the entanglement of formation may also be treated but require extensions of the approach presented here that go beyond the scope of this paper.

Markov Process Quantum Monte Carlo Methods. Our results possess another interpretation that is of interest for the numerical study of quantum-many-body systems. One numerical approach to classical spin systems is to evolve spin configurations randomly according to the Metropolis rule, i.e. always accepting moves that decrease energy and accepting moves that increase energy with a probability proportional to $\exp\{-\Delta E/kT\}$. This reproduces the correct thermal average. One may of course consider a similar approach in the quantum setting applying random two-qubit gates to progress the state. Our present results fall into this category but apply for infinite temperatures as we draw our unitaries from the invariant Haar measure. However, we may adapt our analysis to the finite temperature regime employing stabilizer states at the expense of having to analyse a more complicated Markov process. Basic ideas of our approach carry over and open the possibility for rigourous statements concerning the convergence rate of such a Markov process quantum Monte Carlo approach.

Continuous Variables. These questions are made additionally complicated in the continuous variable setting through the lack of a Haar measure for non-compact groups. However in [52] an approach to proceed is introduced. There one notes that it is reasonable to assume that the maximum energy of the global pure state is finite. This tames the non-compactness and one can define a way to pick states at random. References [52, 53] give an explicit method to achieve that for Gaussian states, which probably will take on the role analogous to stabilizer states in the general setting. This opens up the possibility of studying the questions dealt with here in the continuous variable regime.

Two phases. The relation between the three moments of time that can be used to separate the two phases we observe should be further investigated. We believe that analytical tools for studying the cutoff effect are sufficiently developed to undertake an analysis with the aim of proving relations between τ_{vol}, τ_{cutoff} and τ_{sat}, respectively.

Relation to other work. It would be interesting to investigate how our results relate to existing work on spin-gases [54] which are similarly semi-quantal systems. We also hope to relate this line of enquiry to work touching on what the typical entanglement of the universe is [55, 56]. The main difference to the approach here would be to consider closed systems.

Experimental studies. Optical lattices appear to provide a suitable experimental setting to test the results obtained here. There is also hope of using Bose–Einstein condensates or linear optics to study the continuous variable case.

Acknowledgments

We gratefully acknowledge initial discussion with Jonathan Oppenheim as well as discussions with Koenraad Audenaert, Fernando Brandao, Benoit Darquie, Jens Eisert, David Gross, Aram Harrow, Konrad Kieling, Terry Rudolph, Alessio Serafini, Graeme Smith, John Smolin, and Andreas Winter. We acknowledge support for MBP by the EPSRC QIP-IRC, the EU Integrated Project QAP and the Royal Society, for RO by the NSA, the ARDA through ARO contract number W911NF-04-C-0098, and for OCOD by the Institute for Mathematical Sciences at Imperial College London.

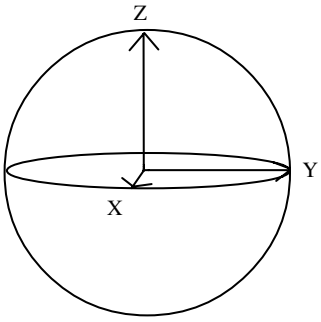

Figure A1. Bloch sphere: the unitarily invariant distribution corresponds to a uniform distribution over this sphere.

Appendix A. Uniform (Haar) measure on the unitary group

We here briefly introduce the often used uniform measure on pure states, sometimes called the unitarily invariant measure. This is a particular instance of a Haar measure, which can be viewed as a generalization of the 'flat distribution'. A flat probability distribution is the one reflecting no bias with respect to any object.

Such a distribution on a set of objects is invariant under permutations of the said objects. Therefore, if there is a group of transformations associated with the set of elements, the distribution on those elements should be invariant under application of elements of the group. The simplest non-trivial example is probably a coin. Here there are two group elements, the identity and the flip. Only the unbiased probability distribution P(Head) = 1/2 and P(Tail) = 1/2 is invariant under those transformations.

Bearing this in mind, consider picking pure general states at random. The unbiased distribution on pure states is here required to be invariant under unitary transforms, i.e. $P(|\Psi\rangle) = P(U|\Psi\rangle)$, where it is implicit that we are in a continuous setting. This requirement uniquely defines the distribution. For a single qubit this can be nicely visualized as a uniformly dense distribution on the Bloch sphere, see figure A1.

The normal method to pick pure states from this distribution is to fix an arbitrary pure state and apply a unitary picked at random from the associated measure on unitary matrices. See for example [51] for the explicit procedure.

Appendix B. Randomizing properties of Haar measure

The following lemma shows that requirement 1 is satisfied by our example with Haar measure.

Lemma 5. *Suppose $F = F(A, B)$ is a bilinear function of one-qubit operators A, B, that σ_a, σ_b are two Pauli matrices, and that I is randomly drawn from Haar measure on $U(2)$. Then*

$$\mathbb{E}[F(T\sigma_a T^\dagger, T\sigma_b T^\dagger)] = \begin{cases} F(I, I), & a = b = 0; \\ \frac{1}{3}\sum_{w\in\{x,y,z\}} F(\sigma_w, \sigma_w), & a = b \neq 0 \\ 0, & \text{otherwise.} \end{cases}$$

To prove it, we need an intermediate result.

Lemma 6. *Assume that σ_a, $a \neq 0$, is a Pauli operator and T is a random unitary drawn from Haar measure on $U(2)$. Then*

$$T\sigma_a T^\dagger = r_x\sigma_x + r_y\sigma_y + r_z\sigma_z,$$

where (r_x, r_y, r_z) is a random vector whose distribution is invariant under permutations. Moreover, $\mathbb{E}[r_u r_{u'}] = \delta_{u,u'}/3$ for all $u, u' \in \{x, y, z\}$.

Proof. $T\sigma_a T^\dagger$ is a 2×2 Hermitian matrix. Therefore, there exist real numbers (in fact random variables) $r_u = \mathrm{Tr}(\sigma_u T\sigma_a T^\dagger)/2$ ($u \in \{0, x, y, z\}$) such that

$$T\sigma_w T^\dagger = r_0 I + r_x\sigma_x + r_y\sigma_y + r_z\sigma_z.$$

$\mathrm{Tr}\,(T\sigma_a T^\dagger) = \mathrm{Tr}(\sigma_a) = 0$ implies $r_0 = 0$. Before we continue, let us state a simple fact we will use repeatedly.

Claim (conjugation trick)**.** *For any unitary $Q \in U(2)$, $T\sigma_a T^\dagger$ and $QT\sigma_a T^\dagger Q^\dagger = QT\sigma_a(QT)^\dagger$ have the same probability distribution. Therefore, the distribution of (r_x, r_y, r_z) is the same as that of (r'_x, r'_y, r'_z), where*

$$r'_u = \tfrac{1}{2}\mathrm{Tr}(\sigma_u QT\sigma_w T^\dagger Q^\dagger) = \tfrac{1}{2}\mathrm{Tr}((Q\sigma_u Q^\dagger)T\sigma_w T^\dagger). \tag{B.1}$$

In fact, this follows directly from the invariance property of Haar measure. Since T and QT have the same probability distribution, so do $T\sigma_a T^\dagger$ and $QT\sigma_a(QT)^\dagger$ (which are the same deterministic function of Q and QT, respectively) and the same holds for (r_x, r_y, r_z) is the same as that of (r'_x, r'_y, r'_z).

We now apply the trick as follows.

(i) Take $Q = H$ (the Hadamard matrix). Then $Q\sigma_x Q^\dagger = \sigma_z$, $Q\sigma_z Q^\dagger = \sigma_x$ and $Q\sigma_y Q^\dagger = \sigma_y$. This implies that $2r'_x = \mathrm{Tr}((Q\sigma_x Q^\dagger)T\sigma_a T^\dagger) = \mathrm{Tr}(\sigma_y T\sigma_w T^\dagger) = 2r_y$, $2r'_y = \mathrm{Tr}((Q\sigma_y Q^\dagger)T\sigma_a T^\dagger) = \mathrm{Tr}(\sigma_z T\sigma_a T^\dagger) = 2r_x$ and $r'_z = r_z$. Hence $(r'_x, r'_y, r'_z) = (r_z, r_y, r_x)$ and (r_x, r_y, r_z) have the same distribution.

(ii) Now take $Q = |0\rangle\langle 0| - \mathrm{i}|1\rangle\langle 1|$. $Q\sigma_x Q^\dagger = \sigma_y$, $Q\sigma_y Q^\dagger = \sigma_x$ and $Q\sigma_z Q^\dagger = \sigma_z$. It follows from the above reasoning that (r_y, r_x, r_z) and (r_x, r_y, r_z) have the same distribution.

(iii) Take $Q = \sigma_z$ this time. Then $Q\sigma_x Q^\dagger = -\sigma_x$, $Q\sigma_y Q^\dagger = -\sigma_y$, $Q\sigma_z Q^\dagger = \sigma_z$, so $(-r_x, -r_y, r_z)$ and (r_x, r_y, r_z) have the same distribution. Similarly, we can take $Q = \sigma_y$ or $Q = \sigma_x$ to show that $(-r_x, r_y, -r_z)$ and $(r_x, -r_y, -r_z)$ also have the same distribution.

The first two items show that the distribution of (r_x, r_y, r_z) is invariant by transposition of the x and z coordinates and of the x and y coordinates. It follows that the distribution is also invariant under transposition of the y and z coordinates (which is a composition of a xz transposition followed by a xy transposition and another xz transposition). Since any permutation is a composition of transpositions, we have shown that the distribution of (r_x, r_y, r_z) is invariant under permutations of the coordinates. Moreover, it also follows that

$$1 = \frac{\mathrm{Tr}\left(\sigma_w^2\right)}{2} = \frac{1}{2}\mathbb{E}[\mathrm{Tr}[(T\sigma_w T^\dagger)^2]] = \mathbb{E}\left[r_x^2 + r_y^2 + r_z^2\right] = 3\mathbb{E}\left[r_u^2\right] \quad \text{for all} \quad u \in \{x, y, z\}.$$

Thus it only remains to show that $\mathbb{E}[r_u r_{u'}] = 0$ if $u \neq u'$. To this end, we use item 3. If for instance $u = x, u' = z$ we recall that $(-r_x, -r_y, r_z)$ and (r_x, r_y, r_z) have the same distribution, hence $r_x r_z$ and $-r_x r_z$ also have the same distribution, which implies our claim. The other cases follow similarly.

We can now prove lemma 5.

Proof. First assume that $a \neq b$. Without loss of generality, assume that $b \neq 0$. We apply an idea based on the conjugation trick from the previous proof. There exists $c \in \{x, y, z\}$ such that σ_a and σ_c commute and σ_b anti-commutes with σ_c. As U and $U\sigma_c$ have the same distribution, $(U\sigma_c)\sigma_a(U\sigma_c)^\dagger = U\sigma_a U^\dagger$ and $(U\sigma_c)\sigma_b(U\sigma_c)^\dagger = -U\sigma_b U^\dagger$, we have

$$\begin{aligned}\mathbb{E}[F(U\sigma_a U^\dagger, U\sigma_b U^\dagger)] &= \mathbb{E}[F((U\sigma_c)\sigma_a(U\sigma_c)^\dagger, (U\sigma_c)\sigma_b(U\sigma_c)^\dagger)] \\ &= -\mathbb{E}[F(U\sigma_a U^\dagger, U\sigma_b U^\dagger)],\end{aligned}$$

thus the expected value is 0. Now if $a = b = 0$, the result is trivial. If $a = b \neq 0$, we can write $U\sigma_a U^\dagger = r_x\sigma_x + r_y\sigma_y + r_z\sigma_z$ as in lemma 6, and by bilinearity

$$\mathbb{E}[F(U\sigma_a U^\dagger, U\sigma_a U^\dagger)] = \sum_{w', w''} \mathbb{E}[r_{w'} r_{w''}] F(\sigma_{w'}, \sigma_{w''}). \tag{B.2}$$

Applying lemma 6 finishes the proof. □

Appendix C. Stabilizer states

Stabilizer states are a discrete subset of general quantum states, which can be described by a number of parameters scaling polynomially with the number of qubits in the state [8, 29, 30].

A *stabilizer operator* on N qubits is a tensor product of operators taken from the set of Pauli operators

$$\sigma_1 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_2 := \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \qquad \sigma_3 := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{C.1}$$

and the identity I. An example for $N = 3$ would be the operator $g = \sigma_1 \otimes I \otimes \sigma_3$. A set $G = \{g_1, \ldots, g_K\}$ of K mutually commuting stabilizer operators that are independent, i.e. $\prod_{i=1}^K g_i^{s_i} = I$ exactly if all s_i are even, is called a *generator set*. For $K = N$, a generator set G uniquely determines a single state $|\psi\rangle$ that satisfies $g_k|\psi\rangle = |\psi\rangle$ for all $k = 1, \ldots, N$. Such a generating set generates the stabilizer group. Each unique such group in turn defines a unique *stabilizer state*.

For example the GHZ state $|000\rangle + |111\rangle$ is defined by the generator set $\langle \sigma_1 \otimes \sigma_1 \otimes \sigma_1, \sigma_0 \otimes \sigma_3 \otimes \sigma_3, \sigma_3 \otimes \sigma_3 \otimes \sigma_0 \rangle$.

A key observation that is useful for the considerations here is the fact that the bipartite entanglement of a stabilizer state, i.e. the entanglement across any bipartite split, takes only integer values [32, 33].

Finally, we note the fact that in order for the stabilizer state to be non-trivial it is necessary and sufficient that the elements of the stabilizer group (a) commute, and (b) are not equal to $-I$ [8].

References

[1] Oliveira R, Dahlsten O C O and Plenio M B 2007 Efficient generation of generic entanglement *Phys. Rev. Lett.* **98** 130502

[2] Plenio M B and Virmani S 2007 An introduction to entanglement measures *Quantum Inf. Comput.* **7** 1–51 (*Preprint* quant-ph/0504163)

[3] Horodecki M 2001 Entanglement measures *Quantum Inf. Comput.* **1,1** 3–26

[4] Wootters W 2001 Entanglement of formation and concurrence *Quantum Inf. Comput.* **1,1** 27–44

[5] Horodecki P and Horodecki R 2001 Distillation and bound entanglement *Quantum. Inf. Comput.* **1,1** 45–75

[6] Plenio M B and Vedral V 1998 Entanglement in quantum information theory *Contemp. Phys.* **39** 431–46

[7] Eisert J and Plenio M B 2003 Introduction to the basics of entanglement theory of continuous-variable systems *Int. J. Quantum Inf.* **1** 479–506

[8] Nielsen M and Chuang I 2000 *Quantum Information and Computation* (Cambridge: Cambridge University Press)
[9] Eisert J and Gross D 2005 *Preprint* quant-ph/0505149
[10] Bennett C H, Popescu S, Rohrlich D, Smolin J A and Thapliyal A V 2001 Exact and asymptotic measures of multipartite pure state entanglement *Phys. Rev.* A **63** 012307
[11] Linden N, Popescu S, Schumacher B and Westmoreland M 1999 Reversibility of local transformations of multiparticle entanglement *Preprint* quant-ph/9912039
[12] Galvão E, Plenio M B and Virmani S 2000 Tripartite entanglement and quantum relative entropy *J. Phys. A: Math. Gen.* **33** 8809
[13] Wu S and Zhang Y 2001 Multipartite pure-state entanglement and the generalized Greenberger–Horne–Zeilinger states *Phys. Rev.* A **63** 012308
[14] Ishizaka S 2004 Bound entanglement provides convertibility of pure entangled states *Phys. Rev. Lett.* **93** 190501
[15] Ishizaka S and Plenio M B 2005 Multiparticle entanglement manipulation under positive partial transpose preserving operations *Phys. Rev.* A **71** 052303
[16] Hayden P, Leung D W and Winter A 2006 Aspects of generic entanglement *Commun. Math. Phys.* **265** 95
[17] Lubkin E 1978 Entropy of an n-system from its correlation with a k-reservoir *J. Math. Phys.* **19** 1028–31
[18] Lloyd S and Pagels H 1988 Complexity as thermodynamic depth *Ann. Phys.* **188** 186–213
[19] Page D N 1993 Average entropy of a subsystem *Phys. Rev. Lett.* **71** 1291–4
[20] Foong S K and Kanno S 1994 Proof of Page's conjecture on the average entropy of a subsystem *Phys. Rev. Lett.* **72** 1148–51
[21] Emerson J, Livine E and Lloyd S 2003 Random circuits and pseudo-Random unitary operators for quantum information processing *Science* **302** 2098
[22] Emerson J, Livine E and Lloyd S 2005 Convergence conditions for random quantum circuits *Phys. Rev.* A **72** 060302
[23] Emerson J 2004 Random circuits and pseudo-Random unitary operators for quantum information processing *QCMC04 (AIP Conf. Proc.* vol 734*)* p 139
[24] Smith G and Leung D W 2006 Typical entanglement of stabilizer states *Phys. Rev.* A **74** 062314
[25] Aldous D J and Fill J *Reversible Markov Chains and Random Walks on Graphs* at press, http://statwww. berkeley. edu/users/aldous/RWG/book.html
[26] Buhrman H, Christandl M, Hayden P, Lo H K and Wehner S 2005 On the (im)possibility of quantum string commitment (*Preprint* quant-ph/0504078)
[27] Harrow A, Hayden P and Leung D 2004 Superdense coding of quantum states *Phys. Rev. Lett.* **92** 187901
[28] DiVincenzo D P, Leung D W and Terhal B M 2002 Quantum data hiding *IEEE Trans. Inf. Theory* **48** 580598
[29] Gottesman D 1997 Stabilizer codes and quantum error correction *PhD Thesis* Caltech
[30] Gottesman D 1998 The Heisenberg representation of quantum computers *Preprint* quant-ph/9807006
[31] Dahlsten O C O and Plenio M B 2006 Entanglement probability distribution of bi-partite randomised stabilizer states *Quantum Inf. Comput.* **6** 527–38
[32] Audenaert K M R and Plenio M B 2005 Entanglement on mixed stabilizer states: normal forms and reduction procedures *New J. Phys.* **7** 170
[33] Fattal D, Cubitt T S, Yamamoto Y, Bravyi S and Chuang I L 2004 Entanglement in the stabilizer formalism *Preprint* quant-ph/0406168
[34] Hein M, Eisert J and Briegel H J 2004 Multiparty entanglement in graph states *Phys. Rev.* A **69** 062311
[35] Anders S, Plenio M B, Dür W, Verstraete F and Briegel H-J 2006 Ground state approximation for strongly interacting systems in arbitrary dimension *Phys. Rev. Lett.* **97** 107206
[36] Dankert C, Cleve R, Emerson J and Livine E 2006 Exact and approximate unitary 2-designs: constructions and applications *Preprint* quant-ph/0606161
[37] Gross D, Audenaert K and Eisert J 2007 Evenly distributed unitaries: on the structure of unitary designs *J. Math. Phys.* **48** 052104
[38] Emerson J, Alicki R and Zyczkowski K 2005 Scalable noise estimation with random unitary operators *J. Opt. B: Quantum Semiclass. Opt.* **7** S347
[39] Diaconis P 1996 The cutoff phenomenon in finite Markov chains *Proc. Natl Acad. Sci. USA* **93** 1659–64
[40] Diaconis P 1988 *Group Representations in Probability and Statistics (Lecture Notes Monograph Series* vol 11*)* (London: Institute of Mathematical Statistics)
[41] Montenegro R and Tetali P 2006 Mathematical aspects of mixing times in Markov chains *Series Foundations and Trends in Theoretical Computer Science* vol 1 ed M Sudan (Boston-Delft: NOW Publishers) http://www. ravimontenegro.com/research/TCS008-journal.pdf

[42] Abeyesinghe A, Hayden P and Smith G 2006 Optimal superdense coding of entangled states *IEEE Trans. Inf. Theory* **52** 3635–41
[43] Audenaert K, Eisert J, Plenio M B and Werner R F 2002 Entanglement properties of the harmonic chain *Phys. Rev.* A **66** 042327
[44] Plenio M B, Eisert J, Dreissig J and Cramer M 2005 Entropy, entanglement, and area: analytical results for harmonic lattice systems *Phys. Rev. Lett.* **94** 060503
[45] Cramer M, Eisert J, Plenio M B and Dreissig J 2006 An entanglement-area law for general bosonic harmonic lattice systems *Phys. Rev.* A **73** 012309
[46] Keating J P and Mezzadri F 2004 Random matrix theory and entanglement in quantum spin chains *Commun. Math. Phys.* **252** 543–79
[47] Vidal G and Werner R F 2002 A computable measure of entanglement *Phys. Rev.* A **65** 32314
[48] Plenio M B 2005 Logarithmic negativity: a full entanglement monotone that is not convex *Phys. Rev. Lett.* **95** 090503
[49] Eisert J 2001 *PhD Thesis* Universität Potsdam
[50] Groisman B, Popescu S and Winter A 2005 On the quantum, classical and total amount of correlations in a quantum state *Phys. Rev.* A **72** 032317
[51] Diaconis P 2005 What is a random matrix *Not. Am. Math. Soc.* **52** 11
[52] Serafini A, Dahlsten O C O and Plenio M B 2007 Teleportation fidelities of squeezed states from thermodynamical state space measures *Phys. Rev. Lett.* **98** 170501
[53] Serafini A, Dahlsten O C O, Gross D and Plenio M B 2007 Canonical and micro-canonical typical entanglement of continuous variable systems (*Preprint* quant-ph/0701051)
[54] Calsamiglia J, Hartmann L, Dür W and Briegel H-J 2005 Entanglement and decoherence in spin gases *Phys. Rev. Lett.* **95** 180502
[55] Gemmer J, Otte A and Mahler G 2001 Quantum approach to a derivation of the second law of thermodynamics *Phys. Rev. Lett.* **86** 1927
[56] Popescu S, Short A and Winter A 2006 The foundations of statistical mechanics from entanglement: Individual states versus averages *Nature Phys.* **2** 754