The Structure and Interpretation of Quantum Programs I: *Foundations*

David Wakeham ◇ Torsor Labs

ABSTRACT

Qubits are a great way to build a quantum computer, but a limited way to program one. We replace the usual "states and gates" formalism with a "props and ops" (propositions and operators) model in which

- the C*-algebra of observables supplies the syntax;
- states, viewed as linear functionals, give the semantics; and
- a novel diagrammatic calculus unifies the two.

The first part develops the basic objects of the framework, encoding consistent patterns of operator correlation, recovering Hilbert space via the GNS construction, and re-deriving the Bloch sphere as the set of all consistent correlations of operators in the Pauli algebra.

We then turn to intervention, showing how measurement modifies state, proving an operator-algebraic version of the Knill-Laflamme conditions, and expressing stabilizer codes with the same diagrammatic machinery. This provides a concise, representation-agnostic account of quantum error correction. The result is a self-contained foundation in which C*-algebras, and their dual Hilbert spaces, offer a rich and universal substrate for quantum programming; forthcoming papers will build a high-level language and quantum software applications on top of this substrate.

Torsor Labs
YAW-ב$_0$-4-25

Reading Guide

What is this document? The first installment of *Structure and Interpretation of Quantum Programming (SIQP)*, a systematic reconstruction of quantum computing from algebraic foundations. Like its spiritual predecessor *SICP*,[1] we seek not merely to explain quantum programming, but to understand it more deeply via the interplay of computational pragmatics, design principles, and logical structure.

[1] *Structure and Interpretation of Computer Programs* (1985), Harold Abelson, Gerald Jay Sussman and Julie Sussman.

Why reinvent the wheel? The standard formulation of quantum computing—based on Hilbert space, qubits, and circuits—plays the same role in quantum computing that truth tables do in classical. This is an important part of the story, but by no means all! By starting with C*-algebras and observables as our primitive, we not only recover the standard treatment, but enable new and flexible patterns of abstraction that stand a better chance of scaling with hardware.

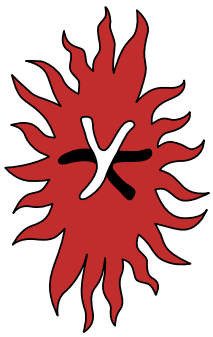
How do you do that? Algebras are more flexible than qubits. You can specify qudits, groups, harmonic oscillators, or open systems with the same ease you work with one and zeros. You can naturally incorporate error correction via the stabilizer formalism, and use the same tools for simulation via Gottesman-Knill. You can apply all this to near-term (e.g., classical shadows), medium-term (e.g., Hadamard tests) and long-term applications (e.g. hidden subgroup problem).

Who should read this? We envision a few different segments:

- *For the traditionalist*: Start with the Bloch sphere (§8) to see the familiar from a new angle, then work backward to understand how it emerges algebraically. Who knows, you might like it!
- *For the information theorist*: Focus on the GNS construction (§5-7), sharpness (§10-12) and error correction (§13-16). Together, these tell a nice story about the algebraic dynamics of information.
- *For the computer scientist*: The abstract wiring diagrams (§3) are the combinators of our upcoming quantum programming language. Get familiar with them and await Part II!
- *For the neophyte*: Focus on foundations (§2-7) and reward yourself with a deeper understanding of the Bloch sphere (§8).
- *For the iconoclast*: Read only marginalia and infer the rest.

Where can I learn more? For those eager to see the programming framework in action, visit `torsor.io/#community` to see some basic quantum algorithms implemented in pseudocode.

When will the real thing be ready? Hopefully Winter 2025! Visit `torsor.io` to stay up to date.



ACKNOWLEDGMENTS

Thanks to those who have helped me refine my ideas over the course of the project, including Shovon Biswas, Leon Di Stefano, Stepan Fomichev, Achim Kempf, Filippo Miatto, and Petar Simidzija. I'm especially grateful to Jon Male, Martine Wakeham, and Clara Weill, whose emotional and financial support made this project possible.

COLOPHON

This document is typeset using the Tufte-LaTeX document class, with *Palatino* as the body font, `IBM Plex Mono` for teletype, and AMS Euler for math. The primary visual inspirations were the Life Nature Library, the books of Edward Tufte, and pre-LaTeX mathematics textbooks. Illustrations were created with a combination of Midjourney and Inkscape. Finally,

Introduction

The digital age was born in a 1937 master's thesis by Claude Elwood Shannon,[2] in which he showed that the physical operations of placing switches in parallel and series could be mapped onto the Boolean operations of AND and OR. This makes the physics of circuits isomorphic to Boolean algebra, with "conducting" corresponding to 1 and "non-conducting" corresponding to 0. Shannon was interested in circuit design; conversely, each time we fiddle with switches and zap them with electricity, we are doing an experiment. Algebraic manipulation lives in the realm of *proof*, called "syntax" by logicians. Physical manipulation belongs to the realm of *truth*, also called "semantics". They are two halves of a single logical coin.

[2] *A Symbolic Analysis of Relay and Switching Circuits.*

Quantum computing was born in the 1981 *Physics and Computation* keynote address by Richard Feynman.[3] There are parallels to Shannon's work, but Feynman was motivated by the distinct philosophy of *reversible computing*, initiated by Landauer, Bennett, Fredkin and Toffoli, among others.[4] They, in turn, were inspired by physics, noting that Nature was not merely lazy but frugal, a habit that perhaps could be adapted to computing. Feynman hung out with Fredkin long enough for some of these ideas to rub off, and his 1981 keynote—delivered to a room full of reversibility enthusiasts—clearly framed quantum as a generalization of reversible computing.

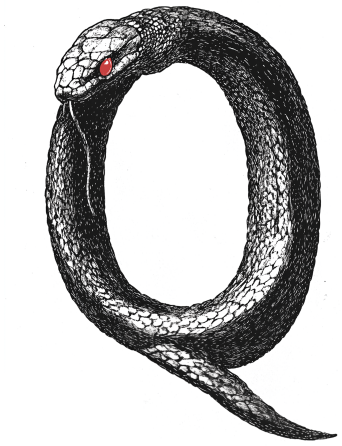
[3] "Simulating physics with computers" (1982). Yuri Manin book, *Computable and Uncomputable* (1980), slightly predates Feynman and contains similar ideas.

[4] See for instance "Irreversibility and Heat Generation in the Computing Process" (1961), Rolf Landauer; "Logical Reversibility of Computation" (1973), Charles Bennett; "Conservative Logic" (1982), Edward Fredkin and Tomaso Toffoli.

In the same way classical information is built up from binary digits (bits), Feynman proposed that quantum state should be built up from a superposition of bits:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \in \mathbb{C}^2, \quad |\alpha|^2 + |\beta|^2 = 1, \tag{1}$$

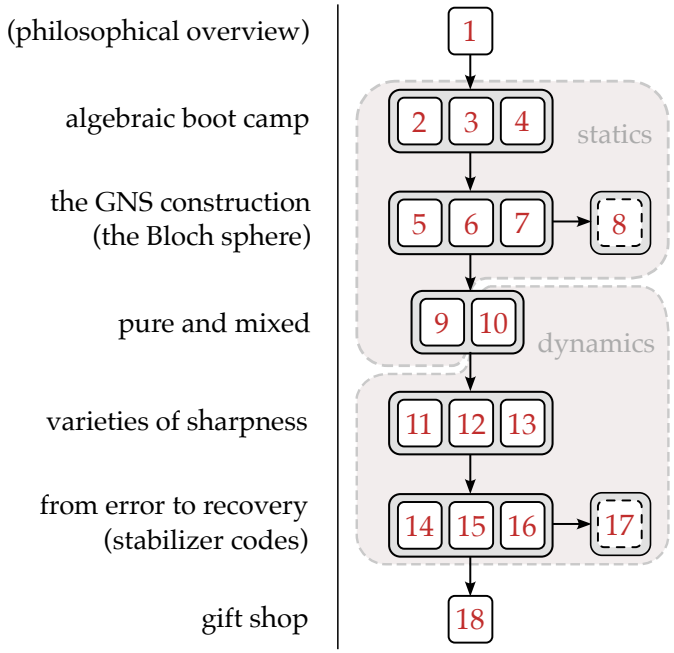
better known as the *quantum bit (qubit)*. If we want to preserve information, we should stick to *unitary* operations until performing a final measurement. This conventional approach is both inspired and realized by quantum physics. This is partly a serendipitous fit that required Feynman's genius and interdisciplinary synergy to realize.



The quantum uroburos: great band name, limited computational formalism.

At the same time, it is a snake eating its own tail. We have been thinking for 45 years in terms of states, qubits, and unitary gates. This may be a good analogy to classical reversible computing, but it is only half of Shannon's two-part harmony; we can run the electricity and do the experiments, but there is no algebra to fool around with. Hilbert space is great way to do models, or in logical terms, *semantics*, but a poor way to proofs in the same way that proof by truth table generically involves writing an exponentially long list.

This monograph proposes that, in the context of quantum computing, C^*-algebras should play the syntactic role of Boolean algebras, and Hilbert space the semantic dual. This provides a structure and interpretation for quantum programs. The paper is essentially linear:



- *Algebraic boot camp.* Defines C*-algebras (§2), translates algebraic expressions into diagrams (§3), and introduces states as a way of encoding correlations (§4).
- *The GNS construction.* Explores the correlation structure (§5), bootstraps an associated Hilbert space (§6), and identifies the familiar unit norm vectors (§7). We illustrate with the Bloch sphere (§8).
- *Pure and mixed.* Mixed states are constructed as convex combinations (§9) and pure states as maximally sharp states (§10).
- *Varieties of sharpness.* We introduce tensor products and entanglement (§11), commutative subalgebras (§12), and use this to give a Gleason-style "derivation" of Born and Lüders rules (§13).
- *From error to recovery.* Introduces quantum operations and channels (§14), the Knill-Laflamme conditions for algebraic error correction (§15), and a general construction of error correcting codes (§16). We concretely demonstrate with the five-qubit code (§17).
- *Gift shop.* Finally, §18 is a standalone showcase that can be consulted for motivation, before, after, or *in medias res*.

We can loosely split the material into "statics" (what stuff is) and "dynamics" (what stuff does).

A note on prerequisites. We assume no exposure to functional analysis, but we do take readers to have first (and maybe second) course on quantum computing. Chapters 1, 2 and perhaps 10 of Mike and Ike[5] is more than sufficient. We also assume some "mathematical maturity" which, for lack of a better definition, is the ability to take definitions on faith until the evidence arrives.

[5] *Quantum Computation and Quantum Information* (2000), Michael Nielsen and Isaac Chuang.

1. Ways of computing

In quantum computing, all roads lead back to John von Neumann. Von Neumann started his career as a graduate student with David Hilbert in Göttingen, where he co-invented Hilbert space.[6] As a young mathematician, he pioneered the study of *operator algebras*,[7] and after "losing faith" in the Hilbert space formulation of quantum mechanics,[8] turned to operator algebras as an alternative foundation. Von Neumann was particularly interested in the logic of *projectors* $\Pi : \mathcal{H} \to \mathcal{H}$, operators on a Hilbert space $\mathcal{H}$ satisfying

$$\Pi^2 = \Pi, \quad \Pi^\dagger = \Pi. \tag{2}$$

Physically, these correspond to "atomic" yes/no measurements; mathematically, they closely parallel propositional variables in a Boolean algebra, which satisfy an *idempotence* condition $p^2 = p$.

Von Neumann pursued this hint rather literally, erecting a whole theory of "quantum logic" around it.[9] Although it provides some hints, quantum logic is too restrictive to compute with; for instance, it has no notion of conditional, which makes programming difficult! Von Neumann was drawn away from foundational questions by wartime work on the Manhattan Project, where he crossed paths with Feynman. After the war, he devoted himself energetically to practical applications like game theory, nuclear strategy and large-scale simulation.[10] Things could have been different (as envisioned in the companion piece, *A Short History of Rocks*), but the task of making quantum compute would be left to his bongo-playing colleague from Los Alamos.

Reversible computing—either classical or quantum—is based on the schema of encoding input data into a state, processing it with a circuit, then reading output from state once more. Computation is reversible if there is a circuit which always produces the initial from the final state. We picture this as a flowchart in Fig. 1.

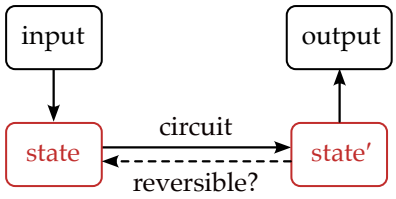


Figure 1: The basic schema of reversible computing separates circuit and state.

By focusing on conservation or non-conservation of information, we implicitly make the courier of that information—state–the key object.

State is classically unproblematic. For a single wire, the logical state is identified with whether current is flowing (1) or not (0), a question we can answer with an ammeter. For a collection of wires,

[6] "Uber die Grundlagen der Quantenmechanik" (1927), David Hilbert and John von Neumann.

[7] "On Rings of Operators I/II" (1936/7), Murray and von Neumann.

[8] "Why John von Neumann did not Like the Hilbert Space Formalism of Quantum Mechanics (and What he Liked Instead)" (1996), Miklós Rédei.

[9] "The Logic of Quantum Mechanics" (1936), Garrett Birkhoff and John von Neumann. This gives a fragment of linear logic, see "Linear Logic for Generalized Quantum Mechanics" (1992), Vaughan Pratt.

[10] "Numerical integration of the barotropic vorticity equation" (1950), Charney, Fjörtoft and von Neumann.

we use a "cross-section" of ammeters to obtain a list of Booleans. From combinatorics, reversibility means we cannot allow wires to split ("fan out") or join ("fan in"). Otherwise, there isn't enough space on the small side to accomodate the possibilities of the large.

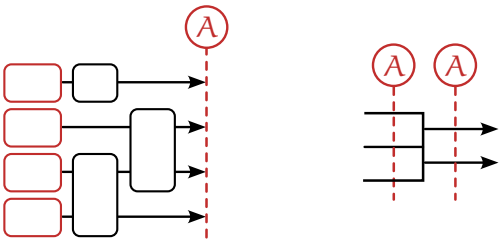


Figure 2: LEFT. Measuring state with ammeters. RIGHT. If we branch or join, state on one side becomes irrecoverable from the other.

The quantum case is superficially similar. All fundamental laws are reversible, so it seems reasonable to expect the circuits in a quantum computer to be. We replace wires with qubits and ammeters with measurement of the Pauli Z, which yields outcomes $\lambda \pm 1$.

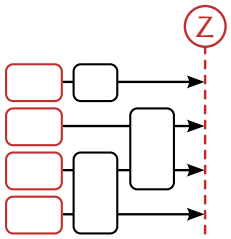


Figure 3: Measuring quantum state with the Pauli Z gives binary outcomes for each qubit.

There is a loose combinatorial intuition that fan outs and fan ins should also be forbidden, an intuition made precise by the NO CLONING THEOREM.[11] We seem to have an excellent analogy, then, between reversible computing in classical and quantum cases; indeed, this analogy has been guiding the field since its inception.

But under careful observation, the analogy breaks down. First of all, we never see the quantum state itself, only the measurement outcomes. A single measurement typically tells us *nothing* about the state. Next, state itself is ill-defined because of phase ambiguity, i.e. the fact that $|\psi\rangle \sim e^{i\theta}|\psi\rangle$ are equivalent. This is what led to von Neumann's crisis of faith in Hilbert space! Finally, measurement *changes* the state; we know what it is now, but not before! Measurement becomes part of the computation. See Fig. 4.

[11] "The concept of transition in quantum mechanics" (1970), James Park; "A single quantum cannot be cloned" (1982), Wootters and Zurek.

If we measure $Z = +1$, all we learn is that $\alpha \neq 0$ in (1). This is generically uninformative unless we do something clever, which to be fair, is the whole point of quantum algorithm design.

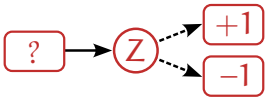


Figure 4: Measuring quantum state is a computational act.

These subtleties are of course familiar to anyone who has taken an undergraduate course in modern physics. But familiar or not, they make the primitive abstractions of Fig. 1 poorly suited to quantum

computing. This would be a quibble if no other approach existed. As it turns out, however, we can repurpose Shannon's classical insights for the quantum realm.

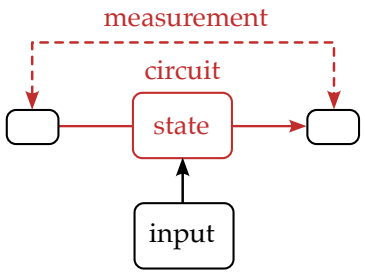


Figure 5: In a switching circuit, state and circuit are combined into a single object and jointly measured.

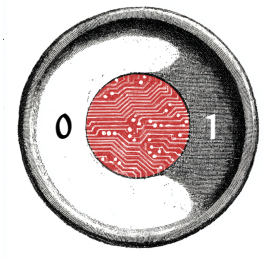
Shannon used a model of computation called a *switching circuit*, shown in Fig. 5. This is an electrical relay which encodes state into a configuration of *switches* rather than a pattern of current. This incorporates state into the circuit itself, and we test if the combined structure is connected. Physically, we can use a voltmeter and a small test current; unlike the ammeter, which interferes with the circuit, the voltmeter (in the limit of zero current) has no effect.

Mathematically, the idea of using voltmeters rather than ammeters suggests that measurement should be viewed as a *torsor*.

The measurement across the whole circuit is determined by measurement of individual switches. For instance, consider the formula $\mathsf{OR}(x,y) = x + y - xy$. The switches are x and y; if we leave y open and close x, for instance, we get a Boolean *valuation* $v(x) = 1, v(y) = 0$. The overall value is

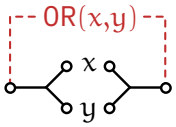
$$v[\mathsf{OR}(x,y)] = v(x) + v(y) - v(x)v(y) = 1.$$

The math faithfully reflects our intuition that the circuit is closed.



Circuits and truth tables: two sides of the same coin.

Let's repackage these ideas using *syntax* and *semantics*. A circuit with indeterminate switches is represented by an expression like $\mathsf{OR}(x,y)$. This is purely syntactic, and can be manipulated symbolically using Boolean algebra. States, on the other hand, live in the semantic realm. A state assigns a Boolean value in $\mathbb{B} = \{0,1\}$ to each propositional variable, or equivalently, flips each switch on or off. Measurement across the whole circuit is determined compositionally from measurements of individual switches, as in Fig. 6:



x	y	OR(x,y)
0	0	0
0	1	1
1	0	1
1	1	1

Figure 6: A truth table lists all assignment of Boolean values to switches (valuations), and the induced value of the circuit.

The full list of states, and correponding status of the circuit, is called a *truth table*. Our goal now is to find an analogous bipartite structure for quantum computing, with bonus points if measurement uses a voltmeter rather than an ammeter.

To keep things concrete, we'll develop a simple example in parallel with the general case, and a particularly revealing choice (both for its familiarity and unexpected depth) is the qubit. Recall that a qubit state (1) lives in a two-dimensional Hilbert space $\mathcal{H} = \mathbb{C}^2$. The Pauli Z we have been measuring with can either be viewed as living in an abstract space of *bounded linear operators* $\mathcal{B}(\mathcal{H})$, or concretely, the set $M_2(\mathbb{C})$ of 2×2 complex matrices, with an expression

"Bounded" means the image of any state has bounded length. This is not a problem in finite dimensions, but in infinite dimensions, is equivalent to continuity.

$$Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Below, we'll explain how to think about Z (and its siblings) without reference to matrices, qubits, or Hilbert space at all.

2. *A trip to Disneyland*

In 1947, a century after George Boole[12] published his groundbreaking pamphlet on the algebra of logic, Irving Segal performed an equivalent feat for physics.[13] Before the war, Segal had worked with von Neumann and Einstein at the IAS, where, coincidentally, he was a postdoc at the same time as Shannon. The war commandeered his brain for ballistics research, but in 1946, he spent the summer at Princeton and returned to thinking about his favourite problem: the mathematical foundations of quantum mechanics.

[12] *The Mathematical Analysis of Logic*.

[13] "Irreducible representations of operator algebras"; "On the embedding of normed rings into the ring of operators in Hilbert space" (1943), Israel Gelfand and Mark Naimark.

Segal was particularly concerned about particle physics, where theoreticians like Feynman were playing three cup monte with infinity. The existing Hilbert space methods were unwieldy and ill-defined, permitting decidedly unphysical operations; the binary projectors of quantum logic, on the other hand, were too tightly constrained. Segal needed something in between. He found it in the work of two Russian mathematicians, Gelfand and Naimark, who had studied certain rings of operators. Segal realized that their structure—which he baptized a *C^*-algebra*—was perfect for modelling the behaviour of quantum-mechanical observables and measurements.[14] The rest of this section gently introduces the math; readers may skip it and return later as needed.

Wave of hand to a physicist is sleight of hand to a mathematician.

[14] "Irreducible representations of operator algebras" (1947), Irving Segal.

A *C^*-algebra* $\mathcal{A}$ is a set of operators which wears many hats. First of all, it is a vector space over $\mathbb{C}$, so closed under *linear combinations*:

$$\alpha A + \beta B \in \mathcal{A} \quad \text{for} \quad A, B \in \mathcal{A} \quad \text{and} \quad \alpha, \beta \in \mathbb{C}.$$

This encodes the linearity of quantum mechanics. Unlike states, we can *compose* operators, that is, apply one after the other. We capture this with a product operation $\cdot : \mathcal{A} \times \mathcal{A} \to \mathcal{A}$, and borrow the following behaviours from operator composition:

- *associativity*: $A \cdot (B \cdot C) = (A \cdot B) \cdot C$;
- *distributivity*: $A \cdot (B + C) = A \cdot B + A \cdot C$;
- *scalar commutativity,*: $\lambda(A \cdot B) = (\lambda A) \cdot B = A \cdot (\lambda B)$;

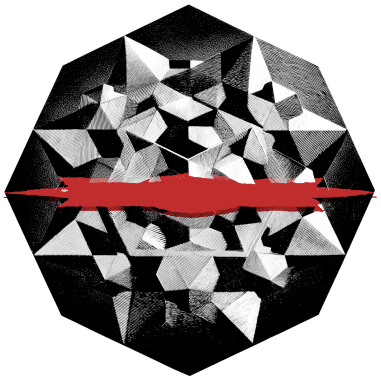
for all $A, B, C \in \mathcal{A}$ and $\lambda \in \mathbb{C}$. We can summarize everything we've listed so far by saying $\mathcal{A}$ is an *associative algebra* over $\mathbb{C}$. There are two further optional criteria: that are important:

- *commutativity*: $A \cdot B = B \cdot A$;
- *unitality*: there exists $I \in \mathcal{A}$ such that $A \cdot I = I \cdot A = A$.

In the first case, we say $\mathcal{A}$ is *commutative*, and otherwise *noncommutative*; in the second that it is *unital*. We stick to unital algebras for technical simplicity, though most of our conclusions hold regardless.

We often use complex numbers to model situations where a real answer is needed; after some analytic magic, we return to the real line by imposing $\bar{z} = z$. Similarly, the linear structure of operators over $\mathbb{C}$ is unphysical, but needed for "analytic magic"; we project back to reality with $A^\dagger = A$. Instead of a concrete Hermitian conjugate †, we use an abstract *adjoint* $* : \mathcal{A} \to \mathcal{A}$, and collect self-adjoint elements into $\mathcal{A}_{\text{sa}}$. The adjoint has the following properties:

Paul Halmos reputedly called complex analysis "the Disneyland of mathematics". C*-algebras are the Disneyland of functional analysis!



Disneyland $\mathcal{A}$ is a fun place to visit, but you don't live there. Reality is the self-adjoint cross-section $\mathcal{A}_{sa}$, in red.

- *involution*: $(A^*)^* = A$;
- *antilinearity*: $(\alpha A + \beta B)^* = \overline{\alpha} A^* + \overline{\beta} B^*$;
- *antimultiplicativity*: $(AB)^* = B^* A^*$;

for all $A, B \in \mathcal{A}$ and $\alpha, \beta \in \mathbb{C}$. This makes our associative algebra into a $*$-algebra. So, we now have everything but the "C" in "C*"!

This comes from the notion of *size*. For matrices, there are many ways to measure size, but perhaps the most natural is the *operator norm*, which is the maximum amount by which it stretches vectors. For matrices acting on a Hilbert space $\mathcal{H}$, we can define

$$\|M\|_{\text{op}} = \inf\{c \geq 0 : \|Mv\| \leq c\|v\|, \forall v \in \mathcal{H}\}.$$

This means that, by definition, $\|Mv\| \leq \|M\|_{\text{op}}\|v\|$. It's not obvious how to port this over to a *$*$-algebra*, without vectors to act on. But for a *diagonalizable* operator, it's simple to show that the operator norm is just the largest (absolute) eigenvalue:

In the diagonal basis, each direction gets scaled by at most $|\lambda_{\text{max}}|$.

$$\|M\|_{\text{op}} = |\lambda_{\text{max}}| = \sup\{|\lambda| : Mv = \lambda v, \exists v \in \mathcal{H} - \{0\}\}.$$

By the spectral theorem, *normal matrices* satisfying $N^\dagger N = NN^\dagger$ are diagonalizable, and in particular, $N = M^\dagger M = |M|^2$ is normal for any M. Some caution is needed here; if a matrix is *not* diagonalizable, the operator norm is not always largest eigenvalue.

For instance, $\left[\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right]$ has eigenvalue $\lambda = 1$, but stretches $(1,1)^\intercal$ by $c \approx 1.6$.

The next step is to define eigenvalues without eigenvectors. This sounds tricky, but luckily, there is a tight connection between eigenvalues and *invertibility* of a matrix:

$$Mv = \lambda v \quad \Longleftrightarrow \quad M - \lambda I \text{ is not invertible.} \tag{3}$$

In algebraic terms, a matrix M is not invertible just in case there is not N such that $MN = NM = I$. So far, we can only port the definition of operator norm to normal operators $A \in \mathcal{A}$. To extend this to "abnormal" operators, we take inspiration from the identities $|z|^2 = |\bar{z}z|$ in complex analysis and $\|M\|_{op}^2 = \|M^\dagger M\|_{op}$ in Hilbert space. Since A^*A is normal, we can use identify the maximum eigenvalue as its operator norm, and then take the square root to obtain the norm of A. Thus, we *define* a C*-algebra as one in which

$$\|A\|^2 = \|A^*A\| = \sup\{|\lambda| : A^*A - \lambda I \text{ is not invertible}\}. \tag{4}$$

This is called the *C* identity*. Remarkably, when this norm exists, it is uniquely determined by the algebraic structure. Pure Disney!

To actually construct C*-algebras, the method of *presentations* will prove extremely useful. The basic idea is to pick a set of unknowns $\mathcal{X}$ (akin to propositional variables $x, y, \ldots$) which we call *generators*. We subject these to identities called *relations* or *identities* $\mathcal{I}$, arranged so that each element $R \in \mathcal{I}$ is equated to zero. For instance, the standard canonical commutation relation (with $\hbar = 1$) is

$$\mathcal{X} = \{X, P\}, \quad \mathcal{I} = \{XP - PX - iI\}. \tag{5}$$

When this algebra exists, we call it the *universal C*-algebra* $C^*\langle\mathcal{X}|\mathcal{I}\rangle$.

Existence is actually nontrivial; for instance, (5) cannot be realized by operators with bounded eigenvalues, and hence never exist in a C*-algebra. We'll see Weyl's workaround for this case in §18 and Appendix A, but for now, we note two sufficient conditions:[15]

- $\mathcal{I}$ is a set of polynomials with no constant term; or
- $\mathcal{I}$ truncates the free $*$-algebra $\mathbb{C}^*[\mathcal{X}]$ to finite dimensions.

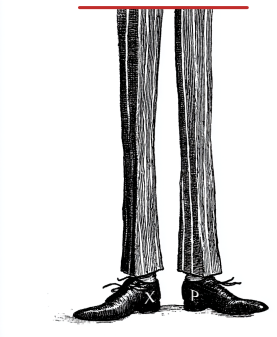
By *free $*$-algebra* $\mathbb{C}^*[\mathcal{X}]$, we mean all finite $\mathbb{C}$-linear combinations of monomials formed from $\mathcal{X}$ and adjoints.

Let's illustrate with our concrete example. Instead of thinking of Z as a matrix, we can treat it as an unknown required to be (a) self-adjoint, $Z = Z^*$, and (b) unitary $Z^2 = I$. Unitarity has a constant term, but luckily, truncates the polynomials $\mathbb{C}^*[Z]$ from arbitrary to linear degree, since higher powers of Z can be reduced. We end up with

$$\mathcal{A}_Z = C^*\langle Z|Z - Z^*, Z^2 - I\rangle = \{\alpha I + \beta Z : \alpha, \beta \in \mathbb{C}\}. \tag{6}$$

To find the norm of Z, note that $Z^2 - I = (Z - I)(Z + I) = 0$ implies eigenvalues ± 1, and hence a norm of $|\pm 1| = 1$.

The "C" stands for "closed", since Segal showed it is *topologically* closed, i.e. there are no missing limit points as measured by $\|\cdot\|$.



In Disneyland, you must be at most *this tall to ride. No canonical commutation relations allowed!*

[15] See "Quantenmechanik und Gruppentheorie" (1927), Hermann Weyl; "C*-algebra relations" (2010), Terry Loring. We also assume $\mathcal{X}$ and $\mathcal{I}$ are finite, and $\mathcal{I}$ is consistent. These conditions, along with either condition on the left, ensure that the quotient $\mathbb{C}^*[\mathcal{X}]/\mathbb{C}^*[\mathcal{I}]$ has well-defined norm satisfying the C* identity (4).

In terms of matrices, we get

$$\begin{bmatrix} \alpha+\beta & \\ & \alpha-\beta \end{bmatrix},$$

which spans *diagonal* elements of M_2.

3. *From prop to op*

Shannon translated Boolean algebra into ordinary electric circuits. We can try to translate C*-algebras into "noncommutative" circuits. We start by recalling Shannon's isomorphism:

- a Boolean variable x is physically encoded by a switch;
- switches x and y in series corresponds to the product xy (AND);
- switches x and y in parallel corresponds to the sum $x \dot{+} y$ (OR);
- scalars (in circles) multiply the adjacent proposition.

Here, $x \dot{+} y = x + y - xy$ isn't quite addition but behaves enough like it for our purposes.

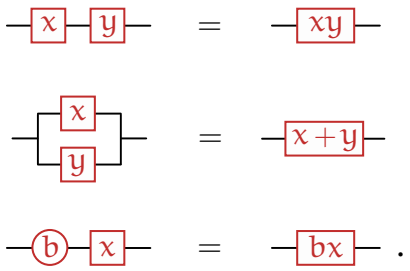


Figure 7: In a Boolean circuit, series corresponds to multiplication, parallel to addition, and circles to scalar multiplication.

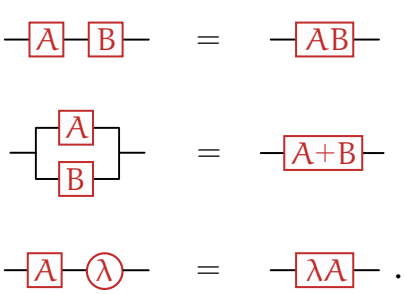
We enclose formulas in squares, Booleans in circles. We don't usually think about scalar multiplication, but it makes the analogy to the noncommutative case nicer. Inserting a Boolean $b \in \mathbb{B}$ either does nothing ($b = 1$) or snips the circuit where inserted ($b = 0$).

An *abstract wiring diagram (awd)* encodes operators of a C*-algebra in an almost identical fashion:

Pronounced "ord". We use "wiring" to differentiate from circuit diagrams (reversible quantum computing) and string diagrams (category theory).

- an operator variable X is physically encoded by an observable;
- expressions in series are multiplied (right to left);
- expressions in parallel are summed;
- scalars multiply the adjacent operator.

This means XY is the same order as the diagram!



Despite the manifest parallels, there are some differences. Boxes don't necessarily commute, scalars are complex, and we read right to left. But most importantly, the word "abstract" signals that, unlike a Boolean circuit, we do not have physical wires connecting switches. The links between them are mathematical.

What kind of spaghetti are we cooking?

The simplest interpretation—that we measure an observable at each box—isn't consistent with the algebra. A simple example is provided by the Stern-Gerlach experiment.[16] An electron approaches a magnetic field in the **z** direction which effectively measures the orientation of its spin; the Pauli Z is a idealized version of this. But although $Z^2 = I$, measuring twice is not the same as doing nothing! The Z deflection, though initially random, is fixed under subsequent Z measurements.

[16] "The experimental proof of directional quantization in the magnetic field" (1922), Walther Gerlach and Otto Stern. We will take this behaviour as an experimental given for now, and discuss how to reflect it in the formalism later.

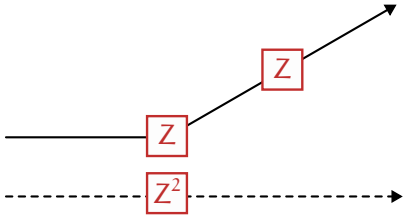


Figure 8: On the solid path, an evenly polarized spin is measured and deflects; the subsequent Z measurement is consistent. On the dotted path, the spin is not measured and hence is undeflected.

If awds are not physical layouts, what are they? At the very least, they are *symbolic models of computation*. They let us manipulate observables algebraically. As a simple illustration of how this can be used, we'll enlarge our Pauli Z example to include the Pauli X and Y. As explicit matrices,

$$\sigma_{(1)} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_{(2)} = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_{(3)} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{7}$$

Along with the identity operator I, these form a basis for the full set of single-qubit operators. We can characterize them without matrices using the *Pauli relations*:

It's straightforward to show they are linearly independent, and since there are four, they form a basis for $M_2(\mathbb{C})$.

$$\sigma_{(i)}\sigma_{(j)} = \delta_{ij} I + i\epsilon_{ijk}\sigma_{(k)}, \tag{8}$$

where δ_{ij} is the Kronecker delta and $\epsilon_{ijk} = \mathrm{Sgn}(ijk)$ the totally antisymmetric tensor. For relations $R_{(ij)}$ encoding (8), the *Pauli algebra* has presentation $\mathcal{A}_{\text{Pauli}} = \mathbb{C}^*\langle \sigma_{(i)} | R_{(jk)} \rangle$, where $i, j, k \in \mathfrak{I} = \{1, 2, 3\}$.

We also use *Einstein convention* of summing over repeated dummy indices. In a non-linear context, it is interpreted with smallest possible scope.

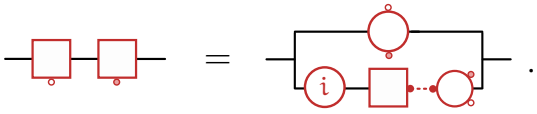
Swapping two indices produces a factor of -1, and $\epsilon_{123} = +1$. This implies cyclic symmetry, $\epsilon_{ijk} = \epsilon_{jki}$.

When objects are indexed by a set $\mathfrak{I}$, we decorate their containers with *index nodes*; usually $\mathfrak{I}$ is clear from context. In our Pauli example, for instance, $\mathfrak{I} = \{1, 2, 3\}$. A square with a single node denotes $\sigma_{(i)}$, a circle with two nodes δ_{ij}, and a circle with three nodes ϵ_{ijk}:

Figure 9: Special objects in the Pauli algebra, decorated by index nodes.

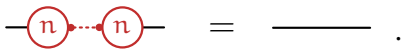
Colour and shading act as dummy indices for index nodes, and in the rare situation with multiple index sets at play, distinct indices get distinct node shapes. Containers without glyphs (as above) refer to special families like $\sigma_{(i)}$ or δ_{ij}. To specify other objects, or particular index values, we label explicitly.

The last operation we introduce for now is a *tether*, which pairs two indices and sums over them. In awds, we simply join two nodes of the same colour and shape with a dotted line. For instance, the Pauli relations (8) can be diagrammatically recast as



This trades complex of products for complexity of sums.

To illustrate, we'll show how to point a Pauli in any direction. Let $\mathbf{n} = (n_1, n_2, n_3)$ be a real vector of unit length, so $n_i n_i = 1$:



We claim the *generalized Pauli operator* $\sigma(\mathbf{n}) = n_i \sigma_{(i)}$ is self-adjoint and unitary. Self-adjointness is easy, so it remains to show that $\sigma(\mathbf{n})^2 = I$. We can do this diagrammatically, indicating $\sigma(\mathbf{n})$ by a box labelled with $\mathbf{n}$. The proof goes as follows:

We have a real linear combination of self-adjoint operators.

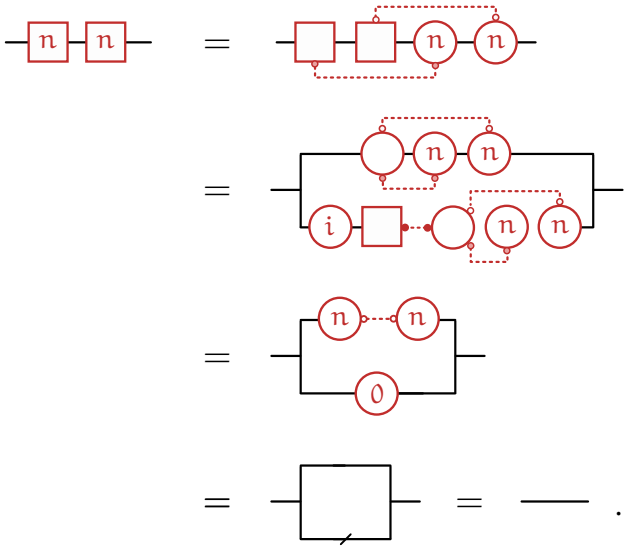


Figure 10: Proof that $\sigma(\mathbf{n})^2 = I$. A small modification shows that

$$\sigma(\mathbf{m})\sigma(\mathbf{n}) = (\mathbf{m}\cdot\mathbf{n})I + i\sigma(\mathbf{m}\times\mathbf{n}),$$

where $\mathbf{n}\cdot\mathbf{m} = n_i m_i$ is the dot and $(\mathbf{n}\times\mathbf{m})_i = \epsilon_{ijk} n_j m_k$ the cross product.

The role of the δ_{ij} (line 2) is to connect incoming tether lines, while ϵ_{ijk} vanishes (line 3) when it detects repetition. Finally, as in the Boolean setting, inserting 0 "snips" the circuit, leaving the identity.

Details for ϵ_{ijk}: $n_i n_j \delta_{ij} = n_k n_k$, and $n_i n_j \epsilon_{ijk} = -n_i n_j \epsilon_{jik} = -n_i n_j \epsilon_{ijk}$, where we used antisymmetry, relabelled $i \leftrightarrow j$ and commuted n_i, n_j. It follows that $n_i n_j \epsilon_{ijk} = 0$.

This shows that awds can be useful for syntactic manipulation. But what does it all *mean*? Like truth tables in the Boolean case, we will introduce a semantics of quantum measurement, with switches we can toggle and compose. But this apparently simple goal will take us deeper into the realms of algebra and functional analysis than might have been expected (or hoped); for better and worse, we feel the complications of living in a noncommutative world.

Whoop-de-doo Basil.

4. States as functionals

In a classical circuit, a *state* flips each switch into the on or off position and sets the corresponding value of a conductance measurement. Let $\mathcal{P} = \{x_1, x_2, \ldots, x_n\}$ denote a set of n propositional switches, $\mathbb{B}[\mathcal{P}]$ the set of all circuits built using them. We can define a state mathematically as an assignment $\nu_0 : \mathcal{P} \to \mathbb{B}$, and the *truth table* as the set of all such maps, which we can view, suggestively, as a binary vector space $\mathbb{B}^{\mathcal{P}}$. States can be uniquely extended from switches to any circuit via

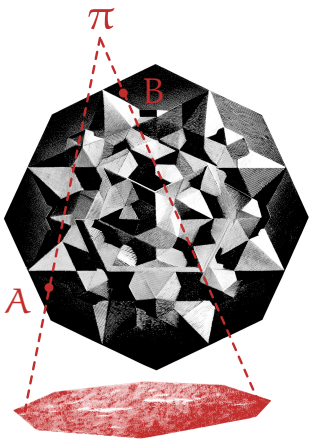
$$\nu(b\eta + b'\zeta) = b\nu(\eta) + b'\nu(\zeta), \quad \nu(\eta\zeta) = \nu(\eta)\nu(\zeta), \quad \nu|_{\mathcal{X}} = \nu_0 \tag{9}$$

for $b, b' \in \mathbb{B}$ and $\eta, \zeta \in \mathbb{B}[\mathcal{P}]$. In words, ν must be *linear* over $\mathbb{B}$, *multiplicative*, and agree on switches. This ensures that each state in $\mathbb{B}^{\mathcal{P}}$ uniquely determines measurements of circuits.

The notation C^D, for sets C and D, means the set of all functions $f : D \to C$.

Analogously, we want our noncommutative states to consistently assign measurements to each operator $A \in \mathcal{A}$. Our job will be to figure out what "consistent", "assign" and "measurement" mean! We take as an axiom that measurement of A yields a random element of the set of eigenvalues, or *spectrum* of A:

$$\mathfrak{S}(A) = \{\lambda \in \mathbb{C} : (A - \lambda I)^{-1} \text{ does not exist}\}. \tag{10}$$



A state projects the full algebra onto a consistent set of expectations (red).

For a quantum state π, we let $A|_\pi$ denote the random variable corresponding to an operator A.

Random variables are a little unwieldy to work with directly; instead, we'll use the averages of A_π, and think of a state as an *expectation functional* $\pi : \mathcal{A} \to \mathbb{C}$ giving

$$\pi(A) = \mathbb{E}\left[A|_\pi\right]. \tag{11}$$

A functional is any map from a vector space to its underlying scalars, e.g. a Boolean valuation $\nu : \mathbb{B}[\mathcal{P}] \to \mathbb{B}$.

We picture this as an dotted circle around a circuit:

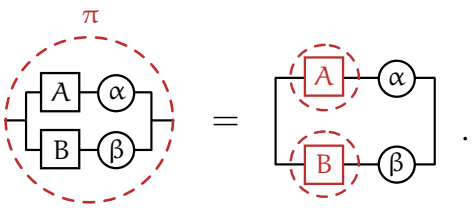
The circle reminds us that the end result is a scalar, while the dots tell us it takes an operator as a argument.

This isn't a precise definition, since we haven't explained how to determine the random variable $A|_\pi$; instead, it is an "intuition pump" we can use to motivate properties we want π to have. First of all, expectation is linear, so

$$\pi(\alpha A + \beta B) = \alpha\pi(A) + \beta\pi(B). \tag{12}$$

This will be consistent with our later notation for a *channel*, which takes an operator to an operator. See Fig. 17.

We represent by extending our awd notation:



The sum is still indicated by the parallel structure, but we remove the trailing wires to indicate this is no longer an operator.

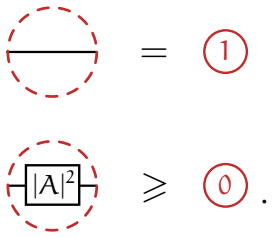
Since the identity I has a unique eigenvalue 1, we measure it with certainty. Thus, the fact that probabilities are normalized requires

$$\pi(I) = 1. \tag{14}$$

Finally, the average over a set of positive values should be positive. A positive operator is one with nonnegative eigenvalues, or equivalently, of the form $B = A^*A = |A|^2$, often written $B \geq 0$. Since the associated measurements are nonnegative, so is the expectation:

$$\pi(|A|^2) \geq 0. \tag{15}$$

Again, we depict normalization and positivity as follows:



Nonpositive operators can have arbitrary complex expectations.

In axiomatic approaches[17] to expectation, the properties (12)–(15) are not only reasonable but complete; we can use them to define which functions behave like expected values of random variables. This suggests that we can define a state on a C*-algebra $\mathcal{A}$ as any map $\pi : \mathcal{A} \to \mathbb{C}$ satisfying (12)–(15), or in words, a *positive linear functional of unit norm*. See Fig. 11 for a cartoon.

Equivalently, we say π has *unit norm* since the operator norm on linear functionals is given by

$$\|\pi\|_* = \sup_{\|A\| \leq 1} \pi(A). \tag{13}$$

This obeys $|\pi(A)| \leq \|\pi\|_* \|A\|$, and hence for a unital algebra, $\|\pi\|_* = \pi(I)$.

[17] See for instance *Foundations of the theory of probability* (1933), Andrei Kolmogorov. These follow from the probability axioms and the measure-theoretic definition of expectation.

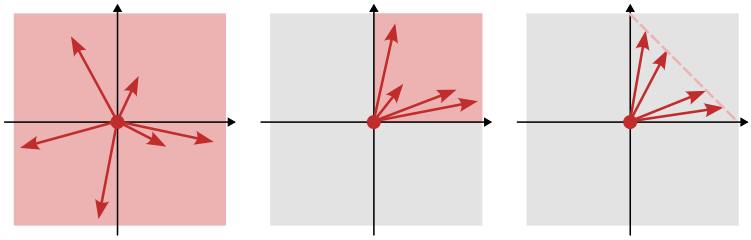


Figure 11: A cartoon of real linear functionals. Left. Linear functionals as vectors in the plane. Middle. Positive functionals, where each component ≥ 0. Right. Unit norm positive functionals.

We call the full collection of states $S(\mathcal{A})$. Note that, unlike Boolean valuations, where we demand $\nu(xy) = \nu(x)\nu(y)$, we say *nothing* about products in the noncommutative case. What gives?

The basic problem is that it is not clear how to define the value of π on some set of "basic switches" and extend it globally. This happens for a variety of reasons. First, because measurement of a product does not equal a product of measurements, the random variable structure is not multiplicative:

$$AB|_\pi \neq A|_\pi \cdot B|_\pi. \tag{16}$$

Even if we had equality in (16), the expectations wouldn't factorize unless $A|_\pi$ and $B|_\pi$ were uncorrelated as random variables. Finally, we might hope for some clever way to compose or relate the two sides of (16), short of equality. Unfortunately, there is *no simple relationship* between the spectra, i.e. the outcomes $\mathfrak{S}(AB)$ and $\mathfrak{S}(A)$, $\mathfrak{S}(B)$.

To illustrate, consider an invertible A with eigenvalues $\lambda_i > 0$, and inverse A^{-1} with eigenvalues $\lambda_i^{-1} > 0$. The product $\mathfrak{S}(AA^{-1}) = \{1\}$.

Evidently, we're thinking about this the wrong way. Progress will require us to zoom out and think about the global shape of any functional satisfying (12)–(15). By carefully studying the multiplicative patterns in such a functional, we can reverse engineer a switch.

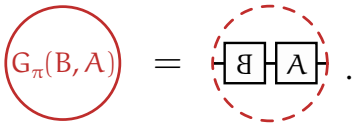
5. *Patterns of correlation*

We'll start by defining a simple object which encodes those multiplicative patterns. The *correlation* of A and B in state π is the expectation of the random variable associated with their product:

$$G_\pi(B, A) = \pi(B^* A). \tag{17}$$

We use the convention that reflecting an object in the x-axis corresponds to taking the adjoint. Thus:

If necessary, we add a dot or other marker in the corner to disambiguate.



Similarly, a reversed scalar λ is the complex conjugate λ^*.

We use B^* in definition (17) rather than B to ensure that an observable is positively correlated with itself, i.e. (15) tells us that

$$G_\pi(A, A) = \pi(|A|^2) \geq 0. \tag{18}$$

The correlation is linear in the second argument,

Note that this is consistent with the physics literature, but the opposite of the math literature. Notation is the ultimate zero sum game!

$$\begin{aligned} G_\pi(B, \alpha_1 A_1 + \alpha_1 A_1) &= \pi[B^*(\alpha_1 A_1 + \alpha_2 A_2)] \\ &= \alpha_1 \pi(B^* A_1) + \alpha_2 \pi(B^* A_2) \\ &= \alpha_1 G_\pi(B, A_1) + \alpha_2 G_\pi(B, A_2), \end{aligned} \tag{19}$$

and similarly *antilinear* in the first. These three properties—nonnegativity, linearity, and antilinearity—make correlation a "positive-semidefinite

sesquilinear form". The name is a mess, but the properties guarantee something neat: the *Cauchy-Schwarz inequality*.[18] Using metric-style notation, $G_\pi(A,A) = \|A\|_\pi^2$, we have

$$|G_\pi(B,A)|^2 \leq \|A\|_\pi^2 \|B\|_\pi^2. \tag{20}$$

The most important application of (20) is to operators Θ with *vanishing* self-correlation, $G_\pi(\Theta,\Theta) = 0$. We call these operators *null*, and the collection of such operators the *null space* or *kernel* of π, $\mathcal{K}_\pi$.

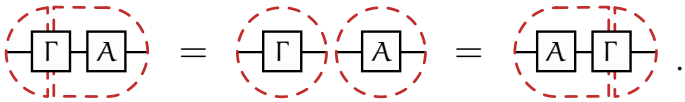
It follows that the correlation with any operator $A \in \mathcal{A}$ vanishes:

$$\Theta \in \mathcal{K}_\pi \quad \Longleftrightarrow \quad |G_\pi(\Theta,A)|^2 \leq \|A\|_\pi^2 \|\Theta\|_\pi^2 = 0. \tag{21}$$

Equivalently, if $\Delta\Gamma = \Gamma - \pi(\Gamma)I$ is null, then linearity of G_π implies that expectations factorize:

$$\pi(\Gamma A) = \pi(\Gamma)\pi(A) = \pi(A\Gamma). \tag{22}$$

We call such a Γ *sharp* or *definite*. In awds, we indicate this with a dotted channel passing through Γ, to suggest we can factorize the state contour through the operator. For instance, (22) corresponds to slicing, swapping, and sewing the operators back together:



We have to be careful, though: we cannot factorize any product along Γ, since $\pi(A\Gamma B) \neq \pi(A)\pi(\Gamma)\pi(B)$ in general. We'll discuss the general interpretation below.

Sharp operators arise from deterministic measurements. If Γ takes value $\pi(\Gamma)$ with certainty, then $\Delta\Gamma$ is null:

$$\|\Delta\Gamma\|_\pi^2 = \pi(|\Gamma|^2) - |\pi(\Gamma)|^2 = \mathrm{var}_\pi(\Gamma) = 0, \tag{23}$$

since Γ has zero variance. The *definite set* is the set of sharp operators which are also self-adjoint:

$$\mathcal{D}_\pi = (\mathbb{C}I + \mathcal{K}_\pi) \cap \mathcal{A}_{\mathrm{sa}} = \mathbb{R}I + (\mathcal{K}_\pi \cap \mathcal{K}_\pi^*), \tag{24}$$

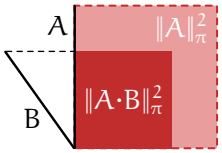
where the first equality is a definition, and the second a theorem.[19] Let's see how this plays out in the Stern-Gerlach experiment. Suppose we measure $Z \leftarrow +1$ with probability p, where $\Lambda \leftarrow \lambda$ means that measuring Λ yields outcome λ. The average is

$$\pi(Z) = (+1)p + (-1)(1-p) = 2p - 1.$$

A little algebra then shows

$$\pi(\Delta Z^2) = \pi(I) - (2p-1)^2 = 1 - (2p-1)^2 = 4p(1-p). \tag{25}$$

[18] A nice blog post: "Amplification, arbitrage, and the tensor power trick" (2007), Terry Tao. Even with nonnegativity, the proof goes through. We can also give a visual "proof":



where (without loss of generality) B is unit norm. The intuition is that projecting is always norm-decreasing.

Figure 12: Factorizing an expectation through a sharp operator.

For instance, for an eigenstate of Z and the Hadamard $H = \frac{1}{\sqrt{2}}(X+Z)$, we have $\pi(HZH) = \pi(X) = 0$, but $\pi(H)\pi(Z)\pi(H) = \frac{1}{2}$. For the correct intepretation, see (29).

[19] "Extensions of pure states" (1959), Kadison and Singer. This paper is famous for spawning the *Kadison-Singer problem*, only resolved in 2017.

Measurement has many similarities to variable assignment, so we borrow the assignment operator notation $\leftarrow$ used in pseudocode.

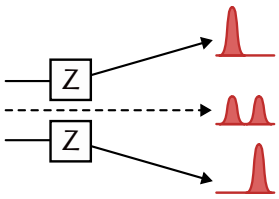


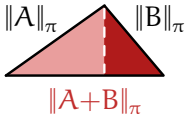
Figure 13: Cartoons of zero variance measurements when $Z \leftarrow \pm 1$ has a deterministic value.

We find that Z is definite just in case $p = 0$ or $p = 1$, i.e. we measure $Z \leftarrow \pm 1$ with certainty. This makes sense! We picture this in Fig. 13.

Returning to the main thread, how can we use these tools to understand the behaviour of measurements? The first step will be to unpack the structure of the kernel. First, note that since G_π satisfies Cauchy-Schwarz (20), it also obeys the *triangle inequality*:

$$\|A + B\|_\pi^2 \leq \|A\|_\pi^2 + \|B\|_\pi^2. \tag{26}$$

Expand $\|A + B\|_\pi^2$, upper bound $\mathrm{Re}[G_\pi(A, B)]$ with Cauchy-Schwarz, and factorize into $(\|A\|_\pi^2 + \|B\|_\pi^2)^2$. There is an even simpler visual "proof":



The sum projects A and B in a common direction, hence decreases both norms.

It follows immediately that $\mathcal{K}_\pi$ is closed under sums, and hence any linear combination. This makes it a vector subspace of $\mathcal{A}$. Similarly, $\mathcal{K}_\pi$ is closed under products. In fact, it is closed under left multiplication by *arbitrary* operators. Taking $\Theta \in \mathcal{K}_\pi, A \in \mathcal{A}$,

$$\|A\Theta\|_\pi^2 = G_\pi(A\Theta, A\Theta) = \pi[(A\Theta)^* A\Theta] = 0 \tag{27}$$

using (21). Thus, $\mathcal{K}_\pi$ is a vector subspace (closed under linear combinations), a *subalgebra* (closed under products), and a *left ideal* (closed under left-multiplication by any element of $\mathcal{A}$).

To see why this might be useful, suppose we can write a projector $\Pi_\mathcal{K}$ onto the kernel of π, with an orthogonal projector $\Pi_\mathcal{K}^\perp$. The null projector has the property that, for any $A \in \mathcal{A}$,

Consult (2) if you've forgotten what a projector is.

$$\pi(\Pi_\mathcal{K} A) = \pi(A\Pi_\mathcal{K}) = 0.$$

Notice that we can split A into four pieces using these projectors:

$$A = \Pi_\mathcal{K} A \Pi_\mathcal{K} + \Pi_\mathcal{K} A \Pi_\mathcal{K}^\perp + \Pi_\mathcal{K}^\perp A \Pi_\mathcal{K} + \Pi_\mathcal{K}^\perp A \Pi_\mathcal{K}^\perp.$$

Since anything with $\Pi_\mathcal{K}^\perp$ on either side vanishes inside π, we can write

$$\pi(A) = \pi(\Pi_\mathcal{K}^\perp A \Pi_\mathcal{K}^\perp). \tag{28}$$

More prosaically, this restricts A to the part that acts nontrivially on the state π. Incidentally, this leads to the generalization of (22): if $\Delta\Gamma$ is null, then $\Delta\Gamma = \Pi_\mathcal{K}(\Delta\Gamma)\Pi_\mathcal{K}$, and hence (with a little algebra)

$$\Gamma = \Pi_\mathcal{K} \Gamma \Pi_\mathcal{K} + \Pi_\mathcal{K}^\perp \pi(\Gamma) \Pi_\mathcal{K}^\perp. \tag{29}$$

This is what the gap in (12) instructs us to sum over!

On the outside of a product, the product of $\Pi_\mathcal{K}$ and $\Pi_\mathcal{K}^\perp$ vanishes, and we can replace Γ with its expectation. But for $\pi(A\Gamma B)$, both components of (29) survive. Shuffling around these projectors seems to involve a lot of bookkeeping! But we are in luck: we can revive Hilbert space to do our bookkeeping for us.

6. Hilbert space redux

Let's assemble what we know so far. A state $\pi \in S(\mathcal{A})$ is a positive, linear functional of unit norm, $\pi : \mathcal{A} \to \mathbb{C}$. For any state, the correlation $G_\pi(B, A)$ is *almost* an inner product, but has vanishing norm on the kernel $\mathcal{K}_\pi$. What it lacks is *positive-definiteness*:

$$\|v\| = 0 \quad \iff \quad v = 0.$$

To make it positive-definite, we need to somehow get rid of $\mathcal{K}_\pi$, or identify it with the zero operator.

Luckily, there is a mathematical procedure for identifying things with zero called a *quotient*. First, we lump together everything that differs by a null operator, called a *null equivalence class*:

$$[A]_\pi = \{B \in \mathcal{A} : A - B \in \mathcal{K}_\pi\} = A + \mathcal{K}_\pi. \tag{30}$$

An *equivalence relation* $R \subseteq S \times S$ is a relation which is (a) *reflexive*, $(s, s) \in R$ for all $s \in S$; (b) *symmetric*, $(s, t) \in R$ implies $(t, s) \in R$; and (c) *transitive*, $(s, t) \in R$ and $(t, u) \in R$ implies $(s, u) \in R$. An *equivalence class* under R is a set of objects related to each other. We write $s \equiv_R t$ for $(s, t) \in R$.

The *quotient space* $\mathcal{A}/\mathcal{K}_\pi = \{[A]_\pi\}_{A \in \mathcal{A}}$ is the set of all equivalence classes. This is a vector space, and a linear combination of classes

$$\alpha[A]_\pi + \beta[B]_\pi = [\alpha A + \beta B]_\pi, \tag{31}$$

is well-defined because $\mathcal{K}_\pi$ is itself closed under linear combinations. On this vector space, G_π "lifts" to a nondegenerate inner product:

$$\langle [B]_\pi, [A]_\pi \rangle = G_\pi(B, A) = \pi(B^* A), \tag{32}$$

with $[0]_\pi = \mathcal{K}_\pi$ playing the role of zero. The first equality follows because we can replace B with any element of its null equivalence class and get the same correlation, due to (21).

A vector space with an inner product is almost, but not quite, a Hilbert space, since it may have "holes" with respect to the metric $\|A\|_\pi$. We just used one math hack—quotients—to get rid of the pesky null operators. We can use another hack—*completion*—to fill in the holes.[20] The result is a Hilbert space

$$\mathcal{H}_\pi = \overline{\mathcal{A}/\mathcal{K}_\pi}^{\|\cdot\|_\pi}, \tag{33}$$

where $\overline{\mathcal{M}}^{\|\cdot\|}$ denotes the completion of a metric space $\mathcal{M}$ with respect to a norm $\|\cdot\|$. This *completion* is unique. Before you get too worried about all the hacks at play, you can take solace in the fact that *finite-dimensional* examples never have any holes to fill!

[20] See, e.g. *Principles of Mathematical Analysis* (1953), Walter Rudin. The basic idea is to form equivalence classes of Cauchy sequences in $\mathcal{M}$, the same way we can build $\mathbb{R}$ from sequences of approximations in $\mathbb{Q}$. This yields a space $\overline{\mathcal{M}}$ which is complete since (roughly speaking) a Cauchy sequence converges to itself!

Since $\mathcal{K}_\pi$ is a left ideal, $A\mathcal{K}_\pi \subseteq \mathcal{K}_\pi$ for any $A \in \mathcal{A}$, and hence we can left-multiply an equivalence class:

$$\begin{aligned} A \cdot [B]_\pi &= A(B + \mathcal{K}_\pi) \\ &= AB + A\mathcal{K}_\pi \\ &= AB + \mathcal{K}_\pi = [AB]_\pi. \end{aligned} \tag{34}$$

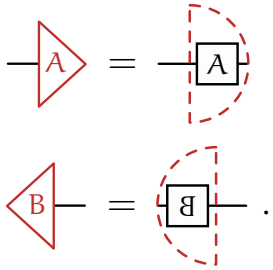
Since A is linear and maps $\mathcal{H}_\pi$ to itself, it acts as a linear operator we denote A_π (not to be confused with the random variable $A|_\pi$) on $\mathcal{H}_\pi$. It's not hard to see that it's bounded, so each $A_\pi \in \mathcal{B}(\mathcal{H}_\pi)$.

Boundedness follows from the fact that $|\pi(A)| \leq \|A\|$; see margins near (14).

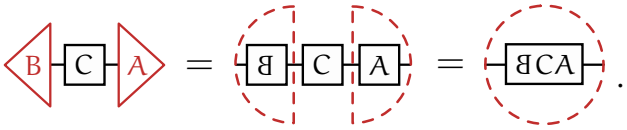
We can restore some of our Hilbert space intuition by using bra-ket notation for (not necessarily normalized) vectors $[A]_\pi$. The kets live in the second slot of the inner product, and the bras in the first:

$$|A\rangle_\pi = G_\pi(\cdot, A) = \langle [\cdot]_\pi, [A]_\pi \rangle, \quad {}_\pi\langle B| = G_\pi(B, \cdot) = \langle [B]_\pi, [\cdot]_\pi \rangle. \tag{35}$$

This makes kets linear and bras antilinear. We can capture these visually by splitting a state into two dotted semicircles:



A general matrix element is then written as follows:



This is consistent with our notation for sharp operators in Fig. 12.

The procedure we've just outlined is the Gelfand-Naimark-Segal (GNS) Construction. It takes a state π on an abstract space of operators $\mathcal{A}$, and cooks up a concrete Hilbert space $\mathcal{H}_\pi$ on which those operators act. Since each operator is bounded, we can say more precisely that it maps $\mathcal{A}$ to a subalgebra of $\mathcal{B}(\mathcal{H}_\pi)$ in a structure-preserving way called a *$*$-homomorphism*. Thus, instead of a positive, normed linear functional, a state can also be viewed as a $*$-homomorphism $\Phi_\pi : \mathcal{A} \to \mathcal{B}(\mathcal{H}_\pi)$ given by

$$\Phi_\pi(A)[B]_\pi = A_\pi[B]_\pi = [AB]_\pi. \tag{36}$$

The map Φ_π is called a *representation* of $\mathcal{A}$. We won't emphasize this perspective here, and simply write $\Phi_\pi(A) = A_\pi$.

Let's see how this works for the Pauli algebra $\mathcal{A}_{\text{Pauli}}$. Since we can anticommute and square any repeated operators to unity by (8), each element takes the form

$$A = \alpha_0 I + \alpha_1 X + \alpha_2 Y + \alpha_3 Z, \quad \alpha_0, \alpha_1, \alpha_2, \alpha_3 \in \mathbb{C}. \tag{37}$$

Thus, as a vector space $\dim(\mathcal{A}_{\text{Pauli}}) = 4$. Measuring $Z \leftarrow +1$ in the Stern-Gerlach experiment leads to a post-measurement state $\pi_{(0)}$.

In contrast to random variables (16), this map *is* multiplicative,

$$(AB)_\pi = A_\pi B_\pi,$$

which follows immediately from (34):

$$\begin{aligned}(AB)_\pi[C]_\pi &= [ABC]_\pi \\ &= A_\pi[BC]_\pi = A_\pi B_\pi[C]_\pi,\end{aligned}$$

for an arbitrary vector $[C]_\pi \in \mathcal{H}_\pi$.

You can shift Γ to either side of the channel to form a bra or ket, with

$$\langle [\Gamma^*]_\pi, [I]_\pi \rangle = \langle [I]_\pi, [\Gamma]_\pi \rangle = \pi(\Gamma)$$

telling us the results are consistent.

It preserves linear combinations, products, and adjoints:

$$\begin{aligned}\Phi_\pi(\alpha A + \beta B) &= \alpha\Phi_\pi(A) + \beta\Phi_\pi(B) \\ \Phi_\pi(AB) &= \Phi_\pi(A)\Phi_B(B) \\ \Phi_\pi(A^*) &= \Phi_\pi(A)^*.\end{aligned}$$

We choose $Z \leftarrow (-1)^b$ to correspond to a functional $\pi_{(b)}$, and later, a ket $|b\rangle$.

The kernel $\mathcal{K}_{(0)}$ contains, at a minimum, $\Delta Z = Z - I$ and all its scalar multiples. But since it is a left ideal, it also contains the results of left-multiplying by X and Y:

The Z product $Z\Delta Z = -\Delta Z$ is boring.

$$X\Delta Z = XZ - X = -iY - X = i(iX - Y) = iY\Delta Z.$$

Both products are proportional to $iX + Y$, so the kernel is

$$\mathcal{K}_{(0)} = \left\{ \tfrac{1}{2}\delta(I - Z) + \tfrac{1}{2}\gamma(X + iY) : \gamma, \delta \in \mathbb{C} \right\}. \tag{38}$$

If we express $(Z - I)/2$ and $(X + iY)/2$ as matrices in the usual basis, we find they correspond to the pink entries:

We add the factor of 2 so operators are unit norm. For instance, $Z - I$ has maximum absolute eigenvalue $|-2| = 2$.

$$\mathcal{H}_{(0)} \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix} \mathcal{K}_{(0)}.$$

Figure 14: The GNS construction for $Z \leftarrow +1$ in the Pauli algebra.

This leads us to expect that the quotient space $\mathcal{H}_{(0)} = \mathcal{A}_{\text{Pauli}}/\mathcal{K}_{(0)}$ lives in the first column, and therefore looks like a qubit!

For completeness, let's check. The generators of (38) let us equate Z with I, and Y with $-iX$. This modifies the coefficients in (37):

$$\alpha_0 \mapsto \alpha_0 + \alpha_3, \quad \alpha_1 \mapsto \alpha_1 - i\alpha_2.$$

Defining $\alpha = (\alpha_0 + \alpha_3)/2$ and $\beta = (\alpha_1 - i\alpha_2)/2$, we obtain

$$\mathcal{H}_{(0)} = \left\{ \tfrac{1}{2}\alpha(I + Z) + \tfrac{1}{2}\beta(X - iY) : \alpha, \beta \in \mathbb{C} \right\}. \tag{39}$$

In terms of matrices, this is indeed the first column of the matrix! To show this without using the matrix representation is straightforward. From (39), the computational basis states will be

These have unit norm, e.g.

$$\|X\|^2_{(0)} = G_{(0)}(X^*, X) = \pi_{(0)}(I) = 1.$$

$$|0\rangle = \frac{1}{2}[Z + I]_{(0)} = [I]_{(0)}, \quad |1\rangle = \frac{1}{2}[X - iY]_{(0)} = [X]_{(0)}. \tag{40}$$

The Pauli algebra $\mathcal{A}_{\text{Pauli}}$ acts by left-multiplication, so after a little work, one obtains $X|b\rangle = |\neg b\rangle$ and $Z|b\rangle = (-1)^b|b\rangle$. This not only looks but behaves like a qubit!

Here is the work:

$X|b\rangle = X[X^b]_{(0)} = [X^{\neg b}]_{(0)} = |\neg b\rangle$

$Z|b\rangle = Z[X^b]_{(0)} = [(-X)^b]_{(0)} = (-1)^b|b\rangle$

where $\neg b = 1 - b$, we use $ZX = iY$ and the null relation $X + iY \equiv_+ 0$.

7. Changing basis

We've shown how to construct a state π and the associated Hilbert space for which Γ is definite; we call π an *eigenfunctional* of Γ. But how much depends on the precise eigenvalue it takes? Let's repeat our Hilbert space construction for $Z \leftarrow -1$, which swaps columns:

We use *eigenvector* for elements of $\mathcal{H}_\pi$, and *eigenstate* when we want to hedge.

$$\mathcal{K}_{(1)} \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix} \mathcal{H}_{(1)}.$$

Figure 15: The GNS construction for $Z = -1$ in the Pauli algebra.

This state, which we call $\pi_{(1)}$, embeds the qubit differently but seems to act the same way. To make "looks the same" more precise, note

that we translate the $\pi_{(1)}$ embedding into the $\pi_{(0)}$ embedding by moving second column to first with a Pauli X, applying a target operator, then swapping back with X once more, as in Fig. 16.

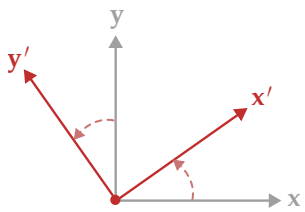
$$\begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix} \xrightarrow{X} \begin{bmatrix} \gamma & \alpha \\ \delta & \beta \end{bmatrix} \xrightarrow{A} \begin{bmatrix} \gamma' & \alpha' \\ \delta' & \beta' \end{bmatrix} \xrightarrow{X} \begin{bmatrix} \alpha' & \gamma' \\ \beta' & \delta' \end{bmatrix}.$$

Figure 16: We can "simulate" $\pi_{(1)}$ with $\pi_{(0)}$ using a change of basis.

For target operators Y and Z, this yields $XZX = -Z$ and $XYX = -Y$, exchanging (38) and (39) as expected.

This is the phenomenon of basis change, familiar from linear algebra. It's often convenient to map vectors to a new coordinate system where an operation is simpler, e.g. an eigenbasis. If U is the *change of basis matrix*, then $A' = U^{-1}AU$ applies the simple operation in the old basis. Instead of shifting operators, we can shift states:

$$\pi'(A) = \pi(A') = \pi(U^{-1}AU).$$



Changing basis from $(\mathbf{x}, \mathbf{y})$ *to* $(\mathbf{x}', \mathbf{y}')$ *with a rotation matrix.*

We can think of π' as the state π in the new basis. But we have to be careful; while every invertible U gives rise to well-defined operators $A' = U^{-1}AU$, π' is not always a well-defined state. Although linearity and the norm condition are guaranteed, positivity can fail. Evaluating π' on a positive operator gives

$$\pi'(A^*A) = \pi[(A^*A)'] = \pi(U^{-1}A^*AU), \tag{41}$$

The norm condition follows because $\pi'(I) = \pi(U^{-1}U) = \pi(I) = 1$.

and $U^{-1}A^*AU$ need not be positive. We can fix this by requiring U to be *unitary*, so the adjoint and inverse coincide: $U^* = U^{-1}$. In this case, the argument $U^{-1}A^*AU = |AU|^2$ of (41) is positive, and hence the new state π' is positive by the positivity of π.

We could consider $A' = U^*AU$ from the outset, which would ensure positivity, but then the norm condition might fail.

Let $\mathcal{U}(\mathcal{A})$ denote the set of unitary operators in $\mathcal{A}$, and define *conjugation* by any $U \in \mathcal{U}(\mathcal{A})$ for operators and hence states as

$$\mathcal{C}_U[A] = U^*AU, \quad \mathcal{C}^U[\pi](A) = \pi(\mathcal{C}_U[A]). \tag{42}$$

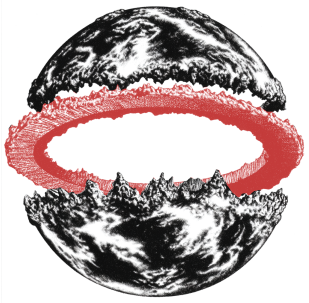
We call $\mathcal{C}^U$ the "pullback" of $\mathcal{C}_U$, since it "pulls $\mathcal{C}_U$ back" onto states.

These maps are "multiplicative" over the change of basis:

$$\mathcal{C}_{UV} = \mathcal{C}_V \circ \mathcal{C}_U, \quad \mathcal{C}^{UV} = \mathcal{C}^U \circ \mathcal{C}^V, \tag{43}$$

with an identity transformation $U = I$. Returning to our original question, we say π and π' are *unitarily equivalent* if they are conjugate, $\pi' = \mathcal{C}^U[\pi]$. As the name suggests, being unitarily equivalent is an equivalence relation between states, as is easily checked:

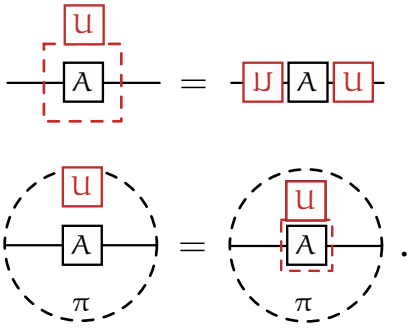
- for reflexivity, setting $U = I$ gives $\mathcal{C}^I[\pi] = \pi$;
- for symmetry, if $\mathcal{C}^U[\pi] = \pi'$, then $\mathcal{C}^{U^*}[\pi'] = \pi$;
- for transitivity, if $\mathcal{C}^U[\pi] = \pi'$ and $\mathcal{C}^{U'}[\pi'] = \pi''$, then $\mathcal{C}^{UU'}[\pi] = \pi''$.



Much confusion about phase symmetry is avoided by thinking about the unitary equivalence from which it is born.

This lumps states into *unitary equivalence classes* which have the same operator correlations, up to a change of basis.

Diagrammatically, we'll attach U outside a dotted box to conjugate an operator. Within a state, we can detach it from the box and attach to the inside of the dotted circle to indicate state conjugation:



This may seem a bit ad hoc, but is an instance of a more general convention that dotted contours are functions, with contour shape indicating output type. We give various examples in Fig. 17.

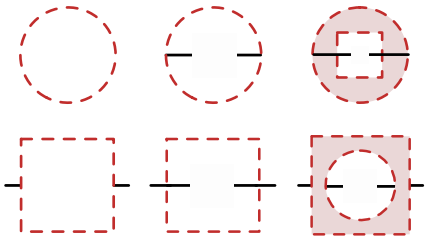


Figure 17: LEFT. Scalar and operator-valued functions of $\mathbb{C}$, e.g. $\lambda \mapsto 2\lambda$ and $\lambda \mapsto \lambda I$. MIDDLE. Functions of $\mathcal{A}$, e.g. a state π or a channel like $\mathcal{C}_U$. RIGHT. Scalar function of a channel, e.g. channel capacity, and operator-valued function of a state, e.g. $\pi \mapsto A_\pi$.

Input type is a bit trickier, but we use leads for operators, blank space for scalars, and otherwise nest contours.

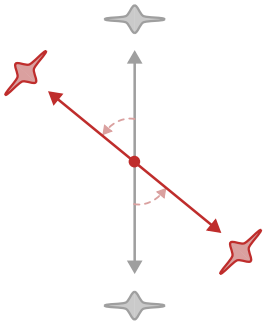
After this pictorial detour, let's return to the question of equivalence. It's clear that distinct but unitarily related states are different; they lead to different experimental outcomes. But it's easy to plonk them into the *same* Hilbert space! For instance, consider the GNS Hilbert space $\mathcal{H}_\pi$ associated with π, and another state $\pi' = \mathcal{C}^U[\pi]$ in the same unitary equivalence class. We can associate these states with respective unit vectors $\pi \mapsto [I]_\pi, \pi' \mapsto [U]_\pi$, since, using the bra-ket notation (35),

$$\pi(A) = {}_\pi\langle I|A|I\rangle_\pi, \quad \pi'(A) = {}_\pi\langle I|U^*AU|I\rangle_\pi = {}_\pi\langle U|A|U\rangle_\pi.$$

Thus, we recover the standard notion of expectations with respect to a unit norm vector inside a Hilbert space! The unitary orbit

$$\mathcal{V}_\pi = \mathcal{U}(\mathcal{A}) \cdot [I]_\pi \subseteq \mathcal{H}_\pi \tag{44}$$

is the set of *vector states* in $\mathcal{H}_\pi$, since these are precisely the vectors of unit norm. In the Hilbert space formulation, these are what we usually think of as states!



Rotating the Stern-Gerlach apparatus changes the physical outcomes.

Clearly ${}_\pi\langle U|U\rangle_\pi = \pi(U^*U) = \pi(I) = 1$, and though we won't show it, every unit vector has this form.

8. The Bloch sphere

As a particularly elegant illustration, we will show that the unitary equivalence class of $\pi_{(0)}$ looks like a sphere. To do this, we will:

- describe all unitaries $\mathcal{U}(\mathcal{A}_{\text{Pauli}})$;
- show that conjugation maps Z to operators of the form $\sigma(\mathbf{n})$;
- conclude $\pi_{(0)}$ is conjugate to the $\sigma(\mathbf{n}) \leftarrow +1$ eigenfunctional $\pi_{(\mathbf{n})}$;
- finally, interpret these statements geometrically.

This will give us a particularly clear derivation of the Bloch sphere and the mysterious half-angles associated with it.

We start by parameterizing the unitaries. Let's extend our previous notation and write $\sigma(\mathbf{v}) = v_i \sigma_{(i)}$ for arbitrary $\mathbf{v} \in \mathbb{C}^3$. Then any $A \in \mathcal{A}_{\text{Pauli}}$ can be written $A = \alpha I + \sigma(\mathbf{v})$, with square

$$\begin{aligned} |A|^2 &= \big(\alpha I + \sigma(\mathbf{v})\big)^* \big(\alpha I + \sigma(\mathbf{v})\big) \\ &= |\alpha|^2 I + \sigma(\overline{\alpha}\mathbf{v} + \alpha\mathbf{v}^*) + |\sigma(\mathbf{v})|^2. \end{aligned}$$

The proof in Fig. 10 goes through as before, except that I is replaced by $|\mathbf{v}|^2 = v_i v_i^*$. For A to be unitary, this means the middle term must vanish and the remaining coefficients sum to unity:

$$\text{Re}[\overline{\alpha}\mathbf{v}] = 0, \quad |\alpha|^2 + |\mathbf{v}|^2 = 1.$$

We can solve these constraints with $\alpha = e^{i\delta}\cos\theta$ and $\mathbf{v} = ie^{i\delta}\sin\theta\,\mathbf{n}$ for phases $\delta, \theta \in [0, 2\pi)$ and unit vector $\mathbf{n} \in \mathbb{R}^3$. We call the resulting expression a *Pauli exponential* and write

$$e^{i\delta I + i\theta\sigma(\mathbf{n})} = e^{i\delta}\big(I\cos\theta + i\sin\theta\,\sigma(\mathbf{n})\big), \tag{45}$$

since it agrees with the results of exponentiating $i\delta I + i\theta\sigma(\mathbf{n})$:

$$\begin{aligned} e^{i\delta I + i\theta\sigma(\mathbf{n})} &= e^{i\delta} \sum_{k=0}^{\infty} \frac{[i\theta\sigma(\mathbf{n})]^k}{k!} \\ &= e^{i\delta} \sum_{k=0}^{\infty} \frac{(-1)^k\theta^{2k}}{(2k)!} + i\sum_{k=0}^{\infty} \frac{(-1)^k\theta^{2k+1}}{(2k+1)!}\sigma(\mathbf{n}) \\ &= e^{i\delta}\big(I\cos\theta + i\sin\theta\,\sigma(\mathbf{n})\big). \end{aligned}$$

This follows from straightforward power series manipulation. More generally, if $f(x)$ has a power series, $f(A)$ denotes the result of formally replacing x with A:

$$f(x) = \sum_{k=0}^{\infty} c_k x^k \quad \Longrightarrow \quad f(A) = \sum_{k=0}^{\infty} c_k A^k. \tag{46}$$

We don't worry about convergence or other analytic niceties here.

Unitary operators beget vector states.

We split the series into even and odd terms, simplify using $\sigma(\mathbf{n})^{2k} = I$ and $\sigma(\mathbf{n})^{2k+1} = \sigma(\mathbf{n})$, and invoke the Taylor series for sine and cosine. Note that we pull $e^{i\delta I}$ out the front since I commutes with everything else.

These niceties lead to the *continuous functional calculus*. We give details and applications in Appendix A.

Now that we have a general unitary of the form $U = e^{-i\theta\sigma(\mathbf{n})}$, we can use it to build the equivalence class of Z. For simplicity, we'll assume $\mathbf{n} = (n_1, n_2, 0)$ is orthogonal to the $\mathbf{z}$-axis. Algebraically, this implies that Z and $\sigma(\mathbf{n})$ *anticommute*:

$$Z\sigma(\mathbf{n}) = Z(n_1 X + n_2 Y) = -(n_1 X + n_2 Y)Z = -\sigma(\mathbf{n})Z.$$

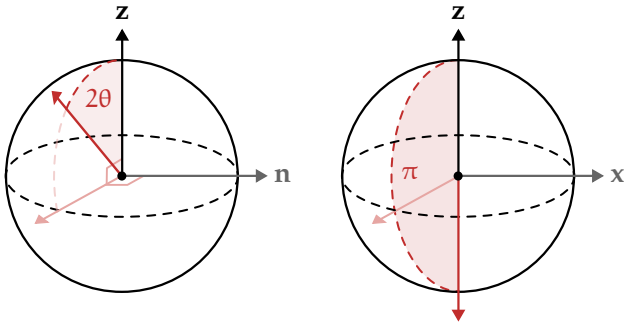
Conjugating by $U = e^{-i\theta\sigma(\mathbf{n})}$ then gives

$$\begin{aligned}\mathcal{C}_U[Z] &= \big(I\cos\theta + i\sin\theta\,\sigma(\mathbf{n})\big)Z\big(I\cos\theta - i\sin\theta\,\sigma(\mathbf{n})\big)\\ &= \big(\cos^2\theta - \sin^2\theta\big)Z + 2i\sin\theta\cos\theta\sigma(\mathbf{n})Z\\ &= \cos(2\theta)Z + i\sin(2\theta)\sigma(\mathbf{n})Z\\ &= \sin(2\theta)\,[-n_2 X + n_1 Y + \cot(2\theta)Z] = \sigma(\mathbf{n}'),\end{aligned} \tag{47}$$

where $\mathbf{n}' = \sin(2\theta)(-n_2, n_1, \cot(2\theta))$ is a unit vector:

$$|\mathbf{n}'|^2 = \sin^2(2\theta)\big(n_1^2 + n_2^2 + \cot^2(2\theta)\big) = 1.$$

Geometrically, conjugation rotates $\mathbf{z}$ by an angle 2θ towards the vector $(-n_2, n_1, 0)$ orthogonal to both $\mathbf{z}$ and $\mathbf{n}$ (Fig. 18, left):



As a special case, setting $U = e^{-i(\pi/2)X} = X$ recovers the column-swap change of basis $\mathcal{C}_U[Z] = -Z$ from the previous section.

We see on geometric grounds that we can map $\mathbf{z}$ to any unit vector $\mathbf{m}$, and hence Z to any generalized Pauli $\sigma(\mathbf{n})$: choose $\mathbf{n}$ orthogonal to our target vector $\mathbf{m}$ and θ half the angle between $\mathbf{z}$ and $\mathbf{m}$. Thus, every generalized Pauli operator is conjugate. We've only considered $\mathbf{n} \perp \mathbf{z}$, but the geometry of the general case is similar: conjugation by $U = e^{-i\theta\sigma(\mathbf{n})}$ rotates $\mathbf{z}$ an angle θ counter-clockwise around the $\mathbf{n}$-axis, tracing out a cone at fixed azimuthal angle to $\mathbf{n}$, shown in Fig. 19. It doesn't add much, but if you're curious, here is the general result in algebraic form:

$$\begin{aligned}\mathcal{C}_U[\sigma(\mathbf{m})] = \cos(2\theta)\sigma(\mathbf{m}) + i\sin(2\theta)\sigma(\mathbf{m}\times\mathbf{n})\\ + 2\sin^2\theta(\mathbf{m}\cdot\mathbf{n})\,\sigma(\mathbf{n}),\end{aligned}$$

where $\mathbf{m}\cdot\mathbf{n} = m_i n_i$ and $(\mathbf{m}\times\mathbf{n})_i = \epsilon_{ijk} m_j n_k$ are our old friends from vector analysis, the dot and cross product. We recover (47) by setting $\mathbf{m} = \mathbf{z}$ and $\mathbf{m}\cdot\mathbf{n} = 0$.

We can ignore the $e^{i\delta}$ in (45) since it is cancelled by the adjoint phase $e^{-i\delta}$.

Remember that $\mathbf{n} = (n_1, n_2, 0)$ is a unit vector so $n_1^2 + n_2^2 = 1$.

Figure 18: Left. Conjugating Z by $e^{-i\theta\sigma(\mathbf{n})}$ (for orthogonal $\mathbf{n}$) rotates the unit vector $\mathbf{z}$ ccw by 2θ in the plane perpendicular to $\mathbf{n}$. Right. A special case $XZX = -Z$.

To derive this, you can extend the proof in Fig. 10 to obtain the neat identity

$$\sigma(\mathbf{m})\sigma(\mathbf{n}) = (\mathbf{m}\cdot\mathbf{n})I + i\sigma(\mathbf{m}\times\mathbf{n}).$$

Applying this a few times gives the result on the left.

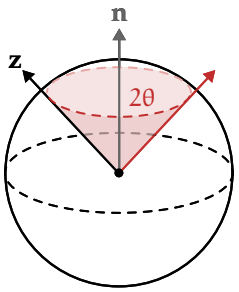


Figure 19: Conjugation by $e^{-i\theta\sigma(\mathbf{n})}$ rotates $\mathbf{z}$ counter-clockwise by 2θ in the plane perpendicular to $\mathbf{n}$.

Equivalence of operators translates into equivalence of states. To prove this, we actually need to give a more careful definition of the kernel $\mathcal{K}_\pi$, which fully characterizes the state π. Recall that $\mathcal{K}_\pi$ consists of operators uncorrelated with anything else in π, as per (21). There is some minimal set $\mathcal{I}_\pi$ of identifications that produces all such operators, usually corresponding to definite measurement outcomes, e.g. $Z \equiv_{(0)} I$ for $\pi_{(0)}$. We can regard $\mathcal{K}_\pi$ as the left ideal generated by $\mathcal{I}_\pi$,

$$\mathcal{K}_\pi = \mathcal{A}\mathcal{I}_\pi \subseteq \mathcal{A}, \tag{48}$$

consisting of all terms of the form $AR \in \mathcal{A}\mathcal{I}_\pi$, simplified using the "ambient" algebra $\mathcal{A}$. Formally, we can view this as the $*$-algebra generated by the (redundant) set $\mathcal{A}\mathcal{I}_\pi$, subject to any complete set of relations for $\mathcal{A}$.

The result of this somewhat abstract characterization is a simple sufficient condition for the equivalence of states:

$$\mathcal{C}_U[\mathcal{I}_\pi] = \mathcal{I}_{\pi'} \implies \mathcal{C}_U[\mathcal{K}_\pi] = \mathcal{K}_{\pi'} \iff \mathcal{C}^U[\pi] = \pi'. \tag{49}$$

It's important to note that this condition isn't *necessary*: the sets $\mathcal{I}_\pi$ are not unique, so $\mathcal{C}_U[\mathcal{I}_\pi] \neq \mathcal{I}_{\pi'}$ does not mean π and π' are inequivalent. But once we have a set $\mathcal{I}_\pi$ we can "sweep out" the vector states by seeing what left ideals are generated by $\mathcal{C}_U[\mathcal{I}_\pi]$. So, returning to Pauli operators, the $Z \leftarrow +1$ eigenfunctional has a generating set $\mathcal{I}_{(0)} = \{Z - I\}$. The hard work we did conjugating operators above now pays dividends, since

For instance, $Z \equiv_{(0)} I$ is equivalent to $\alpha Z \equiv_{(0)} \alpha I$.

$$\mathcal{C}_U[\mathcal{I}_{(0)}] = \{\mathcal{C}_U[Z] - I\} = \{\sigma(\mathbf{m}) - I\}$$

for some unit vector $\mathbf{m}$ determined by U. It follows from (49) that $\mathcal{C}^U[\pi_{(0)}] = \pi_{(\mathbf{m})}$, the $\sigma(\mathbf{m}) \leftarrow +1$ eigenfunctional, just as we expect.

To cast this in a more familiar form, let's work out the vector states (44) of $\mathcal{H}_{(0)}$, i.e. the unitary orbit of $[I]_{(0)} = |0\rangle$. Recall that we can rotate $\mathbf{z}$ to any target vector using an orthogonal vector

$$\mathbf{n} = (n_1, n_2, 0) = (\cos\phi, \sin\phi, 0)$$

for some polar angle $\phi \in [0, 2\pi)$. Applying a general Pauli exponen-

tial and simplifying with $[X]_+ = i[Y]_+ = |1\rangle$, we find

$$\begin{aligned} e^{i\delta}e^{-i\theta\sigma(\mathbf{n})}|0\rangle &= e^{i\delta}\left[\cos\theta|0\rangle + i\sin\theta(\cos\phi\, X + \sin\phi\, Y)|0\rangle\right] \\ &= e^{i\delta}\left[\cos\theta|0\rangle + i\sin\theta(\cos\phi + i\sin\phi)|1\rangle\right] \\ &= e^{i\delta}\left[\cos\theta|0\rangle + i\sin\theta e^{i\phi}|1\rangle\right] = |\psi\rangle. \end{aligned}$$

Now for some bookkeeping. We want to make sure we don't double count vector states, and a crucial observation is that θ is *doubled* on the sphere of unit vectors (see Fig. 18). We therefore define a new angle $\vartheta = 2\theta$ restricted to $[0, 2\pi]$. This leads to the familiar expression

$$|\psi\rangle = e^{i\delta}\left[\cos\left(\tfrac{1}{2}\vartheta\right)|0\rangle + i\sin\left(\tfrac{1}{2}\vartheta\right)e^{i\phi}|1\rangle\right], \tag{50}$$

which parametrizes the Bloch Sphere,[21] shown in Fig. 20:

[21] "Nuclear induction" (1946), Felix Bloch.

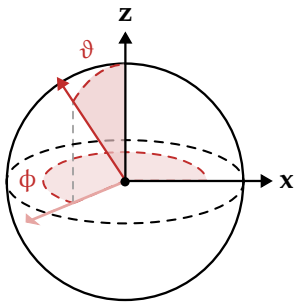


Figure 20: The Bloch sphere of state vectors $\mathcal{V}_{(0)} \subseteq \mathcal{H}_{(0)}$, with global phase suppressed.

As usual, we suppress global phase, but it still has mathematical utility. Noting that $\cos[(2\pi - \vartheta)/2] = -\cos(\vartheta/2)$, and similarly for sin, we see that substituting $\vartheta \mapsto 2\pi - \vartheta$ adds an overall minus sign. We can absorb this into the phase $\delta \mapsto \delta + \pi$, and thus reduce the range of $\vartheta \in [0, \pi]$. It now properly resembles an azimuthal angle!

If you like, you can imagine a circle attached to each point of the sphere. If you are particularly adept at higher-dimensional visualization, you can braid those to form a three-sphere S^3! This is called the *Hopf fibration*.

Hopefully, the origin of the half-angles in (50) is now clear: when we conjugate an operator $\sigma(\mathbf{m})$ by a Pauli exponential $e^{-i\theta\sigma(\mathbf{n})}$, the vector $\mathbf{m}$ is rotated by 2θ since we have a rotation on either side. Ultimately, this means the vector states $\mathcal{V}_{(\mathbf{m})}$ of $\sigma(\mathbf{m})$ covers the unit sphere *twice*. The second cover is associated with a factor of -1, but when we have a global phase we can absorb it.

9. Mixed states

In building a truth table for a noncommutative circuit, we want to make sure we find all the states. We've just explained, for instance, how to populate a whole sphere of states in the case of the Pauli algebra. You can show that all states of $\mathcal{A}_{\text{Pauli}}$ given by definite measurements of (self-adjoint) observables are captured on the Bloch sphere. But while this lists all states with *definite* measurements, we have completely neglected the *indefinite* states, where measurements are fuzzy or information is partial. These are no less important!

The argument goes as follows: if $\Lambda = \Lambda^*$ is self-adjoint, then it is a real linear combination:

$$\Lambda = \alpha I + \sigma(\mathbf{v}) \tag{51}$$

for some $(\alpha, \mathbf{v}) \in \mathbb{R}^4$. Writing $\mathbf{v} = |v|\hat{\mathbf{v}}$, then a sharp measurement $\Lambda \leftarrow \lambda$ is equivalent to $\sigma(\hat{\mathbf{v}}) \equiv_\pi |v|^{-1}(\lambda - \alpha)I$, so we are done.

Luckily, we can build indefinite states by mixing definite ones. In the language of expectation functionals, any "probabilistic" combination of two states forms a state, i.e.

$$\pi = p_1\pi + p_2\pi' \tag{52}$$

is a state provided the coefficients are positive (to ensure π is a positive functional) and sum to one (to ensure π is normalized). More generally, a positive, normalized linear combination is called a *convex combination*, and the resulting functional a *mixed state*. We can view this as randomly choosing a state $\pi_{(j)}$ with probability p_j, or equivalently, associating any operator A to a random variable $A|_{\pi_{(j)}}$.

To make a mixture, just toss a (potentially biased) coin and pick a state.

Consider any set of elements K in a vector space. We call the set of all convex combinations of elements in K the *convex hull*:

$$\mathrm{conv}(K) = \left\{ \sum_j p_i k_j : k_j \in K, p_j \geq 0, \sum_j p_j = 1 \right\}. \tag{53}$$

An *extreme point* is one that cannot be obtained by mixing, or equivalently, such that removing it from K gives a different hull. We will abuse notation and write

$$\partial K = \{k \in K : \mathrm{conv}(K - \{\pi\}) \neq \mathrm{conv}(K)\}. \tag{54}$$

In the algebraic setting, and $K = S(\mathcal{A})$ is the set of states, the extreme points $\partial S(\mathcal{A})$ are called *pure states*.

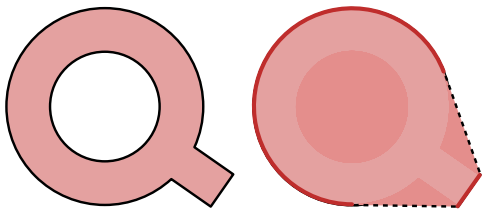


Figure 21: A set and its convex hull, with extreme points in red.

Fig. 21 gives a simple example. Note that the convex hull is in fact the hull of the extreme points. This fact is true in general, a result called the KREIN-MILMAN THEOREM:[22]

$$\mathrm{conv}(K) = \mathrm{conv}(\partial K). \tag{55}$$

[22] "On extreme points of regular convex sets" (1940), Mark Krein and David Milman.

The intuition is that you can "unmix" $k \in \mathrm{conv}(K)$ to obtain "more extreme" points that combined to give it, unmix those, and so on, in such a way that the process terminates with a set of extreme points. In finite dimensions this argument works nicely, but in infinite dimensions, you need the Axiom of Choice.

The Axiom of Choice postulates that you can pick an element from a nonempty set. Surprisingly, this is one of the most controversial statements in mathematics!

The Krein-Milman theorem (55) tells us that we can obtain everything by combining pure states, so in a sense we are done. With

unitary orbits, we were also "done" once we had single representative; it turned out to be helpful to explicitly construct the $\mathcal{V}_\pi$, and in particular we recovered the standard Hilbert space description of states. Similarly, it is useful here to explicitly construct and work with mixed states, and recover the conventional formalism. To guide us, we'll work with a concrete example on $\mathcal{A}_{\text{Pauli}}$:

$$\pi_{(p)} = p\pi_{(0)} + (1-p)\pi_{(1)} = p_b\pi_{(b)}, \tag{56}$$

which flips a coin and uses $\pi_{(0)}$ with probability $p_0 = p$ and $\pi_{(1)}$ with $p_1 = 1-p$. These states form an axis inside the Bloch sphere:

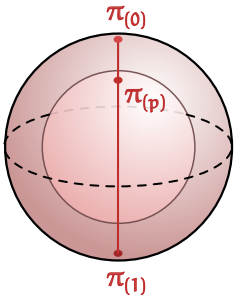


Figure 22: The *Bloch ball* is the convex hull of the Bloch sphere, and encodes all states of $\mathcal{A}_{\text{Pauli}}$. The state $\pi_{(0)}$ is at the north pole, $\pi_{(1)}$ at the south, and $\pi_{(1)}$ lies on the axis between. The orbit of $\pi_{(p)}$ is a sphere of mixed states with equivalent mixedness.

Since every convex line lies inside, we know from the Krein-Milman theorem that the *Bloch ball* captures all states of $\mathcal{A}_{\text{Pauli}}$.

As we showed in (25), in this state Z has variance $4p(1-p)$, so it is not sharp when $0 < p < 1$. In fact, you can show that for any operator $A \in \mathcal{A}_{\text{Pauli}}$, the variance is

$$\pi_{(p)}(|\Delta A|^2) = p_b\pi_{(b)}(|\Delta A|^2) + p_0p_1|\Delta\pi(A)|^2, \tag{57}$$

where $\Delta\pi = \pi_{(0)} - \pi_{(1)}$. The only way for an operator to be definite, then, is to be definite for both $\pi_{(0)}$ and $\pi_{(1)}$, with the same average. But the only such operator is the identity! Thus, the kernel is trivial, $\mathcal{K}_{(p)} = \{0\}$, and the GNS Hilbert space is the algebra itself:

$$\mathcal{H}_{(p)} = \mathcal{A}_{\text{Pauli}}/\mathcal{K}_{(p)} \cong \mathcal{A}_{\text{Pauli}}.$$

The *Law of Total Variance* implies that, for mixed states $\pi_p = p_i\pi_{(i)}$, we have

$$\text{var}_p(\Gamma) = p_i\text{var}_i(\Gamma) + \text{var}_{\pi_{(i)}\sim p_i}[\pi_{(i)}(\Gamma)].$$

The variance of operators in a mixture is the mean variance plus the variance of means. For coin flips, this gives (57).

Since Z is sharp but the means differ, while X and Y both have vanishing mean but nonzero variance.

Effectively, this lumps together the columns from Figs. 14 and 15 to form a matrix. This is what we might have guessed would result from combining $\pi_{(0)}$ and $\pi_{(1)}$!

Lumping together two vector spaces V and W (over the same field) is called the *direct sum*. We just stack vectors from each space on top of each other, and define operations component-wise:

$$V \oplus W = \{(v, w) : v \in V, w \in W\} \tag{58}$$

$$\alpha_i(v_i, w_{\hat{\imath}}) = (\alpha_i v_i, \alpha_{\hat{\imath}} w_{\hat{\imath}}). \tag{59}$$

In (59), we introduce the convention that hatted indices like are *not* summed over, so that $(v_i, w_{\hat{\imath}})$ represents a single vector in the direct

sum, and only once we place α_i out front do we sum over i. Returning to our story, the Hilbert space associated with π_p is then

$$\mathcal{H}_{(p)} = \mathcal{H}_{(0)} \oplus \mathcal{H}_{(1)}. \tag{60}$$

Since $\mathcal{H}_{(p)}$ can be broken down into smaller pieces, we say it is *reducible*; this holds for mixed states in general, since we can break them down into the component pure Hilbert spaces. The pure state Hilbert spaces $\mathcal{H}_{(b)}$ are *irreducible*, i.e. cannot be so reduced. Again, this statement generalizes, so a state $\pi \in S(\mathcal{A})$ is pure just in case $\mathcal{H}_\pi$ is irreducible.[23] We give a more useful purity criterion below.

To reduce them, we'd need a three-dimensional kernel, which we show is impossible in the next section.

[23] If you don't believe me, you may believe Theorem 10.2.3 of *Fundamentals of the Theory of Operator Algebras II* (1997), Richard Kadison and John Ringrose.

Although $\pi_{(1)}$ is associated with a distinct GNS Hilbert space $\mathcal{H}_{(1)}$, recall from Fig. 16 that it can be embedded in $\mathcal{H}_{(1)}$ using a change-of-basis matrix X, or equivalently, as a vector state $X|0\rangle = |1\rangle$. This lets us evaluate $\pi_{(p)}$ entirely within $\mathcal{H}_{(0)}$:

$$\pi_{(p)}(A) = p_0\pi_{(0)}(A) + p_1\pi_{(1)}(XAX) = p_b\langle b|A|\hat{b}\rangle, \tag{61}$$

where $b \in \{0, 1\}$ and we apply our summation convention to basis vector labels. It's convenient to rewrite this using two new operations. First, given two vector states $|\phi\rangle = U|0\rangle, |\psi\rangle = U'|0\rangle$, we define an operator on $\mathcal{H}_{(0)}$ called the *outer product* which scales $|\psi\rangle$ by the overlap with $|\phi\rangle$:

$$|\psi\rangle\langle\phi|(|\zeta\rangle) = \langle\phi|\zeta\rangle|\psi\rangle. \tag{62}$$

This is linear in $|\zeta\rangle$ because the inner product is linear. A closely related operation is the *trace*, defined by

$$\mathrm{tr}\big[|\psi\rangle\langle\phi|A\big] = \langle\phi|A|\psi\rangle = \langle 0|U^*AU'|0\rangle = \pi(U^*AU'), \tag{63}$$

where $|0\rangle$ is the vector state associated with π.

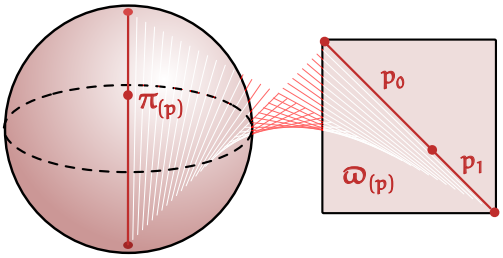


Figure 23: A fanciful depiction of how a mixed state $\pi_{(p)}$ in the Bloch ball transforms into a density matrix $\varpi_{(p)}$: invert the Bloch axis and map onto the principal diagonal. In higher dimensions, we unravel a simplex; left as an exercise to readers who successfully wove the Hopf fibration.

Using outer products and the trace, we can write (61) as

$$\pi_{(p)}(A) = \mathrm{tr}\left[\left(p_b|b\rangle\langle\hat{b}|\right)A\right] = \mathrm{tr}[\varpi_{(p)}A], \tag{64}$$

where $\varpi_{(p)}$ is an operator called the *density matrix*. Fig. 23 visualizes how this matrix is formed. Note that, in general, the set of bounded operators $\mathcal{B}(\mathcal{H})$ is spanned by outer products. This implies that the trace is defined for arbitrary matrices and is *cyclic*, in the sense that

If $\{|i\rangle\}_{i\in\mathcal{I}}$ is an orthonormal basis of $\mathcal{H}$, then $\{|i\rangle\langle i|\}_{i,i'\in\mathcal{I}}$ is an orthonormal basis of $\mathcal{B}(\mathcal{H})$. Orthonormality is easy, and there are enough to span the space.

$$\mathrm{tr}[AB] = A_{ij}B_{ji} = B_{ji}A_{ij} = \mathrm{tr}[BA] \tag{65}$$

for $A = A_{ij}|i\rangle\langle j|$ and $B = B_{ij}|i\rangle\langle j|$.

These results generalize nicely. Suppose we mix $\pi_{(p)} = p_i \pi_{(i)}$, where the states $\pi_{(i)} = \mathcal{C}^{U_{(i)}}[\pi]$ are unitarily equivalent to some fiducial representative π, analogous to $\pi_{(0)}$. We can write

"Fiducial" means we choose it; "canonical" means God chooses it.

$$\pi_{(p)}(A) = p_i \mathcal{C}^{U_{(i)}}[\pi](A) = p_i \pi\big(U^*_{(i)} A U_{(\hat{\imath})}\big) = \mathrm{tr}\big[\varpi_{(p)} A\big], \tag{66}$$

for a density matrix

$$\varpi_{(p)} = p_i |U_{(i)}\rangle\langle U_{(\hat{\imath})}|. \tag{67}$$

This is the conventional formalism of mixed states.

We can repeat the argument in the awd formalism. To mix unitarily equivalent states, we tether p_i to U_i and conjugate a ficudial state. It's then straightforward to obtain $\pi_{(p)}(A) = p_i \langle U_{(i)}|A|U_{(\hat{\imath})}\rangle$:

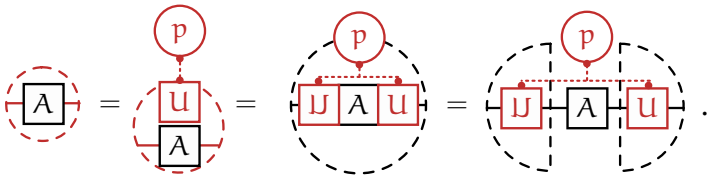


Figure 24: For $\pi_i = \mathcal{C}^{U_{(i)}}[\pi_0]$, the state $\pi_{(p)} = p_i \pi_{(i)}$ takes a simple form. Note that the tether line can "split" when different objects carry the same index.

Finally, we drag $|U_{(i)}\rangle$ around to form an outer product:

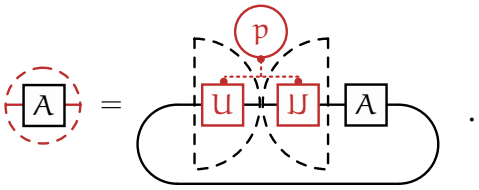


Figure 25: Deforming Fig. 24 to obtain (67), with the awd of a trace.

This is (66), which embeds (67) as the operator next to A, and represents the trace as a solid loop with operators strung along it anticlockwise; this manifests (65) visually.

For now, think of this as a convention for the trace; we will interpret the loop in terms of entanglement below.

10. *Pure states*

According to the Krein-Milman theorem (55), everything is made of pure states. We stated above that these had irreducible GNS Hilbert spaces, but this turns out to be a cumbersome way of testing purity. We will introduce a method which uses observables directly, led, as usual, by the Pauli algebra. In this case, the pure states live on the Bloch sphere, and correspond to sharp measurements of a self-adjoint operator $\sigma(\mathbf{n}) \leftarrow \pm 1$. This measurement fixes all expectations; e.g., $Z \leftarrow +1$ implies $\pi_{(0)}(X) = \pi_{(0)}(Y) = 0$, so all expectations can be computed. Once you fix everything, nothing more can be made sharp, so $\mathcal{D}_{(\mathbf{n})}$ is *maximal*. Conversely, in a mixed state $\pi_{(p)}$, the definite set $\mathcal{D}_{(p)} = \mathbb{R}I$ *can* be extended. Any measurement will do!

Particularly given our efforts to escape the inconvenience of Hilbert space!

An ad hoc way to see this:

$$\pi_{(0)}(X) = \pi_{(0)}(XZ^2) = -\pi_{(0)}(ZXZ),$$

using Pauli relations. Since Z is sharp, it factors out to give $-\pi_{(0)}(X)$, hence $\pi_{(0)}(X) = 0$. We give a more general argument below.

This equivalence holds in general. To state it precisely, recall that the definite set $\mathcal{D}_\pi$ consists of all self-adjoint operators with zero

variance under π, i.e. $\pi(\Gamma^2) = \pi(\Gamma)^2$. STØRMER'S THEOREM[24] shows that π is pure just in case $\mathcal{D}_\pi$ is *maximal*, that is, there is no other state π' such that $\mathcal{D}_{\pi'} \supsetneq \mathcal{D}_\pi$. Operationally, if we try to sharpen a blurry observable, another will become indefinite! This is an uncertainty principle, but for a whole collection of operators $\mathcal{D}_\pi$; see Fig. 26.

[24] "A characterization of pure states of C*-algebras" (1967), Erling Størmer.

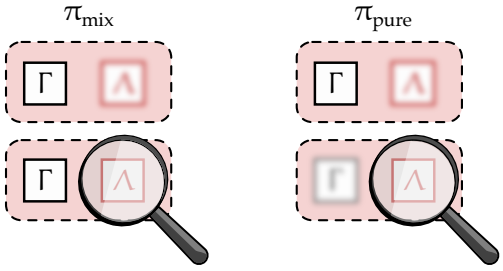


Figure 26: LEFT. In a mixed state, we can increase the set of sharp observables. RIGHT. In pure states, defining one observable blurs an observable elsewhere.

In contrast to the kernel $\mathcal{K}_\pi$, which is closed under products and complex linear combinations, the definite set $\mathcal{D}_\pi$ is closed under neither. However, it is closed under *real* linear combinations, since $\Gamma, \Lambda \in \mathcal{D}_\pi$ and $a, b \in \mathbb{R}$ implies

$$(a\Gamma + b\Lambda)^* = \overline{a}\Gamma^* + \overline{b}\Lambda^* = a\Gamma + b\Lambda,$$

and linear combinations of sharp observables are sharp by the triangle inequality (26). And although $\mathcal{D}_\pi$ is not closed under products, it is closed under a peculiar operation called the *Jordan product*:

$$\Gamma \circ \Lambda = \frac{1}{2}(\Gamma\Lambda + \Lambda\Gamma) = \frac{1}{2}\{\Gamma, \Lambda\}, \tag{68}$$

where $\{\cdot,\cdot\}$ is the *anticommutator*. This is commutative by construction, and ensures that $(\Gamma \circ \Lambda)^* = \Gamma \circ \Lambda$ for self-adjoint Γ, Λ. Moreoever, the Jordan product is sharp by equation (21), with

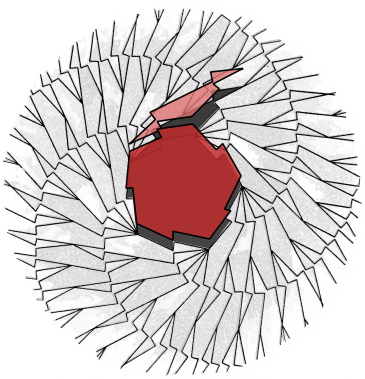
$$\pi(\Gamma \circ \Lambda) = \frac{1}{2}\pi(\Gamma\Lambda) + \frac{1}{2}\pi(\Lambda\Gamma) = \pi(\Gamma)\pi(\Lambda). \tag{69}$$

Crucially, (69) tells us that if we restrict the state π to its definite set $\mathcal{D}_\pi$, the resulting real-valued functional $\chi_\pi = \pi|_{\mathcal{D}_\pi}$ is also *multiplicative* with respect to the Jordan product.

Remarkably, the functionals χ_π fully determine the state when it is pure, as we prove in Appendix B. We say that $\mathcal{D}_\pi$ is *rigid* for pure states. This terminology is borrowed from complex analysis and hyperbolic geometry, but simple illustrations are provided by planar tilings[25] and even recurrence relations. In the Fibonacci sequence, $F_0 = 0$ and $F_1 = 1$ determine everything else! Consider a family of functions $\mathcal{F} = \{f : S \to T\}$ with restrictions $\mathcal{F}|_I = \{f|_R : R \to T\}$. We say $R \subseteq S$ is *rigid for* $\mathcal{F}$ if, for all $f' \in \mathcal{F}|_R$, there is a unique $f \in \mathcal{F}$ such that $f' = f|_R$. Thus, definite sets $\mathcal{D}$ are rigid for pure states, $\mathcal{F} = \partial S(\mathcal{A})$, using our notation for the convex extreme points.

Self-adjointness is easily spoiled! For $\Gamma, \Lambda \in \mathcal{D}_\pi$, $(i\Gamma)^* = -i\Gamma \neq i\Gamma$ unless $\Gamma = 0$, and $(\Gamma\Lambda)^* = \Lambda\Gamma \neq \Gamma\Lambda$ unless Γ and Λ commute. We'll see in a moment that the Jordan product gets around this by symmetrizing.

"Peculiar" because it is *non-associative*, i.e. $A \circ (B \circ C) \neq (A \circ B) \circ C$ in general.



[25] "Patch-determined tilings" (1977), Grünbaum and Shephard. Above, we modify their Fig. 10: a tiling is determined by the placement of a single sawtooth pentagon and rhomb.

Rigidity is central to Størmer's theorem. Intriguingly, rigidity also characterizes *truth tables*. Recall from (9) that lines of the table are Boolean vectors $v_0 \in \mathbb{B}^{\mathcal{P}}$, and they uniquely extend to valuations $v : \mathbb{B}[\mathcal{P}] \to \mathbb{B}$. Put differently, $\mathcal{P}$ is rigid for the family of Boolean valuations. The set $\mathcal{P}$ of propositional variables, or more physically, basic switches, is also *minimal*, since removing a switch x will leave the truth value of many circuits (such as x itself!) undefined. This would be analogous to a mixed state. Unlike $\mathcal{X}$, a definite set $\mathcal{D}$ is a whole vector space over $\mathbb{R}$ and hence far from minimal; χ is determined by the basis of $\mathcal{D}$. Since we can also multiply elements of the basis, this is not usually minimal either! We will define a *switch set* $\mathcal{Q}$ for $\mathcal{D}$ as one which minimally generates it in the algebraic sense:

$$\mathcal{Q} \in \min \arg\max_{\mathcal{G}\subseteq\mathcal{D}} \mathrm{J}^{\circ}\langle\mathcal{G}\rangle, \tag{70}$$

where the argmax means that $\mathcal{G}$ generates $\mathcal{D}$, the J° indicates that generation is with respect to real linear combinations and the Jordan product, and finally, the min means that we pick a *minimal* generating set, such that removing any element no longer generates $\mathcal{D}$.

Note "min" indicates the collection of inclusion-minimal elements. There is always more than one (we can rescale generators), and these need not have the same size!

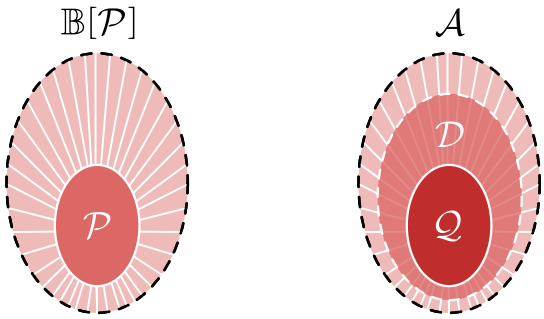


Figure 27: Left. Values on the Boolean switches $\mathcal{P}$ determine the value of any circuit. Right. Quantum switches $\mathcal{Q}$ determine $\mathcal{D}$ determine the full state for pure π.

Let's flesh this out with an example. Suppose we rotate our Stern-Gerlach apparatus in the $\mathbf{n}$-direction and measure $\sigma(\mathbf{n}) \leftarrow +1$. Pauli rotating the $Z \leftarrow +1$ result (38), the kernel takes the form

$$\mathcal{K}_{(\mathbf{n})} = \mathrm{span}_{\mathbb{C}}\{\sigma(\mathbf{n}) - I, \Theta_{(\mathbf{n})}\}, \quad \Theta_{(\mathbf{n})} = \sigma(\mathbf{n}_{\perp}) + i\sigma(\mathbf{n}\times\mathbf{n}_{\perp}),$$

where $\mathbf{n}_{\perp}$ is any unit vector orthogonal to $\mathbf{n}$. Since $\Theta_{(\mathbf{n})}$ is not self-adjoint, the definite set is $\mathcal{D}_{(\mathbf{n})} = \mathrm{span}_{\mathbb{R}}\{I, \sigma(\mathbf{n})\}$. To check rigidity, note that since $\Theta_{(\mathbf{n})}$ has vanishing mean,

To be clear, real and imaginary components are *separately* determined due to positivity of states.

$$\mathrm{Re}\left\{\pi_{(\mathbf{n})}[\Theta_{(\mathbf{n})}]\right\} = \frac{1}{2}\pi_{(\mathbf{n})}[\Theta_{(\mathbf{n})} + \Theta^{*}_{(\mathbf{n})}] = \pi_{(\mathbf{n})}[\sigma(\mathbf{n}_{\perp})] = 0,$$

and similarly $\pi_{\pm\mathbf{n}}[\sigma(\mathbf{n}\times\mathbf{n}_{\perp})] = 0$. Since I, $\sigma(\mathbf{n})$, $\sigma(\mathbf{n}_{\perp})$ and $\sigma(\mathbf{n}\times\mathbf{n}_{\perp})$ span the whole Pauli algebra, the values are indeed determined! The last thing to do is find a switch set. Since $\sigma(\mathbf{n})\circ\sigma(\mathbf{n}) = I$ is the other basis element of $\mathcal{D}_{(\mathbf{n})}$, we can just take $\mathcal{Q}_{(\mathbf{n})} = \{\sigma(\mathbf{n})\}$. Unsurprisingly, we learn that the value $\sigma(\mathbf{n}) \leftarrow \pm 1$ determines everything else.

11. Tensor products

In a Boolean circuit, we can configure a switch by hand, and each switch is independent. In a quantum circuit, we configure a switch by measurement. Measurement is more unruly than manual configuration, since outcomes are random, and later measurements can mess up the result of earlier measurements. We can't configure $\mathcal{Q}$ at will!

It will be instructive to consider a situation where measurements are, like classical switches, completely independent. Suppose our system consists of separate parts called *factors* or *locations*, indexed by $\ell \in \mathfrak{L}$, each of which has associated "local" observables $A^{(\ell)} \in \mathcal{A}^{(\ell)}$. Local measurement is *guaranteed* to be uncorrelated in the sense that

$$\pi_{\text{CDP}}\left(\prod_{\ell\in\mathfrak{L}} \Lambda^{(\ell)}\right) = \prod_{\ell\in\mathfrak{L}} \pi_{\text{CDP}}\left(\Lambda^{(\ell)}\right), \quad \Lambda^{(\ell)} \in \mathcal{A}^{(\ell)}, \tag{71}$$

provided the systems are far enough away from each other and have always been so. This is called the Cluster Decomposition Principle.[26] Since a product of scalars commutes, the order in which we measure on each factor is irrelevant. Hence, if we measure a set of observables $\Lambda^{(\ell)} \in \mathcal{A}^{(\ell)}$, they will be *simultaneously sharp*, since we can suppose that each was the last to be measured. We will make this argument more precise in the next section.

We use the bracketed superscript (ℓ) to indicate locations, as distinguished from unbracketed superscripts (powers) or bracketed subscripts (labelled operator family). We omit these location superscripts in explicit tensor products.

[26] See "Cluster Decomposition Properties of the S Matrix" (1963), E.H. Wichmann and J.H. Crichton, or Weinberg's magisterial *The Quantum Theory of Fields, Volume 1: Foundations*.

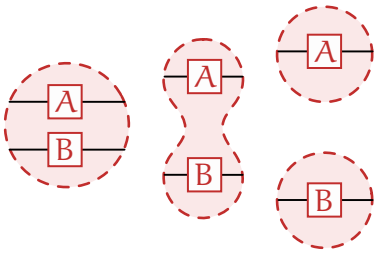


Figure 28: As we separate systems (and their histories), their expectations factorize according to the cluster decomposition principle. This implies measurements on the different systems are jointly sharp.

We can describe the algebra acting on this collection of systems in two stages. For two C*-algebras $\mathcal{A}^{(1)}$ and $\mathcal{A}^{(2)}$, we first define the *algebraic tensor product* $\mathcal{A}^{(1)} \odot \mathcal{A}^{(2)}$ as the set of pairs $\mathcal{A}^{(1)} \times \mathcal{A}^{(2)}$ subject to a *bilinearity relation* (see Fig. 29):

$$\mathcal{A}^{(1)} \odot \mathcal{A}^{(2)} = \text{span}_{\mathbb{C}}(\mathcal{A}^{(1)} \times \mathcal{A}^{(2)})/\mathcal{I}_{\text{BL}}, \tag{72}$$

where $\mathcal{I}_{\text{BL}}$ encodes linearity in each component:

$$\begin{aligned}(\alpha A_1 + \beta B_1, A_2) &\equiv_{\text{BL}} \alpha(A_1, A_2) + \beta(B_1, A_2)\\ (A_1, \alpha A_1 + \beta B_2) &\equiv_{\text{BL}} \alpha(A_1, A_2) + \beta(A_1, B_2).\end{aligned} \tag{73}$$

We picture additivity in Fig. 29. We write $A_1 \odot A_2$ for the equivalence class $[(A_1, A_2)]_{\mathcal{I}_{\text{BL}}}$. The product is defined component-wise:

$$(A_1 \odot A_2)(B_1 \odot B_2) = A_1 B_1 \odot A_2 B_2. \tag{74}$$

With these definitions, $\mathcal{A}^{(1)} \odot \mathcal{A}^{(2)}$ is a $*$-algebra.

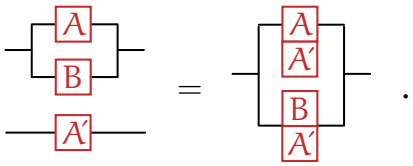


Figure 29: In awds, the tensor product is vertical concatenation, read downwards. Linearity "copies" from one wire into the additive structure of another.

To promote this to a genuine C*-algebra, we need a norm satisfying (4). Here is a place where Hilbert space is very handy! We can use the GNS construction to identify abstract operators $A \in \mathcal{A}$ with bounded operators $\Phi_\pi(A) = A_\pi \in \mathcal{B}(\mathcal{H}_\pi)$, and the tensor product with the familiar Hilbert space tensor product $\otimes$, so

$$\|A_1 \odot A_2\|_{(\pi_1,\pi_2)} = \|\Phi_{\pi_1}(A_1) \otimes \Phi_{\pi_2}(A_2)\|_{\mathrm{op}}, \tag{75}$$

where $\|\cdot\|_{\mathrm{op}}$ is the operator norm. An arbitrary state can lose information about operators, and thereby decrease the operator norm; however, it can never *increase* it. We are thus led to define the *spatial norm* as the supremum of (75) over choice of states:

Since it involves a supremum over vectors, losing vectors only decreases it.

$$\|A_{1(i)} \odot A_{2(i)}\|_{\min} = \sup_{(\pi_1,\pi_2)} \left\{ \|\Phi_{\pi_1}(A_{1(i)}) \otimes \Phi_{\pi_2}(A_{2(i)})\|_{\mathrm{op}} \right\}, \tag{76}$$

where, as usual, we sum over the repeated index i, and unbracketed subscripts denote copies. Finally, we define the *spatial tensor product*

$$\mathcal{A}^{(1)} \otimes \mathcal{A}^{(2)} = \overline{\mathcal{A}^{(1)} \odot \mathcal{A}^{(2)}}^{\|\cdot\|_{\min}}, \tag{77}$$

i.e. the completion of the algebraic tensor product with respect to the spatial norm. It can be shown[27] not only that this is a C*-norm, but the smallest possible norm, hence the subscripted "min".

[27] *C*-Algebras and Finite-Dimensional Approximations* (2008), N. Brown and N. Ozawa.

The state space for $\mathcal{A}^{(1)} \otimes \mathcal{A}^{(2)}$ consists of positive, normalized functionals, with linearity on the tensor product equivalent to linearity in each slot. We can always rescale a positive, normalized product so that the factors are positive and normalized. Hence:

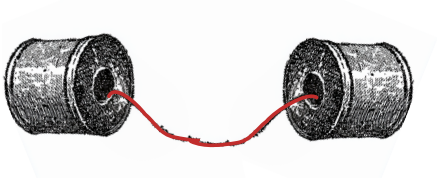
The identity is $I^{(1)} \otimes I^{(2)}$, and since $(\pi_1 \otimes \pi_2)(I^{(1)}I^{(2)}) = \pi_1(I^{(1)})\pi_2(I^{(2)})$, if we rescale π_1 to ensure $\pi_1(I^{(1)}) = 1$, then $\pi_2(I^{(2)}) = 1$. A similar argument works for positivity.

$$S(\mathcal{A}^{(1)} \otimes \mathcal{A}^{(2)}) \cong S(\mathcal{A}^{(1)}) \otimes S(\mathcal{A}^{(2)}). \tag{78}$$

Iterating the tensor product to multiple factors $\mathcal{A} = \bigotimes_{\ell \in \mathfrak{L}} \mathcal{A}^{(\ell)}$ gives an associated state space $S(\mathcal{A}) = \bigotimes_{\ell \in \mathfrak{L}} S(\mathcal{A}^{(\ell)})$. In the language of tensor products, cluster decomposition (71) says that systems with historically well-separated factors have states of the form

$$\pi = \bigotimes_{\ell \in \mathfrak{L}} \pi^{(\ell)}, \tag{79}$$

called *separable* or *product states*. A state which is not separable (with respect to the labelling $\ell \in \mathfrak{L}$) is called *entangled*.



An entangled pair, aka Bell telephones.

To see how it works, let's take two copies of the Pauli algebra. If we choose a state $\pi^{(1)}$ on the first copy, the algebra becomes isomorphic to $M_2(\mathbb{C})$. For a basis matrix $E_{(ab)} = |a\rangle\langle b|$ in the first copy, we

think of $E_{(ab)} \otimes A$ as placing A in entry (a, b), and hence

$$\begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix} \otimes A \cong \begin{bmatrix} \alpha A & \gamma A \\ \beta A & \delta A \end{bmatrix}.$$

This is our usual nested matrix notation. This generalizes so that the tensor product of $\mathcal{A}$ and complex-valued square matrices gives $\mathcal{A}$-valued squares matrices:

$$M_n(\mathbb{C}) \otimes \mathcal{A} \cong M_n(\mathcal{A}). \tag{80}$$

Back to our example, choosing $\pi^{(1)}$ on the second copy makes it, too, equivalent to $M_2(\mathbb{C})$, so we get a matrix-valued matrix equivalent to $M_4(\mathbb{C})$. This identification does not depend on having an entangled state, but simply makes the isomorphism more complicated.

You can literally put an X in your X, so you can swap when you swap.

Just as we concatenate boxes, we concatenate (or even fuse) wires, as in Fig. 29. For instance, consider the *CNOT operator*. We can express this in terms of the *projectors* $E_{(00)} = |0\rangle\langle 0|$ and $E_{(11)} = |1\rangle\langle 1|$ associated with Z. We define

These projectors already showed up as generators of the kernel (38) and Hilbert space (39) for $\pi_{(0)}$.

$$\text{CNOT} = E_{(b\hat{b})} \otimes X^b = |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes X \cong \begin{bmatrix} I & \\ & X \end{bmatrix}, \tag{81}$$

where $\hat{b}$ means that this occurrence of b is exempt from the Einstein convention. CNOT applies X^b to the second qubit if the first is $|b\rangle = X^b|0\rangle$. Since $E_{(b\hat{b})} = \mathcal{C}_{X^b}[|0\rangle\langle 0|]$, we can compactly write

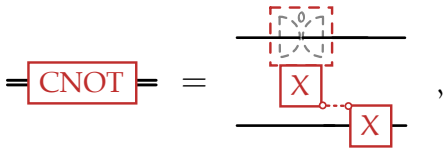


Figure 30: CNOT in awd notation. Index nodes on corners indicate a power X^b. Above, $|0\rangle\langle 0|$ is conjugated by X^b, tethered X^b below. Note that we give a more explicit way to bunch and unbunch wires in §15.

where, by convention, index nodes on box corners stand for powers, and the 0 indicates fiducial state $\pi_{(0)}$.

A crucial application of the CNOT is building entangled states. The *Hadamard operator* $H = \frac{1}{\sqrt{2}}(X + Z)$ is a generalized Pauli operator which obeys $XH = HZ$. Starting both qubits in state $|0\rangle$, applying the Hadamard to the first, then CNOT, and simplifying, gives Fig. 31. There is a fair bit going on in the diagram (see caption for a blow-by-blow summary), but the final result is a vector

Since $X(X + Z) = I + XZ = (Z + X)Z$.

$$|\Psi_{\text{Bell}}\rangle = \frac{1}{\sqrt{2}}|bb\rangle \tag{82}$$

we call the *Bell (vector) state*,[28] or the equivalent functional

$$\pi_{\text{Bell}}(A) = \text{tr}\left[\varpi_{\text{Bell}} A\right], \quad \varpi_{\text{Bell}} = \frac{1}{2}|b\rangle\langle a| \otimes |b\rangle\langle a|. \tag{83}$$

[28] "On the Einstein Podolsky Rosen paradox" (1964), John Stewart Bell; see also "Can Quantum-Mechanical Description of Physical Reality be Considered Complete?" (1935), Einstein, Podolsky and Rosen.

To confirm this state is entangled, there is a nifty shortcut. Recall that the trace (63) was defined for outer products $|\psi\rangle\langle\phi|$ and extended to arbitrary operators by linearity, since the outer products span $\mathcal{B}(\mathcal{H})$. Similarly, we can define the *partial trace* for $A_1 \otimes A_2 \in \mathcal{A}^{(1)} \otimes \mathcal{A}^{(2)}$ by

$$\mathrm{tr}_{(1)}\big[A_1 \otimes A_2\big] = \mathrm{tr}[A_1]A_2, \quad \mathrm{tr}_{(2)}\big[A_1 \otimes A_2\big] = \mathrm{tr}[A_2]A_1, \tag{84}$$

where the subscript indicates the system we trace over. We then extend to arbitrary operators in $\mathcal{A}^{(1)} \otimes \mathcal{A}^{(2)}$ by linearity.

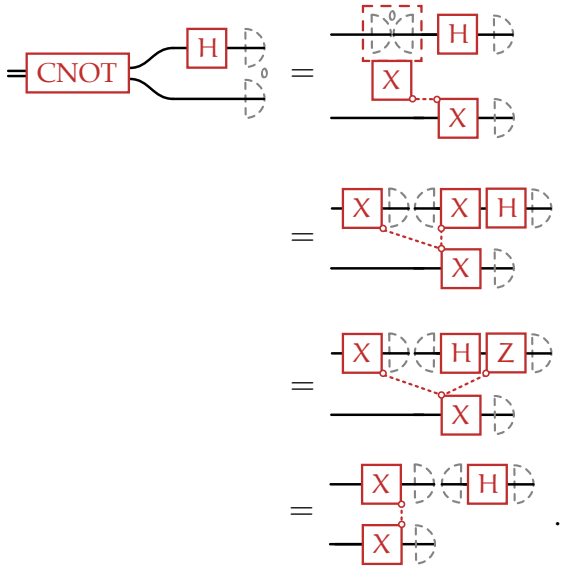


Figure 31: Creation of the Bell state. Line 1. We insert the definition of CNOT from Fig. 30. Line 2. We expand the red conjugator, keeping track of tether lines. Line 3. We use the identity $XH = HZ$. Line 4. We absorb $Z^b|0\rangle = |0\rangle$, leaving $\langle 0|H|0\rangle = \frac{1}{\sqrt{2}}$.

The *reduced density matrix* is the result of performing a partial trace (84) on a density matrix ϖ and seeing what's left over. We write

$$\varpi^{(1)} = \mathrm{tr}_{(2)}[\varpi], \quad \varpi^{(2)} = \mathrm{tr}_{(1)}[\varpi], \tag{85}$$

so $\varpi^{(\ell)}$ traces over the locations $\ell' \neq \ell$. It follows from (84) and (85) that, if $\varpi = \rho_1 \otimes \rho_2$ is separable, then

$$\varpi^{(1)} \otimes \varpi^{(2)} = \mathrm{tr}[\rho_1]\mathrm{tr}[\rho_2]\rho_1 \otimes \rho_2.$$

From (66), it follows that $\mathrm{tr}[\rho_1] = \pi_{\rho_1}(I^{(1)}) = 1$ for any density matrix ρ_1. Thus, a separable density ϖ is the product of its reduced densities:

$$\varpi^{(1)} \otimes \varpi^{(2)} = \rho_1 \otimes \rho_2 = \varpi. \tag{86}$$

Applying this prescription to (83), we find that the reduced density

$$\varpi^{(\ell)}_{\mathrm{Bell}} = \frac{1}{2}\mathrm{tr}\big[|a\rangle\langle b|\big]|a\rangle\langle b| = \frac{1}{2}\delta_{ab}|a\rangle\langle b| = \frac{1}{2}I^{(\ell)}. \tag{87}$$

This is proportional to the identity, and reveals nothing about the state. It is called the *maximally mixed state*, since measurements of it are maximally random, and corresponds to the centre of the Bloch ball. The tensor product of $I^{(1)}$ and $I^{(2)}$ is just the identity $I \in \mathcal{A}^{\otimes 2}_{\mathrm{Pauli}}$, *not* the Bell density, so $|\Psi_{\mathrm{Bell}}\rangle$ is entangled as claimed.

12. *Commuting factors*

Tensor products are a good way to describe systems spread across different locations, when state is separable as in cluster decomposition (71), or entangled like Bell (82). We explained in the former case how to make jointly sharp observables, simply by measuring on each factor. In fact, this remains true even when a state is entangled! But the expectations need not factorize. To illustrate this subtlety, we can use the Bell state (83), and measure $Z \otimes Z$ with average

$$\pi_{\text{Bell}}(Z \otimes Z) = \frac{1}{2}\langle a|Z|b\rangle\langle a|Z|b\rangle = \frac{1}{2}\delta_{ab}\delta_{ab} = 1.$$

On the other hand, the individual measurements $Z^{(1)} = Z \otimes I$ and $Z^{(2)} = I \otimes Z$ have vanishing expectation:

$$\pi_{\text{Bell}}(Z^{(\ell)}) = \frac{1}{2}\langle a|Z|b\rangle\langle a|b\rangle = \frac{1}{2}(-1)^b\langle a|b\rangle\delta_{ab} = 0.$$

Thus, $\pi_{\text{Bell}}(Z^{(1)}Z^{(2)}) \neq \pi_{\text{Bell}}(Z^{(1)})\pi_{\text{Bell}}(Z^{(2)})$, and entanglement manifests as non-factorization.

To see why measuring on each factor nevertheless leads to jointly sharp $Z^{(\ell)}$, we need to explain how to measure a subsystem. We give full details in the next section, but for now, we get by with a rule for *post-selected partial measurement (PPM)*, consisting of two steps. First, for a global state π, measuring location $\ell \in \mathfrak{L}$ factorizes the density into the local reduced density and its complement:[29]

$$\omega_\pi \mapsto \text{tr}_{(\bar{\ell})}\omega_\pi \otimes \text{tr}_{(\ell)}\omega_\pi = \omega_\pi^{(\ell)} \otimes \omega_\pi^{(\bar{\ell})}, \tag{88}$$

generalizing the partial trace notation of (85). For the second stage, measuring $\Lambda \leftarrow \lambda$ on ℓ further modifies (88) to

$$\omega_\pi^{(\ell)} \otimes \omega_\pi^{(\bar{\ell})} \mapsto p_\lambda^{-1}\Pi_\lambda \otimes \omega_\pi^{(\bar{\ell})} \tag{89}$$

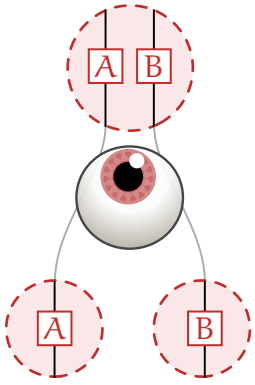
where $p_\lambda = \text{tr}[\omega_\pi^{(\ell)}\Pi_\lambda]$, and Π_λ projects onto the λ eigenspace, i.e.

$$\Pi_\lambda\Lambda = \lambda\Pi_\lambda = \Lambda\Pi_\lambda. \tag{90}$$

Equation (88) is the "partial" and (89) the "post-selected" of PPM. From (88)–(90), we see that measurements of system ℓ disentangle it from $\bar{\ell}$, and no measurement of $\bar{\ell}$ can re-entangle it. Moreoever, the measured local observables $\Lambda^{(\ell)}$ will be sharp, since

$$\pi_{\text{PPM}}^{(\ell)}(\Lambda^k) = p_\lambda^{-1}\text{tr}\left[\Pi_\lambda\Lambda^k\right] = p_\lambda^{-1}\text{tr}\left[\Pi_\lambda\Lambda^k\Pi_\lambda\right] = \lambda^k,$$

and hence $\pi_{\text{PPM}}^{(\ell)}(\Delta\Lambda^2) = \pi_{\text{PPM}}^{(\ell)}(\Lambda^2) - \pi_{\text{PPM}}^{(\ell)}(\Lambda)^2 = 0$. This shows that, given a tensor factorization $\bigotimes_{\ell\in\mathfrak{L}}\mathcal{A}^{(\ell)}$, measurement of a set of local observables $\Lambda^{(\ell)}$ makes them jointly sharp.



Observer-induced factorization.

When we superscript an operator like Z which is associated with a single factor, we will in general tensor it with identities in the remaining locations. This is in contrast to an expression like $\omega^{(\ell)}$, which lives on factor ℓ.

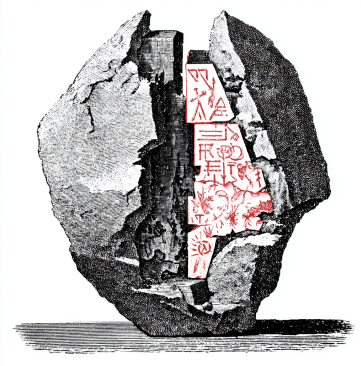
[29] Why? Loosely speaking, because measurement *maximally entangles* the measuring apparatus with system ℓ, and entanglement is *monogamous*. See for instance "Distributed Entanglement" (2000), Coffman, Kundu and Wootters.

Using idempotence of Π_λ, cyclicity of the trace, and the definition of p_λ.

So, like cluster decomposition (71), we seems to have a good way to engineer switches. But also like cluster decomposition, there is a catch: we need our system to be carved into tensor factors *already*. Typically, Nature hands us a big messy algebra $\mathcal{A}$ and *we* do the carving, and since all we have is algebra, we must exploit algebraic clues. Our work with tensor products gives us one such clue: observables associated with different local algebras *commute*, e.g.

$$Z^{(1)}Z^{(2)} = Z \otimes Z = Z^{(2)}Z^{(1)}. \tag{91}$$

We say the local algebras $\mathcal{A}^{(1)}_{\text{Pauli}}$ and $\mathcal{A}^{(2)}_{\text{Pauli}}$ themselves commute, since each observable does. This is an entirely algebraic statement; hopefully it prescribes sensible pieces to carve the algebra into.



To carve an algebra into pieces, we first need to understand its structure.

Let's see what joint measurement of commuting observables yields. Suppose that Λ and Γ are self-adjoint, with vanishing *commutator* $[\Lambda, \Gamma] = \Lambda\Gamma - \Gamma\Lambda = 0$. In the finite-dimensional case, there is a standard argument[30] that they can be simultaneously diagonalized:

$$\Lambda = \sum_{i\in\mathcal{I}} \lambda_i \Pi_i, \quad \Gamma = \sum_{i\in\mathcal{I}} \gamma_i \Pi_i, \tag{92}$$

for some set of orthogonal and hence commuting projectors $\Pi_i, i \in \mathcal{I}$. Thus, any sum of projectors $\Pi = \sum_{i'\in\mathcal{I}'} \Pi_{i'}$, where $\mathcal{I}' \subseteq \mathcal{I}$ is a subset of the full index set, commutes with both Λ and Γ. We prove Λ:

$$\Pi\Lambda = \sum_{i'} \Pi_{i'} \sum_{i} \lambda_i \Pi_i = \sum_{i,i'} \lambda_i \Pi_i \Pi_{i'} = \Lambda\Pi,$$

with a similar result for Γ. In the general case, the same conclusion follows from the continuous functional calculus (Appendix A).

[30] See *Linear Algebra Done Right* (1995), Sheldon Axler, for instance. Instead of defining a unitary U, use project onto the eigenbasis with Π_i directly.

Since $\Pi_i\Pi_{i'} = \delta_{ii'}\Pi_{\hat{i}} = \Pi_{i'}\Pi_i$.

There is a simple generalization of PPM (88)–(90) that captures the structure of measurements in this case. If we measure $\Lambda \leftarrow \lambda$, the associated projector is just the sum of those projectors with the matching eigenvalue. Using the indicator function $\mathbb{I}$, we can write

$$\Pi_\lambda = \sum_{i\in\mathcal{I}} \Pi_i \mathbb{I}[\lambda = \lambda_i], \tag{93}$$

The indicator function $\mathbb{I}[P] = 1$ if the condition P is true, and 0 otherwise.

and note that Λ satisfies (90). Then the post-measurement state is

$$\pi \overset{\lambda}{\longmapsto} \pi' = p_\lambda^{-1} \mathcal{C}^\lambda[\pi], \quad p_\lambda = \pi(\Pi_\lambda), \tag{94}$$

where $\mathcal{C}^\lambda = \mathcal{C}^{\Pi_\lambda}$. We explain in more detail below, but let us check this satisfies our requirements that Λ is sharp with average λ:

$$\pi'(\Lambda^k) = p_\lambda^{-1}\pi(\Pi_\lambda \Lambda^k \Pi_\lambda) = p_\lambda^{-1}\pi(\Pi_\lambda^2 \lambda^k) = \lambda^k p_\lambda^{-1}\pi(\Pi_\lambda) = \lambda^k,$$

where dragging Π_λ through Λ k times creates λ^k via (90). From $k = 1$ we get mean λ, and from $k = 2$ we get sharpness.

Returning to our observables Λ and Γ, suppose we measure $(\Lambda, \Gamma) \leftarrow (\lambda, \gamma)$.[31] The projectors Π_λ, Π_γ, given by (93), are built from the same set of commuting projectors and therefore commute. Let's see what happens when we measure in either order:

$$\begin{aligned} \pi &\xmapsto{\lambda} \pi^\lambda = p_\lambda^{-1} \mathcal{C}^\lambda[\pi] \xmapsto{\gamma} \pi^{\gamma\lambda} = p_{\lambda\gamma}^{-1} p_\lambda^{-1} \mathcal{C}^{\gamma\lambda}[\pi] \\ \pi &\xmapsto{\gamma} \pi^\gamma = p_\gamma^{-1} \mathcal{C}^\gamma[\pi] \xmapsto{\lambda} \pi^{\lambda\gamma} = p_{\gamma\lambda}^{-1} p_\gamma^{-1} \mathcal{C}^{\lambda\gamma}[\pi], \end{aligned} \tag{95}$$

using (43). The superscripts stand for $\Pi_\gamma \Pi_\lambda$ and $\Pi_\lambda \Pi_\gamma$, which are equal as commented above. This means that the states are the same up to a constant, and if normalized, these constants are equal.

We learn that when observables commute, their measurements commute. In contrast, non-commuting observables have values of λ and γ where Π_λ and Π_γ do not commute, so the states in (95) are not equal. This introduces path dependence into the process,

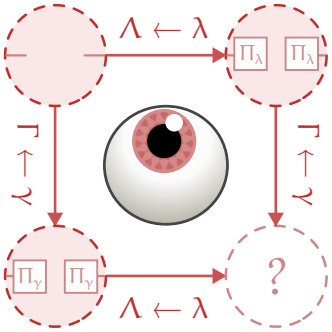


Figure 32: The measurement square associated with two observables. Whether measurements commute (the square closes) depends on whether the observables commute.

so the "measurement square" of Fig. 32 does not close. Take the Pauli algebra, for instance. The projectors for $Z \leftarrow -1$, $X \leftarrow +1$ are

$$\Pi_{Z,-1} = \tfrac{1}{2}(I + Z), \quad \Pi_{X,-1} = \tfrac{1}{2}(I - X).$$

Since X and Z have a nonzero commutator $[Z, X] = 2iY$,

$$\Pi_{Z,-1}\Pi_{X,-1} - \Pi_{X,-1}\Pi_{Z,-1} = -\tfrac{1}{4}[Z, X] = -\tfrac{1}{2}iY.$$

Hence, $Z \leftarrow -1$ does not commmute with $X \leftarrow +1$.

Sometimes we get lucky, and measure two non-commuting observables in such a way that both are sharp. To bound *how* sharp they can simultaneously be, we can use the *Robertson–Schrödinger relation*

$$|\pi([\Lambda, \Gamma])|^2 + 4\,|\pi(\Lambda \circ \Gamma) - \pi(\Lambda)\pi(\Gamma)|^2 \leq 4\|\Delta\Lambda\|_\pi^2 \|\Delta\Gamma\|_\pi^2 \tag{96}$$

proved in Appendix B. The commutator term tells us that, for joint sharpness, Λ and Γ must *commute in expectation*:

$$|\pi([\Lambda, \Gamma])|^2 = 0 \implies \pi(\Lambda\Gamma) = \pi(\Gamma\Lambda). \tag{97}$$

This is equivalent to requiring Π_λ and Π_γ to commute. The anticommutator term, on the other hand, is guaranteed to vanish by (69).

[31] We use a tuple rather than a set because order matters. At a deeper level, this is suggested by the *Curry–Howard correspondence*, but that is beyond the scope of this marginalium. See *Combinatory Logic* (1958), Haskell B. Curry; "The Formulae-as-Types Notion of Construction" (1980), William A. Howard.

Alternatively, you can verify that

$$p_{\lambda\gamma} p_\lambda = p_{\gamma\lambda} p_\gamma = \pi(\Pi_\gamma \Pi_\lambda),$$

so the constants are the same.

Measurement is fundamentally random; to maximize control over sharpness, it's therefore best to work with commuting variables, and focus on *abelian switch sets* $\mathcal{Q}$. Along with the definite set $\mathcal{D}$, we can consider the *abelian* C^**-subalgebra* $\mathcal{M}$ built from elements of $\mathcal{Q}$:

$$\mathcal{D} = J^\circ\langle\mathcal{Q}\rangle, \quad \mathcal{M} = C^*\langle\mathcal{Q}\rangle. \tag{98}$$

$\mathcal{M}$ is also abelian because the property of commuting with everything is closed under linear combinations and products.

A state $\pi(\mathcal{Q})$ associated with measuring all switches in $\mathcal{Q}$ is pure just in case $\mathcal{D}_{\pi(\mathcal{Q})}$ is maximal, by Størmer's theorem; in turn, this is equivalent to $\mathcal{M}_{\pi(\mathcal{Q})}$ being a *maximal abelian subalgebra (MASA)*, since $\mathcal{M}_{\pi(\mathcal{Q})}$ is just the "complexification" of $\mathcal{D}_{\pi(\mathcal{Q})}$.[32]

"Maximal" means nothing else can be added to $\mathcal{M}$ while remaining abelian, and hence the only operators to commute with everything in $\mathcal{M}$ are the elements of $\mathcal{M}$ itself. We capture this neatly with the *commutant* $\mathcal{M}'$, the set of all operators commute with all of $\mathcal{M}$:

$$\mathcal{M}' = \{A \in \mathcal{A} : [M, A] = 0 \text{ for all } M \in \mathcal{M}\}. \tag{99}$$

[32] Complexification just means we replace $\mathbb{R}$ with $\mathbb{C}$; the Jordan product is the same as the regular product when operators commute, $\Lambda \circ \Gamma = \Lambda\Gamma$.

Then $\mathcal{M}$ is abelian just in case $\mathcal{M} \subseteq \mathcal{M}'$, and a MASA when $\mathcal{M} = \mathcal{M}'$. For instance, as we saw in §11, the switch set $\mathcal{Q} = \{\sigma(\mathbf{n})\}$ corresponds to measuring $\sigma(\mathbf{n})$, with definite set and hence MASA

$$\mathcal{D}_{\mathbf{n}} = \text{span}_{\mathbb{R}}\{I, \sigma(\mathbf{n})\}, \quad \mathcal{M}_{\mathbf{n}} = \text{span}_{\mathbb{C}}\{I, \sigma(\mathbf{n})\},$$

since $\mathcal{M}_{\mathbf{n}}$ an abelian subalgebra that complexifies $\mathcal{D}_{\mathbf{n}}$.

Clearly, $\mathcal{M}_{\mathbf{n}}$ is abelian, but it is also maximal since no other $\sigma(\mathbf{m})$ commutes with $\sigma(\mathbf{n})$.

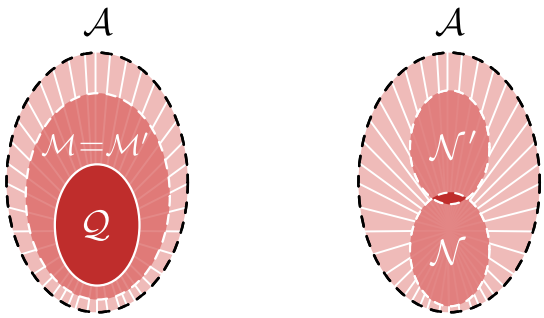


Figure 33: Left. The MASA $\mathcal{M}$ associated to an abelian switch set $\mathcal{Q}$ is its own commutant. Right. In finite dimensions, an arbitrary subalgebra $\mathcal{N}$ is a tensor factor if it spans the full algebra with its commutant; in this case, the overlap is trivial.

Curiously, tensor products like (91) are on the other end of the spectrum. Consider arbitrary subalgebras $\mathcal{N}_1, \mathcal{N}_2 \subseteq \mathcal{A}$, and define their *join* as the subalgebra they generate:

$$\mathcal{N}_1 \vee \mathcal{N}_2 = C^*\langle\mathcal{N}_1 \cup \mathcal{N}_1\rangle. \tag{100}$$

The Bicommutant Theorem of von Neumann[33] implies that, if $\mathcal{N} \vee \mathcal{N}'$ is the full algebra, they tensor factorize $\mathcal{A}$:

$$\mathcal{N} \vee \mathcal{N}' = \mathcal{A} \quad \Longrightarrow \quad \mathcal{A} \cong \mathcal{N} \otimes \mathcal{N}'. \tag{101}$$

[33] "Zur Algebra der Funktionaloperatoren und Theorie der normalen Operatoren" (1929). This result only holds for finite-dimensional systems.

It additionally follows that $\mathcal{N} \cap \mathcal{N}' = \mathbb{C}I$ is trivial. As a sanity check, in $\mathcal{A}_{\text{Pauli}}^{\otimes 2} = \mathcal{A}_{\text{Pauli}} \otimes \mathcal{A}_{\text{Pauli}}$, the first factor $\mathcal{A}_{\text{Pauli}}^{(1)}$ has commutant $\mathcal{A}_{\text{Pauli}}^{(2)}$, they join to give the full space, and overlap trivially. Since it gives reliable switches (via MASAs) and useful algebraic chunks (tensor factors), commutation is clearly the right structural lens.

13. How to measure

We have argued that quantum mechanics is about observables and measurement; states are a bookkeeping device for self-consistently keeping track of things. We've been coy about the measurements themselves, however. We presented post-selected partial measurement in equations (88)–(90), and a general post-selected measurement in (94). But what induces the value we post-select onto? The answer involves *quantum randomness* we have yet to explain.

To rectify this, let's return—as we so often do—to von Neumann, who argued that the true atoms of quantum mechanics are the "yes-no" measurements, or projectors Π satisfying (2). They have a simple binary spectrum, obtained from the defining equation:

$$\Pi^2 - \Pi = \Pi(\Pi - I) = 0 \quad \Longrightarrow \quad \mathfrak{S}(\Pi) = \{0, 1\}. \tag{102}$$

Physically, a projector represents a *filter*, allowing some things through unchanged ("yes", $\Pi \leftarrow 1$) and blocking others ("no", $\Pi \leftarrow 0$). A classic example is a polarizing filter, which lets through only the light aligned with the filter. If we observe the light on the other side, we know that $\Pi \leftarrow 1$; otherwise $\Pi \leftarrow 0$.

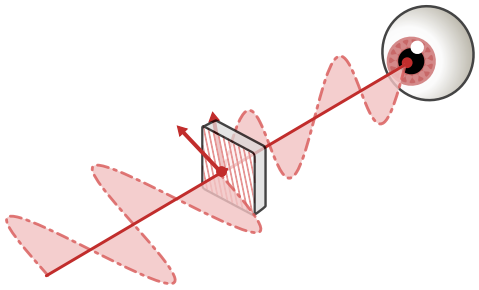


Figure 34: A polarizing filter where light is partially aligned and therefore passes through ($\Pi \to 1$) to the observer.

If we see the light has been polarized, we can post-select, i.e. use our knowledge of the measurement outcome to conditionally update the state. Recall from (28) that, for any state π, we can identify

$$\pi(A) = \pi(\Pi_{\mathcal{K}}^{\perp} A \Pi_{\mathcal{K}}^{\perp}),$$

where $\Pi_{\mathcal{K}}^{\perp}$ projects onto the orthogonal complement of the kernel $\mathcal{K}_\pi$. Once the light passes through the filter, the kernel includes states of light that did *not* pass through, so we update

$$\pi(A) \quad \longmapsto \quad \pi'(A) = p_1^{-1} \pi(\Pi_{\mathcal{K}}^{\perp} \Pi A \Pi \Pi_{\mathcal{K}}^{\perp}),$$

since $\Pi_{\mathcal{K}}^{\perp}$ projects away from the kernel of π, and Π projects us away from the kernel $\Pi^{\perp} = I - \Pi$ of the latest measurement. We're assuming that the filtering doesn't do anything funny to $\Pi_{\mathcal{K}}^{\perp}$; we also need a normalization constant $p_1 = \pi(\Pi)$ so that $\pi'(I) = 1$.

"Not doing anything funny" is an important physical criterion we return to below.

Filtering is not quite measurement in the usual sense: when light doesn't make it through the polarizer, it gets absorbed. This is bad for the lab budget, so instead of throwing it away, we can make a copy and pass it through the second filter if it doesn't make it through the first that it is *guaranteed* to get through. If the first filter is Π, the second, orthogonal filter guaranteed to work is $\Pi^{\perp}$. The measurement outcome tells us which filter we applied.

"Making a copy" really means sampling multiple times from a mixed state, which is a distribution like a coin flip where outcomes are pure states.

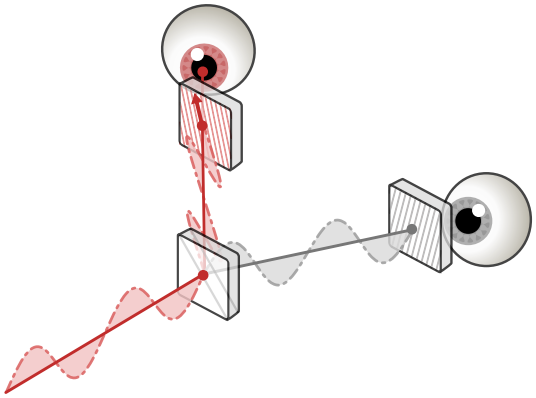


Figure 35: Measuring polarization with orthogonal filters: if it doesn't get through the first, we apply the second.

Suppose we make a measurement, but *don't* observe the outcome, a little like Schrödinger's cat.[34] The state conditioned on filter $\Pi_{(b)}$, where $\Pi_{(0)} = \Pi^{\perp}$ and $\Pi_{(1)} = \Pi$, is

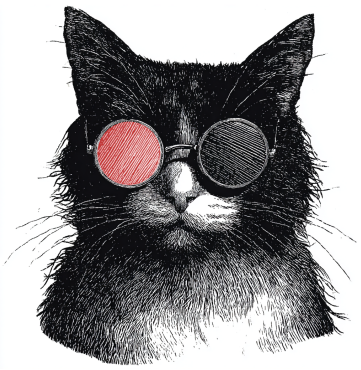
$$\pi_{(b)}(A) = p_b^{-1}\pi(\Pi_{(\hat{b})}A\Pi_{(\hat{b})}). \tag{103}$$

[34] "Die gegenwärtige Situation in der Quantenmechanik" (1935), Erwin Schrödinger.

Thus, the mixed state corresponding to our unseen measurement is

$$\pi' = q_0\pi_{(0)} + q_1\pi_{(1)}, \tag{104}$$

where $q_b = \mathbb{P}[\Pi \leftarrow b]$ is the probability it passes through $\Pi_{(b)}$. Since any complementary probabilities leads to a valid mixed state, it seems like we need to know more about the dynamics to find π'.



Only Schrödinger's cat knows which filter lit up.

The extra piece of information, ironically, is that *we need nothing else*. We assume that Nature is indifferent between the outcomes, in the sense that q_b should only be built out of the ingredient π and Π_b at hand. The unique solution is the normalization constant

$$q_b = \pi(\Pi_{(b)}) = p_b. \tag{105}$$

Why does this follow? By assumption, the only state we have is π, and the only operators are the $\Pi_{(b)}$. The $\pi(\Pi_{(b)}) = p_b$ are complementary, and when $p_b = 0$, setting $q_b = p_b$ cancels the divergence in p_b^{-1}, so we are done. It follows that the measured state is

$$\pi'(A) = \pi(\Pi_{(0)}A\Pi_{(0)}) + \pi(\Pi_{(1)}A\Pi_{(1)}) = \sum_{b=0,1} \mathcal{C}^b[\pi](A), \tag{106}$$

where $\mathcal{C}^b$ conjugates by $\pi_{(b)}$. The full update rule is very simple!

At this point, we should reveal our cards. You may have already recognized (105) as the celebrated BORN RULE[35], while the post-selected update (103) is LÜDERS RULE. The indifference between the eigenvalues of Π that led to Born's rule a baby version of what is called *noncontextuality*—no additional context need apply—and makes our argument a dramatically simplified version of GLEASON'S THEOREM, which derives Born's rule without post-selection.

We can use the awd conjugation notation (Fig. 17) to capture the effect of measurement. Nested circles take states to states, and by convention, circles are normalized. The Born rule is represented by the first circle on the RHS, the Lüders rule by the second:

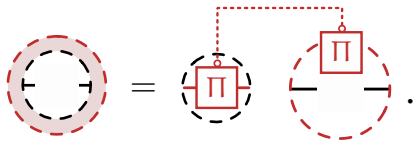


Figure 36: Born and Lüders in awd form. Note that the black leads on the LHS are "transferred" to the red state.

It's easy to generalize all this to *projection-valued measures (PVMs)*. For a complete, orthogonal set of filters labelled by $\lambda \in \mathfrak{F}$, we have

$$\sum_{\lambda\in\mathfrak{F}} \Pi_{(\lambda)} = I, \quad \Pi_{(\mu)}\Pi_{(\lambda)} = \delta_{\mu\lambda}\Pi_{(\hat{\lambda})}. \tag{107}$$

We'll write $\mathcal{C}^{\mathfrak{F}}$ for the sum of filters acting on states by conjugations, and $\mathcal{C}_{\mathfrak{F}}$ for operators. Assuming indifference between eigenvalues, as above, gives

$$\pi' = \mathcal{C}^{\mathfrak{F}}[\pi], \quad \pi_{(\lambda)} = p_\lambda^{-1}\mathcal{C}^{\lambda}[\pi], \quad p_\lambda = \pi(\Pi_{(\lambda)}). \tag{108}$$

This is exactly the Born and Lüders rules suggested by Fig. 36.

This would all be academic if these PVMs were hard to find, but Nature makes them for us! Every self-adjoint operator secretly encodes a set $\mathfrak{F}$ by virtue of the SPECTRAL THEOREM.[36] This guarantees that in a well-behaved Hilbert space $\mathcal{H}$, a bounded, self-adjoint operator $\Lambda \in \mathcal{B}(\mathcal{H})_{sa}$ can be decomposed into an "eigenweighted" PVM:

$$\Lambda = \sum_{\lambda\in\mathfrak{S}(\Lambda)} \lambda\Pi_{(\lambda)}, \quad \Pi_{(\mu)}\Pi_{(\lambda)} = \delta_{\mu\lambda}\Pi_{(\hat{\lambda})}. \tag{109}$$

The index set is the spectrum $\mathfrak{S}(\Lambda) \subseteq \mathbb{R}$. To port this back to an abstract C*-algebra $\mathcal{A}$, we need to ensure our Hilbert space faithfully embeds $\mathcal{A}$. Gelfand and Naimark showed[37] that the direct sum of pure states, identified up to unitary equivalence, is enough:

$$\tilde{\pi} = \bigoplus_{\pi\in \mathrm{Prim}(\mathcal{A})} \pi, \quad \mathrm{Prim}(\mathcal{A}) = \partial S(\mathcal{A})/\mathcal{U}(\mathcal{A}), \tag{110}$$

where $\mathrm{Prim}(\mathcal{A})$ is called the *primitive spectrum* of $\mathcal{A}$. For the Pauli algebra, $\partial S(\mathcal{A}_{\mathrm{Pauli}})$ is the Bloch sphere, and it has a single equivalence class, so any pure state will do.

[35] "On the quantum mechanics of collisions" (1926), Max Born; "Concerning the state-change due to the measurement process" (1950), Gerhart Lüders; "Measures on the closed subspaces of a Hilbert space" (1957), Andrew Gleason.

[36] *Mathematical Foundations of Quantum Mechanics* (1932), John von Neumann. The Hilbert space $\mathcal{H}$ must be *separable*, meaning it has a dense countable subset.

[37] "On the embedding of normed rings into the ring of operators in Hilbert space" (1943), Israel Gelfand and Mark Naimark. Recall that $\partial S(\mathcal{A})$ is the set of extreme points, i.e. pure states, and $\mathcal{U}(\mathcal{A})$ the unitaries which act by conjugation on $S(\mathcal{A})$.

We can spell out in more detail what it means for this construction to "work". The state $\tilde{\pi}$ induces a map $\tilde{\phi} : \mathcal{A} \to \tilde{\phi}(\mathcal{A}) \subseteq \mathcal{B}(\tilde{\mathcal{H}})$, where $\tilde{\mathcal{H}} = \mathcal{H}_{\tilde{\pi}}$. This map $\tilde{\phi}$ has two key properties:

- it is an *isomorphism*, i.e. linear, multiplicative, and invertible;
- it is an *isometry*, so $\|A\| = \|\tilde{\phi}(A)\|_{\mathrm{op}}$ for the operator norm $\|\cdot\|_{\mathrm{op}}$.

Thus, $\tilde{\phi}$ is an *isometric isomorphism* which faithfully embeds both the algebraic and metric structure as a subalgebra of $\mathcal{B}(\tilde{\mathcal{H}})$. For this reason, it is called the *universal representation*. To apply the spectral theorem to $\Lambda \in \mathcal{A}_{\mathrm{sa}}$, we just map to Hilbert space and back:

$$\begin{array}{ccc} \Lambda & \xrightarrow{\text{spec}} & \sum_\lambda \lambda \Pi_{(\lambda)} \\ \tilde{\phi} \downarrow & & \uparrow \tilde{\phi}^{-1} \\ \Lambda & \xrightarrow[\widetilde{\text{spec}}]{} & \sum_{\tilde{\lambda}} \tilde{\lambda} \tilde{\Pi}_{(\tilde{\lambda})} \end{array} . \tag{111}$$

The main obstacle to this argument is that, if $\mathrm{Prim}(\mathcal{A})$ is uncountable, the universal Hilbert space $\tilde{\mathcal{H}}$ is too big for the spectral theorem to hold. Luckily, we are interested in less exotic spaces. We are not so lucky with measurement, where even the simplest experiments are more exotic than the standard formalism.

14. How to err

The spectral theorem associates to each observable Λ a complete, orthogonal set of projectors, or PVM (107). But even for something as simple as a polarizer or a beamsplitter, we need a richer formalism! We call these more general measurements *quantum operations*. They can be incomplete (e.g. a lone polarizer), non-orthogonal (polarizers which are not orthogonal), and non-projective (the beamsplitter itself). We picture some these behaviours in Fig. 37.

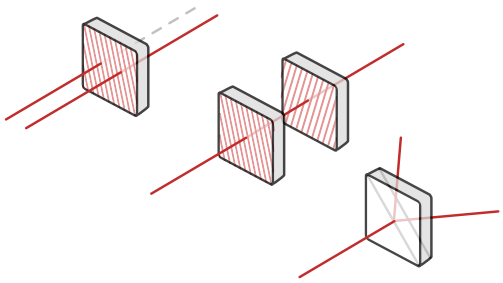


Figure 37: Problems with PVMs. LEFT. Some photons are absorbed. MIDDLE. Some filters are compatible. RIGHT. Some things are not filters.

Happily, the formalism for PVMs generalizes without much effort. Instead of projectors $\Pi_{(\lambda)}$ labelled by $\lambda \in \mathfrak{S}(\Lambda)$, we will suppose a collection of operators $B_{(k)}$ labelled by $k \in \mathfrak{K}$. Our goal will be to figure out what properties the $B_{(k)}$ should have.

First off, we define $\mathfrak{K}$ as a *classical sample space*, from which we randomly select a single element k each time we perform the operation; this is what $\mathfrak{K} \leftarrow k$ means. If k is observed, then the only way to update a state π using $B_{(k)}$ is to conjugate (to preserve positivity) and normalize (to preserve unitality, i.e. mapping $I \mapsto 1$):

There is no loss of generality in sampling a single k. If you want to sample a pair, for instance, you replace $\mathfrak{K}$ by the Cartesian product $\mathfrak{K} \times \mathfrak{K}$.

$$\pi_{(k)} = p_k^{-1} \mathcal{C}^{\hat{k}}[\pi], \quad p_k = \mathcal{C}^k[\pi](I) = \pi(B_{(k)}^* B_{(\hat{k})}). \tag{112}$$

This is a generalization of Lüders rule to arbitrary $B_{(k)}$, called *Kraus operators*.[38] The Born rule is subtler. We start with a general mixture of unobserved operations of the form (112):

[38] See *Effects and Operations: Fundamental Notions of Quantum Theory* (1983), Karl Kraus.

$$\chi_\pi = q_k \pi_{(k)} = q_k p_{\hat{k}}^{-1} \mathcal{C}^k[\pi].$$

Since each term is individually normalized, the whole is normalized provided the q_k sum to unity. But normalizing the state is not necessarily what the Born rule suggests!

This is why we have χ_π rather than π'; we do not assume χ_π is a state.

To see why, consider the case of the lone polarizer. Removing the orthogonal filter from (104) gives a single term, $q_1 \pi_{(1)}$, and normalization would suggest we set $q_1 = 1$. But this means light passes through the filter with certainty, which is nonsense; in fact, we know it gets through with probability $q_1 = \pi_{(1)}(\Pi)$. Extending this, we can make the argument from indifference that

$$\pi \longmapsto \chi_\pi = q_k \pi_{(k)} = \mathcal{C}^{\mathfrak{K}}[\pi], \quad q_k = p_k = \pi(B_{(k)}^* B_{(k)}). \tag{113}$$

The sum of conjugators is called the *Kraus form*. We depict the generalized Born and Lüders rules in Fig. 38.

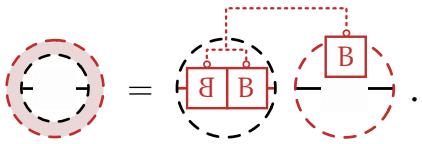


Figure 38: Born and Lüders rules for a quantum operation $\mathcal{C}^{\mathfrak{K}}$.

It's clear that χ_π is a positive linear functional (it is a positive linear combination of positive functionals), but not necessarily normalized. We quantify this with the *trace*:

This equals the trace of the corresponding density matrix. It helps to keep the notation separate even if we muddy the namespace.

$$\tau_{\mathfrak{K}}(\pi) = \chi_\pi(1) = \sum_{k \in \mathcal{K}} p_k = \pi(B_{(k)}^* B_{(k)}), \tag{114}$$

We can interpret $\tau_{\mathfrak{K}}(\pi)$ as the probability we don't destroy our system, like a photon getting absorbed by a polarizer, for input state π. A sensible operation therefore obeys $\tau_{\mathfrak{K}}(\pi) \leq 1$, or

$$\tau_{\mathfrak{K}}(\pi) = \pi(B_{(k)}^* B_{(k)}) \leq 1 \quad \Longrightarrow \quad \pi(I - B_{(k)}^* B_{(k)}) \geq 0.$$

Only positive operators have positive expectation in all states; there is a negative eigenvalue, hence a negative eigenvector. Thus,

$$I - B_{(k)}^* B_{(k)} \geq 0. \tag{115}$$

Together, (112), (113) and (115) define quantum operations.

If $\tau_{\mathfrak{K}}(\pi) = 1$ for all π (i.e. our experiments don't destroy lab supplies) the operation is called a *quantum channel*. Channels linearly map states to states, but not every such map is a channel. An example is provided by the familiar matrix transpose, $A \mapsto A^\top$. To see how it acts on states, we pull it back in the usual way:

$$\pi^\top(A) = \pi(A^\top).$$

The functional $\pi^\top$ is linear by construction, and normalized since $\pi^\top(I) = \pi(I^\top) = 1$. To show that it's positive, we write $A = R^*R$ for some R and transpose. The catch is that, to invoke the usual properties of $^\top$, we must work in the universal representation $\tilde{\mathcal{H}}$ where everything is a matrix. In this case,

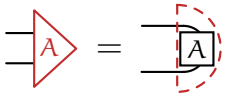
$$A^\top = (R^*R)^\top = R^\top \overline{R} = \overline{R}^* \overline{R},$$

We omit the ~ hat on operators to avoid notational clutter. It's a convenient sin to pretend they are matrices.

which implies that $A^\top$ is positive. This shows the transpose is indeed a linear map from states to states.

Showing that the transpose is not a quantum operation (113) requires more effort. First, we pick an orthonormal basis $\{|i\rangle\}_{i \in \mathcal{I}}$ of the universal Hilbert space $\tilde{\mathcal{H}}$, and form a corresponding basis of outer products $E_{(ij)} = |i\rangle\langle j|$ for $\mathcal{B}(\tilde{\mathcal{H}})$. On these outer products, transposition swaps indices, $E_{(ij)}^\top = E_{(ji)}$, so

$$\Delta = E_{(ij)} - E_{(ji)} \quad \longmapsto \quad E_{(ji)} - E_{(ij)} = -\Delta.$$



We can think of the vectorization $|A\rangle\!\rangle$ as a ket formed by bending a wire through A; we make this precise below.

Since it takes Δ to $-\Delta$, transposition appears to have a negative eigenvalue. To make this idea precise and compare to the general Kraus form (113), we must turn maps between matrices into maps between vectors. A simple option is to flip a bra into a ket:

$$E_{(ij)} = |i\rangle\langle j| \quad \longmapsto \quad |E_{(ij)}\rangle\!\rangle = |i\rangle \otimes |j\rangle.$$

where $|\cdot\rangle\!\rangle$ denotes the *vectorization* or *Choi-Jamiołkowski dual*[39] of a matrix. We can extend this to arbitrary $A = \alpha_{ij} E_{(ij)}$ by linearity:

$$\begin{aligned} |A\rangle\!\rangle = \alpha_{ij}|i\rangle \otimes |j\rangle &= A|j\rangle \otimes |j\rangle \\ &= (A \otimes I)|I\rangle\!\rangle \\ &= (I \otimes A^\top)|I\rangle\!\rangle, \end{aligned} \tag{116}$$

[39] "Linear transformations which preserve trace and positive semidefiniteness of operators" (1972), Andrzej Jamiołkowski; "Completely Positive Linear Maps on Complex Matrices" (1975), Man-Duen Choi.

using $A|j\rangle = A_{ij}|i\rangle$ and $A^\top|i\rangle = A_{ij}|j\rangle$. The vectors $|A\rangle\!\rangle$ live in the Hilbert space $\tilde{\mathcal{H}} \otimes \tilde{\mathcal{H}}$.

To "matricize" a linear transformation $\mathcal{E} : \mathcal{B}(\tilde{\mathcal{H}}) \to \mathcal{B}(\tilde{\mathcal{H}})$, we define the *Choi matrix* $\mathcal{J}(\mathcal{E}) \in \mathcal{B}(\tilde{\mathcal{H}} \otimes \tilde{\mathcal{H}})$ with matrix elements

$$\langle\!\langle E_{(ij)}|\mathcal{J}(\mathcal{E})|E_{(\ell m)}\rangle\!\rangle = \langle\!\langle E_{(ij)}|\mathcal{E}[E_{(\ell m)}]\rangle\!\rangle \tag{117}$$

$$= \langle\!\langle \mathcal{E}^*[E_{(ij)}]|E_{(\ell m)}\rangle\!\rangle. \tag{118}$$

We can now talk rigorously about the eigenvalues of the matrix $\mathcal{J}(\mathcal{E})$ acting on vectorized $|A\rangle\!\rangle$. Letting $\mathcal{E}_\top$ denote the transpose operation, we can rerun our argument from earlier and find that

$$\mathcal{J}(\mathcal{E}_\top)|\Delta\rangle\!\rangle = -|\Delta\rangle\!\rangle.$$

On the other hand, we now show that quantum operations $\mathcal{C}_{\mathfrak{K}}$ in Kraus form correspond to Choi matrices $\mathcal{J}(\mathcal{C}_{\mathfrak{K}})$ with positive eigenvalues. Taking (117) as the definition:

$$\langle\!\langle A|\mathcal{J}(\mathcal{C}_{\mathfrak{K}})|A\rangle\!\rangle = \langle\!\langle A|B^*_{(k)}AB_{(k)}\rangle\!\rangle = \langle\!\langle AB_{(k)}|AB_{(k)}\rangle\!\rangle \geq 0, \tag{119}$$

swapping a Kraus operator over with (118). Thus, $\mathcal{J}(\mathcal{C}_{\mathfrak{K}})$ is a positive matrix with eigenvalues $\lambda \geq 0$, so the transpose cannot be a channel!

To be clear, the adjoint of left-multipication by $B^*_{(k)}$ is right-multiplication by $B_{(k)}$.

We can capture these ideas more perspicuously using awds. Recall the Bell state from Fig. 31, defined as the sum $|\Psi_{\text{Bell}}\rangle = \frac{1}{\sqrt{2}}|bb\rangle$ for $b \in \{0,1\}$. We have generalized to the (non-normalized) sum $|ii\rangle$, $i \in \mathfrak{I}$. Suppose $|i\rangle = U_{(i)}|0\rangle$ for some collection of unitaries $U_{(i)}$. Then we depict the sum

$$|I\rangle\!\rangle = (U_{(i)} \otimes U_{(i)})|00\rangle.$$

simply as a bent line, as in Fig. 39:

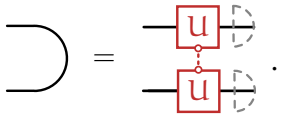


Figure 39: The vectorized identity, aka the Bell state.

The vectorization identities of (116) can then be expressed as

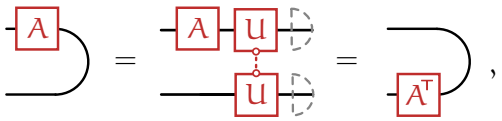


Figure 40: The identities of vectorization $|A\rangle\!\rangle$.

so sliding an operator around the bend is the same as transposing. This is now a vector state, with the Choi matrix capturing the effect of applying $\mathcal{E}$ on the top lead.

Before moving on, we make two diagrammatic observations. First, the fact that the transpose corresponds to sliding an operator around the bend suggests the follow notation:

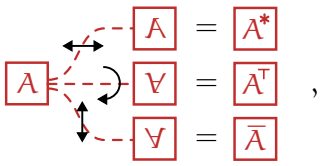


Figure 41: A mapping between involutions of a square and involutions of an operator. See "Efficient decoding for the Hayden-Preskill protocol" (2017), Beni Yoshida and Alexei Kitaev.

This is an isomorphism between symmetries of the square and symmetries of the operator, in the sense that, e.g. a rotation following by a horizontal reflection equals a vertical reflection, and $(A^\top)^* = \overline{A}$.

Second, we already saw the bend of Fig. 39 appear in the trace of Fig. 25. If the notation is consistent, we should be able to write

$$\mathrm{tr}[\varpi A] = \langle\!\langle I|\varpi I\rangle\!\rangle. \tag{120}$$

We can easily check this is true:

$$\langle\!\langle I|\varpi A\rangle\!\rangle = \langle ii|(\varpi A \otimes I)|jj\rangle = \delta_{ij}\langle i|\varpi A|j\rangle = \mathrm{tr}[\varpi A],$$

using the expressions in (65). A cute observation: the second line enforces δ_{ij}, so we our sum $\langle i|\varpi A|i\rangle = \mathcal{C}_{|i\rangle}[\varpi A]$ takes the form of a sum of conjugators, but with bras and kets! This means that the trace is itself a quantum operation $\mathcal{C}_{\mathfrak{I}}$, and in particular a channel since the density matrix ϖ is normalized. The same is true of the partial trace, since $\mathfrak{I}$ can label the basis of a tensor factor rather than the full space. We will see in a moment how viewing states as channels is the natural perspective.

A map with a positive Choi matrix is called *completely positive (CP)*. This equivalence between CP maps and Kraus forms, first proved by Choi, leads to the terminology *completely positive trace-preserving (CPTP) map* as a synonym for "channel". Though it is sometimes presented as a physical necessity, CP violations result from certain physical interventions, just like the loss of purity or trace:

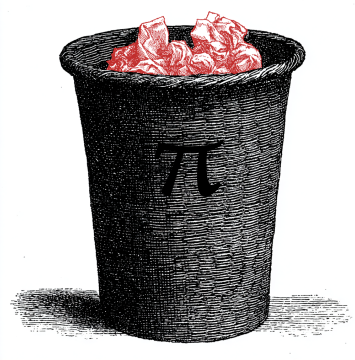
physical intervention	mathematical effect
add classical randomness	pure states become mixed
destroy subsystems	trace not conserved
reduced dynamics with correlation	CP violations

In particular, when a system is connected to an environment, it is easy for the *reduced* system dynamics to be non-CP in the presence of correlations with the environment.[40]

While complete positivity is a convenience rather than a necessity, this convenience is not to be underestimated. A prime example is the *Stinespring dilation theorem*,[41] generalizing both the GNS construction and Choi's theorem. To motivate this, recall that a state $\pi : \mathcal{A} \to \mathbb{C}$ is associated with a Hilbert space $\mathcal{H}_\pi$ where vanishing correlations have been quotiented out. The algebra $\mathcal{A}$ acts via matrices $\Phi_\pi(A)$ on $\mathcal{H}_\pi$, and to take expectations, we can either directly evaluate $\pi(A)$, or in the Hilbert space $\mathcal{H}_\pi$, compute $\mathbb{E}[\Phi_\pi(A)] = \langle 0|\Phi_\pi(A)|0\rangle$ where $|0\rangle$ is our fiducial representative satisfying $\langle 0|0\rangle = 1$. Thus, we find

$$\pi(A) = \langle 0|\Phi_\pi(A)|0\rangle \quad\Longrightarrow\quad \pi = \mathcal{C}_{|0\rangle} \circ \Phi_\pi, \tag{121}$$

since the LHS holds for arbitrary A. We can represent this more



States can be compromised in various ways: classically mixed, partially lost, or correlated with the environment.

[40] On initial environment-system correlations and reduced dynamics, see "Reduced Dynamics Need Not Be Completely Positive" (1994), Philip Pechukas Hi! and "Dynamics beyond completely positive maps: some properties and applications" (2006), Hilary Carteret, Daniel Terno, and Karol Życzkowski.

[41] "Positive Functions on C*-Algebras" (1955), W. Forrest Stinespring; *Effects and Operations: Fundamental Notions of Quantum Theory* (1983), Karl Kraus.

explicitly with domains and codomains shown:

$$\begin{array}{ccc} \mathcal{A} & \xrightarrow{\Phi_\pi} & \mathcal{B}(\mathcal{H}_\pi) \\ & \searrow^{\pi} & \downarrow{\scriptstyle \mathcal{C}_{|0\rangle}} \\ & & \mathbb{C} \end{array} .$$

Stinespring's theorem states that, for CP map $\mathcal{E} : \mathcal{A} \to \mathcal{B}(\mathcal{H})$, there is an associated Hilbert space $\mathcal{H}_\mathcal{E}$, representation $\Phi_\mathcal{E} : \mathcal{A} \to \mathcal{B}(\mathcal{H}_\mathcal{E})$, and map $V : \mathcal{H} \to \mathcal{H}_\mathcal{E}$, exhibiting the same structure $\mathcal{E}[A] = V^* \Phi_\mathcal{E}(A) V$, or

$$\begin{array}{ccc} \mathcal{A} & \xrightarrow{\Phi_\mathcal{E}} & \mathcal{B}(\mathcal{H}_\mathcal{E}) \\ & \searrow^{\mathcal{E}} & \downarrow{\scriptstyle \mathcal{C}_V} \\ & & \mathcal{B}(\mathcal{H}) \end{array} . \tag{122}$$

If $\mathcal{E}$ is unital, with $\mathcal{E}(I) = I_\mathcal{H}$, then V is an *isometry* $V^*V = I_\mathcal{H}$, replacing the condition $\langle 0|0\rangle = 1$. The construction is similar in spirit to GNS (§5).[42]

The appendix also shows how Stinespring's theorem generalizes Choi's equivalence result.

[42] For details, see Nielsen and Chuang (2000) or Kadison and Ringrose (1983).

Thus, a channel $\mathcal{E}$ can always be decomposed into (a) a representation $\Phi_\mathcal{E}$ of $\mathcal{A}$ acting on an auxiliary Hilbert space $\mathcal{H}_\mathcal{E}$; and (b) conjugation by an isometry V from the target Hilbert space $\mathcal{H}$ to the auxiliary $\mathcal{H}_\mathcal{E}$. So much for the math. Physically, we act on some bigger system ($\mathcal{H}_\mathcal{E}$) but observe only the smaller one ($\mathcal{H}$). Instead of dwelling further on the subtleties of information loss, we turn to something a little more exciting: how to win it back again.

If we replace the isometry V with a projector Π, we leak probability; if we replace conjugation $\mathcal{C}_V$ with another map, we lose complete positivity.

15. How to recover

Quantum information is fragile. By performing operations (or more general non-positive updates), we can easily mix, absorb and entangle observables with the environment. In order to compute usefully and not blow our lab budget, we need ways to keep information in the system. We visualize the high-level strategy in Fig. 42 (left):

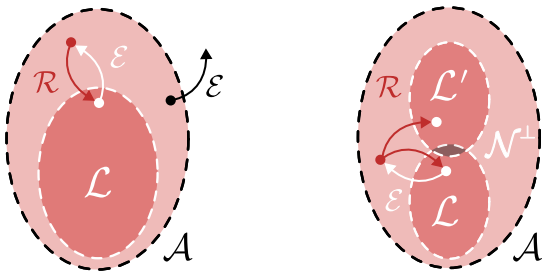


Figure 42: Left. The error process $\mathcal{E}$ kicks the black data point out of the system. The white point is perturbed but can be recovered with $\mathcal{R}$. Right. A weaker condition for recovery is that $\mathcal{R}$ restores a linear combination of the original point and an operator in the commutant.

Under a physical *error process* $\mathcal{E}$, with Kraus operators $E_{(k)}$ for $k \in \mathfrak{K}$, some data will be removed from the system altogether and become unrecoverable. But there may be a smaller, protected subsystem

called the *logical subalgebra* $\mathcal{L} \subseteq \mathcal{A}$ where information remains recoverable in principle. The map $\mathcal{R}$ that restores the data back to its rightful place is called the *recovery map*.

Quantum error correction (QEC) is the art and science of building recovery maps. Let's try to fill in the details. For any $A \in \mathcal{L}$, we should be able to undo the error channel with the recovery map. Equivalently, $\mathcal{R}$ is the left inverse of $\mathcal{E}$ on $\mathcal{L}$:

$$(\mathcal{R} \circ \mathcal{E})[A] = A \quad \Longrightarrow \quad (\mathcal{R} \circ \mathcal{E})|_{\mathcal{L}} = \mathrm{id}_{\mathcal{L}}, \tag{123}$$

where $\mathrm{id}_{\mathcal{L}} = \mathcal{C}_{I_{\mathcal{L}}}$ is the identity channel. But there is a subtlety lurking in the restriction $|_{\mathcal{L}}$. Fig. 42 (right) shows a situation where $(\mathcal{R} \circ \mathcal{E})[A]$ has a component *outside* $\mathcal{L}$. This shouldn't matter, since all we care about is behaviour *inside* $\mathcal{L}$. Let's parse this more carefully.

To start with, assume $\mathcal{L}$ and $\mathcal{L}'$ join to give the full algebra $\mathcal{A}$. By von Neumann's theorem (101), they are tensor factors $\mathcal{A} \cong \mathcal{L} \otimes \mathcal{L}'$, so we can ignore $\mathcal{L}'$ using the partial trace. In general, $\mathcal{L} \vee \mathcal{L}' = \mathcal{N}$ is some subalgebra of $\mathcal{A}$, leaving a non-unique complement $\mathcal{N}^{\perp}$ so that

$$\mathcal{A} = \mathcal{L} \otimes \mathcal{L}' \oplus \mathcal{N}^{\perp}. \tag{124}$$

In this case, we need to restrict to the tensor product with a projector $\Pi_{\mathcal{N}}$ before we can use the partial trace. This modifies (126) to

$$\mathrm{tr}_{\mathcal{L}'}\left[\Pi_{\mathcal{N}}(\mathcal{R} \circ \mathcal{E})[A]\Pi_{\mathcal{N}}\right] = A \text{ for all } A \in \mathcal{L}, \tag{125}$$

or slightly more cleanly, $(\mathbb{E}_{\mathcal{L}} \circ \mathcal{R} \circ \mathcal{E})\big|_{\mathcal{L}} = \mathrm{id}_{\mathcal{L}}$, where $\mathbb{E}_{\mathcal{L}} = \mathrm{tr}_{\mathcal{L}'} \circ \mathcal{C}_{\Pi_{\mathcal{N}}}$. Showing the source and target of maps:

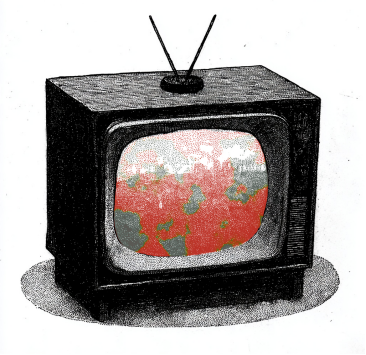
$$\begin{array}{ccc} \mathcal{L} & \xrightarrow{\ \mathcal{E}\ } & \mathcal{A} \\ {\scriptstyle \mathrm{id}}\downarrow & & \downarrow{\scriptstyle \mathcal{R}} \\ \mathcal{L} & \xleftarrow[\ \mathbb{E}_{\mathcal{L}}\ ]{} & \mathcal{A} \end{array}. \tag{126}$$

When $\mathcal{R}$ exists, we say $\mathcal{E}$ is *correctable* on $\mathcal{L}$.

What are some reasonable conditions for correctability? The projective measurements from §13 provide a helpful illustration. The spectral theorem gives rise to *orthogonal* projectors; this is helpful because it uniquely tells you which filter was applied. On the other hand, there is no way correct wavefunction collapse since (nontrivial) projectors cannot be inverted; the cat cannot be revived.[43] We can write orthogonality as

$$\Pi^{*}_{(\lambda)}\Pi_{(\mu)} = \delta_{\lambda\mu}\Pi_{(\hat{\lambda})}.$$

This suggests correctability requires two things. First, the Kraus operators $E_{(k)}$ should be injective when restricted to $\mathcal{L}$, as in Fig. 43 (left); we take "injective" to mean our Kraus operators are *scaled*



Error correction is an attempt to change a noisy channel.

We can define the complement $\mathcal{N}^{\perp}$ as the subspace of lowest dimension such that $\mathcal{N} \oplus \mathcal{N}^{\perp} = \mathcal{A}$. Unlike an orthogonal complement, it is not unique since it can "tilt".

[43] For a formal proof, see "On the Hardness of Detecting Macroscopic Superpositions" (2020), Scott Aaronson, Yosi Atia, and Leonard Susskind.

isometries. Second, that they are orthogonal, so we know which error "fired". We write both conditions in one fell swoop:

$$\mathbb{E}_{\mathcal{L}}[E^*_{(k)}E_{(q)}] = \nu_k \delta_{kq} I_{\mathcal{L}},$$

where $\nu_k > 0$ allows errors to rescale parts of the space.

This is close, but a little too strong: we don't actually need the $E_{(k)}$ to be orthogonal, provided they *overlap perfectly*, as depicted in Fig. 43 (right). We can apply the same correction to different errors!

In other words, $E_{(k)}|_{\mathcal{L}} = \sqrt{\nu_k} V_{(k)}$ for $V_{(k)} : \mathcal{L} \to \mathcal{A}$ satisfing $V^*_{(k)} V_{(k)} = I_{\mathcal{L}}$.

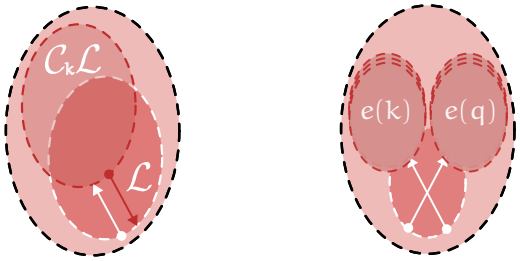


Figure 43: The KL criterion for operator error correction. Left. Errors shift $\mathcal{L}$ invertibly without collapsing it. Right. Errors must either be disjoint or identical, in which case they form a little stack called a syndrome.

These error classes are called *syndromes*, and we can define a function $e : \mathfrak{K} \to \mathfrak{E}$ which assigns a Kraus index $k \in \mathfrak{K}$ to a syndrome $e(k) \in \mathfrak{E}$. Then we can write our guess at as

$$\mathbb{E}_{\mathcal{L}}[E^*_{(k)}E_{(q)}] = \nu_{e(k)} \delta_{e(\hat{k})e(q)} I_{\mathcal{L}}. \tag{127}$$

These are the *Knill-Laflamme (KL) conditions*,[44], and it can be shown they are both necessary and sufficient for correctability. It is usually written in the equivalent, but basis-independent, form

$$\mathbb{E}_{\mathcal{L}}[E^*_{(k)}E_{(q)}] = \nu_{kq} I_{\mathcal{L}}, \tag{128}$$

with ν_{kq} a positive *syndrome matrix*. Since positive matrices are unitarily diagonalizable, this recovers (127).

[44] "Theory of Quantum Error Correction for General Noise" (2000), Emanuel Knill, Raymond Laflamme and Lorenza Viola. Necessity in the case of OQEC was shown in "Unified and Generalized Approach to Quantum Error Correction" (2005), David Kribs, Raymond Laflamme, David Poulin; sufficiency in "Algebraic and information-theoretic conditions for operator quantum error-correction" (2005), Michael Nielsen and David Poulin.

Before we turn to explicit recovery channels, we need to add to our visual toolbox. The algebraic structure (124) relevant to OQEC involves both direct sums and tensor products, so our diagrams must accomodate both. In Fig. 44, we show our proposal. We have tensor and direct sum "fan-outs", indicated by $\oplus$ and $\otimes$, and optional "fan-ins" indicated by $\bullet$. Below, the initial direct sum splits the algebra into $\mathcal{A} = \mathcal{N} \oplus \mathcal{N}^\perp$, and the tensor product splits $\mathcal{N} = \mathcal{L} \otimes \mathcal{L}'$:

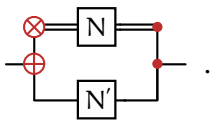


Figure 44: Fan-out notation for direct sums and tensor products.

This helps us illustrate the difference between projection and partial trace (Fig. 45). Loosely speaking, projection is the inverse of direct sums, while the partial trace is the inverse of a tensor product.

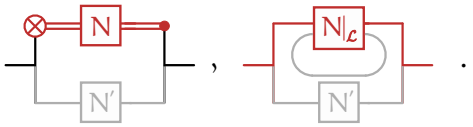


Figure 45: Undoing the operations of Fig. 44. LEFT. Projection isolates components of a direct sum. RIGHT. The partial trace isolates components of a tensor product.

We encode the correctability condition (125) using fan-out notation in Fig. 46. The error channel $\mathcal{E}$ is nested inside the recovery map $\mathcal{R}$; since the channels are linear, they split over the direct summands, and we can ignore their effect on $\mathcal{N}^\perp$ when we ignore $\mathcal{N}^\perp$. We then loop out $\mathcal{L}'$, leaving the operator A on the $\mathcal{L}$ wire trapped.

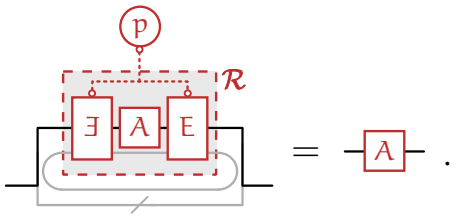


Figure 46: The definition of a correctable channel. We suppress the explicit, state-dependent form of the Kraus coefficients $p_k = \mathcal{C}^{(k)}[\pi]$.

A recovery map $\mathcal{R}$ frees A from this somewhat elaborate cage. The basis-independent KL conditions (128) for the existence of $\mathcal{R}$ are expressed more compactly in Fig. 47:

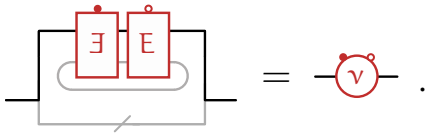


Figure 47: Basis-independent KL conditions for operator error correction.

It's not at all obvious from these diagrams that Fig. 47 should be equivalent to the existence of a box that realizes Fig. 46, but a purely diagrammatic proof is possible.

Given a set of Kraus operators $E_{(k)}$ satisfying the KL conditions, it's easy to build the recovery map: if we measure $\mathcal{E} \leftarrow k$, we apply some local inverse $E^*_{(k')}$ where $e(k') = e(k)$ is some fiducial representative from the syndrome $e(k)$. In other words,

$$\mathcal{R} = \mathcal{C}_{\mathfrak{E}} = \sum_{e \in \mathfrak{E}} \mathcal{C}_e, \quad \mathcal{C}_e[A] = E_e^* A E_{\hat{e}}. \tag{129}$$

What is much less obvious is how to construct channels satisfying (128) and match them to the real-life patterns of computational error. We'll leave the error characterization to the experimentalists, but construct general *quantum error-correcting codes (QECCs)*, which explicitly realize (128), in the next section.

16. How to communicate

We started with Shannon's graduate work on digital circuits. It is fitting that we finish with his "mature" work on the theory of communication. Published in 1948 in the Bell Labs technical journal and

turned into a book with Warren Weaver the next year,[45] it is widely regarded as the founding document of the digital age. One of its key contributions is a formal picture of communication called the *Shannon* or *Shannon-Weaver model*, shown in Fig. 48:

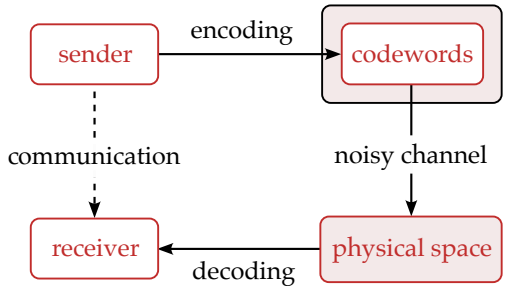


Figure 48: Components of the Shannon-Weaver model.

Although largely self-explanatory, we give a brief gloss. The sender wants to convey a message in some source alphabet to the receiver; they must use a noisy channel, so their strategy is to first encode letters into longer strings called codewords, which live in some protected logical region of physical message space. Codewords are transmitted through the noisy channel and decoded at the receiving end.

The *repetition code* illustrates these ideas in a crucial way. Suppose someone wants to send a sequence of binary digits $b \in \mathbb{B}$, and encodes each bit with its n-fold repetition $c_n(b) \in \mathbb{B}^n$:

$$b \mapsto c_n(b) = bbb\ldots b = b^n.$$

The strings $c_n(b) = b^n$ are the codewords. A reasonable choice of decoder is majority vote for each n-bit block, or equvalently, the closet codeword in Hamming distance

$$c \mapsto d_n(c) = \arg\min_b \|c_n(b) - c\|_1.$$

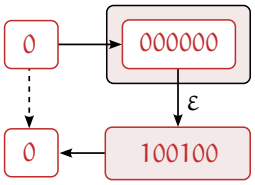
If bits are independently corrupted with probability ε, then applying Hoeffding's inequality[46] to the number of bitflips gives the following bound on decoding errors:

$$\mathbb{P}[\text{decoding error}] \leq e^{-n\left(\frac{1}{2}-\varepsilon\right)}.$$

Provided $\varepsilon < \frac{1}{2}$, the probability of such an error decreases exponentially with n, and can be made arbitrarily unlikely.

The repetition code is simple but deeply instructive. We adapt it to the quantum setting by first making some general observations. First, the decoding step suggests that, instead of identifying a bit with a point $c_n(b) \in \mathbb{B}^n$, we should morally view it as the *code neighbourhood* $\mathbb{H}_b^n = d_n^{-1}(b)$. Below (Fig. 49), we illustrate with neighbourhoods for the $n = 3$ repetition code. The codewords are important,

[45] "A Mathematical Theory of Communication" (1948), Claude E. Shannon; *A Mathematical Theory of Communication* (1949), Claude E. Shannon and Warren Weaver.



Shannon-Weaver model of repetition.

The *Hamming norm* is another name for the ℓ_1 norm on $\mathbb{B}^n$.

[46] "Probability inequalities for sums of bounded random variables" (1963), Wassily Hoeffding. Here, we apply the inequality to the number of bitflips, which for our channel is binomially distributed with mean $n\varepsilon$.

but their role is to define the boundaries of the neighbourhood: like a seed in a Voronoi diagram,[47] we partition the hypercube according to whichever codeword is closest.

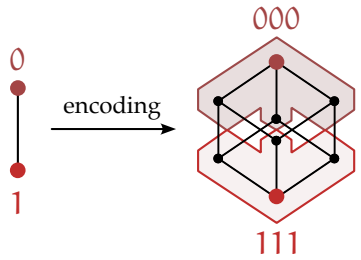


Figure 49: The repetition codes maps bits b to complementary code neighbourhoods $\mathbb{H}^n_b$ on a hypercube $\mathbb{B}^n$.

The second observation is that this partition is *symmetric*. Each code neighborhood has the same size and shape, so we can factorize the auxiliary physical space $\mathbb{B}^n$ into a product of the "Hamming ball" $\mathbb{H}^n$, and the source alphabet $\mathbb{B}$, as in Fig. 50:

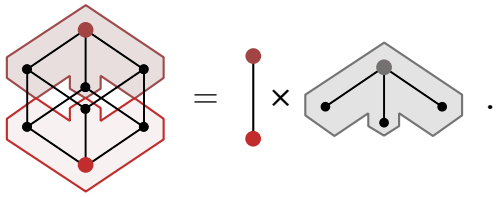


Figure 50: Factorizing a hypercube into the Booleans and a "Hamming ball" $\mathbb{H}^n$, centered on a codeword.

As an equation, we can write

$$\text{physical space} = \text{alphabet} \times \text{code neighbourhood}.$$

The alphabet is the set of things we want to be able to send. The neighbourhood tells us the sorts of errors we can make without corrupt the information we were trying to encode. The physical message space is the product of the two.

Let's now extend the reptition code from classical alphabets to "quantum" alphabets. Instead of classical letters, we can send letters from some alphabet of *operators* $\Sigma \subseteq \mathcal{A}$. For instance, we could try to send Pauli operators $\Sigma = \{X, Z\} \subseteq \mathcal{A}_{\text{Pauli}}$, and choose

$$B \mapsto C_n[B] = B^{\otimes n} \in \mathcal{A}_{\text{phys}} = \mathcal{A}_{\text{Pauli}}^{\otimes n},$$

for $B \in \Sigma$. In general, $\mathcal{A}_{\text{phys}}$ will denote the *physical algebra* we encode our operators in. The decoding step is more involved. We choose neighbourhoods shaped like a *group*.[48] For now, we'll just define a *stabilizer group* as a subset $\mathcal{S} \subseteq \mathcal{U}(\mathcal{A}^{\otimes n})$ such that

- *closure*: if $S, S' \in \mathcal{S}$ then $S \cdot S' \in \mathcal{S}$, or equivalently $\mathcal{S} \cdot \mathcal{S} = \mathcal{S}$;
- *inversion*: for $S \in \mathcal{S}$, then $S^* = S^{-1} \in \mathcal{S}$, or equivalently $\mathcal{S}^* = \mathcal{S}$;
- *commutativity*: for $S, S' \in \mathcal{S}$, $SS' = S'S$.

The name comes from the fact that it stabilizes (leaves invariant) the shape of the neighbourhood.

[47] "Nouvelles applications des paramètres continus à la théorie des formes quadratiques" (1908), Georges Voronoi.

[48] For a short introduction to group theory, see e.g. *Algebra* (1998), Michael Artin.

Suppose that the stabilizer group is generated by a set of operators $\mathcal{X}_\mathcal{S}$. Since $\mathcal{S}$ is commutative, each generator $S \in \mathcal{X}_\mathcal{S}$ commutes with every other, so we can simultaneously measure them. We want to make sure these measurements also don't disturb the information we're trying to send, so each S must commute with every element of Σ. The set of operators that multiplicatively commute with every element of $\mathcal{S}$ is called the *centralizer*, and denoted

$$Z(\mathcal{S}) = \left\{ A \in \mathcal{A}_{\text{phys}}^\times : [A, S]_\times = I \text{ for all } S \in \mathcal{S} \right\}. \tag{130}$$

By "linear generation", we mean $\langle \mathcal{X}_\mathcal{S} \rangle_{\text{phys}}^\times = \mathcal{S}$, where $\langle \cdot \rangle_{\text{phys}}$ is the subalgebra of $\mathcal{A}_{\text{phys}}$ generated by the argument, and superscript $\times$ indicates invertible elements.

The commutant is closely related to the centralizer; in fact, it generates it in the sense that $\langle \mathcal{A}' \rangle^\times = Z(\mathcal{A})$.

So, for stabilizer measurements to play nicely with encoding, we need the centralizer to contain our alphabet, $\Sigma \subseteq Z(\mathcal{S})$.

We can expand Σ to a group $\Sigma^\times$ by multiplying and taking inverses. This group still lives inside the centralizer, $\Sigma^\times \subseteq Z(\mathcal{S})$, since it was made of things that commuted with $\mathcal{S}$. But if $\Sigma^\times$ is strictly smaller than $Z(\mathcal{S})$, there is arbitrage; we can either transmit more operators (make Σ bigger) or make more mutually compatible measurements (make $Z(\mathcal{S})$ smaller). So let's assume they are equal! We let $\Pi_\mathcal{S}$ denote the projector onto the subspace fixed by $\mathcal{S}$, usually called the *code subspace*:

$$\Pi_\mathcal{S} = \frac{1}{|\mathcal{S}|} \sum_{S \in \mathcal{S}} S. \tag{131}$$

We prove this has the desired effect in Appendix B.

So, we have operators Σ we can send and measurements $\mathcal{X}_\mathcal{S}$ we can make. What errors can we correct? We will show that the KL criterion (127) is satisfied when the component Kraus operators E of an error channel $\mathcal{E}$ *projectively commute* with each element of $\mathcal{S}$, i.e. they commute up to a phase. We call the set of all such elements the *projective centralizer*:

$$PZ(\mathcal{S}) = \{ E \in \mathcal{A}_{\text{phys}}^\times : [E, S]_\times \in U(1) I \text{ for all } S \in \mathcal{S} \}. \tag{132}$$

We summarize the phases with an E-dependent function

$$[E, S]_\times = \chi_E(S) I, \quad \chi_E(S) \in U(1), \tag{133}$$

which we call a *stabilizer character*. It follows that χ_E is multiplicative:

$$\chi_E(SS') = [E, SS']_\times = [E, S]_\times [E, S']_\times = \chi_E(S)\chi_E(S'). \tag{134}$$

We let $\widehat{\mathcal{S}}$ denote the set of all multiplicative functions $\chi : \mathcal{S} \to U(1)$. This can be turned into a group (isomorphic to $\mathcal{S}$ itself[49]) with group operation the pointwise product, $[\chi\chi'](s) = \chi(s)\chi'(s)$.

As with the Jordan characters (69) we encountered earlier, multiplicativity corresponds to well-defined measurements, in this case of the elements of $\mathcal{S}$. The moral is that correctable errors are *patterns*

Identifying distinct characters.

[49] By Pontryagin duality. See e.g. *Fourier Analysis on Groups* (1962), Walter Rudin.

of sharp stabilizer measurement. These patterns are slightly redundant, due to $\mathcal{S}$ itself. Suppose $E' = ES$ for some $S \in \mathcal{S}$, with associated characters $\chi_{E'}$ and χ_E. We find

$$\begin{aligned}\chi_{E'}(S') = [E', S']_\times &= [ES, S'] \\ &= ES'S(S'ES)^{-1} \\ &= \chi_E(S')S'ES(S'ES)^{-1} = \chi_E(S'),\end{aligned}$$

using commutativity on the middle line. Since this holds for arbitrary S', the characters are equal, $\chi = \chi'$. These operators are in the same coset $E\mathcal{S}$; it turns out this condition is not only sufficient, but necessary for equality of characters, so distinct characters correspond to the quotient

This follows from the fact that only one character is trivial.

$$\mathfrak{E} = PZ(\mathcal{S})/\mathcal{S}, \tag{135}$$

where $\mathfrak{E}$ anticipates that these will serve as our syndromes. We now have the stabilizer genuinely serving as a neighbourhood, just like the repetition code, though there is no reason for the associated alphabet $\mathfrak{E}$ to equal Σ, or equivalently, for $PZ(\mathcal{S})/\mathcal{S} \cong Z(\mathcal{S})$.

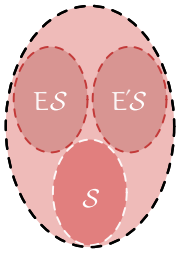
Let's verify the Knill-Laflamme conditions are met. Since characters are multiplicative and projectors idempotent, evaluating the latter on the former gives

$$\chi(\Pi_\mathcal{S}) = \chi(\Pi_\mathcal{S}^2) = \chi(\Pi_\mathcal{S})^2 \quad \Longrightarrow \quad \chi(\Pi_\mathcal{S}) = 0, 1 \tag{136}$$

for all $\chi \in \widehat{\mathcal{S}}$. The only way to obtain unity in (136) is to have constant $\chi(S) = 1$, since otherwise we get less than 1. But since the only other option is 0, the sum over any nonconstant character vanishes! Earlier, we considered a physical algebra of the form

$$\mathcal{A}_{\text{phys}} = \mathcal{L} \otimes \mathcal{L}' \oplus \mathcal{N}^\perp.$$

In this case, our physical space takes the form $\mathcal{A}_{\text{phys}} = \mathcal{L} \oplus \mathcal{L}^\perp$ for $\mathcal{L} = Z(\mathcal{S})$, so we need only use the projector $\mathbb{E}_\mathcal{S} = \mathcal{C}_{\Pi_\mathcal{S}}$. We can explicitly describe the remaining summands $\mathcal{L}^\perp$ using the *Peter-Weyl decomposition*.[50] Projecting onto these instead of the centralizer gives rise to *Clifford codes*.



A cartoon version of (135): we split the projective centralizer into blobs of shape $\mathcal{S}$.

[50] "Die Vollständigkeit der primitiven Darstellungen einer geschlossenen kontinuierlichen Gruppe" (1927), Fritz Peter and Hermann Weyl.

The expected overlap of two operators E, E' which projectively commute with $\mathcal{S}$ is then

$$\begin{aligned}\mathbb{E}_\mathcal{S}[E^*E'] &= \Pi_\mathcal{S}E^*E'\Pi_\mathcal{S} \\ &= [\chi'\overline{\chi}](\Pi_\mathcal{S})E^*E'\Pi_\mathcal{S} \\ &= \delta_{\chi,\chi'}E^*E'\Pi_\mathcal{S} \qquad (137) \\ &= \delta_{\chi,\chi'}\nu_E I_\mathcal{S}, \qquad (138)\end{aligned}$$

Since $\chi\chi'$ is multiplicative, (136) applies and $[\overline{\chi}\chi'](\Pi_\mathcal{S}) = 1$ just in case $\chi = \chi'$.

where (138) assumes E, E' are equivalent scaled isometries on $\mathcal{L}$ when they have the same character. In particular, if distinct Kraus operators belong to different classes in (135), then we only need to make sure they are scaled isometries to satisfy the KL criterion (127).

17. A game of codes

We'll finish by looking at the simplest nontrivial examples in detail, and throw in some diagrammatic shorthands at no extra cost. To warm up, let's encode $m = 1$ copy of the Pauli algebra in $n = 2$ copies, called the $[[n, m]] = [[2, 1]]$ code for short. The codewords are $C_2[L] = L \otimes L$ for $L \in \{X, Z\}$, and C_n is the *repetition map*. We denote this using a box with tensor power superscript, as in Fig. 51:

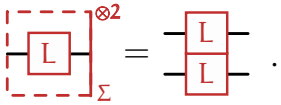


Figure 51: The twofold repetition map.

The solid line on the right of the box is a "filter", restricting the inputs to elements $L \in \Sigma$, with the filtering set Σ in the subscript. To extend this to a channel, i.e. a *linear* map on inputs, we require that Σ be linearly independent, so that *coherent repetition*

$$C_n\left[\lambda_i L_i\right] = \lambda_i L_i^{\otimes n} \tag{139}$$

is well-defined. Note that is *not* the same as the n-fold tensor product of an arbitrary input, which leads to a nonlinear map (no cloning). In the language of awds, our coherent repetition channel is

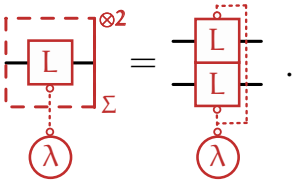
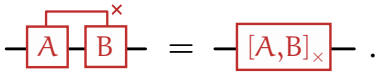


Figure 52: The two-fold coherent repetition channel.

We now turn to the stabilizer $\mathcal{S} \subseteq \mathcal{A}_{\text{phys}} = \mathcal{A}_{\text{Pauli}}^{\otimes 2}$. We need it to (a) commute with Σ, and (b) commute with itself. If we stick to tensor products of operators I, X, Y, Z, the only nontrivial combination that meet requirements (a) and (b) is $Y \otimes Y$, since only Y anticommutes with both X and Z. Thus, we have a lone generator $Y \otimes Y$. We'll use awd notation for multiplicative commutators,



We omit the $\times$ to indicate the usual additive commutator. The fact that codewords commute with stabilizer elements is expressed by Fig. 53. Borrowing some intuition from particle physics, we can think of them as "annihilating" each other.

The next step is to find the projective stabilizer, and identify stabilizer characters according to (132). For our single stabilizer generator $Y \otimes Y$, this is fairly easy. First, the elements which give phase $\chi = +1$

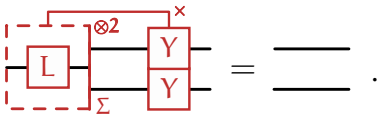


Figure 53: Codewords and stabilizer generators annihilate.

when commuted with $Y \otimes Y$:

$$\Sigma \otimes \Sigma, \quad X\Sigma \otimes X\Sigma,$$

where a tensor product of sets indicates the set of tensor products. This tells us that an error on both copies ($\Sigma \otimes \Sigma$) yields the same character as no error ($X\Sigma \otimes X\Sigma$). Put differently, the $[[2,1]]$ code cannot *detect* two errors. Similarly, the elements with commutator $\chi = -1$ are

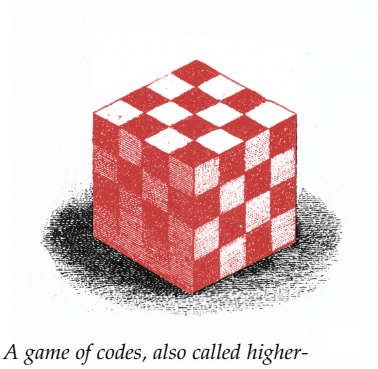
$$X\Sigma \otimes \Sigma, \quad \Sigma \otimes X\Sigma.$$

Technically, we can view $I \otimes Y$ and $Y \otimes I$ as two errors, since $Y \propto XZ$.

Thus, the $[[2,1]]$ code can detect a single error on either copy of the Pauli algebra, but because the errors yield the same character, we cannot tell which copy is corrupted! So it can *detect but not correct* a single error. Since two errors suffice to undetectably corrupt logical data, we say this code has *distance* $d = 2$ between codewords, and append this to our shorthand, $[[n,m,d]] = [[2,1,2]]$.

This isn't the greatest code; error-correction is a high-stakes game of checkers against Nature, and we need to do better! First, it doesn't faithfully represent the Pauli algebra, since the codewords *commute* rather than *anticommute*:

$$C_2[X]C_2[Z] = (-1)^2 C_2[Z]C_2[X] = C_2[Z]C_2[X].$$

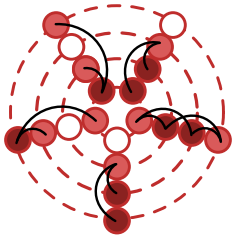


A game of codes, also called higher-dimensional checkers.

For this reason, it makes more sense to consider odd n. The second problem is that we can't actually correct any errors! For $n = 3$, a very similar construction uses stabilizer generators $Y \otimes Y \otimes I$ and $I \otimes Y \otimes Y$, and allows us to not only detect, but correct a single X (*bit flip*) error or a single Z (*phase*) error, but not both.

We'll set our sights a little higher, and construct an $n = 5$ code which can correct both. This is essentially possible because, when we encode $X \mapsto C_5[X]$ and $Z \mapsto C_5[Z]$, a new option emerges for stabilizer generators: build them of X and Z rather than Y. For instance,

$$S_{(0)} = I \otimes X \otimes Z \otimes Z \otimes X$$



Proof by necklace that $S_{(k)}$, $k = 0,1,2,3$, form a commutative group: any pair of circles has precisely two light/dark red (XZ) clashes, indicated by black arcs.

always gets a factor of $(-1)^2 = 1$ when multiplicatively commuted with a codeword. To form a commutative group, we need them to have an even number of XZ overlaps; given $S_{(0)}$, only its cyclic shifts S_k have this property. More formally, for $k = 1,2,3$, we cyclically permute forward with the operator T:

$$S_{(k)} = T^k S_{(0)}, \quad T\left[\bigotimes_{\ell \in \mathfrak{L}} A^{(\ell)}\right] = \bigotimes_{\ell \in \mathfrak{L}} A^{(\ell+1)},$$

so T increments each tensor label (assuming we index cyclically). Note that $S_{(4)} = S_{(0)}S_{(1)}S_{(2)}S_{(3)}$ is automatically included.

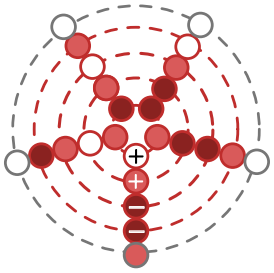
Next, we have rustle up some errors. We'll check that we can distinguish X, Y and Z errors on the first Pauli algebra; the character pattern and symmetry will show that these errors can be distinguished, hence corrected, on any copy. Let's first set

$$E_{(1,i)} = \sigma_{(i)} \otimes I \otimes I \otimes I \otimes I = \sigma^{(1)}_{(i)}.$$

It's easy to read off the associated character

$$\chi_{(1,i)}(S_{(k)}) = \begin{cases} +1,+1,-1,-1,+1 & i = 1 \\ +1,-1,-1,-1,-1 & i = 2 \\ +1,-1,+1,+1,-1 & i = 3, \end{cases}$$

where we range over $k = 0,1,2,3,4$. Since X, Y and Z all have different characters, they are all distingushable; moreover, changing the Pauli algebra on which the error is applied simply shifts the all-1 column, so single Pauli errors on different algebras can be told apart. Since we can detect X, Y and Z, we can detect *any* single-Pauli error. This is called the *five-qubit stabilizer code*,[51] or the $[[5,1,3]]$ since it encodes $n = 1$ copy of the Pauli algebra into $n = 5$, with a distance of $d = 3$ between them. Our checkers game just leveled up.



Entries of the character associated to $E_{(1,1)}$*, read vertically down.*

[51] "Perfect Quantum Error Correcting Code" (1996), Ray Laflamme, Cesar Miquel, Juan Pablo Paz, and Wojciech Zurek; "Mixed-state entanglement and quantum error correction" (1996), Charles Bennett, David DiVincenzo, John Smolin, and William Wootters.

Exit through the gift shop

We hope you enjoyed your visit to the quantum quarter of Disneyland, jumped on a few rides, and laughed at the marginalia. But you may be looking for something useful to remember your visit by. In this final section, we'll preview applications that take us beyond neatness of analogy, or stability of foundation, and into the realm of genuine real-world use. We defer a detailed treatment to future instalments of SIQP.

Harmonic oscillators

Abstract wiring diagrams were initially introduced as an alternative to circuit diagrams. But we were secretly building a programming language all along! `Yaw` is a high-level, functional framework quantum for programming whose formal semantics is based on awds. This makes it considerably more flexible than any existing quantum language, a fact we illustrate with two examples.

First of all, recall the harmonic oscillator, with ladder operators a, a^* satisfying the *canonical commutation relations*

$$[a, a^*] = i. \tag{140}$$

This is pictured in Fig. 54 (above). As we discussed in §2, this does not lead to a well-defined C*-algebra, so we need to employ Weyl's trick (Appendix A) of replacing (146) with the exponentiated form

$$[e^{ita}, e^{isa^*}]_\times = e^{-ist}. \tag{141}$$

We can easily capture this with an awd, Fig. 54 (below):

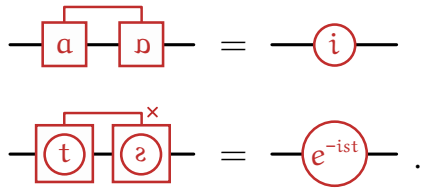


Figure 54: Above. The canonical commutation relation for ladder operators. Below. Weyl's exponentiated comutation relation.

Here, a square with a circle is just an exponentiated a or a^* depending on the orientation of the label. These objects are infinite-dimensional, but an advantage of the functional paradigm is *lazy evaluation*, which lets us handle infinite data structures.[52]

The fact that we can easily specify both continuous- and discrete-variable algebras in the same computations lets us perform "mixed-variable" protocols, such as Brenner et al.'s remarkable algorithm[53] for factorizing an arbitrary integer using three oscillators and a qubit. To the best of our knowledge, hybrid protocols like this one cannot be implemented on any other platform.

[52] "A lazy evaluator" (1976), Peter Henderson and James Morris; "Cons should not evaluate its arguments" (1976), Dan Friedman and David Wise.

[53] "Factoring an integer with three oscillators and a qubit" (2024), Lukas Brenner, Libor Caha, Xavier Coiteux-Roy, and Robert Koenig.

The Weyl commutation relations (141) are not only useful for harmonic oscillators. The familiar defining relations of the Pauli algebra can be written

$$X^2 = Z^2 = I, \quad [X, Z]_\times = -1,$$

with X and Z playing the role of a and a^* respectively, and we note that -1 is a square root of unity. Replacing 2 with d, and square roots with dth root, we get the *generalized Pauli algebra*

$$X^d = Z^d = I, \quad [X, Z]_\times = e^{2\pi i/d} = \omega_d. \tag{142}$$

Under the GNS construction, this is isomorphic to the algebra of $d \times d$ complex matrices $M_d(\mathbb{C})$. We use white and red squares for X and Z, and a white circle for ω_d, so the governing relations are:

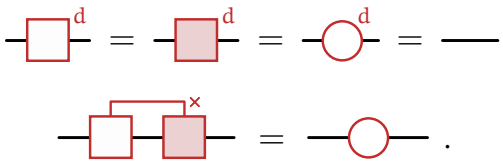


Figure 55: Above. The canonical commutation relation for ladder operators. Below. Weyl's exponentiated comutation relation.

This is a compact and appealing presentation of the algebra of qudits.

Stabilizer circuits

Compactness helps minimize cognitive load, but has other advantages. Suppose we wish to multiply two Pauli strings:

$$A = \bigotimes_{\ell \in \mathfrak{L}} \sigma_{(i_\ell)}, \quad B = \bigotimes_{\ell \in \mathfrak{L}} \sigma_{(j_\ell)}. \tag{143}$$

Since each $i_\ell, j_\ell \in \{0, 1, 2, 3\}$, it takes two bits to specify the Pauli at a given location, and for $n = |\mathfrak{L}|$, this means $4n$ bits total to store both A and B. Multiplication is "diagonal" in the Pauli basis, with n products evaluated according to (8). Thus, multiplying the operators takes $\mathcal{O}(n)$ space and $\mathcal{O}(n)$ time. This is exponentially faster than a naive attempt using matrices, where storage and steps both scale with the dimension, $\mathcal{O}(2^n)$.

It's possible to push this much further. Instead of a single string of Pauli operators, we can utilize stabilizer groups—the basis of the error-correction games we played above—to characterize and efficiently simulate the wider class of *Clifford operations*. This is the celebrated Gottesman-Knill Theorem; the *tableau algorithm* for simulating stabilizer circuits, devised by Aaronson and Gottesman,[54] takes $\mathcal{O}(n^2)$ time and space. The tableau method also goes beyond stabilizer circuits. For instance, adding d non-stabilizer operations, which each distribute across at most b algebras, costs

$$\mathcal{O}(4^{2bd} n + n^2)$$

in both time and space complexity. This definitely adds some overhead, but is not automatically exponential in n.

Some other quantum languages tack stabilizer simulation on, at least for qubits. But tacked on things can fall off again; put differently, core language features dictate what a language is good at. We have chosen to make stabilizer-based simulation idiomatic on the premise that *integrated error correction* is the medium-term fate of quantum programming. Another crucial distinction is that our stabilizer simulation is broader than the Pauli group. Above, we gave an algebra-agnostic version of stabilizer codes, and in future work provide an algorithm for classically simulating these operations. This fundamentally different from existing languages!

Aka the projective centralizer of the group of Pauli strings, aka the muggle (non-magical) operations.

[54] "Improved simulation of stabilizer circuits" (2004), Scott Aaronson and Daniel Gottesman.

Tricks with measurement

Stabilizer operations are not only reserved for error correction and simulation. They are also useful for near-term applications, most notably Classical Shadows Method[55] which which cobbles together a few measurements to approximate the expectations of exponentially many more. Say we start with an unknown state π and

[55] "Predicting many properties of a quantum system from very few measurements" (2020), Hsin-Yuan Huang, Richard Kueng, and John Preskill.

we want to find the means $\pi(A_i)$ for some collection of operators $A_i, i \in \mathfrak{I}$. The intuition is that traditional tomography, which attempts to reconstruct π, is overkill; data about expectations should lie in a smaller space because expectations are related!

The technical insight is that we can fix a PVM Λ, with a complete, orthogonal set of projectors Π_λ labelled by $\lambda \in \mathfrak{F}$, and conjugate it by a random unitary $U \sim \mathrm{Uni}(\mathcal{G})$, where $\mathcal{G} \subseteq \mathcal{U}(\mathcal{A})$ is a finite group of unitaries. This has the effect of measuring our input state π according to the Born-Lüders rules

$$\pi \mapsto p_{U,\lambda}^{-1}\left(C^{\Pi_{\hat{\lambda}}} \circ C^{\hat{U}}\right)[\pi] = \pi_{U,\lambda}, \quad p_{U,\lambda} = \mathcal{C}^U[\pi](\Pi_\lambda).$$

Since both the measurement and the conjugation are random, we can form the expectation (using the uniform distribution over $\mathcal{G}$):

$$\mathcal{M}(\pi) = \mathbb{E}_{U,\lambda}\left[\pi_{\hat{U},\hat{\lambda}}\right] = \frac{1}{|\mathcal{G}|}\left(C^{\mathfrak{F}} \circ C^{\mathcal{G}}\right)[\pi].$$

When the group is large enough that this set of random measurements is tomographically complete, we can reverse $\mathcal{M}$. We then treat specific random post-measurement states $\pi_{U,\lambda}$ as samples of this process, and invert to get *classical shadows* $\mathcal{M}^{-1}(\pi_{U,\lambda})$ of π we can use for esimating our desired averages $\pi(A_i)$.

By batching (to deal with outliers) and taking a median of batch-averaged shadows, we can estimate the averages to additive precision ϵ using $\Theta(C_{\mathcal{G}} \log|\mathfrak{I}|/\epsilon^2)$ shadows, where $C_{\mathcal{G}}$ is the constant

$$\begin{aligned} C_{\mathcal{G}} &= \max_{i\in\mathfrak{I}} \max_{\kappa\in S(\mathcal{A})} \mathbb{E}_{U,\lambda|\kappa}\left[\widetilde{\mathcal{M}}^{-1}(A_i)^2[\mathcal{C}_U\Pi_\lambda]\right] \\ &= \max_{i\in\mathfrak{I}} \|A_i\|_{\text{shadow}}^2. \end{aligned}$$

Here, $\kappa \in S(\mathcal{A})$ is an arbitrary state, and $\widetilde{\mathcal{M}}^{-1}$ the result of pulling the shadow map onto density matrices, then pushing the results back to the space of (not necessarily positive) linear functionals on $\mathcal{A}$. We also define the *shadow metric* $\|\cdot\|_{\text{shadow}}^2$, which measures the complexity of an expectation with respect to our chosen shadows.

All of this can be made analytically explicit in the case of random Cliffords and Paulis, highlighted for different reasons above. The algebraic framework naturally suggests several extensions:

- the stabilizer connection should make it possible to explore randomization schemes related to early fault-tolerant architectures;
- the ability to work with non-qubit analogues of the Clifford and Pauli groups may lead to hardware-adapted shadow ansatzae;
- the expressivity of the language should make it easier to vary, extend, and modularize the protocol, e.g. swapping out classical shadows for estimation via regularized least squares.[56]

We leave these developments to future work.

Making tomography cool again.

"Classical" since this is all done using classical post-processing.

[56] "On the connection between least squares, regularization, and classical shadows" (2024), Zhihui Zhu, Joseph Lukens, and Brian Kirby.

A. Commutative C*-algebras

A.1. Continuous functional calculus

Recall that $\sigma(A)$ is the spectrum of an element $A \in \mathcal{A}$. Define a commutative C*-algebra of continuous functions on the spectrum,

$$C^*(\sigma(A)) = C^0(\sigma(A)).$$

When A is normal, $AA^* = A^*A$, there is a unique assignment

$$\Phi_A : C^*(\sigma(A)) \to \mathcal{A} \quad \text{with} \quad \Phi_A(\mathbf{1}) = I_{\mathcal{A}}, \Phi_A(\mathrm{id}_{\sigma(A)}) = A, \tag{144}$$

where $\mathbf{1}(\lambda) = 1$ is constant and $\mathrm{id}_{\sigma(A)}(\lambda) = \lambda$ is the identity. This allows us to define $f(A) = \Phi_A(f)$, which is called the *continuous functional calculus* for A.[57]

The continuous functional calculus lets us prove the spectral theorem without any need for Hilbert space. For normal A, suppose the spectrum consists of K disjoint closed sets, $\sigma(A) = \bigsqcup_{k\in\mathfrak{K}} \sigma_k$. Then there are associated projectors $\Pi_{(k)}$ with the following properties:

- each projector has spectrum $\sigma(\Pi_{(k)}) = \sigma_k$;
- projectors commute with A, $\Pi_{(k)}A = A\Pi_{(k)}$;
- projectors are orthogonal, $\Pi_{(j)}\Pi_{(k)} = \delta_{jk}\Pi_{(k)}$; and
- projectors resolve the identity, $\sum_{k\in\mathfrak{K}} \Pi_{(k)} = I$.

The proof is straightforward. We define a *characteristic function* $\mathbf{1}_k(\lambda) = \mathbb{I}[\lambda \in \sigma_k]$, which is continuous since the inverse image $\{1\} = \mathbf{1}_k^{-1}(\sigma_k)$ of closed set σ_k is closed. Similarly, $\{0\} = \mathbf{1}_k^{-1}(\sigma_k^c)$ is closed, with

$$\sigma_k^c = \bigsqcup_{j\neq k} \sigma_k$$

a finite union of closed sets by assumption, hence closed. The remaining properties follow from applying the continuous functional calculus Φ_a to the characteristic functions $\mathbf{1}_k$. Thus, we have the usual spectral theorem for normal operators, with a minimal amount of formal baggage.

[57] For a proof, see Theorem 4.1.3 of *Fundamentals of the Theory of Operator Algebras I* (1983), Richard Kadison and John Ringrose.

A.2. Weyl commutation relations

Another application of the continuous functional calculus is taming wild commutation relations, such as the *canonical commutation relation (CCR)* for self-adjoint Q, P:

$$[Q, P] = iI, \tag{145}$$

in units where $\hbar = 2\pi$. This CCR cannot be satisfied by trace class operators Q and P, since taking the trace of the LHS gives zero (using

cyclicity of trace) while the right does not. A similar argument shows that Q and P are unbounded, and therefore cannot live in any C*-algebra, whose members act as bounded linear operators on some GNS Hilbert space. How do we accomodate them?

Although unbounded, P and Q are nevertheless normal, and we can define the exponentials $q(t) = e^{iQt}$ and $p(t) = e^{iPt}$ via the continuous functional calculus. The *Weyl CCR* is the exponentiated or braided form of (145):

$$[q(t), p(s)]_\times = e^{-ist}. \tag{146}$$

If the *Baker-Campbell-Hausdorff (BCH)* formula holds, then (145) follows from (146) by expanding in the coincident limit $s \to t$.[58] If BCH doesn't hold, all bets are off; if it does, the question of *uniqueness* remains, i.e. whether the operators satisfying (146) are uniquely determined to be $q(t)$ and $p(s)$. It's clear that any unitarily equivalent operators also satisfy (146),

$$[U^* q(t) U, U^* p(s) U]_\times = e^{-ist} U^* U = e^{-ist}.$$

The *Stone-von Neumann theorem*[59] asserts the converse, i.e. any pair of one-parameter families $q'(t)$ and $p'(s)$ must be unitarily equivalent to $q(t)$ and $p(s)$.

[58] See for instance *Quantum Computation and Quantum Information* (2000), Michael Nielsen and Isaac Chuang.

[59] For a (perhaps excessively) pedagogical treatment, we refer the reader to Theorem 5.6.36 of *Fundamentals of the Theory of Operator Algebras II* (1997), Richard Kadison and John Ringrose.

B. Proof details

B.1. Jordan characters

Building on a result of Kadison and Singer, Størmer[60] proved the following lemma: if $\mathcal{D}'_\pi \supseteq \mathcal{D}_\pi$ and π is pure, then either $\mathcal{D}'_\pi = \mathcal{A}_{\text{sa}}$ (the self-adjoint elements of $\mathcal{A}$) or $\mathcal{D}'_\pi = \mathcal{D}_\pi$. We won't reproduce the proof, but we do note two corollaries: (a) pure states have maximal definite sets, assuming that no state is definite on all $\mathcal{A}_{\text{sa}}$, i.e. no one-dimensional representations; and (b) pure states are rigid on these definite sets.

[60] "Extensions of pure states" (1959), Kadison and Singer; "A characterization of pure states of C*-algebras" (1967), Erling Størmer.

(a) If $\mathcal{D}_{\pi'} \subseteq \mathcal{D}_\pi$ and π is pure, then $\mathcal{D}_{\pi'} = \mathcal{D}_\pi$, assuming the algebras does not possess characters, i.e. one-dimensional representations.

(b) Suppose π' is pure and agrees with pure π on $\mathcal{D}_{\pi'}$. Then for any $A \in \mathcal{D}_{\pi'}$, we have

$$\pi(A^2) = \pi'(A^2) = \pi'(A)^2 = \pi(A)^2,$$

and hence $A \in \mathcal{D}_{\pi'}$. Since $\mathcal{D}_\pi \supseteq \mathcal{D}_{\pi'}$, Størmer's lemma implies $\mathcal{D}_\pi = \mathcal{D}_{\pi'}$ with identical values, and hence the kernels agree, $\mathcal{K}_\pi = \mathcal{K}_{\pi'}$. This forces the same GNS Hilbert space up to a unitary, $\mathcal{K}_\pi = U^* \mathcal{K}_{\pi'} U$ and hence $[A]_\pi = [U^* A U]_{\pi'}$ for all $A \in \mathcal{A}$.

Sharpness implies that, for all $\Gamma \in \mathcal{D}_{\pi'}$,

$$[\Gamma\, U]_{\pi'} = [U\Gamma]_{\pi'}$$

so the commutator of U and Γ is null. It follows that U can be assigned a sharp value without disturbing the remaining measurements in $\mathcal{D}_{\pi'}$. This contradicts the maximality of $\mathcal{D}_{\pi'}$ unless we already have $U \in \mathcal{D}_{\pi'}$. Thus, we have

$$[U^* A U]_{\pi'} = [U^*]_{\pi'}[A]_{\pi'}[U]_{\pi'} = e^{-i\theta}[A]_{\pi'}e^{i\theta} = [A]_{\pi'},$$

since $[U]_{\pi'} = e^{i\theta} I$ for unitary sharp U. It follows that $[A]_{\pi'} = [A]_\pi$ for all $A \in \mathcal{A}$, hence $\pi' = \pi$.

B.2. Robertson-Schrödinger uncertainty relation

Definite operators, the mechanism by which we specify and control state, are the quantum equivalent of a Boolean switch. But unlike Boolean circuits, where we are free to simultaneously configure every switch, making one operator definite will make others blurry by the uncertainty principle. We can use our correlator to place some concrete, state-dependent bounds on this blurriness. First, note that

$$G_\pi(\Delta A, \Delta B) = \pi[(\Delta B)^* \Delta A] = \pi(B^* A) - \pi(B^*)\pi(A).$$

A little algebra gives

$$\begin{aligned} G_\pi(\Delta A, \Delta B) - G_\pi(\Delta B^*, \Delta A^*) &= \pi(B^\star A) - \pi(A B^\star) \\ &= \pi([B^\star, A]) \qquad (147) \\ G_\pi(\Delta A, \Delta B) + G_\pi(\Delta B^*, \Delta A^*) &= \pi(B^* A) + \pi(A B^*) - 2\pi(B^*)\pi(A) \\ &= \pi(\{B^\star, A\}) - 2\pi(B^*)\pi(A). \qquad (148) \end{aligned}$$

where $[A, B] = AB - BA$ is the *commutator* and $\{A, B\} = AB + BA$ the *anticommutator* of operators.

When A and B are self-adjoint, (147) and (148) are respectively proportional to the imaginary and real part of $G_\pi(\Delta A, \Delta B)$, so

$$|G_\pi(\Delta A, \Delta B)|^2 = \frac{1}{4}|\pi([A, B])|^2 + |\pi(A \circ B) - \pi(A)\pi(B)|^2\,, \qquad (149)$$

where $2A \circ B = \{A, B\}$ is the Jordan product. Finally, using the Cauchy-Schwarz inequality (20), we can lower bound the product of variances:

$$\frac{1}{4}|\pi([A, B])|^2 + |\pi(A \circ B) - \pi(A)\pi(B)|^2 \leq \|\Delta A\|_\pi^2 \|\Delta B\|_\pi^2. \qquad (150)$$

In words, the product of variances is bounded below by the squared commutator plus squared gap between regular and Jordan product.

B.3. Code subspace projection

We claimed above that the operator (131), given by

$$\Pi_{\mathcal{S}} = \frac{1}{|\mathcal{S}|} \sum_{S \in \mathcal{S}} S,$$

projects onto the code subspace

$$\mathcal{H}_{\mathcal{S}} = \{\psi \in \mathcal{H} : S\psi = \psi \text{ for all } S \in \mathcal{S}\},$$

where we work in a fixed representation $\mathcal{H}$. First, we show this is a projector. It is Hermitian since $\mathcal{S}$ is closed under adjoints, $S \in \mathcal{S}$ implies $S^* \in \mathcal{S}$:

$$\Pi_{\mathcal{S}}^* = \frac{1}{|\mathcal{S}|} \sum_{S \in \mathcal{S}} S^* = \frac{1}{|\mathcal{S}|} \sum_{S \in \mathcal{S}} S = \Pi_{\mathcal{S}}.$$

Similarly, it is a projector since $S \cdot \mathcal{S} = \mathcal{S}$, i.e. multiplication by a fixed element is a bijection on $\mathcal{S}$:

$$\Pi_{\mathcal{S}}^2 = \frac{1}{|\mathcal{S}|^2} \sum_{S \in \mathcal{S}} \sum_{S' \in S \cdot \mathcal{S}} S' = \frac{1}{|\mathcal{S}|^2} \sum_{S \in \mathcal{S}} \sum_{S' \in \mathcal{S}} S' = \frac{1}{|\mathcal{S}|} \sum_{S' \in \mathcal{S}} S' = \Pi_{\mathcal{S}}.$$

Thus, $\Pi_{\mathcal{S}}$ is indeed a projector.

To verify the domain of projection, first note that by the same logic as above,

$$S\Pi_{\mathcal{S}} = \Pi_{\mathcal{S}} = \Pi_{\mathcal{S}} S,$$

and hence $\Pi_{\mathcal{S}}$ projects into the set of fixed points of $\mathcal{S}$. Finally, we confirm that every element of the code subspace $\psi \in \mathcal{H}_{\mathcal{S}}$ is fixed:

$$\Pi_{\mathcal{S}}\psi = \frac{1}{|\mathcal{S}|} \sum_{S \in \mathcal{S}} S\psi = \frac{1}{|\mathcal{S}|} \sum_{S \in \mathcal{S}} \psi = \psi.$$

Thus, $\Pi_{\mathcal{S}}$ is a projector with range $\mathcal{H}_{\mathcal{S}}$.

References

[1] Aaronson, S., and Gottesman, D. Improved simulation of stabilizer circuits. *Physical Review A 70*, 5 (Nov. 2004).

[2] Artin, M. *Algebra*. Birkhäuser, 1998.

[3] Axler, S. J. *Linear Algebra Done Right*. Undergraduate Texts in Mathematics. Springer, New York, 1997.

[4] Belavkin, V., and Staszewski, P. A Radon-Nikodym theorem for completely positive maps. *Reports on Mathematical Physics 24*, 1 (1986), 49–55.

[5] Bell, J. S. On the Einstein Podolsky Rosen paradox. *Physics Physique Fizika 1* (Nov 1964), 195–200.

[6] Bennett, C. H. Logical reversibility of computation. *IBM J. Res. Develop. 17*, 6 (1973), 525–532.

[7] Bennett, C. H., DiVincenzo, D. P., Smolin, J. A., and Wootters, W. K. Mixed-state entanglement and quantum error correction. *Phys. Rev. A 54* (Nov 1996), 3824–3851.

[8] Birkhoff, G., and Neumann, J. V. The logic of quantum mechanics. *Annals of Mathematics 37*, 4 (1936), 823–843.

[9] Bloch, F. Nuclear induction. *Phys. Rev. 70* (October 1946), 460–474.

[10] Boole, G. *The Mathematical Analysis of Logic: Being an Essay Towards a Calculus of Deductive Reasoning*. Cambridge University Press, 1847.

[11] Born, M. On the quantum mechanics of collision processes. *Zeitschrift für Physik 37* (1926), 863–867.

[12] Brenner, L., Caha, L., Coiteux-Roy, X., and Koenig, R. Factoring an integer with three oscillators and a qubit, 2024.

[13] Brown, N., and Ozawa, N. *C*-algebras and Finite-dimensional Approximations*. Graduate Studies in Mathematics. American Mathematical Society, 2008.

[14] Carteret, H. A., Terno, D. R., and Życzkowski, K. Dynamics beyond completely positive maps: Some properties and applications. *Phys. Rev. A 77* (Apr 2008).

[15] Charney, J. G., Fjörtoft, R., and von Neumann, J. Numerical integration of the barotropic vorticity equation. *Tellus 2* (1950), 237–254.

[16] Choi, M.-D. Completely positive linear maps on complex matrices. *Linear Algebra Appl. 10*, 3 (1975), 285–290.

[17] Coffman, V., Kundu, J., and Wootters, W. K. Distributed entanglement. *Physical Review A 61*, 5 (Apr. 2000).

[18] Curry, H. B. *Combinatory Logic*. North-Holland Pub. Co., Amsterdam,, 1958.

[19] Einstein, A., Podolsky, B., and Rosen, N. Can quantum-mechanical description of physical reality be considered complete? *Physical Review 47*, 10 (1935), 777–780.

[20] Feynman, R. P. Simulating physics with computers. *International Journal of Theoretical Physics 21*, 6/7 (1982), 467–488.

[21] Fredkin, E. F., and Toffoli, T. Conservative logic. *International Journal of Theoretical Physics 21*, 3/4 (1982), 219–253.

[22] Friedman, D. P., and Wise, D. S. CONS Should Not Evaluate its Arguments. In *Proceedings of the Third International Colloquium on Automata, Languages and Programming* (1976), pp. 257–284.

[23] Gelfand, I., and Naimark, M. On the imbedding of normed rings into the ring of operators in Hilbert space. *Sbornik Mathematics 54*, 2 (1943), 197–217.

[24] Gleason, A. M. Measures on the closed subspaces of a Hilbert space. *Journal of Mathematics and Mechanics 6*, 6 (1957), 885–893.

[25] Grünbaum, B., and Shephard, G. C. Patch-determined tilings. *The Mathematical Gazette 61*, 415 (1977), 31–38.

[26] Hamming, R. W. Error detecting and error correcting codes. *The Bell System Technical Journal 29*, 2 (1950), 147–160.

[27] Henderson, P., and Morris, J. H. A lazy evaluator. In *Proceedings of the 3rd ACM SIGACT-SIGPLAN Symposium on Principles on Programming Languages* (New York, NY, USA, 1976), POPL '76, Association for Computing Machinery, p. 95–103.

[28] Hilbert, D., von Neumann, J., and Nordheim, L. Über die grundlagen der quantenmechanik. *Mathematische Annalen 98* (1928), 1–30.

[29] Hoeffding, W. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association 58*, 301 (1963), 13–30.

[30] Hopf, H. Über die abbildungen der dreidimensionalen sphäre auf die kugelfläche. *Mathematische Annalen 104* (1931), 637–665.

[31] Howard, W. A. The formulae-as-types notion of construction. In *To H. B. Curry: Essays on Combinatory Logic, Lambda Calculus, and Formalism*, H. Curry, H. B., S. J. Roger, and P. Jonathan, Eds. Academic Press, 1980.

[32] Huang, H.-Y., Kueng, R., and Preskill, J. Predicting many properties of a quantum system from very few measurements. *Nature Physics 16*, 10 (June 2020), 1050–1057.

[33] Jamiołkowski, A. Linear transformations which preserve trace and positive semidefiniteness of operators. *Reports on Mathematical Physics 3*, 4 (1972), 275–278.

[34] Kadison, R., and Ringrose, J. *Fundamentals of the Theory of Operator Algebras I: Elementary Theory*. Fundamentals of the Theory of Operator Algebras. American Mathematical Society, 1983.

[35] Kadison, R., and Ringrose, J. *Fundamentals of the Theory of Operator Algebras II: Advanced Theory*. Fundamentals of the Theory of Operator Algebras. American Mathematical Society, 1997.

[36] Kadison, R. V., and Singer, I. M. Extensions of pure states. *Amer. J. Math. 81*, 2 (April 1959), 383–400.

[37] Knill, E., Laflamme, R., Martinez, R., and Negrevergne, C. Benchmarking quantum computers: The five-qubit error correcting code. *Physical Review Letters 86*, 25 (June 2001), 5811–5814.

[38] Knill, E., Laflamme, R., and Viola, L. Theory of quantum error correction for general noise. *Physical Review Letters 84*, 11 (Mar. 2000), 2525–2528.

[39] Kolmogorov, A. N. *Foundations of the Theory of Probability*, second english edition ed. Chelsea Publishing Company, 1950.

[40] Kraus, K., Böhm, A., Dollard, J., and Wootters, W. *States, Effects, and Operations: Fundamental Notions of Quantum Theory*. Lecture Notes in Physics. Springer, 1983.

[41] Krein, M., and Milman, D. On extreme points of regular convex sets. *Studia Mathematica 9*, 1 (1940), 133–138.

[42] Kribs, D., Laflamme, R., and Poulin, D. Unified and generalized approach to quantum error correction. *Physical Review Letters 94*, 18 (May 2005).

[43] Kribs, D. W., Laflamme, R., Poulin, D., and Lesosky, M. Operator quantum error correction. *Quantum Info. Comput. 6*, 4 (July 2006), 382–399.

[44] Laflamme, R., Miquel, C., Paz, J. P., and Zurek, W. H. Perfect quantum error correcting code. *Phys. Rev. Lett. 77* (Jul 1996), 198–201.

[45] Landauer, R. Irreversibility and heat generation in the computing process. *IBM Journal of Research and Development 5*, 3 (1961), 183–191.

[46] Loring, T. A. C*-algebra relations. *Mathematica Scandinavica 107*, 1 (2010), 43–72.

[47] Lüders, G. Concerning the state-change due to the measurement process. *Annalen der Physik 518*, 9 (1951), 663–670.

[48] Manin, Y. *Computable and noncomputable*. Sovet. Radio, 1980.

[49] Murray, F. J., and von Neumann, J. On rings of operators. *Bulletin of the American Mathematical Society 42* (1936).

[50] Murray, F. J., and von Neumann, J. On rings of operators (II). *Transactions of the American Mathematical Society 41*, 2 (1937), 208–248.

[51] Nielsen, M. A., and Chuang, I. L. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[52] Osborne, T. J., and Verstraete, F. General monogamy inequality for bipartite qubit entanglement. *Physical Review Letters 96*, 22 (June 2006).

[53] Park, J. L. The concept of transition in quantum mechanics. *Foundations of Physics* (1970), 23–33.

[54] Pechukas, P. Reduced dynamics need not be completely positive. *Phys. Rev. Lett. 73* (Aug 1994), 1060–1062.

[55] Pratt, V. Linear logic for generalized quantum mechanics. In *Proceedings of the IEEE Workshop on Physics and Computation* (1992).

[56] Rudin, W. *Principles of Mathematical Analysis*, 3rd ed. McGraw-Hill Book Company, Auckland, 1976.

[57] Rudin, W. *Fourier Analysis on Groups*. Wiley Classics Library. Wiley, 1991.

[58] Rédei, M. Why John von Neumann did not like the Hilbert space formalism of quantum mechanics (and what he liked instead). *Studies in History and Philosophy of Modern Physics 27*, 4 (1996), 493–510.

[59] Schmidt, E. On the solution of linear equations with infinitely many unknowns. *Rendiconti del Circolo Matematico di Palermo 25*, 1 (1908), 53–77.

[60] Schrödinger, E. Die gegenwärtige situation in der quantenmechanik. *Naturwissenschaften 23* (1935), 807–812.

[61] Segal, I. E. Irreducible representations of operator algebras. *Bulletin of the American Mathematical Society 53*, 2 (1947), 73 – 88.

[62] Shaji, A., and Sudarshan, G. Who's afraid of not completely positive maps? *Phys. Lett. A 341* (2005), 48–54.

[63] Shannon, C. E. *A symbolic analysis of relay and switching circuits*. MIT masters thesis, 1938.

[64] Shannon, C. E. A mathematical theory of communication. *The Bell System Technical Journal 27* (1948), 379–423.

[65] Shannon, C. E., and Weaver, W. *The Mathematical Theory of Communication*. University of Illinois Press, 1949.

[66] Simmons, G. *Introduction to Topology and Modern Analysis*. McGraw-Hill, 1963.

[67] Stern, O., and Gerlach, W. The experimental proof of directional quantization in the magnetic field. *Zeitschrift für Physik 9*, 1 (1922), 349–352.

[68] Stinespring, W. F. Positive functions on C*-algebras. *Proceedings of the American Mathematical Society 6*, 2 (1955), 211–216.

[69] Størmer, E. A characterization of pure states of C*-algebras. *Proceedings of the American Mathematical Society 19* (10 1968).

[70] Tao, T. Amplification, arbitrage, and the tensor power trick. Blog post, September 2007. *What's New*.

[71] von Neumann, J. Zur algebra der funktionaloperatoren und theorie der normalen operatoren. *Mathematische Annalen 102* (1929), 370–427.

[72] VON NEUMANN, J. *Mathematical Foundations of Quantum Mechanics*. Princeton University Press, 1932.

[73] WEINBERG, S. *The Quantum Theory of Fields, Volume 1: Foundations*. Cambridge University Press, 2005.

[74] WEYL, H. *Gruppentheorie und Quantenmechanik*. Hirzel, 1928.

[75] WEYL, H., AND PETER, P. Die vollständigkeit der primitiven darstellungen einer geschlossenen kontinuierlichen gruppe. *Mathematische Annalen 97* (1927), 737–755.

[76] WICHMANN, E. H., AND CRICHTON, J. H. Cluster decomposition properties of the S-matrix. *Phys. Rev. 132* (Dec 1963), 2788–2799.

[77] WOOTTERS, W. K., AND ZUREK, W. H. A single quantum cannot be cloned. *Nature 299*, 5886 (1982), 802–803.

[78] YOSHIDA, B., AND KITAEV, A. Efficient decoding for the Hayden-Preskill protocol, 2017.

[79] ZHU, Z., LUKENS, J. M., AND KIRBY, B. T. On the connection between least squares, regularization, and classical shadows. *Quantum 8* (Aug. 2024), 1455.