Trabajo de grado

(Proyecto de investigación)

Facultad de Ciencias Naturales y Exactas

Departamento de Física (3146)

Universidad del Valle

Dependencia de la arquitectura en la estructura de entrelazamiento de muchos cuerpos en circuitos cuánticos locales

Architecture Dependence in the Many-Body Entanglement Structure of Local Quantum Circuits

By:
Andres Felipe Valencia Fonseca

Advisor:
Prof. Javier Madroñero Pabón

Co-advisor:
William Esteban Salazar

Santiago de Cali, diciembre 01 del 2025

Agradecimientos

En primer lugar, doy gracias a Dios por haberme permitido, aún bajo toda circunstancia, salir triunfante sin yo merecerlo. A mi padre, le doy gracias por ser quien me enseña a vivir constantemente, y por impulsar mis sueños, haciéndome entender que no hay límite para quienes confían en Dios. A mi madre, le doy gracias por enseñarme el amor y la perseverancia, de quien aprendí la responsabilidad y comprendí que el esfuerzo trae frutos. A ambos, les agradezco por su educación, sus consejos y sus oraciones, que me han mantenido en pie. A mi hermano Juan, le agradezco por enseñarme la generosidad, a pensar en grande y por inspirarme con su determinación para alcanzar propósitos; gracias por su cuidado y apoyo. A mi hermana Sofía, por motivarme a ser un mejor ejemplo cada día, por su amor y abrazos que me animan a avanzar. A toda mi familia en general, y también a mi abuela Isabel, quien ya no está con nosotros, les agradezco por su amor y por ser mi motivación.

Agradezco a mis amigos de infancia Diego y Harold, quienes me acompañaron durante toda la universidad en Cali, viviendo juntos y compartiendo noches de risas. Debo decir que sin ellos, toda mi estancia en Cali mientras estudiaba habría sido más difícil; me ayudaron cuando estuve enfermo y cuando escasamente tenía tiempo para dormir ellos se preocuparon por mí (a Diego doy disculpas por las tantas noches en que me pasaba estudiando con la luz prendida).

También quiero agradecer a los hermanos Migue y Gabriel, por ser amistades fundamentales en mi vocación por las ciencias, y en especial a Miguel, por motivar mi curiosidad a través de libros y nuestras largas y profundas charlas durante el colegio. Agradezco a mi profesor Carlos William del colegio, quien me alentó y me brindó confianza para iniciar la carrera. A mis amigos de cohorte, Nicolás y Kevin, agradezco por la amistad que hemos cultivado a lo largo de los semestres; estudiar juntos nunca hubiera sido mejor; con ellos disfruté y aprendí durante toda la carrera. También agradezco a todos mis amigos de Quantic —Scarpetta, Jeison, Rafael, Anita, Valen, Krysti, Carlos y Esteban— porque reímos, jugamos, estudiamos, y cada momento en la oficina hacía más ameno todo.

A mi director, Javier, muchas gracias por integrarme al grupo de investigación desde tan temprano. Ha sido una referencia fundamental, de quien admiro no solo su excelencia como físico, sino también su compromiso y calidad humana. A Mateo, por su dedicación con nosotros, Nicolás y Kevin, en el semillero, introduciéndonos al grupo. Finalmente a William, que más que mi codirector se convirtió en un amigo, quiero dedicarle un agradecimiento especial, ya que a él le debo, en gran medida, la culminación de este trabajo y buena parte de mi formación en el área; le reconozco profundamente el tiempo y el conocimiento que me brindó para estudiar y desarrollar este proyecto. Es sin duda alguna un ejemplo admirable a seguir.

Sin más que decir, quiero cerrar citando las palabras de *Jeremías (33:3)*:

קְרָא אֵלַי וְאֶעֱנֶךָּ וְאַגִּידָה לְךָ גְּדֹלוֹת וּבְצָרוֹת לֹא יְדַעְתָּם: פ

Contents

Introduction 1

1 Bipartite entanglement in random quantum circuits 2

1.1 Quantum entanglement . . . 2

1.2 Entanglement entropy in bipartite pure states . . . 4

1.2.1 Rényi entropy . . . 4

1.2.2 Properties of Rényi entropy . . . 5

1.3 Schmidt decomposition . . . 7

1.4 Random quantum circuits . . . 8

1.4.1 Modeling quantum many-body systems . . . 9

1.4.2 Circuit architectures and locality . . . 10

1.4.3 Clifford random evolution . . . 12

2 Minimal cut 16

2.1 Minimal cut formalism . . . 16

2.1.1 KPZ Model . . . 17

2.1.2 Cut growth model . . . 19

2.2 Graph-based algorithm for the minimal cut . . . 25

2.2.1 The ZX-calculus . . . 26

2.2.2 Algorithm design . . . 27

2.2.3 Runtime analysis . . . 28

3 Results and analysis 30

3.1 Numerical results . . . 30

4 Conclusions and perspectives 34

Appendices 35

A Proofs of key properties of Rényi entropy . . . 35

B Minimal cut proof . . . 36

C Derivations for the cut growth model . . . 38

C.1 Cut-increase distribution . . . 38

C.2 Cut number distribution . . . 39

Bibliography 41

Resumen

En este trabajo se estudia la dependencia de la arquitectura en la estructura de entrelazamiento en circuitos cuánticos con compuertas locales de hasta dos qubits, utilizando el formalismo del corte mínimo [1]. En una primera etapa, se desarrolla un modelo de crecimiento de corte para establecer una cota superior clásica en la estimación de las entropías de Rényi en circuitos cuánticos aleatorios de tipo Clifford [2]. Posteriormente, se desarrolla un algoritmo computacional original, basado en grafos y propuesto por el autor, para calcular de manera eficiente los cortes mínimos en circuitos cuánticos locales; además, se realizan simulaciones numéricas que permiten contrastar estos resultados con aquellos obtenidos mediante el cálculo directo de las entropías de Rényi. Se observa una correspondencia numérica consistente entre ambos enfoques, lo que valida el corte mínimo como un método eficiente para caracterizar la dinámica de entrelazamiento y recuperar las fluctuaciones características de la clase de universalidad Kardar–Parisi–Zhang (KPZ).

Los resultados obtenidos muestran que la dependencia de la arquitectura se refleja en la escala de saturación del entrelazamiento, la cual se alcanza para circuitos cuya área crece proporcionalmente al cubo del número de qubits. De este modo se establece una ley de escalamiento que vincula la geometría del circuito con la eficiencia en la generación de entrelazamiento. En conjunto, el estudio aquí presentado contribuye a una mejor comprensión del comportamiento genérico del entrelazamiento en sistemas cuánticos de muchos cuerpos.

Palabras claves: Computación cuántica, Información cuántica, Corte mínimo, Entrelazamiento, Circuitos aleatorios.

Abstract

This work studies the architecture dependence on the entanglement structure in quantum circuits with at most two-qubit local gates, using the minimal cut formalism [1]. In the first stage, a cut growth model is developed to establish a classical upper bound for the estimation of Rényi entropies in random Clifford circuits [2]. Subsequently, an original graph-based computational algorithm proposed by the author is developed to efficiently compute minimal cuts in local quantum circuits, and numerical simulations are carried out to compare these results with those obtained through the direct calculation of Rényi entropies. A consistent numerical correspondence is observed between both approaches, which validates the minimal cut as an efficient method for characterizing entanglement dynamics and for recovering the characteristic fluctuations of the Kardar–Parisi–Zhang (KPZ) universality class.

The results show that the dependence on architecture is reflected in the saturation scale of entanglement, which is reached in circuits whose area grows proportionally to the cube of the number of qubits. In this way, a scaling law is established that links the geometry of the circuit with the efficiency of entanglement generation. Overall, the study presented here contributes to a deeper understanding of the generic behavior of entanglement in quantum many-body systems.

Keywords: Quantum computing, Quantum information, Minimal cut, Entanglement, Random

circuits.

Introduction

Undoubtedly, one of the most fascinating properties of quantum systems is quantum entanglement [3]. This property exclusively of composite quantum systems implies the existence of global states that cannot be expressed as a separable product of states of the individual subsystems. Entanglement is a powerful resource in quantum computing, despite the fact that there are still fundamental open questions in entanglement theory, such as, how can entanglement be optimally detected, both theoretically and in the laboratory, and how can it be characterized and quantified [4, 5].

These questions have been approached in various ways [6–9], generally by assigning a metric to quantify entanglement via the system's entropy, known as entanglement entropy [5, 10, 11]. It is known that Rényi entropy generalizes a family of entropies [12], such as the von Neumann entropy and Hartley entropy [13, 14]. Computing the Rényi entanglement entropy in a quantum circuit is, in general, a non-trivial task. As a result, approximation techniques have been introduced, in particular we focus on the minimal cut formalism [1, 15, 16], which provides a classical bound for the Rényi entropy based on the geometry of the quantum circuit. However, the correlation between a circuit's entanglement and its architectural layout remains unclear.

For this reason, the present work aims to study the dependence of entanglement on the architecture of quantum circuits, using the minimal cut technique. To draw general conclusions, we focus on random circuits [2] composed of gates acting on up to two qubits. The goal is to extrapolate the behavior of such systems to the thermodynamic limit, where determining the exact degree of entanglement either numerically or experimentally becomes extremely challenging.

In this thesis, Chapter 1 introduces the fundamental concepts of bipartite entanglement in random quantum circuits, including definitions, entropic measures, and the role of circuit architectures. Chapter 2 develops the minimal cut formalism through two complementary approaches: the cut growth model, a Markov-chain–inspired framework that allows one to infer the evolution of the minimal cut, and a graph-based algorithm that maps the circuit geometry to a weighted graph to compute minimal cut efficiently. In chapter 3, we benchmark both approaches by testing them against the predictions made by the KPZ universality class. Additionally, we explore the extent of our cut growth model by comparing the predictions made on entanglement Rényi entropies with the exact numerical simulations on Clifford circuits. Finally, the appendices provide detailed proofs and derivations that support the results presented in the main text.

Chapter 1

Bipartite entanglement in random quantum circuits

1.1 Quantum entanglement

Quantum entanglement[1], in essence, refers to the profound correlation between two or more quantum systems, in which the state of one is intrinsically linked to the state of the other, regardless of the spatial separation between them. This implies that measuring a certain property of one entangled system reveals the corresponding property of the other.

Within the study of quantum mechanics, formally we have that for multipartite systems, the total Hilbert space $\mathcal{H}$ is constructed as the tensor product of each of the n subsystems, *i.e.* $\mathcal{H} = \otimes_{i=1}^{n} \mathcal{H}_i$. A pure state $|\psi\rangle$ in $\mathcal{H}$ is written as a superposition of basis states of $\mathcal{H}$, and in general, it cannot be expressed as a separable product of individual states for each of the subsystems [5], *i.e.*

$$|\psi\rangle = \sum_{i_1,\ldots,i_n} c_{i_1,\ldots,i_n} |i_1\rangle \otimes \cdots \otimes |i_n\rangle \neq |\psi_1\rangle \otimes \cdots \otimes |\psi_n\rangle \,. \tag{1.1}$$

This observation indicates that, in general, it is not possible to assign a single state vector to any of the n subsystems; this phenomenon is interpreted as quantum entanglement. This leads to the following definition.

Definition 1.1 (Quantum entanglement for pure states) *A pure state $|\psi\rangle$ in a multipartite system is said to be entangled if it cannot be written as a product of states in each subsystem.*

The definition of entanglement extends naturally to mixed states [10], and within the density matrix formalism, a mixed state is said to be entangled if it cannot be written as a convex combination of product states, *i.e.*

[1] Initially, Einstein et al. [3], who introduced this phenomenon, referred to such a connection as "spooky action at a distance" and regarded it as counterintuitive. Later on, although the phenomenon had already been observed and acknowledged, it was Schrödinger who coined the term *Verschränkung* (translated as *entanglement* in English) to describe this property of quantum systems. He contributed significantly to its theoretical formulation, characterizing it as the most distinctive feature of quantum mechanics.

Definition 1.2 (Quantum entanglement for mixed states) *Let ρ be a density operator describing a multipartite quantum system composed of subsystems $A_1, A_2, \ldots, A_n$. The state ρ is said to be* separable *(*i.e.*, not entangled) if it can be written as a convex combination of product states:*

$$\rho = \sum_i p_i \, \rho_{A_1}^{(i)} \otimes \rho_{A_2}^{(i)} \otimes \cdots \otimes \rho_{A_n}^{(i)}, \quad \textit{with } p_i \geq 0, \sum_i p_i = 1. \tag{1.2}$$

If no such decomposition exists, the state ρ is said to be entangled.

Quantum entanglement reveals that certain systems, when composed in specific ways, cease to be described solely by their individual components and instead behave as an inseparable whole. This foundational idea has led to significant scientific and technological advances, including quantum cryptography [17], quantum teleportation [18], and entanglement-based metrology [19]. Most notably, in the current era, entanglement plays a crucial role in quantum computing [20], enabling protocols and algorithms that fundamentally outperform their classical counterparts.

Despite its usefulness, entanglement introduces significant theoretical and practical challenges. One such challenge is the task of identifying whether a given quantum system is entangled and, if so, determining how much entanglement it contains. Based on the definition of entanglement, this task would seem to reduce to checking whether the state can be written as a product of subsystem states.

However, this separability problem becomes computationally intractable for large systems with many degrees of freedom. A closer classical analogue is integer factorization: while no efficient general algorithm is known, there exist special cases where the problem can be efficiently resolved. Similarly, although separability is NP-hard in general [21], certain families of states (*e.g.*, stabilizer states [22]) admit efficient entanglement witnessing.

In pursuit of a more practical characterization of entanglement, John von Neumann, inspired by Shannon's notion of entropy in classical information theory [23], developed a quantum analogue to quantify information in quantum systems. In classical information theory, **Shannon entropy** H quantifies the average information content or surprise of a message and is given by [23]:

$$H(X) = -\sum_i p_i \log_2 p_i, \tag{1.3}$$

where p_i is the probability of the i-th outcome. The higher the entropy, the more information is encoded in the distribution, reflecting greater uncertainty or variability.

Von Neumann extended this idea to quantum systems by defining the **von Neumann entropy** for a quantum state described by a density matrix ρ as:

$$S(\rho) = -\mathrm{Tr}(\rho \log_2 \rho). \tag{1.4}$$

Entanglement naturally emerges in this framework, as it is the entangled states that exhibit the highest von Neumann entropy when one considers the reduced density matrix of a subsystem. This makes intuitive sense: in an entangled system, subsystems contain less information individually, as the total information is encoded nonlocally across the whole system. Maximally entangled states, such as the Bell states for two qubits, and their multipartite generalizations like

GHZ (Greenberger–Horne–Zeilinger) states [20], maximize the entropy of the reduced density matrix in the appropriate bipartition. For instance, for the Bell state

$$|\Phi^+\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}, \tag{1.5}$$

tracing out one qubit yields a maximally mixed state $\rho_A = \frac{I}{2}$, with von Neumann entropy $S(\rho_A) = 1$, the maximum possible for a single qubit. Similarly, GHZ states exhibit maximal entropy when reduced to the bipartition of one qubit versus the remaining $N-1$ qubits [20]. In contrast, separable states yield zero entropy upon reduction, indicating the absence of entanglement. Between these two extremes lies a continuous spectrum of entanglement, where the von Neumann entropy serves as a quantitative measure of how entangled a state is.

1.2 Entanglement entropy in bipartite pure states

Entanglement entropy is introduced as a fundamental tool for quantifying the degree of entanglement in quantum systems and in general as a measure of quantum correlations[2]. However, this definition is not unique, multiple entanglement measures exist apart from von Neumann Eq. (1.4), each based on the density matrix formalism and calibrated according to well-established reference points: maximally entangled states and separable states. These metrics provide complementary insights into different operational or theoretical aspects of entanglement.

To illustrate this, consider the **Hartley entropy** [12], which serves as a simple entanglement measure for pure states. It is defined as:

$$S_0(\rho_A) = \log_2(\mathrm{rank}(\rho_A)), \tag{1.6}$$

where $\rho_A = \mathrm{Tr}_B(\rho_{AB})$ is the reduced density matrix of subsystem A. The Hartley entropy counts the number of nonzero eigenvalues of the reduced density matrix, providing a rough estimate of how many orthogonal states are involved in the entanglement. Although it lacks sensitivity to the distribution of these eigenvalues, it is useful for detecting whether a pure bipartite state is entangled or not.

A natural generalization of both the Hartley entropy Eq. (1.6) and von Neumann entropy Eq. (1.4) is the **Rényi entropy** [13], which forms a one-parameter family of entanglement measures indexed by a real number $\alpha > 0$, $\alpha \neq 1$. Conceptually, Rényi entropy interpolates between various entropic quantities and allows for a tunable sensitivity to the spectrum of the reduced density matrix.

1.2.1 Rényi entropy

The Rényi entropy S_α [12] is introduced for a bipartite system composed of subsystems A and B, as follows:

[2] In this thesis we restrict our attention to bipartite systems, in particular those generated by Clifford circuits acting on computational basis states, which always yield pure global states, however, upon bipartition, each subsystem is generally described by a mixed state. In this setting, the von Neumann entropy of the reduced density matrix fully characterizes entanglement.

Definition 1.3 (Rényi entropy) *Let ρ_A be the reduced density matrix of a subsystem A obtained by tracing out the complementary subsystem B from a global pure state ρ_{AB},* i.e., *$\rho_A = \mathrm{Tr}_B(\rho_{AB})$. The Rényi entropy of order $\alpha \in \mathbb{R}^+ \setminus \{1\}$ is defined as:*

$$S_\alpha(\rho_A) = \frac{1}{1-\alpha} \log_2 \left(\mathrm{Tr}(\rho_A^\alpha) \right), \tag{1.7}$$

where α is a real parameter that characterizes the family of Rényi entropies.

In terms of the eigenvalues $\{\lambda_i\}$ of ρ_A, this can be equivalently written as:

$$S_\alpha(\rho_A) = \frac{1}{1-\alpha} \log_2 \left(\sum_i \lambda_i^\alpha \right). \tag{1.8}$$

This generalization avoids limiting the analysis to a single entropy definition and enables the study of entanglement under different operational perspectives, such as the distinguishability of quantum states or the scaling behavior in many-body systems [1, 13]. In particular, in the limit $\alpha \to 1^+$, the Rényi entropy recovers the von Neumann entropy, whereas in the limit $\alpha \to 0$, it approaches the Hartley entropy.

Thus, entanglement entropy (1.4) and its generalizations provide a rich and flexible framework for quantifying entanglement in both foundational studies and practical applications such as quantum communication, condensed matter physics, and quantum computing [5, 20, 24].

1.2.2 Properties of Rényi entropy

The Rényi entropy $S_\alpha(\rho_A)$, exhibits several important mathematical properties [25], which are particularly relevant in the analysis of the *minimal cut formalism* later in this work (see Chapter 2). This subsection presents a summary of those properties (see Appendix A for the corresponding derivations).

First, the entropy is bounded:

$$0 \leq S_\alpha(\rho_A) \leq \log_2 D, \tag{1.9}$$

where $D = \min(D_A, D_B)$, with D_A and D_B the Hilbert space dimensions of the respective subsystems. The lower bound is attained when ρ_A is a pure state, *i.e.*, $\rho_A = |\phi\rangle\langle\phi|$, while the upper bound corresponds to a maximally mixed state, *i.e.*, $\rho_A = I_D/D$, where I_D is the identity operator on a D-dimensional Hilbert space.

The case $\alpha = 0$, which corresponds to Hartley entropy Eq. (1.6), is particularly simple in the pure bipartite setting, since it serves mainly as an indicator of the presence or absence of entanglement. It is a discontinuous function: even an infinitesimal interaction between initially unentangled subsystems can cause a sudden jump in S_0 from 0 to $\log_2 D$, where D is the Hilbert space dimension of the subsystem. Thus, S_0 acts as a binary indicator that distinguishes "some entanglement" from "no entanglement" rather than providing a gradual quantification.

Despite its abrupt behavior, S_0 has practical utility, especially when interpreted in classical terms [1]. Furthermore, it plays an important role as an upper bound for all other Rényi entropies (see Fig. 1.1):

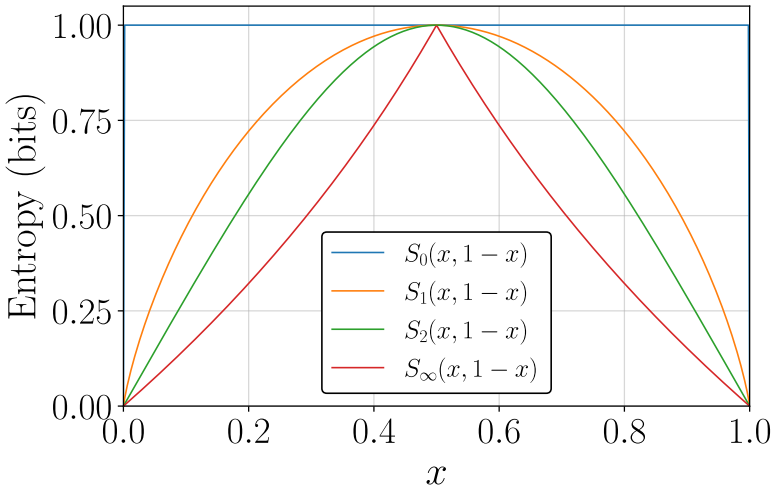


Figure 1.1: Rényi entropies of orders 0, 1, 2, and ∞ for the Bernoulli distribution (binary distribution) $P = (x, 1-x)$ [26]. The horizontal axis denotes the probability x of one outcome, while the vertical axis represents the entropy in bits (logarithm base 2). In the quantum information context, these correspond respectively to the Hartley entropy (S_0), the von Neumann entropy (S_1), the collision entropy (S_2), and the min-entropy (S_∞).

$$S_0(\rho_A) \geq S_\alpha(\rho_A) \geq S_\infty, \quad \text{for all } \alpha \geq 0. \tag{1.10}$$

In the limit $\alpha \to \infty$, the sum in equation (1.8) is dominated by the single largest eigenvalue $\lambda_{\max}$ of ρ_A, and so

$$S_\infty = \log_2(1/\lambda_{\max}) \tag{1.11}$$

This version of the entropy emphasizes the contribution of the most probable eigenvalue of the reduced density matrix. In particular, S_∞ constitutes the lower bound of the family of Rényi entropies.

More precisely, for any $\alpha > 1$, the following inequality holds:

$$S_\infty(\rho_A) \leq S_\alpha(\rho_A) \leq \frac{\alpha}{\alpha - 1} S_\infty(\rho_A). \tag{1.12}$$

Importantly, the inequality above implies that the convergence or boundedness of any Rényi entropy with $\alpha > 1$ determines the behavior of all others in that regime. That is, if $S_\alpha(\rho_A)$ converges for some $\alpha > 1$, then $S_\alpha(\rho_A)$ also converges for all $\alpha > 1$. This justifies the notion of a "generic" behavior for Rényi entropies for $\alpha > 1$, under which the scaling properties of entanglement are robust and largely parameter-independent. While S_1 typically aligns with this generic behavior [15], such alignment is empirical and not a consequence of the bound. By contrast, S_0 can exhibit qualitatively different behavior due to its sensitivity to the rank rather than the spectrum of ρ_A.

Another essential feature of the Rényi entropy is its invariance under local unitary transformations that act exclusively on subsystem A or its complement B [13]. Since these operations do not alter the spectrum of the reduced density matrix ρ_A, the entropy $S_\alpha(\rho_A)$ remains unaffected. This makes it a well-suited measure for entanglement, which should depend only on nonlocal correlations.

These results are key for justifying the use of minimal cut approximations to estimate entanglement entropy in quantum systems with large Hilbert spaces.

1.3 Schmidt decomposition

The Schmidt decomposition is a fundamental result in linear algebra and provides a powerful tool for analyzing entanglement in bipartite quantum systems. In the study of random quantum circuits and the minimal cut formalism [15,16], it offers both an operational and a geometric perspective on entanglement entropy. In particular, it clarifies the role of the Hartley entropy S_0 and its connection to classical bounds such as the minimal cut.

Theorem 1.1 *Let $\mathcal{H}_A$ and $\mathcal{H}_B$ be finite-dimensional Hilbert spaces, with $\dim \mathcal{H}_A = n$ and $\dim \mathcal{H}_B = m$, and assume without loss of generality that $n \geq m$. For any pure state $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$, there exist orthonormal sets $\{|u_i\rangle\} \subset \mathcal{H}_A$ and $\{|v_i\rangle\} \subset \mathcal{H}_B$, and strictly positive real coefficients $\{\lambda_i\}$ such that:*

$$|\psi\rangle = \sum_{i=1}^{r} \sqrt{\lambda_i}\, |u_i\rangle \otimes |v_i\rangle\,, \tag{1.13}$$

where $r \leq m$ is called the Schmidt rank of $|\psi\rangle$. The coefficients $\sqrt{\lambda_i}$ are known as Schmidt coefficients, and are unique up to permutations.

This decomposition is essentially a reformulation of the singular value decomposition of the matrix representation of the state vector. Although its proof can be obtained via standard techniques in linear algebra (see, e.g., [20, 24]), it suffices here to highlight that the Schmidt decomposition reveals that any bipartite pure state can be expressed in a basis that diagonalizes the correlations between subsystems. Furthermore, the number of non-zero Schmidt coefficients provides a measure of entanglement. The state is separable if and only if $r = 1$.

Connection to entanglement entropy

Given the expansion

$$|\psi\rangle = \sum_{i=1}^{r} \sqrt{\lambda_i}\, |u_i\rangle \otimes |v_i\rangle\,, \tag{1.14}$$

we can form the reduced density matrix $\rho_A = \mathrm{Tr}_B\, |\psi\rangle\langle\psi|$, which becomes

$$\rho_A = \sum_{i=1}^{r} \lambda_i\, |u_i\rangle\langle u_i|\,. \tag{1.15}$$

Hence, the eigenvalues of ρ_A are precisely the squared Schmidt coefficients λ_i. This implies that the von Neumann entropy of subsystem A,

$$S_1(\rho_A) = -\sum_{i=1}^{r} \lambda_i \log_2 \lambda_i, \tag{1.16}$$

which vanishes if and only if the state is separable.

Of particular interest in this work is the Hartley entropy Eq. (1.6), which under this decomposition simply counts the number of non-zero Schmidt coefficients, in other words, S_0 reflects the minimal number of terms required in the Schmidt decomposition to represent $|\psi\rangle$ exactly,

$$S_0 = \log_2 \left(\#\text{ of non-zero Schmidt coefficients of } |\psi\rangle\right). \tag{1.17}$$

This motivates the interpretation of S_0 as measuring the "minimal quantum complexity" of the entangled state across a bipartition [15, 16]. For instance, if $|\psi\rangle$ is separable, then the Schmidt decomposition has a single nonzero coefficient, and $S_\alpha = 0$ for all α according to equation (1.8). If $|\psi\rangle$ is maximally entangled, then all $\lambda_i = 1/r$, and $S_\alpha(\rho_A)$ is maximized.

A key observation here is that the Schmidt decomposition provides the most efficient representation of $|\psi\rangle$ in a bipartition: any other decomposition into product states involves at least as many terms. Therefore, Eq. (1.17) implies that

$$S_0 \leq \log_2\big(\#\text{ of non-zero terms in } \textit{any} \text{ decomposition of } |\psi\rangle\big). \tag{1.18}$$

Combining this with inequality (1.10) we have more generally that

$$S_\alpha \leq S_0 \leq \log_2(\#\text{of non-zero terms in } \textit{any} \text{ decomposition of } |\psi\rangle). \tag{1.19}$$

This efficiency becomes particularly useful when analyzing tensor networks or circuit representations of quantum states. The minimal number of terms in the Schmidt decomposition directly bounds the bond dimension required for an exact matrix product representation [15, 16], and the *minimal cut formalism* leverages this connection, we will explore this further in Chapter 2.

1.4 Random quantum circuits

Random quantum circuits emerged in analogy with other applications of randomness, such as random matrix theory [27, 28], motivated by the idea that the entanglement dynamics they generate may exhibit generic universal features. These models form an analytically tractable class of quantum many-body systems, offering valuable insight into nonequilibrium dynamics and the spreading of quantum information [29–31]. Unlike deterministic circuits with fixed gate structures, random circuits are constructed by applying local gates independently drawn from a probability distribution [2]. This stochasticity fosters universality, enabling the development of scaling theories for entanglement growth, statistical analysis over ensembles, and mappings to effective classical models [1, 2, 32, 33].

Random circuits generate entanglement whose behavior can be captured by an effective minimal membrane theory [34]. In this picture, the entanglement entropy $S_A(t)$ of a region A in $d+1$

spatial dimensions is determined by minimizing an effective "free energy", or entanglement cost, for a d-dimensional membrane [35, 36].

For a given realization at early times, the membrane spans the full time interval, leading to linear growth of entropy in time until saturating,

$$S_A(t) \propto \partial A\, t, \tag{1.20}$$

where ∂A is the size of the boundary between region A and its complement divided by the membrane. At later times, the membrane transitions to a time-independent configuration whose cost reproduces the thermal entropy,

$$S_A(\infty) \propto V_A, \tag{1.21}$$

reaching a saturation proportional to the volume (V_A) of A [37, 38]. As in this work we are interested in one dimensional systems of qubits evolving through time, *i.e.* a $(1+1)d$ dynamics, the volume V_A corresponds to the total number of sites (qubits) in region A.

Moreover, this membrane picture not only captures the mean entanglement growth, but also provides analytic control over its fluctuations. For the $(1+1)d$ dimension considered in this work, these fluctuations are governed by the KPZ universality class [39], as it is discussed in greater detail in Section 2.1.1.

Finally, random circuits have been generalized to incorporate additional structures such as global symmetries, Floquet driving, and measurement protocols [2]. These variants display diverse dynamical phenomena, including conserved-charge transport, localization, and measurement-induced phase transitions [16, 40–43].

For now, we introduce the essential tools and conceptual framework required to model quantum many-body systems using random quantum circuits.

1.4.1 Modeling quantum many-body systems

The unitary dynamics of many-body systems governed by a Hamiltonian H can be simulated in quantum circuits using the Lie–Trotter decomposition [44, 45]. In this scheme, the continuous evolution operator $U(t) = \exp(-\mathrm{i}Ht/\hbar)$ is approximated by a sequence of discrete time steps, where $t = N\Delta t$, as

$$|\psi(t)\rangle \approx \left(\mathrm{e}^{-\mathrm{i}H\Delta t/\hbar}\right)^N |\psi(0)\rangle\,. \tag{1.22}$$

Quantum circuits model time evolution by sequentially applying unitary gates, drawn from a universal gate set, to a product state $|\psi(0)\rangle$. Each gate acts locally on a small subset of qubits, typically one or two, corresponding to single-qubit and two-qubit gates, respectively. This local structure induces a light-cone-like propagation of quantum correlations, constrained by a maximum velocity known as the Lieb-Robinson bound [46, 47].

This discretized description of quantum dynamics is well-suited for both theoretical modeling and experimental implementation [20]. In many-body settings, the quantum circuit formalism allows us to perform quantum simulation. The underlying Hilbert space of these systems grows exponentially with system size. For a chain of L qubits, the corresponding Hilbert space is given by $\mathcal{H} = (\mathbb{C}^2)^{\otimes L}$.

In addition, the quantum circuits provide an intuitive and compact graphical representation of quantum computations. Each diagram consists of horizontal lines, known as *wires*, representing individual qubits. Quantum gates are depicted as symbols placed on these wires, indicating the sequence of operations applied to the qubits as the computation progresses from left to right. In particular single-qubit gates are represented by labeled boxes acting on individual wires, while two-qubit gates are depicted using vertical connections between wires. Fig. 1.2 depicts a simple example.

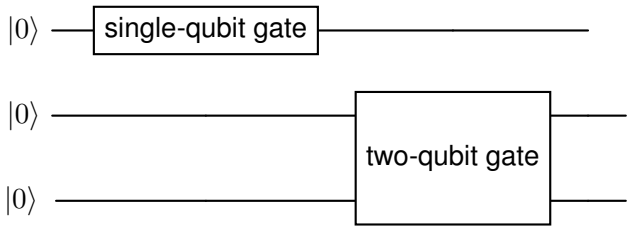


Figure 1.2: Example of a quantum circuit with single and two-qubit gates.

This formalism allows for an immediate visualization of the structure and depth of a quantum algorithm, and is particularly useful for translating different circuit architectures in a simple and unambiguous manner. Furthermore, it provides a significant geometrical tool within the *minimal cut formalism*, as we discuss in the Section 2.

Until quantum computers reach the capability required to efficiently simulate generic quantum dynamics, classical simulations remain indispensable despite facing significant limitations. In this context, exploring how circuit architecture influences entanglement patterns becomes essential for evaluating classical simulability, as highly non-local architectures are harder to simulate, while local ones are more amenable to efficient approximation.

1.4.2 Circuit architectures and locality

The structure of the circuit determines how information and entanglement propagate through the system, so the type of architecture used is important when it is necessary to control entanglement and complexity. This thesis focuses exclusively on local quantum circuits composed of two-qubit gates acting on adjacent qubits (locality). In general, any two-qubit gate acting on non-neighboring qubits can ultimately be decomposed into a sequence of two-qubit gates acting only on adjacent qubits by introducing SWAP gates. Therefore, considering locality here does not reduce the generality of our analysis.

The architecture implemented here follows the "*brickwork*" pattern (see Fig. 1.3), a common scheme where two-qubit gates are applied in alternating layers on even and odd bonds. This structure introduces interactions between neighboring qubits in a regular, staggered manner. In addition, single-qubit gates are applied at every layer, adding superposition and phase to individual qubits.

The general structure of the evolution can be written as:

$$U(t) = \prod_{t=1}^{T} \left[\bigotimes_{i}^{\text{even}} U^{(2)}_{2i,\,2i+1}(t) \bigotimes_{i}^{\text{odd}} U^{(2)}_{2i+1,\,2i+2}(t) \bigotimes_{i} U^{(1)}_{i}(t) \right], \tag{1.23}$$

where, $U_i^{(1)}(t)$ and $U_{i,j}^{(2)}(t)$ denote the single-qubit and two-qubit gates, respectively, applied at layer t to qubit i (and j in the case of two-qubit gates). The brickwork architecture is minimal yet sufficient to ensure the generation of volume-law entanglement and to emulate thermalization-like behavior in systems with no conserved quantities [2]. The constraint of nearest-neighbor interactions reflects realistic physical constraints in quantum hardware and many-body simulation [48].

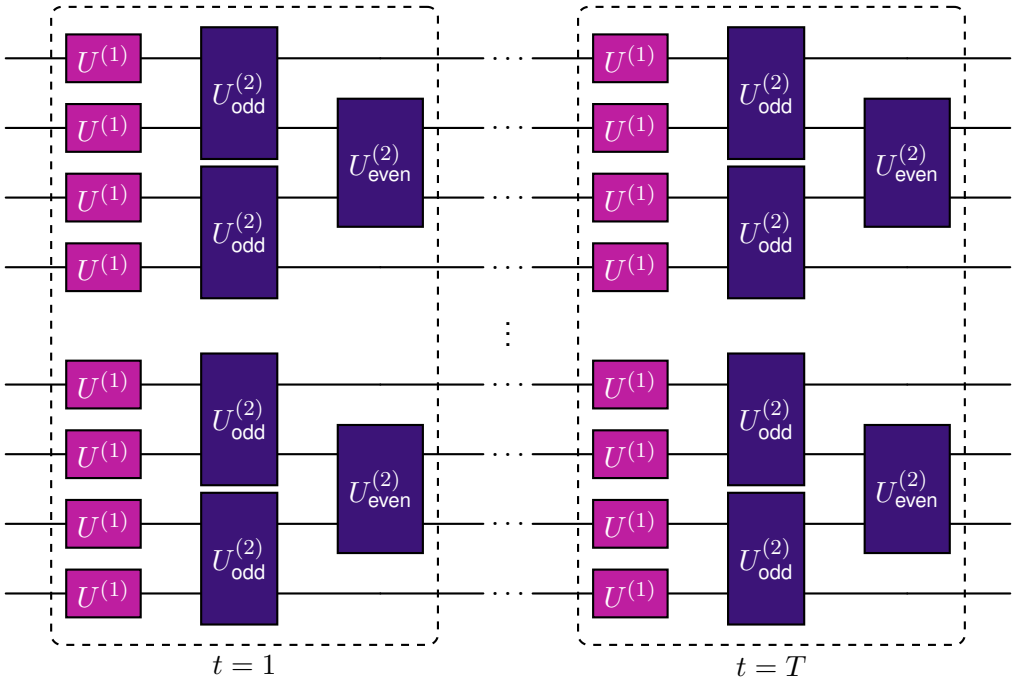


Figure 1.3: Brickwork quantum circuit architecture. Each layer t consists of alternating single-qubit gates $U^{(1)}$ and two-qubit gates $U^{(2)}$ acting on neighboring qubits.

In the present work we perform the distinction between the following two regimes:

Dilute regime – In this regime, each layer contains $\Theta(1)$[3] two-qubit gates, and the average total number of two-qubit gates grows linearly with the circuit depth T.

Dense regime – In this regime, the total number of two qubit gates grows with the volume $\mathcal{A} = LT$ of the circuit.

For both regimes, the probability $g(x,t)$ of finding a two-qubit gate at position x (*i.e.*, between qubits x and $x+1$) and layer t in the circuit is determined by sampling gates from a universal gate set $\mathcal{G}$, according to a specific measure that is explained shortly later in Section 1.4.3. For the dense regime, a naive guess would be to assume that $g(x,t) = \frac{1}{2}$ (see Fig. 1.3). However, g cannot be exactly $1/2$, since that would overcount configurations with overlapping gates. A solution is to slightly shift the probability, *e.g.* $g(x,t) \to \left(\frac{1}{2} - \epsilon\right)$, with ϵ a regulator. In order to fix ϵ, notice that the maximum number of two-qubit gates that a realistic circuit supports[4] is

[3]Here, $\Theta(g(n))$ denotes the set of functions that grow at the same rate as $g(n)$ asymptotically. In particular, $\Theta(1)$ indicates a constant quantity, independent of the system size.

[4]For simplicity we assume both L, T even. The distinction between even/odd becomes irrelevant in the large-circuit

Max(2-qubit gates) $= \frac{1}{2}LT - \frac{T}{2}$; thus, the cutoff ϵ is bounded by $1 \geq 2\epsilon \geq L^{-1}$. Therefore, we choose to parametrize the two-qubit gate probability g as

$$g(x,t) \to \frac{1}{2}\left(1 - \frac{\alpha}{L}\right), \tag{1.24}$$

such that we can smoothly interpolate from the real-hardware (physical) dense regime ($\alpha = 1$) to the dilute one ($\alpha \sim L$).

For simplicity, and without loss of generality, we assume that each layer contains only a single two-qubit gate between adjacent qubits. This places us in the dilute regime. Changing regimes amounts to rescaling the circuit depth, leaving the underlying physics invariant. This assumption also yields a natural time-step structure in which the circuit depth equals the total number of gates applied, simplifying both analytical and numerical analyses.

1.4.3 Clifford random evolution

To study entanglement dynamics, we implement a class of random quantum circuits built exclusively from Clifford gates, inspired by the construction proposed by Nahum *et al.* [1]. Although it is known that Haar-random unitaries generate stronger scrambling and rapid volume-law entanglement by exploring the full Hilbert space, they are infeasible for classical simulation [49]. For this reason, Clifford circuits are preferred, as they enable efficient simulation while still capturing universal aspects of entanglement dynamics [22, 50].

The Clifford group

In our framework, all unitaries are sampled from a prescribed gate set. A natural and analytically tractable choice is the Clifford group $\mathcal{G}$ [50], which is an important subgroup of the unitary group that preserves the Pauli group under conjugation,

$$UP_nU^\dagger \in \mathcal{P}_n, \tag{1.25}$$

where $\mathcal{P}_n$ denotes the n-qubit Pauli group.

Definition 1.4 (Pauli group) *The n-qubit Pauli group, denoted by $\mathcal{P}_n$, is the group generated by tensor products of the single-qubit Pauli matrices*

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

together with multiplicative factors $\{\pm 1, \pm \mathrm{i}\}$. Explicitly,

$$\mathcal{P}_n = \left\{ \omega\, P_1 \otimes P_2 \otimes \cdots \otimes P_n \,\middle|\, P_j \in \{I, X, Y, Z\},\ \omega \in \{\pm 1, \pm \mathrm{i}\} \right\}. \tag{1.26}$$

size limit.

Clifford circuits are efficiently classically simulable (in polynomial time) via the stabilizer formalism [22, 51], yet they are capable of simulating entanglement and information spreading. Therefore, the Clifford group is generated by:

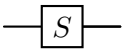
- The **Hadamard gate** H acts on a single qubit and maps the computational basis states $|0\rangle$ and $|1\rangle$ into superpositions with equal amplitudes:

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad H|0\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \quad H|1\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle). \tag{1.27}$$

$$H \quad = \quad$$ 

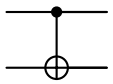
- The **phase gate** S, also known as the $\pi/2$ phase gate, adds a relative phase of i to the state $|1\rangle$ while leaving $|0\rangle$ unchanged:

$$S = \begin{pmatrix} 1 & 0 \\ 0 & \mathrm{i} \end{pmatrix}, \quad S|0\rangle = |0\rangle, \quad S|1\rangle = \mathrm{i}|1\rangle. \tag{1.28}$$

$$S \quad = \quad$$ 

- The **controlled-NOT gate** CNOT is a two-qubit gate that flips the state of the target qubit if and only if the control qubit is in the state $|1\rangle$. In the computational basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$, it is represented as:

$$\mathrm{CNOT} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{1.29}$$

$$\mathrm{CNOT} \quad = \quad$$ 

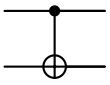
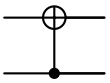
These gates, together with the Pauli group and classical control, form the basis of Clifford circuits. Thus, in the context of the brickwork architecture, the inclusion of Clifford gates allows for the construction of circuits that can generate a variety of quantum states, including maximally entangled states.

Random evolution

We consider Clifford circuits on a one-dimensional chain of qubits, with gates arranged in a brickwork pattern as introduced earlier. The two-qubit gate set is defined as

$$U^{(2)} \in \{\mathrm{CNOT}^{(\mathrm{L})}, \mathrm{CNOT}^{(\mathrm{R})}\}, \tag{1.30}$$

where (L) and (R) denote the direction of the CNOT gate[5], specifying the roles of control and target qubits:

$\mathrm{CNOT}^{(\mathrm{L})}$ =  , $\mathrm{CNOT}^{(\mathrm{R})}$ =  .

The single-qubit gate set is given by

$$U^{(1)} \in \{H, S, I\}. \tag{1.31}$$

These gates constitute the two-qubit Clifford group, denoted as $\mathcal{G}_{|2|} = U^{(1)} \cup U^{(2)}$. The evolution begins with an initial layer of Hadamard gates applied to every qubit. This initialization step places all qubits in a superposition basis and is not counted as a time step; its sole purpose is to slightly accelerate entanglement growth.

At each discrete time step, a single two-qubit gate $U^{(2)}$ is selected at random and applied to a randomly chosen bond. Then, for each qubit, a single-qubit gate $U^{(1)}$ is independently selected and applied. This defines one layer (or time step) of the dynamics (see Fig. 1.4).

The selection of both gate type and gate location follows a uniform probability distribution. Under the constraint of having $\Theta(1)$ two-qubit gates per layer, the corresponding probability $g(x,t)$ must therefore be uniform,

$$g(x,t) = \frac{1}{L-1}. \tag{1.32}$$

As previously mentioned, each layer includes a single two-qubit gate, this choice does not affect the dynamics, as demonstrated in [1], since time can be rescaled accordingly, which is equivalently controlled by the parameter α. In our case, this convention ensures that time corresponds to the number of applied two-qubit gates, and is particularly useful in the *minimal cut model* where single-qubit gates do not contribute. Nevertheless, including single-qubit gates increases the randomness of the generated states, which improves the convergence of entropy averages when only a limited number of realizations can be sampled. In addition, their inclusion yields circuit architectures that more closely reflect practical algorithmic implementations [20].

[5]Note that $\mathrm{CNOT}^{(\mathrm{L})}$ and $\mathrm{CNOT}^{(\mathrm{R})}$ are equivalent up to a SWAP operation, yet in our simulations each is regarded as a distinct gate, since both are implemented as individual gates.

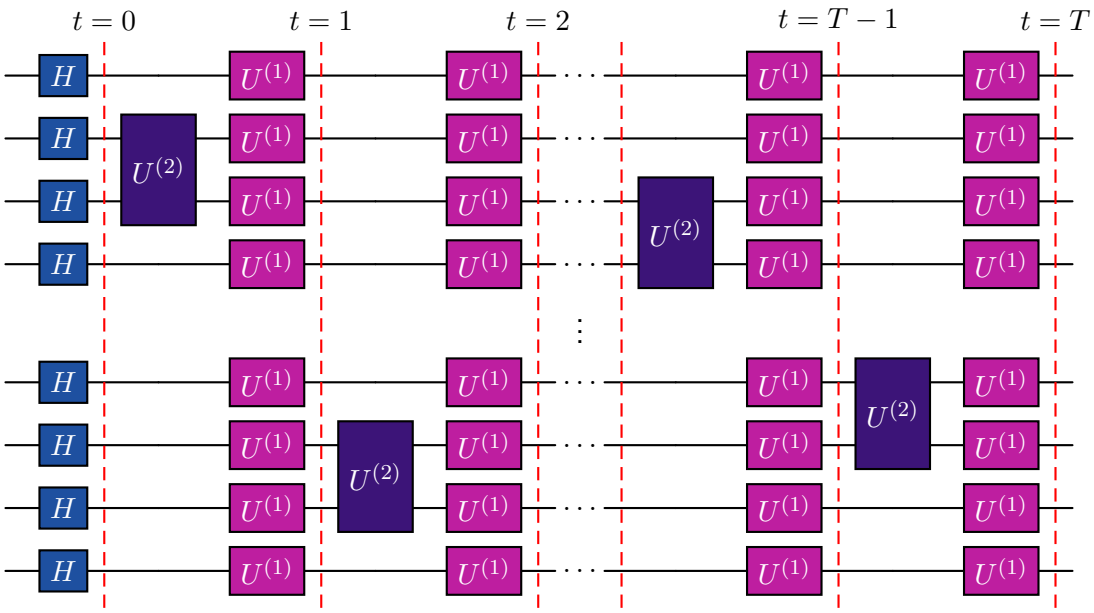


Figure 1.4: Clifford random evolution. The evolution begins with Hadamard gates on all qubits, preparing them in a superposition state. At each discrete time step t, a two-qubit Clifford unitary $U^{(2)}$ is chosen uniformly at random and applied to a randomly selected pair of neighboring qubits. This is followed by independent single-qubit Clifford unitaries $U^{(1)}$, also drawn uniformly at random, applied to all qubits.

Chapter 2

Minimal cut

2.1 Minimal cut formalism

The minimal cut is a classical method for bounding entanglement entropy in tensor networks [52–55]. Continuing the previous discussion on the minimal membrane in Section 1.4, we focus on the $(1+1)d$ systems studied in this thesis, where the minimal membrane becomes a one-dimensional curve (a "cut") that bisects the circuit and provides a geometric upper bound for entanglement entropy.

To illustrate this idea, consider a quantum circuit that evolves an initial product state of dimension q^L into an entangled state at time or layer $\tau = t$. We analyze the Rényi entropy $S_\alpha(x,t)$ for a bipartition at position x, where subsystem A includes the degrees of freedom from 1 to x and B from $x+1$ to L. Any cut through the circuit at position x divides the circuit into two parts, connected by N_{cut} bonds, *i.e.*, legs intersected by the cut (see Fig 2.1). Viewing these as composite tensors *Left* and *Right*, the state takes the form

$$|\psi(t)\rangle = \sum_{i=1}^{q^{N_{\text{cut}}}} \sum_{A,B} \textit{Left}_A^i \ \textit{Right}_B^i \ |A\rangle\,|B\rangle\,, \tag{2.1}$$

where $|A\rangle$ and $|B\rangle$ are basis states for subsystems A and B. This expression implies that the Schmidt rank across the bipartition is bounded by $q^{N_{\text{cut}}}$ (according Eq. (1.19)), hence

$$S_\alpha \leq S_0 \leq \log_2 q^{N_{\text{cut}}} = N_{\text{cut}} \log_2 q. \tag{2.2}$$

The minimal cut, defined as the path intersecting the fewest legs, yields the tightest classical upper bound (the complete proof is provided in Appendix B). In our case, since we work with qubits, we take $q = 2$, giving

$$S_\alpha \leq N_{\text{min cut}} = S_0, \tag{2.3}$$

thus $N_{\text{min cut}} \equiv \mathcal{N}$ coincides exactly with S_0, and therefore the entanglement entropy is bounded above by the minimal number of tensor legs that must be separated to partition the circuit, $\mathcal{N}(x,t)$.

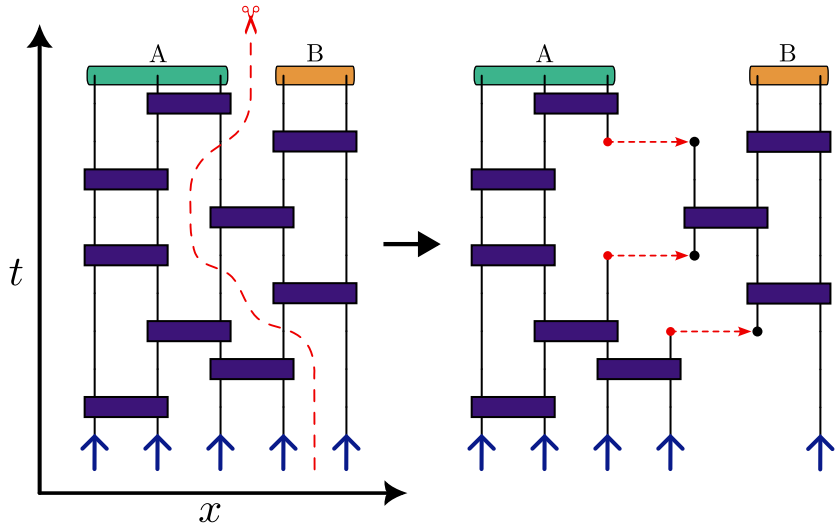


Figure 2.1: Illustration of the minimal cut construction. (Left) Example of a random circuit with five qubits, initialized in superposition states (blue arrows) after Hadamard gates. A candidate cut (dashed red line) intersects three tensor legs, providing an upper bound to the entanglement across the bipartition $A|B$. (Right) Equivalent representation obtained by splitting the circuit into two sub-networks *Left* and *Right*, connected through the shared indices (red dashed arrows), which act as both outputs and inputs of *Left* and *Right*. The number of such shared bonds determines the maximal Schmidt rank allowed by the cut.

This dimensionless quantity is of the same order as S_0 and depends on the discrete site index x in the microscopic description and on the discrete evolution time t, as defined in Section 1.4.3. Notably, this bound become powerful in the limit $q \to \infty$, where for Haar-random states the inequality becomes tight [1].

Moreover, even for finite q, the minimal cut remains a robust estimator for entanglement entropy, serving as a classical computationally efficient approximation to its behavior. The *minimal cut formalism* recasts the problem of computing S_0 to the classical optimization task of minimizing the total number of bonds that must be cut to separate A from B.

2.1.1 KPZ Model

Determining the minimal cut corresponds to a classical problem extensively studied in non-equilibrium statistical mechanics, namely the directed polymer in a random environment (DPRE) [56]. In this framework, the entanglement entropy $S(x,t)$ is mapped to the minimal energy configuration of a directed polymer crossing a disordered landscape. The discrete analogue to the polymer's energy corresponds to the number of circuit legs severed by the minimal cut in a quantum circuit.

The DPRE problem belongs to the universality class of the Kardar–Parisi–Zhang (KPZ) equation [39], which governs the dynamics of growing interfaces. In our context, the KPZ equation does not follow from a microscopic derivation of circuit dynamics, rather emerges from the mapping to the DPRE, placing random circuits in the same universality class as KPZ growth [30]. The

KPZ equation describes the stochastic evolution of the entanglement profile $S_\alpha(x,t)$ [1], which we extrapolate to the minimal cut by the correspondence $S_\alpha(x,t) \sim \mathcal{N}(x,t)$, as both quantities capture the coarse-grained growth of an effective entanglement "surface". $\mathcal{N}(x,t)$ thus follows the KPZ equation:

$$\partial_t \mathcal{N}(x,t) = \nu \partial_x^2 \mathcal{N}(x,t) - \frac{\lambda}{2}(\partial_x \mathcal{N}(x,t))^2 + \eta(x,t) + c, \tag{2.4}$$

where ν, λ, and c are non-universal, model-dependent constants. The first term represents diffusion, the second term accounts for the nonlinearity associated with the local growth rate (or, equivalently, the effective "particle size"), and the last term, $\eta(x,t)$, is a Gaussian noise term representing randomness in the circuit.

In terms of the minimal cut $\mathcal{N}(x,t)$, the scaling translates into the following behaviors:

$$\begin{aligned} \langle \mathcal{N}(x,t) \rangle &\sim v_{\mathrm{E}} t, \\ \Delta(\mathcal{N}(x,t)) &\sim t^{\beta}, \\ \xi(t) &\sim t^{1/z}, \end{aligned} \tag{2.5}$$

where v_{E} is the entanglement velocity, $\Delta\mathcal{N}(x,t)$ is the deviation that quantifies the size of fluctuations around the mean, and $\xi(t)$ is the spatial correlation length. The first relation describes the ballistic growth of mean entanglement entropy in time, while the second and third equations encode its universal fluctuation and correlations properties. In most analyses, these quantities are expressed in rescaled units, so that time, length, and height become dimensionless, emphasizing the universal nature of the KPZ scaling exponents.

The solution to the equation (2.4) exhibits universal features characterized by two critical exponents. The exponent $\beta = 1/3$ governs the magnitude of temporal fluctuations in the interface height, while $\alpha = 1/2$ governs the roughness or spatial correlations of the interface [1, 57]. Together they define the dynamical exponent $z = \alpha/\beta = 3/2$, which controls the space–time scaling of fluctuations.

At early times ($t \ll x$), both the entanglement entropy and the minimal cut grow ballistically, while at later times ($t \gg x$), it saturates to a volume-law value $S(x,t) \sim x$. This crossover yields the characteristic "entanglement pyramid" profile (see Fig 2.2): for a finite region of length x, the entropy increases linearly, until the saturation time $t_{\mathsf{sat}} = x/v_{\mathrm{E}}$ is reached, after which it remains constant. This linear-to-saturated behavior is also referred to as an "entanglement light cone" in the literature [1].

This piecewise behavior is captured by:

$$\langle \mathcal{N}(x,t) \rangle = \begin{cases} v_{\mathrm{E}} t, & \text{if } t < t_{\mathsf{sat}}, \\ x, & \text{if } t \geq t_{\mathsf{sat}}, \end{cases} \tag{2.6}$$

The saturation time t_{sat} marks the point at which entanglement entropy ceases to grow and reflects the maximum entanglement attainable for the given region.

This mapping provides a physical interpretation of entanglement dynamics in d-dimensional random quantum circuits, in which it behaves as a fluctuating domain wall (or membrane) whose position minimizes a random energy functional. The KPZ exponents describe both the average

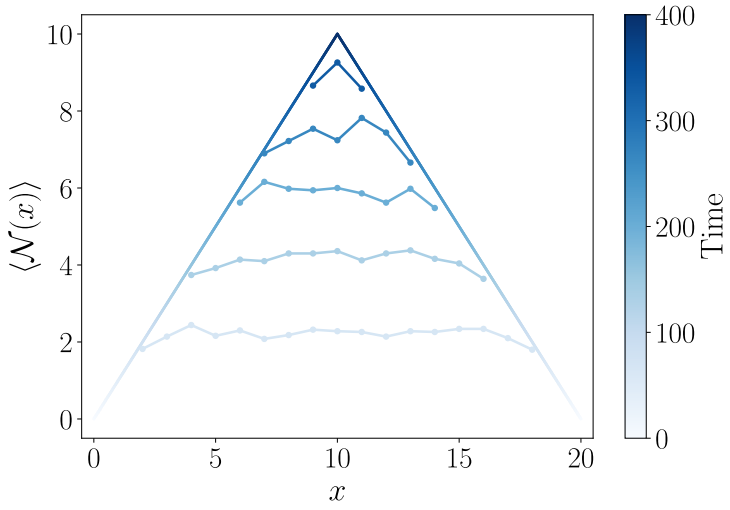


Figure 2.2: Spatial profile of the minimal cut $\mathcal{N}(x,t)$ for a random Clifford circuit with $L = 20$ qubits at evolution times $t = 66, 133, 200, 266,$ and 333. The profile exhibits the characteristic *pyramidal* shape: at late times, $t \gg x$, the minimal cut exits through the lateral boundary of the circuit indicating that the entropy has reached saturation, whereas for $t \ll x$ it exits through the bottom boundary. Reproduced from [1].

and fluctuating behavior of this membrane in a generic case, offering a universal framework to understand entanglement growth in its most general form.

However, this scheme captures only fluctuation statistics and not the exact entropy dynamics. The limitation stems from the fact that KPZ, formulated for continuous systems such as polymers, cannot be directly mapped to discrete quantum circuits. In this sense, KPZ combined with the *replica trick* yields only universal scaling exponents [58].

2.1.2 Cut growth model

In this section, we study the statistical properties of cuts in random quantum circuits to extract information beyond the generic fluctuations described by KPZ. By mapping cuts to the trajectory of classical walkers navigating the circuit geometry, we obtain an intuitive framework to follow the evolution of cuts over time.

Any cut can be viewed as the world-line traced by a classical point-particle (walker) which avoids the obstacles (gates) (see Fig. 2.3). Concretely, from this perspective, we are interested in the evolution of the probability $p(x,t)$ that the "walker" is in the x-th rail at time t. In the most general scenario, the evolution of this probability is described by the Markov process,

$$p(x, t+1) = \sum_{x' \in \text{rails}} \Pr(x' \to x|t)\, p(x', t), \tag{2.7}$$

where $\Pr(x' \to x|t)$ is the transition probability of moving from x' to x in step t. For circuits

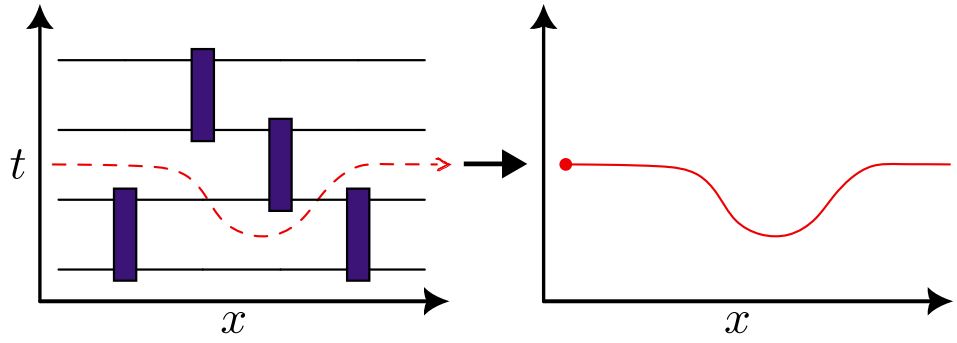


Figure 2.3: Minimal cut, and worldline equivalence. The Minimal cut can be mapped to the worldline traced by a classical walker (red circle) moving in geometry imposed by the obstacles (purple gates).

composed of two-qubit gates, the Markov process reads

$$\begin{aligned} p(x,t+1) =& c(x+1,x)\, g(x+1,t)\, p(x+1,t) \\ &+ c(x-1,x)\, g(x-1,t)\, p(x-1,t) + (1-g(x,t))\, p(x,t), \end{aligned} \tag{2.8}$$

where $c(x',x)$ is the probability that the walker chooses to jump from rail x' to x. This is $\frac{1}{2}$ if $|x'-x|=1$, except at the boundaries, where $c(-\frac{L}{2}+1,-\frac{L}{2}+2) = c(\frac{L}{2}-1,\frac{L}{2}-2) = 1$, and 0 otherwise. The function $g(x,t)$ represents the probability that a two-qubit gate is in the x-th rail at time t. As mentioned in Section 1.4.3, the gates are distributed according to a uniform distribution, which gives $g(x,t) = g = \frac{1}{L-1}$, with L the number of qubits.

This yields the following tridiagonal transition matrix:

$$\begin{pmatrix} p(\frac{L}{2}-1,t+1) \\ \vdots \\ p(x,t+1) \\ \vdots \\ p(-\frac{L}{2}+1,t+1) \end{pmatrix} = \begin{pmatrix} 1-g & \frac{1}{2}g & & & & \\ g & 1-g & \frac{1}{2}g & & & \\ & \frac{1}{2}g & 1-g & \ddots & & \\ & & \ddots & \ddots & \frac{1}{2}g & \\ & & & \frac{1}{2}g & 1-g & g \\ & & & & \frac{1}{2}g & 1-g \end{pmatrix} \begin{pmatrix} p(\frac{L}{2}-1,t) \\ \vdots \\ p(x,t) \\ \vdots \\ p(-\frac{L}{2}+1,t) \end{pmatrix}. \tag{2.9}$$

Rather than tracking the position of the walker via its probability distribution $p(l,t)$, the figure of merit, as long as the minimal cut concerns is the statistics for the total number of cuts (walker crossings) $\mathcal{N}(t)$ at the time t.

Definition 2.1 (Cut counting function) *Given an L-qubit, depth-T circuit $\mathcal{C}$, the cut counting function $\mathcal{N}[\mathcal{P}_\mathcal{C}]$ takes a path $\mathcal{P}_\mathcal{C} \in \mathcal{P}[\mathcal{C}]$, with $P[\mathcal{C}]$ the set of all paths compatible with the circuit $\mathcal{C}$,* i.e. *paths that don't pass through any gate and cut the circuit $\mathcal{C}$ across a given size x bi-partition (all possible ways for a walker), and returns a positive integer counting the number of times the path $\mathcal{P}$ cuts the circuit,* i.e.

$$\mathcal{N} : (\mathcal{C},\mathcal{P}) \to \mathbb{Z}^+. \tag{2.10}$$

In this definition we are implicitly assuming that the transverse path $\mathcal{P}_\perp$, i.e. *the one that horizontally cuts x times the circuit $\mathcal{C}$ is a valid path. Therefore the cut function achieves the maxima $\max_{\mathcal{C}} \mathcal{N}[P_{\mathcal{C}}] = \max(T, x)$.*

Closely related with $\mathcal{N}(t)$, is the cut-increase distribution $\mathcal{Q}(t)$ defined as the probability that the random variable $\mathcal{N}(t)$ increases at time t, *i.e.*

$$\mathcal{Q}(t) = (\Delta\mathcal{N}(t) = 1), \tag{2.11}$$

with $\Delta\mathcal{N}(t) = \mathcal{N}(t+1) - \mathcal{N}(t)$ the spacings. Restricting to the dilute-regime, and far from the boundaries, the walker-probability (2.8) follows the discrete diffusion equation,

$$p(x, t+1) = \frac{1}{2(L-1)} \left(p(x+1, t) + p(x-1, t)\right) + \left(1 - \frac{1}{L-1}\right) p(x, t). \tag{2.12}$$

Then the cut-increase probability (2.11) is constant, and equal to the probability of finding a two-qubit gate at the position x (see Appendix C.1),

$$\mathcal{Q}(t) = \sum_{x \in \text{rails}} g(x, t)\, p(x, t) = \frac{1}{L-1}. \tag{2.13}$$

As a final remark from this example, and to gain a physical picture, notice that the continuum analog of (2.12) is a Diffusion/Advection system with diffusion constant $D \sim \frac{1}{L}$, and associated characteristic velocity $v = \frac{D}{l_{\text{typ}}}$, being l_{typ} the typical size of the generators (n-qubit gates) for the Brownian kicks that makes the particle diffuse, which in our case is $l_{\text{typ}} = 2$. This typical velocity will be related to the velocity of the entanglement growth as we will show later.

2.1.2.1 Cut-Number distribution

Computing the exact distribution of the minimal cut in a random quantum circuit is a highly nontrivial problem. This difficulty arises because the task amounts to solving an optimization problem, which is computationally hard in the presence of disorder. Consequently, rather than attempting an exact evaluation, we adopt an approximate strategy.

In particular, we employ the *quenched approach*[1], in which both the optimization and averaging are on the same footing. This procedure means that one averages over the quantity $\mathcal{N}[\mathcal{P}_{\mathcal{C}}]$. Thus, the quenched average is defined as

$$\Big\langle \mathcal{N}[\mathcal{P}_{\mathcal{C}}] \Big\rangle_{\mathcal{U}_{L,T}}, \tag{2.14}$$

where the brackets denote the average over $\mathcal{U}_{L,T}$, the ensemble of all random circuits $\mathcal{C} \in \mathcal{U}_{L,T}$ with L qubits and depth T.

This approach allows us to characterize the distribution of cut numbers for a generic cut (not necessarily the minimal one). Importantly, any such cut still constitutes a valid upper bound on the Hartley entropy S_0, though not the sharpest bound. Nevertheless, by analyzing the statistics

[1] The term "quenched" is borrowed from the disorder community, where it refers to averaging procedures in systems with frozen randomness (quenched disorder), in contrast to "annealed" average.

of generic cuts we may extrapolate their behavior to infer properties of the minimal one, which remains the central object of interest in bounding entanglement growth. To summarize, the study of cut-number distributions offers a practical route to approximate the scaling of the minimal cut.

In terms of the cut gaps $\Delta\mathcal{N}$, the number of cuts is defined by the random process $\mathcal{N}(t) = \sum_{\tau=0}^{t} \Delta\mathcal{N}(\tau)$. Since we are interested in time-independent (stationary) circuit architectures, each cut spacing $\Delta\mathcal{N}(t)$ is an independent binary random variable sampled according to

$$\varrho(\Delta\mathcal{N}(t)) = \mathcal{Q}(t)\delta(\Delta\mathcal{N}(t)-1) + (1-\mathcal{Q}(t))\delta(\Delta\mathcal{N}(t)), \tag{2.15}$$

and the joint distribution $P[\mathcal{N}(t)]$ for the total number of cuts at depth t (modulo normalization pre-factor) is given by (see Appendix C.2),

$$P[\mathcal{N}(t)] = \int_{\mathbb{R}} \mathrm{e}^{ik\mathcal{N}(t)+\sum_{\tau=0}^{t}\log\left(1+\mathcal{Q}(\tau)\left[e^{-ik}-1\right]\right)} dk. \tag{2.16}$$

2.1.2.2 Gaussian Approximation

Before delving into the full number of cuts distribution, let us make the following approximation to grasp intuition regarding the qualitative behavior of (2.16). For large enough number of steps, the number of cuts $\mathcal{N}$ increases as well, and only small values of k contribute inside (2.16) making the distribution sharply Gaussian, *i.e.*

$$P[\mathcal{N}(t)] \approx \exp\left(-\frac{1}{2t}\frac{(\mathcal{N}(t)-t\,\overline{\mathcal{Q}}(t))^2}{(\overline{\mathcal{Q}}(t)-\overline{\mathcal{Q}^2}(t))}\right), \tag{2.17}$$

where the overline denotes the time-average, *e.g.* $\overline{\mathcal{Q}^m}(t) = \frac{1}{t}\sum_{\tau=0}^{t}\mathcal{Q}^m(\tau)$. Within this approximation, and as expected, the average cut-number at long-times is given by $\overline{\mathcal{N}}(t) = t\,\overline{\mathcal{Q}}(t)$[2]. Using the transition matrix in the dilute, uniform case, the moments satisfy $\overline{\mathcal{Q}^m} = g^m$ (see Appendix C.1).

2.1.2.3 Exact Distribution

After integrating out k in (2.16) (see Appendix C.2), the distribution for the number of cuts made by the walker (without the normalization factor) reads

$$P\left[\mathcal{N}(t)\right] \propto \frac{1}{\mathcal{N}(t)!}\left(\sum_{\tau=1}^{t}\frac{\mathcal{Q}(\tau)}{1-\mathcal{Q}(\tau)}\right)^{\mathcal{N}(t)}\prod_{\tau=1}^{t}(1-\mathcal{Q}(\tau)). \tag{2.18}$$

Therefore, for constant $\mathcal{Q}(\tau)\equiv\mathcal{Q}$, the number of cuts $\mathcal{N}$ at large t follows the Binomial distribution, $\mathcal{N}\sim\mathrm{Bin}(t,\mathcal{Q})$ with $\mathcal{Q}=\frac{1}{L-1}$. *i.e.*, the number-of-cuts distribution reads

$$P\left[\mathcal{N}(t)\right] = \binom{t}{\mathcal{N}}(1-\mathcal{Q})^{t-\mathcal{N}}\;\mathcal{Q}^{\mathcal{N}}, \tag{2.19}$$

[2]Note that $\overline{\mathcal{N}}$ denotes the average cut over all paths $\mathcal{P}_{\text{all}} = \cup_{\mathcal{C}\in\mathcal{U}_{L,T}}\mathcal{P}[\mathcal{C}]$ for a fixed ensemble $\mathcal{U}_{L,T}$. This is interpreted within the quenched average as $\overline{\mathcal{N}} = \langle\mathcal{N}[\mathcal{P}]\rangle_{\mathcal{U}_{L,T}}$.

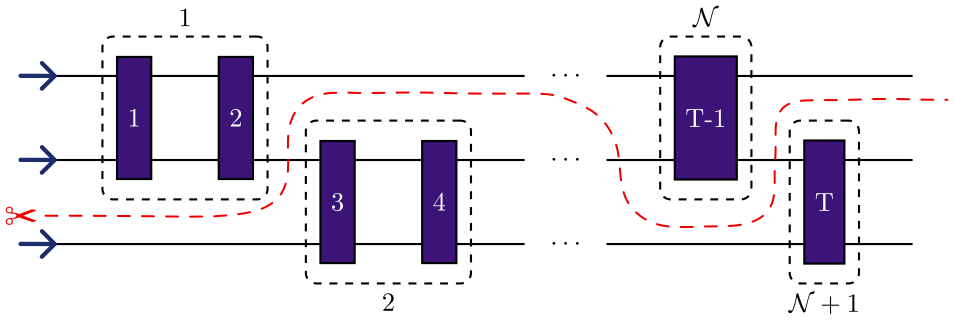


Figure 2.4: Minimal cut for a three-qubit circuit ($L = 3$), an analytically solvable case. The number of ways to obtain $\mathcal{N}$ cuts follows from distributing t gates into $\mathcal{N}+1$ partitions, with the additional contribution from allowing obstacle 1 to remain empty. Summing both yields the binomial distribution shown. With only two rails, the cut is symmetric about x, and so is the entanglement entropy.

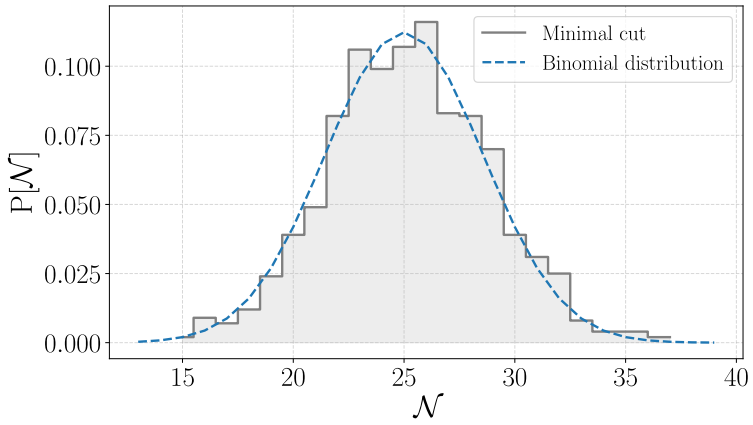


Figure 2.5: Normalized histogram for $L = 3$ qubits at $t = 50$, obtained from 1000 ensembles of random Clifford circuits without considering saturation effects. The data clearly follows the binomial distribution given in Eq. (2.20), with mean value $t\mathcal{Q} = \frac{t}{L-1}$.

and we find that the mean cut scales as $\overline{\mathcal{N}}(t) = t\mathcal{Q}(t)$.

To gain further insight, let us examine the behavior of the system in the regime close to saturation. In particular, it is useful to consider a simple case where exact results can be obtained. As a test of the approximation: for $L = 3$ qubits, the minimal-cut problem is exactly solvable. In this scenario, there is a cut for each obstacle (two-qubit gate) in the circuit, since we have two possible positions available (see Fig. 2.4). Distributing t gates among these $\mathcal{N}+1$ obstacles leads precisely to the binomial form given in Eq. (2.19),

$$P\left[\mathcal{N}(t)\right] = \frac{\text{\# Circuits with } \mathcal{N}+1 \text{ obstacles}}{\text{\# Total circuits}} = \frac{1}{2^t}\binom{t}{\mathcal{N}}. \tag{2.20}$$

This result is consistent with the random-walker approach underlying the minimal cut formulation, providing the exact connection between a generic cut and the minimal one, and is supported

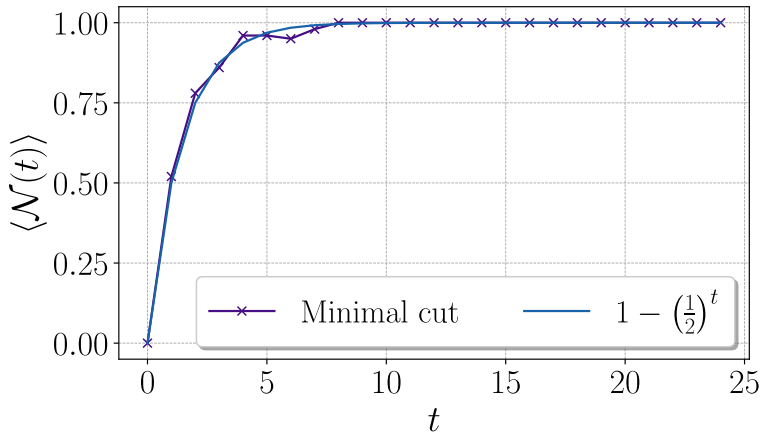


Figure 2.6: Average minimal cut for random Clifford circuits with $L = 3$ qubits, computed over 50 ensembles. The data shows clear agreement with the analytical description given by Eq. 2.22, once saturation effects are taken into account. The bipartition position x has only two possible choices, $x = 1$ or $x = 2$, and both cases are symmetric; therefore, the results are identical. Here we present the case $x = 1$.

by numerical evidence (see Fig.2.5). However, the analysis above ignores the fact that the entanglement entropy saturates at its maximal value (Eq. 1.9). In physical circuits, the minimal cut $\mathcal{N}(x,t)$ cannot grow indefinitely; instead, it saturates at the subsystem size x (with $x \leq L - x$). Consequently, the cut growth analysis must be modified: if $\mathcal{N}(x,t) \geq x$, then $\mathcal{N}(x,t) = x$. Thus, the mean cut is given by the following truncated binomial form:

$$\begin{aligned} \langle \mathcal{N}(x,t) \rangle &= \sum_{n=0}^{x-1} n \binom{t}{n} (1-\mathcal{Q})^{t-n} \, \mathcal{Q}^n + x \left(1 - \sum_{n=0}^{x-1} \binom{t}{n} (1-\mathcal{Q})^{t-n} \, \mathcal{Q}^n \right) \\ &= \sum_{n=0}^{x-1} (n-x) \binom{t}{n} (1-\mathcal{Q})^{t-n} \, \mathcal{Q}^n + x. \end{aligned} \tag{2.21}$$

Following the analytic example for $L = 3$, the maximum saturation value is $x = 1$, yielding the exact result

$$\langle \mathcal{N}(t) \rangle = 1 - \left(\frac{1}{2} \right)^t . \tag{2.22}$$

The correction term $(1/2)^t$ decays exponentially and thus becomes negligible at late times. However, it substantially alters the pre-saturation behavior, rendering a naive linear approximation inadequate, as illustrated in Fig. 2.6.

For large L, however, Eq. (2.19) provides an accurate description prior to saturation for the cut. The deviation becomes relevant only near saturation, where contributions with $\mathcal{N} > x$ accumulate, shifting probability weight toward x. Nevertheless, before this regime, the expected linear growth is preserved.

Returning to our initial discussion, this binomial behavior corresponds to the *generic* cut and not to the minimal one. This can be easily seen by finding the minima of the cut-number distribution

which trivially yields zero, *i.e.*,

$$\min_{\mathcal{P}\in\mathcal{P}_{\text{all}}} \left\langle \mathcal{N}\left[\mathcal{P}_{\mathcal{C}}\right] \right\rangle_{\mathcal{U}_{L,T}} = \min\left(\text{Bin}(T,\mathcal{Q})\right) = 0, \quad \text{with} \quad \mathcal{P}_{\text{all}} = \cup_{\mathcal{C}\in\mathcal{U}_{L,T}}\mathcal{P}[\mathcal{C}]. \tag{2.23}$$

This result follows because, as we first average over the ensemble $\mathcal{U}_{L,T}$, there always exist a family of circuits $\mathcal{C}$ with no gates intersecting the bipartition at position x, thereby yielding a simple cut equal to zero. Therefore, searching for the minimal cut from this point becomes trivial. Nevertheless, we have gained important information about the generic cut under the Markov-chain model. In this model, we only assume that the random walker avoids obstacles when encountering one, which does not deviate significantly from the behavior of a true minimal cut. Indeed, we can think of the walker as attempting to minimize the energy (or cut) by switching between paths only when necessary, the minimum cut also seeks this same optimization, but in the most efficient manner possible. Hence, we can expect that the generic behavior of the cut brings us closer to understanding the minimal one.

In summary, the analysis of the cut-number distribution reveals two central insights. First, in the pre-saturation regime, the mean cut grows linearly with time, in agreement with the predictions of the KPZ universality class, where the entanglement entropy scales as $S_\alpha(x,t) \sim v_{\text{E}}t$.

Second, by employing the binomial approximation for a generic cut, we predict that the entanglement velocity scales inversely with the number of qubits, $v_{\text{E}} \propto 1/L$. This already goes beyond the KPZ framework, which does not account for such microscopic system-specific dependence. Moreover, under the diffusion equation picture introduced earlier (Eq. 2.12), this inverse proportionality must be determined by the gate size. Since our circuits consist of two-qubit gates, we conjecture that the proportionality constant is is given by $1/2$, yielding

$$v_{\text{E}} = \frac{1}{2L}. \tag{2.24}$$

Thus, our approach not only recovers the linear growth predicted by KPZ but also refines it by identifying a precise dependence on system size and microscopic gate structure. This constitutes the main contribution of this picture, a quantitative prediction for the entanglement velocity that extends beyond the reach of the KPZ universality class.

2.2 Graph-based algorithm for the minimal cut

As discussed in the previous sections, computing the minimal cut in random quantum circuits amounts for solving a nontrivial optimization problem, closely related to classical disordered systems. This intrinsic difficulty motivates the development of specialized computational methods tailored to the structure of quantum circuits. In this spirit, we propose an algorithmic strategy designed to efficiently determine the minimal cut.

This section presents, as far as we are aware, an original algorithm, developed by the author and not previously reported in the literature, to compute the minimal cut in quantum circuits represented as graphs. The approach is inspired by the classical algorithm known as *Dijkstra's shortest path* [59], but it is modified to incorporate the geometric and physical constraints inherent to quantum circuit architectures.

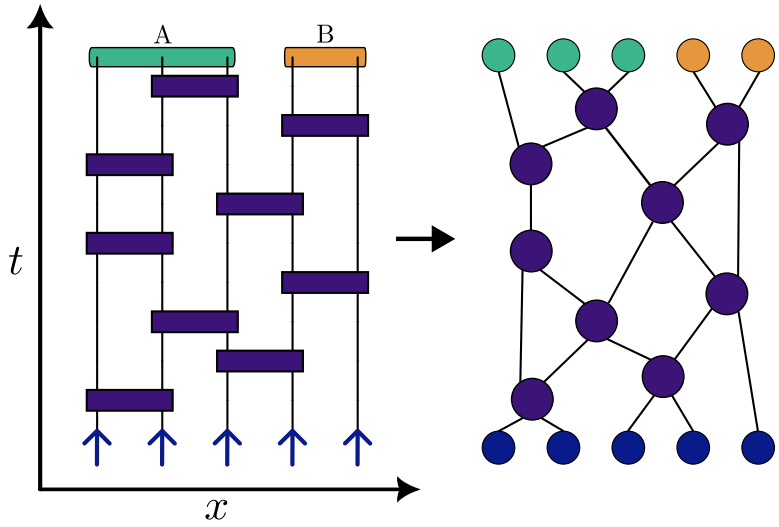


Figure 2.7: Graph representation of a quantum circuit. Each two-qubit gate is mapped to a four-legged node (spider). When two gates act consecutively on the same pair of qubits, the connecting edges are fused and their weight is increased by one. Qubits are represented as nodes with a single edge, and the final state is depicted by evolved qubit nodes at the output layer.

The significance of this algorithm lies in its runtime complexity, which is polynomial in both the circuit depth and the number of qubits, as demonstrated later. This property provides an efficient computational tool for evaluating S_0 and, more generally, for establishing upper bounds on the Rényi entropies S_α. We now detail the construction, motivation, and computational characteristics of this approach.

2.2.1 The ZX-calculus

To understand the mentioned algorithm, it is necessary to introduce the ZX-calculus, a rigorous graphical language for reasoning about linear maps between qubits, using a form of string diagrams called ZX-diagrams [60]. To build the graph representation of a circuit, we follow a construction inspired by the ZX-calculus formalism, in which quantum circuits are viewed as tensor networks composed of spiders (nodes) and wires (edges).

Each two-qubit gate is encoded as an edge connecting the qubit nodes through an intermediate spider, ensuring the correct causal and spatial structure. Single-qubit gates, though naturally associated with individual nodes, are omitted since they do not affect the minimal cut. Qubits appear as horizontal sequences of nodes with time flowing from bottom to top, and gates are placed along their trajectories.

This graphical encoding provides a framework to handle quantum operations; however, our analysis focuses only on the geometric structure rather than on the explicit gate representation. A schematic example of such a representation is provided in Fig. 2.7.

2.2.2 Algorithm design

A proper minimal cut $\mathcal{N}(x, T)$ in this context must respect two key constraints:

1. The cut must separate the circuit into two valid subsystems A and B, beginning between specified top nodes a and b at $t = T$, corresponding to the qubits located at positions x and $x + 1$, respectively, which define the boundary between the subsystems.
2. The cut must remain entirely within the interior of the circuit, avoiding any exit through the top layer or the lateral boundaries[3].

To implement these constraints, we propose the following algorithm:

Min-cut algorithm. The main idea is to reframe the cut as a path within the cycle space of the circuit graph. In graph theory, a *cycle* is defined as a closed path that starts and ends at the same vertex. The cycle space is generated by the so-called basis cycles, *i.e.*, a minimal set of simple cycles from which all other cycles in the graph can be constructed.

Definition 2.2 (Simple cycle) *A simple cycle in a graph is a closed path that does not repeat vertices or edges, except for the initial and final vertex which coincide.*

As illustrated in Fig. 2.7, the qubits represented by the top and bottom nodes are not directly connected to each other. Consequently, it is not possible to construct cycles involving only these nodes and the first or last gates. To overcome this, we introduce additional edges connecting the corresponding top and bottom nodes; however, these edges are assigned weight zero, since they do not represent physical connections.

Furthermore, in order to reduce the number of cycles, we note that any pair of contiguous two-qubit gates acting on the same qubits can be merged into a single edge with weight two. Recall that, by default, each edge is assigned weight one, which corresponds to the cost of cutting that edge. Therefore, it is consistent to assign weight two to the merged edge, as it accounts for the cost of cutting two equivalent edges that would otherwise form a simple cycle involving only two gate nodes.

Following this line of ideas, we can construct the cycle-graph that represents the original quantum circuit graph (see Fig. 2.8), subject to the following conditions:

- Each node represents a simple cycle in the quantum circuit graph, hereafter referred to as a “cycle-vertex”.
- An edge connects two cycle-vertices if they share a common physical edge in the quantum circuit graph; its weight reflects the cost of cutting that shared edge.
- Source node s corresponds to the cycle including top nodes a and b and their last quantum gate.

[3] Although the cut can in principle exit through the lateral boundaries, within the algorithm we adopt the convention that whenever $\mathcal{N}(x, t) > x$ (with $x \leq L - x$), we simply set $\mathcal{N}(x, t) = x$

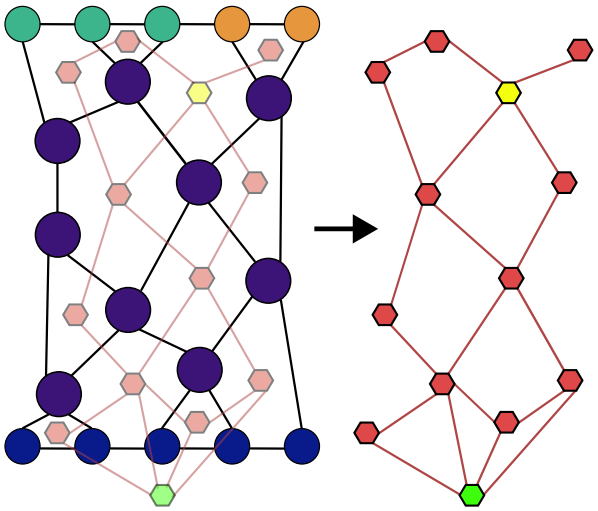


Figure 2.8: Cycle-graph representation. On the left, the quantum circuit graph is shown, where all qubits are connected by weightless edges. Simple cycles are identified and represented as hexagonal nodes. Among them, the yellow node marks the simple cycle along which the system is bipartitioned into $A|B$. An additional green hexagonal node is introduced at the bottom, outside the circuit, to which all possible cut outputs converge. From this construction, the cycle-graph (right) is extracted, where the task reduces to find the shortest path from the yellow partition node to the green output node.

- Target node e represents an artificial cycle added to the quantum circuit graph, to which all cuts converge as the output.

Once the cycle-graph representation of the quantum circuit is constructed, encoding the possible flow of connectivity across different cutting configurations, we proceed to employ *Dijkstra's shortest path algorithm*. This classical algorithm, widely used in graph theory, determines the minimal-cost path between a source s and a target node e by iteratively updating distance estimates and selecting the node with the smallest tentative distance until the destination is reached [59]. In our context, the "cost" assigned to each edge corresponds to a unit contribution to the cut. Hence, the algorithm finds the path of minimal edge weight, interpreted as the minimal number of cuts in the quantum circuit graph.

2.2.3 Runtime analysis

The runtime of the proposed minimal cut algorithm can be decomposed into three main contributions: (i) the search for the cycle basis of the quantum circuit graph, (ii) the generation of the cycle-graph, and (iii) the execution of Dijkstra's shortest path algorithm on the cycle-graph.

i) We begin by noting that the initial input is the quantum circuit graph. The first step is to compute a minimum cycle basis. For this task we adopt the algorithm of Kavitha *et al.* [61], which runs in time,

$$\mathcal{O}(E^2V), \tag{2.25}$$

with V and E the number of vertices and edges of the quantum circuit graph, respectively.

ii) Once the simple cycle basis is determined, the second step is the construction of the cycle-graph. Each basis cycle is promoted to a node (a cycle-vertex) in this auxiliary graph. Edges are then added between two cycle-vertices if the corresponding cycles in the original quantum circuit graph share an edge, in which case the new edge inherits the weight of the shared edge. This procedure requires comparing every pair of cycles, yielding $\mathcal{O}(V_{\mathrm{c}}^2 E_{\mathrm{c}})$ operations, where V_{c} is the number of basis cycles and E_{c} the typical cycle size. In the quantum circuit graphs, cycles are highly local and typically consist of only four nodes and four edges, *i.e.*, $E_{\mathrm{c}} = \Theta(1)$. Consequently, the practical runtime of constructing the cycle-graph reduces to

$$\mathcal{O}(V_{\mathrm{c}}^2). \tag{2.26}$$

iii) Finally, the minimal cut is obtained by applying Dijkstra's algorithm on the cycle-graph. The standard runtime of Dijkstra's algorithm using an adjacency matrix representation is

$$\mathcal{O}(V_{\mathrm{c}}^2), \tag{2.27}$$

scaling quadratically in the number of cycle-vertices.

Collecting all contributions, the total runtime of the algorithm is bounded by

$$\mathcal{O}\big(E^2 V + V_c^2\big)\,, \tag{2.28}$$

Inserting the explicit scalings in terms of the number of qubits (L) and the circuit depth (T), we have $V = 2L + T$, $E = 3L + 2T - 2$, and $V_{\mathrm{c}} \leq L + T$ for circuits in $\mathcal{U}_{L,T}$, this last bound is not always saturated since contiguous two-qubit gates on the same qubits merge into a single edge, eliminating a simple cycle.

Therefore, the runtime is polynomial in L and T, and is asymptotically dominated by the $\mathcal{O}(E^2 V)$ term from the minimum cycle basis algorithm. This establishes that the overall complexity of the minimal cut algorithm scales as

$$\mathcal{O}\big((L+T)^3\big). \tag{2.29}$$

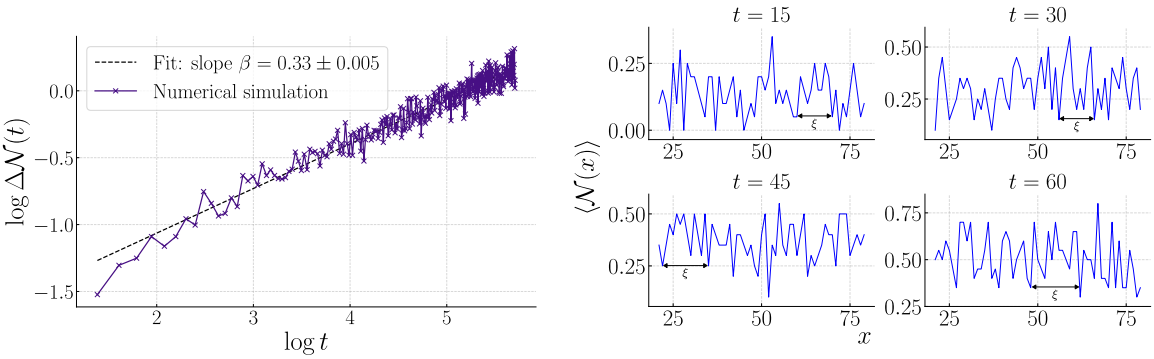
Thus, the runtime analysis demonstrates that the computational bottleneck lies in the determination of the cycle basis, while the subsequent construction of the cycle-graph and execution of Dijkstra's algorithm contribute only quadratic terms in $L + T$.

Chapter 3

Results and analysis

3.1 Numerical results

Computing the exact entanglement entropy requires evolving the full quantum state, which is exponentially costly in the number of qubits L, scaling as $\mathcal{O}(2^L)$. In contrast, the Graph-Based Algorithm for the Minimal Cut proposed earlier offers a polynomial-time classical approximation to study entanglement, making it substantially faster and more scalable.



(a) (b)

Figure 3.1: (a) Log–log scaling of the minimal cut fluctuations $\Delta\mathcal{N}(t)$ for a random circuit with $L = 50$ qubits, evolved up to depth $t = 300$ and averaged over 100 ensembles. The linear regression in the asymptotic regime yields a slope $\beta = 0.333 \pm 0.005$, in agreement with the KPZ universality class. (b) Spatial fluctuations sampled over 100 random circuits of $L = 100$ qubits at times $t = 20, 40, 60, 80$, showing collapse under KPZ scaling with exponent $\xi \sim t^{2/3}$.

KPZ fluctuations in minimal cut

Building on this computational advantage, we begin by verifying the emergence of KPZ scaling in random quantum circuits using our minimal cut algorithm. First, the growth exponent β is extracted from the variance of the minimal cut as a function of time, which exhibits the expected scaling $\Delta\mathcal{N} \sim t^{1/3}$, in agreement with KPZ theory. This behavior is illustrated in Fig. 3.1(a).

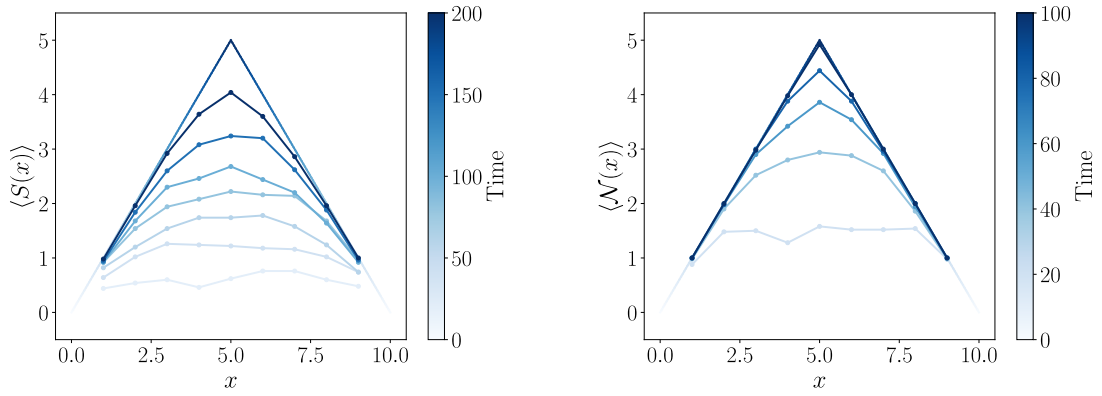


Figure 3.2: Pyramidal entanglement profile for a chain of $L = 10$ qubits, obtained from an ensemble of 50 random circuits. (a) Entanglement entropy $\langle S(x,t) \rangle$ and (b) minimal cut $\langle \mathcal{N}(x,t) \rangle$ are shown as functions of position x for circuit depths $t = 20, 40, 60, 80$, where the minimal cut has already saturated while the entropy continues to grow. Results at longer times $t = 100, 150, 200$ further illustrate the growth of entropy.

Next, we examine the spatial variation of the minimal cut $\mathcal{N}(x,t)$ as a function of the bipartition position x for a fixed circuit depth t. For a chain of $L = 100$ qubits, this analysis reveals that the spatial fluctuations scale as $\xi \sim t^{2/3}$, as shown in Fig. 3.1(b). These results validate our algorithm, as they reproduce the established KPZ scaling behavior.

Continuing with the analysis of the spatial dependence of the minimal cut, scanning over the bipartition position x reveals the "pyramidal" spatial profile of both the entanglement entropy $S_\alpha(x,t)$[1] and the minimal cut $\mathcal{N}(x,t)$. Both quantities reach their maximum at $x = L/2$ and decay linearly toward the boundaries, as introduced earlier in Sec. 2.1.1. This behavior is shown in Fig. 3.2 for both the minimal cut and the entropy.

Temporal dependence of the minimal cut

Since the spatial profile is nearly flat (apart from KPZ fluctuations) until saturation, we restrict our analysis to the central bipartition at $x = L/2$[2] in order to study the time dependence of $S(t) \equiv S(L/2,t)$ and $\mathcal{N}(t) \equiv \mathcal{N}(L/2,t)$. We present numerical results for system sizes ranging from $L = 3$ to $L = 10$ qubits (see Fig. 3.3). The exponential cost of full entanglement simulations limits them to $L = 10$ qubits, by contrast, the minimal cut approach scales only polynomially with L, enabling the simulation of systems of increased size.

From Fig. 3.3, it is evident—as expected—that the minimal cut provides an upper bound to the entanglement entropy, *i.e.*, $S(t) \leq \mathcal{N}(t)$. Moreover, the temporal evolution of both $S(t)$

[1] We restrict to $\alpha = 1$, the von Neumann entropy ($S_1(x,t) \equiv S(x,t)$), the standard measure of bipartite entanglement. This choice does not reduce the generality of our results.

[2] For odd L, the integer floor of $L/2$ is used.

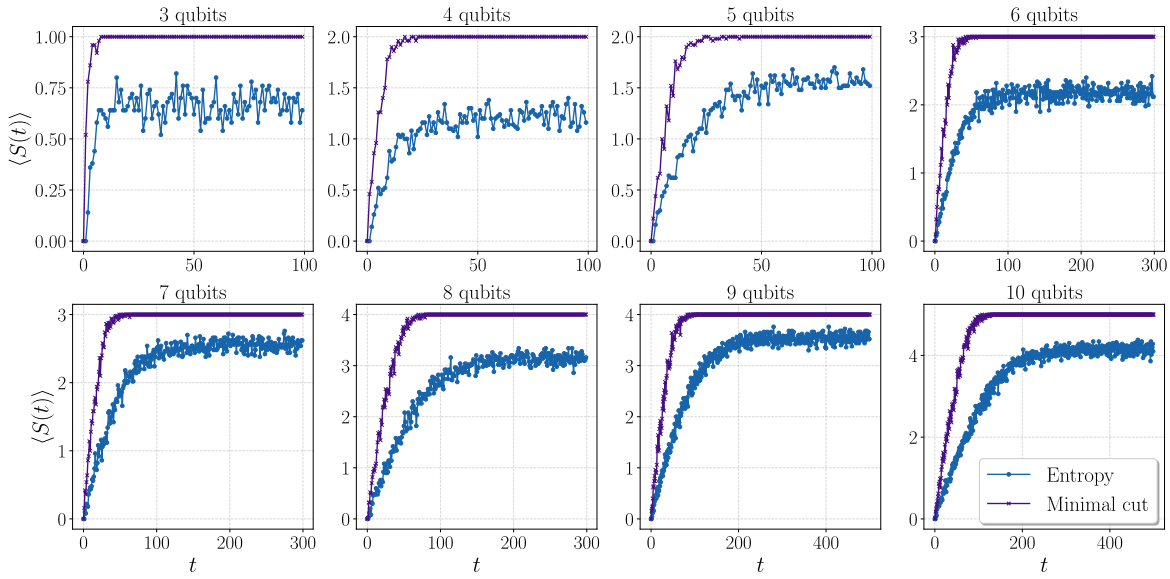


Figure 3.3: Time evolution of the entanglement entropy and the minimal cut at the central bipartition $x = L/2$, for system sizes ranging from $L = 3$ to $L = 10$ qubits. Each curve corresponds to an average over an ensemble of 100 uniformly distributed random Clifford circuits.

and $\mathcal{N}(t)$ reveals two distinct regimes, as discussed in the previous sections: an initial linear growth with slope v_{E} (the entanglement velocity), followed by saturation around $L/2$. The minimal cut $\mathcal{N}(t)$ saturates sharply to exactly $L/2$ due to its discrete, integer-valued, and combinatorial definition. In contrast, the entropy $S(t)$ exhibits residual fluctuations near saturation, reflecting its continuous nature and sensitivity to subleading finite-size effects. Nevertheless, this fluctuations is expected to decrease with increasing system size, as such effects become less relevant in the thermodynamic limit, while still maintaining a finite gap from the maximum entropy. This distinction accounts for the small but persistent deviation observed in simulations with a limited number of qubits, where the ensemble-averaged entanglement entropy $\langle S(t)\rangle$ does not achieve the maximum saturation value exactly.

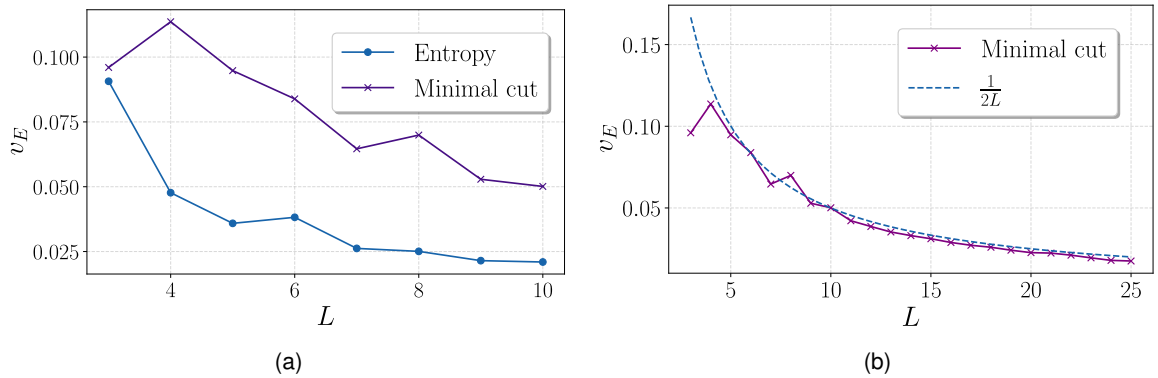
Entanglement velocity

To quantify the growth rate, we compute the entanglement velocity by performing a linear regression up to the saturation point, both for the entropy and for the minimal cut. As shown in Fig. 3.4(a), we find $v_{\mathrm{E}}^{(S)} \leq v_{\mathrm{E}}^{(\mathcal{N})}$, a direct consequence of the inequality $S(t) \leq \mathcal{N}(t)$. Therefore, the velocity of the minimal cut also provides an upper bound for the entanglement velocity.

In the thermodynamic limit $L \to \infty$, we expect $v_{\mathrm{E}}^{(\mathcal{N})}$ and $v_{\mathrm{E}}^{(S)}$ to converge. In practice, this allows us to study $v_{\mathrm{E}}^{(\mathcal{N})} \equiv v_{\mathrm{E}}$ for large L using only minimal cut simulations, bypassing the costly full quantum evolution. Using this approach, we simulate $\mathcal{N}(t)$ for L up to 25 qubits (Fig. 3.5).

The extracted velocity, shown in Fig. 3.4(b), scales with L as

$$v_{\mathrm{E}}(L) = \frac{1}{2L}, \tag{3.1}$$



(a) (b)

Figure 3.4: Entanglement velocity v_{E} extracted from (a) entanglement entropy and minimal cut obtained via linear regression of $S(t)$ and $\mathcal{N}(t)$ versus depth t up to the saturation point, and (b) minimal cut for larger system sizes.

thereby confirming the expectation from the cut growth model discussed in Sec. 2.1.2. This scaling implies that the saturation time for a bipartition at position x is,

$$t_{\mathrm{sat}} = \frac{x}{v_{\mathrm{E}}} = 2Lx. \tag{3.2}$$

For the central cut $x = L/2$, this yields $t_{\mathrm{sat}} = L^2$, indicating that saturation occurs when the total number of two-qubit gates applied is proportional to L^2. This defines an effective "area" of the circuit that leads to a maximal entanglement scaling as L^3.

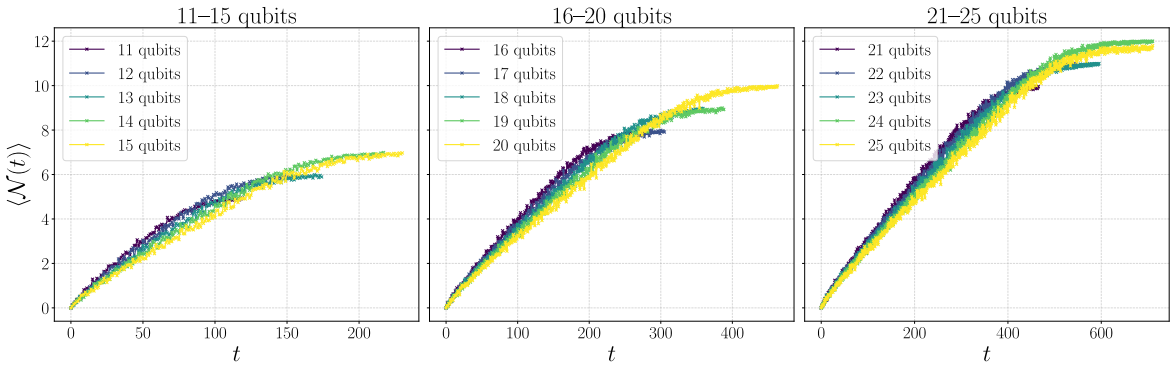


Figure 3.5: Time evolution of the minimal cut for systems with $L = 11$ up to $L = 25$ qubits, considering the bipartition at $x = L/2$, averaged over an ensemble of 100 uniformly distributed random circuits. Each system is evolved well beyond its saturation point to confirm that the entanglement has fully saturated. However, for clarity, the plots are shown only up to the saturation time, as later evolution adds no further insight. A linear regime prior to saturation is observed, with a gradual decrease in the slope as the number of qubits increases.

Chapter 4

Conclusions and perspectives

In this work, we established—based on the results presented in Section 3.1—that for random Clifford circuits, the bipartite entanglement between subsystems of sizes x and $L-x$ saturates at a depth $t_{\mathrm{sat}} = 2Lx$, reducing to $t_{\mathrm{sat}} = L^2$ for the central bipartition. This scaling law constitutes one of the main results of the thesis, as it provides a concrete upper bound on the circuit depth required to reach maximal entanglement.

The graph-based algorithm for the minimal cut developed here constitutes one of the main contributions of this work, as it introduces a new and efficient original alternative—with polynomial runtime, in contrast to the exponential cost associated with simulating the full quantum state outside the stabilizer formalism—for computing the Hartley entropy S_0. This capability is particularly valuable not only for the direct study of S_0, but also for analyzing entanglement dynamics through the von Neumann entropy S_1 in the limit of large system size, as developed throughout this work. The algorithm successfully reproduces pyramidal entanglement profiles, recovers the KPZ fluctuation exponents, and demonstrates that the entanglement velocity scales as $v_{\mathrm{E}} \sim 1/2L$.

An important methodological contribution concerns the statistical interpretation of the minimal cut (Section 2.1.2). Throughout this work, we adopted the quenched method (Eq. 2.23), where the average is performed over circuits and paths on the same footing, and only at the end the minimum is taken. This arises naturally within the Markov-chain formulation. In contrast, the annealed average

$$S_0 = \left\langle \min_{\mathcal{P}\in\mathcal{P}[\mathcal{C}]} \mathcal{N}\left[\mathcal{P}_{\mathcal{C}}\right] \right\rangle_{\mathcal{U}_{L,T}}, \tag{4.1}$$

where the minimization is taken over all paths compatible with a given circuit and only afterwards averaged over the ensemble of circuits, treats circuits and paths separately have been left as open problems. Future work should therefore extend the present framework towards annealed statistics to recover the full distribution of the minimal cut.

Another natural extension is to incorporate a finite density of three-qubit gates, which are present in realistic architectures and could substantially modify entanglement growth. Finally, moving beyond Clifford dynamics, it will be essential to perform statistical analyses under the Haar measure, since Haar-random circuits provide the most appropriate ensemble for capturing the genuinely quantum features of entanglement statistics [2, 62, 63]. Exploring this direction will enable a deeper understanding of entanglement growth and may improve upper bounds for

quantum entanglement, thereby offering greater control over its generic behavior.

Appendices

A Proofs of key properties of Rényi entropy

The following derivations follow standard arguments for Rényi entropies, which can be found in [12].

Proof of bounded Rényi entropy:

Since the rank of ρ_A equals the number of non-zero eigenvalues $\{\lambda_i\}$, and the Rényi entropy for $\alpha > 0$ is defined as:

$$S_\alpha(\rho_A) = \frac{1}{1-\alpha} \log_2 \left(\sum_i \lambda_i^\alpha \right), \tag{2}$$

the function $\sum_i \lambda_i^\alpha$ reaches its maximum value when the eigenvalues are uniformly distributed, i.e., $\lambda_i = 1/r$ for r non-zero terms. Substituting:

$$\sum_i \lambda_i^\alpha = r \left(\frac{1}{r} \right)^\alpha = r^{1-\alpha}, \tag{3}$$

we obtain:

$$S_\alpha(\rho_A) = \frac{1}{1-\alpha} \log_2(r^{1-\alpha}) = \log_2 r = S_0(\rho_A), \tag{4}$$

which proves that:

$$S_\alpha(\rho_A) \leq S_0(\rho_A). \tag{5}$$

Proof of the $\alpha \to \infty$ limit:

As $\alpha \to \infty$, the sum $\sum_i \lambda_i^\alpha$ becomes dominated by the largest eigenvalue $\lambda_{\max}$, since all other terms become negligible in comparison. Thus:

$$\lim_{\alpha \to \infty} S_\alpha(\rho_A) = \lim_{\alpha \to \infty} \frac{1}{1-\alpha} \log_2 \left(\lambda_{\max}^\alpha \right) = \log_2 \left(1/\lambda_{\max} \right) = S_\infty(\rho_A). \tag{6}$$

This entropy S_∞ is often called the min-entropy, and in fact it naturally serves as a lower bound for all Rényi entropies with $\alpha > 1$, as can be shown through the following inequalities:

For $\alpha > 1$, we have $\lambda_i^\alpha \le \lambda_{\max}^{\alpha-1}\lambda_i$. Summing over i and using $\sum_i \lambda_i = 1$ yields

$$\operatorname{Tr}\rho_A^\alpha = \sum_i \lambda_i^\alpha \le \lambda_{\max}^{\alpha-1}. \tag{7}$$

Since $1-\alpha < 0$, applying $\log$ flips the inequality:

$$S_\alpha(\rho_A) = \frac{1}{1-\alpha}\log_2\big(\operatorname{Tr}\rho_A^\alpha\big) \ge \frac{1}{1-\alpha}\log_2\big(\lambda_{\max}^{\alpha-1}\big) = -\log_2\lambda_{\max} = S_\infty(\rho_A). \tag{8}$$

For the opposite direction, $\operatorname{Tr}\rho_A^\alpha \ge \lambda_{\max}^\alpha$, hence

$$S_\alpha(\rho_A) \le \frac{1}{1-\alpha}\log_2\big(\lambda_{\max}^\alpha\big) = -\frac{\alpha}{\alpha-1}\log_2\lambda_{\max} = \frac{\alpha}{\alpha-1}\,S_\infty(\rho_A). \tag{9}$$

Combining both inequalities proves the two-sided bound for all $\alpha > 1$.

$$S_\infty(\rho_A) \le S_\alpha(\rho_A) \le \frac{\alpha}{\alpha-1}\,S_\infty(\rho_A), \qquad \forall\,\alpha > 1. \tag{10}$$

B Minimal cut proof

Operator factorization across a cut

Fix a bipartition $A|B$ at position x at the final layer t, and let a cut through the circuit intersect N_{cut} internal legs. Denote by $\mathcal{H}_A$ and $\mathcal{H}_B$ the output Hilbert spaces. Along any vertical leg that is intersected $m \ge 1$ times, order the intersections from bottom (input boundary) to top (output boundary). Define two disjoint sets of shared indices by parity along each leg: the odd ("lower") intersections form $\mathcal{S}^\downarrow$ and the even ("upper") intersections form $\mathcal{S}^\uparrow$. The shared Hilbert space then factorizes as

$$\mathcal{H}_{\text{sh}} = \mathcal{H}_{\text{sh}}^\downarrow \otimes \mathcal{H}_{\text{sh}}^\uparrow, \qquad \mathcal{H}_{\text{sh}}^\downarrow := \bigotimes_{s\in\mathcal{S}^\downarrow}\mathcal{H}_s, \quad \mathcal{H}_{\text{sh}}^\uparrow := \bigotimes_{s\in\mathcal{S}^\uparrow}\mathcal{H}_s.$$

By construction, intersections alternate the ownership of the leg between the left and right subnetworks; consequently the odd (even) intersections serve as outputs (inputs) of the left subcircuit and as inputs (outputs) of the right subcircuit.

With this convention, the operator defined by the left subcircuit is

$$U_A:\ \mathcal{H}_{A,\text{left}}^{(\text{in})} \otimes \mathcal{H}_{\text{sh}}^\uparrow \ \longrightarrow\ \mathcal{H}_A \otimes \mathcal{H}_{\text{sh}}^\downarrow,$$

while the right subcircuit defines

$$U_B:\ \mathcal{H}_{B,\text{right}}^{(\text{in})} \otimes \mathcal{H}_{\text{sh}}^\downarrow \ \longrightarrow\ \mathcal{H}_B \otimes \mathcal{H}_{\text{sh}}^\uparrow.$$

(If a leg is intersected only once, that intersection belongs to $\mathcal{S}^\downarrow$ by convention, so $\mathcal{H}_{\text{sh}}^\uparrow$ is trivial.)

For an initial product state $|\psi_0\rangle$ at time 0 (supported on $\mathcal{H}^{(\mathrm{in})}_{A,\mathrm{left}} \otimes \mathcal{H}^{(\mathrm{in})}_{B,\mathrm{right}}$), the final state admits the decomposition

$$|\psi(t)\rangle = \sum_{\sigma^{\downarrow}_{\mathrm{sh}},\, \sigma^{\uparrow}_{\mathrm{sh}}} \left(\sum_{\sigma_A} a\big(\sigma_A, \sigma^{\downarrow}_{\mathrm{sh}} \,\big|\, \sigma^{\uparrow}_{\mathrm{sh}}\big)\ |\sigma_A\rangle \right) \otimes \left(\sum_{\sigma_B} b\big(\sigma_B, \sigma^{\uparrow}_{\mathrm{sh}} \,\big|\, \sigma^{\downarrow}_{\mathrm{sh}}\big)\ |\sigma_B\rangle \right), \tag{11}$$

with coefficients given explicitly by the matrix elements of U_A and U_B:

$$\begin{aligned} a\big(\sigma_A, \sigma^{\downarrow}_{\mathrm{sh}} \,\big|\, \sigma^{\uparrow}_{\mathrm{sh}}\big) &= \ \langle \sigma_A, \sigma^{\downarrow}_{\mathrm{sh}}|\ U_A\, |\psi_{A,0}, \sigma^{\uparrow}_{\mathrm{sh}}\rangle\,, \\ b\big(\sigma_B, \sigma^{\uparrow}_{\mathrm{sh}} \,\big|\, \sigma^{\downarrow}_{\mathrm{sh}}\big) &= \ \langle \sigma_B, \sigma^{\uparrow}_{\mathrm{sh}}|\ U_B\, |\psi_{B,0}, \sigma^{\downarrow}_{\mathrm{sh}}\rangle\,. \end{aligned}$$

Here $\{|\sigma_A\rangle\}$ and $\{|\sigma_B\rangle\}$ are computational bases for $\mathcal{H}_A$ and $\mathcal{H}_B$, and $\{|\sigma^{\downarrow}_{\mathrm{sh}}\rangle\}$, $\{|\sigma^{\uparrow}_{\mathrm{sh}}\rangle\}$ are bases for $\mathcal{H}^{\downarrow}_{\mathrm{sh}}$ and $\mathcal{H}^{\uparrow}_{\mathrm{sh}}$, respectively. This is the generic, architecture-independent version of the construction, illustrated by an example in Fig 2.1.

Classical bound from the cut

The shared index $\sigma_{\mathrm{sh}} = \sigma^{\downarrow}_{\mathrm{sh}} \cup \sigma^{\uparrow}_{\mathrm{sh}}$ takes $q^{N_{\mathrm{cut}}}$ values, so the above (Eq. 11) is a decomposition of $|\psi(t)\rangle$ into at most $q^{N_{\mathrm{cut}}}$ product terms across $A|B$. Therefore the Schmidt rank obeys [1]

$$\mathrm{rank}(\rho_A) \le q^{N_{\mathrm{cut}}} \quad \Rightarrow \quad S_\alpha \le S_0 \le N_{\mathrm{cut}} \log q \quad (\forall\, \alpha \ge 0). \tag{12}$$

Minimizing over all admissible cuts yields the *minimal-cut* bound

$$S_\alpha(x,t) \le N_{\min\ \mathrm{cut}}(x,t) \log q \quad (\forall\, \alpha \ge 0). \tag{13}$$

Random circuits

Let $S_0(x,t)$ denote the Hartley entropy across bond x at layer t, and let $N_{\min\ \mathrm{cut}}(x,t)$ be the length of the minimal cut ending at (x,t). For brickwork (or any local, layer-by-layer) circuits with two-site gates, the following two statements hold:

(i) (***Update rule for*** S_0.) If the gate at layer $t \to t{+}1$ acts across bond x, then for a Haar-typical choice of the two-site unitary (indeed, in the specific case of random Clifford circuits)

$$S_0(x, t{+}1) = \min\{S_0(x{-}1, t),\ S_0(x{+}1, t)\} + 1. \tag{14}$$

Otherwise $S_0(y, t{+}1) = S_0(y,t)$ for $y \neq x$. This expresses that S_0 increases to the maximum allowed by subadditivity, and is a standard fact for random local circuits.

(ii) (***Update rule for the minimal cut***.) Under the same gate placement, the minimal-cut length obeys the identical recursion

$$N_{\min\ \mathrm{cut}}(x, t{+}1) = \min\{N_{\min\ \mathrm{cut}}(x{-}1, t),\ N_{\min\ \mathrm{cut}}(x{+}1, t)\} + 1, \tag{15}$$

with $N_{\min\ \mathrm{cut}}(y, t{+}1) = N_{\min\ \mathrm{cut}}(y,t)$ for $y \neq x$.

[1] the logarithms are taken base q (the local Hilbert-space dimension). With this choice, the Hartley entropy is $S_0 = \log_q \mathrm{rank}(\rho_A)$ and has units of "qudits".

Both recursions start from the same initial condition $S_0(x,0) = N_{\min\text{ cut}}(x,0) = 0$ for a product input state. Therefore, by induction on t,

$$S_0(x,t) = N_{\min\text{ cut}}(x,t) \quad \text{for all } x,t \text{ (almost surely for random gates).} \tag{16}$$

In particular, for qubits ($q = 2$),

$$S_\alpha(x,t) \;\le\; S_0(x,t) \;=\; N_{\min\text{ cut}}(x,t). \tag{17}$$

This establishes not only the upper bound but also its saturation for S_0 in generic local random circuits, without relying on any specific example.

Remark. In large local dimension q, higher Rényi entropies also coincide with S_0, and the same minimal-cut value controls S_α as well.

Note. The material and proofs in this section are adapted from the lecture notes by Brian Skinner [15].

C Derivations for the cut growth model

C.1 Cut-increase distribution

Starting from the transition matrix in Eq. (2.9), we can sum over all possible positions x for a fixed t, obtaining:

$$\begin{aligned}\sum_x p(x,t+1) &= p\left(\tfrac{L}{2}-1,t\right)\left[(1-g)+g\right] + \sum_{x=-L/2+2}^{L/2-2} p(x,t)\left(\tfrac{1}{2}g+(1-g)+\tfrac{1}{2}g\right)\\ &\quad + p\left(-\tfrac{L}{2}+1,t\right)\left[g+(1-g)\right]\\ &= \sum_x p(x,t).\end{aligned}$$

Since at $t = 0$ we impose the initial condition that the walker starts at the center of the circuit, $p(x,0) = \delta_{x,0}$, we have $\sum_x p(x,0) = 1$. Combining this with the previous result, we conclude that the classical particle following the worldline remains localized in the probability distribution at all times:

$$\sum_x p(x,t) = 1, \quad \forall t. \tag{18}$$

Continuing with the cut-increase distribution, defined as $\mathcal{Q}(t) = \text{Prob}(\Delta\mathcal{N}(t) = 1)$, we note that the probability of increasing the cut by one is given by the probability of having a gate at (x,t) times the probability of being at (x,t). Since we take $g(x,t) = g$ to be constant,

$$\mathcal{Q}(t) = \sum_x g(x,t)p(x,t) = g\sum_x p(x,t) = g,$$

which follows directly from Eq. (18). Consequently, the moments follow trivially as powers of g:

$$\overline{\mathcal{Q}^m(t)} = \frac{1}{t}\sum_{\tau=0}^{t}\mathcal{Q}^m(\tau) = g^m.$$

C.2 Cut number distribution

Since $\varrho(\Delta\mathcal{N}(t))$ is independent and *i.i.d.* as defined in Eq. (2.15), and given $\mathcal{N}(t) = \sum_{\tau=0}^{t}\Delta\mathcal{N}(\tau)$, the distribution $P[\mathcal{N}(t)]$ is obtained through convolution.

$$P[\mathcal{N}(t)] = \prod_{\tau}^{t}\varrho(\Delta\mathcal{N}(\tau)) = \prod_{\tau}^{t}\left[\mathcal{Q}(\tau)\delta(\Delta\mathcal{N}(\tau)-1) + (1-\mathcal{Q}(\tau))\delta(\Delta\mathcal{N}(\tau))\right],$$

using the integral representation of the delta function $\delta(\Delta\mathcal{N}(t)) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{i}k\Delta\mathcal{N}(t)}dk$, and mapping $z \to \exp(\mathrm{i}k)$, we obtain:

$$P[\mathcal{N}(t)] = \prod_{\tau}^{t}\left[\frac{1}{2\pi}\oint_{S^1}\left(\mathcal{Q}(\tau)\,z^{\Delta\mathcal{N}(\tau)-1} + (1-\mathcal{Q}(\tau))\,z^{\Delta\mathcal{N}(\tau)}\right)\frac{dz}{\mathrm{i}z}\right], \tag{19}$$

where we have performed a complex contour integral over the unit circle S^1. Distributing certain terms and using the relation $\Delta\mathcal{N}(\tau) = \mathcal{N}(\tau+1) - \mathcal{N}(\tau)$, the powers of z cancel except at $\tau = t$, which allows us to change the order between the integrand and the product, yielding:

$$P[\mathcal{N}(t)] = \oint_{S^1} z^{\mathcal{N}(t)}\prod_{\tau}^{t}\frac{1}{2\pi}\left([1-\mathcal{Q}(\tau)] + \frac{\mathcal{Q}(\tau)}{z}\right)\frac{dz}{\mathrm{i}z}, \tag{20}$$

where the product inside the integral can be further expanded as

$$\prod_{\tau}^{t}\left([1-\mathcal{Q}(\tau)] + \frac{\mathcal{Q}(\tau)}{z}\right) = \prod_{\tau}^{t}(1-\mathcal{Q}(\tau))\sum_{\tau}^{t}\frac{F_\tau}{\tau!}z^{-\tau},$$

with

$$F_\tau = \sum_{m_1\neq m_2\neq\ldots\neq m_\tau}\frac{\mathcal{Q}(m_1)\mathcal{Q}(m_2)\ldots\mathcal{Q}(m_\tau)}{(1-\mathcal{Q}(m_1))(1-\mathcal{Q}(m_2))\ldots(1-\mathcal{Q}(m_\tau))} \approx \left(\sum_{m=1}^{t}\frac{\mathcal{Q}(m)}{1-\mathcal{Q}(m)}\right)^{\tau}. \tag{21}$$

Here we have used the fact that for large circuit depth (correspondingly large τ), diagonal terms become sub-leading. After inserting (21) into (20), we obtain:

$$P[\mathcal{N}(t)] = \prod_{\tau}^{t}(1-\mathcal{Q}(\tau))\frac{1}{2\pi}\oint_{S^1}\sum_{\tau}^{t}\left[\frac{z^{\mathcal{N}(t)-\tau-1}}{\tau!}\left(\sum_{m=1}^{t}\frac{\mathcal{Q}(m)}{1-\mathcal{Q}(m)}\right)^{\tau}\right]\frac{dz}{\mathrm{i}},$$

which, upon expanding into a Laurent series, corresponds to extracting the coefficient a_{-1}, *i.e.* requiring $\mathcal{N}(t) - \tau - 1 = -1$. This leads to the result:

$$P\left[\mathcal{N}(t)\right] = \frac{1}{\mathcal{N}(t)!}\left(\sum_{\tau=1}^{t}\frac{\mathcal{Q}(\tau)}{1-\mathcal{Q}(\tau)}\right)^{\mathcal{N}(t)}\prod_{\tau=1}^{t}(1-\mathcal{Q}(\tau)). \tag{22}$$

C.2.1 Gaussian approximation

Starting from integral (19), substituting $z = \mathrm{e}^{\mathrm{i}k}$ gives

$$P(\mathcal{N}(t)) = \int_{\mathbb{R}} \exp\left(\mathrm{i}k\mathcal{N}(t) + \sum_{\tau}^{t} \log\left(1 + \mathcal{Q}(\tau)(\,\mathrm{e}^{-\mathrm{i}k} - 1)\right)\right) dk.$$

For large t, only small k significantly contribute. We therefore expand $\log(1 + \mathcal{Q}(t)(\,\mathrm{e}^{-\mathrm{i}k} - 1))$ and $\mathrm{e}^{-\mathrm{i}k} - 1$ up to second order. The integral then becomes

$$P(\mathcal{N}(t)) \approx \int_{\mathbb{R}} \exp\left(\mathrm{i}k\Big(\mathcal{N}(t) - \sum_{\tau}^{t} \mathcal{Q}(\tau)\Big) - \tfrac{k^2}{2}\Big(\sum_{\tau}^{t} Q(\tau) - \sum_{\tau}^{t} \mathcal{Q}^2(\tau)\Big)\right) dk.$$

Introducing the moment representation $\overline{\mathcal{Q}^m}(t) = \frac{1}{t}\sum_{\tau=1}^{t} \mathcal{Q}^m(\tau)$, we can rewrite

$$P(\mathcal{N}(t)) \approx \int_{\mathbb{R}} \exp\left(\mathrm{i}k\Big(\mathcal{N}(t) - t\overline{\mathcal{Q}}(t)\Big) - \tfrac{k^2}{2}\Big(t\overline{\mathcal{Q}}(t) - t\overline{\mathcal{Q}^2}(t)\Big)\right) dk,$$

which corresponds to the Fourier transform of a displaced Gaussian. Therefore,

$$P(\mathcal{N}(t)) \approx \exp\left(-\frac{1}{2t}\frac{(\mathcal{N}(t) - t\overline{\mathcal{Q}}(t))^2}{\overline{\mathcal{Q}}(t) - \overline{\mathcal{Q}^2}(t)}\right). \tag{23}$$

C.2.2 Binomial distribution

Since in our case $\mathcal{Q}(t)$ is constant, we can revisit Eq. (22), obtaining

$$P\left[\mathcal{N}(t)\right] = \frac{(t\mathcal{Q})^{\mathcal{N}(t)}}{\mathcal{N}(t)!}(1-\mathcal{Q})^{t-\mathcal{N}(t)} = \binom{t}{\mathcal{N}(t)}\mathcal{Q}^{\mathcal{N}(t)}(1-\mathcal{Q})^{t-\mathcal{N}(t)}\frac{t^{\mathcal{N}(t)}(t-\mathcal{N}(t))!}{t!},$$

where the last term $\frac{t^{\mathcal{N}(t)}(t-\mathcal{N}(t))!}{t!} \approx 1$ for large t, leaving the binomial distribution. Substituting $\mathcal{Q} = \frac{1}{L-1}$ we obtain:

$$P[\mathcal{N}(T)] = \binom{t}{\mathcal{N}(t)}(1-\ \mathcal{Q})^{t-\mathcal{N}(t)}\ \mathcal{Q}^{\mathcal{N}(t)} \to \binom{t}{\mathcal{N}(t)}\left(1-\frac{1}{L-1}\right)^{t-\mathcal{N}(t)}\left(\frac{1}{L-1}\right)^{\mathcal{N}(t)} \tag{24}$$

Bibliography

[1] Adam Nahum, Jonathan Ruhman, Sagar Vijay, and Jeongwan Haah. *Quantum Entanglement Growth under Random Unitary Dynamics*. *Phys. Rev. X* **7**, 031016 (2017). DOI: 10.1103/PhysRevX.7.031016.

[2] Matthew P.A. Fisher, Vedika Khemani, Adam Nahum, and Sagar Vijay. *Random Quantum Circuits*. *Annu. Rev. Condens. Matter Phys.* **14**(Volume 14, 2023), 335–379 (2023). DOI: https://doi.org/10.1146/annurev-conmatphys-031720-030658.

[3] A. Einstein, B. Podolsky, and N. Rosen. *Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?* *Phys. Rev.* **47**, 777–780 (1935). DOI: 10.1103/PhysRev.47.777.

[4] Anton Zeilinger. *Quantum entanglement: a fundamental concept finding its applications*. *Physica Scripta* **1998**(T76), 203 (1998). DOI: 10.1238/Physica.Topical.076a00203.

[5] Ryszard Horodecki, Paweł Horodecki, Michał Horodecki, and Karol Horodecki. *Quantum entanglement*. *Rev. Mod. Phys.* **81**, 865–942 (2009). DOI: 10.1103/RevModPhys.81.865.

[6] Zhi-Kang Lin, Yao Zhou, Bin Jiang, Bing-Quan Wu, Li-Mei Chen, Xiao-Yu Liu, Li-Wei Wang, Peng Ye, and Jian-Hua Jiang. *Measuring entanglement entropy and its topological signature for phononic systems*. *Nature Comm.* **15**(1), 1601 (2024). DOI: 10.1038/s41467-024-45887-8.

[7] Matthew B. Hastings, Iván González, Ann B. Kallin, and Roger G. Melko. *Measuring Renyi Entanglement Entropy in Quantum Monte Carlo Simulations*. *Phys. Rev. Lett.* **104**, 157201 (2010). DOI: 10.1103/PhysRevLett.104.157201.

[8] G. Vidal and R. F. Werner. *Computable measure of entanglement*. *Phys. Rev. A* **65**, 032314 (2002). DOI: 10.1103/PhysRevA.65.032314.

[9] Sumiyoshi Abe and A. K. Rajagopal. *Quantum entanglement inferred by the principle of maximum nonadditive entropy*. *Phys. Rev. A* **60**, 3461–3466 (1999). DOI: 10.1103/PhysRevA.60.3461.

[10] Ryszard Horodecki and Michal/ Horodecki. *Information-theoretic aspects of inseparability of mixed states*. *Phys. Rev. A* **54**, 1838–1843 (1996). DOI: 10.1103/PhysRevA.54.1838.

[11] Benjamin Schumacher. *Quantum coding*. *Phys. Rev. A* **51**, 2738–2747 (1995). DOI: 10.1103/PhysRevA.51.2738.

[12] Alfréd Rényi. *On measures of entropy and information*. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability, volume 1: contributions to the theory of statistics*, volume 4, pages 547–562. University of California Press (1961).

[13] Rajibul Islam, Ruichao Ma, Philipp M Preiss, M Eric Tai, Alexander Lukin, Matthew Rispoli, and Markus Greiner. *Measuring entanglement entropy in a quantum many-body system*. *Nature* **528**(7580), 77–83 (2015). DOI: 10.1038/nature15750.

[14] John von Neumann and Robert T. Beyer. *Mathematical Foundations of Quantum Mechanics: New Edition*. Princeton University Press, ned - new edition edition (2018).

[15] B Skinner. *Introduction to random unitary circuits and the measurement-induced entanglement phase transition*. (2023). arXiv:2307.02986, DOI: 10.48550/arXiv.2307.02986.

[16] Brian Skinner, Jonathan Ruhman, and Adam Nahum. *Measurement-Induced Phase Transitions in the Dynamics of Entanglement*. *Phys. Rev. X* **9**, 031009 (2019). DOI: 10.1103/PhysRevX.9.031009.

[17] Charles H. Bennett and Gilles Brassard. *Quantum cryptography: Public key distribution and coin tossing*. *Theor. Comput. Sci.* **560**, 7–11 (2014). Theoretical Aspects of Quantum Cryptography – celebrating 30 years of BB84, DOI: https://doi.org/10.1016/j.tcs.2014.05.025.

[18] Charles H. Bennett, Gilles Brassard, Claude Crépeau, Richard Jozsa, Asher Peres, and William K. Wootters. *Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels*. *Phys. Rev. Lett.* **70**, 1895–1899 (1993). DOI: 10.1103/PhysRevLett.70.1895.

[19] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. *Quantum-Enhanced Measurements: Beating the Standard Quantum Limit*. *Science* **306**(5700), 1330–1336 (2004). DOI: 10.1126/science.1104149.

[20] M.A. Nielsen and I.L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press (2010).

[21] Leonid Gurvits. *Classical deterministic complexity of Edmonds' Problem and quantum entanglement*. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, STOC '03, page 10–19, New York, NY, USA (2003). Association for Computing Machinery.

[22] Scott Aaronson and Daniel Gottesman. *Improved simulation of stabilizer circuits*. *Phys. Rev. A* **70**, 052328 (2004). DOI: 10.1103/PhysRevA.70.052328.

[23] C. E. Shannon. *A mathematical theory of communication*. *The Bell System Technical Journal* **27**(3), 379–423 (1948). DOI: 10.1002/j.1538-7305.1948.tb01338.x.

[24] John Watrous. *The theory of quantum information*. Cambridge university press (2018).

[25] Noah Linden, Milán Mosonyi, and Andreas Winter. *The structure of Rényi entropic inequalities*. *Proc. R. Soc. A: Math. Phys. Eng. Sci.* **469**(2158), 20120737 (2013). DOI: 10.1098/rspa.2012.0737.

[26] W. Feller. *An Introduction to Probability Theory and Its Applications, Volume 1*. An Introduction to Probability Theory and Its Applications. Wiley (1968).

[27] Madan Lal Mehta. *1 - Introduction*. In MADAN LAL MEHTA, editor, *Random Matrices (Revised and Enlarged Second Edition)*, pages 1–35. Academic Press, San Diego, revised and enlarged second edition edition (1991).

[28] Giacomo Livan, Marcel Novaes, and Pierpaolo Vivo. *Introduction to Random Matrices*. (2018). DOI: 10.1007/978-3-319-70885-0.

[29] Amos Chan, Andrea De Luca, and J. T. Chalker. *Solution of a Minimal Model for Many-Body Quantum Chaos*. *Phys. Rev. X* **8**, 041019 (2018). DOI: 10.1103/PhysRevX.8.041019.

[30] Tianci Zhou and Adam Nahum. *Emergent statistical mechanics of entanglement in random unitary circuits*. *Phys. Rev. B* **99**, 174205 (2019). DOI: 10.1103/PhysRevB.99.174205.

[31] Tianci Zhou and Adam Nahum. *Entanglement Membrane in Chaotic Many-Body Systems*. *Phys. Rev. X* **10**, 031066 (2020). DOI: 10.1103/PhysRevX.10.031066.

[32] Vedika Khemani, Ashvin Vishwanath, and David A. Huse. *Operator Spreading and the Emergence of Dissipative Hydrodynamics under Unitary Evolution with Conservation Laws*. *Phys. Rev. X* **8**, 031057 (2018). DOI: 10.1103/PhysRevX.8.031057.

[33] C. W. von Keyserlingk, Tibor Rakovszky, Frank Pollmann, and S. L. Sondhi. *Operator Hydrodynamics, OTOCs, and Entanglement Growth in Systems without Conservation Laws*. *Phys. Rev. X* **8**, 021013 (2018). DOI: 10.1103/PhysRevX.8.021013.

[34] Márk Mezei and Julio Virrueta. *Exploring the membrane theory of entanglement dynamics*. *Journal of High Energy Physics* **2020**(2), 13 (2020). DOI: 10.1007/JHEP02(2020)013.

[35] Pasquale Calabrese and John Cardy. *Entanglement entropy and conformal field theory*. *J. Phys. A Math. Theor.* **42**(50), 504005 (2009). DOI: 10.1088/1751-8113/42/50/504005.

[36] Hong Liu and S. Josephine Suh. *Entanglement Tsunami: Universal Scaling in Holographic Thermalization*. *Phys. Rev. Lett.* **112**, 011601 (2014). DOI: 10.1103/PhysRevLett.112.011601.

[37] O C O Dahlsten, R Oliveira, and M B Plenio. *The emergence of typical entanglement in two-party random processes*. *J. Phys. A Math. Theor.* **40**(28), 8081 (2007). DOI: 10.1088/1751-8113/40/28/S16.

[38] Hyungwon Kim and David A. Huse. *Ballistic Spreading of Entanglement in a Diffusive Nonintegrable System*. *Phys. Rev. Lett.* **111**, 127205 (2013). DOI: 10.1103/PhysRevLett.111.127205.

[39] Mehran Kardar, Giorgio Parisi, and Yi-Cheng Zhang. *Dynamic Scaling of Growing Interfaces*. *Phys. Rev. Lett.* **56**, 889–892 (1986). DOI: 10.1103/PhysRevLett.56.889.

[40] Jonas Richter, Oliver Lunt, and Arijeet Pal. *Transport and entanglement growth in long-range random Clifford circuits*. *Phys. Rev. Res.* **5**, L012031 (2023). DOI: 10.1103/PhysRevResearch.5.L012031.

[41] Chao-Ming Jian, Yi-Zhuang You, Romain Vasseur, and Andreas W. W. Ludwig. *Measurement-induced criticality in random quantum circuits*. *Phys. Rev. B* **101**, 104302 (2020). DOI: 10.1103/PhysRevB.101.104302.

[42] Yimu Bao, Soonwon Choi, and Ehud Altman. *Theory of the phase transition in random unitary circuits with measurements*. *Phys. Rev. B* **101**, 104301 (2020). DOI: 10.1103/PhysRevB.101.104301.

[43] Grace M. Sommers, Sarang Gopalakrishnan, Michael J. Gullans, and David A. Huse. *Zero-temperature entanglement membranes in quantum circuits*. *Phys. Rev. B* **110**, 064311 (2024). DOI: 10.1103/PhysRevB.110.064311.

[44] Sophus Lie. *Theorie der transformationsgruppen I*. *Mathematische Annalen* **16**(4), 441–528 (1880). DOI: 10.1007/bf01446218.

[45] Hale F. Trotter. *On the product of semi-groups of operators*. *Proc. Am. Math. Soc.* **10**(4), 545–551 (1959). DOI: 10.2307/2033649.

[46] Daniel A. Roberts and Brian Swingle. *Lieb-Robinson Bound and the Butterfly Effect in Quantum Field Theories*. *Phys. Rev. Lett.* **117**, 091602 (2016). DOI: 10.1103/PhysRevLett.117.091602.

[47] Elliott H Lieb and Derek W Robinson. *The finite group velocity of quantum spin systems*. *Commun. Math. Phys.* **28**(3), 251–257 (1972). DOI: 10.1007/BF01645779.

[48] Gushu Li, Yufei Ding, and Yuan Xie. *Tackling the Qubit Mapping Problem for NISQ-Era Quantum Devices*. In *Proceedings of the Twenty-Fourth International Conference on Architectural Support for Programming Languages and Operating Systems*, ASPLOS '19, page 1001–1014, New York, NY, USA (2019). Association for Computing Machinery.

[49] Zi-Wen Liu, Seth Lloyd, Elton Yechao Zhu, and Huangjun Zhu. *Generalized Entanglement Entropies of Quantum Designs*. *Phys. Rev. Lett.* **120**, 130502 (2018). DOI: 10.1103/PhysRevLett.120.130502.

[50] Daniel Gottesman. *The Heisenberg Representation of Quantum Computers*. (1998). arXiv:quant-ph/9807006, DOI: 10.48550/ARXIV.QUANT-PH/9807006.

[51] Maarten Van Den Nest. *Classical simulation of quantum computation, the Gottesman-Knill theorem, and slightly beyond*. *Quantum Info. Comput.* **10**(3), 258–271 (2010). DOI: 10.48550/arXiv.0811.0898.

[52] Brian Swingle. *Entanglement renormalization and holography*. *Phys. Rev. D* **86**, 065007 (2012). DOI: 10.1103/PhysRevD.86.065007.

[53] Claudio Chamon and Eduardo R. Mucciolo. *Virtual Parallel Computing and a Search Algorithm Using Matrix Product States*. *Phys. Rev. Lett.* **109**, 030503 (2012). DOI: 10.1103/PhysRevLett.109.030503.

[54] F. Verstraete, V. Murg, and J. I. Cirac. *Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems*. *Adv. Phys.* **57**(2), 143–224 (2008). DOI: 10.1080/14789940801912366.

[55] Horacio Casini, Hong Liu, and Márk Mezei. *Spread of entanglement and causality*. *J. High Energy Phys.* **2016**(7), 1–63 (2016). DOI: 10.1007/JHEP07(2016)077.

[56] David A. Huse and Christopher L. Henley. *Pinning and Roughening of Domain Walls in Ising Systems Due to Random Impurities*. *Phys. Rev. Lett.* **54**, 2708–2711 (1985). DOI: 10.1103/PhysRevLett.54.2708.

[57] J. M. Kim, M. A. Moore, and A. J. Bray. *Zero-temperature directed polymers in a random potential*. *Phys. Rev. A* **44**, 2345–2351 (1991). DOI: 10.1103/PhysRevA.44.2345.

[58] Yaodong Li, Sagar Vijay, and Matthew P.A. Fisher. *Entanglement Domain Walls in Monitored Quantum Circuits and the Directed Polymer in a Random Environment*. *PRX Quantum* **4**, 010331 (2023). DOI: 10.1103/PRXQuantum.4.010331.

[59] E. W. Dijkstra. *A Note on Two Problems in Connexion with Graphs*, page 287–290. Association for Computing Machinery, New York, NY, USA, 1 edition (2022).

[60] Bob Coecke and Ross Duncan. *Interacting quantum observables: categorical algebra and diagrammatics*. *New J. Phys.* **13**(4), 043016 (2011). DOI: 10.1088/1367-2630/13/4/043016.

[61] Telikepalli Kavitha, Kurt Mehlhorn, Dimitrios Michail, and Katarzyna E Paluch. *An algorithm for minimum cycle basis of graphs*. *Algorithmica* **52**(3), 333–349 (2008). DOI: 10.1007/s00453-007-9064-z.

[62] Oscar C O Dahlsten, Cosmo Lupo, Stefano Mancini, and Alessio Serafini. *Entanglement typicality*. *J. Phys. A Math. Theor* **47**(36), 363001 (2014). DOI: 10.1088/1751-8113/47/36/363001.

[63] Bill Fefferman, Soumik Ghosh, and Wei Zhan. *Anti-Concentration for the Unitary Haar Measure and Applications to Random Quantum Circuits*. (2024). arXiv:2407.19561, DOI: 10.48550/ARXIV.2407.19561.