Randomness in Quantum Information and Many-Body Physics

Xhek Turkeshi* **and Markus Heinrich†**

Institute for Theoretical Physics, University of Cologne, Zülpicherstraße 77a, 50937 Köln, Germany

* xturkesh@uni-koeln.de , † markus.heinrich@uni-koeln.de

Abstract

These lecture notes develop a self-contained route from Haar-random unitaries to quantitative diagnostics of randomness in quantum many-body systems. As major tools, we introduce permutation representations, graphical replica calculus, Weingarten calculus, and unitary designs. We then apply these to understand classical shadows as means to measure properties of many-body states. Random quantum circuits are introduced to analyze the emergence of quantum randomness from local random dynamics. We concretely study anticoncentration, magic, entanglement growth, operator spreading, and out-of-time-ordered correlators. The presentation targets Master's students with minimum prerequisities; concepts are introduced when needed and companying exercises are provided throughout.

Contents

1 About this course **2**
1.1 How to use these notes 3
1.2 Global notation 3

2 Computing integrals over the Haar measure **4**
2.1 What is a Haar measure? 4
2.2 Permutations and their combinatorics 8
2.2.1 Permutations and Cycles 8
2.2.2 The Action of Permutations on Quantum States 13
2.3 Weingarten calculus 18
2.4 Approximating the Haar measure: Unitary designs 21
2.5 The Clifford group 22
2.6 Further reading 24
2.7 Additional exercises 24

3 Many-Body Properties from Classical Shadows **27**
3.1 Reconstructing quantum states from measurements 28
3.2 Shadow estimation with randomized measurements 30
3.2.1 Clifford measurements 32
3.2.2 Pauli measurements 36
3.2.3 Beyond linear observables: estimating purity 38
3.3 Further reading 40
3.4 Exercises on classical shadows 40

4 Quantum Randomness from Local Gates **44**

4.1 Random quantum circuits and their moment operators 44
4.2 Convergence of random quantum circuits and approximate designs 48
4.2.1 Limit behaviour of random circuits 48
4.2.2 Notions of approximate unitary designs 49
4.3 Lightcones and lower bounds 51
4.4 Logarithmic-depth designs by gluing 53
4.5 Further reading 54
4.6 Exercises on random circuits and approximate designs 55

5 Random Dynamics in Many-Body Systems 58

References 59

A Some linear algebra 62
A.1 States, operators, superoperators 62
A.2 Non-orthonormal bases 67

B Unsigned Stirling numbers of the first kind 68

1 About this course

Quantum information science is the research field that studies how information is stored, processed, and transmitted when it is governed by the laws of quantum mechanics. It includes areas such as *quantum computing*, *quantum communication, and quantum cryptography*. Although still relatively young, quantum information science has already had a profound impact on many areas of physics, especially on the study of quantum systems composed of many interacting particles, commonly referred to as *many-body quantum systems*.

Crucially, the interplay between quantum information and many-body physics has not only refined our understanding of traditional problems, but has also uncovered entirely new dynamical phases of matter – phases that arise uniquely in *programmable quantum devices*. This so-called *synthetic quantum matter* cannot be characterized by conventional local order parameters such as magnetization or current. Instead, its defining features are quantum informational, such as the structure of entanglement or nonstabilizer (magic state) resources, or the system's ability to preserve quantum information against noise and local errors. Understanding and classifying such phases requires a shift in perspective –from symmetry and energetics to *information content* and *computational complexity*.

Quantum circuits have emerged as essential tools to tackle fundamental questions in quantum dynamics, both traditional and synthetic: from the mechanisms of *thermalization* and the emergence of *statistical mechanics* in isolated systems, to the onset of *quantum chaos* and its deep connections with *black hole physics and holography* in quantum gravity. Studying these phenomena in a concrete setting is notoriously difficult, due to the exponential growth of the Hilbert space and the inherently non-linear structure of quantum correlations. Random dynamics, in particular *random quantum circuits*, offer a powerful way to overcome these challenges: they enable analytical and numerical progress through disorder averaging, while capturing the generic behavior of quantum systems thanks to *typicality arguments*.

At the same time, random quantum circuits have become indispensable tools in quantum computing. They provide efficient and versatile frameworks for a wide range of applications, including the benchmarking and verification of quantum computations, the characterization of noise in experimental platforms, and the estimation of observables via shadow tomography.

They have also been central to demonstrating a *quantum-computational advantage* over classical computations in specific problems, most notably *random circuit sampling*. Far from being purely theoretical constructs, these methods are implemented across various platforms – from superconducting qubits to cold atoms – and are central to the ongoing development of quantum technologies.

This course provides a pedagogical introduction to random unitaries and to several key methods from quantum information theory, with a focus on their application to many-body physics. A substantial part of the course covers research-level topics introduced only in the past few years, offering a unique opportunity to engage with current questions at the interface between two rapidly evolving fields. It is also intended to serve as a solid preparation to pursue research in these areas, starting from the Master's level. The only prerequisities are master-level quantum mechanics and basic linear algebra.

1.1 How to use these notes

We introduce the core randomization concepts in two parts, each followed by a section that immediately connects the developed methods to applications in quantum information and many-body physics. The introduced methods and concepts are illustrated in numerous educational examples. Starred sections contain useful extensions but are not needed for a first reading. Each section ends with a set of exercises, as well as a 'further reading' section providing additional information and references, primarily in the form of original articles. This allows to deepen the learned concepts and to connect the course material to current research in the field. The outline of these notes is as follows:

Sec. 2 first introduces the Haar measure and motivates the study of random quantum systems and their statistics. We give a pedagogical treatment of permutations and a graphical calculus, which will provide the foundation for *Weingarten calculus*, the central toolbox throughout the course. We also define *unitary designs*, which provide efficient approximations of Haar randomness and play an important role in practical implementations.

Sec. 3 treats the *framework of classical shadows*, an efficient way to extract information about quantum states using only a small number of randomized measurements. This technique has already found widespread use in experimental platforms and continues to inspire a growing body of research.

Sec. 4 explores random quantum circuits in detail. These models provide an efficient and physically motivated approach to generate randomness in quantum many-body systems, and they serve as minimal models for chaotic quantum dynamics.

Sec. 5 applies the preceding tools to anticoncentration, Pauli spectra and magic, entanglement growth, operator spreading, and out-of-time-order correlators.

1.2 Global notation

The following conventions are used throughout. Subscripts indicate a subsystem only when this is stated explicitly; otherwise they are ordinary indices. All logarithms are natural unless a base is displayed explicitly.

Symbol	Meaning
$\mathcal{H} = (\mathbb{C}^q)^{\otimes n}$	Hilbert space of n qudits of local dimension q.
d	Total Hibert space dimension: $d = q^n$.
$\mathrm{L}(\mathcal{H})$	The vector space of linear operators on $\mathcal{H}$.
$\mathbb{1}_A$	Identity operator on subsystem or space A; the subscript is omitted when unambiguous.
$\|X\|_p$	Schatten p-norm; in particular $p = 1, 2, \infty$ give the trace, Hilbert–Schmidt, and operator norms.
tr_A	Partial trace over subsystem A; tr without a subscript is the full trace.
$\mathrm{U}(\mathcal{H}) = \mathrm{U}(d)$	The unitary group on $\mathcal{H}$: the group of unitary $d \times d$ matrices.
$\mathrm{Cl}_n(q)$	The Clifford group on n qudits of local (prime) dimension q.
S_k	The symmetric group on k objects.
k	Number of copies/replicas; also order of a moment.
R_π	Operator representing the permutation $\pi \in \mathrm{S}_k$ on $\mathcal{H}^{\otimes k}$.
$\mathsf{F} = R_{(12)}$	Two-copy swap operator; it exchanges the two replica factors.
μ, ν	Probability measures on a compact group G (typically $G = \mathrm{U}(d)$). The normalized Haar measure is denoted by μ_G.
$\mathbb{E}_{U\sim\nu}[\,\cdot\,]$	Expectation over a random unitary sampled from the measure ν.

2 Computing integrals over the Haar measure

In this section, we lay the foundation for the study of random unitaries in quantum information theory and many-body systems. By 'random' we mean mainly *Haar-random*, this is, the unitaries are sampled *uniformly* from the unitary group on our system's Hilbert space. We formally introduce this distribution, the Haar measure, in Sec. 2.1. Eventually, we are interested in statistical quantities – expectation values, correlations, and generally moments of our random unitaries – and will develop tools to compute those within this section. Central to these are permutations and their combinatorics of which we will cover the necessary background in Sec. 2.2. Based on permutations, we then introduce Weingarten calculus in Sec. 2.3 as a way to evaluate statistical quantities. Finally, we address practical problems with implementing Haar-random unitaries in Sec. 2.4 and present unitary designs as means to overcome these obstacles. We then define the Clifford group in Sec. 2.5 as the most important example of such a unitary design.

2.1 What is a Haar measure?

On the real line $\mathbb{R}$, there is a unique measure dx satisfying two simple but fundamental properties

$$\mathrm{d}(x+y) = \mathrm{d}x \qquad \text{(translation invariance)}\,, \qquad \int_0^1 \mathrm{d}x = 1 \qquad \text{(unit volume)}\,. \tag{1}$$

This measure represents what we intuitively mean by a *uniform* distribution: translation invariance ensures that no point is preferred over any other. However, since $\mathbb{R}$ is non-compact, this measure assigns infinite volume to the full space, and thus cannot be normalized to a probability measure. To work around this, we typically restrict to a compact interval such as $[0, 1]$, which provides a natural normalization.

Remarkably, this idea of defining an invariant, uniform measure generalizes to far more abstract settings. Whenever the underlying set carries a suitable group structure, one can often define a natural measure that is invariant under group multiplication. More precisely, if G is a

compact (Hausdorff topological) group then there exists a unique measure dg on G that is both

$$\text{left/right invariant} \qquad \mathrm{d}g = \mathrm{d}(hg) = \mathrm{d}(gh) \tag{2}$$

$$\text{normalized} \qquad \int_G \mathrm{d}g = 1\,. \tag{3}$$

This unique measure is called the *Haar (or uniform) measure* on G.

In this course, our primary interest lies in the unitary group $\mathrm{U}(d)$, which is compact and thus satisfied the condition above. Concretely, the Haar measure dU on $\mathrm{U}(d)$ satisfies

$$\mathrm{d}U = \mathrm{d}(UV) = \mathrm{d}(VU) \qquad \int_{\mathrm{U}(d)} \mathrm{d}U = 1\,. \tag{4}$$

This measure provides a rigorous definition of what it means to sample a unitary matrix uniformly at random. Such samples are typically referred to as being 'Haar-random' in the literature and we will simply write $U \sim \mathrm{U}(d)$ for a random matrix U distributed according to the Haar measure on $\mathrm{U}(d)$.

An important property of the unitary group is its *transitive* action on quantum states: For every two pure quantum states $|\psi\rangle$ and $|\varphi\rangle$, there is a unitary U such that $U|\psi\rangle = |\varphi\rangle$. This also means that every state can be expressed using a reference state, say $|0\rangle$, as $U|0\rangle$. In this way, the Haar measure on $\mathrm{U}(d)$ induces a probability measure on quantum states, samples of which have the form $U|0\rangle$ with $U \sim \mathrm{U}(d)$. In a minor abuse of language, these random quantum states are also referred to as *Haar-random states*.

Haar integrals and moments. As mentioned in Sec. 1, we intend to use random unitaries as a tractable proxy for more complicated dynamics, or as means to accomplish certain non-trivial tasks in quantum information processing. The central ideal is to sample random unitaries and then ask questions about the *average behaviour of our quantum system*. This means that we want to evaluate statistical quantities, i.e. expectation values $\mathbb{E}_{U\sim\mathrm{U}(d)}[f(U)]$ of suitable functions f defined by the system we want to analyze.

A very simple example would be a quantum system initialized in $|\psi\rangle$ which undergoes a Haar-random unitary U and is subsequently measured in a basis $|i\rangle$. The probability of observing outcome i is then $f(U) = |\langle i\,|U|\psi\rangle|^2$, and its expectation value is given by

$$\mathbb{E}_{U\sim\mathrm{U}(d)} f(U) = \int_{\mathrm{U}(d)} \mathrm{tr}(|i\rangle\langle i|U|\psi\rangle\langle\psi|U^\dagger)\mathrm{d}U = \frac{1}{d}\,\mathrm{tr}(|i\rangle\langle i|\mathbb{1}) = \frac{1}{d}\,. \tag{5}$$

Here, we used $\mathbb{E}_{U\sim\mathrm{U}(d)} U|\psi\rangle\langle\psi|U^\dagger = \mathbb{1}/d$, a simple fact that we will show at the end of this section. Hence, on average, the outcome distribution will be a uniform one. To get a better understanding of how strongly the random outcome distribution concentrates around its mean (the uniform distribution), we can now look at the variance of $f(U)$:

$$\mathrm{Var}_{U\sim\mathrm{U}(d)}[f(U)] = \mathbb{E}_{U\sim\mathrm{U}(d)} f(U)^2 - \left(\mathbb{E}_{U\sim\mathrm{U}(d)} f(U)\right)^2 = \int_{\mathrm{U}(d)} \mathrm{tr}(|i\rangle\langle i|U|\psi\rangle\langle\psi|U^\dagger)^2\mathrm{d}U - \frac{1}{d^2}\,. \tag{6}$$

Note that

$$\mathrm{tr}(|i\rangle\langle i|U|\psi\rangle\langle\psi|U^\dagger)^2 = \mathrm{tr}(|i\rangle\langle i|^{\otimes 2} U^{\otimes 2}|\psi\rangle\langle\psi|^{\otimes 2} U^{\otimes 2,\dagger})\,, \tag{7}$$

hence, we would now have to compute $\mathbb{E}_{U\sim\mathrm{U}(d)} U^{\otimes 2}|\psi\rangle\langle\psi|^{\otimes 2} U^{\otimes 2,\dagger}$, involving tensor powers of U. This is slightly less trivial than the previous computation and requires a few more techniques that we will develop in Sec. 2.3. There, we will also show that the resulting value for the

variance is just slightly less than $1/d^2$, showing that the fluctuations of the outcome distribution around the mean are comparably small. We will return to this discussion in **??**.

Another example is given by analyzing the (average) entanglement in the final state $U|\psi\rangle$ accross some bipartition of the system. A measure of entanglement is the purity of the reduced state: $f(U) = \mathrm{tr}(\rho_1^2)$ where $\rho_1 = \mathrm{tr}_2(U|\psi\rangle\langle\psi|U^\dagger)$ is the reduced state on the first subsystem. Clearly, if $U|\psi\rangle$ is maximally entangled, the reduced state is maximally mixed, thus $\mathrm{tr}(\rho_1^2) = 1/d_1$. In contrast, if $U|\psi\rangle$ is a product state, than the reduced state is pure and $\mathrm{tr}(\rho_1^2) = 1$. Interestingly, we can again re-express $f(U)$ using tensor powers of U by exploiting the *swap trick* $\mathrm{tr}(X^2) = \mathrm{tr}(X^{\otimes 2}\mathbb{F})$ where $\mathbb{F}$ is the swap (or flip) operator that exchanges the two copies of our system. We then see that

$$\begin{aligned}\mathrm{tr}(\rho_1^2) &= \mathrm{tr}_{11}\left([\mathrm{tr}_2(U|\psi\rangle\langle\psi|U^\dagger)]^{\otimes 2}\mathbb{F}\right) \\ &= \mathrm{tr}_{11}\left(\mathrm{tr}_{22}[(U|\psi\rangle\langle\psi|U^\dagger)^{\otimes 2}]\mathbb{F}\right) = \mathrm{tr}\left((U|\psi\rangle\langle\psi|U^\dagger)^{\otimes 2}\mathbb{F}\otimes\mathbb{1}_{22}\right), \quad (8)\end{aligned}$$

where we used the defining property of the partical trace, namely $\mathrm{tr}(XO_1\otimes\mathbb{1}_2) = \mathrm{tr}(\mathrm{tr}_2(X)O_1)$, and $\mathbb{F}\otimes\mathbb{1}_{22}$ is the operator that swaps only the two copies of the first subsystem (and leave the other one invariant).

The reason why we can re-express the above functions f using tensor powers of U is because they are *polynomials*. This should remind us of classical probability theory, where the central objects describing a (real-valued) random variable X are its *moments* $\mathbb{E}[X^k]$. Here, the analogous function f is simply the monomial $f(x) = x^k$. The study of moments is of course motivated by the fact that we can expand more complicated (analytical) functions into a Taylor series and thus, the moments essentially capture all the details of the random variable's distribution (at least if it is sufficiently well behaved). For random unitaries, we likewise restrict our attention to polynomial functions $f(U,\bar{U})$ which are complex polynomials in the entries of U and its complex conjugate $\bar{U}$. In this context, the role of monomials is taken by so-called *homogeneous polynomials* of degree k – those for which $f(\lambda U,\lambda\bar{U}) = \lambda^{2k}f(U,\bar{U})$. A simple example of a homogeneous polynomial of degree 3 in two complex variables would be $x^2\bar{y}+y^3$. A standard result in the theory of polynomials says that they can be expressed in terms of symmetric tensors in the variables. Applied to our case, that means that any homogeneous polynomial of degree k is a linear combination of functions of the form

$$p_{A,B}(U,\bar{U}) := \mathrm{tr}(BU^{\otimes k}AU^{\otimes k,\dagger}) \quad (9)$$

We refer to the expectation value of the homogeneous polynomial (9) simply as a k-th moment (of the Haar measure).

Mathematically, the expectation value of (9) can be interpreted as a *matrix element* of the following linear map, acting on linear operators:

$$\mathsf{M}_k : A \mapsto \int_{\mathrm{U}(d)} U^{\otimes k}AU^{\otimes k,\dagger}\mathrm{d}U\,. \quad (10)$$

This map, known as the k-th *Haar twirl* or k-th Haar *moment operator*, appears repeatedly throughout these notes and plays a fundamental role in both quantum information theory and many-body physics. It arises in contexts ranging from randomized benchmarking and quantum designs to entanglement theory, thermalization, and quantum chaos. The reason is that $\mathrm{tr}(B\mathsf{M}_k(A))$ has an operational interpretation if B is an Hermitian operator and $A=\rho$ is a quantum state: Preprate the state ρ, apply a random unitary $U^{\otimes k}$ to it and measure the expectation value of B. Thus, these moments can in principle be accessed by a quantum experiment (but not necessarily in an efficient way).

Computing M_k –or its matrix elements– would allow us to predict the statistical behaviour of Haar-random unitaries, for instance in the case of the previously given examples. In principle,

one can perform integration over the Haar measure by explicitly parametrizing the unitaries and integrating over the associated coordinates. We illustrate this for $d = 2$ and $k = 1$ in Example 2.1. However, this approach becomes highly impractical as the dimension d increases, due to the complicated geometry and rapidly growing number of parameters. Instead, we will introduce more powerful techniques – based on symmetry and representation theory – that allow us to compute Haar integrals in an efficient and conceptually clean way.

Example 2.1: Computing a Haar integral, the pedestrian way

Consider $d = 2$ and let us compute the Haar average

$$\int_{\mathrm{U}(2)} \mathrm{d}U\, UAU^\dagger \qquad \text{for} \quad A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}. \tag{11}$$

A generic unitary on $\mathrm{U}(2)$ is given by

$$U(\theta,\phi,\psi) = \begin{pmatrix} \cos\theta & e^{i\phi}\sin\theta \\ e^{i\psi}\sin\theta & e^{i(\phi+\psi)}\cos\theta \end{pmatrix} \tag{12}$$

with the Haar measure $\mathrm{d}U = (4\pi^2)^{-1}\sin(2\theta)\mathrm{d}\theta\mathrm{d}\phi\mathrm{d}\psi$ with $\theta \in [0,\pi/2]$ and $\phi,\psi \in [0,2\pi]$. (We ignore the global phase since it cancels trivially in $UAU^\dagger$. The integral acts on each element of the total matrix $(UAU^\dagger)_{ij} = \sum_{k,l} U_{i,k}A_{k,l}\bar{U}_{j,l}$. For concreteness, let us compute only the entry $i = j = 1$: $(U^\dagger AU)_{11} = u_{11}a_{11}\bar{u}_{11} + u_{11}a_{12}\bar{u}_{12} + u_{12}a_{21}\bar{u}_{11} + u_{22}a_{22}\bar{u}_{22}$. Using the well known formula $\int \mathrm{d}\alpha e^{in\alpha} = \delta_{n,0}$, we find that the terms a_{12} and a_{21} simplify. The remaining computation requires

$$\int_0^{\pi/2} \mathrm{d}\theta\, \sin(2\theta)(\cos^2\theta a_{11} + \sin^2\theta a_{22}) = \frac{1}{2}(a_{11} + a_{22}). \tag{13}$$

Working out the details for the other indices, we obtain

$$\int_{\mathrm{U}(2)} \mathrm{d}U\, UAU^\dagger = \frac{1}{2}\begin{pmatrix} a_{11}+a_{22} & 0 \\ 0 & a_{11}+a_{22} \end{pmatrix} = \frac{1}{2}\mathrm{tr}(A)\mathbb{I}. \tag{14}$$

The unitary commutant. The key observation in evaluating the integral (10) is that $\mathsf{M}_k(A)$ commutes with all unitaries $U^{\otimes k}$ for $U \in \mathrm{U}(d)$. The subspace of operators that commute with all $U^{\otimes k}$ is called the (k-fold) commutant of $\mathrm{U}(d)$:

$$\mathrm{Comm}_k = \{A \in \mathrm{L}(\mathcal{H}^{\otimes k}) \,|\, U^{\otimes k}A = AU^{\otimes k}\ \forall U \in \mathrm{U}(d)\}. \tag{15}$$

We then show the following lemma:

Lemma 2.1. $\mathsf{M}_k(A) \in \mathrm{Comm}_k$ *for all* $A \in \mathrm{L}(\mathcal{H}^{\otimes k})$. *In fact, every element in* Comm_k *is of the form* $\mathsf{M}_k(A)$.

Proof. Clearly, if A commutes with all $U^{\otimes k}$, then $\mathsf{M}_k(A) = A$, i.e. every element in Comm_k is of the form $\mathsf{M}_k(A)$. Vice versa, if $B = \mathsf{M}_k(A)$, then, by the invariance of the Haar measure:

$$U^{\otimes k}BU^{\otimes k,\dagger} = \int_{\mathrm{U}(d)} (UV)^{\otimes k}A(UV)^{\otimes k,\dagger}\mathrm{d}V = \int_{\mathrm{U}(d)} V^{\otimes k}AV^{\otimes k,\dagger}\mathrm{d}V = B, \quad \forall U \in \mathrm{U}(d). \tag{16}$$

Thus, $B \in \mathrm{Comm}_k$. □

A central step in understanding the commutant is to realize that every $U^{\otimes k}$ commutes with any operator R_π that permutes the k copies of our Hilbert space $\mathcal{H}$ according the permutation π. For instance, $U^{\otimes 2}$ clearly commutes with a swap of the two copies. Hence, the permutation operators R_π are contained in the k-th fold unitary commutant. These permutation operators form what is called a *representation* of the *symmetric group* S_k. A powerful result from representation theory – the *Schur-Weyl duality* – then tells us that the groups $\mathrm{U}(d)$ and S_k are *dual to each other* which, among others, implies that every element in the commutant Comm_k can be written as a linear combination of permutation operators R_π:

Theorem 2.1 (Schur-Weyl duality). *The k-fold unitary commutant Comm_k is spanned by $\{R_\sigma \,|\, \sigma \in \mathrm{S}_k\}$. Vice versa, the commutant of $\{R_\sigma \,|\, \sigma \in \mathrm{S}_k\}$ is spanned by $\{U^{\otimes k} \,|\, U \in \mathrm{U}(d)\}$*

We will not seek a deeper understanding of Schur-Weyl duality in this course but merely take this result as a motivation to develop techniques to evaluate $\mathsf{M}_k(A)$. In particular, Schur-Weyl duality allows us to express $\mathsf{M}_k(A)$ as a linear combination over permutations, with coefficients determined by their combinatorics. This leads to the framework known as *Weingarten calculus*, which enables the exact evaluation of many relevant integrals over the unitary group. As a result, permutations will play a key role throughout this course – not for abstract mathematical reasons, but because they offer a concrete and computable handle on random quantum processes.

2.2 Permutations and their combinatorics

In this section, we will introduce the essential properties of permutations needed for our purposes. While there is a rich and beautiful mathematical structure behind these ideas, we will focus only on the aspects that are directly relevant for our discussion and applications.

2.2.1 Permutations and Cycles

A *permutation* is a reordering of a finite set of elements. In this course, we consider permutations of the set $\{1, 2, \ldots, k\}$. We will denote permutations by Greek letters such as π, σ, τ, and so on. The set of all permutations of k elements forms a group under composition, called the *symmetric group*, and denoted by S_k.

Given a permutation $\sigma \in \mathrm{S}_k$ and an element $x \in \{1, \ldots, k\}$, we write $\sigma(x)$ to indicate the image of x under σ. A common and explicit way to write a permutation is the *two-line notation*, where the first row lists the original elements and the second row gives their images under the permutation:

$$\sigma = \begin{pmatrix} 1 & 2 & \ldots & k \\ \sigma(1) & \sigma(2) & \ldots & \sigma(k) \end{pmatrix}. \tag{17}$$

Example 2.2: Explicit Notation of Permutations

An example of a permutation of four elements is:

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix}. \tag{18}$$

This means that the permutation acts as:

$$\sigma(1) = 3, \quad \sigma(2) = 1, \quad \sigma(3) = 4, \quad \sigma(4) = 2. \tag{19}$$

The group operation in S_k is the composition of permutations, denoted by $\sigma \cdot \tau$ for $\sigma, \tau \in \mathrm{S}_k$, and defined by

$$(\sigma \cdot \tau)(x) = \sigma(\tau(x)) \qquad \text{for all } x \in \{1, \ldots, k\}. \tag{20}$$

Note that permutation composition is applied from right to left: τ acts first, followed by σ.

Example 2.3: Product of Permutations

Consider the following two permutations of $k = 4$ elements:

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix}, \qquad \tau = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 2 & 3 & 1 \end{pmatrix}. \tag{21}$$

To compute the composition $(\tau \cdot \sigma)(x) = \tau(\sigma(x))$, we apply σ first and then τ:

$$\begin{aligned} &\sigma(1) = 3, \quad \tau(\sigma(1)) = \tau(3) = 3, \\ &\sigma(2) = 1, \quad \tau(\sigma(2)) = \tau(1) = 4, \\ &\sigma(3) = 4, \quad \tau(\sigma(3)) = \tau(4) = 1, \\ &\sigma(4) = 2, \quad \tau(\sigma(4)) = \tau(2) = 2. \end{aligned} \tag{22}$$

Putting everything together, we find:

$$\tau \cdot \sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix}. \tag{23}$$

It is important to emphasize that for $k \geq 3$, the symmetric group S_k is *non-abelian*, meaning that the order of composition matters – in general, $\sigma \cdot \tau \neq \tau \cdot \sigma$.

Exercise 2.1. *Consider the permutations from Example 2.3. Compute the product $\sigma \cdot \tau$, and verify that it differs from $\tau \cdot \sigma$.*

The symmetric group contains a special element called the *identity permutation*, denoted by ι, which leaves all elements unchanged:

$$\iota = \begin{pmatrix} 1 & 2 & \dots & k \\ 1 & 2 & \dots & k \end{pmatrix}. \tag{24}$$

By definition, composition with the identity does not change the permutation:

$$\iota \cdot \sigma = \sigma = \sigma \cdot \iota. \tag{25}$$

Moreover, every permutation $\sigma \in S_k$ has an *inverse* σ^{-1} such that:

$$\sigma \cdot \sigma^{-1} = \iota = \sigma^{-1} \cdot \sigma. \tag{26}$$

To compute the inverse of a permutation π, apply the rule

$$\pi^{-1}(\pi(i)) = \iota(i) = i, \qquad i \in \{1, 2, \dots, k\}. \tag{27}$$

The resulting list π^{-1} is the inverse permutation.

Example 2.4: Inverse of a Permutation

Consider the following two permutations of $k = 5$ elements:

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 5 & 1 & 2 & 4 \end{pmatrix}. \tag{28}$$

To compute the inverse σ^{-1} we use the rule $\sigma^{-1}(\sigma(i)) \equiv i$

$$\begin{aligned} \sigma(1) = 3 &\Rightarrow \sigma^{-1}(3) = 1 \\ \sigma(2) = 5 &\Rightarrow \sigma^{-1}(5) = 2 \\ \sigma(3) = 1 &\Rightarrow \sigma^{-1}(1) = 3 \\ \sigma(4) = 2 &\Rightarrow \sigma^{-1}(2) = 4 \\ \sigma(5) = 4 &\Rightarrow \sigma^{-1}(4) = 5 \end{aligned} \tag{29}$$

Reordering the list in terms of the argument, we find

$$\sigma^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 4 & 1 & 5 & 2 \end{pmatrix}. \tag{30}$$

A *cyclic permutation*, or simply a *cycle*, is a specific type of permutation in which a subset of elements is permuted in a closed loop, while all remaining elements remain fixed. Formally, an l-cycle is a permutation that permutes l elements cyclically and leaves the other $k-l$ elements unchanged. The number l is called the *length of the cycle*.

Concretely, a cycle of length l means that there exists a subset $\{i_1, i_2, \ldots, i_l\} \subset \{1, 2, \ldots, k\}$ such that

$$\sigma(i_1) = i_2, \quad \sigma(i_2) = i_3, \quad \ldots, \quad \sigma(i_{l-1}) = i_l, \quad \sigma(i_l) = i_1, \tag{31}$$

and for all other elements $x \notin \{i_1, \ldots, i_l\}$, we have $\sigma(x) = x$.

Exercise 2.2 (Cyclic permutations). *The following are examples of cyclic permutations:*

$$\tau = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix}, \quad \sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 2 & 1 & 3 & 5 \end{pmatrix}. \tag{32}$$

Can you identify the subset $\{a_1, \ldots, a_l\}$ that is cyclically permuted in each case? What is the length l of each cycle? Write both permutations in cycle notation, including the fixed points.

One fundamental property of cycles is that any permutation can be decomposed into a product of disjoint cycles. That is, for any $\sigma \in S_k$, there exists a unique set of cycles that act on mutually disjoint subsets of $\{1, 2, \ldots, k\}$. This decomposition is unique up to the order in which the cycles are written. This motivates the *cycle notation* of permutations:

$$\sigma = (a_1\, a_2\, \ldots\, a_{j_1})(a_{j_1+1}\, a_{j_1+2}\, \ldots\, a_{j_2}) \ldots (a_{j_{r-1}+1}\, a_{j_{r-1}+2}\, \ldots\, a_{j_r})\,. \tag{33}$$

Here, each tuple represents a cycle, and all elements a_m are drawn from $\{1, 2, \ldots, k\}$ without repetition. The integer $r = \#(\sigma)$ denotes the *number of disjoint cycles* in the decomposition of σ.

Let us now describe an explicit algorithm to obtain the cycle decomposition of a permutation, as in Eq. (33). The idea is simple: we iteratively follow the action of the permutation until we return to the starting point, keeping track of all visited elements.

(i) **Start from the smallest unvisited element.** Initially, this is $x = 1$. If 1 has already been included in a previous cycle, move to the next smallest unvisited element.

(ii) **Construct a cycle by iterating the permutation.** Begin by writing down x. Then repeatedly apply the permutation π to generate the sequence

$$x,\ \pi(x),\ \pi(\pi(x)),\ \pi(\pi(\pi(x))),\ \ldots$$

Continue this process until you return to the starting point x. The list of elements

$$(x\ \ \pi(x)\ \ \pi^2(x)\ \ \ldots\ \ \pi^{j-1}(x)) \tag{34}$$

forms a cycle of length j. Mark all of these elements as visited.

(iii) **Repeat the process.** Find the next smallest unvisited element and return to step (ii). Continue until all elements have been visited. The full cycle decomposition of π is then obtained by combining the individual cycles found in each iteration.

Example 2.5: Cycle decomposition

Consider the permutation

$$\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 5 & 3 & 1 & 2 \end{pmatrix}. \tag{35}$$

We apply the cycle decomposition algorithm step by step:

- Start with 1:
$$1 \overset{\tau}{\mapsto} 4 \overset{\tau}{\mapsto} 1.$$
This gives the first cycle: $(1\ 4)$. Mark 1 and 4 as visited.
- Next smallest unvisited element is 2:
$$2 \overset{\tau}{\mapsto} 5 \overset{\tau}{\mapsto} 2.$$
This gives the second cycle: $(2\ 5)$. Mark 2 and 5 as visited.
- The last unvisited element is 3, and since
$$3 \overset{\tau}{\mapsto} 3,$$
this is a fixed point (a 1-cycle), written as (3).

Combining the above, the full cycle decomposition is:

$$\tau = (1\ 4)(2\ 5)(3). \tag{36}$$

When the total number of elements k is clear from the context (e.g., $k = 5$ in this case), it is common to omit one-cycles, also denoted *fixed points*, because these elements are understood to remain unchanged under the permutation. Using this convention, the permutation in Example 2.5 is simply written as:

$$\tau = (1\ 4)(2\ 5). \tag{37}$$

With this notation, the identity permutation is denoted by $\iota = ()$, which is shorthand for $\iota = (1)(2)\cdots(k)$ – that is, all elements are fixed.

In many computations throughout this course, we will be interested in the number of permutations in S_k with a fixed number of cycles $r = \#(\sigma)$. This quantity is given by the

unsigned Stirling numbers of the first kind, denoted by $c(k,r)$. These numbers satisfy the recursive relation:

$$c(k+1,r) = k \cdot c(k,r) + c(k,r-1), \tag{38}$$

This recurrence computes the numbers inductively without listing all permutations. They form a triangle analogous to Pascal's triangle; the values through $k = 10$ are tabulated in Sec. B.

Starting with the base case $k = 1$, where $\mathrm{S}_1 = \{\iota\}$, we find:

$$c(1,0) = 0, \qquad c(1,1) = 1. \tag{39}$$

Using the recurrence, the next values for $k = 2$ are:

$$c(2,0) = 0, \qquad c(2,1) = 1, \qquad c(2,2) = 1. \tag{40}$$

Exercise 2.3. *Compute the values $c(k,r)$ for $1 \le r \le k$ when $k = 3$ and $k = 4$. Verify that:*

$$\sum_{r=1}^{k} c(k,r) = k!\,, \qquad \sum_{r=1}^{k} c(k,r)x^r = x(x+1)\cdots(x+k-1)\,. \tag{41}$$

The first identity reflects the fact that summing over all permutations with a fixed number of cycles r recovers the total number of permutations in S_k, while the second gives another combinatorial interpretation of $c(k,r)$ as the coefficients in the power series of the 'rising factorial'.

Consider now the following permutations of six elements:

$$\sigma_1 = (1\,2\,3)(5\,6), \quad \sigma_2 = (2\,4\,6\,1), \quad \sigma_3 = (1\,2)(3\,4)(5\,6). \tag{42}$$

All three permutations have exactly three disjoint cycles, but the individual lengths of their cycles differs greatly. A finer classification of permutations can thus be obtained by specifying the full *cycle structure*. Concretely, the cycle structure of a permutation, denoted by $\lambda(\pi)$, is the list of the lengths of its disjoint cycles. For example, the permutation $\tau = (1\,4)(2\,5)$ has cycle structure $\lambda(\tau) = (2,2,1)$. The number of disjoint cycles is then given by the length of this list:

$$\#(\pi) = |\lambda(\pi)|, \qquad \text{for any } \pi \in \mathrm{S}_k\,. \tag{43}$$

Example 2.6: Cycle Structure

Consider the following permutations of six elements

$$\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 4 & 1 & 6 & 3 & 2 \end{pmatrix}, \quad \sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 2 & 6 & 4 & 1 & 3 \end{pmatrix}. \tag{44}$$

It is a simple exercise to show that their cycle decomposition is $\tau = (1,5,3)(2,4,6)$ and $\sigma = (1,5)(3,6)$ [which is a shorthand notation for $\sigma = (1,5)(3,6)(2)(4)$]. The permutation τ has two cycles of length 3, hence the cycle structure is $\lambda(\tau) = (3,3)$. Instead, σ is composed of two 2-cycles and two 1-cycles, so the cycle structure is $\lambda(\sigma) = (2,2,1,1)$.

Exercise 2.4. *Show that conjugation preserves the cycle structure of a permutation. That is, for any $\sigma, \pi \in \mathrm{S}_k$, prove that $\lambda(\pi\sigma\pi^{-1}) = \lambda(\sigma)$.*

Exercise 2.5. *Show that any two permutations with the same cycle structure are related by conjugation. In particular, for any $\sigma \in \mathrm{S}_k$ with $\lambda(\sigma) = (l_1, \ldots, l_r)$ there is a $\pi \in \mathrm{S}_k$ such that*

$$\pi\sigma\pi^{-1} = (1,\ldots,l_1)(l_1+1, l_1+2, \ldots, l_1+l_2)\cdots\Big(\sum_{i=1}^{r-1} l_i + 1, \ldots, \sum_{i=1}^{r} l_i\Big)\,.$$

Transpositions, also known as *swaps*, are a special class of permutations that exchange exactly two elements and leave all others unchanged. A transposition has the form:

$$\sigma = (i\ j) = \begin{pmatrix} 1 & \dots & i & \dots & j & \dots & k \\ 1 & \dots & j & \dots & i & \dots & k \end{pmatrix}. \tag{45}$$

It is straightforward to verify that *any permutation* can be written as a product of transpositions. However, unlike the decomposition into disjoint cycles, this representation is *not unique* – a given permutation can be written in many different ways as a product of transpositions. While this makes the transposition decomposition less suitable for labeling permutations, it is extremely useful in algebraic manipulations. For instance, it turns out that the number of transpositions in any representation is always odd or always even, which justifies the definition of the *sign of a permutation*:

$$\mathrm{sgn}(\sigma) = (-1)^{\#\text{transpositions in } \sigma}. \tag{46}$$

The sign function is important for the construction of representations of S_k and plays an important role in multilinear algebra, in particular in the definition of the *determinant of a matrix*.

We conclude this section by reviewing some structural aspects of the symmetric groups S_k for varying values of k. A key property is that the group S_k naturally embeds into S_{k+1}: that is, every permutation of k elements can be viewed as a permutation of $k+1$ elements that leaves the $(k+1)$-th element fixed. Since the symmetric group is generated by transpositions, we can even write

$$\mathrm{S}_{k+1} = \mathrm{S}_k \sqcup \{(j\ k+1)\cdot\sigma \mid \sigma \in \mathrm{S}_k,\ j = 1, 2, \dots, k\}, \tag{47}$$

where the union is disjoint and the second term represents all permutations obtained by composing an element of S_k with a transposition that swaps $k+1$ with one of the first k elements.

This recursive structure is useful for establishing many properties of permutations by induction on k. A simple but important example is the total number of elements in the symmetric group.

Theorem 2.2 (Counting of Permutations). *For any $k \geq 1$, the total number of permutations is $|\mathrm{S}_k| = k!$.*

Proof. We prove the statement by induction over k. For $k = 1$, $\mathrm{S}_1 = \{\iota\}$ consists only of the identity permutation, so $|\mathrm{S}_1| = 1 = 1!$. Assume now that $|\mathrm{S}_k| = k!$ for some $k \geq 1$. Equation (47) then implies that

$$|\mathrm{S}_{k+1}| = |\mathrm{S}_k| + k \cdot |\mathrm{S}_k| = k! + k \cdot k! = (k+1)\cdot k! = (k+1)!. \tag{48}$$

This concludes the proof. □

2.2.2 The Action of Permutations on Quantum States

Throughout this course, we work with a d-dimensional Hilbert space $\mathcal{H} = \mathbb{C}^d$, with orthonormal basis $\{|x\rangle\}_{x=0}^{d-1}$. Its k-fold tensor power $\mathcal{H}^{\otimes k}$ is sometimes called the *replica space* in many-body physics. Quantum information scientists prefer to talk about *copies* instead of *replicas*.

We can let a permutation $\sigma \in \mathrm{S}_k$ act on a basis state $|x_1, \dots, x_k\rangle \in \mathcal{H}^{\otimes k}$ by moving the content of slot j to slot $\sigma(j)$:

$$R_\sigma |x_1, \dots, x_k\rangle = |x_{\sigma^{-1}(1)}, \dots, x_{\sigma^{-1}(k)}\rangle. \tag{49}$$

Note that the inverse in the definition is essential to get the claimed action. For example, let $\sigma = (1\,4\,3)(2)$, then $1 \mapsto 4$, $4 \mapsto 3$, and $3 \mapsto 1$; therefore

$$|x_1, x_2, x_3, x_4\rangle \overset{\sigma}{\mapsto} |x_3, x_2, x_4, x_1\rangle. \tag{50}$$

This is exactly the action in Eq. (49).

The map $R: \pi \mapsto R_\pi$ defines a so-called *unitary representation* of the symmetric group S_k on $\mathcal{H}^{\otimes k}$. This means that the operators R_π behave exactly as the according permutations in S_k, i.e., the fulfill the following identities:

$$R_{\sigma\pi} = R_\sigma R_\pi, \qquad R_\iota = \mathbb{1}, \qquad R_{\sigma^{-1}} = R_\sigma^{-1} = R_\sigma^\dagger. \tag{51}$$

Here, $(\sigma\pi)(j) := \sigma(\pi(j))$ and ι is the identity permutation. We leave the verification of Eq. (51) as an exercise.

Exercise 2.6. *Verify that R defines a unitary representation of the symmetric group S_k, i.e., check Eq. (51).*

Exercise 2.7. *Define $L_\sigma|x_1, x_2, \ldots, x_k\rangle := |x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(k)}\rangle$. Show that L fulfills all identities in Eq. (51) except the first. What is the correct identity instead?*

Exercise 2.8. *Show that permutations act on tensor products of operators as follows:*

$$R_\sigma(A_1 \otimes A_2 \otimes \cdots \otimes A_k)R_\sigma^\dagger = A_{\sigma^{-1}(1)} \otimes A_{\sigma^{-1}(2)} \otimes \cdots \otimes A_{\sigma^{-1}(k)}. \tag{52}$$

Hint: *Apply both sides to a product basis state.*

Composite Systems In practice, we often work with *multi-qudit systems* described by a Hilbert space of the form $\mathcal{H} = (\mathbb{C}^q)^{\otimes n}$, corresponding to n qudits of local dimension q. In this setting, the k-fold copy of the system is given by the tensor product

$$((\mathbb{C}^q)^{\otimes n})^{\otimes k}, \tag{53}$$

which can be naturally visualized as a $k \times n$ grid of qudits (see Fig. 1). Each row represents one replica of the full system, and each column represents the k copies of a single local qudit.

Quantum operations such as global unitaries typically act *row-wise*, that is, identically and independently on each copy. Such operations are of the form

$$U^{\otimes k}, \qquad \text{where } U \in \mathrm{U}((\mathbb{C}^q)^{\otimes n}), \tag{54}$$

meaning that the unitary acts in parallel across the k rows.

In contrast, permutations act by permuting the *rows*, i.e., the k copies of each local qudit. This operation is performed *column-wise*, and can be implemented by applying the same permutation operator to each column in parallel. This leads to a convenient factorized structure: if we reinterpret the total space via the isomorphism

$$((\mathbb{C}^q)^{\otimes n})^{\otimes k} \simeq ((\mathbb{C}^q)^{\otimes k})^{\otimes n}, \tag{55}$$

then the permutation operator R_π acting on the full system decomposes as

$$R_\pi = r_\pi^{\otimes n}, \tag{56}$$

Here r_π permutes the k replicas of only one qudit: it acts on $(\mathbb{C}^q)^{\otimes k}$, a space of dimension q^k. This "horizontal" factorization is exactly what is depicted in Fig. 1 and will be essential in constructing efficient representations of randomized operations throughout the course.

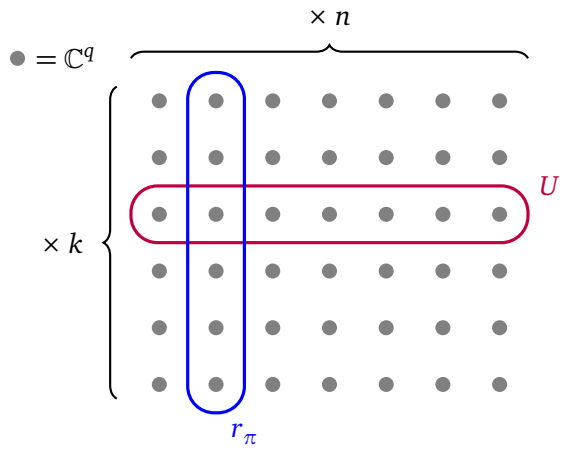


Figure 1: The Hilbert space $((\mathbb{C}^q)^{\otimes n})^{\otimes k}$ depicted as a $k \times n$ grid where every point corresponds to a copy of $\mathbb{C}^q$. Unitaries $U \in \mathrm{U}(q^n)$ act row-wise on the grid, while permutations $\pi \in \mathrm{S}_k$ act column-wise.

Traces Traces will play a fundamental role for explicit computations. Here, we derive a formula for the trace of permutation operators and for the trace of a product operator multiplied with a permutation. Let us first consider the cyclic permutation $\gamma = (1, \dots, k-1, k)$. Then, we compute

$$\begin{aligned}\mathrm{tr}(R_\gamma) &= \sum_{x_1,\dots,x_k=0}^{d-1} \langle x_1, \dots, x_k | R_\gamma | x_1, \dots, x_k \rangle \\ &= \sum_{x_1,\dots,x_k} \langle x_1, \dots, x_k | x_k, x_1 \dots, x_{k-1} \rangle = \sum_{x_1,\dots,x_k} \delta_{x_1,x_k} \delta_{x_2,x_1} \dots \delta_{x_k,x_{k-1}} = d. \end{aligned} \tag{57}$$

Next, consider an arbitrary permutation σ. By Exercise 2.5, we can find a permutation π such that $\pi\sigma\pi^{-1} = (1, \dots, b_1)(b_1+1, b_1+2, \dots, b_2)\cdots(b_{r-1}+1, b_{r-1}+2, \dots, b_r)$ where $r = \#\sigma$ is the number of cycles in σ. Note that $R_{\pi\sigma\pi^{-1}}$ is simply a tensor product of cyclic permutations on $\mathcal{H}^{\otimes b_1}, \dots, \mathcal{H}^{\otimes(b_2-b_1)}, \dots, \mathcal{H}^{\otimes(b_r-b_{r-1})}$ and thus

$$\mathrm{tr}(R_\sigma) = \mathrm{tr}(R_{\pi\sigma\pi^{-1}}) = \prod_{i=1}^{r} d = d^r = d^{\#\sigma}. \tag{58}$$

Beyond this simple situation, we often need to compute the trace of product operators multiplied with a permutation operator, i.e. an expression of the form $\mathrm{tr}(A_1 \otimes A_2 \otimes \cdots \otimes A_k R_\sigma)$. Here, we can follow the same arguments as above: First, if $\sigma = \gamma = (1, \dots, k)$ is again the cyclic permutation, then

$$\mathrm{tr}(A_1 \otimes A_2 \otimes \cdots \otimes A_k R_\gamma) = \sum_{x_1,\dots,x_k} \langle x_1, \dots, x_k | A_1 \otimes A_2 \otimes \cdots \otimes A_k | x_k, x_1 \dots, x_{k-1} \rangle \tag{59}$$

$$= \sum_{x_1,\dots,x_k} (A_1)_{x_1,x_k} (A_2)_{x_2,x_1} \cdots (A_k)_{x_k,x_{k-1}} = \mathrm{tr}(A_k A_{k-1} \cdots A_1). \tag{60}$$

Note that the matrices are exactly ordered according to $\gamma^{-1} = (k, k-1, \dots, 1)$. Next, for an arbitrary $\sigma = (a_1, \dots, a_{j_1})(a_{j_1+1}, \dots, a_{j_2}) \dots (a_{j_{r-1}+1}, \dots, a_k)$, find again a permutation π such that $\pi\sigma\pi^{-1} = (1, \dots, b_1)(b_1+1, b_1+2, \dots, b_2)\cdots(b_{r-1}+1, b_{r-1}+2, \dots, b_r)$. Using that $R_{\pi\sigma\pi^{-1}}$

factorizes as above and Exercise 2.8, we find

$$\begin{aligned}\operatorname{tr}(A_1 \otimes A_2 \otimes \cdots \otimes A_k R_\sigma) &= \operatorname{tr}(A_{\pi^{-1}(1)} \otimes \cdots \otimes A_{\pi^{-1}(k)} R_{\pi\sigma\pi^{-1}}) && (61)\\ &= \operatorname{tr}(A_{\pi^{-1}(b_1)} \cdots A_{\pi^{-1}(1)}) \cdots \operatorname{tr}(A_{\pi^{-1}(b_r)} \cdots A_{\pi^{-1}(b_{r-1}+1)}) && (62)\\ &= \operatorname{tr}(A_{a_{j_1}} \cdots A_{a_1}) \cdots \operatorname{tr}(A_{a_k} \cdots A_{a_{j_{r-1}+1}+1}). && (63)\end{aligned}$$

For the important case when $A_1 = A_2 = \cdots = A_k$, the final result is simplified to

$$\operatorname{tr}(A^{\otimes k} R_\sigma) = \prod_{c \in \lambda(\sigma)} \operatorname{tr}(A^c), \tag{64}$$

where $\lambda(\sigma)$ is the cycle structure of the permutation σ.

Symmetric subspace Throughout this course, symmetries under permutations will play a fundamental role, in particular the *symmetric subspace* of $(\mathbb{C}^d)^{\otimes k}$ composed of vectors that are left invariant by permutations:

$$\operatorname{Sym}_{k,d} \equiv \operatorname{Sym}((\mathbb{C}^d)^{\otimes k}) := \{\psi \in (\mathbb{C}^d)^{\otimes k} \mid R_\sigma |\psi\rangle = |\psi\rangle \ \forall \sigma \in \mathrm{S}_k\}. \tag{65}$$

We will now show that the projector onto $\operatorname{Sym}_{k,d}$ is

$$P_{\operatorname{Sym},k,d} = \frac{1}{k!} \sum_{\sigma \in \mathrm{S}_k} R_\sigma . \tag{66}$$

To see this, we first check that $P_{\operatorname{Sym},k,d}$ is an orthogonal projector:

$$P^2_{\operatorname{Sym},k,d} = \frac{1}{(k!)^2} \sum_{\sigma,\pi \in \mathrm{S}_k} R_{\sigma\pi} = \frac{1}{k!} \sum_{\sigma \in \mathrm{S}_k} \frac{1}{k!} \sum_{\tau \in \mathrm{S}_k} R_\tau = P_{\operatorname{Sym},k,d}\,, \tag{67}$$

$$P^\dagger_{\operatorname{Sym},k,d} = \frac{1}{k!} \sum_{\sigma \in \mathrm{S}_k} R_{\sigma^{-1}} = \frac{1}{k!} \sum_{\pi \in \mathrm{S}_k} R_\pi = P_{\operatorname{Sym},k,d}\,, \tag{68}$$

where we substituted variables as $\tau = \sigma\pi$ and $\pi = \sigma^{-1}$, respectively, and used that the sum is invariant under the change of variables. Next, note that for all $\psi \in \operatorname{Sym}_{k,d}$:

$$P_{\operatorname{Sym},k,d}|\psi\rangle = \frac{1}{k!} \sum_{\sigma \in \mathrm{S}_k} R_\sigma |\psi\rangle = \frac{1}{k!} \sum_{\sigma \in \mathrm{S}_k} |\psi\rangle = |\psi\rangle\,, \tag{69}$$

thus, $\operatorname{Sym}_{k,d}$ is in the range of $P_{\operatorname{Sym},k,d}$. Moreover, if $P_{\operatorname{Sym},k,d}|\psi\rangle = |\psi\rangle$, then

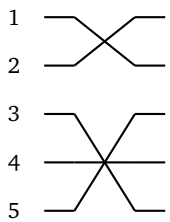
$$R_\pi|\psi\rangle = R_\pi P_{\operatorname{Sym},k,d}|\psi\rangle = \frac{1}{k!} \sum_{\sigma \in \mathrm{S}_k} R_{\pi\sigma}|\psi\rangle = \frac{1}{k!} \sum_{\tau \in \mathrm{S}_k} R_\tau |\psi\rangle = P_{\operatorname{Sym},k,d}|\psi\rangle = |\psi\rangle\,, \tag{70}$$

and thus the range of $P_{\operatorname{Sym},k,d}$ is exactly $\operatorname{Sym}_{k,d}$. We can now compute the dimension of the symmetric subspace using Eq. (58) and Exercise 2.3 as

$$\begin{aligned}\dim \operatorname{Sym}_{k,d} = \operatorname{tr} P_{\operatorname{Sym},k,d} &= \frac{1}{k!} \sum_{\sigma \in \mathrm{S}_k} d^{\#\sigma} \\ &= \frac{1}{k!} \sum_{r=1}^{k} c(k,r) d^r = \frac{d(d+1)\cdots(d+k-1)}{k!} = \binom{d+k-1}{k}.\end{aligned} \tag{71}$$

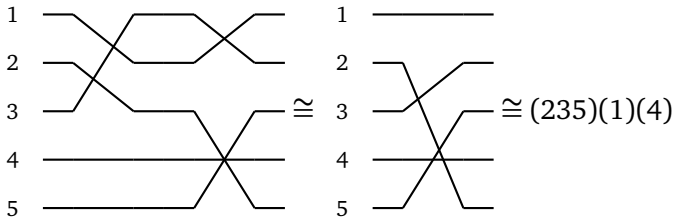
Graphical representation While the above methodologies are generic and straightforward, the algebra is often cumbersome. For this reason, it is useful to introduce a graphical notation to represent permutations. Similar to the Feynman diagrammatics for perturbative expansions, this is simply a bookkeeping of the operation previously described.

We draw a permutation by connecting input position j to output position $\sigma(j)$. For example, the five-element permutation $\tau = (12)(35)(4)$ is represented by

$(12)(35)(4) \cong$ 

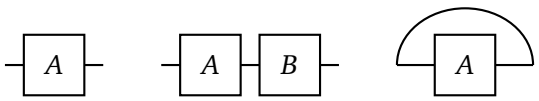
(72)

This notation makes computing products particularly easy. In the next diagram the lines for $\sigma = (123)(4)(5)$ are encountered first and those for $\tau = (12)(35)(4)$ second; it therefore represents $\tau \cdot \sigma$:

$\tau \cdot \sigma \cong$ 

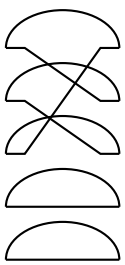
$\cong (235)(1)(4)$ (73)

We can use the same notation for the representation R of S_k with the wires now representing the replicas of the Hilbert space $\mathcal{H}$. Concretely, the diagram for R_π now encodes the action on a computational basis state $|x_1, \ldots, x_k\rangle$. This notation is in fact a special case of a *tensor network diagram*. Here, a matrix $A = (A_{ij})$ is represented by a box with two 'legs' representing the indices. If two such boxes are connected along a leg, the corresponding index is summed over. In this way, we can write down the multiplication of two matrices A and B. It also allows us to write down the trace by connecting an input and an output leg.



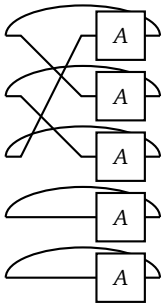
(74)

Combining these diagrammatic allows us to reason about products of permutations with operators and (partial) traces. For example:

$\mathrm{tr}(R_\sigma) =$  $= \left(\bigcirc\right)^{\#(\sigma)} = d^{\#(\sigma)}.$ (75)

Similarly, we can include operators as follows:

$$\mathrm{tr}(A^{\otimes k}R_\sigma) = \quad = \mathrm{tr}(A^3)\,\mathrm{tr}(A)^2\,. \tag{76}$$



2.3 Weingarten calculus

We recall the definition of the k-th Haar *moment operator* M from Sec. 2.1. It is the map that averages an input operator A over the k-copy action of $\mathrm{U}(d)$:

$$\mathsf{M}_k : A \mapsto \int_{\mathrm{U}(d)} U^{\otimes k} A U^{\otimes k,\dagger} \mathrm{d}U\,. \tag{77}$$

As shown in Lem. 2.1, $\mathsf{M}_k(A)$ lies in the unitary commutant for all A and can thus be written as a linear combination

$$\mathsf{M}_k(A) = \sum_{\pi\in \mathrm{S}_k} c_\pi(A) R_\pi\,, \tag{78}$$

for some coefficients $c_\pi(A)$. A natural question to ask is whether these coefficients are unique, i.e. whether permutations form a basis for the commutant. Intriguingly, this is the case if the dimension d is large enough:

Lemma 2.2. *The set* $\{R_\sigma \,|\, \sigma \in \mathrm{S}_k\}$ *is linearly independent for* $d \geq k$.

Proof. We consider the standard basis of $\mathbb{C}^d$, which we here denote as $|1\rangle, \ldots, |d\rangle$. Since $k \leq d$, we can consider the action of permutations on $|1, \ldots, k\rangle \in (\mathbb{C}^d)^{\otimes k}$:

$$R_\pi |1, \ldots, k\rangle = |\pi^{-1}(1), \ldots, \pi^{-1}(k)\rangle\,. \tag{79}$$

Suppose that $\sum_{\pi\in \mathrm{S}_k} \alpha_\pi R_\pi = 0$. Applying this operator to $|1, \ldots, k\rangle$ gives

$$\sum_\pi \alpha_\pi |\pi^{-1}(1), \ldots, \pi^{-1}(k)\rangle = 0. \tag{80}$$

Because $k \leq d$, all these kets are distinct standard-basis vectors and are therefore linearly independent. Hence $\alpha_\pi = 0$ for all π, proving that the permutation operators are linearly independent. □

We note that the condition $d \geq k$ is sharp: As soon as $d < k$, the permutations become linearly dependent, cf. Exercise 2.9. In the following, we will always assume that $d \geq k$, sometimes called the *stable regime* in the literature.

Exercise 2.9. *Show that permutations become linearly dependent as soon as* $k > d$.
Hint: *Consider the* antisymmetric *subspace* $\mathrm{Alt}_{k,d} \subset (\mathbb{C}^d)^{\otimes k}$ *which is the joint* -1 *eigenspace of all* transpositions $R_{(ij)}$ *for* $(ij) \in \mathrm{S}_k$. *The projector onto* $\mathrm{Alt}_{k,d}$ *has the general form*

$$P_{\mathrm{Alt},k,d} = \frac{1}{k!} \sum_{\pi\in \mathrm{S}_k} \mathrm{sgn}(\pi) R_\pi\,. \tag{81}$$

Consider the action of $P_{\mathrm{Alt},k,d}$ *on an orthonormal basis* $|x_1, \ldots, x_k\rangle$ *of* $(\mathbb{C}^d)^{\otimes k}$ *for* $k > d$. *What does the result imply for* $P_{\mathrm{Alt},k,d}$ *and the set of permutation operators?*

Let us proceed by deriving a formula for the expansion coefficicients in Eq. (78). Note that taking the trace inner product of $\mathsf{M}_k(A)$ with a fixed permutation $R_\sigma^\dagger$ yields

$$\operatorname{tr}(R_\sigma^\dagger \mathsf{M}_k(A)) = \int \operatorname{tr}(R_\sigma^\dagger U^{\otimes k} A U^{\otimes k,\dagger}) dU = \int \operatorname{tr}(U^{\otimes k,\dagger} R_\sigma^\dagger U^{\otimes k} A) dU = \operatorname{tr}(R_\sigma^\dagger A)\,. \tag{82}$$

However, we also have

$$\operatorname{tr}(R_\sigma^\dagger A) = \operatorname{tr}(R_\sigma^\dagger \mathsf{M}_k(A)) = \sum_{\pi\in S_k} c_\pi(A) \operatorname{tr}(R_\sigma^\dagger R_\pi) =: \sum_{\pi\in S_k} c_\pi(A) G_{\sigma,\pi}\,, \tag{83}$$

where we defined the *Gram matrix*

$$G_{\pi,\sigma} := \operatorname{tr}(R_\pi^\dagger R_\sigma) = \operatorname{tr}(R_\pi^\dagger R_\sigma) = \operatorname{tr}(R_{\pi^{-1}\sigma}) = d^{\#(\pi^{-1}\sigma)}\,. \tag{84}$$

Here, we used that R is a representation to combine the product of permutation operators, and the trace formula (58). Setting $a_\sigma := \operatorname{tr}(R_\sigma^\dagger A)$, we can write the above equation in matrix form as $a = Gc$, which we could hope to invert to get an expression for the coefficient vector c. Note that the permutations are not orthogonal with respect to the trace inner product $(A|B) = \operatorname{tr}(A^\dagger B)$, and hence the Gram matrix is not simply diagonal. However, we know that the permutations span the commutant and that $\mathsf{M}_k(A)$ lies in the commutant. Hence, the equation $a = Gc$ always has a solution and it is unique if and only if the permutations form a basis, i.e. iff $d \geq k$, which is what we agreed to assume for the remainder of this course.[1] Then, this solution is simply $c = G^{-1}a$, or put differently,

$$\mathsf{M}_k(A) = \sum_{\pi,\sigma\in S_k} W_{\pi,\sigma} \operatorname{tr}(R_\sigma^\dagger A) R_\pi\,, \tag{85}$$

where we defined $W := G^{-1}$, the so-called *Weingarten matrix*. Knowing the Weingarten matrix allows us to compute integrals of the form (10) using the Weingarten expansion (85).

Example 2.7: Weingarten matrix for $k = 2$

Let us consider $k = 2$. Then the two permutation operators are $\mathbb{1} = R_\iota$ and the flip/swap $\mathsf{F} = R_{(12)}$. There is only one non-trivial matrix element, namely $G_{\iota,(12)} = \operatorname{tr}(\mathsf{F}) = d$. Hence, the Gram and Weingarten matrices are

$$G = d^2 \begin{pmatrix} 1 & d^{-1} \\ d^{-1} & 1 \end{pmatrix}, \qquad W = \frac{1}{d^2-1} \begin{pmatrix} 1 & -d^{-1} \\ -d^{-1} & 1 \end{pmatrix}. \tag{86}$$

General properties of the Gram and Weingarten matrix. The Gram and Weingarten matrix have a substantial structure which directly relates to the representation theory of the symmetric group. We will not dive into these details in this course, but instead prove some concrete relations. We summarize them in the following.

Lemma 2.3. *The Gram and Weingarten matrix fulfill the following properties.*

(a) $G_{\pi,\sigma}$ *and* $W_{\pi,\sigma}$ *only depend on* $\pi^{-1}\sigma$.

(b) The row and column sums of G are constant:

$$\mathcal{G}_{k,d} := \sum_\sigma G_{\pi,\sigma} = \sum_\pi G_{\pi,\sigma} = \frac{(d+k-1)!}{(d-1)!} = d(d+1)\cdots(d+k-1)\,. \tag{87}$$

[1] It is however not terribly complicated to make this work for $d < k$, see Sec. 2.6.

(c) *The row and column sums of W are constant:*

$$\sum_\sigma W_{\pi,\sigma} = \sum_\pi W_{\pi,\sigma} = \mathcal{G}_{k,d}^{-1} = \frac{(d-1)!}{(d+k-1)!}\,. \tag{88}$$

Proof. (a) Clearly, $G_{\pi,\sigma}$ depends only on $\pi^{-1}\sigma$ by definition (cf. Eq. (84)). Now note that this implies that G is invariant under simultaneous row and column permutations. Indeed, if T_τ is the permutation matrix acting as $T_\tau|e_\pi\rangle = |e_{\tau\pi}\rangle$, then $(T_\tau^{-1}GT_\tau)_{\sigma,\pi} = G_{\tau\sigma,\tau\pi} = G_{\sigma,\pi}$. Inverting $G = T_\tau^{-1}GT_\tau$ yields $W = T_\tau^{-1}WT_\tau$ and thus $W_{\pi,\sigma} = W_{\tau\pi,\tau\sigma}$ for all τ, in particular $W_{\pi,\sigma} = W_{\mathrm{id},\pi^{-1}\sigma}$ for $\tau = \pi^{-1}$. For (b), we compute

$$\mathcal{G}_{k,d} = \sum_\sigma \mathrm{tr}(R_{\pi^{-1}\sigma}) = \sum_\sigma \mathrm{tr}(R_\sigma) = k!\,\mathrm{tr}(P_{\mathrm{Sym},k,d}) = k!\binom{d+k-1}{k} = \frac{(d+k-1)!}{(d-1)!}\,. \tag{89}$$

Here, we used that the multiplication by π^{-1} can be absorbed into the sum (variable change), and the definition of the projector onto the symmetric subspace, $P_{\mathrm{Sym},k,d} = \frac{1}{k!}\sum_\sigma R_\sigma$, and the value of its trace, cf. Eqs. (66) and (71). For (c), we note that the definition of W as inverse of G implies

$$\sum_\pi W_{\sigma,\pi}G_{\pi,\tau} = \delta_{\sigma,\tau} \quad\Rightarrow\quad 1 = \sum_{\pi,\tau} W_{\sigma,\pi}G_{\pi,\tau} = \mathcal{G}_{k,d}\sum_\pi W_{\sigma,\pi} \quad\Rightarrow\quad \sum_\pi W_{\sigma,\pi} = \mathcal{G}_{k,d}^{-1}\,. \tag{90}$$

□

Example 2.8: Average collision probability

Measuring $U|0\rangle$ in the computational basis gives outcome x with probability $p(x|U) = |\langle x|U|0\rangle|^2$. A measure of flatness of this distribution is the *collision probability*:

$$Z_U := \sum_x p(x|U)^2 = \sum_x |\langle x|U|0\rangle|^4\,. \tag{91}$$

Here, we are interested on how flat this distribution is *on average*, over Haar-random unitaries U. Due to the invariance of the Haar measure, we can simply absorb the X gates that prepare $|x\rangle = X^{x_1}\otimes\cdots\otimes X^{x_n}|0\rangle =: X(x)|0\rangle$ into the average:

$$Z := \int Z_U \mathrm{d}U = \sum_x \int_U |\langle 0|X(x)^\dagger U|0\rangle|^4 \mathrm{d}U = d\int_U |\langle 0|U|0\rangle|^4 \mathrm{d}U\,. \tag{92}$$

To compute the integral, we use second-order Weingarten calculus:

$$\begin{aligned} Z &= d\int_U \mathrm{tr}(|0\rangle\langle 0|^{\otimes 2}U^{\otimes 2}|0\rangle\langle 0|^{\otimes 2}U^{\otimes 2,\dagger})\mathrm{d}U && (93)\\ &= d\sum_{\pi,\sigma\in S_2} W_{\pi,\sigma}\,\mathrm{tr}(R_\sigma^\dagger|0\rangle\langle 0|^{\otimes 2})\,\mathrm{tr}(R_\pi|0\rangle\langle 0|^{\otimes 2}) && (94)\\ &= d\sum_{\pi,\sigma\in S_2} W_{\pi,\sigma} = 2d\mathcal{G}_{2,d}^{-1} = 2d\frac{(d-1)!}{(d+1)!} = \frac{2}{d+1}\,. && (95)\end{aligned}$$

Here, we used Lem. 2.3. Note that we haven't actually used the exact form of the Weingarten matrix from Example 2.7. In fact, along the same lines we find that the average of $\sum_x p(x|U)^k$ is

$$\mathrm{IPR}_k := \sum_x \int p(x|U)^k \mathrm{d}U = k!d\mathcal{G}_{k,d}^{-1} = \frac{k!d!}{(d+k-1)!}\,. \tag{96}$$

Exercise 2.10. *Using Weingarten calculus, show that for all states $|\psi\rangle$, we have*

$$\int (U|\psi\rangle\langle\psi|U^\dagger)^{\otimes k} \mathrm{d}U = \frac{P_{\mathrm{Sym},k,d}}{\binom{d+k-1}{k}}, \tag{97}$$

where $P_{\mathrm{Sym},k,d}$ is the projector onto the symmetric subspace. Hint: *Follow the derivation of Example 2.8.*

2.4 Approximating the Haar measure: Unitary designs

We have seen how Haar integrals over the unitary group can be computed using Weingarten calculus. These tools will help us design randomized protocols and algorithms, beginning with the application in the next section. In practice, Haar-random unitaries are however way too *expensive* to be useful: On a system with n qudits, a quantum circuit implementing a Haar-random unitary requires a circuit depth that is exponential in n. This means that any quantum algorithm that utilizes Haar-random unitaries necessarily has exponential runtime.[2]

In the following, we thus briefly discuss a concept introduced to lower the requirements for quantum randomness: *unitary k-designs*. These are sets of unitaries that mimic the Haar measure up to the k-th moment and can thus be used as a replacement in applications that rely on finite moments only. Unitary designs can be seen as a set of points in the unitary group that are "sufficiently equally distributed" to reproduce the first moments of the Haar measure, cf. Fig. 2. Importantly, unitary designs may be constructed using significantly less resources than Haar-random unitaries, in particular using polynomial-sized circuits.

Formally, we define a unitary design as follows. Let $\mathcal{G} \subset \mathrm{U}(d)$ be a (finite) set of unitaries (we can extend this to infinite sets equipped with a suitable probability measure). Then, we call $\mathcal{G}$ a *unitary k-design* if

$$\frac{1}{|\mathcal{G}|}\sum_{U\in\mathcal{G}} U^{\otimes k} A U^{\otimes k,\dagger} = \mathrm{M}_k(A), \tag{98}$$

for all matrices A. We call the largest k for which Eq. (98) holds the *order* of the unitary design. A k-design is automatically an exact j-design for every $j \leq k$. To see this, simply take the partial trace of Eq. (98) over a sufficient number of copies.

It is useful to keep two first examples in mind: A 1-design reproduces the depolarizing identity

$$\frac{1}{|\mathcal{G}|}\sum_{U\in\mathcal{G}} UAU^\dagger = \frac{\mathrm{tr}(A)}{d}\,\mathbb{1}. \tag{99}$$

Weingarten expansion for $k = 2$, see Example 2.7, shows that a 2-design must obey

$$\frac{1}{|\mathcal{G}|}\sum_{U\in\mathcal{G}} U^{\otimes 2} A U^{\otimes 2,\dagger} = \frac{1}{d(d^2-1)}\left[(d\,\mathrm{tr}(A) - \mathrm{tr}(\mathbb{F}A))\mathbb{1} - (\mathrm{tr}(A) - d\,\mathrm{tr}(\mathbb{F}A)\mathbb{F}\right]. \tag{100}$$

Note that the left hand side of Eq. (98) can be seen as the k-fold twirl over the set $\mathcal{G}$. This means that averages over the set $\mathcal{G}$ can be computed using averages over the full unitary group – for instance using Weingarten calculus. Vice versa, in applications that only depend on k-fold twirls, Haar-random unitaries can be replaced by unitary k-designs, thereby enabling more resource-efficient and flexible implementations.

One might hope that one could find sufficiently symmetric subgroups of $\mathrm{U}(d)$ that give natural candidates for unitary designs. However, it turns out that Eq. (98) together with the group structure imposes very strict conditions on such a subgroup, resulting in the fact that

[2] At least if we are not willing to use an exponential number of ancilla qubits. Very recent work showed that one can in fact trade the exponential runtime for exponential space (qubits) [1].

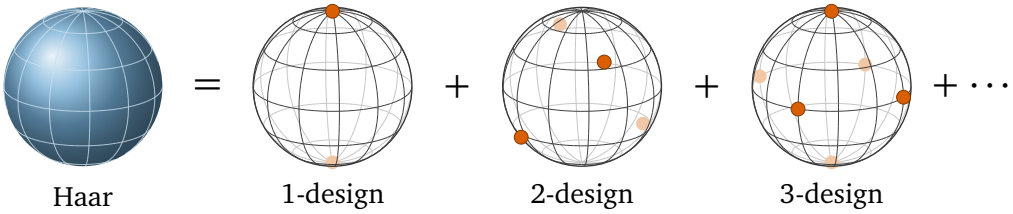


Figure 2: Caricature showing a "series expansion" of the Haar measure on the full unitary group (here depicted by a sphere) into finite subsets that agree with the k-th Haar moment. The point sets shown are the antipodal pair, the tetrahedron and the octahedron, which are 1-, 2- and 3-designs on the Bloch sphere, respectively.

these do not exist for $k \geq 4$ (and $d \geq 5$). Despite this, the most important example of a unitary design (with $k = 3$) is in fact a group, the *Clifford group*, which we will introduce in the next section.

Beyond subgroups, unitary k-designs exist for all k and dimensions d. Unfortunately, explicit constructions of general unitary k-designs are incredibly rare and the known ones are highly inefficient. To overcome these obstacles, it has been fruitful to demand that Eq. (98) holds only approximately – such *approximate unitary designs* can be realized much more easily and enable the efficient implementation of quantum randomness also for higher moments k. We will come back to this point in Sec. 4.

2.5 The Clifford group

The prototypical example of a unitary design is the *Clifford group*. Besides designs, the Clifford group plays a major role in quantum error correction, where it typically constitutes the set of "easy" gates in fault-tolerant quantum computing. Clifford operations can also be efficiently simulated on a classical computer, making them the starting point for classical simulation algorithms and investigations of the "non-classicality" of quantum mechanics.

The simplest way to define the Clifford group is via its local generators: The single-qubit *phase gate* S and *Hadamard gate* H, as well as the two-qubit CNOT gate CX, given by

$$S|x\rangle = i^x|x\rangle \qquad H|x\rangle = \frac{1}{\sqrt{2}}\big(|0\rangle + (-1)^x|1\rangle\big) \qquad CX|x,y\rangle = |x, y \oplus x\rangle\,, \tag{101}$$

where $y \oplus x$ denotes addition modulo 2. The group that is generated by S and H on every qubit and CX on every pair of qubits is the n-qubit *Clifford group* Cl_n. It is a finite subgroup of $\mathrm{U}(2^n)$ with $2^{O(n^2)}$ elements. Importantly, every Clifford unitary can be implemented using $O(n^2)$ generators S, H, and CX.

An important subgroup of the Clifford group is the *Pauli group*. Recall the Pauli operators

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{102}$$

which we complement with the identity $\mathbb{1}$. Then, the multi-qubit Pauli operators on $\mathcal{H} = (\mathbb{C}^2)^{\otimes n}$ are simply given by all possible tensor products of the single-qubit Pauli operators, in formula:

$$\sigma = \sigma_1 \otimes \cdots \otimes \sigma_n\,, \qquad \sigma_i \in \{\mathbb{1}, X, Y, Z\}\,. \tag{103}$$

The n-qubit Pauli operators form a group up to phases, the so-called n-qubit Pauli group:

$$\mathrm{P}_n := \{i^t \sigma_1 \otimes \cdots \otimes \sigma_n \mid t \in \mathbb{Z}_4,\, \sigma_i \in \{\mathbb{1}, X, Y, Z\}\}\,. \tag{104}$$

To see that P_n forms a subgroup of Cl_n, note that $Z = S^2$, $HZH = X$, and $Y = iXZ$. It turns out that we can characterize the Clifford group uniquely through the Pauli group: Clifford unitaries are exactly those unitaries that conjugate Paulis into Paulis, up to a phase. In group-theoretic terms, the Clifford group is the normalizer of the Pauli group within the unitary group, at least if we ignore global phases. In formula, we thus have

$$\mathrm{U}(1)\mathrm{Cl}_n = \left\{U \in \mathrm{U}(2^n) \mid U\mathrm{P}_n U^\dagger = \mathrm{P}_n\right\}. \tag{105}$$

Here, we explicitly added arbitrary global phases ($\mathrm{U}(1)$) to Cl_n as these are present on the right hand side as well.

The normalizer definition shows that a Clifford unitary is fully determined by its action on the Pauli generators X_j and Z_j, which act as X and Z on the j-th qubit, respectively. Every Pauli operator σ is a product of these generators, say

$$\sigma \propto \prod_{j=1}^{n} X_j^{x_j} Z_j^{z_j}, \tag{106}$$

where we ignored the global phase for now and the $2n$ bits x_j and z_j indicate whether σ features a X or Z (or both, a Y) on the j-th qubit. The action of a Clifford unitary U on σ can then be computed from its action on generators:

$$U\sigma U^\dagger \propto \prod_{j=1}^{n} (UX_j^{x_j}U^\dagger)(UZ_j^{z_j}U^\dagger), \tag{107}$$

The action on generators can be encoded in a binary $2n \times 2n$ matrix, a enormous compression over a generic complex $2^n \times 2^n$ matrix! This is the elementary mechanism behind efficient stabilizer simulation; the binary convention is reviewed in more detail in Sec. A.1.

For the above defined Clifford gates, the action on generators is as follows:

$$HXH = Z, \qquad HZH = X, \qquad SXS^\dagger = Y, \qquad SZS^\dagger = Z. \tag{108}$$

For a controlled-NOT with control c and target t,

$$X_c \longmapsto X_c X_t, \qquad Z_c \longmapsto Z_c, \qquad X_t \longmapsto X_t, \qquad Z_t \longmapsto Z_c Z_t. \tag{109}$$

The qudit case. The definition of the Clifford group can be readily extended to higher-dimensional qudits of dimension q. We will here focus on the case that $q > 2$ is **prime**, as this will matter for the design properties of the Clifford group.

We start by generalizing the generators of the qubit Clifford group. To this end, let $\omega_q := e^{2\pi i/q}$ be a primitive q-th root of unity, and let 2^{-1} be the inverse of 2 modulo q. Define

$$H_q|x\rangle := \frac{1}{\sqrt{q}} \sum_{y=0}^{q-1} \omega_q^{xy}|y\rangle \qquad S_q|x\rangle := \omega_q^{2^{-1}x(x-1)}|x\rangle \qquad CX_q|x,y\rangle := |x, y \oplus x\rangle, \tag{110}$$

where $y \oplus x$ denotes addition modulo q. Then, we again define the n-qudit Clifford group $\mathrm{Cl}_n(q)$ to be the group generated by H_q, S_q on every qudit and CX_q on every pair of qudits. This is again a finite subgroup of $\mathrm{U}(q^n)$ of order $q^{O(n^2)}$.

Equivalently, we can define $\mathrm{Cl}_n(q)$ in terms of a qudit version of Pauli operators, defined by

$$Z_q|x\rangle := \omega_q^x|x\rangle, \qquad X_q|x\rangle := |x \oplus 1\rangle, \qquad Y_q := \omega_q^{2^{-1}} X_q^\dagger Z_q^\dagger. \tag{111}$$

The qudit Pauli group is then given as

$$\mathrm{P}_n(q) := \left\{ \omega_q^c \bigotimes_{j=1}^{n} X_q^{a_j} Z_q^{b_j} \,\middle|\, c, a_j, b_j \in \mathbb{Z}_q \right\}. \tag{112}$$

As in the qubit case, qudit Clifford unitaries map qudit Paulis to qudit Paulis, up to a phase.

The Clifford group as a design. Finally, we have the following result on the design properties of the Clifford group.

Theorem 2.3. *Let q be prime. Then, the Clifford group $\mathrm{Cl}_n(q)$ is a unitary 2-design but not a 3-design if q is odd, and a unitary 3-design but not a 4-design if $q = 2$.*

We will not prove this theorem here. It is not particularly difficult (see e.g. Ref. [2, Sec. 12.2] for a summary), but it requires a slightly deeper analysis of the structure of the Clifford group which shall not be the focus of this course.

2.6 Further reading

Weingarten calculus is related to Wick calculus for Gaussian integrals, which was found in the 50s by physicists in the context of quantum field theories (and decades earlier by the mathematician Isserli). The analogous calculus for Haar integrals was first studied by Weingarten [3] in the late 1970s. Driven by the increasing interest in random matrix theory, Collins [4] then developed the calculus systematically in the early 2000s and named it after Weingarten. Collins's works contain deep insights on the properties and combinatorics of the Weingarten function and related objects, based on group and representation theory. An introduction by him and others can be found in Ref. [5].

We note that Weingarten calculus also works for $d < k$ (the "unstable regime") with the caveat that $a = Gc$ no longer has a unique solution. A somewhat convenient solution is given by replacing the inverse with the *pseudo-inverse* $W = G^+$.

The concept of unitary designs was introduced by Dankert, Cleve, Emerson, and Livine [6], and its mathematical foundation was subsequently refined by Gross, Audenaert, and Eisert [7]. The name 'unitary design' is adapted from the analogous concept over the sphere, so-called spherical designs. The fact that the Clifford group is a unitary 2-design even predates the invention of this concept [8, 9]. Later, it was shown that the multi-qubit Clifford group is a unitary 3-design but not a 4-design, and that in higher local dimensions, the barrier is already at $k = 2$ [10, 11]. An answer to the puzzling question of why essentially the only natural group-based example of a high-order unitary design is the Clifford group was given by Bannai, Navarro, Rizo, and Tiep [12], but the heavy lifting was already done way earlier by Guralnick and Tiep [13] (see also the summary in Ref. [14, Sec. 5]). These results imply that the Clifford groups are the only unitary designs given by groups on multi-qudit systems. In particular, we cannot find higher-order designs with group properties.

Most of the early work was focused on unitary 2-designs where the corresponding integrals can be computed explicitly with Schur's lemma instead of Weingarten calculus. Nevertheless, the ideas of Weingarten calculus have also proven useful for second moments (cf. **??**). Higher-order calculations can seldom be performed without Weingarten calculus and thus it has become a central tool in analyzing random unitaries in quantum information and many-body physics. A pedagogical introduction to Haar integration, Weingarten calculus, and unitary designs from this perspective can also be found in Ref. [15].

2.7 Additional exercises

This section contains additional exercises, supplementing the ones in the main text. The first collection of exercises is meant to train the basic concepts introduced in Sec. 2.2.

Exercise 2.11 (Permutation fluency: commuting, inverses, and cycles)**.** *The notes already contain a noncommuting product and elementary cycle examples. Complete that picture as follows.*

(a) *For each of the following pairs, compute $\sigma\tau$ and $\tau\sigma$, write the answer in cycle notation, and decide whether the two permutations commute:*

$$\sigma_1 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 4 & 3 \end{pmatrix}, \qquad \tau_1 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix}, \tag{113}$$

$$\sigma_2 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 4 & 2 & 3 & 5 \end{pmatrix}, \qquad \tau_2 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 2 & 3 & 4 & 1 \end{pmatrix}. \tag{114}$$

Explain the second answer using the supports of the two permutations.

(b) *Consider*

$$\rho = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 3 & 2 & 8 & 5 & 7 & 1 & 4 & 6 \end{pmatrix}. \tag{115}$$

Find ρ^{-1} in two-line notation. Then find the disjoint-cycle decomposition, the cycle structure $\lambda(\rho)$, and the number of cycles $\#(\rho)$. Use the cycle decomposition as an independent check of your inverse.

(c) *Decompose*

$$\eta = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 8 & 2 & 4 & 9 & 10 & 5 & 7 & 6 & 3 & 1 \end{pmatrix} \tag{116}$$

into disjoint cycles, including fixed points. Determine $\lambda(\eta)$ and $\#(\eta)$.

Exercise 2.12 (Cycle contractions for replica traces). *Let $A_1, \dots, A_k \in \mathcal{L}(\mathcal{H})$ and $\pi \in \mathrm{S}_k$.*

(a) *Starting from the definition of the trace in a product basis, prove*

$$\mathrm{tr}[(A_1 \otimes \cdots \otimes A_k) R_\pi] = \prod_{(a_1 \dots a_\ell) \in \mathcal{C}(\pi)} \mathrm{tr}(A_{a_1} A_{a_\ell} \cdots A_{a_2}), \tag{117}$$

where $\mathcal{C}(\pi)$ is the set of disjoint cycles, including fixed points. Explain why the order inside each trace is reversed by our convention for R_π.

(b) *Set every $A_j = \mathbb{1}$. Find the cycle decomposition and compute $\mathrm{tr}(R_\pi)$ for*

$$\pi_1 = (1\,4\,2)(3), \qquad \pi_2 = (1\,2)(3\,4\,5), \qquad \pi_3 = (1)(2\,6\,4\,3\,5), \tag{118}$$

$$\pi_4 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 6 & 7 & 2 & 1 & 5 & 3 & 4 \end{pmatrix}, \tag{119}$$

$$\pi_5 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 3 & 8 & 5 & 2 & 4 & 1 & 7 & 6 \end{pmatrix}. \tag{120}$$

(c) *Let k be even and set $\mathcal{O}_k = (A \otimes B)^{\otimes k/2}$, with no assumption that A and B commute. Use Eq. (117) to compute $\mathrm{tr}(R_\pi \mathcal{O}_k)$ for*

$$(12) \in \mathrm{S}_2, \tag{121}$$

$$(12)(35)(46), \quad (1532)(46), \quad (135)(426) \in \mathrm{S}_6, \tag{122}$$

$$(28)(16)(735)(4), \quad (281)(67345) \in \mathrm{S}_8\,. \tag{123}$$

Give the order of the factors inside every nontrivial trace explicitly.

Exercise 2.13 (The $k = 3$ Gram and Weingarten matrices). *Assume $d \geq 3$. Order the elements of S_3 as*

$$e,\ (12),\ (13),\ (23),\ (123),\ (132). \tag{124}$$

Recall

$$G_{\sigma,\pi} = \mathrm{tr}(R_\sigma^\dagger R_\pi) = d^{\#(\sigma^{-1}\pi)}. \tag{125}$$

(a) *Construct the full* 6×6 *Gram matrix. Explain why its entries can only be* d^3, d^2, d.

(b) *By conjugation symmetry, the entries of* $W = G^{-1}$ *depend only on whether* $\sigma^{-1}\pi$ *is the identity, a transposition, or a three-cycle. Denote these three values by* a, b, c*, respectively. Solve for* a, b, c*, write the full matrix* W*, and verify one diagonal and the two inequivalent off-diagonal equations in* $GW = \mathbb{1}$.

(c) *Determine the leading power of* d^{-1} *of* a, b, c*. Relate this hierarchy to the approximate orthogonality of distinct permutation operators at large* d.

Exercise 2.14 (Symmetric-subspace contractions). *Let*

$$P_{\mathrm{Sym},k,d} = \frac{1}{k!} \sum_{\pi \in \mathrm{S}_k} R_\pi \tag{126}$$

be the projector onto the symmetric subspace of $(\mathbb{C}^d)^{\otimes k}$*. Compute the contraction* $C_{\mathrm{Sym}}(\mathcal{A}_k) := \mathrm{tr}(\mathcal{A}_k P_{\mathrm{Sym},k,d})$ *in the following cases.*

(a) $k = 2$ *and* $\mathcal{A}_2 = A \otimes B$.

(b) $k = 4$ *and* $\mathcal{A}_4 = A^{\otimes 4}$*. Organize the sum by the five cycle types of* S_4*, rather than listing all 24 permutations.*

(c) *Generic* k *and* $\mathcal{A}_k = (|0\rangle\langle 0|)^{\otimes k}$.

(d) $k = 2, 4$ *and* $\mathcal{A}_k = P^{\otimes k}$*, where* $P = P^\dagger$*,* $P^2 = \mathbb{1}$*, and* $\mathrm{tr}\, P = 0$.

(e) $k = 2m$ *and*

$$\mathcal{A}_{2m} = (|0\rangle\langle 0|)^{\otimes m} \otimes (|1\rangle\langle 1|)^{\otimes m}, \qquad \langle 0|1\rangle = 0. \tag{127}$$

Derive a formula valid for every m*, then evaluate it for* $k = 2, 4, 6$.

The following three exercises revisit the main ideas of this section at increasing levels of difficulty. They are designed to be attempted after completing the exercises on permutations and Haar integration.

Exercise 2.15 (Basic: the first Haar moment from invariance). *For* $A \in \mathcal{L}(\mathbb{C}^d)$ *we have*

$$\mathrm{M}_1(A) = \int_{\mathrm{U}(d)} UAU^\dagger \, \mathrm{d}U.$$

(a) *Use left invariance of Haar measure to prove that* $V\,\mathrm{M}_1(A)V^\dagger = \mathrm{M}_1(A)$ *for every* $V \in \mathrm{U}(d)$.

(b) *Give an elementary matrix proof that an operator commuting with every unitary is a scalar multiple of* $\mathbb{1}$*. One possible route is to commute first with all diagonal phase unitaries and then with all basis permutations.*

(c) *Take the trace to determine the scalar and conclude that*

$$\mathrm{M}_1(A) = \frac{\mathrm{tr}(A)}{d}\,\mathbb{1}.$$

What does this formula give when A *is a density matrix, and when* A *is traceless?*

Exercise 2.16 (Intermediate: a 1-design that is not a 2-design). *Let* μ_{P} *be the uniform distribution on the four one-qubit Pauli unitaries* $\{\mathbb{1}, X, Y, Z\}$.

(a) *Prove that μ_{P} is an exact unitary 1-design, this is that*

$$\frac{1}{4}\sum_{P\in\{\mathbb{1},X,Y,Z\}} PAP^\dagger = \frac{\mathrm{tr}(A)}{2}\mathbb{1}.$$

Hint: Use that $\{\mathbb{1},X,Y,Z\}$ is a basis for the 2×2 matrices.

(b) *Apply the same random Pauli to both copies of $|0\rangle\langle 0|$ and compute*

$$\frac{1}{4}\sum_{P}(P|0\rangle\langle 0|P^\dagger)^{\otimes 2}.$$

(c) *For a Haar-random one-qubit pure state, use*

$$\int |\psi\rangle\langle\psi|^{\otimes 2}\,\mathrm{d}\psi = \frac{\mathbb{1}+R_{(12)}}{2(2+1)} = \frac{P_{\mathrm{Sym},2,2}}{3}$$

to show that the operator in part (b) is not the Haar second moment. Exhibit one normalized two-qubit vector on which the two operators have different expectation values. Conclude that μ_{P} is not a unitary 2-design.

Exercise 2.17 (Advanced: the frame potential as a design certificate)**.** *Let μ be a probability measure on $\mathrm{U}(d)$ and let $k\geq 1$. On the vectorized k-replica operator space define*

$$R_k(U) := U^{\otimes k}\otimes \overline{U}^{\otimes k}, \qquad M_{\mu,k} := \mathbb{E}_{U\sim\mu}[R_k(U)].$$

The order-k frame potential is

$$\mathcal{F}_k(\mu) := \mathbb{E}_{U,V\sim\mu}\left[|\,\mathrm{tr}(U^\dagger V)|^{2k}\right],$$

where U and V are independent.

(a) *Prove that $\mathcal{F}_k(\mu) = \mathrm{tr}(M_{\mu,k}^\dagger M_{\mu,k})$.*

(b) *Let $M_{\mathrm{Haar},k}$ denote the Haar moment operator. Show directly from Haar invariance that it is the orthogonal projector onto the vectors fixed by every $R_k(U)$ and that*

$$M_{\mu,k}M_{\mathrm{Haar},k} = M_{\mathrm{Haar},k}M_{\mu,k} = M_{\mathrm{Haar},k}.$$

(c) *In the stable range $d\geq k$, use the permutation description of the commutant to justify $\mathrm{rank}(M_{\mathrm{Haar},k}) = k!$ and prove the identity*

$$\mathcal{F}_k(\mu) - k! = \mathrm{tr}\left[(M_{\mu,k}-M_{\mathrm{Haar},k})^\dagger(M_{\mu,k}-M_{\mathrm{Haar},k})\right].$$

Deduce both the lower bound $\mathcal{F}_k(\mu)\geq k!$ and the equality criterion: equality holds precisely when μ is an exact unitary k-design.

3 Many-Body Properties from Classical Shadows

Modern experiments can prepare and coherently manipulate quantum states at historically unmatched levels of quality. However, even with perfect control, we are left with the fundamental problem of extracting information from a quantum system: physical quantitities such as an energy, correlation functions, fidelities, or entanglement measures. Although we can, in

principle, reconstruct all d^2-1 real parameters of a quantum state using full tomography, this is prohibitively expensive for large system sizes where $d = q^n$.

In this section, we consider the problem of experimentally estimating an expectation value of the form $\mathrm{tr}(O\rho)$, where ρ is the quantum state that is experimentally accessible and O is the observable of interest. Importantly, we do not assume that O can be measured directly. In the following, we will show that for some interesting classes of observables this problem can be solved using *classical shadows*. The main idea is to randomize measurement bases using a suitable ensemble of random unitaries, resulting in a partial classical representation of the quantum state (the "shadow"). The shadow can then be used by a classical computer to predict expectation values. Intriguingly, the same classical shadow can be reused for many observables at once, at a moderate (logarithmic) overhead in the size of the shadow (i.e. the number of measurements).

We will first review some basics on quantum measurements and the (partial) reconstruction of quantum states from measurement data using linear inversion. Afterwards, we introduce the idea of classical shadows and treat the most common instances of the protocol based on Clifford unitaries. Finally, we explain carefully how classical shadows can be used to estimate a nonlinear quantity, the purity.

3.1 Reconstructing quantum states from measurements

Suppose we have access to many copies of a quantum state ρ and we want to obtain a classical description of ρ from measurement statistics, e.g., in the form of a density matrix. This reconstruction task is also called *quantum state tomography*. The central question is how we should choose our measurements so that this reconstruction succeeds for any state, and how many measurements we need in total.

Let us consider quantum measurements in a basis $|\varphi_1\rangle, \ldots, |\varphi_d\rangle$ of the Hilbert space $\mathcal{H}$. The probability distribution over outcomes of this measurement follows Born's rule:

$$\text{Born's rule:} \qquad p(i|\rho) := \langle\varphi_i|\rho|\varphi_i\rangle = \mathrm{tr}(E_i\rho) \quad \text{where} \quad E_i := |\varphi_i\rangle\langle\varphi_i|\,. \tag{128}$$

Clearly, measuring in a basis cannot be enough for state reconstruction because we can only assess the diagonal elements of ρ in the chosen basis. Hence, we are insensitive to any coherence in the basis: For instance, we cannot distinguish between any two normalized states of the form $(|0\rangle + e^{i\alpha}|1\rangle)/\sqrt{2}$ by looking at the diagonal elements of their density matrices only.

However, it turns out that if we combine measurements in sufficiently many bases, state reconstruction is possible and we will explicitly describe such a reconstruction algorithm in a moment. Before, we introduce some new notation that will simplify both the following computations as well as those in the later sections.

Operator bra-ket notation. Recall that the vector space $\mathrm{L}(\mathcal{H})$ of linear operators on $\mathcal{H}$ is equipped with the *trace inner product*:

$$(A|B) := \mathrm{tr}(A^\dagger B)\,. \tag{129}$$

With this notation, we can now express Born probabilities as

$$p(i|\rho) = (E_i|\rho)\,, \tag{130}$$

with the interpretation of projecting the state ρ onto the pure state $E_i = |\varphi_i\rangle\langle\varphi_i|$. Analogous to the ordinary bra-ket notation based on the inner product $\langle\cdot|\cdot\rangle$ on $\mathcal{H}$, we now introduce *operator bra-ket notation* by defining *operator kets* and *operator bras* as

$$|B) := B\,, \qquad\qquad (A| : B \equiv |B) \mapsto (A|B)\,. \tag{131}$$

Note that operator bras $(A|$ are linear forms on $\mathrm{L}(\mathcal{H})$ (dual vectors), just as ordinary bras $\langle\psi|$ are dual vectors on $\mathcal{H}$. Again similar to the ordinary bra-ket notation, we introduce *operator bra-kets* as the outer products

$$|A)(B|:\ C\equiv|C)\mapsto|A)(B|C)\,, \tag{132}$$

As we will see in a moment, this gives us a convenient way to write down linear maps on operators, similar to expressions of the form $A=\sum_{i,j}A_{i,j}|i\rangle\langle j|$ for the ordinary bra-ket notation. For instance, the Weingarten expansion (85) attains the following appealing form in this notation:

$$\mathsf{M}_k=\sum_{\pi,\sigma\in S_k}W_{\pi,\sigma}|R_\pi)(R_\sigma|\,, \tag{133}$$

Following the quantum information language, we will refer to linear maps on operators, such as (132), as *superoperators* (as these are "operators on operators"). Important examples of superoperators are *quantum channels*.

Reconstruction via linear inversion. Equipped with the new notation, let us come back to the reconstruction of quantum states from measurements. Consider the following superoperator:

$$S:=\sum_{i=1}^{d}|E_i)(E_i|\,. \tag{134}$$

We call S the *frame (super)operator* associated to the basis $(\varphi_i)_{i\in[d]}$. In the classical shadows literature, S is also called the *measurement channel*. Note that we have $S(\rho)=\sum_{i=1}^{d}p(i|\rho)E_i$ for any state ρ and in this sense, the inability of reconstructing a state from a basis measurement is encoded in the non-injectivity of S.

However, it turns out that if we combine measurements in sufficiently many bases, we can perform successful state reconstruction. Suppose we are given bases $(\varphi_{i,j})_{i\in[d]}$ for $j=1,\dots,m$ and we perform measurements in any of those. The frame operator of this combined measurement strategy is a convex combination of the single frame operators:

$$S=\frac{1}{m}\sum_{j=1}^{m}S_j=\frac{1}{m}\sum_{j=1}^{m}\sum_{i=1}^{d}|E_{i,j})(E_{i,j}|\,. \tag{135}$$

Suppose that this combined frame operator S is invertible, which is –as it turns out– necessary for reconstruction (cf. Exercise 3.1). Then, we can perform the simple manipulation

$$\rho=S^{-1}S(\rho)=\frac{1}{m}\sum_{j=1}^{m}\sum_{i=1}^{d}S^{-1}|E_{i,j})(E_{i,j}|\rho)=\frac{1}{m}\sum_{j=1}^{m}\sum_{i=1}^{d}p(i,j|\rho)\tilde{E}_{i,j}\,, \tag{136}$$

where $p(i,j|\rho)=(E_{i,j}|\rho)$ are again the Born probabilities and $\tilde{E}_{i,j}:=S^{-1}(E_{i,j})$ are the *dual* measurement elements. This gives a simple recipe for reconstructing ρ from the measured Born probabilities $p(i,j|\rho)$, called *linear inversion tomography*.

The number of copies needed for tomography depends on the allowed measurement model. For nonadaptive protocols that measure each copy separately, learning an arbitrary rank-r state to trace-distance error ε requires $\Omega(dr^2/\varepsilon^2)$ copies in the standard constant-success regime [16]. Collective measurements on many copies can improve the rank dependence: a general upper bound is $O[dr\log(d/\varepsilon)/\varepsilon^2]$ copies [17]. The distinction matters operationally, but both costs contain the Hilbert-space dimension $d=q^n$ and are therefore exponential in the number of qudits without additional structural promises.

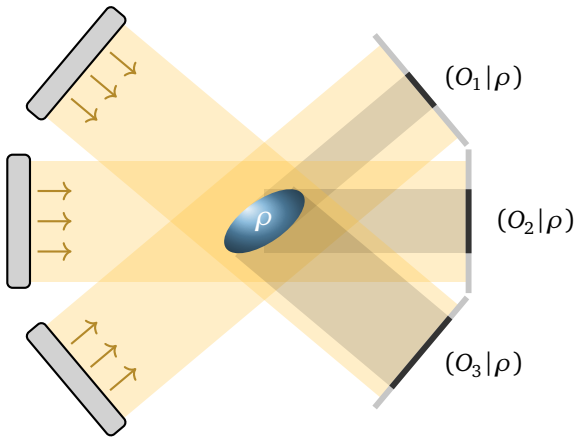


Figure 3: Shadow tomography is the task of determining only some of the features of a quantum state without performing full quantum state tomography. We do this by considering only few measurements of the state akin to the shadows of an object under few illumination angles as shown in this cartoon: each shadow reveals only one feature $(O_s|\rho)$ of the object ρ.

Exercise 3.1 (Informationally complete measurements). *In general, state reconstruction can only be successful if the Born probabilities differ for any two states $\rho \neq \rho'$. If this is the case, we call the measurement* informationally complete *(IC).*

(a) Show that measurements in several bases are informationally complete if and only if every operator $X \in \mathbb{C}^{d\times d}$ is a linear combination of the $E_{i,j}$, i.e., $X = \sum_{i,j} x_{i,j} E_{i,j}$. Equivalently, the $E_{i,j}$ span $\mathrm{L}(\mathcal{H})$. Hint: *Consider the linear map $V(x) := \sum_{i,j} x_{i,j} E_{i,j}$ from $\mathbb{C}^{d\times m}$ to* $\mathrm{L}(\mathcal{H})$ *and its adjoint $V^\dagger$ (bra-ket notation might be useful).*

(b) Show that measurements in several bases are informationally complete if and only if the frame operator (134) *is invertible.* Hint: *Show that $S = \frac{1}{m} VV^\dagger$ where V is the map from part (a).*

Exercise 3.2. *For those interested in quantum information: Show that the frame operator* (134) *is indeed a quantum channel by computing its Choi matrix.*

3.2 Shadow estimation with randomized measurements

In the previous section, we have seen that quantum state tomography can be realized based on simple linear inversion, but it requires exponentially many measurements rendering it very inefficient. But what if we are not interested in reconstructing the full state ρ, but only some of its *features*? Concretely, let us say that we want to reconstruct M linear functions of ρ, which we can write as $(O_s|\rho)$ $(s = 1, \ldots, M)$. We call this task *shadow tomography* as we do not observe the full state, but only some of features, similar to the shadow of an object that is illuminated from a certain direction, cf. Fig. 3. Can shadow tomography be done more efficiently than performing full quantum state tomography?

The answer is *yes, but it depends*, namely on the allowed measurement strategy (can we access only single copies of ρ, or some of them at once) and also on the observables O_s. In the following, we will focus on single-copy measurements and show how randomized measurements can be used to estimate some classes of expectation values very efficiently.

We thus study measurements in random rotations of the computational basis, i.e. measurements are performed in the bases

$$|U,x\rangle := U^\dagger|x\rangle\,, \qquad E_{U,x} := |U,x\rangle\langle U,x| = U^\dagger|x\rangle\langle x|U\,, \qquad x \in [d]\,, \tag{137}$$

where the unitary U is sampled uniformly at random from a finite set $\mathcal{G} \subset \mathrm{U}(d)$ (here the adjoint comes from using the Heisenberg picture). The frame operator is then a probabilistic mixture of the frame operators in the rotated bases (137):

$$S = \frac{1}{|\mathcal{G}|} \sum_{U\in\mathcal{G}} \sum_{x=1}^{d} |E_{U,x})(E_{U,x}| \,. \tag{138}$$

The underlying observation is that we can use the linear inversion trick (136) to obtain the following identity for expectation values:

$$\begin{aligned} (O|\rho) &= (O|S^{-1}S|\rho) \\ &= \frac{1}{|\mathcal{G}|} \sum_{U\in\mathcal{G}} \sum_{x=1}^{d} (O|S^{-1}|E_{U,x})(E_{U,x}|\rho) \\ &= \sum_{U\in\mathcal{G}} \sum_{x=1}^{d} (O|\tilde{E}_{U,x}) p(x|U,\rho) p(U) \,, \end{aligned} \tag{139}$$

where we set $p(x|U,\rho) = (E_{U,x}|\rho)$ and $p(U) = 1/|\mathcal{G}|$. We chose the later notation to highlight that we are sampling pairs (U,x) according to the joint distribution $p(x|U,\rho)p(U)$. This suggests to estimate the expectation value $(O|\rho)$ using the following protocol.

Definition 3.1: Shadow estimation protocol

Repeat the following steps N times:

(i) Sample a unitary U uniformly at random from the set $\mathcal{G}$ and apply it to ρ

(ii) Measure in the computational basis resulting in outcome x

(iii) Record the pair (U,x)

Estimate $(O|\rho)$ using the mean estimator of $\hat{o}_{U,x} := (O|\tilde{E}_{U,x})$ on the N samples.

Equation (139) then guarantees that shadow estimation (Def. 3.1) converges in expectation to $(O|\rho)$ (i.e. $\hat{o}$ is an unbiased estimator). But how fast does this estimator converge, i.e. how many measurements N do we have to perform? As we will see below, this depends a lot on the used ensemble $\mathcal{G}$ and on the observable O (as well as on the state).

To answer the question on sample complexity, as well as other questions on the efficiency and practicability of the shadow estimation protocol (Def. 3.1), we have to analyze it in more depth. As a first step we will compute the frame operator (134) and its inverse, as it is needed in the construction of the estimator $\hat{o}_{U,x} = (O|S^{-1}|E_{U,x})$. The frame operator involves two invocations of U and $U^\dagger$, respectively, which we can explicitly see by considering the matrix elements

$$(A|S|B) = \sum_{x=1}^{d} \frac{1}{|\mathcal{G}|} \sum_{U\in\mathcal{G}} (A|E_{U,x})(E_{U,x}|B) = \sum_{x=1}^{d} \frac{1}{|\mathcal{G}|} \sum_{U\in\mathcal{G}} \mathrm{tr}\big[A^\dagger \otimes B\, (U^\dagger)^{\otimes 2} |x\rangle\langle x|^{\otimes 2} U^{\otimes 2}\big] \tag{140}$$

Hence, if we take $\mathcal{G}$ to be a unitary 2-design, we can replace the average over $\mathcal{G}$ in Eq. (140) by an average over the unitary group, computable via Weingarten calculus. In the following, we will see that this brings S into a very simple form.

3.2.1 Clifford measurements

Computing the shadow estimator. Let us thus assume that we are randomizing over a unitary 2-design, or equivalently, over Haar-random unitaries, and compute the frame operator by $k = 2$ Weingarten calculus. To this end, we evaluate the integral (cf. Example 2.8):

$$(A|S|B) = \sum_{x=1}^{d} \int_{\mathrm{U}(d)} \mathrm{tr}(A^\dagger \otimes B\, U^{\otimes 2}|x\rangle\langle x|^{\otimes 2} U^{\otimes 2,\dagger})\, \mathrm{d}U \tag{141}$$

$$= \sum_{x=1}^{d} \sum_{\pi,\sigma\in \mathrm{S}_2} W_{\pi,\sigma}\, \mathrm{tr}(R_\sigma^\dagger A^\dagger \otimes B)\, \mathrm{tr}(R_\pi |x\rangle\langle x|^{\otimes 2}) \tag{142}$$

$$= d \sum_{\pi,\sigma\in \mathrm{S}_2} W_{\pi,\sigma}\, \mathrm{tr}(R_\sigma^\dagger A^\dagger \otimes B) \tag{143}$$

$$= d\frac{(d-1)!}{(d+1)!} \sum_{\sigma\in \mathrm{S}_2} \mathrm{tr}(R_\sigma^\dagger A^\dagger \otimes B) \tag{144}$$

$$= \frac{1}{d+1}\left(\mathrm{tr}(A^\dagger)\,\mathrm{tr}(B) + \mathrm{tr}(A^\dagger B)\right). \tag{145}$$

In the last step, we used that $\mathrm{tr}(R_{(12)} A^\dagger \otimes B) = \mathrm{tr}(A^\dagger B)$ which can be easily verified using the graphical notation (cf. Exercise 2.12). Alternatively, we could have also used the 2-design formula (100). Next, we note that

$$\frac{1}{d+1}\left(\mathrm{tr}(A^\dagger)\,\mathrm{tr}(B) + \mathrm{tr}(A^\dagger B)\right) = \frac{1}{d+1}\Big((A|\mathbb{1})(\mathbb{1}|B) + (A|\mathrm{id}|B)\Big) \tag{146}$$

$$= \frac{1}{d+1}\Big(d(A|\mathcal{D}|B) + (A|\mathrm{id}|B)\Big), \tag{147}$$

where $\mathcal{D} = \frac{1}{d}|\mathbb{1})(\mathbb{1}|$ is the completely depolarizing channel, acting as $\mathcal{D}(X) = \mathrm{tr}(X)\mathbb{1}/d$. Hence, we have shown that

$$S = \frac{1}{d+1}\big(d\mathcal{D} + \mathrm{id}\big) \equiv \mathcal{D}_{\frac{d}{d+1}}\,, \tag{148}$$

is a convex combination of the completely depolarizing channel and the identity, and thus a depolarizing channel of strength $d/(d+1)$. Moreover, as both channels are trace-preserving (TP), so is S. It is then staightforward to invert S:

$$Y = S(X) = \frac{\mathrm{tr}(X)\mathbb{1} + X}{d+1} \quad \Rightarrow \quad X = (d+1)Y - \mathrm{tr}(Y)\mathbb{1} \quad \Rightarrow \quad S^{-1} = (d+1)\mathrm{id} - d\mathcal{D}, \tag{149}$$

where we used that S is TP and thus $\mathrm{tr}X = \mathrm{tr}\,S(X) = \mathrm{tr}\,Y$.

Using the explicit form of S^{-1}, we can write the shadow estimator $\hat{o}$ as follows

$$\hat{o}_{U,x} = (O|S^{-1}|E_{U,x}) = (d+1)(O|E_{U,x}) - \mathrm{tr}(O) = (d+1)\,\mathrm{tr}(OU^\dagger|x\rangle\langle x|U) - \mathrm{tr}(O). \tag{150}$$

Thus, the evaluation of $\hat{o}_{U,x}$ requires us to classically compute the expectation value of O in the rotated basis states $U^\dagger|x\rangle$. We will comment on the efficiency of this computation later.

Number of measurements. A central question remains: *How efficient is shadow estimation in terms of the required number of measurements (*sample complexity*)?* This requires us to bound the convergence of the mean estimator associated to $\hat{o}$ for which we will use Chebyshev's inequality.

Lemma 3.1 (Chebyshev's inequality). *Let X be a random variable and $\varepsilon > 0$. Then, we have*

$$\Pr[|X - \mathbb{E}[X]| > \varepsilon] \leq \frac{\mathrm{Var}[X]}{\varepsilon^2}. \tag{151}$$

In particular, if $X_1, \dots, X_N$ are independent and identically distributed (iid) random variables with mean μ and variance σ^2, and $\overline{X} := \frac{1}{N}\sum_{i=1}^N X_i$, then

$$\Pr[|\overline{X} - \mu| > \varepsilon] \leq \frac{\sigma^2}{N\varepsilon^2}. \tag{152}$$

Chebyshev's inequality implies that we need $N \geq \varepsilon^{-2}\delta^{-1}\mathrm{Var}[\hat{o}]$ many samples to get an ε-approximate estimate of $\mathrm{tr}(O\rho)$ with probability at least $1-\delta$. Hence, we need the variance of $\hat{o}$ to be sufficiently small to get a decent bound on the sample complexity.

Before we compute $\mathrm{Var}[\hat{o}]$, we note that

$$\mathrm{Var}[\hat{o}] = \mathbb{E}[(\hat{o} - \mathbb{E}[\hat{o}])^2]. \tag{153}$$

but $\hat{o} - \mathbb{E}[\hat{o}]$ only depends on the traceless part of O, i.e. $O_0 = O - \mathrm{tr}(O)\mathbb{1}/d$, since

$$\begin{aligned}
\hat{o}_{U,x} - \mathbb{E}[\hat{o}] &= (d+1)\,\mathrm{tr}(OU^\dagger|x\rangle\langle x|U) - \mathrm{tr}(O) - \mathrm{tr}(O\rho) && (154)\\
&= (d+1)\,\mathrm{tr}(O_0U^\dagger|x\rangle\langle x|U) - \mathrm{tr}(O_0\rho) + \mathrm{tr}(O)\left(\frac{d+1}{d} - 1 - \frac{1}{d}\right) && (155)\\
&= (d+1)\,\mathrm{tr}(O_0U^\dagger|x\rangle\langle x|U) - \mathrm{tr}(O_0\rho). && (156)
\end{aligned}$$

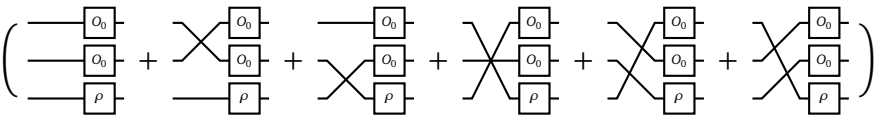
Thus, we can use the traceless observable O_0 instead of O in the future computations. Next, we use $\mathrm{Var}[\hat{o}] = \mathbb{E}[\hat{o}^2] - \mathbb{E}[\hat{o}]^2$ and focus on the computation of $\mathbb{E}[\hat{o}^2]$.

Using the explicit form (150) of $\hat{o}$, we then compute

$$\begin{aligned}
\mathbb{E}[\hat{o}_0^2] &= \sum_{U\in\mathcal{G}}\sum_{x=1}^d (O_0|\tilde{E}_{U,x})^2 p(x|U,\rho)p(U) && (157)\\
&= \frac{(d+1)^2}{|\mathcal{G}|}\sum_{U\in\mathcal{G}}\sum_{x=1}^d (O_0|E_{U,x})^2 (E_{U,x}|\rho) && (158)\\
&= \frac{(d+1)^2}{|\mathcal{G}|}\sum_{U\in\mathcal{G}}\sum_{x=1}^d (O_0^{\otimes 2}\otimes\rho\,|E_{U,x}^{\otimes 3}) && (159)\\
&= \frac{(d+1)^2}{|\mathcal{G}|}\sum_{U\in\mathcal{G}}\sum_{x=1}^d \mathrm{tr}\left(O_0^{\otimes 2}\otimes\rho\, U^{\otimes 3}|x\rangle\langle x|^{\otimes 3}U^{\otimes 3,\dagger}\right). && (160)
\end{aligned}$$

To compute the latter twirl, we now assume that $\mathcal{G}$ is a *unitary 3-design* such that we can apply Weingarten calculus. We then find, analogous to the computation of the frame operator (cf. Lem. 2.3):

$$\begin{aligned}
\mathbb{E}[\hat{o}_0^2] &= (d+1)^2\sum_{x=1}^d\sum_{\pi,\sigma\in S_3} W_{\pi,\sigma}\,\mathrm{tr}\left(O_0^{\otimes 2}\otimes\rho\,R_\pi\right)\mathrm{tr}(R_\sigma^\dagger|x\rangle\langle x|^{\otimes 3}) && (161)\\
&= d(d+1)^2\frac{(d-1)!}{(d+2)!}\sum_{\pi\in S_3}\mathrm{tr}\left(O_0^{\otimes 2}\otimes\rho\,R_\pi\right) && (162)\\
&= \frac{d+1}{d+2}\,\mathrm{tr}\Bigg(\quad + \quad + \quad + \quad + \quad + \quad \Bigg) && (163)\\
&= \frac{d+1}{d+2}\left(\mathrm{tr}(O_0^2) + 2\,\mathrm{tr}(O_0^2\rho)\right), && (164)
\end{aligned}$$



where we used that O_0 is traceless and thus the first, third, and fourth term in Eq. (163) vanishes. We thus get the following bound on the variance for a unitary 3-design:

$$\mathrm{Var}[\hat{o}] \leq \mathbb{E}[\hat{o}_0^2] = \frac{d+1}{d+2}\left(\mathrm{tr}(O_0^2) + 2\,\mathrm{tr}(O_0^2\rho)\right) \leq \frac{d+1}{d+2}\left(\|O_0\|_2^2 + 2\|O_0\|_\infty^2\right) \leq 3\|O_0\|_2^2. \quad (165)$$

Here, $\|X\|_2^2 = (X|X) = \mathrm{tr}(X^\dagger X)$ is the *Hilbert-Schmidt norm* and $\|X\|_\infty = \sup_\psi \|X|\psi\rangle\|$ is the *spectral* or *operator norm*. In the last step, we used Hölder's inequality, $\mathrm{tr}(O_0^2\rho) \leq \|O_0\|_\infty^2 \,\mathrm{tr}\,\rho$, as well as the general inequality $\|X\|_\infty \leq \|X\|_2$.

Discussion of shadow estimation with Clifford unitaries. In the previous derivations we assumed that $\mathcal{G}$ forms a unitary 3-design and for concreteness we will take $\mathcal{G} = \mathrm{Cl}_n$ to be the n-qubit Clifford group.[3] Based on the variance (165), we can see that shadow estimation with Clifford unitaries is sample-efficient for observables O with bounded Hilbert-Schmidt norm, $\|O_0\|_2 \leq \|O\|_2 \leq \text{const.}$ (in the sense that the number of measurements N depends only on the precision, not on the number of qubits). The most important examples of such observables are *quantum states* for which $\|\sigma\|_2 = \sqrt{\mathrm{tr}(\sigma^2)} \leq 1$. The associated expectation values then have the form $\mathrm{tr}(\rho\sigma) = (\rho|\sigma)$ and can be interpreted as *overlaps between the states*. Moreover, if $\sigma = |\psi\rangle\langle\psi|$, then $\mathrm{tr}(\rho\sigma) = \langle\psi|\rho|\psi\rangle \equiv F(\rho,\sigma)$ coincides with the usual definition of fidelity between two quantum states (for mixed states, the formula is a bit more complicated). Hence, we can use shadow estimation with Clifford unitaries to efficiently perform *fidelity estimation* with pure target states.

While shadow fidelity estimation is sample-efficient, the *classical post-processing* of the measurement data requires the evaluation of

$$\hat{o}_{U,x} = (d+1)\langle\phi|U^\dagger|x\rangle|^2 - 1. \quad (166)$$

This involves the classical simulation of the unitary evolution of $|x\rangle$ under $U^\dagger$. For the considered Clifford unitaries, it turns out that $U^\dagger|x\rangle$ can be efficiently computed on a classical computer (in time $O(n^3)$) and has an efficient classical description ($O(n^2)$ bits) that can be stored (*Gottesman-Knill theorem*).

To see why, first note that the computational-basis state $|x\rangle$ is the simultaneous $+1$ eigenstate of the n commuting Pauli operators $(-1)^{x_i}Z_i$. Indeed, we can write

$$|x\rangle\langle x| = \prod_{i=1}^n \left(\frac{\mathbb{1} + (-1)^{x_i}Z_i}{2}\right). \quad (167)$$

This means that the n Pauli operators $(-1)^{x_i}Z_i$ completely fix the density matrix of $|x\rangle$. We can now compute how this description updates under a Clifford unitary V. We have:

$$|\psi\rangle := V|x\rangle = V(-1)^{x_i}Z_i|x\rangle = (-1)^{x_i}VZ_iV^\dagger V|x\rangle = g_i|\psi\rangle\,, \quad (168)$$

with $g_i := (-1)^{x_i}VZ_iV^\dagger$. Since V is Clifford, the g_i are again (signed) Pauli operators and they commute because $[g_i,g_j] \propto [VZ_iV^\dagger, VZ_jV^\dagger] = [Z_i,Z_j] = 0$. The updated state $|\psi\rangle$ is then fully specified by the g_i, in particular

$$|\psi\rangle\langle\psi| = V|x\rangle\langle x|V^\dagger = \prod_{i=1}^n \frac{\mathbb{1}+g_i}{2}\,. \quad (169)$$

Such states are called *stabilizer states*. As noted in Sec. 2.5, a Pauli operator is fully described by two binary vectors and a phase,

$$P = i^\varphi \bigotimes_{j=1}^n X^{a_j}Z^{b_j}, \qquad a_j, b_j \in \mathbb{Z}_2, \quad \varphi \in \mathbb{Z}_4. \quad (170)$$

[3] See Sec. 3.3 for the qudit case.

Clifford conjugation is described by a simple $2n \times 2n$ matrix acting on (a, b). In particular, this update can be computed in time $O(n^2)$. To updated the entire state $|x\rangle$, we have to perform this n times, resulting in time $O(n^3)$.

However, the evaluation of $\hat{o}_{U,x}$ also involves taking the inner product with O which is efficient if the target state $|\phi\rangle$ is also a stabilizer state, i.e., it has the form (169). For general states, however, this inner product is inefficient to compute, i.e. we have to expect that this scales exponentially with the number of qubits n.

But how does this compare to other methods of fidelity estimation? The *direct fidelity estimation* protocol achieves the same task, but requires a number of measurements that depend on the target state $|\psi\rangle$ and typically scale exponentially in the number of qubits n, while the demands on the classical computer are negligible. In this sense, shadows move the complexity of fidelity estimation from measurements to the classical post-processing, which may be advantageous in near-term devices where taking measurements is more costly than running computations on a powerful computer. For a detailed comparison of fidelity estimation based on classical shadows and direct fidelity estimation, see also Ref. [18].

Predicting many expectation values at once. We note that the above estimation strategy can also be extended to estimate several expectation values $(O_1|\rho), \ldots, (O_M|\rho)$ at once. To make this precise, note that for the simultaneous estimation of all expectation values we have to choose the failure probability for the i-th mean estimator to be δ/M such that the joint failure probability is uniformly bounded by δ (union bound). Hence, we need the following number of measurements in total:

$$N \geq \frac{M}{\varepsilon^2 \delta} \max_{i=1,\ldots,M} \mathrm{Var}[\hat{o}_i], \qquad \text{(mean estimators)}, \tag{171}$$

Depending on the size of M, it may be beneficial to replace the mean estimator for each observable by a so-called *median-of-means estimator*. The idea behind the median-of-means estimator is that the data set is decomposed into K equally sized batches of size N'; within each batch an independent mean estimator $\hat{o}_i^{(j)}(N')$ is computed. Finally, we take the median over all these mean estimators, in formula

$$\hat{o}_i(N', K) = \mathrm{median}\left\{\hat{o}_i^{(1)}(N'), \ldots, \hat{o}_i^{(K)}(N')\right\}. \tag{172}$$

The advantage of this approach is that it is more robust against deviations from the mean. Indeed, a statistical analysis of the median-of-means estimator shows that the dependence on the failure probability of estimating all M expectation values is improved from M/δ to $\log(M/\delta)$ at the cost of larger constants. More precisely, choose an odd number of blocks

$$K \geq 2 \log\left(\frac{2M}{\delta}\right) \tag{173}$$

(rounded up to the next odd integer) and a block size

$$N' \geq \frac{34}{\varepsilon^2} \max_{i=1,\ldots,M} \mathrm{Var}[\hat{o}_i]. \tag{174}$$

Then $N = KN'$ samples suffice to estimate all M expectation values to additive error ε with probability at least $1 - \delta$. Ignoring the harmless integer rounding gives the compact bound

$$N \geq \frac{68}{\varepsilon^2} \log\left(\frac{2M}{\delta}\right) \max_{i=1,\ldots,M} \mathrm{Var}[\hat{o}_i]. \tag{175}$$

Here K is the number of blocks and N' is the number of samples per block. Finally, the median-of-means estimator is beneficial when

$$68\log(2M/\delta) \leq M/\delta, \tag{176}$$

which holds for $M/\delta \geq 464.76$. Assuming $\delta = 0.01$ (1% failure probability), we thus obtain $M \geq 5$.

3.2.2 Pauli measurements

In the following, we consider the ensemble given by *local* Clifford unitaries: these are unitaries of the form $U = U_1 \otimes \cdots \otimes U_n$ where $U_i \in \mathrm{Cl}_1$ are single-qubit Clifford unitaries. We denote this set by $\mathrm{LCl}_n := \mathrm{Cl}_1^{\otimes n}$. The effect of local Clifford unitaries is to locally change the basis: Instead of measuring each qubit in the Z basis, each qubit is measured in either the X, Y, or Z basis, depending on the Clifford unitary U_i. This is why these are also referred to as *Pauli measurements*. The hope in this choice of ensemble is that it may be better adjusted to measuring *local observables* of the form $O = O_1 \otimes \cdots \otimes O_r \otimes \mathbb{1}^{\otimes(n-r)}$ which only act non-trivially on r qubits.

Computing the shadow estimator. Because all of the expressions factorize, we can re-use our computations from Sec. 3.2.1. For instance,

$$\begin{aligned} E_{U,x} &= (U_1^\dagger \otimes \cdots \otimes U_n^\dagger)|x_1,\ldots,x_n\rangle\langle x_1,\ldots,x_n|(U_1 \otimes \cdots \otimes U_n) \\ &= \bigotimes_{i=1}^n U_i^\dagger |x_i\rangle\langle x_i| U_i = \bigotimes_{i=1}^n E_{U_i,x_i}\,. \end{aligned} \tag{177}$$

Hence, the frame operator becomes

$$S = \frac{1}{|\mathrm{Cl}_1|^n} \sum_{U_1,\ldots,U_n \in \mathrm{Cl}_1} \sum_{x\in\{0,1\}^n} \bigotimes_{i=1}^n |E_{U_i,x_i})(E_{U_i,x_i}| = S_1^{\otimes n}\,, \tag{178}$$

where S_1 is the local frame operator for which we can use Eq. (148) with $d = 2$:

$$S_1 = \frac{1}{|\mathrm{Cl}_1|} \sum_{U\in\mathrm{Cl}_1} \sum_{x\in\{0,1\}} |E_{U,x})(E_{U,x}| = \frac{1}{3}(2\mathcal{D} + \mathrm{id})\,. \tag{179}$$

In particular, Eq. (149) gives

$$S^{-1} = (3\mathrm{id} - 2\mathcal{D})^{\otimes n}. \tag{180}$$

For the r-local observable

$$O = O_1 \otimes \cdots \otimes O_r \otimes \mathbb{1}^{\otimes(n-r)}, \tag{181}$$

we can then use that each factor $(3\mathrm{id} - 2\mathcal{D})$ preserves the trace, so that

$$\hat{o}_{U,x} = (O|S^{-1}|E_{U,x}) \tag{182}$$

$$= \prod_{i=1}^r (O_i|3\mathrm{id} - 2\mathcal{D}|E_{U_i,x_i}) \prod_{i=r+1}^n (\mathbb{1}|E_{U_i,x_i}) \tag{183}$$

$$= \prod_{i=1}^r \hat{o}_{i,U_i,x_i} \tag{184}$$

$$= \prod_{i=1}^r \left(3\,\mathrm{tr}(O_i U_i^\dagger |x_i\rangle\langle x_i| U_i) - \mathrm{tr}(O_i)\right). \tag{185}$$

Number of measurements. Next, we compute the variance of $\hat{o}$. Because the estimator only depends on the first r qubits we can resum the remaining local terms inside the Born probability using that $\frac{1}{|\mathrm{Cl}_1|}\sum_{V\in\mathrm{Cl}_1}\sum_{y\in\{0,1\}} V^\dagger|y\rangle\langle y|V = \mathbb{1}$

$$\mathbb{E}[\hat{o}^2] = \frac{1}{|\mathrm{Cl}_1|^n}\sum_{U\in\mathrm{Cl}_1^n}\sum_{x\in\{0,1\}^n}\prod_{i=1}^r \hat{o}^2_{i,U_i,x_i}\,\mathrm{tr}\left(\bigotimes_{i=1}^n U_i^\dagger|x_i\rangle\langle x_i|U_i\rho\right) \tag{186}$$

$$= \frac{1}{|\mathrm{Cl}_1|^r}\sum_{U\in\mathrm{Cl}_1^r}\sum_{x\in\{0,1\}^r}\prod_{i=1}^r \hat{o}^2_{i,U_i,x_i}\,\mathrm{tr}\left(\bigotimes_{i=1}^r U_i^\dagger|x_i\rangle\langle x_i|U_i\otimes\mathbb{1}^{\otimes(n-r)}\rho\right) \tag{187}$$

$$= \frac{1}{|\mathrm{Cl}_1|^r}\sum_{U\in\mathrm{Cl}_1^r}\sum_{x\in\{0,1\}^r}\prod_{i=1}^r \hat{o}^2_{i,U_i,x_i}\,\mathrm{tr}\left(\bigotimes_{i=1}^r U_i^\dagger|x_i\rangle\langle x_i|U_i\,\rho'\right), \tag{188}$$

where $\rho' = \mathrm{tr}_{r+1,\dots,n}\rho$ is the reduced state on the first r qubits.

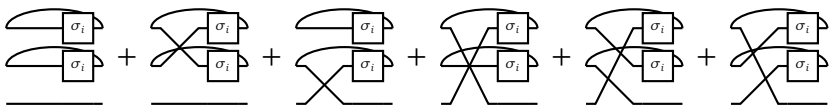
In the following, we concentrate on the case that $O = \sigma_1\otimes\cdots\otimes\sigma_r\otimes\mathbb{1}^{\otimes(n-r)}$ is a Pauli operator supported on r qubits (here $\sigma_i\in\{X,Y,Z\}$) and refer to Ref. [19] for general r-local observables. Using the explicit form of the shadow estimator (185), we can then compute in analogy to Sec. 3.2.1:

$$\mathbb{E}[\hat{o}^2] = \frac{1}{|\mathrm{Cl}_1|^r}\sum_{U\in\mathrm{Cl}_1^r}\sum_{x\in\{0,1\}^r}\prod_{i=1}^r 9\,\mathrm{tr}(\sigma_i U_i^\dagger|x_i\rangle\langle x_i|U_i)^2\,\mathrm{tr}\left(\bigotimes_{i=1}^r U_i^\dagger|x_i\rangle\langle x_i|U_i\,\rho'\right) \tag{189}$$

$$= \mathrm{tr}\left[\mathbb{1}_r^{\otimes 2}\otimes\rho'\bigotimes_{i=1}^r\left(\frac{9}{|\mathrm{Cl}_1|}\sum_{U\in\mathrm{Cl}_1}\sum_{x\in\{0,1\}}(U_i^\dagger|x_i\rangle\langle x_i|U_i)^{\otimes 3}\sigma_i^{\otimes 2}\otimes\mathbb{1}_1\right)\right] \tag{190}$$

$$= \mathrm{tr}\left[\mathbb{1}_r^{\otimes 2}\otimes\rho'\bigotimes_{i=1}^r\left(\frac{3}{4}\sum_{\pi\in S_3}R_\pi\sigma_i^{\otimes 2}\otimes\mathbb{1}_1\right)\right]. \tag{191}$$

In the last step, we used the previous computation (162) (for $d=2$). We can now perform the partial trace over the first two systems based on the previous graphical calculus (163):



$$= \mathrm{tr}(\sigma_i^2)\mathbb{1} + 2\sigma_i^2 = 4\mathbb{1}\,. \tag{192}$$

Here, we used that Pauli operators square to the identity: $X^2 = Y^2 = Z^2 = \mathbb{1}$. Hence, we obtain the following variance from Eq. (191):

$$\mathrm{Var}[\hat{o}] \le \mathbb{E}[\hat{o}^2] = \mathrm{tr}\left[\rho'\bigotimes_{i=1}^r(3\mathbb{1})\right] = 3^r\,. \tag{193}$$

Discussion of shadow estimation with local Clifford unitaries. It is worth highlighting that, when using local Clifford unitaries for shadow estimation, the postprocessing become particularly efficient. At the same time, the sampling complexity for $r = O(1)$ is constant, meaning that the whole shadow tomography performance is efficient. In this setting, each measurement outcome corresponds to projecting onto a local Pauli basis element, and the corresponding classical shadow can be reconstructed with computational cost scaling only linearly with the system size. Moreover, for observables that are themselves local (i.e., supported on a constant number of qubits), the postprocessing requires evaluating simple expectation values over these classical shadows. This makes the overall procedure highly scalable in practice, provided the observables are local.

3.2.3 Beyond linear observables: estimating purity

Consider the problem of estimating the purity $\mathcal{P}(\rho) \equiv \mathrm{tr}(\rho^2)$ of a state ρ, for instance the reduced density matrix of a global pure state $|\Psi\rangle$. It is inherently nonlinear in the state, so it does not follow in the shadow framework discussed above. In the following, we discuss how it extends to polynomial functions at the example of the purity.

For a given shot r, define the reconstructed snapshot

$$\widehat{\rho}_r := S^{-1}(E_{U_r,x_r}). \tag{194}$$

Equation (139) implies $\mathbb{E}[\widehat{\rho}_r] = \rho$. Notice that $\widehat{\rho}_r$ need not be a positive operator; it is an unbiased estimator, not a physical density matrix. For two different shots $r \neq s$, statistical independence then gives

$$\mathbb{E}[\mathrm{tr}(\widehat{\rho}_r\widehat{\rho}_s)] = \mathrm{tr}(\mathbb{E}[\widehat{\rho}_r]\mathbb{E}[\widehat{\rho}_s]) \tag{195}$$

$$= \mathrm{tr}(\rho^2) = \mathcal{P}(\rho). \tag{196}$$

Note that the diagonal terms $r = s$ must be excluded: in general, $\mathbb{E}[\mathrm{tr}(\widehat{\rho}_r^2)] \neq \mathrm{tr}(\rho^2)$. Averaging all unordered pairs therefore produces the unbiased U-statistic

$$\widehat{\mathcal{P}}_N := \binom{N}{2}^{-1} \sum_{1\leq r<s\leq N} \mathrm{tr}(\widehat{\rho}_r\widehat{\rho}_s). \tag{197}$$

Next, we want to analzye the variance of $\widehat{\mathcal{P}}_N$, which is explicitly given by

$$\mathrm{Var}[\widehat{\mathcal{P}}_N] = \binom{N}{2}^{-2} \sum_{r<s}\sum_{r'<s'} \left(\mathbb{E}[\mathrm{tr}(\widehat{\rho}_r\widehat{\rho}_s)\,\mathrm{tr}(\widehat{\rho}_{r'}\widehat{\rho}_{s'})] - \mathcal{P}(\rho)^2\right) \tag{198}$$

$$= \binom{N}{2}^{-2} \sum_{r<s}\sum_{r'<s'} \mathrm{Cov}(h_{rs}, h_{r's'}), \tag{199}$$

where we set $h_{rs} := \mathrm{tr}(\widehat{\rho}_r\widehat{\rho}_s)$ and $\theta := \mathrm{tr}(\rho^2)$. We can then isolate three contributions to the double sum for $\mathrm{Var}(\widehat{\mathcal{P}}_N)$:

1. The two pairs are identical. This contributes $\binom{N}{2}\mathrm{Var}(h_{12})$ before the overall normalization.

2. The pairs share exactly one shot, for example $(1,2)$ and $(1,3)$. We can then use the law of of total expectation by conditioning on $\widehat{\rho}_1$:

$$\mathbb{E}[h_{12}, h_{13}] = \mathbb{E}_{\widehat{\rho}_1}[\mathbb{E}[h_{12}h_{13}|\widehat{\rho}_1]] = \mathbb{E}_{\widehat{\rho}_1}[\mathbb{E}[\mathrm{tr}(\widehat{\rho}_1\widehat{\rho}_2)\,\mathrm{tr}(\widehat{\rho}_1\widehat{\rho}_3)|\widehat{\rho}_1]] = \mathbb{E}[\mathrm{tr}(\widehat{\rho}_1\rho)^2], \tag{200}$$

   where we used that $\widehat{\rho}_2$ and $\widehat{\rho}_3$ are independent. Thus we find

$$\mathrm{Cov}(h_{12}, h_{13}) = \mathrm{Var}[\mathrm{tr}(\widehat{\rho}_1\rho)]. \tag{201}$$

   For each unordered pair there are $2(N-2)$ ordered pairs sharing one index.

3. Four distinct shots are independent, so their covariance is zero.

Dividing by $\binom{N}{2}^2$ gives the exact decomposition

$$\boxed{\mathrm{Var}(\widehat{\mathcal{P}}_N) = \binom{N}{2}^{-1}\mathrm{Var}(h_{12}) + \frac{2(N-2)}{\binom{N}{2}}\mathrm{Var}[\mathrm{tr}(\widehat{\rho}_1\rho)]}. \tag{202}$$

The first term is the pair-noise contribution and decays as N^{-2}; the second is the shared-shot contribution and generically decays as N^{-1}. Their constants depend on the measurement ensemble and on ρ.

The same construction applies to $\mathrm{tr}(\rho^m)$: use m independent shadows, contract them with the cyclic permutation on m replicas, and average over tuples with distinct shot indices. The number of replicas is therefore set by the degree of the nonlinear observable.

Exact closed form for global Clifford measurements on qubits. The two variances in Eq. (202) can be evaluated exactly for the global Clifford protocol: We assume a n-qubit system, $d = 2^n$, and measurement shots are independent and identically distributed (i.i.d): For every shot, we sample a uniformly random global Clifford, followed by a computational-basis measurement. We then have

$$E_{U_r,x_r} = U_r^\dagger |x_r\rangle\langle x_r| U_r\,, \qquad \widehat{\rho}_r = (d+1)E_{U_r,x_r} - \mathbb{1}\,. \tag{203}$$

Our goal is to compute the two variances in Eq. (202). To this end, note that the explicit form of $\widehat{\rho}_r$ implies

$$\mathrm{tr}(\widehat{\rho}_1\rho) = (d+1)\,\mathrm{tr}(E_1\rho) - 1 =: (d+1)a_1 - 1\,, \tag{204}$$

$$h_{12} = \mathrm{tr}(\widehat{\rho}_1\widehat{\rho}_2) = (d+1)^2\,\mathrm{tr}(E_1E_2) - (d+2) =: (d+1)^2 s_{12} - (d+2)\,, \tag{205}$$

Thus, computing the variances in Eq. (202) reduces to computing the variances of a_1 and s_{12}.

We compute, using Exercise 2.10:

$$\mathbb{E}[a_1^t] = \mathbb{E}_{U,x}\,\mathrm{tr}(E_{U,x}\rho)^t \tag{206}$$

$$= \mathbb{E}_{U\sim \mathrm{Cl}_n} \sum_{x\in\{0,1\}^n} \mathrm{tr}(E_{U,x}\rho)^t p(x|U,\rho) \tag{207}$$

$$= \mathbb{E}_{U\sim \mathrm{Cl}_n} \sum_{x\in\{0,1\}^n} \mathrm{tr}(U^\dagger|x\rangle\langle x|U\rho)^{t+1} \tag{208}$$

$$= \sum_{x\in\{0,1\}^n} \mathbb{E}_{U\sim \mathrm{Cl}_n}\,\mathrm{tr}((U^\dagger|x\rangle\langle x|U)^{\otimes(t+1)}\rho^{\otimes(t+1)}) \tag{209}$$

$$= \frac{t!d!}{(d+t)!} \sum_{\pi\in \mathrm{S}_{t+1}} \mathrm{tr}(R_\pi \rho^{\otimes(t+1)})\,. \tag{210}$$

With $\mathcal{P} := \mathrm{tr}(\rho^2)$ and $\mathcal{T} := \mathrm{tr}(\rho^3)$, this yields

$$\mathbb{E}_{\mathrm{shot}}[a_r] = \frac{1+\mathcal{P}}{d+1}, \qquad \mathbb{E}_{\mathrm{shot}}[a_r^2] = \frac{1+3\mathcal{P}+2\mathcal{T}}{(d+1)(d+2)}. \tag{211}$$

To compute the variance of s_{12}, condition on E_1 and apply the third moment to the second shot. In the contraction with $\rho\otimes E_1\otimes E_1$, the identity and the transposition (23) contribute 1 each; the other two transpositions and the two 3-cycles contribute a_1 each. Thus the six permutations in S_3 contribute $2+4a_1$, and therefore

$$\mathbb{E}_{\mathrm{shot}}[s_{12}^2 \mid E_1] = \frac{2+4a_1}{(d+1)(d+2)},$$

$$\mathbb{E}_{\mathrm{shot}}[s_{12}^2] = \frac{2(d+3+2\mathcal{P})}{(d+1)^2(d+2)}. \tag{212}$$

The two random variables appearing in the general variance decomposition are now

$$\mathrm{tr}(\widehat{\rho}_1\rho) = (d+1)a_1 - 1, \qquad h_{12} = (d+1)^2 s_{12} - (d+2). \tag{213}$$

Using Eqs. (211) and (212) and $\mathbb{E}[h_{12}] = \mathcal{P}$ yields

$$V_1 := \mathrm{Var}[\mathrm{tr}(\widehat{\rho}_1 \rho)] = \frac{d+1}{d+2}(1+3\mathcal{P}+2\mathcal{T}) - (1+\mathcal{P})^2, \tag{214}$$

$$V_2 := \mathrm{Var}(h_{12}) = \frac{2(d+1)^2}{d+2}(d+3+2\mathcal{P}) - (d+2+\mathcal{P})^2. \tag{215}$$

Consequently, for the N-snapshot estimator with $N \geq 2$ in Eq. (197),

$$\boxed{\mathrm{Var}(\widehat{\mathcal{P}}_N) = \binom{N}{2}^{-1} [V_2 + 2(N-2)V_1].} \tag{216}$$

Here N is consistently the number of measurement shots; n is the number of qubits, and $d = 2^n$ is the Hilbert-space dimension. As checks, $V_1 = 0$ for $\rho = \mathbb{1}/d$, while for a pure state $V_1 = 2(d-1)/(d+2)$. For fixed state moments and large d, $V_2 = d^2 + O(d)$ whereas $V_1 = \mathcal{P} + 2\mathcal{T} - \mathcal{P}^2 + O(d^{-1})$. Hence the pair-noise term scales as d^2/N^2, but the shared-shot term generically scales as $1/N$ and, whenever $V_1 > 0$, eventually dominates when the number of shots is sufficiently large.

3.3 Further reading

The classical-shadow protocol and its median-of-means guarantee were introduced in [19]. The perspective developed here emphasizes the measurement channel and its inverse because this makes the unbiasedness and the observable-dependent variance transparent. For precision guarantees beyond the idealized setting, see Ref. [20]. Shallow-shadow protocols interpolate between local Pauli measurements and global Clifford measurements and expose the tradeoff between circuit depth, sample complexity, and classical post-processing; see [21]. Stabilizer-specific fidelity estimation and its relation to nonstabilizerness are discussed in Ref. [18].

Recall that the n-qudit Clifford group for prime qudit dimension $q > 2$ does not form a unitary 3-design and thus the variance calculations in Sec. 3.2.1 do not apply to this case. However, using more advanced techniques, the variance of shadow estimation was very recently computed in this case as well [22, 23]. Perhaps surprisingly, only a small correction depending on the qudit dimension q was found:

$$\mathrm{Var}[\hat{o}] \leq \frac{d+1}{d+2}\left((2q-3)\|O_0\|_2^2 + 2\|O_0\|_\infty^2\right) \qquad \text{(for Cliffords in prime local dimension } q\text{)}. \tag{217}$$

This implies that the discussion in the qubit case also applies to qudits if the local dimension is not too large.

3.4 Exercises on classical shadows

Exercise 3.3 (The global shadow channel and unbiasedness). *Let $\mathcal{H} = (\mathbb{C}^q)^{\otimes n}$ and $d = q^n$. Draw U uniformly from an exact unitary 2-design, measure $U\rho U^\dagger$ in the computational basis, and denote the outcome by x. Set*

$$E_{U,x} := U^\dagger |x\rangle\langle x| U, \qquad S(X) := \mathbb{E}_U \sum_{x=1}^{d} \mathrm{tr}(E_{U,x} X) E_{U,x}.$$

(a) Using the swap operator $R_{(12)}$, show that

$$\mathbb{E}_U \sum_{x=1}^{d} E_{U,x}^{\otimes 2} = \frac{\mathbb{1}\otimes\mathbb{1} + \mathsf{F}}{d+1}.$$

(b) *Deduce that, for every* $X \in \mathcal{L}(\mathcal{H})$,

$$S(X) = \frac{X + \mathrm{tr}(X)\mathbb{1}}{d+1}, \qquad S^{-1}(X) = (d+1)X - \mathrm{tr}(X)\mathbb{1}.$$

(c) *Define the one-shot shadow*

$$\widehat{\rho}_{U,x} := S^{-1}(E_{U,x}) = (d+1)E_{U,x} - \mathbb{1}.$$

Prove that $\mathbb{E}_{U,x}[\widehat{\rho}_{U,x}] = \rho$, *where* x *is sampled according to Born's rule. Conclude that* $\widehat{o}_{U,x} := \mathrm{tr}(O\widehat{\rho}_{U,x})$ *is an unbiased estimator of* $\mathrm{tr}(O\rho)$ *for every observable* O.

State precisely where the 2-design assumption is used.

Exercise 3.4 (Local Pauli shadows for qubits). *In this exercise assume* $q = 2$. *Independently on every qubit* i, *choose a basis* $P_i \in \{X, Y, Z\}$ *uniformly, measure* P_i, *and record the eigenvalue* $s_i \in \{+1, -1\}$. *Thus*

$$E_{P_i,s_i} = \frac{\mathbb{1} + s_i P_i}{2}.$$

(a) *Show that the inverse of the single-qubit measurement channel gives*

$$\widehat{\rho}_i = 3E_{P_i,s_i} - \mathbb{1} = \frac{\mathbb{1} + 3s_i P_i}{2},$$

and that the n*-qubit snapshot is* $\widehat{\rho} = \widehat{\rho}_1 \otimes \cdots \otimes \widehat{\rho}_n$.

(b) *Let* $Q = Q_1 \otimes \cdots \otimes Q_n$ *be a Pauli string and let* $r := |\mathrm{supp}(Q)|$. *Prove the one-shot formula*

$$\widehat{o}_Q := \mathrm{tr}(Q\widehat{\rho}) = 3^r \prod_{i \in \mathrm{supp}(Q)} s_i \, \mathbb{1}\{P_i = Q_i\}.$$

(c) *Show, for every input state* ρ, *that*

$$\mathbb{E}[\widehat{o}_Q^2] = 3^r, \qquad \mathrm{Var}(\widehat{o}_Q) \le 3^r.$$

(d) *If* N *independent snapshots are averaged, use Chebyshev's inequality to give a sufficient value of* N *for additive error* ε *and failure probability at most* δ.

Exercise 3.5 (Many observables and the median-of-means estimator). *Let* $O_1, \ldots, O_M$ *be* M *observables. A single shadow produces random variables* $Z_1, \ldots, Z_M$ *satisfying*

$$\mathbb{E}[Z_a] = \mu_a := \mathrm{tr}(O_a\rho), \qquad \mathrm{Var}(Z_a) \le V \quad (a = 1, \ldots, M).$$

Assume $V > 0$ *and* $\varepsilon, \delta \in (0, 1)$. *No independence between* Z_a *and* $Z_{a'}$ *obtained from the same snapshot is assumed. Split* $N = Bb$ *independent snapshots into* B *independent blocks of size* b, *compute the mean in each block, and take the median of the* B *block means separately for every observable. Take* B *odd.*

(a) *Prove that* $b \ge 4V/\varepsilon^2$ *makes any fixed block mean bad (farther than* ε *from* μ_a*) with probability at most* $1/4$.

(b) *Prove that the median for a fixed observable is bad with probability at most* $\exp(-B/8)$.

(c) *Use a union bound to show that all* M *estimates are simultaneously* ε*-accurate with probability at least* $1 - \delta$ *if*

$$b = \left\lceil \frac{4V}{\varepsilon^2} \right\rceil, \qquad B \text{ is the least odd integer not smaller than } 8\log\left(\frac{M}{\delta}\right).$$

Explain why correlations among different observables do not invalidate the argument.

Exercise 3.6 (Fidelity estimation: why a third moment appears). *Let $P = |\phi\rangle\langle\phi|$ be a pure target state, let $f := \mathrm{tr}(P\rho)$, and use the global snapshot from Exercise 3.3. Assume now that the unitary ensemble is an exact unitary 3-design. For one snapshot define*

$$Z := \mathrm{tr}(P\widehat{\rho}) = (d+1)\,\mathrm{tr}(PE) - 1, \qquad E = E_{U,x}.$$

(a) First take U Haar random. Show that, for any function g of a rank-one projector,

$$\mathbb{E}_{U,x}[g(E)] = d \int \mathrm{tr}(E_\psi \rho)\, g(E_\psi)\, \mathrm{d}\psi, \qquad E_\psi = |\psi\rangle\langle\psi|,$$

where $\mathrm{d}\psi$ is Haar measure on pure states. Explain why an exact 3-design gives the same answer for the degree-three polynomial moments used in part (b), even though it need not reproduce this identity for an arbitrary non-polynomial g.

(b) Use

$$\int E_\psi^{\otimes t}\, \mathrm{d}\psi = \frac{1}{d(d+1)\cdots(d+t-1)} \sum_{\pi \in S_t} R_\pi, \qquad t = 1,2,3,$$

to prove

$$\mathbb{E}[Z^2] = \frac{d(1+2f)}{d+2}, \qquad \mathrm{Var}(Z) = \frac{d(1+2f)}{d+2} - f^2 < 3.$$

(c) Explain why unbiasedness only needs a 2-design whereas this variance computation needs a 3-design. Give a Chebyshev bound for the average of N independent copies of Z.

Exercise 3.7 (An unbiased purity estimator). *Let $\widehat{\rho}_1, \ldots, \widehat{\rho}_N$ be independent global shadows, with $N \geq 2$, obtained from an exact unitary 2-design. Define the order-two U-statistic*

$$\widehat{\mathcal{P}}_{\mathrm{U}} := \binom{N}{2}^{-1} \sum_{1 \leq i < j \leq N} \mathrm{tr}(\widehat{\rho}_i \widehat{\rho}_j).$$

(a) Prove that $\mathbb{E}[\widehat{\mathcal{P}}_{\mathrm{U}}] = \mathrm{tr}(\rho^2)$.

(b) Show that every global snapshot obeys

$$\mathrm{tr}(\widehat{\rho}_i^2) = d^2 + d - 1.$$

(c) Consider the tempting plug-in estimator

$$\widetilde{\mathcal{P}} := \mathrm{tr}\left[\left(\frac{1}{N}\sum_{i=1}^{N} \widehat{\rho}_i\right)^2\right] = \frac{1}{N^2}\sum_{i,j=1}^{N} \mathrm{tr}(\widehat{\rho}_i \widehat{\rho}_j).$$

Compute its expectation and its bias. Explain why using two independent snapshots in each term removes the diagonal bias.

These exercises move from single-qubit reconstruction to choosing a measurement ensemble and finally to a simple noise-mitigation calculation.

Exercise 3.8 (Basic: explicit tomography of one qubit). *Write an unknown qubit state as*

$$\rho = \frac{1}{2}(\mathbb{1} + r_x X + r_y Y + r_z Z).$$

For each $a \in \{x, y, z\}$, let $p_{a,+}$ be the probability of obtaining the $+1$ outcome when measuring the corresponding Pauli operator σ_a.

(a) *Derive* $r_a = 2p_{a,+} - 1$ *and hence reconstruct* ρ *from the three measurement probabilities.*

(b) *Diagonalize* ρ *and prove that the reconstructed matrix is a physical density matrix if and only if* $r_x^2 + r_y^2 + r_z^2 \leq 1$.

(c) *Empirical frequencies from finitely many shots can give a vector* $\widehat{\boldsymbol{r}}$ *outside the Bloch ball. Give a simple projected estimator that is always physical, and explain why the need for this projection does not contradict unbiasedness of the unprojected linear estimator.*

Exercise 3.9 (Intermediate: which shadow ensemble should one use?). *Suppose one wants to estimate* M *expectation values to additive error* ε *and simultaneous failure probability at most* δ. *Compare the following two families of observables on* n *qubits, where* $d = 2^n$:

(i) *pure-state projectors* $O_j = |\phi_j\rangle\langle\phi_j|$;

(ii) *Pauli strings* Q_j *of weight at most* w.

Use a median-of-means bound of the form

$$N = O\left(\frac{V}{\varepsilon^2}\log\frac{2M}{\delta}\right),$$

where V *bounds the one-shot variance. For global Clifford shadows use* $V \leq 3\|O_0\|_2^2$, *with* $O_0 = O - \mathrm{tr}(O)\mathbb{1}/d$. *For local Pauli shadows use* $V \leq 3^w$ *for a weight-*w *Pauli string.*

(a) *Derive the resulting sample-complexity scaling for each observable family and each measurement ensemble whenever the stated bounds apply.*

(b) *Show that global Clifford shadows give a dimension-independent guarantee for pure-state fidelities, whereas the same global bound is* $O(d)$ *for a nonidentity Pauli string.*

(c) *State the regime in which local Pauli measurements give the better guarantee for Pauli strings. Why is there no analogous, state-independent local-Pauli guarantee for arbitrary pure-state projectors without using additional structure of the target states?*

Exercise 3.10 (Advanced: correcting symmetric readout noise). *In a local Pauli-shadow experiment, suppose every ideal eigenvalue* $s_i \in \{+1, -1\}$ *is independently flipped by the detector with probability* $\eta < 1/2$. *The recorded value is* $\widetilde{s}_i = \xi_i s_i$, *where* $\Pr(\xi_i = -1) = \eta$.

(a) *Show that* $\mathbb{E}[\widetilde{s}_i \mid s_i] = (1-2\eta)s_i$ *and derive the corrected one-qubit snapshot*

$$\widehat{\rho}_i^{(\eta)} = \frac{1}{2}\left(\mathbb{1} + \frac{3\widetilde{s}_i}{1-2\eta}P_i\right),$$

where $P_i \in \{X, Y, Z\}$ *is the randomly selected measurement axis. Verify its unbiasedness by averaging first over readout noise, then over outcomes and axes.*

(b) *Let* Q *be a Pauli string of weight* w. *Write its corrected one-shot estimator explicitly and prove*

$$\mathbb{E}\left[(\widehat{o}_Q^{(\eta)})^2\right] = \left(\frac{3}{(1-2\eta)^2}\right)^w.$$

Give the corresponding variance and sample-complexity upper bounds.

(c) *If the correction factor is omitted, compute the expectation of the noisy estimator. Interpret the exponential dependence on* w *and the divergence as* $\eta \to 1/2$.

4 Quantum Randomness from Local Gates

As argued previously in Sec. 2.4, Haar-random unitaries are notoriously expensive to implement in that they need exponentially deep quantum circuits. For this reason, unitary designs have been introduced as a way to mimic the Haar measure up to a certain moment. *The* prototypical example of a unitary design is the Clifford group which is thus extensively used throughout the literature, and also in our treatment of classical shadows in Sec. 3.

Nevertheless, there are two major problems with unitary designs: First, there are no practical examples of unitary designs beyond the Clifford group and we are thus limited to third moments. Second, Clifford unitaries need linear-depth circuits to be implemented and we will see in this section that linear depth is unavoidable for every exact unitary design in one dimension. However, near-term devices are noisy and thus effectively limited in the depth of the circuits they can execute. We would thus prefer ensembles with a more fine-grained control on the circuit depth and, ideally, sub-linearly sized circuits.

From a practical point of view, it thus seems to be a good idea to take a *constructive* approach to random unitaries by considering families of quantum circuits with variable depth, where the (local) gates are drawn at random, see Fig. 4 for an example. We call the so-constructed ensemble a *random quantum circuit* (RQC). Importantly, these RQCs will generally not form a unitary design in the sense of Sec. 2.4, i.e., moments of the RQC ensemble will not exactly reproduce the Haar moments. However, we expect that they become good approximations if the RQC is sufficiently deep – and this is enough for essentially all the applications we could come up with.

This section is organized around three questions. How do we compute the moments of an RQC (Sec. 4.1)? Do they converge to the Haar moments, and in which sense should a circuit be close to the Haar measure (Sec. 4.2)? How deep does a circuit have to be (Secs. 4.3 and 4.4)? As we will see, the answer to the last question depends sharply on the notion of approximation.

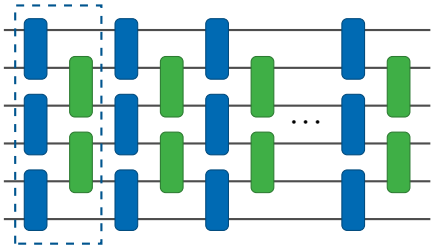


Figure 4: Random quantum circuits are circuits in which the gates are drawn independently at random. Typically, these gates are local and arranged in a specific layout, for instance in the popular *brickwork layout* shown here: The gates of a brickwork circuit are organized in layers of disjoint gates, alternating between odd (blue) and even (green) qubit pairs.

4.1 Random quantum circuits and their moment operators

Let us make it precise what we mean by a *random quantum circuit* (RQC) in the following. The construction of a RQC will be based on an *architecture*, a fixed layout for the circuit constrained by qudit connectivity, but otherwise free of our choosing. This fixed layout will then be populated by independent random gates. We will then see that the moment operator of the RQC precisely reflects the underlying architecture.

Architectures. As before, we work with the n-qudit Hilbert space $\mathcal{H} = (\mathbb{C}^q)^{\otimes n}$ with local dimension q and global dimension $d = q^n$. For a subset $e \subseteq [n] = \{1, \ldots, n\}$ of qudits and

$V \in \mathrm{U}(q^{|e|})$, we write $V_e \simeq V \otimes \mathbb{1}$ for the gate V acting on the qudits in e and trivially on all others.

Definition 4.1: Random Quantum Circuit

An *architecture* is a (finite or infinite) sequence $e_1, \ldots, e_m, \ldots$ with $e_i \subseteq [n]$ that defines the supports of gates in every time step i. A *random quantum circuit* (RQC) with this architecture is defined by choosing *gate distributions* ν_i, i.e., probability measures on $\mathrm{U}(q^{|e_i|})$, from which the gates at step i are sampled. Concretely, the RQC ensemble after m steps is given by the unitaries

$$U_m = (V_m)_{e_m} \cdots (V_2)_{e_2} (V_1)_{e_1}, \qquad V_i \sim \nu_i \text{ independent}. \tag{218}$$

More generally, a random quantum circuit is a probabilistic mixture of random quantum circuits with fixed architecture and gate distributions.

Importantly, it seems that *the literature cannot quite agree on the definition of a random quantum circuit*. Frequently, a brickwork circuit is implicitly meant. However, in these lecture notes, we adopt a deliberately literal definition of this concept. Definition 4.1 does explicitly not constrain the supports of these gates, thus a sequence of global (!) random unitaries would still count as a random quantum circuit. Very often, we want these unitaries to be local, i.e., the support is of constant size, and we will discuss this in more detail below. However, the key property of a random quantum cirucit is instead *independence* of the individual unitaries composing the circuit.

Example 4.1: Popular random circuits

Often, the local distributions ν_i are chosen to be the Haar measure and all the gates are of bounded support, $|e_i| \leq r$ (e.g., 2-qudit gates). Some examples that appear in these notes are:

(a) *Brickwork circuits* (Fig. 4): Here, the qudits are arranged on a line and we assume for simplicity that n is even. The architecture sequence alternates between the finite sequences $(1,2),(3,4),\ldots,(n-1,n)$ and $(2,3),(4,5),\ldots,(n-2,n-1),\ldots$. The 2-local gates are typically sampled Haar-randomly, but uniform sampling from the 2-qudit Clifford group, or other distributions, are also of interest.

(b) *Staircase circuits*: Again on a line, the support of each gate is shifted by a fixed amount compared to its predecessor, leading to "staircase circuits" which generate random matrix product states, see **??** This is represented by a (finite) architecture sequence of the form $(1,2,\ldots,r),(1+s,2+s,\ldots,r+s),\ldots$.

(c) *Random-position circuits*: Consider the set of qudit pairs $e = (i,j)$ which are connected on a given quantum device (i.e., we can apply gates to them). In every time step, a connected qudit pair e is selected with probability p_e and a Haar-random gate is applied to it. Here, also the architecture is random, but it is chosen independently of the gates. The resulting RQC ensemble is formally a probabilistic mixture (convex combination) of deterministic architectures given by sequences of qudit pairs $e_1, e_2, \ldots$ with weight $p_{e_1} p_{e_2} \cdots$.

Gates with disjoint supports commute and we will assume that they can be applied simultaneously. We can therefore group the gates into *layers* of gates with disjoint supports, and call

the number of layers the *depth* of the circuit. For instance, the gates in the brickwork circuits, can be grouped in "odd" and "even" layers, see also Fig. 4.

Moment operators. We will next derive a formula for the moment operator of a RQC. Here, the *k-th moment operator* of a probability measure ν on $\mathrm{U}(d)$ is defined analogously to the Haar moment operator (10):

$$\mathsf{M}_{k,\nu}(X) := \mathbb{E}_{U\sim\nu}\left[U^{\otimes k} X U^{\otimes k,\dagger}\right] = \int_{\mathrm{U}(d)} U^{\otimes k} X U^{\otimes k,\dagger} \mathrm{d}\nu(U). \tag{219}$$

Again, for the Haar measure, we simply write $\mathsf{M}_{k,\nu} = \mathsf{M}_k$. In this section, we also often use the shorthand notation $\Phi_U(X) := U^{\otimes k} X U^{\otimes k,\dagger}$ for the k-copy quantum channel given by a unitary U. We will now prove that the independence of the random gates in a RQC implies that the moment operator factorizes:

Lemma 4.1 (Independence implies factorization). *Consider an ensemble of unitaries $U = V_m \cdots V_1$ where $V_i \sim \nu_i$ are independent random unitaries, and let η be the associated probability measure. Then, we have:*

$$\mathsf{M}_{k,\eta} = \mathsf{M}_{k,\nu_m} \cdots \mathsf{M}_{k,\nu_2} \mathsf{M}_{k,\nu_1}. \tag{220}$$

Proof. Since $(V_2 V_1)^{\otimes k} = V_2^{\otimes k} V_1^{\otimes k}$, we have $\Phi_{V_2 V_1} = \Phi_{V_2}\Phi_{V_1}$ and consequently also $\Phi_U = \Phi_{V_m} \cdots \Phi_{V_1}$. Because the V_i are independent, their joint distribution is the product measure. The integral then splits by Fubini's theorem as follows:

$$\mathsf{M}_{k,\eta} = \int \Phi_{V_m} \cdots \Phi_{V_1} \,\mathrm{d}\nu_1(V_1) \cdots \mathrm{d}\nu_m(V_m) = \int \Phi_{V_m} \mathrm{d}\nu_m(V_m) \cdots \int \Phi_{V_1} \mathrm{d}\nu_1(V_1). \tag{221}$$

Re-expressing these integrals as moment operators then shows the claim. □

Note that only independence of the random unitaries V_i was used, not that they arise from some architectures as the RQCs in Def. 4.1. Moreover, their individual distributions do not need to be identical. If they are, $\nu_i \equiv \nu$, Lem. 4.1 directly shows that the moment operator of the entire sequence is the m-fold power of a single factor: $\mathsf{M}_{k,\eta} = \mathsf{M}_{k,\nu}^m$. In the context of RQCs, we call this a *layered random quantum circuit*.

Definition 4.2: Layered RQC

We call a RQC as in Def. 4.1 *layered*, if gates can be arranged in layers in such a way that the layers are identically distributed.

Example 4.2: Layered random circuits

Consider again the cases in Example 4.1:

(a) Brickwork circuit are layered: The gates can naturally be arranged in *even* and *odd* layers. These are not identical, the moment operator becomes instead $\mathsf{M}_{k,\mathrm{odd}}\mathsf{M}_{k,\mathrm{even}}\mathsf{M}_{k,\mathrm{odd}}\cdots$. Assume that $\mathsf{M}_{k,\mathrm{odd}}^2 = \mathsf{M}_{k,\mathrm{odd}}$, which is the case if the local distribution is Haar-random (see Exercise 4.5). Then, for an odd number of layers, we can group them in the layout odd-even-odd, because $(\mathsf{M}_{k,\mathrm{odd}}\mathsf{M}_{k,\mathrm{even}}\mathsf{M}_{k,\mathrm{odd}})^L$ expands into exactly $2L+1$ even/odd layers.

(b) The staircase circuit is *not* layered, because the gates in every step act on a different subset of qubits and the overlaps prevent them from being compressed any further.

(c) Random-position circuits are layered: In a given time step, the moment operator is $\mathsf{M}_{k,\mathrm{RP}} = \sum_e p_e \mathsf{M}_{k,e}$ where $\mathsf{M}_{k,e}$ is the moment operator of 2-qudit Haar-random unitaries applied on the qudit pair e. After L time steps, the moment operator is $\mathsf{M}_{k,\mathrm{RP}}^L$.

Recall that the defining property of random quantum circuits is *independence*. The radically opposite of this would be if the same random unitary is applied repeatedly, as it is the case in a *Floquet circuit* U^L, where U is sampled randomly. Then, the moment operator is $\mathbb{E}_U[\Phi_U^L]$, which in general differs a lot from $(\mathbb{E}_U\Phi_U)^L$:

Exercise 4.1 (Floquet circuits). *Let U be Haar random on $\mathrm{U}(d)$, $d \geq 2$, and let $\Phi_U = U(\,\cdot\,)U^\dagger$ be the corresponding first-moment channel.*

(a) Show that the trace of the superoperator Φ_U is $\mathrm{Tr}(\Phi_U) = |\operatorname{tr} U|^2$.

(b) Use second-order Weingarten calculus to show $\mathbb{E}|\operatorname{tr} U^2|^2 = 2$. Hint: *Use the swap trick $\operatorname{tr} U^2 = \operatorname{tr}(U^{\otimes 2}\mathsf{F})$ and expand the traces.*

(c) Show that $\mathbb{E}|\operatorname{tr}(U_2U_1)|^2 = 1$ for independent Haar-random unitaries U_1 and U_2. Conclude that $\mathbb{E}[\Phi_U^2]$ differs from M_1^2.

Locality. Independence allows us to decompose the moment operator of a RQC into a product of independent moment operators. If the gates V_i of a RQC (218) also act locally, we can furthermore compute the moment operators locally, i.e., only on the support of these gates. To see this, let us recall from Sec. 2.2.2 that the k-copy Hilbert space $((\mathbb{C}^q)^{\otimes n})^{\otimes k}$ can be arranged as a $k \times n$ grid of qudits, where row i is the i-th copy and column j contains the k replicas of qudit j, see Fig. 5. The k copies of a local gate $V \in \mathrm{U}(q^{|e|})$ then only act on the 'local' k-copy Hilbert space $((\mathbb{C}^q)^{\otimes|e|})^{\otimes k}$. For calculations, it will be convenient to reorder the tensor factors column-wise as $((\mathbb{C}^q)^{\otimes n})^{\otimes k} \simeq \bigotimes_{j=1}^n (\mathbb{C}^q)^{\otimes k}$ such that local gates act on the corresponding qudit factors. Note that if the qudits in the support e are not sequential, formally writing down the action of the gate can be a bit cumbersome. For simplicity, let us group all the qudits in e in the front, then we formally have $(V_e)^{\otimes k} \simeq V^{\otimes k} \otimes \mathbb{1}$. Then, we also have $\Phi_{V_e} \simeq \Phi_V \otimes \mathrm{id}$ and consequently the associated moment operator is a local operator: $(\mathsf{M}_{k,\nu})_e \simeq \mathsf{M}_{k,\nu} \otimes \mathrm{id}$.

As a result, **moments operators exactly reflect the architecture and locality of the circuit**. Concretely, if we want to compute the moment operator of a RQC, Def. 4.1, we can take independent and local averages over the used gate distributions, see Fig. 6. This allows us the write the RQC moment operator as a circuit of local moment operators which are arranged exactly as in the original circuit, but act on the k-copy operator spaces.

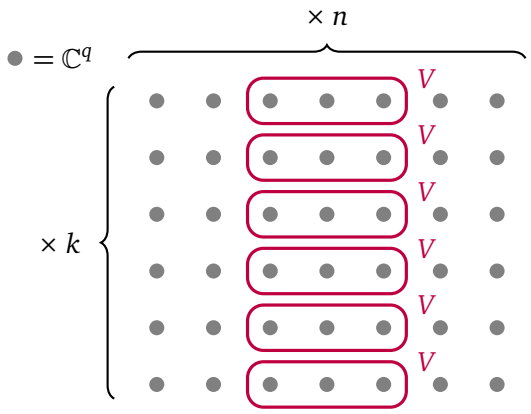


Figure 5: The Hilbert space $((\mathbb{C}^q)^{\otimes n})^{\otimes k}$ depicted as a $k \times n$ grid where every point corresponds to a copy of $\mathbb{C}^q$. Local unitaries $V \in \mathrm{U}(q^{|e|})$ act only on the local k-copy Hilbert space $((\mathbb{C}^q)^{\otimes |e|})^{\otimes k}$.

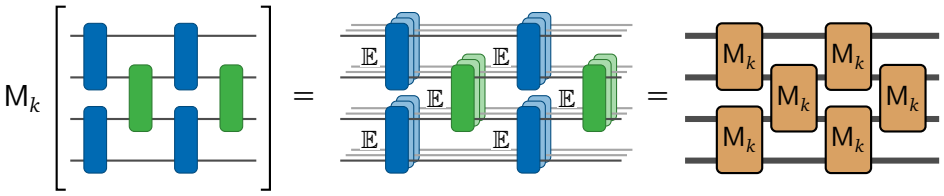


Figure 6: The moment operator of a random quantum circuit (left) is obtained by independent and local averages over the k-copy gates (middle), cf. Lem. 4.1. The result is a circuit of local moment operators (right) with exactly the same architecture as the original circuit, but acting on the k-copy operator spaces (thick wires). Here, all gates are Haar-random such that every local moment operator is the Haar moment operator M_k on two qudits.

Example 4.3: Brickwork circuits are local

The even and odd layers of the brickwork circuit can be written as $\mathsf{M}_{k,\mathrm{odd}} = \mathsf{M}_k \otimes \mathsf{M}_k \otimes \cdots$ and $\mathsf{M}_{k,\mathrm{even}} = \mathrm{id} \otimes \mathsf{M}_k \otimes \mathsf{M}_k \otimes \cdots$ where M_k is the Haar moment operator on 2 qudits.

4.2 Convergence of random quantum circuits and approximate designs

4.2.1 Limit behaviour of random circuits

Intuitively, a sufficiently well designed RQC should convergence to a Haar-random unitary with an increasing depth. If the RQC is layered, this is

$$\left(\mathsf{M}_{k,\nu}\right)^L \to \mathsf{M}_k \quad \text{for } L \to \infty\,. \tag{222}$$

To see this, we can consider the fixed points of $\mathsf{M}_{k,\nu}$. Using a change of variables and the invariance of the Haar measure, cf. Eq. (4), we find:

$$\mathsf{M}_{k,\nu}\mathsf{M}_k = \int \Phi_U \mathrm{d}\,\nu(U) \int \Phi_V \mathrm{d}V = \int\int \Phi_{UV}\,\mathrm{d}\,\nu(U)\mathrm{d}V \tag{223}$$

$$= \int\int \Phi_W \mathrm{d}\,\nu(U)\mathrm{d}(U^\dagger W) = \int \mathrm{d}\,\nu(U) \int \Phi_W \mathrm{d}W = \mathsf{M}_k\,. \tag{224}$$

And likewise $\mathsf{M}_k\mathsf{M}_{k,\nu} = \mathsf{M}_k$. Since $\mathsf{M}_k(A) = A$ for any $A \in \mathrm{Comm}_k$ (see Lem. 2.1) and M_k is self-adjoint, M_k is the identity on Comm_k. With respect to the orthogonal decomposition $\mathrm{Comm}_k \oplus \mathrm{Comm}_k^\perp$, we can then write

$$\mathsf{M}_k \simeq \begin{pmatrix} \mathrm{id} & 0 \\ 0 & 0 \end{pmatrix}. \tag{225}$$

But then, $\mathsf{M}_k\mathsf{M}_{k,\nu} = \mathsf{M}_k = \mathsf{M}_{k,\nu}\mathsf{M}_k$ shows that

$$\mathsf{M}_{k,\nu} \simeq \begin{pmatrix} \mathrm{id} & 0 \\ 0 & * \end{pmatrix}, \tag{226}$$

where $*$ is some operator on $\mathrm{Comm}_k^\perp$. By the block form (226), convergence of an RQC to the Haar measure is a statement about the block $*$, i.e. about the action of $\mathsf{M}_{k,\nu}$ on $\mathrm{Comm}_k^\perp$. For instance, it is sufficient that the operator norm $\|*\|_\infty < 1$ since then $\|*^L\|_\infty \to 0$, thus $*^L \to 0$ and $\mathsf{M}_{k,\nu}^L \to \mathsf{M}_k$. The value $\Delta := 1 - \|*\|_\infty$ is called the *spectral gap* of the RQC.

Under reasonable assumptions on the RQC architecture, for instance that there are no isolated qudit pairs, one can show that any (layered) RQC is gapped and thus eventually converges to the Haar measure. Computing the spectral gap then allows to give a bound at which rate this happens. Indeed, a major part of the RQC literature is devoted to spectral gaps, see Sec. 4.5. However, the spectral gap lacks an operational interpretation and, in fact, fails to predict to correct convergence rates in many settings.

4.2.2 Notions of approximate unitary designs

We say that a RQC (or any other probability measure on $\mathrm{U}(d)$ for that matter) is an *ε-approximate unitary k-design* if $\mathsf{M}_{k,\nu} \approx_\varepsilon \mathsf{M}_k$ in a chosen metric. It is worthwhile to discuss different notions of approximation that capture different operational meaning and are of varying strength. As it turns out, the required circuit depth for different approximations can diverge drastically: While a basic mathematical fact states that all norms are equivalent on finite-dimensional vector spaces, the conversion typically involves the dimensional factor 2^{nk} and can thus change the convergence rate signficantly, leading to different n (and k)-dependencies in circuit depth bounds.

Distinguishing a random circuit from the Haar measure. A natural operational definition would be the following: How deep do we have to make a RQC for it to be almost indistinguishable for any observer that has black-box access to at most k sampled unitaries? This boils down to the problem of *quantum channel discrimination*, a foundational problem in quantum information theory.

If we want to design an experiment that can discriminate between two quantum states ρ_1 and ρ_2, Holevo's theorem shows that this can only succeed with probability at most

$$p_{\mathrm{succ}} = \frac{1}{2} + \frac{1}{4}\|\rho_1 - \rho_2\|_1 . \tag{227}$$

This is exactly the reason why the *trace norm* $\|A\|_1 := \mathrm{tr}(|A|) = \mathrm{tr}(\sqrt{A^\dagger A})$ plays a fundamental role in quantum information: it operationally quantifies the indistinguishability of two quantum states. If the goal is to discriminate between two quantum channels ϕ_1 and ϕ_2, we can optimize over all possible input states and measurements, such that the relevant norm is

$$\sup_{\|\rho\|_1=1} \|\phi_1(\rho) - \phi_2(\rho)\|_1 =: \|\phi_1 - \phi_2\|_{1\to 1} . \tag{228}$$

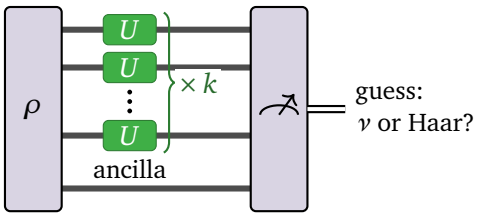


Figure 7: Operational meaning of additive-error designs. An experiment prepares a state ρ, possibly entangled with an ancilla, queries k copies of the unknown unitary U in parallel, and performs a joint measurement in order to decide whether U was drawn from ν or from the Haar measure. If ν is an additive-error design (Def. 4.3), no such experiment succeeds with probability larger than $\frac{1}{2} + \frac{1}{4}\varepsilon$.

Since we can always use ancillary systems, and only apply the channels on a subsystem, the operationally relevant norm is the *diamand norm*:

$$\|\phi_1 - \phi_2\|_\diamond := \sup_D \|\phi_1 \otimes \mathrm{id}_D - \phi_2 \otimes \mathrm{id}_D\|_{1\to 1} \tag{229}$$

In fact, one can show that taking an ancillary system that is as large as the original system (i.e., doubling the original system) is sufficient.

We can now define an approximate design by requiring that the RQC moment operator is close to the Haar one in diamond norm. The operational interpretation is that no experiment that queries a sampled unitary at most k times (in parallel) can discriminate between the two with probability larger than $\frac{1}{2} + \frac{1}{4}\varepsilon$, see Fig. 7.

Definition 4.3: Additive-error designs

We call a probability measure ν on $\mathrm{U}(d)$ an *approximate unitary k-design with additive error ε* (short: *additive-error design*), if

$$\|\mathsf{M}_{k,\nu} - \mathsf{M}_k\|_\diamond \leq \varepsilon\,. \tag{230}$$

While this seems like a natural definition of approximation, it is also a rather weak. Many statistical properties in which we are interested in, are not well controlled by the diamond norm. This is because outcome probabilities are only close in additive error, and this does not say much if the probabilities are very small, for instance the collision probability in Example 2.8 which will play a major role in **??**.

Exercise 4.2 (Additive error misses rare events)**.** *Let ν be the ensemble that applies the identity with probability p and a Haar-random unitary otherwise. Show that $\|\mathsf{M}_{k,\nu} - \mathsf{M}_k\|_\diamond \leq 2p$. Compute $I_2 = \mathbb{E}_{U\sim\nu}|\langle 0|U|0\rangle|^4$ and show that it is not $\leq cd^{-2}$ for any constant c unless $p = \mathrm{O}(d^{-2})$. In the language of* **??***, ν does not* anticoncentrate*.*

Relative errors. In the following, we will introduce a definition that controls such small quantities multiplicatively. However, we restrict this multiplicative approximation to quantities that can be expressed as *Born probabilities*, i.e., are of the form $\mathrm{tr}[E\Phi_U(\rho)]$ where ρ is a quantum state and E is a POVM element. The common property of the latter operators are that they are *positive semidefinite* (psd): they are Hermitian and all eigenvalues are non-negative.

Definition 4.4: Approximate designs, relative error

A probability measure ν on $\mathrm{U}(d)$ is an *approximate unitary k-design with relative error ε* (short: *relative-error design*) if for all positive semidefinite A, B on $\mathcal{H}^{\otimes k} \otimes \mathcal{H}^{\otimes k}$:

$$(1-\varepsilon)\operatorname{tr}\big[B(\mathsf{M}_k \otimes \mathrm{id})(A)\big] \le \operatorname{tr}\big[B(\mathsf{M}_{k,\nu} \otimes \mathrm{id})(A)\big] \le (1+\varepsilon)\operatorname{tr}\big[B(\mathsf{M}_k \otimes \mathrm{id})(A)\big]. \quad (231)$$

Exercise 4.3 (Relative errors with CP ordering). *Equation (231) can be formulated in a more concise way: Recall that a linear map $\phi : \mathrm{L}(\mathcal{H}) \to \mathrm{L}(\mathcal{H})$ is called* completely positive *(CP) if $\phi \otimes \mathrm{id}$ maps psd operators to psd operators. We then write $\phi \preceq \phi'$ iff $\phi' - \phi$ is CP. Show that Def. 4.4 is equivalent to*

$$(1-\varepsilon)\,\mathsf{M}_k \preceq \mathsf{M}_{k\nu} \preceq (1+\varepsilon)\,\mathsf{M}_k\,. \quad (232)$$

Proposition 4.1 (Properties of relative error). *Let ν be an approximate k-design with relative error ε. Then we have:*

(i) Indistinguishability: $\|\mathsf{M}_{k,\nu} - \mathsf{M}_k\|_\diamond \le \varepsilon$.

(ii) Stability: *if $\Phi \preceq \Phi'$ and Λ_1, Λ_2 are completely positive, then $\Lambda_2 \Phi \Lambda_1 \preceq \Lambda_2 \Phi' \Lambda_1$ and $\Phi \otimes \mathrm{id} \preceq \Phi' \otimes \mathrm{id}$. Consequently, if Haar-random blocks in a random circuit are replaced by independent approximate designs with relative errors $\varepsilon_1, \dots, \varepsilon_m$ on the respective subsystems, the new moment operator M' and the old one M satisfy $\prod_i (1-\varepsilon_i)\mathsf{M} \preceq \mathsf{M}' \preceq \prod_i (1+\varepsilon_i)\mathsf{M}$.*

Proof. (i) Let ρ be a state, $\Delta := ((\mathsf{M}_{k,\nu} - \mathsf{M}_k) \otimes \mathrm{id})(\rho)$, and $\sigma := (\mathsf{M}_k \otimes \mathrm{id})(\rho)$. Let P be the projector onto the positive part of Δ. By Eq. (231), both $\varepsilon\sigma - \Delta$ and $\varepsilon\sigma + \Delta$ are positive semidefinite, hence

$$\|\Delta\|_1 = \operatorname{tr}[P\Delta] - \operatorname{tr}[(\mathbb{1} - P)\Delta] \le \varepsilon \operatorname{tr}[P\sigma] + \varepsilon \operatorname{tr}[(\mathbb{1} - P)\sigma] = \varepsilon\,. \quad (233)$$

(ii) Compositions and tensor products of completely positive maps are completely positive. For the consequence, write the moment operator of the circuit as an alternating product of the moment operators of the blocks and of the remaining (completely positive) parts, and replace one block at a time. □

Property (ii) makes relative error a convenient notion for constructions: errors of independent building blocks simply multiply. Further properties are collected in Exercise 4.9.

4.3 Lightcones and lower bounds

Consider a circuit U of depth t on a line, built from nearest-neighbour two-qudit gates. If an operator O is supported on a set X of qudits, then $UOU^\dagger$ is supported on the *lightcone* of X: every layer enlarges the support by at most one site on each side, see Fig. 8. In particular, the lightcone of a single qudit after t layers is an interval of at most $2t+1$ qudits.

Exact designs need linear depth. For simplicity, consider qubits and let A and B be Pauli operators on the first and the last qubit. Set $A_U := U^\dagger A U$ and define

$$C_{A,B}(U) = \frac{1}{2d} \operatorname{tr}\big([A_U, B]^\dagger [A_U, B]\big) = 1 - \operatorname{Re} F_4(U)\,, \qquad F_4(U) = \frac{1}{d} \operatorname{tr}\big(A_U B A_U B\big). \quad (234)$$

As in Sec. 2.2.2, we have the identity

$$F_4(U) = \frac{1}{d} \operatorname{tr}\big((A_U B)^{\otimes 2} \mathsf{F}\big) = \frac{1}{d} \operatorname{tr}\big((U^\dagger A U)^{\otimes 2} B^{\otimes 2} \mathsf{F}\big), \quad (235)$$

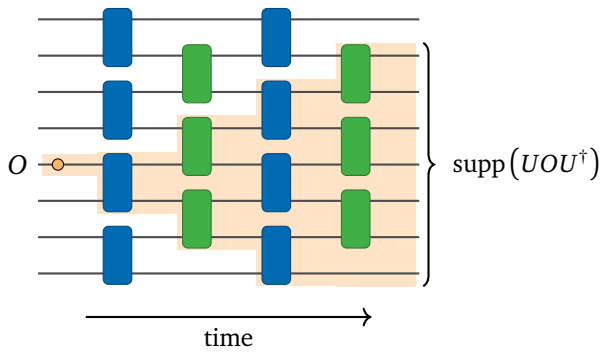


Figure 8: Lightcone in a brickwork circuit of depth $t = 4$. The support of the evolved operator $UOU^\dagger$ (shaded) grows by at most one site on each side per layer. Operators outside the lightcone of O commute with $UOU^\dagger$ for every choice of the gates.

where F is the swap operator between the two copies. Hence, $F_4(U)$ is a polynomial of degree 2 in the entries of U and $\overline{U}$, i.e. a second moment. If $t < n-1$, the lightcone of the first site does not reach the last one, so $[A_U, B] = 0$ and $F_4(U) = 1$ for *every* circuit of depth t. For Haar-random U, instead, it is straightforward to compute that $\mathbb{E}_{U\sim\text{Haar}} F_4 = -1/(d^2-1)$ and we will do so in **??**, see **??**. Hence, no ensemble of circuits of depth $t < n-1$ is an exact unitary 2-design, or an exact k-design for any $k \geq 2$ (Exercise 4.10). For qubits and $k \leq 3$, this linear lower bound is also achieved: every Clifford unitary can be implemented by nearest-neighbour circuits of depth $14n-4$ [24], and random Clifford unitaries form an exact 3-design.

Lightcones are invisible to approximate designs. Does the lightcone argument extend to approximate designs? From Eq. (235), the F_4 moment is

$$\mathbb{E}_{U\sim\nu} F_4(U) = \frac{1}{d} \operatorname{tr}\left[A^{\otimes 2}\, \mathsf{M}_{2,\nu}\left(B^{\otimes 2}\mathsf{F}\right)\right]. \tag{236}$$

If we want to bound the error on $\mathbb{E}_{U\sim\nu} F_4(U)$ for an additive-error design ν, we can make use of Hölder's inequality:

$$|\operatorname{tr}[X\phi(Y)]| \leq \|X\|_\infty \|\phi(Y)\|_1 \leq \|X\|_\infty \|\phi\|_\diamond \|Y\|_1 \tag{237}$$

Now note that both $X \equiv A^{\otimes 2}$ and $Y \equiv B^{\otimes 2}\mathsf{F}$ are unitaries, so while there operator norm is nicely bounded, their trace norm is exactly d since $\operatorname{tr}(U) = \operatorname{tr}\sqrt{U^\dagger U} = \operatorname{tr}\mathbb{1} = d$. Hence, plugging in the above replacement and $\phi = \mathsf{M}_{2,\nu} - \mathsf{M}_2$, we have

$$|\mathbb{E}_{U\sim\nu} F_4(U) - \mathbb{E}_{U\sim\text{Haar}} F_4(U)| \leq \varepsilon d\,. \tag{238}$$

Since $|F_4(U)| \leq 1$, this bound is vacuous unless $\varepsilon < 1/d$. Relative-error designs cannot do better because neither $A^{\otimes 2}$ nor $B^{\otimes 2}\mathsf{F}$ are psd. Expressing them as linear combinations of, e.g., states comes with an exponentially large overhead that cancels the multiplicative error.

Thus, lightcones do not put direct constraints on the depth of approximate designs as we defined them in Sec. 4.2. This means that it might be possible to realize them in much shorter depth than the spreading of information in the system would suggest.

Indeed, to measure F_4 experimentally, the setups considered in Sec. 4.2 are not sufficient, as the only allow for forward-in-time queries to a unitary U.[4] The situation changes if queries to the inverse $U^\dagger$ are also allowed (in the case of a RQC, we may simply run it backwards): Prepare

[4] There are forward-in-time experiment to measure F_4, but they need exponentially many queries to U [25] and this is necessary by the gluing argument in Ref. [26].

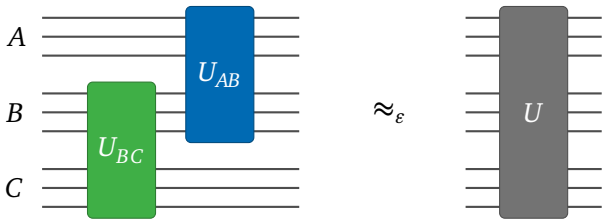


Figure 9: Gluing lemma: two independent random unitaries on AB and BC are indistinguishable from a Haar-random unitary U on ABC up to a small relative error, provided that the overlap B consists of $O(\log k)$ qubits. Time runs from left to right.

a maximally entangled state on the doubled system, apply the unitary $A_U B A_U B$ to one half of it, and measure in the Bell basis. This test accepts with probability $|F_4(U)|^2$. This probability equals one for shallow circuits, while it is $O(d^{-2})$ on average for Haar-random unitaries. With access to $U^\dagger$, lightcones thus enforce linear depth. Ensembles that are indistinguishable from Haar even under inverse queries are called *strong* designs [27].

4.4 Logarithmic-depth designs by gluing

We have seen in Sec. 4.3 that exact designs on a line need linear depth, while lightcones put no direct constraint on approximate designs. How shallow can an approximate design then be? In this section, we sketch a construction due to Schuster, Haferkamp, and Huang [26] which achieves relative-error designs in only logarithmic depth. The central insight is that *designs can be glued*: two random unitaries acting on overlapping subsystems together form an approximate design on the union, as long as the overlap is not too small. For simplicity, we restrict ourselves to qubits ($q = 2$) in this section.

Lemma 4.2 (Gluing lemma [26]). *Let A, B, C be disjoint sets of qubits and $d_B = 2^{|B|}$. Let U_{AB} and U_{BC} be independent Haar-random unitaries acting on AB and BC, respectively, and trivially on the remaining qubits. If $d_B \geq 2k^2$, then both $U_{AB}U_{BC}$ and $U_{BC}U_{AB}$ are approximate unitary k-designs on ABC with relative error*

$$\varepsilon \leq \frac{2k^2}{d_B}\,. \tag{239}$$

We do not prove the gluing lemma here and refer to Ref. [26] for details. The proof is an application of the Weingarten calculus from Sec. 2.3: If the dimension D is large compared to k^2, the permutation operators R_π become approximately orthogonal and, correspondingly, the Weingarten matrix is approximately $D^{-k}\mathbb{1}$. The Haar moment operators on AB, BC and ABC are thus all close to a simple sum over permutations, and the overlap B enters only through the overlaps $\mathrm{tr}(R_\sigma^\dagger R_\tau) = d_B^{\#(\sigma^{-1}\tau)}$ on B. The technical part is to establish these approximations in the completely positive order of Eq. (232). Remarkably, the error (239) only depends on the size of the overlap, not on the size of A and C.

Gluing many patches. We can now iterate the gluing lemma. Divide $n = m\xi$ qubits on a line into $m \geq 3$ consecutive *patches* $P_1, \ldots, P_m$ of ξ qubits each, and let $\chi = 2^\xi$ be the dimension of a patch. Place independent random unitaries $U_{j,j+1}$, which we call *bricks*, on neighbouring patches $P_j \cup P_{j+1}$, and arrange them in two layers: the bricks with odd j form the first layer, those with even j the second layer. This is a brickwork circuit on m thick wires of dimension χ, exactly the *super-brickwork circuits* that we will encounter again in **??**.

Proposition 4.2 (Designs by gluing [26]). *Assume that every brick $U_{j,j+1}$ is drawn independently from an approximate unitary k-design with relative error ε' on $P_j \cup P_{j+1}$, and that $\chi \geq 2k^2$. Then*

the super-brickwork circuit is an approximate unitary k-design on all n qubits with relative error ε, where

$$1+\varepsilon \leq (1+\varepsilon')^{m-1}\left(1+\frac{2k^2}{\chi}\right)^{m-2}. \tag{240}$$

In particular, for any $0<\varepsilon\leq 1$, the relative error is at most ε if $\varepsilon' \leq \varepsilon/(4m)$ and $\chi \geq 8mk^2/\varepsilon$, which is guaranteed by a patch size of $\xi = \lceil \log_2(8nk^2/\varepsilon)\rceil$ qubits.

Proof. By Thm. 4.1(ii), we can first assume that all bricks are Haar random, and replace them by the approximate designs afterwards, which yields the factor $(1+\varepsilon')^{m-1}$ (and analogously $(1-\varepsilon')^{m-1}$ for the lower bound). For Haar-random bricks, we glue the bricks one after the other, from top to bottom. Let $W_2 := U_{12}$, and let $W_{j+1} := W_j U_{j,j+1}$ if j is odd and $W_{j+1} := U_{j,j+1} W_j$ if j is even, i.e., bricks of the first layer are multiplied from the right and those of the second layer from the left. Since bricks within a layer commute, W_m is the full super-brickwork circuit. Now suppose that W_j, which acts on $P_1 \cup \cdots \cup P_j$, is a design with relative error ε_j. By Thm. 4.1(ii), we can replace W_j by a Haar-random unitary on $P_1 \cup \cdots \cup P_j$ at the cost of a factor $(1\pm\varepsilon_j)$. The gluing lemma, Lem. 4.2, with $A = P_1 \cup \cdots \cup P_{j-1}$, $B = P_j$ and $C = P_{j+1}$ then shows that $1+\varepsilon_{j+1} \leq (1+\varepsilon_j)(1+2k^2/\chi)$, and analogously for the lower bound. Starting from $\varepsilon_2 = 0$, we obtain Eq. (240) after $m-2$ gluing steps. For the second statement, use $1+x \leq e^x$ and $m \leq n$ to bound the right-hand side of Eq. (240) by $\exp(\varepsilon/4+\varepsilon/4) \leq 1+\varepsilon$ for $\varepsilon \leq 1$. The lower bound $(1-\varepsilon')^{m-1}(1-2k^2/\chi)^{m-2} \geq 1-\varepsilon/2$ follows in the same way. □

Filling the bricks with random circuits. Theorem 4.2 reduces the construction of a design on n qubits to the construction of designs on only $2\xi = \mathrm{O}(\log(nk/\varepsilon))$ qubits. We can therefore fill each brick with any construction that is efficient on 2ξ qubits, even if it requires depth linear in the number of qubits. For instance, brickwork random quantum circuits on N qubits are approximate unitary k-designs with relative error ε' in depth $\mathrm{poly}(k)\cdot\mathrm{O}(N+\log(1/\varepsilon'))$ [28–30]. Choosing $N = 2\xi$ and $\varepsilon' = \varepsilon/(4m)$, every brick has depth $\mathrm{poly}(k)\cdot\mathrm{O}(\log(nk/\varepsilon))$ and the entire circuit is twice as deep. We thus obtain relative-error k-designs on a line in depth

$$\mathrm{poly}(k)\cdot\mathrm{O}\big(\log(nk/\varepsilon)\big), \tag{241}$$

which is exponentially shallower than the linear depth required by exact designs. Note that the resulting circuit is not itself a brickwork circuit: each brick is a small brickwork circuit, but the bricks of the first layer act on disjoint blocks of 2ξ qubits and do not interact with each other. For fixed k and ε, the logarithmic depth (241) is optimal in n: Ref. [26] also shows that any circuit ensemble on a line that forms a relative-error 2-design requires depth $\Omega(\log n)$. Independently, LaRacuente and Leditzky [31] obtained relative-error designs in logarithmic depth from a similar two-step construction, using different techniques. The k-dependence in Eq. (241) can be reduced to linear using the recent spectral gap results by Baer and Haah [32].

4.5 Further reading

Random quantum circuits. Random quantum circuits are studied in quantum information theory as efficient approximations of the Haar measure, and in many-body physics as minimal models of chaotic quantum dynamics. The latter perspective, including entanglement growth and operator spreading, will be explored in more detail in Sec. 5. A gentle introduction to Haar-measure tools and moment operators is given in Ref. [15].

Convergence and spectral gaps. That local random circuits form approximate 2-designs in polynomial depth was first shown in Ref. [33]. Ref. [28] extended this to k-designs with relative error in the completely positive order of Def. 4.4, and showed that one-dimensional

brickwork circuits achieve it in depth $\mathrm{poly}(k)\cdot O(n+\log(1/\varepsilon))$ by bounding their spectral gap. The dependence on k has since been improved considerably [29, 30], and recently spectral gaps that do not decay with k have been established [32]. Circuits in higher dimensions and with long-range gates are treated in Ref. [34], and general fixed architectures in Ref. [35]. A complementary approach maps the moments of random circuits to statistical-mechanics models of permutations, see Ref. [36] for unitary designs and Ref. [37] for anticoncentration, which we will meet in **??**.

Approximate designs and lightcones. The diamond norm, the Holevo–Helstrom theorem and the Choi–Jamiołkowski isomorphism underlying the definitions in Sec. 4.2 are covered in the textbook [38]. Random Clifford unitaries can be implemented in linear depth on a line [24], which matches the lightcone bound for exact designs in Sec. 4.3. Designs that remain indistinguishable from the Haar measure under queries to both U and $U^\dagger$ (*strong* designs), and their relation to fast scrambling, are discussed in Ref. [27].

Gluing. The gluing lemma, its proof via Weingarten calculus, and the logarithmic-depth construction of Sec. 4.4 are from Schuster, Haferkamp, and Huang [26], which also proves the matching $\Omega(\log n)$ lower bound and discusses other geometries. LaRacuente and Leditzky [31] independently obtained shallow relative-error designs with a different proof technique based on alternating projections. Without the restriction to geometrically local gates, the depth can be reduced even further, see Ref. [39]. Finally, gluing also applies to *pseudorandom unitaries*, i.e., efficiently implementable ensembles that no efficient quantum algorithm can distinguish from the Haar measure: gluing small pseudorandom unitaries yields pseudorandom unitaries in low depth [26], and Ref. [40] shows how to construct pseudorandom unitaries in the first place.

Sublinear-depth designs. From the gluing lemma, we generally expect that additive-error and relative-error designs should be realizable in *logarithmic depth*. Laracuente [41] established almost-logarithmic depth for additive errors and brickwork circuits. First steps towards a similar result for relative errors were made by Heinrich, Haferkamp, Roth, and Helsen [42]. We expect that only strong designs match the bounds obtained by spectral gaps; since the spectral gaps are invariant under (partial) inversion, spectral gaps cannot differentiate between ordinary and strong designs.

4.6 Exercises on random circuits and approximate designs

Exercise 4.4 (Moment operators and spectral gaps). *Let ν be a probability measure on $\mathrm{U}(d)$ and let $\tilde{\nu}$ be the distribution of $U^\dagger$ for $U\sim\nu$.*

(a) Show that the adjoint of $\mathsf{M}_{k,\nu}$ with respect to the Hilbert–Schmidt inner product (cf. Sec. A) is $\mathsf{M}_{k,\tilde{\nu}}$. Conclude that the Haar moment operator M_k is self-adjoint.

(b) Show that every element of Comm_k is a fixed point of $\mathsf{M}_{k,\tilde{\nu}}$, and use (a) to show that $\mathsf{M}_{k,\nu}$ maps $\mathrm{Comm}_k^\perp$ to itself. This gives an alternative derivation of the block form (226).

(c) Suppose that $\mathsf{M}_{k,\nu}$ is self-adjoint, as is the case for the odd-even-odd layers of a Haar-random brickwork circuit in Example 4.2. Show that $\|\mathsf{M}_{k,\nu}^L-\mathsf{M}_k\|_\infty=(1-\Delta)^L$, where Δ is the spectral gap and $\|\cdot\|_\infty$ is the operator norm with respect to the Hilbert–Schmidt norm.

Exercise 4.5 (Brickwork layers as products of projectors). *Consider an open chain of n qudits with Haar-random two-qudit gates. Let $\mathsf{M}_{k,e}$ be the moment operator of a Haar-random gate on the edge $e=(x,x+1)$, and let $\mathsf{M}_{k,\mathrm{odd}}$ and $\mathsf{M}_{k,\mathrm{even}}$ be the moment operators of the odd and even layers.*

(a) *Show that every* $\mathsf{M}_{k,e}$ *is a Hilbert–Schmidt orthogonal projector, and that* $\mathsf{M}_{k,\mathrm{odd}}$ *and* $\mathsf{M}_{k,\mathrm{even}}$ *are orthogonal projectors, too.*

(b) *Let* $n=3$, $k=2$, *and* $X=\mathsf{F}\otimes\mathbb{1}\otimes\mathbb{1}$, *where* F *swaps the two copies of the first qudit. Use the Weingarten formula for* $k=2$ *to compute* $\mathsf{M}_{2,(1,2)}\mathsf{M}_{2,(2,3)}(X)$ *and* $\mathsf{M}_{2,(2,3)}\mathsf{M}_{2,(1,2)}(X)$. *Conclude that the two projectors do not commute.*

(c) *Let* $P_1,\dots,P_m$ *be orthogonal projectors on a Hilbert space. Show that* $\|P_m\cdots P_1 x\|=\|x\|$ *if and only if* $P_i x = x$ *for all* i.

(d) *Assume that the only operators fixed by all* $\mathsf{M}_{k,e}$ *are those in* Comm_k, *i.e.,* $\bigcap_e \mathrm{ran}(\mathsf{M}_{k,e})=\mathrm{Comm}_k$. *Use (c) to show that the brickwork circuit has a positive spectral gap* $\Delta>0$ *and thus converges to the Haar measure.*

Exercise 4.6 (Too many fixed points obstruct a design). *Let* ν *be any probability measure supported on unitaries that are diagonal in the computational basis of* $\mathbb{C}^d$, *with* $d>1$.

(a) *Show that every diagonal operator* A *is a fixed point of the first moment operator,* $\mathsf{M}_{1,\nu}(A)=A$.

(b) *Compare the dimension of this fixed-point space with the fixed-point space of the Haar moment operator* M_1.

(c) *Conclude that* ν *is not a unitary 1-design, no matter how broadly the phases are distributed. Explain the obstruction physically in terms of computational-basis populations.*

(d) *Consider a random quantum circuit whose gates are diagonal two-qudit unitaries. Explain why its moment operator does not converge to* M_1, *however deep the circuit is. What happens to the block form* (226)*?*

Exercise 4.7 (Random-position circuits). *Consider the random-position circuit of Example 4.1(c) on an open chain of* n *qubits, where in every time step one of the* $n-1$ *nearest-neighbour edges is chosen uniformly at random. Its moment operator after one step is* $\mathsf{M}_{k,\mathrm{RP}}=\frac{1}{n-1}\sum_e \mathsf{M}_{k,e}$. *We focus on the first moment,* $k=1$.

(a) *Let* P *be a Pauli string. Show that* $\mathsf{M}_{1,e}(P)=P$ *if* P *acts trivially on both qubits of* e, *and* $\mathsf{M}_{1,e}(P)=0$ *otherwise.*

(b) *Deduce that* $\mathsf{M}_{1,\mathrm{RP}}(P)=\big(1-\frac{w(P)}{n-1}\big)P$, *where* $w(P)$ *is the number of edges that touch the support of* P.

(c) *Show that* $2^{-n}\,\mathrm{tr}[P\,\mathsf{M}_{1,\mathrm{RP}}^{L}(P)]\le\varepsilon$ *for all* $P\neq\mathbb{1}$ *requires* $L\approx(n-1)\log(1/\varepsilon)$ *steps.*

(d) *In contrast, show that a single brickwork layer of* $n/2$ *Haar-random gates (*n *even) is an exact 1-design. Explain the difference.*

Exercise 4.8 (Random Clifford circuits). *Consider qubits and an architecture of two-qubit gates.*

(a) *Show that independent uniformly random two-qubit Clifford gates produce the same* k*-th moment operator as Haar-random gates for* $k\le 3$.

(b) *Consider a super-brickwork circuit as in Sec. 4.4 whose bricks are uniformly random Clifford unitaries on* 2ξ *qubits. Use Thm. 4.2 to show that it is an approximate 3-design with relative error* ε *for a suitable patch size* $\xi=\mathsf{O}(\log(n/\varepsilon))$. *Every Clifford unitary on* N *qubits can be implemented in depth* $\mathsf{O}(N)$ *on a line [24]. What is the resulting circuit depth?*

(c) *Let* $Q:=4^{-n}\sum_P P^{\otimes 4}$, *where the sum runs over all* 4^n *Pauli strings. Show that* Q *is an orthogonal projector and that* $C^{\otimes 4}QC^{\otimes 4,\dagger}=Q$ *for every Clifford unitary* C.

(d) *Conclude that Q is a fixed point of the fourth moment operator of every random circuit with Clifford gates. It is known that $Q \notin \mathrm{span}\{R_\pi : \pi \in \mathrm{S}_4\}$ [43]. Show that random Clifford circuits never converge to a unitary 4-design.*

Exercise 4.9 (Properties of relative error). *Let ν be an approximate unitary k-design with relative error ε.*

(a) *Show that ν is also an approximate j-design with relative error ε for every $j \leq k$.* Hint: *Show that $\mathrm{M}_{j,\nu}(Y) = d^{-(k-j)} \operatorname{tr}_{j+1,\dots,k}\left[\mathrm{M}_{k,\nu}(Y \otimes \mathbb{1}^{\otimes(k-j)})\right]$.*

(b) *Let ν' be an independent approximate k-design with relative error ε' on the same system. Show that the distribution of UU' with $U \sim \nu$, $U' \sim \nu'$ is an approximate k-design with relative error $\min(\varepsilon, \varepsilon')$.*

(c) *Consider the state ensemble $\{U|0\rangle : U \sim \nu\}$ and its moments I_k from* **??**. *Show that relative error ε implies $|I_k - I_k^{\mathrm{Haar}}| \leq \varepsilon I_k^{\mathrm{Haar}}$, while additive error ε only implies $|I_k - I_k^{\mathrm{Haar}}| \leq \varepsilon/d$. Compare both bounds with $I_k^{\mathrm{Haar}} = k!/\mathcal{G}_{k,d}$.*

Exercise 4.10 (Lightcones and exact designs). *Consider a nearest-neighbour circuit U of depth t on a line of n qubits. Let A and B be non-identity single-qubit Pauli operators on two qubits at distance $r > t$, and let $C_{A,B}(U)$ and $F_4(U)$ be defined as in Sec. 4.3.*

(a) *Use the lightcone to show that $[A_U, B] = 0$, and hence $C_{A,B}(U) = 0$ and $F_4(U) = 1$ for every circuit of depth t.*

(b) *Use the Haar value $\mathbb{E}_{U\sim\mathrm{Haar}} F_4 = -1/(d^2-1)$ from* **??** *to show that no ensemble of such circuits is an exact unitary 2-design. Show that every exact k-design is an exact j-design for all $j \leq k$, and conclude that exact k-designs with $k \geq 2$ require depth $t \geq n-1$.*

(c) *Derive Eq. (236) and show that $|\mathbb{E}_{U\sim\nu} F_4(U) - \mathbb{E}_{U\sim\mathrm{Haar}} F_4(U)| \leq \varepsilon d$ for every approximate 2-design ν with additive or with relative error ε. Why does this not contradict the existence of approximate designs of depth $t < r$?*

(d) *Show that the protocol described at the end of Sec. 4.3, which uses two queries to U and two to $U^\dagger$, accepts with probability $|F_4(U)|^2$.*

Exercise 4.11 (The gluing lemma is almost tight). *Let A, B, C be quantum systems of dimensions d_A, d_B, d_C with $d = d_A d_B d_C$, and let $V = U_{AB} U_{BC}$ be glued from independent Haar-random unitaries as in Lem. 4.2. We write M_k^{BC} for the moment operator of U_{BC}, and similarly for AB. Recall $\mathcal{G}_{k,D} = D(D+1)\cdots(D+k-1)$ from Lem. 2.3.*

(a) *Show that $\mathrm{M}_k^{BC}\left(|0\rangle\langle 0|^{\otimes k}\right) = |0_A\rangle\langle 0_A|^{\otimes k} \otimes \mathcal{G}_{k,d_B d_C}^{-1} \sum_{\sigma\in\mathrm{S}_k} R_\sigma^{BC}$.*

(b) *Show that $\mathbb{E}|\langle 0|V|0\rangle|^{2k} = k!\,\mathcal{G}_{k,d_B}/(\mathcal{G}_{k,d_A d_B}\,\mathcal{G}_{k,d_B d_C})$, and compare with the Haar value $k!/\mathcal{G}_{k,d}$.*

(c) *For $d_A, d_C \geq d_B$, show that the ratio of the two values is $1 + k(k-1)/(2d_B) + \mathrm{O}(k^4/d_B^2)$. Conclude that the relative error of V is at least $k(k-1)/(2d_B)$ to leading order, so the bound $2k^2/d_B$ of Lem. 4.2 is tight up to a constant factor.*

(d) *Compare with* **??** *for $m = 3$ thick wires.*

Exercise 4.12 (How deep is a glued design?). *Consider the construction of Sec. 4.4 on n qubits for $k = 2$ and relative error $\varepsilon = 0.01$.*

(a) *Use Thm. 4.2 to compute the patch size ξ and the number of qubits per brick for $n = 10^3$ and $n = 10^6$.*

(b) *Suppose that brickwork circuits on N qubits are approximate 2-designs with relative error ε' in depth $c\,(N+\log_2(1/\varepsilon'))$ for some constant c. Compare the depth of the glued construction with the depth of a brickwork circuit on all n qubits. From which n on does gluing pay off?*

(c) *Why can we not reduce the depth further by filling every brick with a glued circuit itself?* Hint: *Which relative error does every brick need, and what does Thm. 4.2 then require for the patches inside a brick?*

5 Random Dynamics in Many-Body Systems

References

[1] Barak Nehoran, Joseph Slote, and Henry Yuen. *All Unitaries Have Constant Depth Quantum Circuits*. 2026. DOI: 10.48550/arXiv.2609.40351. arXiv: 2609.40351.

[2] Markus Heinrich. *On stabiliser techniques and their application to simulation and certification of quantum devices*. PhD thesis. Universität zu Köln, 2021.

[3] Don Weingarten. *Asymptotic behavior of group integrals in the limit of infinite rank*. In: *Journal of Mathematical Physics* 19.5 (1978), pp. 999–1001. DOI: 10.1063/1.523807.

[4] Benoît Collins. *Moments and cumulants of polynomial random variables on unitary groups, the Itzykson–Zuber integral, and free probability*. In: *International Mathematics Research Notices* 2003.17 (2003), pp. 953–982. DOI: 10.1155/S107379280320917X.

[5] Benoît Collins, Sho Matsumoto, and Jonathan Novak. *The Weingarten Calculus*. In: *Notices of the AMS* 69 (2022), p. 734. DOI: 10.1090/noti2474.

[6] Christoph Dankert, Richard Cleve, Joseph Emerson, and Etera Livine. *Exact and approximate unitary 2-designs and their application to fidelity estimation*. In: *Physical Review A* 80.1 (2009), p. 012304. DOI: 10.1103/PhysRevA.80.012304.

[7] D. Gross, K. Audenaert, and J. Eisert. *Evenly distributed unitaries: on the structure of unitary designs*. In: *Journal of Mathematical Physics* 48.5 (2007), p. 052104. DOI: 10.1063/1.2716992. arXiv: quant-ph/0611002.

[8] D. P. DiVincenzo, D. W. Leung, and B. M. Terhal. *Quantum data hiding*. In: *IEEE Transactions on Information Theory* 48.3 (2002), pp. 580–598. DOI: 10.1109/18.985948.

[9] Hoi Fung Chau. *Unconditionally secure key distribution in higher dimensions by depolarization*. In: *IEEE Transactions on Information Theory* 51.4 (2005), pp. 1451–1468. DOI: 10.1109/TIT.2005.844076.

[10] Zak Webb. *The Clifford Group Forms a Unitary 3-design*. In: *Quantum Info. Comput.* 16.15 (2016), pp. 1379–1400.

[11] Huangjun Zhu. *Multiqubit Clifford groups are unitary 3-designs*. In: *Physical Review A* 96.6 (2017), p. 062336. DOI: 10.1103/PhysRevA.96.062336.

[12] Eiichi Bannai, Gabriel Navarro, Noelia Rizo, and Pham Huu Tiep. *Unitary t-groups*. In: *Journal of the Mathematical Society of Japan* (2020). Publisher: Mathematical Society of Japan. DOI: 10.2969/jmsj/82228222.

[13] Robert Guralnick and Pham Tiep. *Decompositions of small tensor powers and Larsen's conjecture*. In: *Representation Theory of the American Mathematical Society* 9.5 (2005), pp. 138–208. DOI: 10.1090/S1088-4165-05-00192-5.

[14] J. Haferkamp, F. Montealegre-Mora, M. Heinrich, J. Eisert, D. Gross, and I. Roth. *Efficient Unitary Designs with a System-Size Independent Number of Non-Clifford Gates*. en. In: *Communications in Mathematical Physics* 397.3 (2023), pp. 995–1041. DOI: 10.1007/s00220-022-04507-6.

[15] Antonio Anna Mele. *Introduction to Haar Measure Tools in Quantum Information: A Beginner's Tutorial*. In: *Quantum* 8 (2024), p. 1340. DOI: 10.22331/q-2024-05-08-1340.

[16] Angus Lowe and Ashwin Nayak. *Lower Bounds for Learning Quantum States with Single-Copy Measurements*. In: *ACM Transactions on Computation Theory* 17.1, 7 (2025), pp. 1–42. DOI: 10.1145/3717450. arXiv: 2207.14438.

[17] Jeongwan Haah, Aram W. Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. *Sample-Optimal Tomography of Quantum States*. In: *IEEE Transactions on Information Theory* 63.9 (2017), pp. 5628–5641. DOI: 10.1109/TIT.2017.2719044. arXiv: 1508.01797.

[18] Lorenzo Leone, Salvatore F. E. Oliviero, and Alioscia Hamma. *Nonstabilizerness determining the hardness of direct fidelity estimation*. In: *Physical Review A* 107.2 (2023), p. 022429. DOI: 10.1103/PhysRevA.107.022429.

[19] Hsin-Yuan Huang, Richard Kueng, and John Preskill. *Predicting many properties of a quantum system from very few measurements*. In: *Nature Physics* 16.10 (2020), pp. 1050–1057. DOI: 10.1038/s41567-020-0932-7.

[20] Srilekha Gandhari, Victor V. Albert, Thomas Gerrits, Jacob M. Taylor, and Michael J. Gullans. *Precision Bounds on Continuous-Variable State Tomography Using Classical Shadows*. In: *PRX Quantum* 5.1 (2024), p. 010346. DOI: 10.1103/PRXQuantum.5.010346.

[21] Riccardo Cioli, Elisa Ercolessi, Matteo Ippoliti, Xhek Turkeshi, and Lorenzo Piroli. *Approximate Inverse Measurement Channel for Shallow Shadows*. In: *Quantum* 9 (2025), p. 1698. DOI: 10.22331/q-2025-04-08-1698. arXiv: 2407.11813.

[22] Chengsi Mao, Changhao Yi, and Huangjun Zhu. *The Magic in Qudit Shadow Estimation based on the Clifford Group*. 2024. DOI: 10.48550/arXiv.2410.13572. arXiv: 2410.13572.

[23] Huangjun Zhu, Chengsi Mao, and Changhao Yi. *Third moments of qudit Clifford orbits and 3-designs based on magic orbits*. 2024. DOI: 10.48550/arXiv.2410.13575. arXiv: 2410.13575.

[24] Dmitri Maslov and Martin Roetteler. *Shorter stabilizer circuits via Bruhat decomposition and quantum circuit transformations*. In: *IEEE Transactions on Information Theory* 64.7 (2018), pp. 4729–4738. arXiv: 1705.09176.

[25] Benoît Vermersch, Andreas Elben, Lukas M. Sieberer, Norman Y. Yao, and Peter Zoller. *Probing scrambling using statistical correlations between randomized measurements*. In: *Physical Review X* 9.2 (2019), p. 021061. DOI: 10.1103/PhysRevX.9.021061.

[26] Thomas Schuster, Jonas Haferkamp, and Hsin-Yuan Huang. *Random unitaries in extremely low depth*. 2025. DOI: 10.48550/arXiv.2407.07754. arXiv: 2407.07754.

[27] Thomas Schuster, Fermi Ma, Alex Lombardi, Fernando Brandão, and Hsin-Yuan Huang. *Strong random unitaries and fast scrambling*. 2025. arXiv: 2509.26310.

[28] Fernando G. S. L. Brandão, Aram W. Harrow, and Michał Horodecki. *Local Random Quantum Circuits are Approximate Polynomial-Designs*. In: *Communications in Mathematical Physics* 346.2 (2016), pp. 397–434. arXiv: 1208.0692.

[29] Jonas Haferkamp. *Random quantum circuits are approximate unitary t-designs in depth* $O\left(nt^{5+o(1)}\right)$. In: *Quantum* 6 (2022), p. 795. DOI: 10.22331/q-2022-09-08-795. arXiv: 2203.16571.

[30] Chi-Fang Chen, Jeongwan Haah, Jonas Haferkamp, Yunchao Liu, Tony Metger, and Xinyu Tan. *Incompressibility and spectral gaps of random circuits*. 2024. DOI: 10.48550/arXiv.2406.07478. arXiv: 2406.07478.

[31] Nicholas LaRacuente and Felix Leditzky. *Approximate Unitary k-Designs from Shallow, Low-Communication Circuits*. In: *Communications in Mathematical Physics* 407 (2026), p. 51. DOI: 10.1007/s00220-025-05542-9. arXiv: 2407.07876.

[32] Tim Baer and Jeongwan Haah. *Random Unitary Circuits with Constant Spectral Gap*. 2026. DOI: 10.48550/arXiv.2607.20919. arXiv: 2607.20919.

[33] Aram W. Harrow and Richard A. Low. *Random Quantum Circuits are Approximate 2-designs*. In: *Communications in Mathematical Physics* 291.1 (2009), pp. 257–302. DOI: 10.1007/s00220-009-0873-6. arXiv: 0802.1919.

[34] Aram W. Harrow and Saeed Mehraban. *Approximate Unitary t-Designs by Short Random Quantum Circuits Using Nearest-Neighbor and Long-Range Gates*. In: *Communications in Mathematical Physics* 401 (2023), pp. 1531–1626. DOI: 10.1007/s00220-023-04675-z. arXiv: 1809.06957.

[35] Daniel Belkin, James Allen, Soumik Ghosh, Christopher Kang, Sophia Lin, James Sud, Fred Chong, Bill Fefferman, and Bryan K. Clark. *Approximate t-designs in generic circuit architectures*. 2023. arXiv: 2310.19783.

[36] Nicholas Hunter-Jones. *Unitary designs from statistical mechanics in random quantum circuits*. 2019. arXiv: 1905.12053.

[37] Alexander M. Dalzell, Nicholas Hunter-Jones, and Fernando G. S. L. Brandão. *Random quantum circuits anti-concentrate in log depth*. In: *PRX Quantum* 3.1 (2022), p. 010333. DOI: 10.1103/PRXQuantum.3.010333. arXiv: 2011.12277 [quant-ph].

[38] John Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018. DOI: 10.1017/9781316848142.

[39] Laura Cui, Thomas Schuster, Fernando Brandão, and Hsin-Yuan Huang. *Unitary designs in nearly optimal depth*. 2025. arXiv: 2507.06216.

[40] Fermi Ma and Hsin-Yuan Huang. *How to Construct Random Unitaries*. 2024. arXiv: 2410.10116.

[41] Nicholas Laracuente. *Quantum Relative Entropy Decay Composition Yields Shallow, Unstructured k-Designs*. 2025. DOI: 10.48550/arXiv.2510.08537. arXiv: 2510.08537.

[42] Markus Heinrich, Jonas Haferkamp, Ingo Roth, and Jonas Helsen. *Anticoncentration is (almost) all you need*. In: *Physical Review Letters* 137.5 (2026). arXiv:2510.23719 [quant-ph], p. 050601. DOI: 10.1103/z3mp-5gml.

[43] Huangjun Zhu, Richard Kueng, Markus Grassl, and David Gross. *The Clifford group fails gracefully to be a unitary 4-design*. In: *arXiv:1609.08172 [quant-ph]* (2016). arXiv: 1609.08172.

A Some linear algebra

This appendix reviews the finite-dimensional linear algebra used throughout the course. Most of the material should be familiar from introductory linear algebra and quantum mechanics, but we include the details to make the notes self-contained and to fix conventions that will be used in later sections.

A.1 States, operators, superoperators

State space. Quantum mechanics is formulated on a Hilbert space $\mathcal{H}$. Throughout these notes, $\mathcal{H}$ is finite-dimensional, so after choosing an orthonormal basis $\{|x\rangle\}_{x=0}^{d-1}$ we may identify $\mathcal{H} \simeq \mathbb{C}^d$. The inner product of $|\psi\rangle, |\varphi\rangle \in \mathcal{H}$ can then be written as

$$\langle\psi|\varphi\rangle = \sum_{x=0}^{d-1} \overline{\psi_x}\, \varphi_x, \tag{242}$$

where $\psi_x = \langle x|\psi\rangle$ and $\varphi_x = \langle x|\varphi\rangle$. By convenction, quantum state vectors are normalized, $\langle\psi|\psi\rangle = 1$.

Dirac notation distinguishes vectors $|\psi\rangle \in \mathcal{H}$, called *kets*, from dual vectors $\langle\psi| \in \mathcal{H}^*$, called *bras*. The bra associated with $|\psi\rangle$ is the linear functional

$$\mathcal{H} \ni |\varphi\rangle \longmapsto \langle\psi|\varphi\rangle. \tag{243}$$

The finite-dimensional Riesz representation theorem guarantees this correspondence. Pairing a bra and a ket gives the inner product. Reversing their order gives the rank-one operator

$$|\psi\rangle\langle\varphi| : |\chi\rangle \longmapsto |\psi\rangle\langle\varphi|\chi\rangle. \tag{244}$$

Linear operators. The vector space of linear operators $A : \mathcal{H} \to \mathcal{H}$ is denoted by $\mathrm{L}(\mathcal{H})$. For every $A \in \mathrm{L}(\mathcal{H})$, its adjoint $A^\dagger \in \mathrm{L}(\mathcal{H})$ is uniquely defined by

$$\langle\psi|A\varphi\rangle = \langle A^\dagger\psi|\varphi\rangle, \qquad \text{for all } |\psi\rangle, |\varphi\rangle \in \mathcal{H}. \tag{245}$$

In an orthonormal basis, $A^\dagger = \overline{A}^{\mathsf{T}}$ is the conjugate transpose of the matrix representing A.

Definition A.1: Important classes of operators

Let $A \in \mathrm{L}(\mathcal{H})$.

- A is *Hermitian*, or self-adjoint, if $A^\dagger = A$. Its eigenvalues are real and it admits an orthonormal eigenbasis.
- A is *unitary* if $A^\dagger A = AA^\dagger = \mathbb{1}_{\mathcal{H}}$.
- A is *positive semidefinite*, written $A \geq 0$, if it is Hermitian and all its eigenvalues are non-negative. Equivalently, $\langle\psi|A|\psi\rangle \geq 0$ for every $|\psi\rangle \in \mathcal{H}$.
- A is an orthogonal *projector* if $A = A^\dagger = A^2$.
- A *quantum state* is an operator $\rho \in \mathrm{L}(\mathcal{H})$ satisfying

$$\rho \geq 0, \qquad \mathrm{tr}[\rho] = 1. \tag{246}$$

  It is pure if it has rank one, equivalently if $\rho^2 = \rho$. Every pure state can be written $\rho = |\psi\rangle\langle\psi|$ for a normalized ket $|\psi\rangle$.

The unitaries on $\mathcal{H}$ form the unitary group $\mathrm{U}(\mathcal{H}) \simeq \mathrm{U}(d)$.

Tensor products. The tensor product $\mathcal{H}_A \otimes \mathcal{H}_B$ of two Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ describes a composite quantum system. Its elements are linear combinations of *product states* $|\psi\rangle \otimes |\varphi\rangle$. Product states describe quantum systems in which subsystem A does not have any correlations with subsystem B (and vice versa). Quantum state vectors which are *not* product states are called *entangled*. More generally, we say that a (possibly mixed) quantum state ρ is *separable* if it is a convex combination of product states, $\rho = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i| \otimes |\varphi_i\rangle\langle\varphi_i|$ with $\lambda_i \geq 0$. If a quantum state ρ is not separable, it is entangled.

Traces and reduced states. The trace of an operator X is defined as

$$\mathrm{tr}(X) := \sum_{x=0}^{d-1} \langle x\,|X|\,x\rangle\,, \tag{247}$$

where $\{|x\rangle\}_{x=0}^{d-1}$ is again some orthonormal basis of $\mathcal{H}$. The definition does not depend on the choise of basis. For a composite system $\mathcal{H}_A \otimes \mathcal{H}_B$, we can define the partial trace (over B) of an operator $X \in \mathrm{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$ as the unique operator $\mathrm{tr}_B(X) \in \mathrm{L}(\mathcal{H}_A)$ such that

$$\mathrm{tr}_A(O_A \,\mathrm{tr}_B(X)) = \mathrm{tr}(O_A \otimes \mathbb{1}_B X)\,, \quad \forall O_A \in \mathrm{L}(\mathcal{H}_A)\,. \tag{248}$$

Explicitly, we can compute the partial trace by fixing an ONB of B and evaluate

$$\mathrm{tr}_B(X) = \sum_{x=0}^{d_B-1} (\mathbb{1}_A \otimes \langle x|) X (\mathbb{1}_A \otimes |x\rangle)\,. \tag{249}$$

If we are given X as a sum of product operators, we can evaluate Eq. (249) on the individual terms.

For a quantum state ρ on AB, the reduced state on A is $\rho_A = \mathrm{tr}_B(\rho)$ (and analogously on B). Because of Eq. (248), the reduced state encodes all physical information that is needed to evaluate *local* expectation values on subsystem A.

Operator space. The operator space $\mathrm{L}(\mathcal{H})$ is itself a Hilbert space. Its dimension is

$$\dim \mathrm{L}(\mathcal{H}) = (\dim \mathcal{H})^2 = d^2, \tag{250}$$

and its Hilbert–Schmidt, or trace, inner product is

$$(X \,|\, Y) := \mathrm{tr}[X^\dagger Y]. \tag{251}$$

An orthonormal operator basis therefore contains d^2 operators. Writing it as $\{A_\mu\}_{\mu=1}^{d^2}$, orthonormality means

$$(A_\mu | A_\nu) = \delta_{\mu\nu}. \tag{252}$$

Superoperators and quantum channels. A *superoperator* on $\mathcal{H}$ is a linear map

$$\phi : \mathrm{L}(\mathcal{H}) \longrightarrow \mathrm{L}(\mathcal{H}). \tag{253}$$

It is *positive* if $\phi(A) \geq 0$ whenever $A \geq 0$. Positivity alone does not ensure that the map remains positive when it acts on one part of an entangled system. We therefore call ϕ *completely positive* if

$$\phi \otimes \mathrm{id}_{\mathcal{A}} \tag{254}$$

is positive for every finite-dimensional auxiliary Hilbert space $\mathcal{A}$.

Definition A.2: Quantum channel

A quantum channel is a completely positive, trace-preserving linear map. Trace preservation means

$$\mathrm{tr}[\phi(A)] = \mathrm{tr}[A] \qquad \text{for every } A \in \mathrm{L}(\mathcal{H}). \tag{255}$$

The channel is *unital* if

$$\phi(\mathbb{1}_{\mathcal{H}}) = \mathbb{1}_{\mathcal{H}}. \tag{256}$$

Complete positivity guarantees physical action even when an arbitrary auxiliary system is present; trace preservation guarantees that normalized states remain normalized.

Example A.1: Quantum channels

Common examples include:

- *Unitary channels*: $\phi(X) = UXU^\dagger$ for $U \in \mathrm{U}(\mathcal{H})$.
- *Mixed-unitary channels*: $\phi(X) = \sum_i p_i U_i X U_i^\dagger$, where $p_i \geq 0$ and $\sum_i p_i = 1$.
- *Dephasing in the basis* $\{|x\rangle\}$: $\phi(X) = \sum_x \langle x|X|x\rangle |x\rangle\langle x|$.
- *Reset to a fixed state* ρ: $\phi(X) = \mathrm{tr}[X]\rho$.

The reset channel illustrates that trace preservation and unitality are different: $\phi(\mathbb{1}_{\mathcal{H}}) = d\rho$, which equals $\mathbb{1}_{\mathcal{H}}$ only for the maximally mixed reset state $\rho = \mathbb{1}_{\mathcal{H}}/d$.

Pauli operators and binary notation. For one qubit, let

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad Y = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \qquad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{257}$$

We use the binary phase convention

$$\sigma_{x,z} := \mathrm{i}^{xz} X^x Z^z, \qquad (x,z) \in \mathbb{Z}_2^2. \tag{258}$$

It gives

$$\sigma_{0,0} = \mathbb{1}, \qquad \sigma_{1,0} = X, \qquad \sigma_{0,1} = Z, \qquad \sigma_{1,1} = Y. \tag{259}$$

For n qubits, write a binary label in split form

$$\mathbf{a} = (\mathbf{x}, \mathbf{z}) = (x_1, \ldots, x_n, z_1, \ldots, z_n) \in \mathbb{Z}_2^{2n}. \tag{260}$$

The corresponding Pauli string is

$$\boxed{\sigma_{\mathbf{a}} := \bigotimes_{j=1}^{n} \sigma_{x_j,z_j} = \mathrm{i}^{\mathbf{x}\cdot\mathbf{z}} \bigotimes_{j=1}^{n} X_j^{x_j} Z_j^{z_j}.} \tag{261}$$

The dot product in the phase is evaluated as an integer. Binary vector addition below is instead understood modulo two.

The 4^n Pauli strings are Hilbert–Schmidt orthogonal:

$$(\sigma_{\mathbf{a}} | \sigma_{\mathbf{b}}) = \mathrm{tr}[\sigma_{\mathbf{a}} \sigma_{\mathbf{b}}] = 2^n \delta_{\mathbf{a},\mathbf{b}}. \tag{262}$$

Since $d = 2^n$, the normalized strings

$$\hat{\sigma}_{\mathbf{a}} := 2^{-n/2} \sigma_{\mathbf{a}} \tag{263}$$

form an orthonormal basis of $\mathrm{L}((\mathbb{C}^2)^{\otimes n})$.

Generalized Weyl operators give an analogous orthogonal unitary basis for a qudit of arbitrary local dimension q. For $q > 2$ those basis operators are not generally Hermitian, so the binary phase and symplectic conventions used here should not be transferred to qudits without modification.

Exercise A.1 (Properties of Pauli operators). *Using Eq. (261), prove the following statements. Let* $\mathbf{a} = (\mathbf{x}, \mathbf{z})$ *and* $\mathbf{b} = (\mathbf{x}', \mathbf{z}')$.

(a) $\sigma_{\mathbf{a}}^\dagger = \sigma_{\mathbf{a}}$ *and* $\sigma_{\mathbf{a}}^2 = \mathbb{1}$.

(b) $\sigma_{\mathbf{a}} \sigma_{\mathbf{b}}$ *is a phase times* $\sigma_{\mathbf{a}\oplus\mathbf{b}}$, *where* $\oplus$ *denotes componentwise addition in* $\mathbb{Z}_2^{2n}$.

(c) Define the binary symplectic form by

$$[\mathbf{a}, \mathbf{b}] := \mathbf{x}\cdot\mathbf{z}' + \mathbf{z}\cdot\mathbf{x}' \pmod 2. \tag{264}$$

Show that

$$\sigma_{\mathbf{a}} \sigma_{\mathbf{b}} = (-1)^{[\mathbf{a},\mathbf{b}]} \sigma_{\mathbf{b}} \sigma_{\mathbf{a}}. \tag{265}$$

Conclude that two Pauli strings commute exactly when their symplectic product vanishes.

(d) Derive Eq. (262) and verify that there are exactly $\dim \mathrm{L}((\mathbb{C}^2)^{\otimes n}) = 4^n$ *basis elements.*

Operator Dirac notation and Pauli-transfer matrices. We use the Hilbert–Schmidt inner product to identify an operator $X \in \mathrm{L}(\mathcal{H})$ with an operator ket $|X)$. Its dual is $(X|$, defined by

$$(X||Y) = (X|Y) = \mathrm{tr}[X^\dagger Y]. \tag{266}$$

The operator-space outer product acts as

$$|X)(Y||A) = |X)(Y|A). \tag{267}$$

For the normalized Pauli basis,

$$\mathrm{id}_{\mathrm{L}(\mathcal{H})} = \sum_{\mathbf{a}} |\hat{\sigma}_{\mathbf{a}})(\hat{\sigma}_{\mathbf{a}}|. \tag{268}$$

Inserting this resolution of the identity on both sides of a superoperator gives

$$\begin{aligned} \phi &= \sum_{\mathbf{a},\mathbf{b}} |\hat{\sigma}_{\mathbf{a}})(\hat{\sigma}_{\mathbf{a}}|\phi|\hat{\sigma}_{\mathbf{b}})(\hat{\sigma}_{\mathbf{b}}| \\ &= \sum_{\mathbf{a},\mathbf{b}} \phi_{\mathbf{a},\mathbf{b}} |\hat{\sigma}_{\mathbf{a}})(\hat{\sigma}_{\mathbf{b}}|, \end{aligned} \tag{269}$$

where

$$\phi_{\mathbf{a},\mathbf{b}} = (\hat{\sigma}_{\mathbf{a}}|\phi|\hat{\sigma}_{\mathbf{b}}) = \mathrm{tr}[\hat{\sigma}_{\mathbf{a}}\phi(\hat{\sigma}_{\mathbf{b}})]. \tag{270}$$

The second index labels the input column, and the first index labels the output row.

Let $\mathbf{0}$ denote the all-zero binary vector, so that $\hat{\sigma}_{\mathbf{0}} = \mathbb{1}_{\mathcal{H}}/\sqrt{d}$. Trace preservation fixes the *row* with output index $\mathbf{0}$:

$$\begin{aligned} \phi_{\mathbf{0},\mathbf{b}} &= \frac{1}{\sqrt{d}} \mathrm{tr}[\phi(\hat{\sigma}_{\mathbf{b}})] \\ &= \frac{1}{\sqrt{d}} \mathrm{tr}[\hat{\sigma}_{\mathbf{b}}] = \delta_{\mathbf{0},\mathbf{b}}. \end{aligned} \tag{271}$$

Unitality instead fixes the *column* with input index $\mathbf{0}$:

$$\phi_{\mathbf{a},\mathbf{0}} = \delta_{\mathbf{a},\mathbf{0}}. \tag{272}$$

Ordering the identity component first, a trace-preserving map and a trace-preserving unital map therefore have the respective block forms

$$\phi_{\mathrm{TP}} \simeq \begin{pmatrix} 1 & 0 \\ \mathbf{t} & T \end{pmatrix}, \qquad \phi_{\mathrm{TP+unital}} \simeq \begin{pmatrix} 1 & 0 \\ 0 & T \end{pmatrix}. \tag{273}$$

Exercise A.2 (Pauli-transfer representation). *Let ϕ be an n-qubit quantum channel and let $(\phi_{\mathbf{a},\mathbf{b}})_{\mathbf{a},\mathbf{b}}$ be its Pauli-transfer matrix as defined in Eq. (269).*

(a) Show that every entry is real. Use that a quantum channel maps Hermitian operators to Hermitian operators.

(b) Reproduce Eqs. (271) and (272) and explain why trace preservation constrains a row whereas unitality constrains a column in the convention of Eq. (270).

(c) Let ϕ be a Pauli channel,

$$\phi(X) = \sum_{\mathbf{a}} p_{\mathbf{a}} \sigma_{\mathbf{a}} X \sigma_{\mathbf{a}}, \qquad p_{\mathbf{a}} \geq 0, \qquad \sum_{\mathbf{a}} p_{\mathbf{a}} = 1. \tag{274}$$

Show that every $\hat{\sigma}_{\mathbf{b}}$ is an eigenoperator:

$$\phi(\hat{\sigma}_{\mathbf{b}}) = \left[\sum_{\mathbf{a}} p_{\mathbf{a}} (-1)^{[\mathbf{a},\mathbf{b}]}\right] \hat{\sigma}_{\mathbf{b}}. \tag{275}$$

Conclude that its Pauli-transfer matrix is diagonal. Notice that the channel probabilities $p_{\mathbf{a}}$ are not, in general, the diagonal eigenvalues themselves.

Norms. Let $X : \mathcal{V} \to \mathcal{W}$ be a linear map between finite-dimensional Hilbert spaces, and let $\sigma_1, \ldots, \sigma_r$ be its nonzero singular values. Equivalently, the σ_i^2 are the nonzero eigenvalues of $X^\dagger X$, and $|X| := \sqrt{X^\dagger X}$. For $1 \le p < \infty$, the Schatten p-norm is

$$\|X\|_p := \left(\mathrm{tr}[|X|^p]\right)^{1/p} = \left(\sum_{i=1}^{r} \sigma_i^p\right)^{1/p}. \tag{276}$$

The endpoint $p = \infty$ is defined by a limit:

$$\|X\|_\infty := \max_i \sigma_i. \tag{277}$$

The cases $p = 1$, $p = 2$, and $p = \infty$ are respectively the trace, Hilbert–Schmidt, and spectral norms.

A superoperator can itself be regarded as a matrix acting between operator Hilbert spaces, so it also has Schatten norms. These should not be confused with induced norms. For example,

$$\|\phi\|_{2\to 2} := \sup_{X \neq 0} \frac{\|\phi(X)\|_2}{\|X\|_2} \tag{278}$$

is the spectral norm of the matrix representing ϕ, whereas $\|\phi\|_2$ is the Hilbert–Schmidt norm of that matrix.

A.2 Non-orthonormal bases

Let $\{|f_i\rangle\}_{i=1}^m$ be a basis of an m-dimensional Hilbert space $\mathcal{V}$. Every vector has a unique expansion

$$|v\rangle = \sum_{i=1}^{m} v_i |f_i\rangle. \tag{279}$$

For an orthonormal basis, $v_i = \langle f_i | v\rangle$. For a non-orthonormal basis, the correct coefficients are obtained from the *dual basis* $\{|\widetilde{f}_i\rangle\}_{i=1}^m$, defined by

$$\langle \widetilde{f}_i | f_j \rangle = \delta_{ij}. \tag{280}$$

It follows immediately that

$$v_i = \langle \widetilde{f}_i | v\rangle, \qquad |v\rangle = \sum_i |f_i\rangle\langle \widetilde{f}_i | v\rangle, \tag{281}$$

and hence

$$I_{\mathcal{V}} = \sum_i |f_i\rangle\langle \widetilde{f}_i| = \sum_i |\widetilde{f}_i\rangle\langle f_i|. \tag{282}$$

To construct the dual basis, define the Gram matrix of the original basis, not of a new family of vectors, by

$$G_{ij} := \langle f_i | f_j \rangle. \tag{283}$$

For every nonzero coefficient vector $c \in \mathbb{C}^m$,

$$c^\dagger G c = \left\| \sum_{i=1}^{m} c_i |f_i\rangle \right\|^2 > 0. \tag{284}$$

Thus G is positive definite, not merely positive semidefinite, and is invertible. Let $W := G^{-1}$. In a complex Hilbert space it is safest to state the construction first at the level of bras:

$$\boxed{\langle \widetilde{f}_i| = \sum_{j=1}^{m} W_{ij} \langle f_j|.} \tag{285}$$

Indeed,

$$\langle \widetilde{f}_i | f_\ell \rangle = \sum_j W_{ij} G_{j\ell} = (WG)_{i\ell} = \delta_{i\ell}. \tag{286}$$

Since G and W are Hermitian, the corresponding ket is

$$\boxed{|\widetilde{f}_i\rangle = \sum_{j=1}^{m} W_{ji} |f_j\rangle.} \tag{287}$$

The transposed index order in the ket formula is essential over $\mathbb{C}$; omitting it silently assumes that the Gram matrix is real. For an orthonormal basis $G = W = I_m$, so the dual basis coincides with the original basis.

B Unsigned Stirling numbers of the first kind

The entry in row k and column r is $c(k, r)$, the number of permutations of k objects with exactly r cycles. Blank entries are zero. Each row sums to $k!$. The recurrence

$$c(k+1, r) = k\,c(k, r) + c(k, r-1) \tag{288}$$

generates the next row.

$k\backslash r$	1	2	3	4	5	6	7	8	9	10
1	1									
2	1	1								
3	2	3	1							
4	6	11	6	1						
5	24	50	35	10	1					
6	120	274	225	85	15	1				
7	720	1764	1624	735	175	21	1			
8	5040	13068	13132	6769	1960	322	28	1		
9	40320	109584	118124	67284	22449	4536	546	36	1	
10	362880	1026576	1172700	723680	269325	63273	9450	870	45	1