QUANTUM COMPUTING

# Quantum Computing

*Lecture notes for the software half of the course: what a quantum state is, what can be done to it, what a measurement returns, and what an algorithm has to arrange before that measurement is worth making.*

---

| Covers | Level | Assumed background |
| --- | --- | --- |
| Chapters 1 to 6, and Appendix A | Undergraduate | Linear algebra over the real numbers, and some Python |

These notes are an adaptation. Their syllabus and the sequence of topics they teach derive from **Quantum Computing Lectures** by Aleksandr Krasnok, *github.com/AlexKrasnok/quantum-computing-lectures*, used under CC BY 4.0. Every page here — the prose, the figures, the worked examples and the questions — is written for this edition.

### How to read these notes

Each idea arrives in the same order: what it is for, a definition, an equation, a short derivation with every step shown, a worked example, and the mistake that is easiest to make. Worked examples use five headings — Given, Find, Method, Solution, Check. Do the Check step yourself before reading it; it is where a wrong answer is most likely to survive, because a calculation that goes wrong in the middle usually goes wrong in the check as well and agrees with itself.

Four conventions apply everywhere, and the first two are worth fixing now because getting either wrong costs an error that nothing on the page reveals. A register of $n$ qubits is written $|q_{n-1}\ldots q_1 q_0\rangle$, and entry $x$ of its column of amplitudes is the amplitude of $|x\rangle$ with $x$ read as a binary number. A phase on the whole state is not physical and may be dropped; a phase between two terms is physical and may never be dropped. The inner product conjugates its first argument, $\langle u|v\rangle = \sum_k u_k^* v_k$, which in NumPy is `np.vdot(u, v)` and never `np.dot`. And $\hbar = 1$, so a Hamiltonian is measured in angular frequency and evolution is $U(t) = e^{-iHt}$.

The contents below carries a third column. An entry such as **NC CH2.1.4** points into the course textbook, Nielsen and Chuang, *Quantum Computation and Quantum Information*, tenth anniversary edition, where the same material is developed at length. The **NC** marker is what tells the two numbering systems apart, and it is not decorative: this chapter is chapter 1 here and section 2.1 there, and every later chapter disagrees with the book by a different amount.

THE ENGINEERING REGIME THESE NOTES ASSUME

Semiconductor scaling made quantum effects unavoidable at small dimensions. Quantum engineering asks when selected effects can instead be controlled as resources. The field then added a general computational model, algorithms with proved advantages under stated resource models, and today's noisy intermediate-scale quantum (NISQ) processors. NISQ hardware supports calibration studies, small demonstrations and experiments with noise. It is not fault-tolerant hardware, and error mitigation is not error correction. A defensible result names the task, classical baseline, hardware conditions, shot count, uncertainty and mitigation cost.

## 1 The mathematics of quantum states

NC CH2.1 · NC CH2.2.7

States as complex columns; bras and the inner product; length, orthogonality and overlap; orthonormal bases. Amplitude and phase, and the difference between a global and a relative one. Outer products, projectors and the resolution of the identity. Gram–Schmidt. The tensor product and where the exponential comes from. The adjoint, Hermitian and unitary operators, and the exponential that joins them. Eigenvectors, the spectral theorem and functions of an operator. Dirac notation. Square-integrable functions as vectors, complete bases, Parseval and truncation.

## 2 States, measurement and dynamics

NC CH2.2, NC CH2.1.3, NC CH2.1.9

The Born rule, active and passive transformations, and why choosing a measurement basis is choosing an experiment. Projective measurement, the state it leaves behind, and the instrument behind a general measurement. Observables, means and spreads. Compatibility, the commutator, and the uncertainty relation. The Pauli algebra and a measurement along any direction. Position and momentum, the free particle, the infinite square well, stationary states and why a gate is an exponential. What a finite run of shots is worth.

## 3 Mixed states and entanglement

NC CH2.4 · NC CH2.5 · NC CH2.6 · NC CH8.2 · NC CH11.3

The density operator, and the two situations no state vector describes. Which matrices are states, and the purity that says how mixed one is. Quantum channels in Kraus form, amplitude damping and dephasing, and the two decay times they become in continuous time. The partial trace, and a pure pair whose halves are mixed. Separability, the Schmidt decomposition and the entropy that weighs it. The Bell states, CHSH, and why entanglement sends nothing.

## 4 The Bloch sphere and quantum gates

NC CH1.2 · NC CH1.3 · NC CH3.2.5 · NC CH4.2 · NC CH4.5

One qubit drawn: the two angles, the half angle, and why opposite points are the orthogonal ones. Global against relative phase, and the double cover. Every one-qubit gate as a rotation — the Pauli gates, the Hadamard and the phase family. The order a circuit multiplies in, the Euler decomposition, and the three-parameter gate a machine offers. Reversible embeddings, ancillas and uncomputing. Two-qubit gates and the ordering that fails silently. Entangling power and universality.

## 5 Circuits and protocols

NC CH1.3.4 · NC CH1.3.5 · NC CH1.3.7 · NC CH4.4 · NC CH4.6 · NC CH6.1 · NC CH6.6

A circuit as a program, and the model of computation it belongs to. Depth against gate count, and which one a coherence time is spent against. Exact simulation and where it stops; shots, error bars, and the difference between sampling noise and a broken machine. Measurement inside a circuit, classical feedforward, and what a compiler does before the hardware sees anything. Then the five things a resource claim must name, and two

protocols worked end to end against them: teleportation, which finishes nothing until two classical bits arrive, and Grover, whose square root is a count of queries and not a time.

## 6 Quantum algorithms

NC CH1.4.2 · NC CH1.4.3 · NC CH1.4.4 · NC CH5.1 · NC CH5.2 · NC CH5.3 · NC CH6.3

What a query model counts, and what a query count is not. One mechanism — phase kickback — first with a Boolean oracle and then with any unitary, and the rule that a superposition which never interferes has bought nothing. Deutsch and Deutsch–Jozsa. The quantum Fourier transform, its circuit and what it does not return. Phase estimation, the two floors it guarantees, and quantum counting. Order finding, the continued fractions that read it, and the factoring that follows — with the reach of the same mechanism named at the end.

## A Formula summary

Every formula the six chapters establish, in the order they establish it, each entry naming the section that develops it. Nothing is derived there.

1

# The mathematics of quantum states

*Nothing in this chapter is quantum mechanics. It is the language the quantum mechanics of the rest of the course is written in, and it is worth learning on its own terms first: a reader who is decoding the notation cannot also be following the physics.*

Fix a basis — a list of the outcomes a system can be found in — and three things become concrete at once. A **state** is a column of complex numbers, one for each outcome. An **operation** is a matrix acting on that column. A **question** is a single number, formed from two columns by an inner product. For $n$ qubits the column has $2^n$ entries, so the objects are large; they are not complicated, and everything done to them in this course is built from four constructions, all of them in this chapter.

THE FOUR CONSTRUCTIONS

The **inner product** $\langle\phi|\psi\rangle$, which turns two states into a number and is where every probability comes from. The **outer product** $|\phi\rangle\langle\psi|$, which turns two states into an operator. The **tensor product** $\otimes$, which turns two systems into one. The **spectral decomposition**, which turns an operator into its eigenvalues and the projectors that belong to them.

## 1.1 Vectors, dual vectors and the inner product

A quantum state is written $|\psi\rangle$ and is called a **ket**. Once a basis $\{|0\rangle, |1\rangle\}$ has been named, the ket is a column of two complex numbers called **amplitudes**.

A QUBIT STATE

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle = \begin{bmatrix}\alpha\\ \beta\end{bmatrix}, \qquad |\alpha|^2 + |\beta|^2 = 1$$

The space these live in is a complex vector space, which means only that adding two states gives a state and multiplying a state by a complex number gives a state. That closure is the whole content of the word **superposition**. The normalisation is not a mathematical necessity; it is the bookkeeping that makes the probabilities of the next chapter add to one.

WHAT A SUPERPOSITION IS NOT

It is not the statement that the qubit is really in $|0\rangle$ or really in $|1\rangle$ and that we do not yet know which. That description predicts different numbers, and an interference experiment separates the two. A superposition is one state with two amplitudes, and both of them are there at once.

Every ket has a partner. The **bra** $\langle a|$ is the ket $|a\rangle$ transposed and conjugated, so it is a row. A row on the left of a column contracts to a single number, and that number is the **inner product**.

THE INNER PRODUCT

$$\langle a| = |a\rangle^\dagger = \begin{bmatrix} a_1^* & \cdots & a_n^* \end{bmatrix}$$

$$\langle a|b\rangle = \sum_k a_k^*\, b_k$$

The conjugate sits on the **first** argument and on nothing else. Swapping the arguments therefore conjugates the answer, $\langle b|a\rangle = \langle a|b\rangle^*$, and $\langle a|a\rangle = \sum_k |a_k|^2$ is real and never negative — which is what lets it be a squared length.

---

EXAMPLE 1.1 · AN OVERLAP, WITH THE CONJUGATE IN THE RIGHT PLACE

GIVEN $|a\rangle = \frac{1}{\sqrt{2}}(1,\ i)$ and $|b\rangle = \frac{1}{\sqrt{2}}(1,\ -i)$.

FIND $\langle a|b\rangle$.

METHOD Write the bra explicitly — conjugate every entry of the first vector — then multiply term by term and add.

SOLUTION $\langle a| = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -i \end{bmatrix}$, so $\langle a|b\rangle = \frac{1}{2}\left[(1)(1) + (-i)(-i)\right] = \frac{1}{2}\left[1 - 1\right] = 0$. The two states are orthogonal.

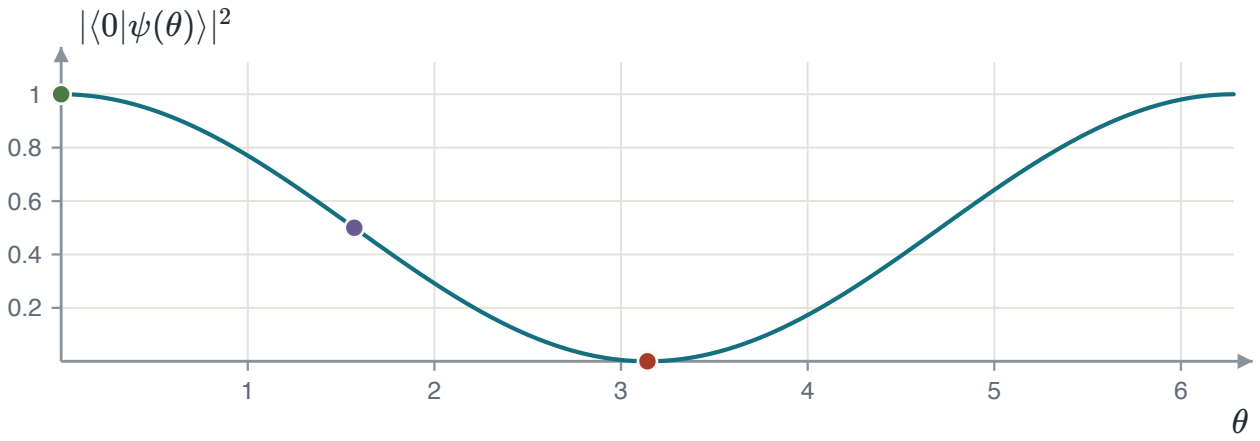
CHECK Both are normalised, so Cauchy–Schwarz allows anything from $0$ to $1$ in modulus. Zero is the smallest it could have been, and it says the two states are perfectly distinguishable.

---

THE MISTAKE THIS SECTION EXISTS TO STOP

Leaving the conjugate out gives $\frac{1}{2}\left[(1)(1) + (i)(-i)\right] = 1$. That is not a small error: an overlap of one says the two are **the same state**, where the truth is that they are as different as two states can be. In NumPy the conjugating product is `np.vdot(a, b)`; `np.dot(a, b)` is the version without it and returns the wrong number without complaint.

The inner product of a state with itself gives its length, $\|a\| = \sqrt{\langle a|a\rangle}$, and the inner product of two different states says how alike they are. The second is bounded: in any inner-product space the **Cauchy–Schwarz inequality** holds, $|\langle a|b\rangle| \le \|a\|\,\|b\|$. For two normalised states that reads $0 \le |\langle a|b\rangle|^2 \le 1$, and the two ends have names. Zero means **orthogonal** and a single measurement separates them with certainty. One means the same state up to a phase on the whole vector, and nothing separates them ever.




The squared overlap of $|0\rangle$ with the family $|\psi(\theta)\rangle = \cos(\theta/2)|0\rangle + \sin(\theta/2)|1\rangle$. The three marks are $\theta = 0$, where the states coincide; $\theta = \pi/2$, where the overlap is one half; and $\theta = \pi$, where they are orthogonal. The half angle is the first sighting of a factor chapter 4 explains.

A set $\{|e_1\rangle, \ldots, |e_n\rangle\}$ is an **orthonormal basis** when $\langle e_i|e_j\rangle = \delta_{ij}$ and it spans the space. Every vector then has one expansion in it, and the coefficients are computed rather than guessed. Apply $\langle e_j|$ to both sides of $|v\rangle = \sum_i v_i|e_i\rangle$:

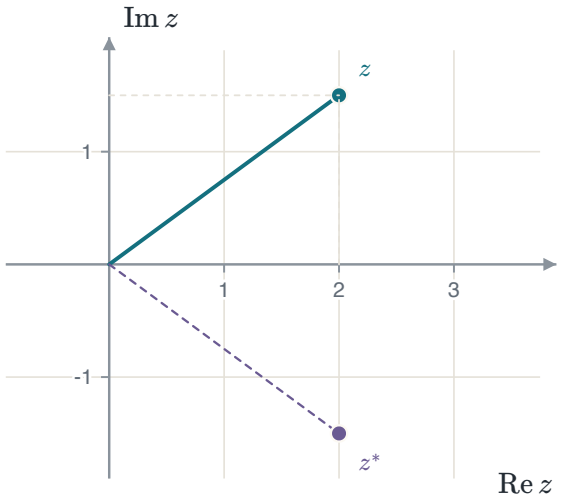
$$\langle e_j|v\rangle = \sum_i v_i \langle e_j|e_i\rangle = \sum_i v_i \, \delta_{ji} = v_j$$

One line, and it is used constantly: **a coefficient is an inner product with the corresponding basis vector**. It is why the basis is chosen orthonormal in the first place, and it fails as soon as the basis is not. It also makes clear that the coefficients belong to the basis and not to the state: writing $|0\rangle = \frac{1}{\sqrt{2}}|+\rangle + \frac{1}{\sqrt{2}}|-\rangle$ changes nothing about the state, and the word "superposition" is therefore incomplete on its own. A state is a superposition **with respect to a named basis**.

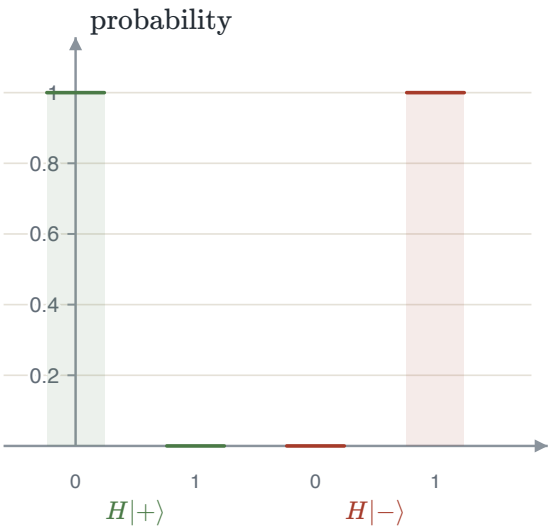
## 1.2 Amplitude, phase and interference

An amplitude is one complex number and it carries two real ones. In Cartesian form it is a real part and an imaginary part; in polar form it is a size and an angle, and Euler's formula $e^{i\varphi} = \cos\varphi + i\sin\varphi$ is the bridge between them.

> MODULUS AND PHASE
>
> $$z = x + iy = r\,e^{i\varphi}, \qquad z z^{*} = x^2 + y^2 = |z|^2$$
>
> Every probability in this course is a $|z|^2$, and every interference effect is a difference of two $\varphi$. The modulus becomes a probability; the phase never becomes a probability on its own, and only ever acts by being added to or subtracted from another phase.




A complex number and its conjugate. Conjugation is a reflection in the real axis. It is not a rotation by $\pi$, which would send $z$ to $-z$ instead.




After one Hadamard, $|+\rangle$ and $|-\rangle$ give opposite certain answers. Before the gate all four bars were one half, which is why the figure shows only the half of the story the relative phase decides.

WHERE THE PHASE GETS LOST

Computing the phase as $\arctan(y/x)$ throws the quadrant away: for $z = -1 + i$ it returns $-\pi/4$, which points at $1 - i$. The function that keeps the quadrant takes the two arguments separately, $\operatorname{atan2}(y, x)$, which is `np.angle` in NumPy. It also survives $x = 0$, where the ratio does not exist at all.

Now multiply a whole state by a phase and ask what changes. Take any other state $|\phi\rangle$ and form the overlap: $\left|\langle\phi|e^{i\gamma}\psi\rangle\right|^2 = \left|e^{i\gamma}\right|^2 |\langle\phi|\psi\rangle|^2 = |\langle\phi|\psi\rangle|^2$. Nothing changes, because $|e^{i\gamma}| = 1$, and every number an experiment can produce is of that form.

A GLOBAL PHASE IS NOT PHYSICAL

$$e^{i\gamma}|\psi\rangle \;\equiv\; |\psi\rangle$$

Put the phase on one term only and the conclusion reverses. The states $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$ differ by a relative phase of $\pi$. Both give $\left|\pm\frac{1}{\sqrt{2}}\right|^2 = \frac{1}{2}$ for each computational outcome, so a measurement in that basis cannot tell them apart. Apply a Hadamard, which sends $|0\rangle \mapsto (|0\rangle + |1\rangle)/\sqrt{2}$ and $|1\rangle \mapsto (|0\rangle - |1\rangle)/\sqrt{2}$, and collect terms:

$$H|+\rangle = \tfrac{1}{2}(|0\rangle + |1\rangle) + \tfrac{1}{2}(|0\rangle - |1\rangle) = |0\rangle$$
$$H|-\rangle = \tfrac{1}{2}(|0\rangle + |1\rangle) - \tfrac{1}{2}(|0\rangle - |1\rangle) = |1\rangle$$

THE RULE TO CARRY FORWARD

A phase on the **whole** state can always be dropped. A phase **between two terms** can never be dropped. Interference is the only mechanism any algorithm in this course has, and interference is entirely a statement about relative phase. The trap inside that rule: a phase that is global for one qubit stops being global as soon as that qubit is one branch of a larger superposition, and chapter 6 builds its whole mechanism out of exactly this.

## 1.3 Outer products, projectors and the identity

Reverse the order of the row and the column. A column on the left of a row does not contract; it spreads into a square array, and that array is an operator. This is the one construction that turns states into the things that act on states.

THE OUTER PRODUCT

$$|a\rangle\langle b| = |a\rangle\,|b\rangle^{\dagger}, \qquad (|a\rangle\langle b|)_{jk} = a_j\, b_k^*$$

What it does is easy to read: $(|a\rangle\langle b|)\,|v\rangle = \langle b|v\rangle\,|a\rangle$, because the bra eats the ket first. Every outer product therefore sends the whole space onto the one line through $|a\rangle$, and the number it multiplies by is the overlap of the input with $|b\rangle$.

| WRITTEN | SHAPE | WHAT IT IS |
|---|---|---|
| $\langle a\|b\rangle$ | $1 \times n$ times $n \times 1$ | a number |
| $\|a\rangle\langle b\|$ | $n \times 1$ times $1 \times n$ | an operator |
| $\|a\rangle \otimes \|b\rangle$ | $n \times 1$ with $m \times 1$ | a longer column, of length $nm$ |

Take the outer product of a normalised state with itself and the result is a **projector**, $P_u = |u\rangle\langle u|$. It is its own adjoint, because reversing an outer product swaps two copies of the same state, and applying it twice is the same as applying it once:

$$P_u^2 = |u\rangle\underbrace{\langle u|u\rangle}_{=1}\langle u| = P_u$$

An operator with $P^2 = P$ can only have eigenvalues $0$ and $1$, since $\lambda^2 = \lambda$. Those two numbers are "kept" and "discarded", and in chapter 2 they become the two things a measurement can report. A projector does not preserve length: for $P = |+\rangle\langle +|$ and $|v\rangle = |0\rangle$ the output $P|0\rangle = \frac{1}{2}(|0\rangle + |1\rangle)$ has length $1/\sqrt{2}$, and the lost length is exactly the probability of the other outcome.

Now take one projector for each vector of an orthonormal basis and add them. Nothing has been thrown away, so nothing has changed.

THE RESOLUTION OF THE IDENTITY

$$\sum_k |e_k\rangle\langle e_k| = I$$

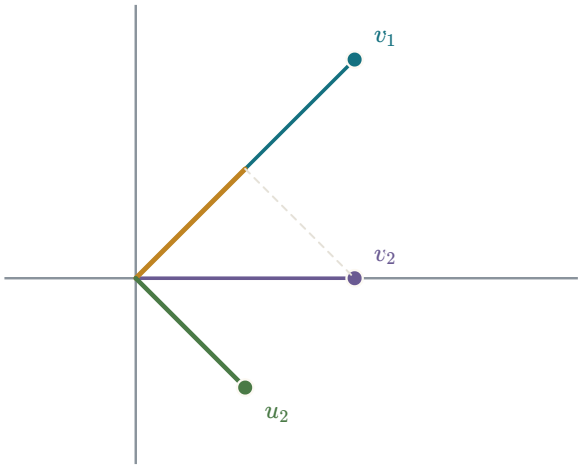
This is used as a **move** rather than as a fact. Inserting it turns a statement about a vector into a statement about its coefficients, and it derives the basis expansion of section 1.1 rather than assuming it: $|v\rangle = I|v\rangle = \sum_k |e_k\rangle\langle e_k|v\rangle$. The same move between two bras turns an inner product into a sum, $\langle a|b\rangle = \sum_k \langle a|e_k\rangle\langle e_k|b\rangle$, and the basis to insert is whichever one makes an object in the expression simple.

## 1.4 Building an orthonormal basis

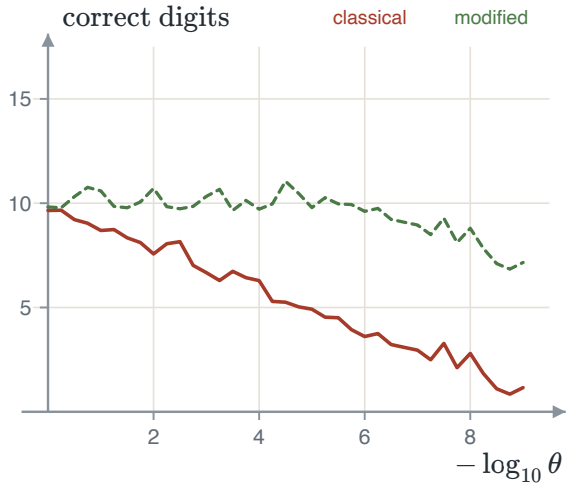
Given independent vectors $v_1, \dots, v_k$, the **Gram–Schmidt procedure** produces an orthonormal set spanning the same space. Each step removes from the next vector everything that already lies along the vectors already built, then divides by the length of what is left.

GRAM–SCHMIDT

$$u_j = v_j - \sum_{i<j} \langle e_i|v_j\rangle\, e_i, \qquad e_j = \frac{u_j}{\|u_j\|}$$




One step of the procedure. The amber arrow is the part of $v_2$ that already lies along $e_1$; removing it leaves $u_2$, which is at a right angle to $e_1$ by construction and only has to be scaled to length one.




How many digits of orthogonality survive three vectors in space, against how nearly parallel the first two are. The recursion above is run at every setting and the result measured; the modified recursion, which subtracts from what is left rather than from the original, is run on the same inputs.

EXAMPLE 1.2 · TWO VECTORS IN THE PLANE

GIVEN $v_1 = (1, 1)$ and $v_2 = (1, 0)$, real.

METHOD Normalise the first; subtract from the second the part along it; normalise what is left.

SOLUTION $\|v_1\| = \sqrt{2}$, so $e_1 = \frac{1}{\sqrt{2}}(1, 1)$. Then $\langle e_1|v_2\rangle = \frac{1}{\sqrt{2}}$ and $u_2 = (1, 0) - \frac{1}{2}(1, 1) = (\frac{1}{2}, -\frac{1}{2})$, of length $\frac{1}{\sqrt{2}}$, so $e_2 = \frac{1}{\sqrt{2}}(1, -1)$.

CHECK $\langle e_1|e_2\rangle = \frac{1}{2}(1 - 1) = 0$ and both have length one. The pair is $|+\rangle$ and $|-\rangle$ in disguise.

WHERE IT BREAKS ON A COMPUTER

If two input vectors are nearly parallel, the subtraction cancels almost everything and $u_j$ is a small difference of large numbers. Rounding error, negligible in the inputs, is then a large fraction of the result, and the "orthonormal" vectors come out neither orthogonal nor normal. A QR factorisation computes the same span by a different arithmetic route and loses far less; in NumPy that is `np.linalg.qr`, and the columns of $Q$ are the orthonormal set. Gram–Schmidt stays the right thing to know because it says **what** is being computed; QR is the right thing to run because it says how.

## 1.5 The tensor product, and where the exponential comes from

Two separate systems are described by one state, and the construction that builds it is the **tensor product**: every product of an entry of the first with an entry of the second.

THE TENSOR PRODUCT OF TWO COLUMNS

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \otimes \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} a_1 b_1 \\ a_1 b_2 \\ a_2 b_1 \\ a_2 b_2 \end{bmatrix}$$

Two qubits are therefore described by four numbers and not by two plus two. For matrices the rule is the same block by block: $A \otimes B$ carries the block $A_{ij}B$ in position $(i, j)$.

THE CONVENTION THAT FAILS SILENTLY

This course writes a register as $|q_{n-1} \dots q_1 q_0\rangle$, and entry $x$ of the column is the amplitude of $|x\rangle$ read as a binary number. So $|10\rangle$ means $q_1 = 1$ and $q_0 = 0$, and it is entry $2$. Courses that count the other way write the same symbols and mean a different state. Nothing looks wrong when the two are mixed; every number afterwards is simply another problem's answer.

The tensor product multiplies dimensions, so adding a qubit doubles the space and $n$ of them give $\dim(\mathbb{C}^2)^{\otimes n} = 2^n$. That is the entire origin of the exponential this subject is famous for. It is not a physical postulate, it is not an assumption about speed, and it is not special to quantum mechanics: it is what "combine two systems" means, applied $n$ times.


Why the count doubles. Each new qubit gives every basis string of the register two continuations rather than one, so the number of strings goes $1, 2, 4, 8$ and the amplitude column grows with it.

> AND WHAT THE SIZE DOES NOT BUY
>
> A measurement of $n$ qubits returns $n$ bits — one string, not $2^n$ amplitudes and not a summary of them. Holding many amplitudes is therefore worth nothing on its own, and an algorithm has to make the unwanted ones cancel **before** the readout. Nor does a large state space prove the physics is hard to simulate: classical methods use locality, symmetry, sparsity, low entanglement and stabilizer structure, and a great many quantum states fall to one of them. The dimension is a fact about a generic dense description, and a reason the problem might be interesting rather than a proof that it is.

## 1.6 Hermitian and unitary operators

The **adjoint** of an operator is its conjugate transpose, $A^\dagger = (A^*)^T$ — the same operation that turned a ket into a bra, applied to a square array. Two things happen and both are needed: every entry is conjugated, and every entry is moved across the diagonal. In NumPy that is `A.conj().T`; `A.T` alone is the real-valued version and drops the conjugation without warning. The adjoint reverses a product, $(AB)^\dagger = B^\dagger A^\dagger$.

An operator equal to its own adjoint is **Hermitian**. Its diagonal entries are real and its off-diagonal entries come in conjugate pairs, both visible at a glance. The consequence that matters is one short argument. Let $A|v\rangle = \lambda|v\rangle$ with $|v\rangle$ normalised, and compute $\langle v|A|v\rangle$ twice. Reading the operator to the right gives $\lambda$; reading it to the left uses $A^\dagger = A$ and gives $\lambda^*$:

$$\langle v|A|v\rangle = \langle v|\lambda v\rangle = \lambda, \qquad \langle v|A|v\rangle = \langle Av|v\rangle = \lambda^* \implies \lambda \in \mathbb{R}$$

A measuring instrument reports a real number, so an operator whose job is to carry the list of readings it can produce must have real eigenvalues, and Hermiticity is the condition that guarantees it. The same theorem gives an orthonormal basis of eigenvectors, which is what makes two different readings perfectly distinguishable.

Ask instead for the operators that leave every overlap alone, and the condition writes itself. Apply $U$ to both arguments and move one copy across the inner product: $\langle Ua|Ub\rangle = \langle a|U^\dagger U|b\rangle$. For this to equal $\langle a|b\rangle$ for every pair, the operator in the middle must be the identity.

UNITARY

$$U^\dagger U = I \qquad \Longleftrightarrow \qquad U^{-1} = U^\dagger$$

Three consequences follow at once. Lengths are preserved, so a normalised state stays normalised and the probabilities of chapter 2 keep adding to one. The operator is invertible, so the evolution is reversible. And the columns of $U$ are an orthonormal basis, which is what the condition says entry by entry.

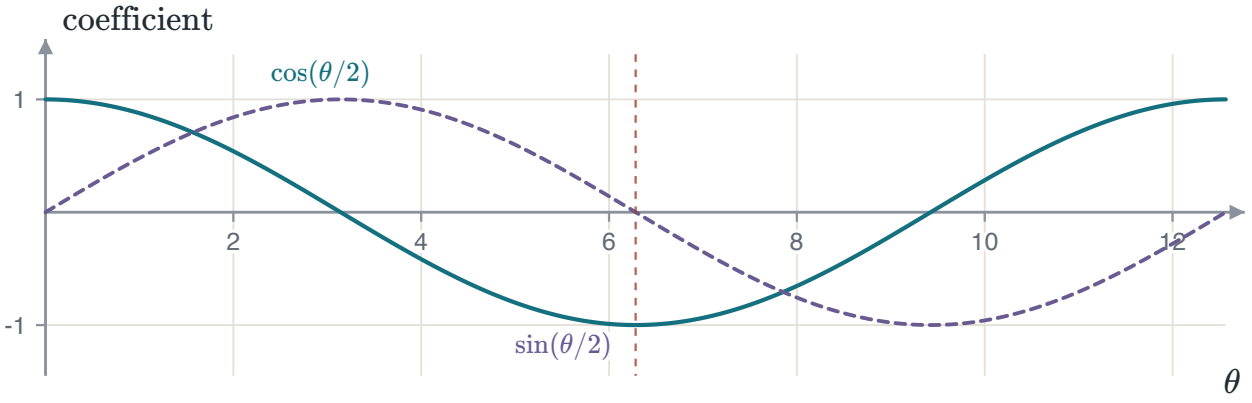
A DETERMINANT OF MODULUS ONE IS NOT ENOUGH

The shear $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ has determinant $1$ and is invertible, and it is not unitary: it sends $(0,1)$ to $(1,1)$, of length $\sqrt{2}$. Unitarity is a statement about every vector, and the only way to test it is $U^\dagger U = I$.

The two kinds of operator are joined by an exponential. If $G$ is Hermitian and $\theta$ is real then $U(\theta) = e^{-i\theta G}$ is unitary, because $U^\dagger = e^{+i\theta G}$ and the two exponentials share a generator, so their exponents add to zero. For a Pauli operator the series closes in two terms: $\sigma^2 = I$ makes every even power the identity and every odd power $\sigma$, and the two resulting sums are a cosine and a sine.

A PAULI ROTATION

$$e^{-i\theta\sigma/2} = \cos\!\left(\tfrac{\theta}{2}\right) I \;-\; i \sin\!\left(\tfrac{\theta}{2}\right) \sigma$$




The two coefficients over two full turns. At $\theta = 2\pi$ the cosine is $-1$ and the sine is $0$, so the operator is $-I$: the matrix takes twice as long to come home as the state it acts on. The leftover sign is a global phase and no experiment sees it, which is why the two are consistent.

---

EXAMPLE 1.3 · A ROTATION ABOUT $z$

GIVEN $\sigma = Z = \mathrm{diag}(1,-1)$ and $\theta = \pi/3$.

METHOD Substitute into the closed form; no series and no diagonalisation are needed.

SOLUTION $\cos(\pi/6) = \frac{\sqrt{3}}{2}$ and $\sin(\pi/6) = \frac{1}{2}$, so $R_z(\pi/3) = \mathrm{diag}\left(e^{-i\pi/6},\, e^{i\pi/6}\right)$.

CHECK Each diagonal entry has modulus one and the matrix is diagonal, so $U^\dagger U = I$. The relative phase between the two entries is $\pi/3$, which is the $\theta$ that went in.

---

Chapter 2 will say that a closed system evolves as $U(t) = e^{-iHt}$ with $H$ the Hamiltonian. None of that is new mathematics: it is this section with $\theta G$ replaced by $Ht$. What chapter 2 adds is the claim that nature does it.

## 1.7 The spectral theorem and functions of an operator

Most vectors are turned by an operator; a few come back pointing the same way, scaled by a number. Those are its **eigenvectors** and the numbers are its **eigenvalues**, and $(A - \lambda I)|v\rangle = 0$ has a non-zero solution only when $\det(A - \lambda I) = 0$. For a $2 \times 2$ that is $\lambda^2 - (\operatorname{tr} A)\lambda + \det A = 0$, whose two coefficients are read straight off the matrix. An eigenvector is a direction rather than a vector: normalising leaves only a phase, and for a quantum state that phase is global and unobservable, so an eigenvector really is one physical state.

A Hermitian operator has an orthonormal basis of eigenvectors. Take the projector onto each eigenspace and weight it by that eigenvalue, and the operator comes back.

> THE SPECTRAL DECOMPOSITION
>
> $$A = \sum_k \lambda_k P_k, \qquad P_k = |v_k\rangle\langle v_k|$$
>
> $$P_j P_k = \delta_{jk} P_k, \qquad \sum_k P_k = I$$
>
> The second line is the resolution of the identity again, now in the eigenbasis of a particular operator. This is the shape every measurement in chapter 2 has: one outcome per eigenvalue, with $P_k$ deciding both how likely that outcome is and what the state becomes afterwards.

Once an operator is written that way, a function of it means one thing only: apply the function to the eigenvalues and leave the projectors alone.

> A FUNCTION OF AN OPERATOR
>
> $$f(A) = \sum_k f(\lambda_k)\, P_k$$
>
> This agrees with the obvious definition wherever the obvious one exists. For a power, $A^2 = \sum_{j,k} \lambda_j \lambda_k P_j P_k$ collapses to $\sum_k \lambda_k^2 P_k$, so every power series follows — and the definition may then be used for functions with no series at all, such as a square root. The case this course needs constantly is $e^{-iAt} = \sum_k e^{-i\lambda_k t} P_k$, which is what makes an evolution operator computable at all.

EXAMPLE 1.4 · A SPECTRAL DECOMPOSITION END TO END

GIVEN $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$.

FIND Its eigenvalues, its projectors, and $e^{-iAt}$.

METHOD Characteristic equation; then $(A - \lambda I)|v\rangle = 0$; then the weighted sum; then the function on the eigenvalues.

SOLUTION $(2-\lambda)^2 - 1 = 0$ gives $\lambda = 3$ on $|+\rangle$ and $\lambda = 1$ on $|-\rangle$. So $A = 3\,|+\rangle\langle+| + 1\,|-\rangle\langle-|$ and $e^{-iAt} = e^{-3it}|+\rangle\langle+| + e^{-it}|-\rangle\langle-|$.

CHECK The eigenvalues add to the trace, $3 + 1 = 4 = 2 + 2$, and multiply to the determinant, $3 \times 1 = 3 = 4 - 1$. At $t = 0$ both exponentials are one and the sum is $P_+ + P_- = I$, as any evolution operator must be at zero time.

A FUNCTION OF AN OPERATOR IS NEVER APPLIED ENTRY BY ENTRY

For $X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ the correct $e^X = \cosh(1)I + \sinh(1)X$ has entries $1.5431$ and $1.1752$ and eigenvalues $e$ and $e^{-1}$. Exponentiating the entries gives $\begin{bmatrix} 1 & e \\ e & 1 \end{bmatrix}$, whose eigenvalues are $1 \pm e$ — one of them negative, so it is not the exponential of any Hermitian matrix. The two agree only for a diagonal matrix, which is why testing on one teaches the wrong lesson.

## 1.8 Dirac notation

Nothing in this chapter needed the notation. Every result was a statement about columns, rows and matrices and could have been written that way throughout. The notation earns its place because it says which object is which without naming a basis, and because the shapes then check themselves.

| WRITTEN | ONCE A BASIS IS CHOSEN | WHAT IT IS |
|---|---|---|
| $\|\psi\rangle$ | a column of $n$ complex numbers | a state |
| $\langle\psi\|$ | the same column, conjugated and laid on its side | a row |
| $\langle\phi\|\psi\rangle$ | row times column | one complex number |
| $\|\phi\rangle\langle\psi\|$ | column times row | an $n \times n$ operator |
| $A\|\psi\rangle$ | matrix times column | another state |
| $\langle\psi\|A\|\psi\rangle$ | row times matrix times column | the expectation value of chapter 2 |

A matrix product acts on a ket from the right, so the operator written **last** in a product is applied **first**: a circuit that runs $H$, then $R_z$, then $H$ is the operator $HR_zH$. Circuit diagrams run left to right and algebra runs right to left, and the two are not in conflict — they are two orders for the same sequence. An operator can also be written entirely in the notation, with no matrix anywhere, by inserting the identity on both sides: $A = \sum_{j,k} |e_j\rangle\langle e_j|A|e_k\rangle\langle e_k|$, where $\langle e_j|A|e_k\rangle$ is the matrix entry $A_{jk}$. The matrix is not a different object from the operator; it is the operator, read in a basis.

## 1.9 Square-integrable functions, completeness and truncation

A column labels its entries with integers. A function labels them with a continuous coordinate. The same vector-space construction therefore uses an integral in place of a sum. A function belongs to $L^2[a,b]$ when its squared modulus has a finite integral, which gives it a finite norm and allows it to be normalised.

THE FUNCTION-SPACE INNER PRODUCT

$$\langle f|g\rangle = \int_a^b f^*(x)g(x)\,\mathrm{d}x$$

$$\|f\|^2 = \int_a^b |f(x)|^2\,\mathrm{d}x$$

For example, $\sin x/\sqrt{\pi}$ and $\cos x/\sqrt{\pi}$ are orthonormal on $[-\pi,\pi]$. Their curves cross, but orthogonality is decided by the integral of their product.

An orthonormal family is **complete** when no non-zero vector is orthogonal to every member. Then the inner products with the basis recover the function. Fourier expansion is this statement for the normalised constant, sine and cosine functions on a finite interval.

COMPLETENESS, PARSEVAL AND TRUNCATION

$$|f\rangle = \sum_{n=1}^{\infty} c_n|u_n\rangle, \qquad c_n = \langle u_n|f\rangle$$

$$\|f\|^2 = \sum_{n=1}^{\infty} |c_n|^2, \qquad \left\| f - \sum_{n=1}^{N} c_n u_n \right\|^2 = \sum_{n>N} |c_n|^2$$

Parseval turns the coefficient tail into an exact squared error for the truncated approximation. It controls convergence in norm, not uniform or pointwise convergence at every coordinate.

TWO TESTS THAT A PLOT CANNOT REPLACE

A function can be finite at every point and still have an infinite squared norm on an unbounded interval. A truncated Fourier plot can also look accurate while its pointwise error is poor at selected coordinates. Test the integral for membership in $L^2$ and use the coefficient tail for the norm error.

## 1.10 Summary

- A state is a normalised column of complex amplitudes. Its coefficients in a basis are inner products with that basis, and they change when the basis does while the state does not.
- The inner product conjugates its first argument. Nearly every wrong answer in this chapter is a missing conjugate.
- A phase on the whole state is not physical and may be dropped. A phase between two terms is physical and may never be dropped.
- A ket beside a bra is an operator; the projectors of an orthonormal basis add to the identity, and inserting that identity is a derivation step with a name.

- Two systems make one by the tensor product, so dimensions multiply and $n$ qubits carry $2^n$ amplitudes. A measurement returns $n$ bits.
- Hermitian means real eigenvalues and an orthonormal eigenbasis, which is what an observable needs. Unitary means every inner product is preserved, which is what a gate needs. The exponential of a Hermitian operator is unitary.
- A function of an operator acts on its eigenvalues and leaves its projectors alone.
- Square-integrable functions are vectors with an integral inner product. In a complete orthonormal basis, Parseval gives both the norm and the exact truncation error.

SIX LINES TO BE ABLE TO WRITE WITHOUT LOOKING

$$\langle a|b\rangle = \sum_k a_k^* b_k \;\cdot\; v_j = \langle e_j|v\rangle \;\cdot\; \sum_k |e_k\rangle\langle e_k| = I \;\cdot\; A = \sum_k \lambda_k P_k \;\cdot\; f(A) = \sum_k f(\lambda_k) P_k \;\cdot\; e^{-i\theta\sigma/2} = \cos(\theta/2) I - i\sin(\theta/2)\sigma.$$

FOUR ERRORS THAT COST A WHOLE QUESTION

A missing conjugate in an inner product. A phase cancelled as global when it sat on one branch of a superposition. A function of a matrix applied entry by entry. And a two-qubit state written in the other bit order, which does not make an answer approximately wrong — it makes it an answer to a different question.

2

# States, measurement and dynamics

*Chapter 1 built a language and nothing in it could have been argued with. This chapter adds the four statements that connect that language to a laboratory. They are postulates: they are not derived from anything, they are what experiment has found to hold, and everything after them follows.*

Two of the four were already used. The **state** postulate says a pure state is a normalised vector, with two vectors differing by a global phase describing the same physical state; the **composition** postulate says two systems combine by the tensor product. This chapter is about the other two: how a state evolves, and what happens when it is read.

EVOLUTION AND MEASUREMENT

**Evolution.** A closed system evolves by a unitary operator, and that operator is the exponential of a Hermitian one: $|\psi(t)\rangle = e^{-iHt}|\psi(0)\rangle$. **Measurement.** A measurement returns one outcome, with a probability fixed by the state, and it leaves the state changed. Everything in this chapter is one of those two.

## 2.1 The Born rule

An amplitude is not observable. The rule that turns one into something a laboratory can count is the only bridge in the subject between the mathematics and an experiment, and it is one line. Measure in an orthonormal basis $\{|n\rangle\}$; the probability of outcome $n$ is the squared modulus of the coefficient.

THE BORN RULE

$$p(n) = |\langle n|\psi\rangle|^2$$

The coefficients of chapter 1 now have a name: **probability amplitudes**. That the probabilities add to one is a theorem rather than a second postulate — insert the resolution of the identity into $\langle\psi|\psi\rangle = 1$ and the sum $\sum_n |c_n|^2 = 1$ falls out. Normalisation was the bookkeeping; this is what it was bookkeeping for.

THE MODULUS, NOT THE SQUARE

$p(n)$ is $|c_n|^2$ and never $c_n^2$. For $c = 4i/5$ the square is $-16/25$, a negative probability. Squaring the coefficient instead of its modulus is the commonest first error in the subject, and a negative or complex probability is always this mistake and no other.

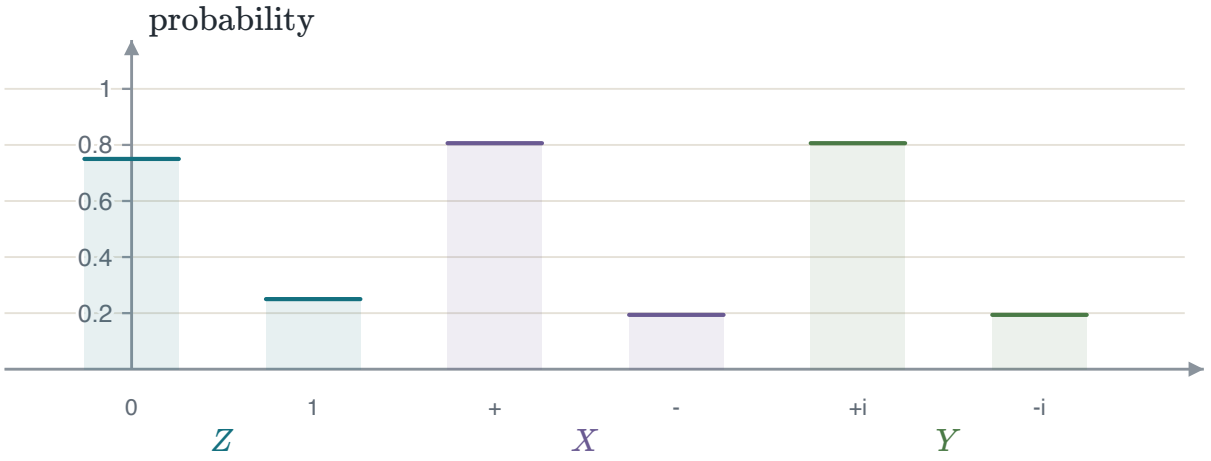
Two things in this course are both called "changing the basis" and they are not the same. A **passive change of coordinates** rewrites the same state and the same operator in a different orthonormal basis, and every physical prediction is unchanged. **Choosing what to measure** is a different matter: an instrument whose outcomes are the $X$ eigenstates is a different instrument from one whose outcomes are the $Z$ eigenstates, and it returns a different distribution from the same state.

PASSIVE COORDINATES AGAINST AN ACTIVE TRANSFORMATION

$$|\psi\rangle' = V^\dagger|\psi\rangle, \qquad A' = V^\dagger A V, \qquad \langle\psi|A|\psi\rangle = \langle\psi'|A'|\psi'\rangle$$

$$|\psi\rangle \longmapsto U|\psi\rangle \qquad \text{with the coordinates and } A \text{ fixed}$$

The first line changes only the numbers used to describe the same state and observable, so every prediction is invariant. The second is active: a gate or physical evolution moves the state, and its probabilities generally change. Choosing a new measurement basis changes the projectors of the apparatus and is a third operation.




One state, read by three different instruments. The three groups are the eigenbases of $Z$, $X$ and $Y$, and the bars are computed from the state itself. Nothing about the state changed between them; the apparatus did.

AND WHAT A CHANGE OF BASIS COSTS

Real hardware measures in one basis only, almost always the computational one. Measuring $X$ means applying a gate that rotates the $X$ eigenstates onto the computational ones and then measuring as usual. A change of measurement basis therefore costs a gate, that gate has an error, and a scheme that needs many bases pays for each of them.

Chapter 1 said that an overlap of zero means two states are perfectly distinguishable. The argument is short. If $\langle a|b\rangle = 0$, measure with $P_a = |a\rangle\langle a|$ and $I - P_a$: on $|a\rangle$ the first outcome is certain, and on $|b\rangle$ it has probability $|\langle a|b\rangle|^2 = 0$. If they are not orthogonal, every outcome is one both states can produce, and one copy is not enough to be certain. The best guess succeeds with probability $\frac{1}{2}(1 + \sin\theta)$, where $\cos\theta = |\langle a|b\rangle|$.

## 2.2 Projective measurement

The Born rule was written for an orthonormal basis. The general form replaces the basis vectors by the projectors of the spectral decomposition, and says the same thing wherever both apply.

A PROJECTIVE MEASUREMENT

$$p(a) = \langle\psi|P_a|\psi\rangle$$

$$P_j P_k = \delta_{jk} P_k, \qquad \sum_a P_a = I$$

The two conditions are not decoration. Orthogonality makes the outcomes exclusive and completeness makes the probabilities add to one; drop either and the numbers stop being a distribution. When every projector has rank one this reduces to $p(a) = |\langle a|\psi\rangle|^2$. The new case is a repeated eigenvalue, whose projector has higher rank: the instrument then reports a number and cannot tell apart the states inside that eigenspace.

A measurement also changes the state. Given the outcome $a$, what is left is the projected state, put back to length one.

THE UPDATE RULE

$$|\psi_a\rangle = \frac{P_a|\psi\rangle}{\sqrt{p(a)}}$$

The division is the renormalisation chapter 1 deliberately did not do: a projector shortens a state, and the length it removes is the probability of the other outcomes. Measure the same observable again immediately and $P_a|\psi_a\rangle = |\psi_a\rangle$, so the same reading comes back with certainty. That is what makes a reading mean something, and it is what the word **projective** names.

---

EXAMPLE 2.1 · THREE MEASUREMENTS IN A ROW

GIVEN — A qubit in $|0\rangle$. Measure $Z$, then $X$, then $Z$ again.

FIND — The probability that the third reading agrees with the first.

METHOD — Follow each branch: probability, updated state, next measurement.

SOLUTION — The first is $+1$ with certainty. The second gives $\pm 1$ with probability $\frac{1}{2}$ each, leaving $|+\rangle$ or $|-\rangle$. From either, the third gives $\pm 1$ with probability $\frac{1}{2}$. So the third agrees with the first with probability $\frac{1}{2}$.

CHECK — Omit the middle measurement and the answer is one, because an immediate repeat is certain. The difference is entirely $[Z, X] \neq 0$, and no noise was added anywhere in the story.

---

A projective measurement is not the most general one an apparatus can perform. The general form keeps only what is needed for the probabilities to be a distribution: a set of positive operators called **effects** that add to the identity, with $p(m) = \langle\psi|E_m|\psi\rangle$. Projectors are the special case. Effects need not be orthogonal and there need not be as many of them as the space has dimensions.

A POVM GIVES PROBABILITIES AND AN INSTRUMENT GIVES THE STATE

$$E_m = \sum_\alpha M_{m\alpha}^\dagger M_{m\alpha}, \qquad p(m) = \mathrm{Tr}(\rho E_m)$$

$$\rho_m = \frac{\sum_\alpha M_{m\alpha}\rho M_{m\alpha}^\dagger}{p(m)}$$

The effects alone do not determine the conditional state. Two devices can have the same POVM and disturb the system differently because their instruments differ. A projective measurement is the special case $M_m = P_m$.

---

EXAMPLE 2.2 · A READOUT THAT SOMETIMES LIES

GIVEN — A symmetric assignment error $\epsilon$: $E_0 = (1-\epsilon)|0\rangle\langle 0| + \epsilon|1\rangle\langle 1|$ and $E_1 = I - E_0$.

FIND — What the instrument reports for a state with true $p(0) = q$.

SOLUTION — $p_{\mathrm{rep}}(0) = (1-\epsilon)q + \epsilon(1-q) = \epsilon + (1-2\epsilon)q$: a straight line of slope $1-2\epsilon$.

CHECK — At $\epsilon = 0$ it is $q$; at $\epsilon = \frac{1}{2}$ it is $\frac{1}{2}$ whatever $q$ is, and the instrument has stopped reporting anything about the state.

---

INVERTING THE CALIBRATION IS NOT FREE

That line can be inverted to recover $q$, which is what readout-error mitigation does. Dividing by $1-2\epsilon$ removes the bias and multiplies the statistical error by the same factor. The correction is exact in expectation and noisier in practice, and a report that gives the corrected number without the widened error bar has not reported the measurement it made.

## 2.3 Observables, means and spreads

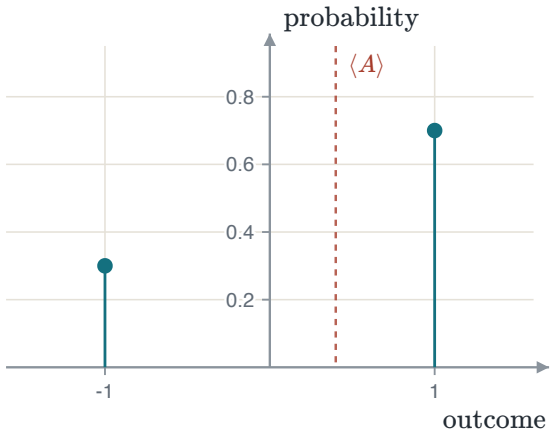
An observable is a Hermitian operator, its eigenvalues are the numbers an instrument can report, and its projectors say how often. The average of many readings is one sandwich, and that it agrees with the average is a calculation rather than a definition: expand $A$ spectrally and use the Born rule on each term.

EXPECTATION AND VARIANCE

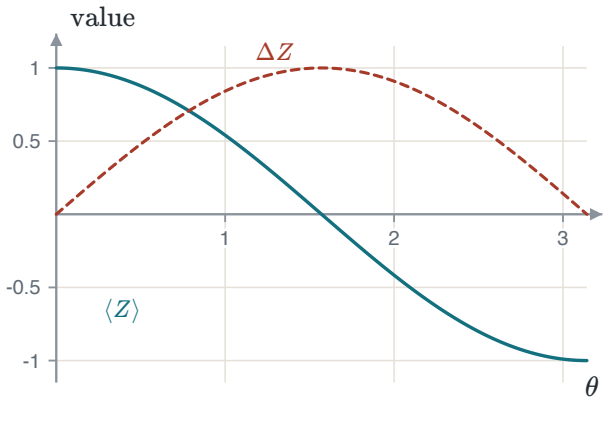
$$\langle A \rangle = \langle\psi|A|\psi\rangle = \sum_a a\, p(a)$$

$$\mathrm{Var}(A) = \langle A^2\rangle - \langle A\rangle^2$$

Note which operator is squared where: $\langle A^2\rangle$ squares the matrix before the sandwich, and $\langle A\rangle^2$ squares the number afterwards. Their difference is the whole quantity. For any Pauli, $A^2 = I$, so $\mathrm{Var}(A) = 1 - \langle A\rangle^2$ and no second sandwich is needed.




A qubit observable with outcomes $-1$ and $+1$, and the mean of many readings. The instrument returns one of the two stems and never the dashed line; the dashed line is where the stems balance.




The mean and the spread of $Z$ over the family $\cos(\theta/2)|0\rangle + \sin(\theta/2)|1\rangle$. At the two ends the state is an eigenstate: the spread is zero and the mean is the eigenvalue.

AN EXPECTATION VALUE IS NOT A READING

An instrument measuring $Z$ returns $+1$ or $-1$. It never returns $0.6$, and no single shot ever will. $\langle Z\rangle = 0.6$ is a statement about a long run of identically prepared systems, and reporting it without the shot count and the error bar is reporting a number nobody measured. The spread vanishes exactly on the eigenstates: $\Delta A = 0$ if and only if $A|\psi\rangle = a|\psi\rangle$.

## 2.4 Compatibility and uncertainty

Two observables can both be sharp on one state exactly when they share an eigenbasis, and there is a test for that which needs no eigenvectors: the commutator $[A, B] = AB - BA$ vanishes. One direction is immediate, since diagonal matrices commute; the other is the theorem, and it holds for Hermitian operators in finite dimensions.

The commutator does more than answer yes or no. It bounds how sharp two readings can be at once, on any state at all. The proof is the Cauchy–Schwarz inequality of chapter 1, applied to the two shifted operators $A - \langle A\rangle I$ and $B - \langle B\rangle I$.

THE ROBERTSON RELATION

$$\Delta A \, \Delta B \;\geq\; \tfrac{1}{2}\,|\langle [A, B]\rangle|$$

For $X$ and $Z$ the commutator is $-2iY$, so the bound is $|\langle Y\rangle|$ — a number that depends on the state. On $|0\rangle$ it is zero and the relation says nothing, correctly, since $\Delta Z = 0$ there makes the left side zero too. On $|+i\rangle$ it is one, and the product of the two spreads is also one: the relation is saturated and no state does better.

IT IS NOT ABOUT CLUMSY INSTRUMENTS

The relation says nothing about disturbing a system by looking at it. $\Delta A$ is the spread of the readings of $A$ over many runs of **one** preparation, and $\Delta B$ the same for $B$ over a separate set of runs of the same preparation. No single experiment measures both, and none has to: the claim is about two distributions. For position and momentum, $[x, p] = i\hbar I$ gives the constant $\hbar/2$; the qubit case, where the bound moves with the state, is the general one and the constant is the accident.

## 2.5 The Pauli algebra

Three matrices carry almost all of the one-qubit course. Each is Hermitian, so each is an observable; each is also unitary, so each is a gate. That coincidence is particular to them, and it is why they appear on both sides of every calculation here.

| OBSERVABLE | $+1$ EIGENSTATE | $-1$ EIGENSTATE |
|---|---|---|
| $X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ | $\lvert+\rangle = \frac{1}{\sqrt{2}}(1, 1)$ | $\lvert-\rangle = \frac{1}{\sqrt{2}}(1, -1)$ |
| $Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$ | $\lvert+i\rangle = \frac{1}{\sqrt{2}}(1, i)$ | $\lvert-i\rangle = \frac{1}{\sqrt{2}}(1, -i)$ |
| $Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ | $\lvert0\rangle = (1, 0)$ | $\lvert1\rangle = (0, 1)$ |

Each squares to the identity and each is traceless, and those two facts force the eigenvalues: a traceless operator whose square is the identity has eigenvalues summing to zero and squaring to one, which leaves $+1$ and $-1$, one of each. Multiplying two of them never needs a matrix.

THE PAULI PRODUCT RULE

$$\sigma_i \sigma_j = \delta_{ij}\, I + i \sum_k \varepsilon_{ijk}\, \sigma_k$$

$$[\sigma_i, \sigma_j] = 2i \sum_k \varepsilon_{ijk} \sigma_k, \qquad \{\sigma_i, \sigma_j\} = 2\delta_{ij} I$$

Read the first line in two halves. Same index: the identity. Different indices: $i$ times the third, with a plus sign in the cyclic order $X \to Y \to Z \to X$ and a minus against it. Two different Pauli operators therefore **anticommute**, and that single fact does most of the work in chapters 4 and 5.

An instrument can be aimed along any direction. For a unit vector $\mathbf{n}$ the observable is $\mathbf{n} \cdot \boldsymbol{\sigma}$, its square is the identity, its eigenvalues are again $\pm 1$, and its projectors are one line: $P_\pm = \frac{1}{2}(I \pm \mathbf{n} \cdot \boldsymbol{\sigma})$. Collecting the three Pauli means of the state into a real vector $r_a = \langle \sigma_a \rangle$ turns every single-qubit measurement question into a dot product.

A MEASUREMENT ALONG A DIRECTION

$$p(\pm) = \tfrac{1}{2}\left(1 \pm \mathbf{n}\cdot\mathbf{r}\right)$$

A pure state has $|\mathbf{r}| = 1$, so three numbers describe it completely, and they are measurable: run the circuit three times over, once in each basis. Writing $\mathbf{n}\cdot\mathbf{r} = \cos\alpha$ gives $p(+) = \cos^2(\alpha/2)$, and $\alpha$ is the angle between two **vectors** — twice the angle between the corresponding states. Two orthogonal states sit at opposite ends of the sphere, not at a right angle. That factor of two is the double cover of chapter 1, appearing as geometry.

## 2.6 Dynamics

The function-space vectors of section 1.9 become physical wavefunctions when the coordinate is position. In one dimension, position multiplies a wavefunction and momentum differentiates it. A free particle has kinetic energy only.

COORDINATE REPRESENTATION AND A FREE PARTICLE

$$(\hat{x}\psi)(x) = x\psi(x), \qquad (\hat{p}\psi)(x) = -i\hbar\frac{\mathrm{d}\psi}{\mathrm{d}x}$$

$$\hat{H}_0 = \frac{\hat{p}^2}{2m} = -\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2}$$

The plane wave $e^{ikx}$ has momentum $p = \hbar k$ and energy $E = \hbar^2 k^2/(2m)$. It extends across all space and is not square-integrable, so a physical free-particle state is a normalisable wave packet built as a continuous superposition of plane waves.

Confining the particle to $0 < x < a$ with infinite walls keeps the free-particle equation inside the well and adds the boundary conditions $\phi(0) = \phi(a) = 0$. Those conditions select a discrete set of standing waves.

THE INFINITE SQUARE WELL

$$\phi_n(x) = \sqrt{\frac{2}{a}}\sin\!\left(\frac{n\pi x}{a}\right), \qquad n = 1, 2, \ldots$$

$$E_n = \frac{\hbar^2\pi^2 n^2}{2ma^2}$$

There is no $n = 0$ state because it is the zero function. Each energy eigenstate has a time-independent probability density. A superposition of two levels has a moving cross term at angular frequency $(E_m - E_n)/\hbar$. For the first two levels, $E_2 = 4E_1$ and the period is $2\pi\hbar/(3E_1)$.

A closed system evolves by a differential equation with one operator in it, the **Hamiltonian**. With $\hbar = 1$, and for a Hamiltonian that does not depend on time, the solution is the exponential of chapter 1 and no new mathematics is needed.

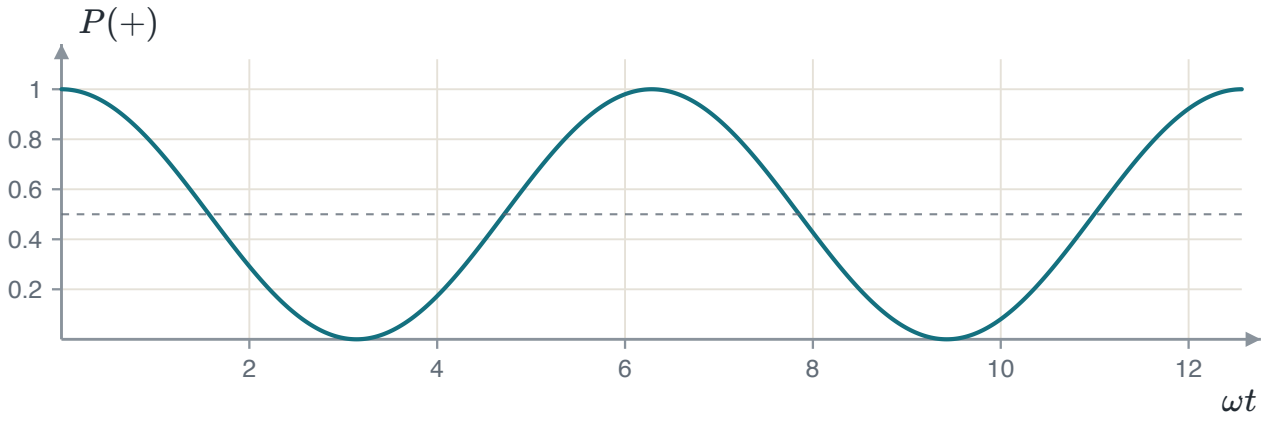
CLOSED-SYSTEM EVOLUTION

$$i\,\frac{\mathrm{d}}{\mathrm{d}t}|\psi(t)\rangle = H\,|\psi(t)\rangle$$

$$|\psi(t)\rangle = U(t)|\psi(0)\rangle, \qquad U(t) = e^{-iHt}$$

$H$ is Hermitian because it is the observable called energy, and the exponential of a Hermitian operator is unitary. So the evolution preserves every inner product and the probabilities keep adding to one at every time, and nothing had to be imposed to make that true. Replacing $H$ by $H + cI$ multiplies $U$ by a global phase, so only energy **differences** are physical.

Start in an eigenstate of the Hamiltonian and the exponential returns a phase, $|E_n(t)\rangle = e^{-iE_nt}|E_n\rangle$. That is a global phase, so no probability changes ever: such a state is **stationary**, and the name is exact. Start in a superposition of two of them and each term picks up its own phase; pulling out the first as a global one leaves a relative phase turning at the difference $E_2 - E_1$.




A qubit started in $|+\rangle$ under $H = \frac{\omega}{2}Z$, watched in the $X$ basis. The two energy eigenstates are each stationary and their sum is not. In the computational basis nothing at all happens: $P(0) = \frac{1}{2}$ at every time, and a reader watching only that basis would report a dead qubit.

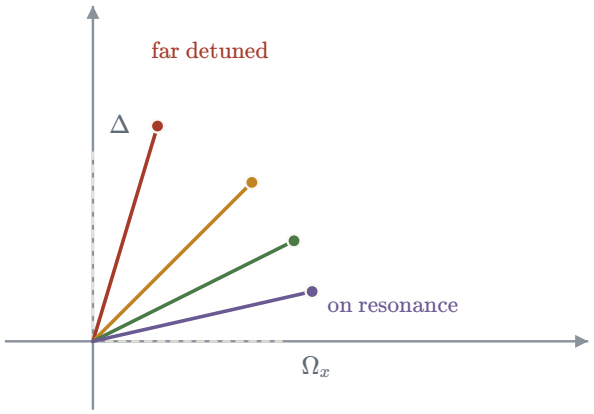
A qubit driven by an engineered field has, in the frame the drive defines, a Hamiltonian built from the three Pauli operators and nothing else. Since $(\mathbf{n}\cdot\boldsymbol{\sigma})^2 = I$, the closed form applies at once, and a gate is not a separate kind of object bolted onto the theory: it is what a Hamiltonian does for a chosen length of time.

A DRIVEN QUBIT

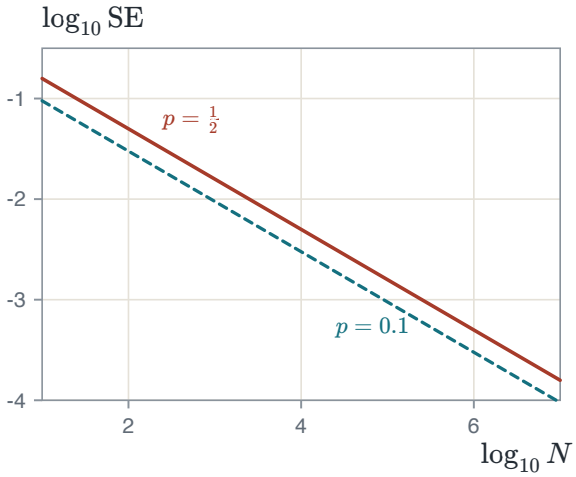
$$H = \tfrac{1}{2}\left(\Omega_x X + \Omega_y Y + \Delta Z\right) = \tfrac{\Omega}{2}\,\mathbf{n}\cdot\boldsymbol{\sigma}$$

$$U(t) = \cos\!\left(\tfrac{\Omega t}{2}\right) I - i\sin\!\left(\tfrac{\Omega t}{2}\right)\mathbf{n}\cdot\boldsymbol{\sigma}$$

Three knobs, three properties of one rotation. The **pulse area** $\Omega t$ sets the angle turned through, so a $\pi$ pulse is a bit flip and a $\pi/2$ pulse makes an even superposition. The **phase of the drive** chooses an axis in the equator. The **detuning** tilts that axis towards $z$, and the population can then reach only $(\Omega_x/\Omega)^2$.




The axis the drive turns the qubit about, for four detunings at one drive strength. On resonance it lies in the equator and the drive moves the population fully; far off it lies almost along $z$ and the drive barely tips the state at all.




The standard error of an estimated probability against the shot count, on decade axes. The slope is one half in both cases: that is the square root, drawn. A rare outcome is cheaper in absolute terms and dearer in relative ones.

WHERE A GATE ERROR COMES FROM

A pulse a few per cent too long turns through a few per cent too much: a **coherent** error, the same every time, which accumulates over a circuit rather than averaging away. A residual detuning does the same about a slightly wrong axis. Both are calibration failures rather than noise, and both are why chapter 5 counts depth.

## 2.7 Finite shots

Every probability in this chapter is exact and none of them is ever observed. What is observed is a count: run the circuit $N$ times, see the outcome $K$ times, report $K/N$. That number is a random variable and its spread is known.

THE COST OF A MEASUREMENT

$$K \sim \mathrm{Binomial}(N, p), \qquad \mathrm{SE}\left(\tfrac{K}{N}\right) = \sqrt{\frac{p(1-p)}{N}}$$

The worst case is $p = \frac{1}{2}$, where the standard error is $1/(2\sqrt{N})$. One more decimal place costs a hundred times as many shots: one per cent needs of order ten thousand shots, one part in a thousand a million. No hardware improvement changes this, because it is counting and not physics. For a $\pm 1$ observable the same statement reads $\mathrm{SE} = \sqrt{(1 - \langle A \rangle^2)/N}$.

SAMPLING ERROR IS NOT NOISE

Sampling error is present on a perfect device and shrinks as $1/\sqrt{N}$. State-preparation, gate and readout errors are **systematic**: they bias the answer, and more shots make the wrong number more precise rather than more right. A report that gives a shot count and no error bar, or an error bar with no statement of which of the two it covers, has not reported the measurement.

## 2.8 Summary

- A measurement returns one outcome with probability $p(a) = \langle\psi|P_a|\psi\rangle$, and leaves the state $P_a|\psi\rangle$ put back to length one. The outcomes are the eigenvalues, not the eigenvectors.
- Choosing a measurement basis is choosing an experiment, not a coordinate system. On real hardware it costs a gate.
- An expectation value is a mean over repetitions and is usually not a reading any instrument can return. A variance vanishes exactly on the eigenstates.
- Two observables can both be sharp exactly when they commute; otherwise $\Delta A\,\Delta B \geq \frac{1}{2}|\langle[A,B]\rangle|$, which is a statement about two distributions and not about a clumsy apparatus.
- The three Pauli operators are Hermitian and unitary at once, square to the identity, and multiply by one cyclic rule.
- A closed system evolves by $U(t) = e^{-iHt}$. An energy eigenstate is stationary; two of them beat at their difference; a gate is a Hamiltonian run for a chosen time.
- A histogram of counts is an estimate with an error bar of $\sqrt{p(1-p)/N}$, and never the distribution itself.

FOUR ERRORS THAT COST A WHOLE QUESTION

Squaring an amplitude instead of its modulus. Reporting an expectation value as though an instrument could return it. Reading the uncertainty relation as a statement about disturbing the system. And treating a histogram of counts as the distribution rather than as an estimate of it.

WHAT COMES NEXT

Everything here assumed the system is alone and its state is a single vector. Chapter 3 drops both: a system that is one half of a larger one has no vector of its own, and describing it needs the density operator. That is also where a measurement whose result was thrown away finally gets a proper description.

3

# Mixed states and entanglement

*Chapters 1 and 2 assumed two things without saying so: that the system is alone, and that its preparation is known exactly. Both fail as soon as a qubit sits beside anything else. This chapter replaces the state vector with an operator that survives both failures, and it needs no new postulate to do it.*

Two ordinary situations break the vector. In the first, a device emits $|0\rangle$ half the time and $|1\rangle$ the other half and nobody records which; that is not the superposition $(|0\rangle + |1\rangle)/\sqrt{2}$, because the superposition gives a certain answer in the $X$ basis and the device gives a coin there. In the second, two qubits are in a pure state of the pair while neither one has a vector of its own. What replaces the vector in both cases is one matrix, the **density operator**.

## 3.1 The density operator

For a pure state the definition is the outer product of chapter 1. For a preparation that produces $|\psi_i\rangle$ with classical probability $p_i$ it is the weighted sum of those. The weights are ordinary probabilities describing a classical coin, and they are added rather than superposed: a superposition of two states is a third state, and a mixture of two states is neither of them and is not a vector at all.

THE DENSITY OPERATOR

$$\rho_\psi = |\psi\rangle\langle\psi|$$

$$\rho = \sum_i p_i\, |\psi_i\rangle\langle\psi_i|, \qquad p_i \geq 0, \qquad \sum_i p_i = 1$$

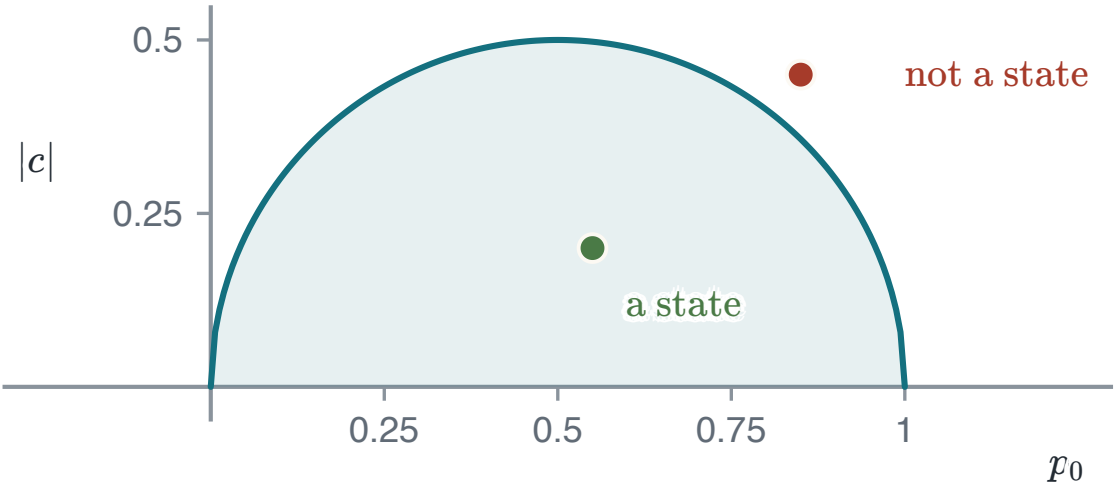
For one qubit the diagonal entries are the probabilities a computational-basis reading returns, and the off-diagonal entry is the **coherence** — the only place a relative phase survives, and therefore the only thing an $X$ or a $Y$ reading can see. Replacing $|\psi\rangle$ by $e^{i\gamma}|\psi\rangle$ leaves $\rho$ unchanged, because the two phases meet as $e^{i\gamma}e^{-i\gamma}$: the global-phase convention of chapter 1 stops being a convention here and becomes a property of the object.

Not every square matrix is a state. Three conditions are needed, and each protects a probability: Hermiticity makes every expectation value real, positivity makes every probability non-negative, and unit trace makes them add to one. Together the eigenvalues of $\rho$ are a probability distribution.

WHICH MATRICES ARE PHYSICAL

$$\rho = \rho^\dagger, \qquad \rho \succeq 0, \qquad \operatorname{Tr}\rho = 1$$

For one qubit these collapse to a single inequality on the coherence, $|\rho_{01}|^2 \leq \rho_{00}\rho_{11}$, which is positivity written as "the determinant is not negative". It is a real physical statement rather than bookkeeping: a nearly-certain qubit is nearly incoherent.




Every one-qubit matrix with a real coherence, drawn as a point. The shaded half disc is the physical region and the curve is its edge, where the matrix is pure. The upper point is Hermitian and has trace one, and is still not a state.

TWO TESTS ARE NOT THREE

Hermitian with unit trace is easy to arrange and means nothing on its own. Positivity is the condition that actually restricts, and it is the one skipped: a matrix estimated from noisy measured data very often fails it by a small amount, and every tomography routine has to project back onto the physical set. A reported state with a negative eigenvalue is a report that has not finished.

Every prediction comes from one trace, and the same formula covers pure and mixed states with no change of case. That it agrees with the average over branches is a calculation rather than a definition: the trace does not care about the order of a product as long as the order is only rotated, so $\mathrm{Tr}(|\psi\rangle\langle\psi|A) = \langle\psi|A|\psi\rangle$, and applying that term by term gives $\sum_i p_i \langle\psi_i|A|\psi_i\rangle$ — the quantum average inside each branch, then the classical average over branches.

EVERY PREDICTION, FROM ONE TRACE

$$\langle A\rangle = \mathrm{Tr}\left(\rho A\right), \qquad p(m) = \mathrm{Tr}\left(\rho E_m\right)$$

So $\rho$ carries exactly what is needed to predict every measurement on the system, and nothing else. It is not a shorthand for a longer story about which state was really made; it is the complete answer to every question that can be asked of the system alone.

EXAMPLE 3.1 · ONE MATRIX, TWO PREPARATIONS

GIVEN — Device A emits $|+\rangle$ or $|-\rangle$, each with probability $\frac{1}{2}$. Device B emits $|0\rangle$ or $|1\rangle$, each with probability $\frac{1}{2}$.

FIND — Whether any measurement separates them.

SOLUTION — Both give $\frac{1}{2}I$. Since every prediction is $\mathrm{Tr}(\rho E)$ and the two operators are equal, every probability of every outcome of every measurement agrees, on any number of copies.

CHECK — Try $X$ on device A. It emits $X$ eigenstates, but which one is a fair coin, so the $X$ reading is a coin as well — exactly as it is for device B.

WHERE THIS IS MISREAD

"The qubit really was in $|0\rangle$ or $|1\rangle$, we just do not know which" is a story, not a fact, and device A shows it is not forced. Nothing in the mathematics picks a preferred decomposition and no experiment does either. The matrix is the physics; the ensemble behind it is a description of a laboratory.

## 3.2 Purity, and the ball of qubit states

The single most useful number about a density operator is the trace of its square. It is one exactly for pure states, where $\rho^2 = \rho$ and the rank is one, and it is $1/d$ for the maximally mixed state $I/d$. It needs no diagonalisation: it is the sum of the squared moduli of all the entries.

PURITY

$$\gamma = \mathrm{Tr}\left(\rho^2\right) = \sum_k \lambda_k^2, \qquad \frac{1}{d} \le \gamma \le 1$$

Purity says **how** mixed and never **which** state. It is one number and a qubit state needs three: for a qubit the trace and the purity fix the two eigenvalues but say nothing about the eigenbasis, so a whole family shares one purity.

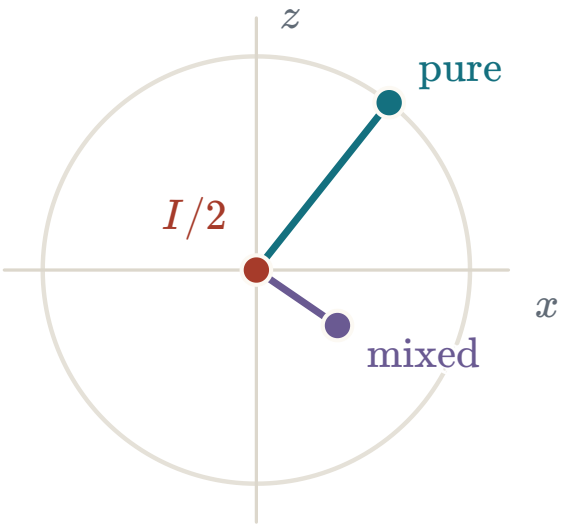
For a qubit there is a complete picture. The three Pauli means chapter 2 collected into a vector are not a summary of the state, they **are** the state: the four matrices $I, X, Y, Z$ span every two-by-two matrix, and the trace condition fixes the coefficient of $I$.

A QUBIT STATE AS A VECTOR

$$\rho = \tfrac{1}{2}\left(I + \mathbf{r}\cdot\boldsymbol{\sigma}\right), \qquad r_a = \mathrm{Tr}\left(\rho\,\sigma_a\right)$$

$$\lambda_\pm = \tfrac{1}{2}\left(1 \pm |\mathbf{r}|\right), \qquad \mathrm{Tr}\,\rho^2 = \tfrac{1}{2}\left(1 + |\mathbf{r}|^2\right)$$

Positivity is now the single statement $|\mathbf{r}| \le 1$, so the set of qubit states is a solid ball: pure states on the surface, mixed states inside, and $I/2$ at the exact centre. Chapter 4 draws the whole sphere and names the two angles; here only the length matters.




A flat cross-section through the ball of qubit states. A pure state reaches the rim, a mixed one falls short, and $I/2$ sits at the centre with no direction at all.

THE CENTRE IS NOT HALF WAY BETWEEN $|0\rangle$ AND $|1\rangle$

$(|0\rangle + |1\rangle)/\sqrt{2}$ is $|+\rangle$, a pure state on the rim. The centre of the ball is the **mixture** of $|0\rangle$ and $|1\rangle$, which is a different object entirely. Adding vectors and adding matrices are two different operations, and this is the picture in which the difference is impossible to miss.

## 3.3 Quantum channels

A closed system evolves by a unitary; a qubit on a chip is not closed. The evolution of an open system is a **quantum channel**, and every channel can be written as a sum of terms of one shape. The second condition below is what keeps the trace at one, and the check is one line: $\operatorname{Tr}\mathcal{E}(\rho) = \sum_k \operatorname{Tr}(K_k^\dagger K_k \rho) = \operatorname{Tr}\rho$. A unitary is the one-term case.

A CHANNEL IN KRAUS FORM

$$\mathcal{E}(\rho) = \sum_k K_k \,\rho\, K_k^\dagger, \qquad \sum_k K_k^\dagger K_k = I$$

The form is general rather than convenient: couple the system to an environment, run one unitary on the pair, then ignore the environment, and every channel arises that way. Nothing in the theory has been weakened — the whole is still a closed system running a unitary, and what makes the map on $\rho$ non-unitary is only that part of the result is never looked at. Two different sets of Kraus operators can describe the same channel, so the operators are not physical and the map is.

Two elementary qubit channels carry almost all of the practice. **Amplitude damping** describes a qubit in its upper level emitting and falling to the lower one, with $\gamma$ the probability that the emission happens. **Dephasing**, in its phase-flip form, applies $Z$ with probability $p$ and does nothing otherwise.

AMPLITUDE DAMPING

$$K_0 = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{bmatrix}, \qquad K_1 = \begin{bmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{bmatrix}$$

$$\rho_{11} \mapsto (1-\gamma)\rho_{11}, \qquad \rho_{01} \mapsto \sqrt{1-\gamma}\,\rho_{01}$$

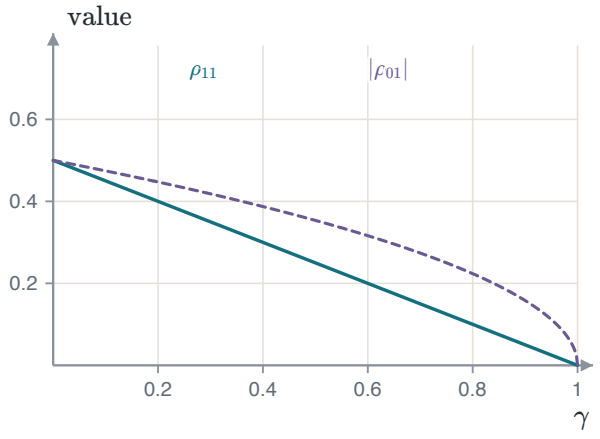
$K_1$ is not a matrix any unitary could be: it sends $|1\rangle$ to $|0\rangle$ and $|0\rangle$ to nothing, describing the branch in which a photon was emitted, and that branch is not reversible because the photon is gone. The two branches together do preserve the trace. At $\gamma = 1$ every state becomes $|0\rangle$, which is why this is the model of relaxation towards a cold equilibrium.

THE PHASE-FLIP CHANNEL

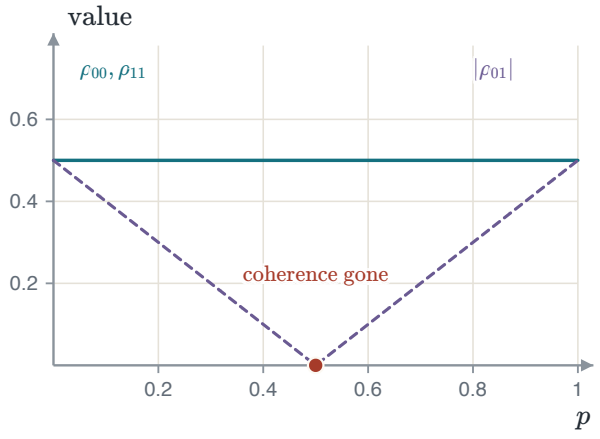
$$\mathcal{E}_Z(\rho) = (1-p)\,\rho + p\,Z\rho Z$$

$$\rho_{00}, \rho_{11} \text{ unchanged}, \qquad \rho_{01} \mapsto (1-2p)\,\rho_{01}$$

At $p = \frac{1}{2}$ the coherence is gone completely and what is left is the diagonal part of what was there. That is exactly the unrecorded measurement chapter 2 promised to describe: measuring $Z$ and throwing the result away is $\rho \mapsto |0\rangle\langle 0|\rho|0\rangle\langle 0| + |1\rangle\langle 1|\rho|1\rangle\langle 1|$, and no new postulate was needed. At $p = 1$ the channel is the gate $Z$ and nothing has been lost, so the destruction is greatest in the middle and not at the end.




The upper population and the coherence of $|+\rangle$ under amplitude damping. The population falls linearly in $\gamma$ and the coherence as its square root, so at every partial damping more coherence is left than population.




The same two quantities under the phase-flip channel. The populations are a flat line at one half for every $p$. Only the coherence moves, and it is what carries every interference effect in this course.

DECOHERENCE IS BASIS-DEPENDENT AND OFTEN MISSTATED

The phase-flip channel destroys coherence **in the $Z$ basis** and does nothing at all to a state already diagonal there: a qubit sitting in $|0\rangle$ is untouched by any amount of dephasing. "The environment destroys superpositions" is only true once the basis is named, and naming it is what the engineering of a qubit is about.

## 3.4 Relaxation and dephasing in time

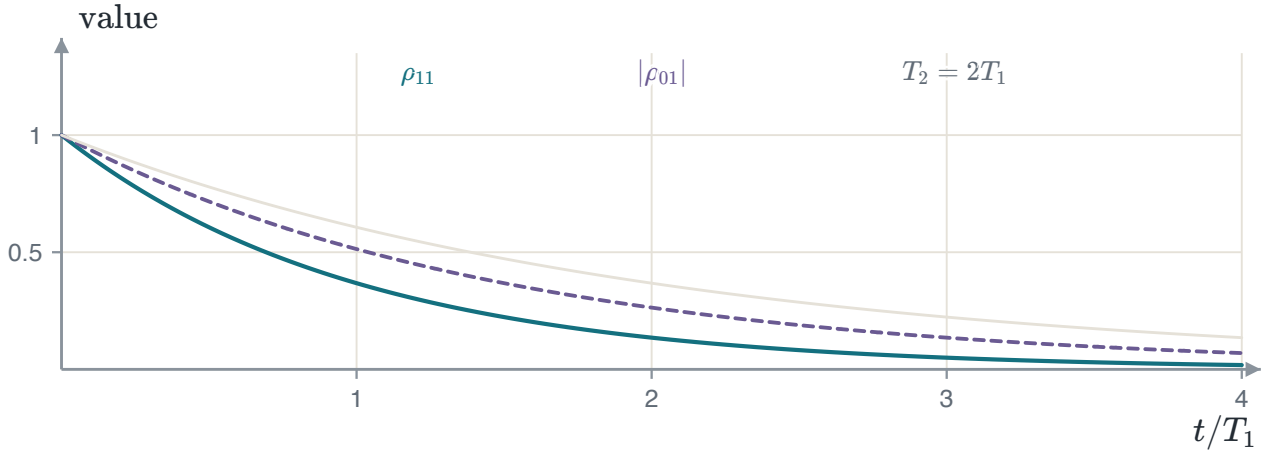
Applied continuously rather than once, the two channels give two exponentials with two time constants. $T_1$ is how long a population survives and $T_2$ how long a relative phase does, and they are not independent: losing the population also destroys the coherence, at half the rate, which is the same $\sqrt{1-\gamma}$ seen above.

THE TWO DECAYS, AND THE INEQUALITY BETWEEN THEM

$$\rho_{11}(t) = \rho_{11}^{\mathrm{eq}} + [\rho_{11}(0) - \rho_{11}^{\mathrm{eq}}]\, e^{-t/T_1}, \qquad \rho_{01}(t) = e^{-t/T_2}\rho_{01}(0)$$

$$\frac{1}{T_2} = \frac{1}{2T_1} + \frac{1}{T_\phi} \qquad \Longrightarrow \qquad T_2 \le 2T_1$$

Equality holds when there is no pure dephasing at all, so $T_2 = 2T_1$ is the best a qubit can do and every real device is below it. These are the time constants of a Markovian model in which the environment has no memory: reporting a $T_1$ and a $T_2$ is reporting the parameters of a fitted model, and a decay that is not exponential does not have them.




The population and the coherence against time, in units of $T_1$, for a device with $T_2 = 1.5\,T_1$. The faint curve is the ceiling $T_2 = 2T_1$: no coherence may decay more slowly than that, whatever else is done.

$T_2$ AND $T_2^*$ ARE DIFFERENT MEASUREMENTS

$T_2^*$ is the decay seen in a plain interference experiment, and it includes drift of the qubit frequency between one run and the next. A spin echo reverses that drift and returns a longer $T_2$. Quoting one where the other was measured overstates or understates the device by a large factor, and the two are routinely confused.

## 3.5 Two systems, and one of them alone

A general two-qubit pure state has four amplitudes, $c_0|00\rangle + c_1|01\rangle + c_2|10\rangle + c_3|11\rangle$, and the product of two single-qubit states is the special case in which they factor. Reading the product column below shows the ordering: the left factor changes the amplitude in pairs and the right factor changes it every entry, so the left factor is the more significant bit and entry $x$ of the column is the amplitude of the bit string $x$ written in binary.

THE ORDERING THIS COURSE FIXES

$$\begin{bmatrix} a \\ b \end{bmatrix} \otimes \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} ac \\ ad \\ bc \\ bd \end{bmatrix}, \qquad |q_1 q_0\rangle$$

This is the convention that fails silently. A library that counts qubits the other way builds $(ac,\ bc,\ ad,\ bd)$ from the same two states: nothing raises an error, the norm is still one, two of the four entries agree, and every number downstream describes a different state. Whenever a state crosses from one piece of software to another, print the four amplitudes and look at them. An operation on the first qubit alone is $A \otimes I$ and on the second alone $I \otimes B$, and those two commute because they act on different factors.

Given a joint state, the operation that answers "what is the state of $A$ by itself" is the **partial trace**. It does not depend on which basis of $B$ is used, and it is the only operation with the property that makes it the right answer: it reproduces every measurement performed on $A$ alone. That second line is the definition worth remembering — $\rho_A$ is not an approximation and not an average over anything, it is the exact state for that class of questions.

THE PARTIAL TRACE

$$\rho_A = \mathrm{Tr}_B\left(\rho_{AB}\right) = \sum_j \left(I_A \otimes \langle j|\right) \rho_{AB} \left(I_A \otimes |j\rangle\right)$$

$$\mathrm{Tr}\left(\rho_A A\right) = \mathrm{Tr}\left[\rho_{AB}\left(A \otimes I_B\right)\right] \quad \text{for every } A$$

For two qubits there is a block rule. Cut the four-by-four matrix into four two-by-two blocks: the traces of the blocks are $\rho_A$, and the sum of the two diagonal blocks is $\rho_B$. Two different operations, and getting them the wrong way round is the usual slip.

---

EXAMPLE 3.2 · A PURE PAIR WITH MIXED HALVES

GIVEN $|\Phi^+\rangle = \frac{1}{\sqrt{2}}\left(|00\rangle + |11\rangle\right)$.

FIND The reduced state of each qubit, and the purity of the pair and of one half.

SOLUTION $\rho_{AB}$ has $\frac{1}{2}$ at the four corners of the four-by-four and zero elsewhere. Its diagonal blocks are $\frac{1}{2}|0\rangle\langle 0|$ and $\frac{1}{2}|1\rangle\langle 1|$ and its off-diagonal blocks are traceless, so $\rho_A = \rho_B = I/2$. The pair has purity one and each half has purity $\frac{1}{2}$.

CHECK Every measurement on one qubit of a Bell pair is a fair coin, in every basis, which is what $I/2$ says. Nothing about the pair is unknown and everything about each qubit is — and that cannot happen classically, where a certain joint description makes every part certain too.

---

A WORKING DEFINITION OF ENTANGLEMENT

A pure joint state is **entangled** exactly when its reduced states are mixed. There are then two sources of a mixed state and they are not distinguishable from inside: either a classical coin decided the preparation, or the system is entangled with something else that is not being looked at. On real hardware the second is usually the honest description — the qubit is entangled with its environment, and $\rho$ is what is left after tracing the environment out.

A PARTIAL TRACE CANNOT BE UNDONE

Two very different joint states can share a reduced state: the Bell pair and the classical mixture $\frac{1}{2}|00\rangle\langle 00| + \frac{1}{2}|11\rangle\langle 11|$ both give $I/2$ on each side. So $\rho_A$ and $\rho_B$ together do not determine $\rho_{AB}$, and everything that distinguishes those two joint states lives in the correlations the partial trace threw away.

## 3.6 Separability, the Schmidt decomposition and entropy

A bipartite pure state is **separable** when it is a product and **entangled** when it is not. For two qubits the test is one determinant, and it vanishes exactly when the four amplitudes factor as $ac$, $ad$, $bc$, $bd$. So $\frac{1}{\sqrt{2}}(|01\rangle + |11\rangle)$ is $|+\rangle \otimes |1\rangle$, while $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ is no product at all, for any choice of the two single-qubit states.

THE AMPLITUDE TEST, AND THE GENERAL FORM

$$|\psi\rangle \text{ is a product} \iff c_0c_3 - c_1c_2 = 0$$

$$|\psi\rangle_{AB} = \sum_{k=1}^{r} \sqrt{\lambda_k}\,|u_k\rangle_A |v_k\rangle_B, \qquad \lambda_k > 0, \quad \sum_k \lambda_k = 1$$

The second line is the **Schmidt decomposition**: two orthonormal sets, chosen for the state, in which only the diagonal terms survive. The number of terms $r$ is the **Schmidt rank**, the reduced states are $\rho_A = \sum_k \lambda_k |u_k\rangle\langle u_k|$ and likewise for $B$, and the state is entangled exactly when $r > 1$. Both sides always have the same non-zero eigenvalues, so the smaller side caps the rank and entanglement of a pure state cannot be more on one side than on the other. For mixed states "separable" means a mixture of products rather than one product, and no test as short as the determinant exists.

The decomposition is computed with a singular value decomposition. Write the amplitudes as a matrix instead of a column, rows indexed by the first system and columns by the second; then $C = U\Sigma V^\dagger$ is the Schmidt decomposition, the singular values are $\sqrt{\lambda_k}$, and the two sets of singular vectors are the two Schmidt bases. Nothing has to be derived twice: $\rho_A = CC^\dagger$, and the singular values of $C$ are the square roots of the eigenvalues of $CC^\dagger$ by definition.

TWO THINGS TO STATE BEFORE RESHAPING

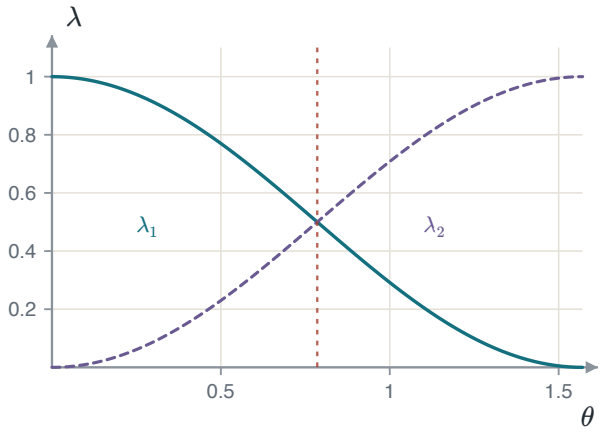
The qubit ordering, and which qubits go into the rows. Reshaping the wrong way round transposes $C$, which leaves the singular values alone and swaps the two Schmidt bases — so the entanglement number looks right while the states attached to it belong to the other system. And the rank is the count of singular values above a stated tolerance: a value of $10^{-16}$ is a zero, and calling it a third Schmidt term is reporting rounding.

The rank counts; the entropy weighs. The von Neumann entropy is the Shannon entropy of the eigenvalues of $\rho$, in bits, and for a **pure** bipartite state the entropy of either reduced state is the amount of entanglement the pair carries. A maximally entangled pair of qubits carries one bit, called one **ebit**, and it is the unit every protocol in chapter 5 is priced in.

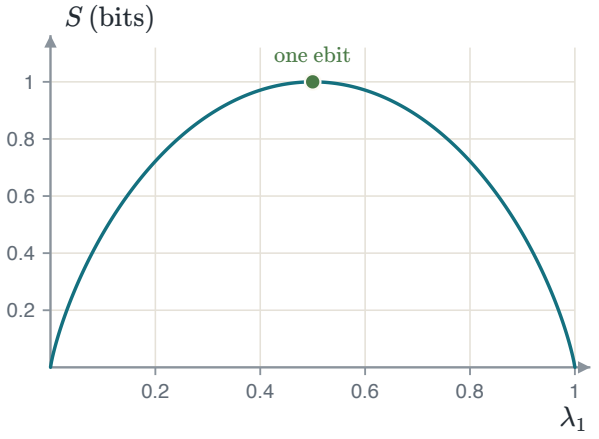
ENTANGLEMENT ENTROPY

$$S(\rho) = -\operatorname{Tr}(\rho \log_2 \rho) = -\sum_k \lambda_k \log_2 \lambda_k, \qquad S(\rho_A) = S(\rho_B)$$

This measure is only for pure pairs. For a mixed joint state $S(\rho_A)$ measures the total uncertainty about $A$, which mixes entanglement together with ordinary classical noise, and a separable mixed state can have a large $S(\rho_A)$ while carrying no entanglement at all. Quoting the reduced entropy of a noisy pair as its entanglement is a common and serious overstatement.




The two Schmidt coefficients of $\cos\theta\,|00\rangle + \sin\theta\,|11\rangle$. At the ends one of them is zero, the rank is one, and the state is a product; where they are equal the pair is maximally entangled.




The entanglement entropy of a two-term Schmidt spectrum, against the larger coefficient. Zero at both ends, where the state is a product, and one bit in the middle.

## 3.7 Bell correlations, CHSH and no signalling

Four maximally entangled two-qubit states are used constantly, and together they are an orthonormal basis of the four-dimensional space. Every one of them has $\rho_A = \rho_B = I/2$, so no measurement on one qubit tells them apart; what separates them is entirely in the joint correlations.

THE BELL STATES, AND THE CORRELATIONS OF ONE OF THEM

$$|\Phi^{\pm}\rangle = \frac{|00\rangle \pm |11\rangle}{\sqrt{2}}, \qquad |\Psi^{\pm}\rangle = \frac{|01\rangle \pm |10\rangle}{\sqrt{2}}$$

$$\langle X \otimes X \rangle = 1, \qquad \langle Y \otimes Y \rangle = -1, \qquad \langle Z \otimes Z \rangle = 1$$

Each correlation is $\pm 1$, so each is a certain statement about the pair: measure both qubits in the $X$ basis and the two answers always agree, even though each answer alone is a fair coin. The classical mixture $\frac{1}{2}|00\rangle\langle 00| + \frac{1}{2}|11\rangle\langle 11|$ reproduces the $Z$ correlation exactly and gives zero for the other two, so any model saying "the pair was made as $00$ or as $11$, we just do not know which" gets the first right and the other two wrong.

That is not yet a proof that no such model works, because a cleverer one might be built. The argument that closes it uses two measurements on each side, each returning $\pm 1$, combined into one number with the fourth term subtracted. Suppose every run carries definite values $a_0, a_1, b_0, b_1 \in \{\pm 1\}$, fixed before the settings are chosen, and group the terms as $a_0(b_0 + b_1) + a_1(b_0 - b_1)$: two numbers each $\pm 1$ either agree or differ, so one bracket is $\pm 2$ and the other is exactly zero.

THE CHSH COMBINATION AND ITS CLASSICAL BOUND

$$S = \langle A_0 B_0 \rangle + \langle A_0 B_1 \rangle + \langle A_1 B_0 \rangle - \langle A_1 B_1 \rangle$$

$$|S| \le 2 \qquad \text{for every local model with pre-existing values}$$

No quantum mechanics went into that derivation. It uses only that the four numbers exist together, which is what "the values were there before anyone looked" means. The minus sign is what makes the bound bite: with four plus signs the classical bound would be four and so would the quantum one.

---

EXAMPLE 3.3 · WHAT QUANTUM MECHANICS REACHES

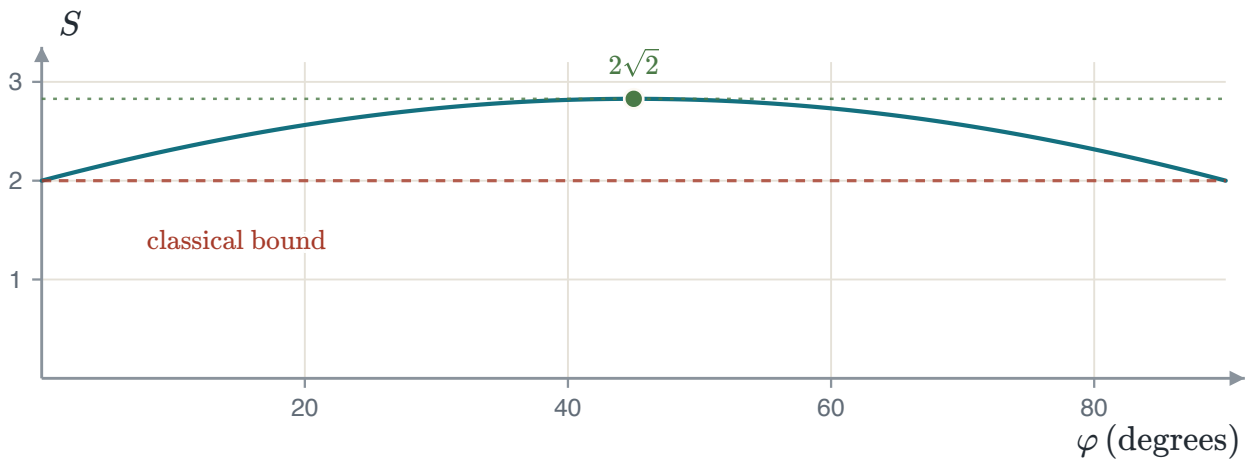
GIVEN — $|\Phi^{+}\rangle$, with $A_0 = Z$, $A_1 = X$, $B_0 = (Z + X)/\sqrt{2}$ and $B_1 = (Z - X)/\sqrt{2}$.

METHOD — For this state $\langle (\mathbf{n}\cdot\boldsymbol{\sigma}) \otimes (\mathbf{m}\cdot\boldsymbol{\sigma}) \rangle = n_x m_x - n_y m_y + n_z m_z$.

SOLUTION — Each of the first three correlations is $1/\sqrt{2}$ and the fourth is $-1/\sqrt{2}$, so $S = 4/\sqrt{2} = 2\sqrt{2} \approx 2.828$.

CHECK — Every term has modulus at most one, so $S$ could in principle be four. Quantum mechanics stops at $2\sqrt{2}$, and that ceiling is a theorem rather than an accident of these four angles.

---




The CHSH value for $|\Phi^+\rangle$ with $A_0 = Z$, $A_1 = X$, and the other party's two settings placed symmetrically at $\pm\varphi$ in the same plane. The curve is above the classical bound over a wide range of angles rather than at one exact setting, which is what makes the experiment possible with imperfect apparatus.

> WHAT IS REFUTED, EXACTLY
>
> Not "the qubits communicate". The assumption that fails is that the four outcomes exist together before the settings are chosen. And $S$ is estimated from four finite runs, so each term carries the sampling error of chapter 2: reporting $S = 2.6$ without a shot count and an interval is not reporting a violation.

It is tempting to read the correlation as a channel. It is not one, and the proof is the partial trace. Whatever the second party measures, the first party's state — averaged over outcomes they have not been told — is the operator it was before, because the effects are complete: $\mathrm{Tr}[\rho_{AB}(A \otimes \sum_m E_m)] = \mathrm{Tr}[\rho_{AB}(A \otimes I)] = \mathrm{Tr}(\rho_A A)$. The second party's choice disappears from the expression before any number is computed.

> WHERE THE CORRELATION ACTUALLY APPEARS
>
> Only when the two records are brought together and compared, which needs an ordinary classical channel and travels no faster than one. Entanglement is a resource that makes two records agree in ways no classical preparation could; it is not a wire. Every protocol in chapter 5 that uses a Bell pair also sends classical bits, and that is why. "Measuring one qubit instantly collapses the other" describes a bookkeeping step: two parties who have not spoken hold different descriptions of the same pair, and both are correct, because a description is a statement about what its holder can predict.

## 3.8 Summary

- A density operator is Hermitian, positive semidefinite and of unit trace, and every prediction is $\mathrm{Tr}(\rho A)$. It does not remember which preparation produced it, and two ensembles giving the same matrix are separated by no experiment.
- Purity $\mathrm{Tr}\,\rho^2$ runs from $1/d$ to one. For a qubit, $\rho = \frac{1}{2}(I + \mathbf{r}\cdot\boldsymbol{\sigma})$ with $|\mathbf{r}| \le 1$, and the states form a solid ball.
- A channel is $\sum_k K_k \rho K_k^\dagger$ with $\sum_k K_k^\dagger K_k = I$: a unitary on a larger system with the rest ignored. Damping moves populations, dephasing moves only coherences, and $T_2 \le 2T_1$.

- The partial trace is defined by reproducing every measurement on one system. A pure pair with mixed parts is entangled; the Schmidt rank counts and $S(\rho_A)$ weighs, in bits.
- The Bell states share one reduced state and differ in every joint correlation. $|S| \leq 2$ for any model with pre-existing values, and quantum mechanics reaches $2\sqrt{2}$.
- No choice of local measurement moves the other party's distribution, so entanglement carries no message on its own.

FOUR ERRORS THAT COST A WHOLE QUESTION

Treating a mixture as a superposition. Checking Hermiticity and trace and calling the result a state. Swapping the two partial traces, so $\rho_A$ is computed where $\rho_B$ was asked for. And reading a Bell violation as communication.

WHAT COMES NEXT

Chapter 4 draws the ball this chapter has been computing in, gives the two angles of a pure state their names, and turns every single-qubit gate into a rotation of it. It also builds the two-qubit gate that makes a Bell state out of a product one, so the pairs used here stop being assumed and start being constructed.

4

# The Bloch sphere and quantum gates

*Three chapters of algebra have said what a qubit state is and what may be done to it. This chapter says the same things again as a drawing, and the drawing loses nothing: every pure state of one qubit is a point of a sphere, every mixed state a point inside it, and every gate a rotation of that sphere.*

The counting is what makes it work, and it is worth doing before anything is drawn. Two complex amplitudes are four real numbers. Normalisation removes one and the fact that a global phase is not physical removes another, so two are left — and two angles are exactly what it takes to name a point on a sphere. The same counting works for the gates: a two-by-two unitary has four real parameters, one of them a phase nobody can see, and the three that remain are exactly the three that name a rotation, being an axis and an angle.

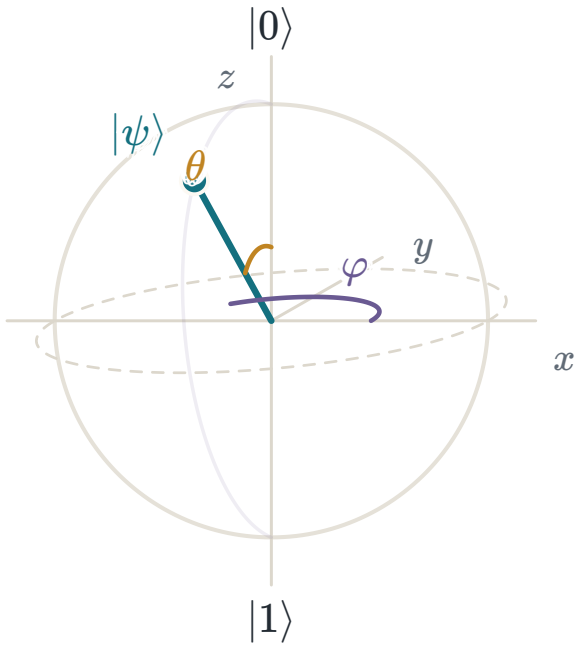
> THE PICTURE IS FOR ONE QUBIT AND STOPS THERE
>
> There is no drawing like this for two qubits: a pure state of a pair needs six real parameters and a mixed one fifteen. Everything in the second half of this chapter — the entangling gates — is therefore done in algebra, and the sphere is used only for what each qubit does alone.

## 4.1 The Bloch sphere, and the half angle

Start from any normalised qubit state and use the freedom of chapter 1: multiply by whatever global phase makes the first amplitude real and not negative. What is left is one real number in the first amplitude and one phase in the second, which is two angles.

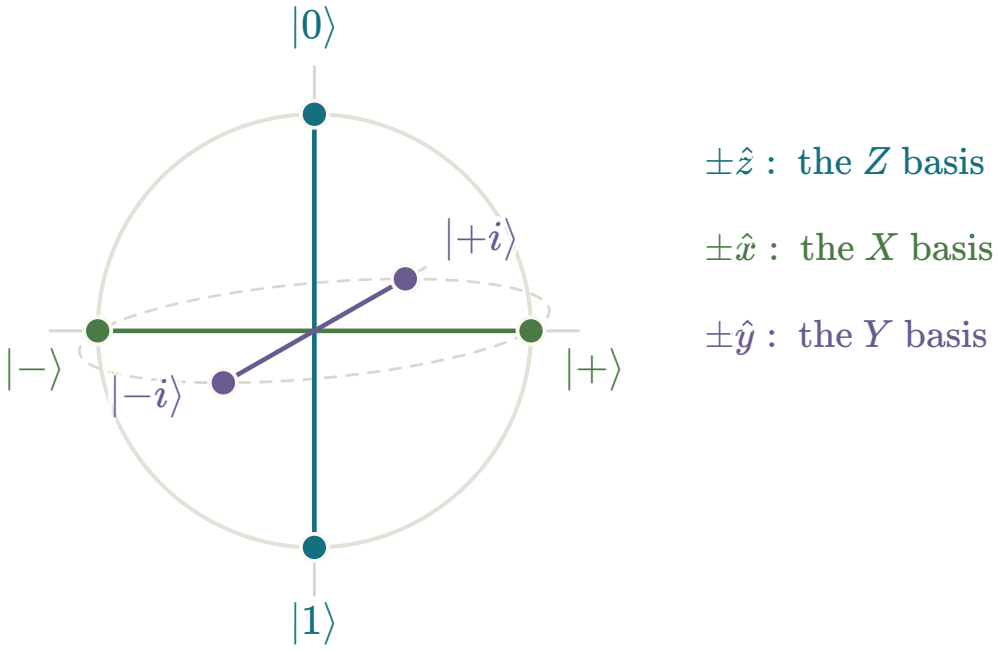
> A PURE QUBIT STATE, AND THE POINT IT NAMES
>
> $$|\psi(\theta,\varphi)\rangle = \cos\frac{\theta}{2}\,|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}\,|1\rangle, \qquad 0 \le \theta \le \pi, \quad 0 \le \varphi < 2\pi$$
>
> $$\mathbf{r} = (\sin\theta\cos\varphi,\ \sin\theta\sin\varphi,\ \cos\theta)\,, \qquad |\mathbf{r}| = 1$$
>
> The second line is the Bloch vector of chapter 3, now written in the two angles rather than in three traces. A mixed state is the same expression with $|\mathbf{r}| < 1$, so the sphere is the surface of the ball chapter 3 was already computing in. At a pole the azimuth means nothing, because $\sin 0 = 0$ leaves nothing for the phase to multiply: like latitude and longitude, the coordinates are degenerate there, and reporting a phase for a state at the north pole is reporting a number no experiment can contain.

The half angle is not a typographic accident and not a normalisation trick. The state must return to itself after $\theta$ has gone round once, and the vector must return after $\theta$ has gone round once as well — but a state and its negative are the same state, so the state may take two turns while the vector takes one. Halving the angle in the amplitudes is what arranges that, and section 4.2 shows the same factor again from the other side.




One state on the sphere. The plane of the page is the $x$–$z$ plane and is drawn undistorted, so the outline is a true circle and $\theta$ is drawn at its real size; the $y$ direction is foreshortened, which is what makes the equator an ellipse.

Six states carry most of a first course, and they are the two eigenstates of each Pauli operator. On the sphere they are the six points where the axes meet the surface: $|0\rangle$ and $|1\rangle$ at $\pm\hat{z}$, $|+\rangle$ and $|-\rangle$ at $\pm\hat{x}$, and $|+i\rangle$ and $|-i\rangle$ at $\pm\hat{y}$. Each pair is one measurement basis, and each pair sits at the two ends of one axis.




The six states. The frame is isotropic, so a pair that looks opposite on the page really is opposite in the data, and the axes really are at right angles inside the plane of the page.

One formula turns the picture into probabilities, and it follows in a line from chapter 3: $|\langle\chi|\psi\rangle|^2 = \mathrm{Tr}(\rho_\psi \rho_\chi)$, and multiplying out $\frac{1}{2}(I + \mathbf{r}\cdot\boldsymbol{\sigma})$ against $\frac{1}{2}(I + \mathbf{s}\cdot\boldsymbol{\sigma})$ leaves the dot product.

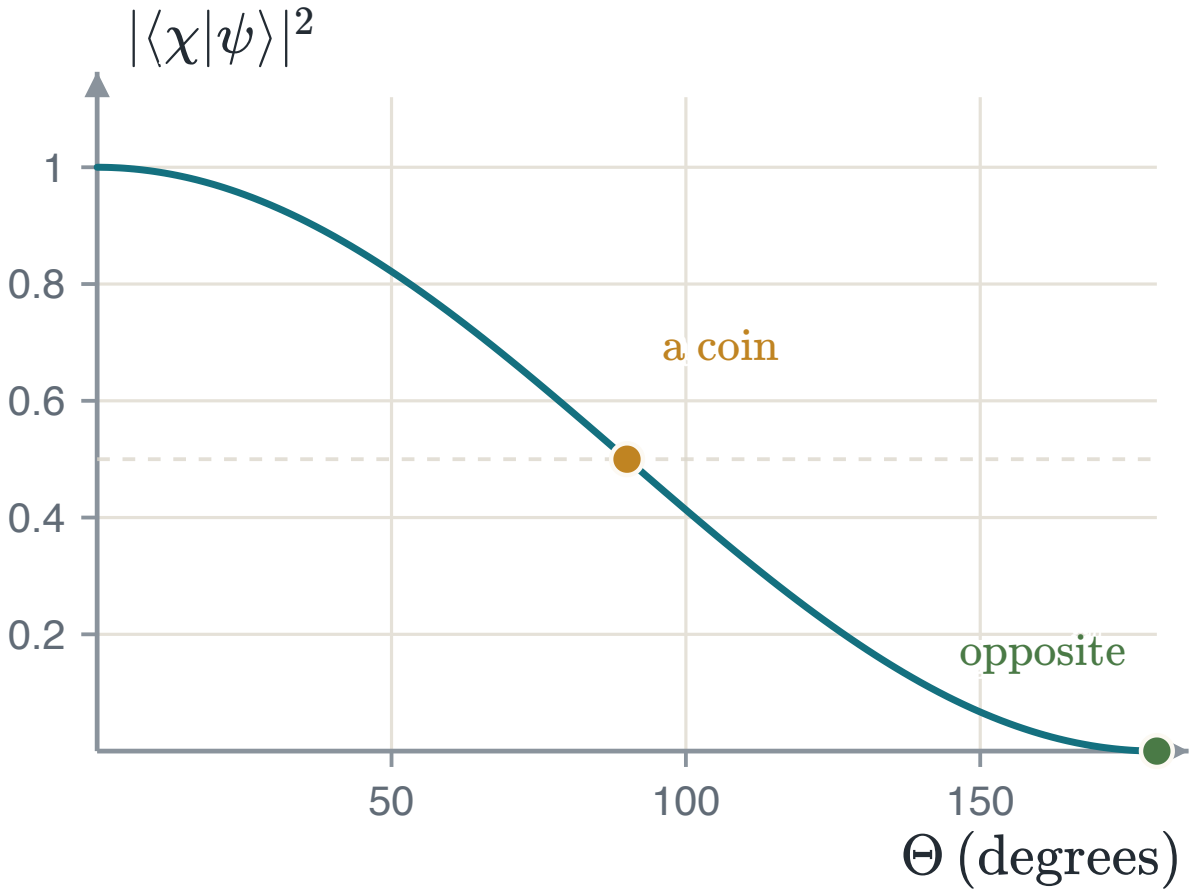
THE OVERLAP OF TWO PURE STATES, FROM THE ANGLE BETWEEN THEIR VECTORS

$$|\langle\chi|\psi\rangle|^2 = \frac{1 + \mathbf{r}\cdot\mathbf{s}}{2} = \cos^2\frac{\Theta}{2}$$

Read that once and the half angle stops being strange. Two states are orthogonal — perfectly distinguishable in one measurement — exactly when their points are **opposite**. Two points at a right angle on the sphere overlap with probability one half, which is a fair coin.

RIGHT ANGLES ON THE SPHERE ARE NOT ORTHOGONALITY

The word is the same and the meaning is not. The states at $\pm\hat{z}$ are orthogonal; the states at $+\hat{z}$ and $+\hat{x}$ are as hard to tell apart as a coin toss allows. Reading the geometry with the wrong meaning of the word is the fastest way to lose a whole question in this chapter.




The overlap against the angle on the sphere. It is one half at a right angle and reaches zero only at the far side, which is exactly what the half angle in the amplitudes is there to produce.

EXAMPLE 4.1 · FROM ANGLES TO NUMBERS

GIVEN $\theta = 60^\circ$ and $\varphi = 135^\circ$.

FIND The state, its Bloch vector, and $p(0)$.

SOLUTION $\cos 30^\circ = 0.8660$ and $\sin 30^\circ = 0.5$, so $|\psi\rangle = 0.8660|0\rangle + 0.5\, e^{i3\pi/4}|1\rangle$ and $\mathbf{r} = (-0.6124,\ 0.6124,\ 0.5)$. Then $p(0) = \cos^2 30^\circ = 0.75$.

CHECK $0.6124^2 + 0.6124^2 + 0.5^2 = 1$, so the point is on the surface as a pure state must be, and $\frac{1}{2}(1 + r_z) = 0.75$ agrees with the amplitude.

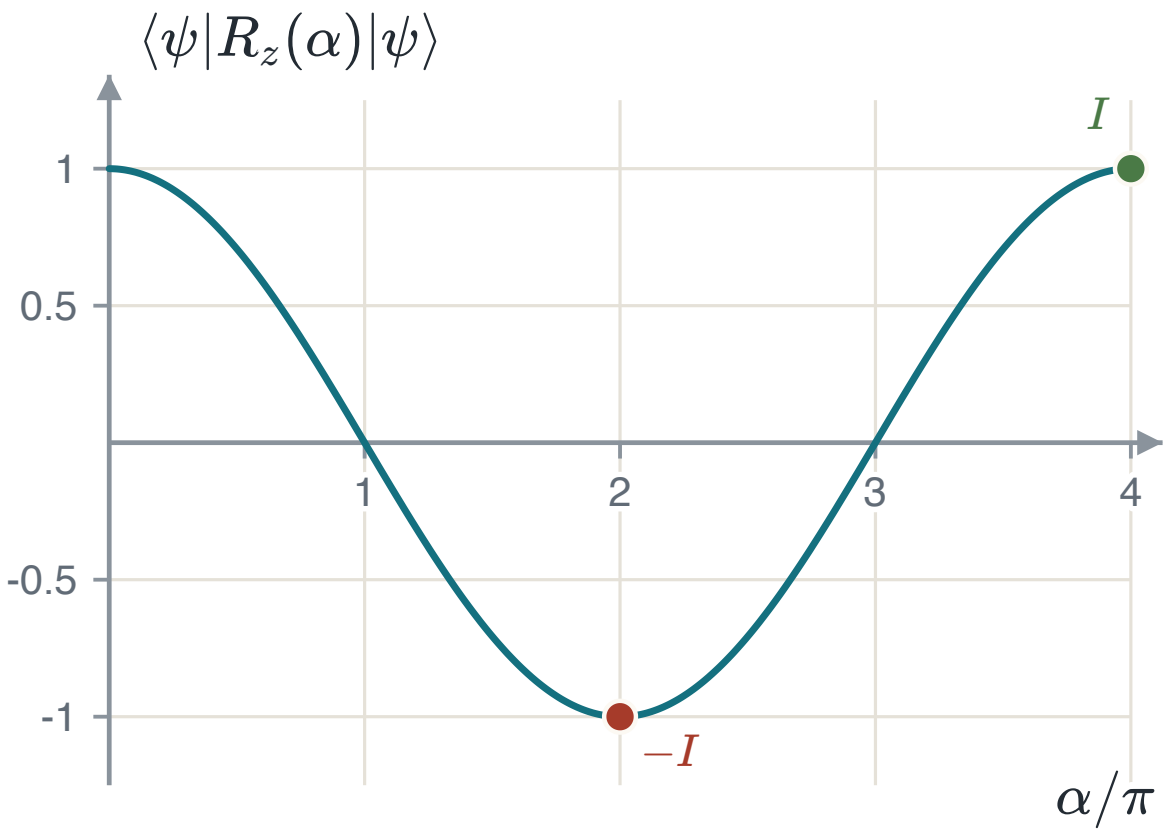
## 4.2 Global phase, relative phase, and the double cover

The rule stated in chapter 1 and proved in chapter 3 is now a fact about a drawing. Multiplying the whole state by a phase leaves $\rho$ alone, so it leaves the point alone. Putting a phase between the two amplitudes is a different operation entirely, and it moves the point round the equator by exactly that angle.

> THE TWO PHASES, DRAWN APART
>
> $$e^{i\gamma}|\psi\rangle \longmapsto \text{the same } \mathbf{r}, \qquad \cos\tfrac{\theta}{2}|0\rangle + e^{i\varphi}\sin\tfrac{\theta}{2}|1\rangle \longmapsto \varphi$$
>
> The two look alike on paper and are opposite in effect. A global phase is a change of nothing; a relative phase carries the state a definite distance, and at $\varphi = \pi$ it has carried $|+\rangle$ all the way to $|-\rangle$, which any $X$ measurement separates from it every single time.

The double cover is what happens when this is pushed. A rotation about an axis $\mathbf{n}$ by an angle $\alpha$ — the operator section 4.3 derives — halves the angle inside, so a full turn of the Bloch vector is a half turn in the amplitudes.

> A FULL TURN IS NOT THE IDENTITY
>
> $$R_{\mathbf{n}}(2\pi) = -I, \qquad R_{\mathbf{n}}(4\pi) = +I$$
>
> The Bloch vector is back where it started after $2\pi$ — it has to be, it is an ordinary vector being turned in ordinary space. The state is not: it has picked up a minus sign, and only after a second full turn does it return. So two different unitaries, $U$ and $-U$, produce the same rotation of the sphere, and the map from the gates to the rotations is two to one. That is the precise sense in which a half angle appears, and it is a fact about the geometry rather than a convention anyone chose.




What is left of the state after a turn about its own axis. At one full turn the amplitude is $-1$: the same state, with a sign. At two full turns the gate is the identity itself.

> "A ROTATION BY $2\pi$ DOES NOTHING" IS FALSE FOR A GATE
>
> It does nothing to that qubit alone. As a gate inside a larger circuit it is $-I$, and $-I$ applied on one branch of a superposition is a sign that survives into the interference. Put a $2\pi$ rotation under a control with the control in $|+\rangle$ and the control comes out as $|-\rangle$ — fully observable. The habit of dropping a global phase is right for a state and wrong for a gate.

## 4.3 Single-qubit gates as rotations

Chapter 2 built the operator $\mathbf{n} \cdot \boldsymbol{\sigma}$ and showed that its square is the identity. That one fact collapses the exponential: split the series into even and odd terms and each becomes an ordinary trigonometric series, exactly as it did for the generator in chapter 1.

THE ROTATION OPERATOR

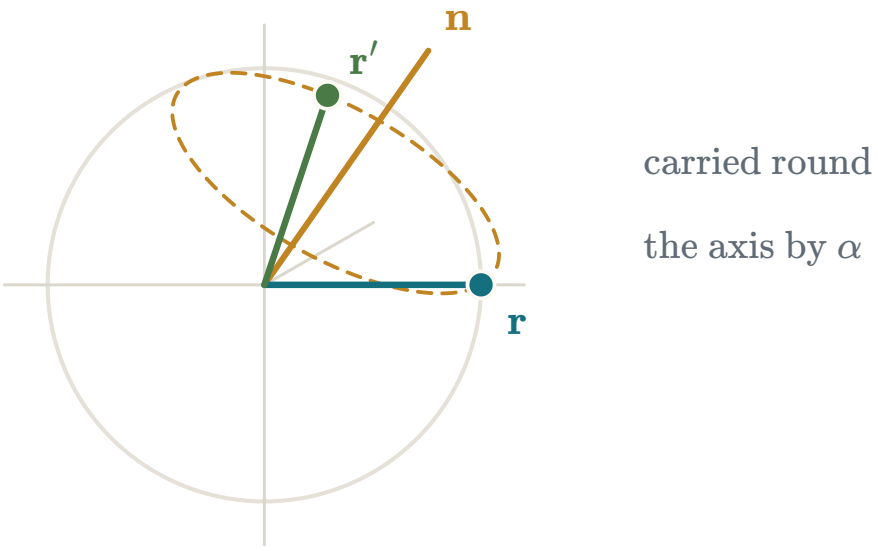
$$R_{\mathbf{n}}(\alpha) = e^{-i\alpha\,\mathbf{n}\cdot\boldsymbol{\sigma}/2} = \cos\frac{\alpha}{2}\,I - i\sin\frac{\alpha}{2}\,\mathbf{n}\cdot\boldsymbol{\sigma}$$

$$\begin{aligned} r_x &\mapsto r_x\cos\alpha - r_y\sin\alpha \\ r_y &\mapsto r_x\sin\alpha + r_y\cos\alpha \\ r_z &\mapsto r_z \end{aligned}$$

The second block is the case $\mathbf{n} = \hat{z}$ written out, and it is a plain rotation of the Bloch vector by $\alpha$. Any other axis is this one with the coordinates renamed, so the statement holds for every $\mathbf{n}$. Conversely every two-by-two unitary is $e^{i\gamma}R_{\mathbf{n}}(\alpha)$ for some phase, axis and angle, so **rotation** is not one family of gates among many: it is all of them.

THE ANGLE IN THE MATRIX IS HALF THE ANGLE ON THE SPHERE

$R_z(\pi)$ contains $\cos(\pi/2) = 0$ and $\sin(\pi/2) = 1$, and it turns the vector by a **half** turn, not a quarter. Reading the number inside the exponential as the angle on the sphere is the most common slip in this chapter, and it is always wrong by a factor of two.




A rotation about a tilted axis. The vector is carried round the amber circle keeping its angle to the axis, so its length never changes — which is unitarity, drawn. The axis sits at its true angle to $z$ because the frame is isotropic.

Putting $\alpha = \pi$ into the formula kills the cosine and leaves one Pauli operator with a phase in front of it. So each Pauli gate is a **half turn about its own axis**, up to a phase no measurement on that qubit can see, and half a turn fixes the two points on the axis while sending every other point to the far side.

THE PAULI GATES

$$R_x(\pi) = -iX, \qquad R_y(\pi) = -iY, \qquad R_z(\pi) = -iZ$$

$$X:\ |0\rangle \leftrightarrow |1\rangle, \qquad |+\rangle \text{ and } |-\rangle \text{ fixed}$$
$$Z:\ |+\rangle \leftrightarrow |-\rangle, \qquad |0\rangle \text{ and } |1\rangle \text{ fixed}$$

"$X$ is a bit flip" is therefore only half the story: it flips a computational-basis state and leaves $|+\rangle$ and $|-\rangle$ exactly where they are, because they lie on its axis. Likewise $Z$ is called a phase flip and does nothing at all to $|0\rangle$ or $|1\rangle$. Which gate looks like a flip depends entirely on the basis the state happens to be written in, and the sphere is where that stops being confusing.

The Hadamard is the same kind of gate with a different axis, and its own definition says which one: it is the average of two Pauli operators, so it is $\mathbf{n}\cdot\boldsymbol{\sigma}$ for the direction halfway between $\hat{x}$ and $\hat{z}$.

THE HADAMARD, AND WHAT A HALF TURN ABOUT A DIAGONAL DOES

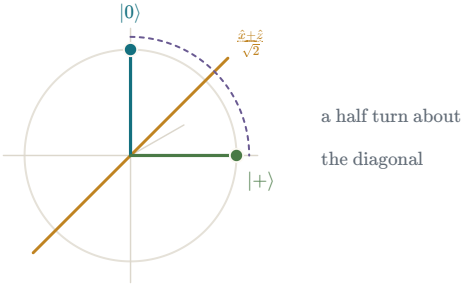
$$H = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \frac{X+Z}{\sqrt{2}}, \qquad H^2 = I$$

$$HXH = Z, \qquad HZH = X, \qquad HYH = -Y$$

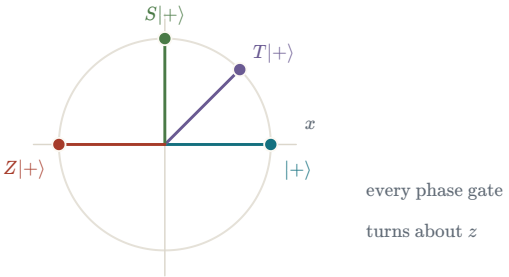
Half a turn about the diagonal exchanges the two axes it sits between and reverses the third. That is why $H$ is the standard change of measurement basis: measuring $X$ on a state is the same experiment as applying $H$ and then measuring $Z$, and a machine that can only read the computational basis needs exactly this gate to read any other.

THE HADAMARD DOES NOT CREATE SUPERPOSITIONS

It creates one from $|0\rangle$ and destroys one from $|+\rangle$: $H|+\rangle = |0\rangle$. Being its own inverse, it can only be a swap of the two bases, never a machine that makes superposition out of nothing. "Apply $H$ to get a superposition" is a sentence about one particular input.




The axis of the Hadamard sits at $45°$ between $x$ and $z$, and the half turn about it carries $|0\rangle$ to $|+\rangle$. The $45°$ on the page is $45°$ in the data.




The equator seen from above, with $|+\rangle$ carried round by $T$, by $S$ and by $Z$. Each is the same gate with a different angle.

The last family is the diagonal one, which puts a phase on the second amplitude and leaves the first alone. It is a turn about $\hat{z}$ carrying a phase, and two of its members have names because they are the two a machine usually provides.

THE PHASE GATES

$$P(\varphi) = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\varphi} \end{bmatrix} = e^{i\varphi/2}\, R_z(\varphi)$$

$$S = P\left(\tfrac{\pi}{2}\right), \qquad T = P\left(\tfrac{\pi}{4}\right), \qquad S^2 = Z, \quad T^2 = S$$

$S$ is a quarter turn of the equator and $T$ is an eighth of one, so eight applications of $T$ are a full turn and give the identity back. The two are not interchangeable in practice: $S$ is a Clifford gate and cheap to correct for, and $T$ is not, which is the subject of section 4.7.

A PHASE GATE DOES NOTHING TO A COMPUTATIONAL-BASIS STATE

Both poles sit on the axis of the turn, so $P(\varphi)$ cannot change any $Z$ probability at all, ever. Its whole effect appears only after a later gate has moved the state off that axis — which in practice means after a Hadamard. That two-step pattern, write a phase and then convert it, is the mechanism of every algorithm in chapter 6.

---

EXAMPLE 4.2 · A PHASE, AND THE GATE THAT REVEALS IT

GIVEN — A qubit in $|+\rangle$; apply $T$, then measure $Z$; then repeat with a Hadamard inserted before the measurement.

FIND — Both probabilities of the outcome $0$.

SOLUTION — $T$ turns $(1, 0, 0)$ to $(0.7071,\ 0.7071,\ 0)$, so $r_z = 0$ and $p(0) = 0.5$: the reading is blind to what happened. The Hadamard sends that vector to $(0,\ -0.7071,\ 0.7071)$, so $p(0) = \tfrac{1}{2}(1 + 0.7071) = 0.8536$.

CHECK — Directly, $|\langle 0|HT|+\rangle|^2 = \cos^2(\pi/8) = 0.8536$. Both lengths are one throughout, as two rotations require.

---

## 4.4 Composing gates: order, Euler angles and one gate with three dials

A circuit diagram is a picture of time, and the first gate is drawn on the left. A matrix acts on a ket standing to its right, so the first gate is the one closest to the ket, which is the one written last. This is the transcription error of the chapter, and it is invisible whenever the example happens to be symmetric.

THE ORDER A CIRCUIT MULTIPLIES IN

$$U_1 \text{ first},\ U_2 \text{ next},\ U_3 \text{ last} \qquad \Longrightarrow \qquad U = U_3\, U_2\, U_1$$

$$(A \otimes I)\,(I \otimes B) = A \otimes B = (I \otimes B)\,(A \otimes I)$$

The second line says gates on different qubits always commute, because they act on different tensor factors. That is why a diagram can show them side by side in any order without ambiguity, and it is what lets a scheduler run them at the same moment. Only gates that share a qubit have an order that has to be respected — and those usually do not commute: $SH|0\rangle = |+i\rangle$ while $HS|0\rangle = |+\rangle$, two states at right angles on the sphere.

Any rotation of a sphere can be reached by three turns about two fixed axes: spin about $z$, tilt about $y$, spin about $z$ again. The same is true of the gates, with one phase left over, and the parameter count shows that nothing is spare — four real degrees of freedom in a two-by-two unitary against three angles and a phase.

THE EULER DECOMPOSITION, AND THE SAME GATE AS ONE MATRIX

$$U = e^{i\alpha}\, R_z(\phi)\, R_y(\theta)\, R_z(\lambda)$$

$$U(\theta,\phi,\lambda) = \begin{bmatrix} \cos\frac{\theta}{2} & -e^{i\lambda}\sin\frac{\theta}{2} \\ e^{i\phi}\sin\frac{\theta}{2} & e^{i(\phi+\lambda)}\cos\frac{\theta}{2} \end{bmatrix}$$

The second form is what most software presents, with the global phase fixed by a convention rather than carried as a fourth parameter. Its first column is the state it prepares from $|0\rangle$, so $\theta$ and $\phi$ are the polar and azimuthal angles of that state and $\lambda$ is what the gate does to everything else. Every named gate is a setting of the three dials: $H = U(\frac{\pi}{2}, 0, \pi)$, $X = U(\pi, 0, \pi)$ and $P(\varphi) = U(0, \varphi, 0)$.

This is what a compiler is doing when it turns a requested gate into hardware instructions. A machine calibrates one or two native rotations well and reaches everything else through this identity, rather than calibrating every gate a user might ask for. It is the reason a small instruction set is not a limitation.

---

EXAMPLE 4.3 · THE EULER FORM OF THE HADAMARD

GIVEN $H = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$.

FIND Angles $\alpha, \phi, \theta, \lambda$ with $H = e^{i\alpha} R_z(\phi) R_y(\theta) R_z(\lambda)$.

SOLUTION $|U_{00}| = 1/\sqrt{2}$ gives $\cos(\theta/2) = 1/\sqrt{2}$, so $\theta = \pi/2$. Matching the phases of the first column and the sign in the first row gives $\lambda = \pi$, $\phi = 0$ and $\alpha = \pi/2$, so $H = e^{i\pi/2} R_y(\pi/2) R_z(\pi)$.

CHECK $R_z(\pi) = -iZ$ and $R_y(\pi/2) = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$, and the product is $i(-i)\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} Z = H$. Note that $R_z(\pi)$ is applied first, because it is written on the right.

---

THE PHASE IS NOT OPTIONAL WHEN THE GATE IS CONTROLLED

Dropping $e^{i\alpha}$ gives a gate with the same effect on any single qubit. Put it under a control and the dropped phase becomes a relative phase between two branches, and the circuit is wrong. A controlled version of a decomposed gate has to carry that phase explicitly, usually as an extra phase gate on the control. Three programs can implement the same rotation and print three different matrices, all correct, differing only in $e^{i\alpha}$ — harmless until one of them is controlled.

## 4.5 Reversible embeddings, ancillas and uncomputing

An AND gate takes two bits and returns one, so four inputs land on two outputs, the inputs cannot be recovered, and the map is not invertible. A unitary is invertible by definition, so no unitary implements AND

as written. The fix is not to change the function but to keep enough of the input alongside the answer.

THE REVERSIBLE EMBEDDINGS

$$(a,\, b) \longmapsto (a,\ a \oplus b) \qquad \text{is exactly CNOT}$$

$$|x\rangle|y\rangle \longmapsto |x\rangle\,|y \oplus f(x)\rangle \qquad \text{for any } f$$

The first map is its own inverse — apply it twice and $a \oplus b \oplus a = b$ comes back — so it is a permutation of the four inputs, which is what a unitary on the computational basis is allowed to be. The second is the general pattern: keep the input, and write the answer into a second register by addition rather than by overwriting. Chapter 6 calls that an oracle and builds every algorithm on it. AND itself needs a third wire, $(a, b, c) \mapsto (a, b, c \oplus ab)$, which is the Toffoli gate.

---

EXAMPLE 4.4 · A REVERSIBLE HALF ADDER

GIVEN — Two input bits $a, b$, with sum $s = a \oplus b$ and carry $c = ab$.

FIND — A reversible circuit that writes both outputs without erasing the inputs.

METHOD — Keep $a$ and $b$, and initialise separate sum and carry wires to zero.

SOLUTION — $(a, b, 0, 0) \mapsto (a, b, a \oplus b, ab)$. Two CNOTs write $a \oplus b$ on the sum wire and one Toffoli writes $ab$ on the carry wire.

CHECK — Run the three gates backwards. The Toffoli clears the carry and the CNOTs clear the sum, returning all four wires to $(a, b, 0, 0)$. The map is invertible because the inputs were preserved.

---

ERASING A BIT COSTS ENERGY, AND THAT IS NOT A METAPHOR

Landauer's argument ties the erasure of one bit to a definite minimum heat, $k_B T \ln 2$, released into the environment. Reversible computation is the way around it, and the reason it matters here is narrower and sharper: a quantum computer cannot erase inside a coherent block at all, because erasure is exactly the operation a unitary is not.

Any real computation needs somewhere to put intermediate results. Those extra qubits are **ancillas**, they start in a known state, and the problem is what they hold at the end. Run a computation on an input in superposition and the ancilla holds a different value in each branch, so the output register is entangled with the workings.

COMPUTE, COPY OUT, UNCOMPUTE

$$\sum_x c_x\,|x\rangle|0\rangle \ \longmapsto\ \sum_x c_x\,|x\rangle\,|g(x)\rangle \qquad \text{(dirty)}$$

$$V_f^\dagger\ (\text{copy})\ V_f\ :\ |x\rangle|0\rangle|y\rangle\ \longmapsto\ |x\rangle|0\rangle|y \oplus f(x)\rangle$$

The repair is three steps: compute, copy the answer out with a CNOT, then run the computation backwards. The ancilla leaves as it arrived and every branch agrees about it, so it is no longer entangled with anything, and the branches can interfere again. The cost is that the work is done twice.

A DIRTY ANCILLA DESTROYS AN ALGORITHM SILENTLY

It raises no error, the state is still normalised, and the circuit still runs. What is lost is the cancellation every algorithm in chapter 6 depends on, and the symptom is a flat output distribution that looks like noise. The smallest example is a two-qubit register left as a Bell pair with its ancilla: the register alone is $I/2$, so a Hadamard followed by a measurement returns a coin where a clean ancilla would have given $0$ with certainty. Counting the ancillas and checking that each returns to $|0\rangle$ is part of reading a circuit, not an optimisation afterwards.

## 4.6 Two qubits, and the ordering a gate is silently wrong about

Chapter 3 fixed the convention for states, and it is repeated here because this is where it starts breaking results rather than merely labels. A pair is written $|q_1 q_0\rangle$ with the higher-numbered qubit on the left, and entry $x$ of the column is the amplitude of the bit string $x$, so $x = 2q_1 + q_0$. A gate on one qubit alone is then written with the identity in the other slot, and **which** slot is the whole content of the statement.

A ONE-QUBIT GATE ON A NAMED QUBIT OF A PAIR

$$\text{on } q_0 : \; I \otimes A, \qquad \text{on } q_1 : \; A \otimes I$$

$$(I \otimes X)\,|10\rangle = |11\rangle, \qquad (X \otimes I)\,|10\rangle = |00\rangle$$

Both are four-by-four unitaries, both are perfectly valid gates, and they do different things. Nothing in the mathematics protects a reader who picks the wrong one: the result is normalised, its probabilities add to one, and it describes a different experiment. Whenever a state or a gate moves between two pieces of software, prepare $|10\rangle$, apply $X$ to the qubit you mean, and print all four amplitudes. It takes a minute and it is the only reliable test — reading the wire order off a circuit picture is not one, because drawing conventions and index conventions are independent.

The controlled-NOT flips one qubit exactly when the other is one. Its action on the computational basis is a permutation, and which qubit is the control changes the matrix, so a CNOT written down without naming its control and its target is not yet a gate. This course draws $q_0$ at the top of a circuit and uses it as the control unless a scene says otherwise.

THE TWO CNOT MATRICES, IN THE BASIS ORDER $|00\rangle, |01\rangle, |10\rangle, |11\rangle$

$$\text{CNOT} : \; |c\rangle|t\rangle \longmapsto |c\rangle|c \oplus t\rangle$$

$$\text{CNOT}_{0\to1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \qquad \text{CNOT}_{1\to0} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

The subscript reads control to target. The first exchanges the second and fourth basis states and the second exchanges the third and fourth, and they are different gates: handed $\frac{1}{\sqrt{2}}(|00\rangle + |10\rangle)$, one of them does nothing at all and the other produces a maximally entangled pair.

THE CONTROL IS NOT A SPECTATOR

Prepare the target in $|-\rangle$ and let the control be $\alpha|0\rangle + \beta|1\rangle$. The gate does nothing in the $|0\rangle$ branch, and in the $|1\rangle$ branch it applies $X$, which gives $X|-\rangle = -|-\rangle$. The target comes out exactly as it went in and the control comes out as $\alpha|0\rangle - \beta|1\rangle$: it has been hit with $Z$. This is **phase kickback**, three chapters early, and it is the whole mechanism of chapter 6. Consistently, $\mathrm{CNOT}_{0\to1} = (H \otimes H)\,\mathrm{CNOT}_{1\to0}\,(H \otimes H)$: in the $X$ basis the gate runs the other way round.

Two more two-qubit gates complete what this course needs. The controlled-$Z$ multiplies one basis state by $-1$ and leaves the other three alone; being diagonal it moves no probability at all, and being symmetric under exchanging its qubits it has no unique target. SWAP exchanges the two qubits, sends products to products, and creates no entanglement whatever it is given.

CZ AND SWAP, AND WHAT EACH COSTS

$$\mathrm{CZ} = \mathrm{diag}(1,1,1,-1) = (H \otimes I)\;\mathrm{CNOT}_{0\to1}\;(H \otimes I)$$

$$\mathrm{SWAP} = \mathrm{CNOT}_{0\to1}\;\mathrm{CNOT}_{1\to0}\;\mathrm{CNOT}_{0\to1}$$

A machine that provides one of CZ and CNOT provides both, at the cost of two one-qubit gates; which is native is a hardware question. SWAP is the expensive one: three two-qubit gates to move a state one step across a chip. Hardware does not let every pair of qubits interact, so a compiler moves qubits along the coupling map with SWAPs until the pair it needs is adjacent, and on some circuits those routing gates outnumber the gates the programmer wrote.

A DIAGONAL GATE IS NOT A HARMLESS GATE

CZ changes no outcome probability of any computational-basis measurement made immediately afterwards, which makes it look inert. It is not: handed $|+\rangle \otimes |+\rangle$ it produces a maximally entangled state. The phase it writes is relative between branches, and one later Hadamard turns it into a population difference. Likewise, relabelling two qubits is free and swapping them is not — a SWAP is required only when the two must physically move relative to a fixed coupling map.

## 4.7 Entangling gates and universality

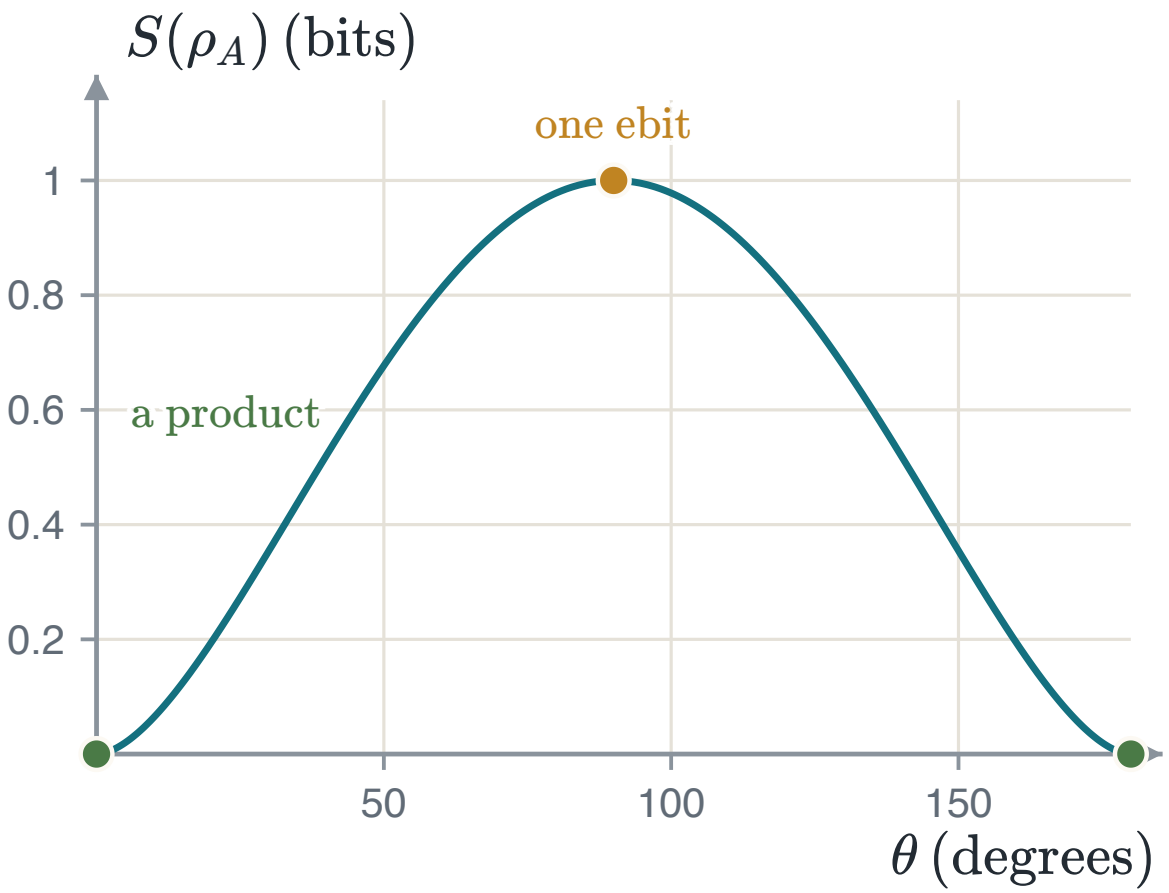
A gate of the form $U_A \otimes U_B$ acts on each half separately. In the Schmidt decomposition of chapter 3 it rotates the two bases and leaves the coefficients exactly where they were, so the entropy is unchanged and **no amount of one-qubit work creates entanglement**. Something acting on both qubits at once is required, and that is what the word entangling gate means.

ONE CNOT, AND HOW MUCH IT MAKES

$$(U_A \otimes U_B)\sum_k \sqrt{\lambda_k}\,|u_k\rangle|v_k\rangle = \sum_k \sqrt{\lambda_k}\,\left(U_A|u_k\rangle\right)\left(U_B|v_k\rangle\right)$$

$$|00\rangle \xrightarrow{I\otimes R_y(\theta)} \cos\tfrac{\theta}{2}|00\rangle + \sin\tfrac{\theta}{2}|01\rangle \xrightarrow{\mathrm{CNOT}_{0\to1}} \cos\tfrac{\theta}{2}|00\rangle + \sin\tfrac{\theta}{2}|11\rangle$$

At $\theta = 90^\circ$ the output is $|\Phi^+\rangle$ and the two halves are maximally entangled; at $\theta = 0$ nothing happened at all and the same gate produced a product state. So a gate is entangling if it entangles **some** input, not every input.




How much entanglement one CNOT produces, against the tilt it is handed. Zero at both ends, where the input is a computational-basis state and the gate is a permutation, and one full bit in the middle.

A CIRCUIT DIAGRAM IS NOT A PROOF OF ENTANGLEMENT

The output of a Bell circuit on real hardware is entangled only if the gates were good enough. Establishing it requires measurements in more than one basis, or a witness, with the uncertainty reported. Naming the circuit is not measuring the state, and a noisy Bell circuit produces a separable mixture long before it stops producing something.

Two universality statements, and they promise different things. The exact one needs a continuum of one-qubit gates; the approximate one needs four fixed gates and a budget. Both need one entangling gate, and no set of one-qubit gates alone is universal for any number of qubits.

TWO UNIVERSAL SETS

$$\{\text{every one-qubit gate}\} \cup \{\text{CNOT}\} \implies \text{every unitary, exactly}$$

$$\{H,\ S,\ T,\ \text{CNOT}\} \implies \text{every unitary, to any } \varepsilon > 0$$

Only approximation is possible in the second case, because a finite set of gates generates a countable set of circuits and the unitaries are not countable. The price of $\varepsilon$ is length: the Solovay–Kitaev result says a one-qubit gate can be reached with a number of gates growing like a power of $\log(1/\varepsilon)$, which is cheap — with the square, dropping $\varepsilon$ from $10^{-2}$ to $10^{-4}$ multiplies the length by four.

WHY $T$ IS SINGLED OUT, AND WHAT "UNIVERSAL" DOES NOT PROMISE

$H$, $S$ and CNOT generate the Clifford group, which maps Pauli operators to Pauli operators. That structure makes Clifford circuits classically simulable in polynomial time, so a Clifford-only machine offers no advantage at all. $T$ is the cheapest gate that leaves the group, and counting $T$ gates is how the cost of a fault-tolerant circuit is usually reported. Universality itself is a reachability claim and not a performance claim: counting shows there are far more $n$-qubit unitaries than short circuits, so a generic one needs a circuit exponential in $n$. It makes a machine programmable; it is never an argument that a particular task is fast.

---

EXAMPLE 4.5 · ONE GATE, TWO DIRECTIONS

GIVEN — $\frac{1}{\sqrt{2}}(|00\rangle + |10\rangle)$, that is $q_1$ in $|+\rangle$ and $q_0$ in $|0\rangle$, and a CNOT whose direction has not been named.

FIND — The output for each direction, and its entanglement.

SOLUTION — With $q_0$ as control the control is $0$ in both terms, so nothing happens and the state stays a product: $S = 0$. With $q_1$ as control, $|10\rangle \to |11\rangle$ and the output is $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) = |\Phi^+\rangle$, with Schmidt weights $\frac{1}{2}, \frac{1}{2}$ and $S = 1$ bit.

CHECK — The reduced states agree: after the first, each qubit is pure with purity one; after the second, both are $I/2$ with purity one half. "Apply a CNOT" was never an instruction — one direction produces a maximally entangled pair and the other produces nothing.

---

## 4.8 Summary

- A pure qubit state is $\cos\frac{\theta}{2}|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}|1\rangle$ at the point $\mathbf{r} = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)$. Opposite points are orthogonal states, and $|\langle\chi|\psi\rangle|^2 = \cos^2(\Theta/2)$.
- A global phase moves nothing and a relative phase moves the azimuth by itself. A $2\pi$ rotation is $-I$, which is invisible on one qubit and observable under a control.
- $R_{\mathbf{n}}(\alpha) = \cos\frac{\alpha}{2}I - i\sin\frac{\alpha}{2}\,\mathbf{n}\cdot\boldsymbol{\sigma}$ turns the Bloch vector about $\mathbf{n}$ by $\alpha$, and every one-qubit gate is one of these up to a phase. Each Pauli is a half turn, $H$ a half turn about $(\hat{x}+\hat{z})/\sqrt{2}$, and $P(\varphi)$ a turn about $z$.
- A circuit reads left to right and its matrices multiply right to left. Three turns and a phase reach every one-qubit gate, and $U(\theta,\phi,\lambda)$ is the same statement as one matrix.
- An irreversible function is embedded as $|x\rangle|y\rangle \mapsto |x\rangle|y \oplus f(x)\rangle$, and the ancillas it needs have to be uncomputed or the interference is lost.
- Two qubits are $|q_1 q_0\rangle$ with $x = 2q_1 + q_0$; a gate on one is $A \otimes I$ or $I \otimes A$; a CNOT is not a gate until its control is named. CZ writes a phase and is symmetric, and SWAP costs three CNOTs.
- No local gate changes the Schmidt coefficients, so one entangling gate is needed and one is enough. Every one-qubit gate plus CNOT is exactly universal, and $H, S, T$ and CNOT reach any accuracy, paid for in depth.

FOUR ERRORS THAT COST A WHOLE QUESTION

Reading the angle inside a rotation as the angle on the sphere, which is always wrong by a factor of two. Multiplying a circuit left to right. Dropping a global phase from a gate that is about to be controlled. And applying a one-qubit gate to the other qubit of a pair, which no later check will catch on its own.

WHAT COMES NEXT

Chapter 5 runs these gates: how a circuit is actually executed, what a shot count buys against an exact statevector, and what a compiler does to a circuit before the machine sees it. Then two protocols are worked end to end on it, teleportation and Grover, using nothing beyond the gates written down here.

5

# Circuits and protocols

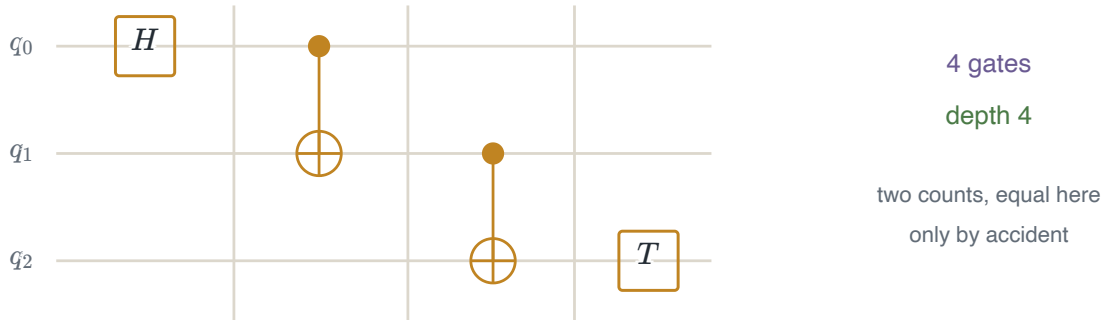
*Chapter 4 finished the gates, and nothing new about quantum mechanics is needed after it. What is still missing is everything about running the gates: what a circuit is as an object, what a machine gives back when it runs one, what a compiler does to it first, and how to say honestly what any of it cost.*

Three objects get confused with each other, and keeping them apart is most of this chapter. A **circuit** is a program: wires, gates, and the order they run in. A **run** is a physical experiment, repeated. The **result** is a table of bit strings and how often each came up. The state in the middle is never handed to anyone: a simulator can print it and a machine cannot, and every claim made here has to survive that.

> THE TWO HALVES
>
> The first half is the machine — circuits, shots, measurement inside a circuit, and what a compiler does. The second half runs two protocols on it end to end: teleportation, which finishes nothing until a classical channel has delivered two bits, and Grover, which is where a resource claim finally has to be written out in full.

## 5.1 The circuit model

A circuit diagram has one horizontal line for each qubit, and time runs left to right along those lines. A box on a line is a gate acting on that qubit; a dot on one line joined to a symbol on another is a gate acting on both. That is the whole notation, and what it describes is a program: a list of instructions and the order they must run in. It is not a stored state and it is not a drawing of anything on a chip.




Three qubits and four gates. The vertical rules are the layer boundaries. Here every gate waits for the one before it, so the depth happens to equal the gate count; that is not the usual case and it is worth noticing when it fails.

The model of computation the diagram belongs to is deliberately narrow, and the narrowness is what makes a cost statement possible at all.

THE CIRCUIT MODEL, IN ONE LINE

$$|0\rangle^{\otimes n} \xrightarrow{U_d \cdots U_2 U_1} |\psi\rangle \xrightarrow{\text{measure}} x \in \{0,1\}^n$$

Every qubit starts in $|0\rangle$, every gate comes from one fixed finite set, the measurement is in the computational basis, and the whole thing is repeated because one run gives one string. Each restriction is doing work. Fixing the input means an algorithm cannot smuggle in an answer through its preparation; fixing the gate set means the cost of a circuit is the number of gates; fixing the measurement basis means a change of basis is a gate and has to be paid for.

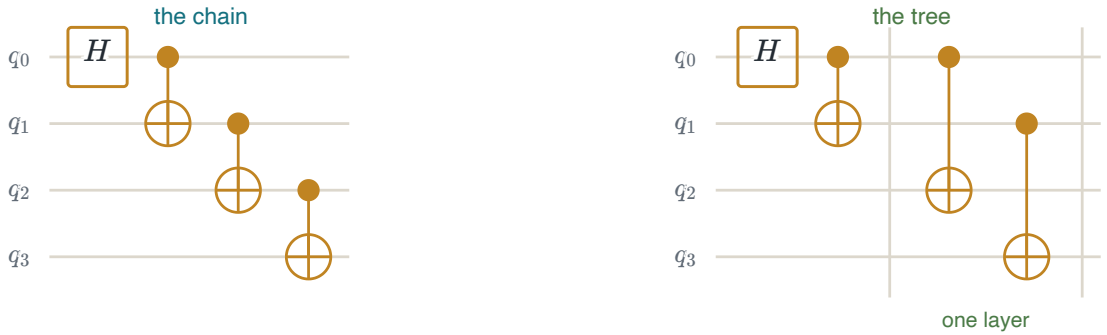
The ordering convention of chapter 4 starts costing money here, because two conventions meet on a diagram and they point in opposite directions. Qubit $q_0$ is drawn on the **top** wire, and $q_0$ is the **last** digit of the ket and the **least significant** bit of the index. So the wire order down the page and the digit order across the ket are reverses of each other, and the printed classical string follows the ket: the top wire is its rightmost character.

ONE STATE, NAMED FOUR WAYS

$$|q_{n-1} \ \cdots \ q_1 q_0\rangle, \qquad x = \sum_k 2^k q_k$$

A run that prints $101$ on three qubits read $q_2 = 1$, $q_1 = 0$ and $q_0 = 1$, which is the ket $|101\rangle$ and entry $5$ of the state vector. Reading the string onto the wires from the top instead names $|101\rangle$ again — a palindrome cannot expose the error — and names $|001\rangle$ where the run printed $|100\rangle$. Both are legal normalised states, so nothing downstream complains.

Two numbers describe how big a circuit is and they answer different questions. The **gate count** is how much work there is; the **depth** is how many layers must run one after another, and therefore how long the circuit takes. Chapter 3 gave the reason the second is the one that hurts: a qubit keeps its phase for a time $T_2$, and a circuit of depth $d$ made of gates lasting $\tau$ occupies about $d\tau$, so it is useful only while $d\tau \ll T_2$.




The same four-qubit GHZ state, two ways. Four gates in both. The chain waits for each CNOT before starting the next; the tree runs two of them together in one layer.

EXAMPLE 5.1 · THE SAME WORK, A THIRD OF THE TIME

GIVEN GHZ on $n = 16$ qubits, gates of $200\,\mathrm{ns}$, $T_2 = 80\,\mu\mathrm{s}$.

FIND The gate count, the depth and the duration of each construction.

SOLUTION Both use $1 + 15 = 16$ gates. The chain has depth $16$ and takes $3.2\,\mu\mathrm{s}$; the tree has depth $1 + \log_2 16 = 5$ and takes $1.0\,\mu\mathrm{s}$.

CHECK Put $n = 1000$ in. The chain needs $200\,\mu\mathrm{s}$ against a $T_2$ of $80$ and is impossible; the tree needs $2.2\,\mu\mathrm{s}$ and is not. The gate count never saw the difference.

"FEWER GATES" IS NOT "FASTER" AND NOT "MORE ACCURATE"

A rewrite that removes gates but serialises what was parallel makes the circuit slower. A rewrite that removes one-qubit gates and adds one two-qubit gate usually makes it less accurate, because a two-qubit gate is the expensive one by an order of magnitude. Compare two-qubit count, depth and duration, and say which of the three the claim is about.

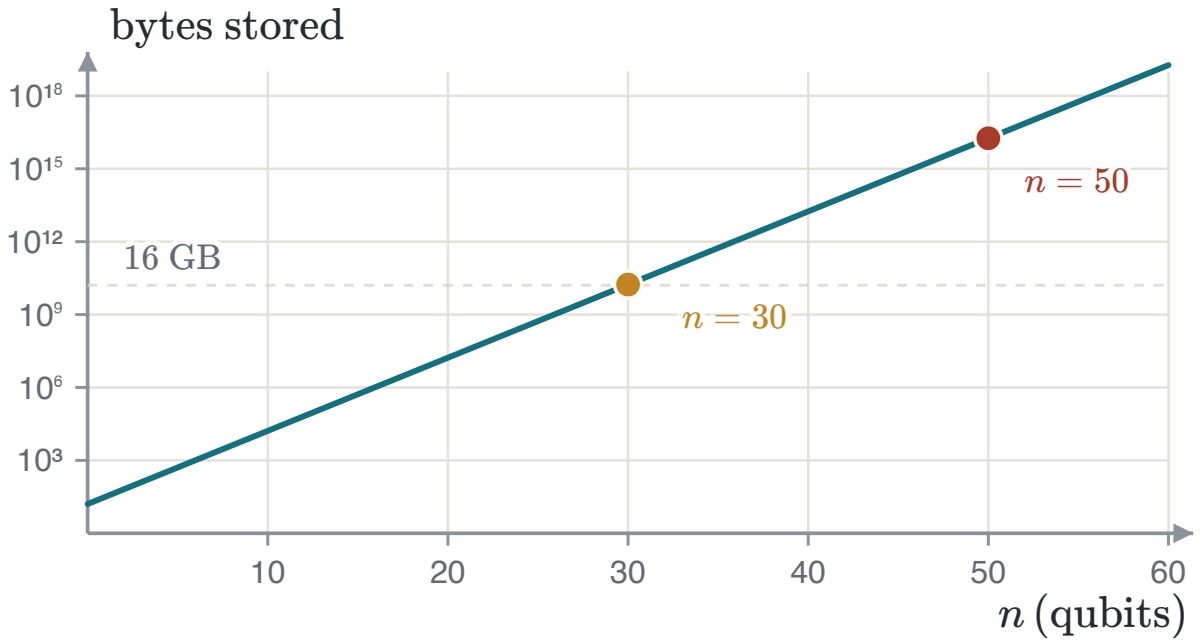
## 5.2 Running a circuit

A classical computer can hold the state vector of a small circuit and apply the gates to it exactly. There are no shots and no noise, and the amplitudes it prints are deterministic up to rounding, which is what makes it useful: it answers whether the circuit is the circuit that was meant. It also stops, because the vector has $2^n$ complex entries at sixteen bytes each.

WHAT AN EXACT SIMULATION COSTS

$$\text{bytes} = 16 \cdot 2^n, \qquad n = 30 \Rightarrow 17\,\mathrm{GB}, \qquad n = 50 \Rightarrow 18\,\mathrm{PB}$$

The unitary matrix of a circuit is worse at $16 \cdot 4^n$: a fifteen-qubit gate matrix already needs $17\,\mathrm{GB}$, which is why a circuit is checked by acting on states rather than by forming its matrix. Note what this argument does **not** establish. "A classical computer cannot store this state" is a statement about one method; many circuits far too large for a state vector are easy to simulate another way, so the size of the vector is an upper bound on the difficulty and never a lower one.




What one state vector costs, against the number of qubits. The vertical axis is logarithmic, so the exponential is a straight line and no amount of extra hardware bends it.

A machine gives back strings, not amplitudes. Counting how often an outcome came up and dividing by the number of shots gives an **estimate** of its probability, and an estimate is not the probability.

THE ERROR BAR ON A PROBABILITY READ FROM A FINITE RUN

$$K \sim \mathrm{Binomial}(N, p), \qquad \mathrm{SE}(\hat{p}) = \sqrt{\frac{p(1-p)}{N}}$$

The square root is the whole economics of running a quantum computer: ten times the accuracy costs a hundred times the shots, and every shot is a full circuit executed on the machine. An estimate near one half to within $\pm 0.005$ needs ten thousand shots. Choosing the shot count from a target accuracy rather than from habit is the difference between a five-minute job and a five-hour one.

MORE SHOTS NEVER REMOVE A BIAS

Sampling error shrinks as $1/\sqrt{N}$ and goes to zero. Device error does not: a faulty gate or a mis-calibrated readout changes the distribution being sampled, so more shots estimate the **wrong** probability more precisely. When a result disagrees with the ideal value by far more than its error bar, the answer is never more shots.

A measurement need not be the last thing a circuit does, and once it is not, one question decides whether the circuit can be rearranged: does anything after the measurement act on the measured wire? If a qubit is measured and its result only ever **controls** later gates, the measurement can be pushed to the end and the classical control replaced by a quantum one, with identical counts. The reason is short: a control acts differently on the two branches of the control qubit and does nothing to mix them, so whether the branches were separated first makes no difference to any later count.

DEFERRED MEASUREMENT

$$\text{measure } q_0,\ \text{then } X^m \text{ on } q_1 \ \equiv\ \text{CNOT}_{0\to1},\ \text{then measure } q_0$$

Put a Hadamard on $q_0$ **after** the measurement and the two circuits stop agreeing at once, because the measurement destroyed the phase that Hadamard needed. The rule is about a measured wire used as a control and nothing else, and a barrier in a compiled circuit is the mark that tells an optimiser it may not move an instruction across a particular place.

A **dynamic circuit** goes the other way: it measures part-way through a shot and uses that classical bit to decide what to do next, inside the same shot, with the decision made by control electronics beside the machine. Teleportation, error correction and state reset all need a gate whose identity is not known until a measurement has happened, and none of them can wait for a result to travel to a host computer and back. Feedforward is not free — a mid-circuit measurement takes far longer than a gate, and the electronics need time to read the bit and act on it — so a dynamic circuit can be worse than the fixed circuit it replaced even with fewer gates.

## 5.3 Compiling for a machine

Chapter 4 showed that a small set of gates is universal. A real machine chooses one such set — its **instruction set** — and can run nothing else, so a circuit written in convenient gates has to be translated first. The translations are the identities of chapter 4 used in the other direction, and none of them is an approximation.

TWO EXACT REWRITES

$$H = e^{i\pi/2} R_z\left(\tfrac{\pi}{2}\right) R_x\left(\tfrac{\pi}{2}\right) R_z\left(\tfrac{\pi}{2}\right)$$

$$\text{CNOT}_{0\to1} = (H \text{ on } q_1)\ \ \text{CZ}\ \ (H \text{ on } q_1)$$

What is lost is only length: one gate becomes three, and a circuit of a thousand becomes a circuit of three thousand. The two-qubit count, which is the number the error budget cares about, does not move. The phase $e^{i\pi/2}$ in the first line is invisible on its own and fully observable the moment the gate is controlled, which is why a compiler tracks it and a person rewriting a controlled gate by hand often does not.

A circuit also assumes that any two qubits can share a gate, and a chip does not: only certain pairs are physically joined, and the list of joined pairs is the **coupling map**. Bringing a circuit onto a chip is four compiler passes. **Layout** chooses which physical qubit plays each qubit of the circuit. **Routing** inserts SWAP gates so that every two-qubit gate acts on a joined pair when it runs. **Translation** rewrites into the instruction set, and **optimisation** then removes what it can.

Routing is the expensive one, and it is expensive in the currency that matters. A SWAP is three CNOTs, and a CNOT is the gate with the largest error, so a gate between qubits $d$ steps apart costs about $3(d-1)$ extra CNOTs to bring them together and possibly the same again to put them back. On a line of five qubits, a CNOT written between the two ends becomes ten two-qubit gates, or nineteen if the original layout has to be restored. Nothing about the algorithm changed; the chip decided the cost, after the circuit was written.

TWO RUNS OF A COMPILER NEED NOT AGREE

Layout and routing are heuristic searches and most of them are randomised, so the same circuit compiled twice can differ in depth by a large factor and both results are correct. A published gate count is incomplete unless it names the compiler, its settings and its seed — the same discipline any numerical result needs.

That brings the third sentence of this course to the place where it becomes a working tool rather than a warning. A statement that one method beats another has to name five things, and a statement missing any of them cannot be checked at all.

| | |
|---|---|
| task | what exactly is to be produced, and from what |
| input model | how the data is reached: a query, a list, a formula |
| accuracy | with what probability, and to what error |
| hardware model | how many qubits, how connected, how noisy |
| baseline | the best classical algorithm for the same task |

The five. The last one is where most claims fail, and it fails in a particular way: the quantum method is compared against the obvious classical method rather than the best one.

A QUERY COUNT IS NOT A RUNTIME

Almost every quantum speedup in a textbook is stated as a count of **queries** to a black box. That is a real and provable statement, and it is not a time. One quantum query is a whole circuit, deep and error-corrected, running for microseconds; one classical query is a memory read taking nanoseconds. A ratio of query counts and a ratio of runtimes can point in opposite directions, and section 5.7 works an example where they do.

### From physical qubits to logical qubits

A **physical qubit** is a device on a chip. A **logical qubit** is quantum information encoded across a block of physical data qubits. Extra ancilla qubits measure checks on the block. Their results form a **syndrome**, which reveals information about an error pattern without reading the unknown amplitudes. A classical decoder maps that syndrome to a correction, and the checks repeat while the computation runs.

A code corrects only the error patterns it was designed to handle. Increasing its **distance** can protect against more faults, but it also needs more physical qubits and more check rounds. Let $p$ be the physical error rate and $p_{\mathrm{th}}$ the threshold for a stated code, noise model and control system. When $p < p_{\mathrm{th}}$, increasing the distance can reduce the logical error rate. Above that threshold, making the same code larger does not provide scalable reliability. The threshold is not one universal hardware number.

LOGICAL RESOURCES ARE NOT HARDWARE RESOURCES

A count of logical qubits and logical gates is only the first line of a hardware estimate. The estimate must name the code and distance, physical error assumptions, ancillas and routing. It must also name the correction-cycle time and target failure probability. Error correction expands both space and time, sometimes by many orders of magnitude.

## 5.4 Turning a phase into counts

Every algorithm in this course and the next works by arranging a relative phase and then converting it into a probability. The smallest circuit that does both is three gates long and is worth being able to write from memory.

A PHASE WRITTEN BETWEEN TWO HADAMARDS

$$|0\rangle \xrightarrow{H} \tfrac{1}{\sqrt{2}}\left(|0\rangle + |1\rangle\right) \xrightarrow{P(\varphi)} \tfrac{1}{\sqrt{2}}\left(|0\rangle + e^{i\varphi}|1\rangle\right)$$

$$\xrightarrow{H} \tfrac{1}{2}\left(1 + e^{i\varphi}\right)|0\rangle + \tfrac{1}{2}\left(1 - e^{i\varphi}\right)|1\rangle, \qquad p(0) = \cos^2\frac{\varphi}{2}$$

The middle state has probability one half on each outcome for every $\varphi$, so measuring there learns nothing about the phase at all. The second Hadamard is what makes the phase readable, and it does so by bringing the two amplitudes into the same outcome so that they can add or cancel. Replace $P(\varphi)$ by a wait of duration $t$ and the phase becomes $\Delta\omega\, t$, so the counts oscillate as the wait grows and their shrinking amplitude is the $T_2$ of chapter 3. The most common measurement made on a real qubit is this circuit with a dial on the middle gate.

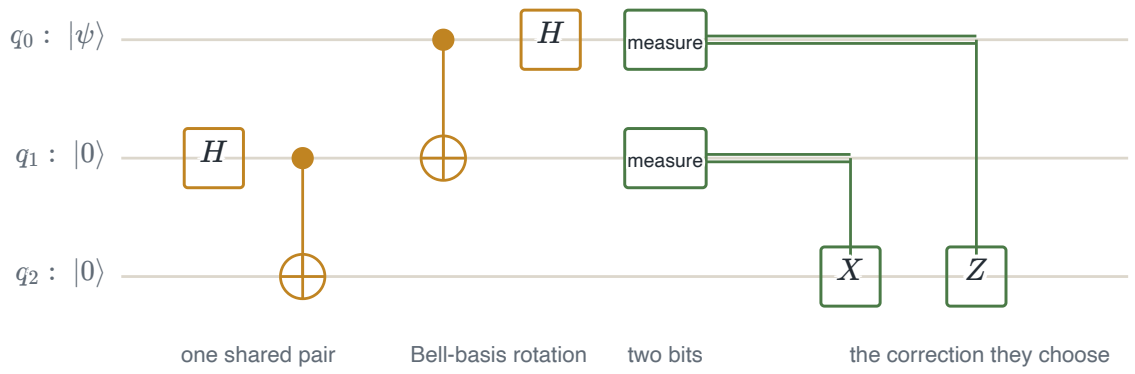
## 5.5 Teleportation

Before asking how to move a quantum state from one place to another, it is worth knowing why the classical answer — copy it and send the copy — is unavailable. Suppose a unitary copied every state, so that $U|\psi\rangle|0\rangle = |\psi\rangle|\psi\rangle$ for all $|\psi\rangle$. Apply it to two states and take the inner product of the results; a unitary preserves inner products, so $\langle\phi|\psi\rangle = \left(\langle\phi|\psi\rangle\right)^2$, which forces the overlap to be zero or one.

WHAT NO-CLONING DOES AND DOES NOT FORBID

A copier can exist only for a set of states that are all either identical or orthogonal, which is exactly a set of classical alternatives — and is why classical bits can be copied. A known state can be prepared again as often as wanted. What cannot be done is copying a state nobody has been told. The obvious circuit shows the failure directly: a CNOT sends $|+\rangle|0\rangle$ to $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, which overlaps the real copy $|+\rangle|+\rangle$ with probability one half and leaves each half maximally mixed. It entangled instead of copying.

Teleportation moves the state instead. Three qubits, and each belongs to somebody: $q_0$ carries the unknown state and belongs to Alice, $q_1$ is Alice's half of an entangled pair, and $q_2$ is Bob's half, which he has taken away with him. The circuit has four stages, and the last is not a quantum gate at all.




The whole protocol. The double lines are classical bits, and everything to the right of them waits for those bits to arrive. That wait is not a detail of the drawing; it is the reason nothing here travels faster than light.

The protocol is one algebraic identity, and it is worth doing in full once, because every claim about teleportation is read straight off the last line. Write $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ on $q_0$ and the shared pair on $q_1$ and $q_2$, with kets written $|q_2 q_1 q_0\rangle$ throughout.

STEP ONE: THE STARTING STATE, MULTIPLIED OUT

$$|\Psi_0\rangle = (\alpha|0\rangle + \beta|1\rangle)_0 \otimes \tfrac{1}{\sqrt{2}}(|00\rangle + |11\rangle)_{21} = \tfrac{1}{\sqrt{2}}\Big[\alpha(|000\rangle + |110\rangle) + \beta(|001\rangle + |111\rangle)\Big]$$

STEP TWO: THE CNOT FROM $q_0$ TO $q_1$ FLIPS $q_1$ WHERE $q_0 = 1$

$$|\Psi_1\rangle = \tfrac{1}{\sqrt{2}}\Big[\alpha(|000\rangle + |110\rangle) + \beta(|011\rangle + |101\rangle)\Big]$$

STEP THREE: THE HADAMARD ON $q_0$ SPLITS EVERY TERM

$$|\Psi_2\rangle = \tfrac{1}{2}\Big[\alpha|000\rangle + \alpha|001\rangle + \alpha|110\rangle + \alpha|111\rangle + \beta|010\rangle - \beta|011\rangle + \beta|100\rangle - \beta|101\rangle\Big]$$

STEP FOUR: COLLECT THE TERMS BY WHAT ALICE WILL READ

$$|\Psi_2\rangle = \tfrac{1}{2}\sum_{m_1, m_0 \in \{0,1\}} |m_1 m_0\rangle_{10} \otimes X^{m_1} Z^{m_0} |\psi\rangle_2$$

That is the identity. It is exact, it holds for every $\alpha$ and $\beta$, and nothing in it depends on knowing what they are — which is the point, because Alice does not. One warning about how to read it: before Alice measures, the four terms are a superposition and not four alternatives that have already happened. Treating them as four equally likely cases gives the right number for every question in this course and is the wrong description of the state, in exactly the way chapter 3 separated a superposition from a mixture.

Since $X$ and $Z$ are their own inverses, undoing the operator is immediate: Bob applies $X^{m_1}$ and then $Z^{m_0}$. Four bit patterns, four corrections, and the choice is made by the bits and by nothing else.

| $m_1 m_0$ | PROBABILITY | WHAT BOB HOLDS | WHAT BOB APPLIES |
|---|---|---|---|
| 00 | 0.25 | $\|\psi\rangle$ | $I$ |
| 01 | 0.25 | $Z\|\psi\rangle$ | $Z$ |
| 10 | 0.25 | $X\|\psi\rangle$ | $X$ |
| 11 | 0.25 | $XZ\|\psi\rangle$ | $X$ then $Z$ |

The two bits are individually fair coins and carry no information at all about $\alpha$ or $\beta$. What they carry is which of four known operations was applied, and that is why two classical bits are enough to move a state described by two continuous parameters: the bits do not describe the state, they select a correction. Skipping the correction does not give a slightly worse state — averaging the four branches with weight one quarter each is the depolarising channel of chapter 3 at full strength, and it destroys the state completely.

The same average is the reason nothing is sent. Chapter 3 proved that no local operation on one half of an entangled pair changes the other half's reduced state, and here that theorem becomes a circuit — the one everybody first suspects of sending something instantly.

WHAT BOB HOLDS BEFORE THE BITS ARRIVE

$$\rho_B = \tfrac{1}{4} \sum_{m_1, m_0} X^{m_1} Z^{m_0} \, \rho \, Z^{m_0} X^{m_1} = \frac{I}{2}$$

The right-hand side has no $\rho$ in it. Whatever Alice was sending, whatever she did, and whether or not she ran the protocol at all, Bob's qubit is maximally mixed until two classical bits reach him. The sum is easiest to check in the Bloch picture: $X$ flips the signs of $r_y$ and $r_z$, $Z$ flips $r_x$ and $r_y$, and $XZ$ flips $r_x$ and $r_z$, so each component appears twice with each sign and the four vectors add to zero. Alice's measurement changes what **she** can say about Bob's qubit; it changes nothing he can observe, and that is why the theory and relativity do not collide.

The account is short and exact. Moving one unknown qubit consumes one shared entangled pair and two classical bits, and destroys the original; nothing is reusable, and a second qubit needs a second pair. Saying that a laboratory has **done** this needs a number, because a noisy apparatus produces something that resembles the input, and the number is the fidelity $F(\psi) = \langle\psi|\rho_{\text{out}}|\psi\rangle$.

WHY THIS IS A QUANTUM-NETWORK PRIMITIVE

A quantum network first distributes an entangled pair across a link. Teleportation then consumes that pair and two classical bits to move a state between nodes without sending the data qubit through the link. Teleportation does not create the pair, remove loss or extend the link by itself. A repeater needs entanglement generation, verification and link joining around it.

A BEST-CASE FIDELITY IS NOT A RESULT

One fidelity on one input proves nothing: a machine that always outputs $|0\rangle$ scores $F = 1$ whenever the input happened to be $|0\rangle$. The claim has to be an average over inputs covering the sphere, and it has to beat $2/3$ — what measuring the qubit and preparing a new one achieves with no entanglement at all. With pairs of singlet fraction $f$ the optimum is $F_{\text{avg}} = (2f + 1)/3$, which meets both ends correctly: $f = 0.5$ gives exactly $2/3$ and $f = 1$ gives $1$. A report must also state the unconditional success rate, what was discarded and by what rule, and a confidence interval.

## 5.6 Grover search

The problem is stated carefully because the statement is what the result is about. There are $N = 2^n$ candidates; a function $f$ answers one question about each, returning $1$ for a solution and $0$ otherwise; exactly $M$ candidates are solutions; and nothing else is known, so the candidates have no order and no structure to exploit. The function is available only as a black box, built as chapter 4 built every reversible embedding: $U_f|x\rangle|y\rangle = |x\rangle|y \oplus f(x)\rangle$. One use of that box is one **query**, and the classical cost in the same model is about $N/2$ queries when $M = 1$.

The oracle flips a target bit and interference needs a phase. One line of preparation converts one into the other, and it is the mechanism every algorithm in this chapter and the next runs on.

PHASE KICKBACK

$$X|-\rangle = -|-\rangle$$

$$U_f\,|x\rangle|-\rangle = (-1)^{f(x)}\,|x\rangle|-\rangle$$

Read the second line carefully. The target register comes out exactly as it went in, so it is not entangled with anything and can be ignored from here on; the answer has landed on the first register, as a sign. Applied to a superposition, one query signs every marked term at once — which is not "trying every answer at once", because no probability has changed and a measurement here still returns one string. What has changed is a set of relative phases, and relative phases are the only thing an algorithm can work with. The argument needs the target to come out unchanged, so an oracle that leaves an ancilla behind entangles the register with it and the interference at the end simply does not happen.

An $n$-qubit state lives in $2^n$ dimensions and following it there would be hopeless. It is not necessary. Define the even mixture of the marked candidates and the even mixture of the rest; they are orthonormal, the starting state lies in the plane they span, and both operations the algorithm uses keep the state inside that plane.

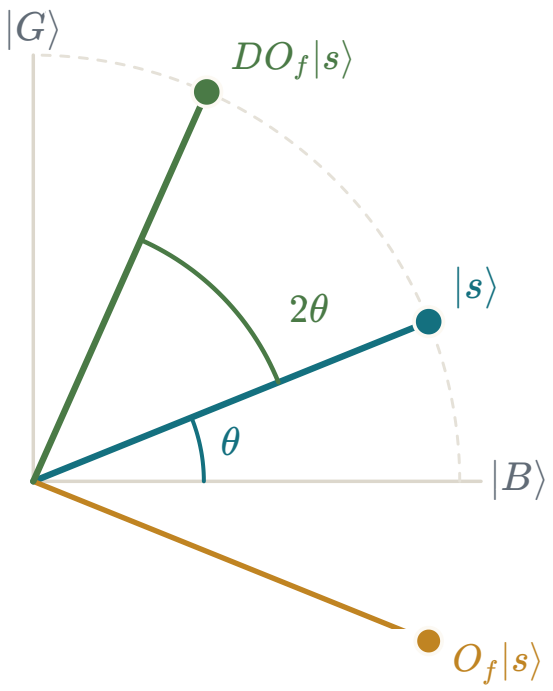
THE PLANE THE ALGORITHM LIVES IN

$$|G\rangle = \frac{1}{\sqrt{M}} \sum_{f(x)=1} |x\rangle, \qquad |B\rangle = \frac{1}{\sqrt{N-M}} \sum_{f(x)=0} |x\rangle$$

$$|s\rangle = \sin\theta\,|G\rangle + \cos\theta\,|B\rangle, \qquad \sin\theta = \sqrt{\frac{M}{N}}$$

A problem in $2^n$ dimensions has become a problem in two, and the two can be drawn on paper. One warning: this is a two-dimensional real plane inside a $2^n$-dimensional complex space. It is not a Bloch sphere, the vectors are not Bloch vectors, and the half-angle rules of chapter 4 do not apply to it.

One iteration is two operations and each of them is a reflection of that plane. The **phase oracle** $O_f = I - 2|G\rangle\langle G|$ puts a minus sign on the marked terms, which is a reflection in the $|B\rangle$ axis. The **diffusion operator** $D = 2|s\rangle\langle s| - I$ reflects in the starting vector. Two reflections of a plane, in lines separated by an angle, compose to a rotation by twice that angle — and the two lines here are separated by $\theta$.




One iteration, drawn isotropically so the angles are the real ones. The oracle takes the vector below the horizontal; the diffusion takes it back above, but past where it started. The net turn is $2\theta$.

Written out in the computational basis, $D$ says: replace each amplitude by twice the mean of all the amplitudes, minus itself. That is the familiar "inversion about the mean", and it is the same operator in a different basis. The reflection picture is the one to reason with and the mean picture is the one to program. Note also what $D$ is not: it contains no information about $f$, so it is not counted as a query — but it is still a circuit, and its multiply-controlled phase needs a chain of Toffolis and their ancillas.

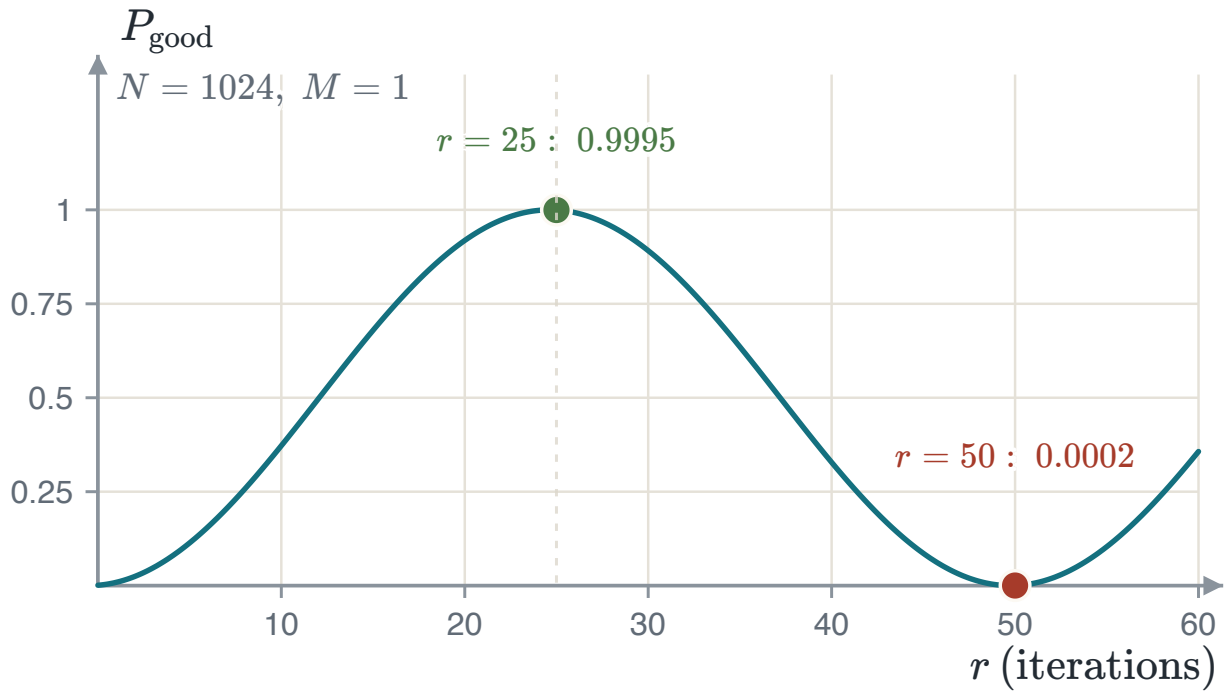
After $r$ iterations the state has turned to $(2r+1)\theta$ from the $|B\rangle$ axis, so the probability of measuring a marked candidate is a sine squared and nothing else.

THE SUCCESS PROBABILITY, AND WHERE IT PEAKS

$$P_{\text{good}}(r) = \sin^2\big((2r+1)\theta\big)$$

$$r_* = \frac{\pi}{4\theta} - \frac{1}{2}, \qquad r_* \approx \frac{\pi}{4}\sqrt{\frac{N}{M}} \ \text{ when } M \ll N$$

The square root arrives here, at the end, and it comes from the size of one step: each iteration adds a fixed angle $2\theta \approx 2\sqrt{M/N}$ and a right angle has to be reached, so the number of steps goes as $\sqrt{N/M}$. Nothing exponential ever happened; a small angle was added many times.




The success probability against the iteration count, for a thousand candidates with one marked. It rises to almost one at twenty-five iterations and falls back to almost nothing at fifty. The curve is a sine squared and it keeps going.

---

EXAMPLE 5.2 · EIGHT CANDIDATES, TWO MARKED

GIVEN $N = 8$ and $M = 2$.

FIND The angle, the best iteration count, and the success probability at it and at the next one.

SOLUTION $\sin\theta = \sqrt{2/8} = 0.5$, so $\theta = 30^\circ$ and $r_* = \frac{90}{60} - \frac{1}{2} = 1$ exactly. Then $P(1) = \sin^2 90^\circ = 1$: the answer is certain after one iteration.

CHECK A second iteration gives $P = \sin^2 150^\circ = 0.25$, which is exactly $P(0)$ — the probability before any iteration at all. Two iterations have undone the work of one.

---

MORE ITERATIONS ARE NOT BETTER

Past $r_*$ the rotation carries the vector beyond $|G\rangle$ and the success probability falls; at about $2r_*$ it is back where it started. Running "as many as the time allows" is therefore worse than running the right number, and it fails in a way that looks like a broken machine — a flat distribution over every outcome — produced by a perfect machine following a bad plan. When $M$ is not known the fix is not more iterations: it is a randomised schedule of iteration counts, or estimating $M$ first.

## 5.7 What the square root claims

The result is real and it is provable. About $\frac{\pi}{4}\sqrt{N}$ queries find a marked candidate among $N$, where a classical search needs about $N/2$, and it is optimal: no quantum algorithm in this model does better than order $\sqrt{N}$, so the quadratic saving is the end of the story rather than the beginning of it. Written out against the five components, three of them turn out to be doing heavy work.

- **Input model.** The saving assumes query access to a reversible circuit for $f$. If the candidates are records in a database they must first be loaded into a quantum memory, and loading $N$ records takes at least $N$ operations — which destroys the saving before the algorithm starts.
- **Hardware model.** The count assumes ideal qubits. A circuit of $\sqrt{N}$ oracle calls, each with its Toffolis and ancillas, needs error correction, and error correction multiplies both the qubit count and the depth by a large factor.
- **Baseline.** The comparison assumes the classical method must scan. Almost no real problem is unstructured enough for that to be true: a sorted list is searched in $\log_2 N$ steps and a hash table in one.

---

EXAMPLE 5.3 · THE QUERY RATIO AND THE TIME RATIO POINT OPPOSITE WAYS

GIVEN — $N = 1024$ with one marked candidate, an oracle circuit taking $10\,\mu\text{s}$ per call, and a classical processor evaluating the same predicate in $10\,\text{ns}$.

FIND — The two query counts and the two runtimes.

SOLUTION — Queries: $25$ against $512$, a saving of about twenty times. Time: $25 \times 10\,\mu\text{s} = 250\,\mu\text{s}$ against $512 \times 10\,\text{ns} = 5.1\,\mu\text{s}$ — the classical run is about fifty times faster.

CHECK — Both ratios are correct. The crossing point is the $N$ at which $\sqrt{N}$ slow queries beat $N$ fast ones, and finding that $N$ is the honest form of the question.

---

WHAT THE RESULT DOES SAY, STATED FAIRLY

For a problem with genuinely no structure, where the predicate is cheap to build as a reversible circuit and the data never has to be loaded, a quantum computer needs quadratically fewer evaluations of that predicate, and this is optimal. That is a strong theorem. It is a theorem about a count, and turning it into a statement about a wall clock needs every one of the other four components filled in.

## 5.8 Summary

- A circuit is a program: prepare $|0\rangle^{\otimes n}$, apply gates from a fixed set, measure in the computational basis, repeat. Depth is layers and is what a coherence time is spent against; the gate count is a different number and only accidentally equal to it.
- $q_0$ is the top wire, the last digit of the ket and the least significant bit of the index, so the wire order and the ket digits run opposite ways.
- An exact state vector costs $16 \cdot 2^n$ bytes and stops near thirty qubits; a circuit matrix costs $16 \cdot 4^n$ and stops much sooner.
- A probability from $N$ shots carries $\sqrt{p(1-p)/N}$, which shrinks as $1/\sqrt{N}$. Device error does not shrink at all, so a large disagreement is never a reason for more shots.
- A measured control may be deferred to the end of a circuit; a gate on the measured wire may not cross the measurement. A compiler rewrites gates exactly and routes them at three CNOTs per SWAP, and two runs of it need not agree.
- A logical qubit is an encoded block of physical data qubits and check ancillas. Syndromes guide corrections. Below the threshold, more distance can suppress logical error, while the encoding and correction cycles expand both space and time.
- Teleportation: $|\Psi\rangle = \frac{1}{2}\sum_{m_1 m_0} |m_1 m_0\rangle\, X^{m_1} Z^{m_0} |\psi\rangle$, four branches of probability $\frac{1}{4}$ for every input, so $\rho_B = I/2$ until two classical bits arrive. A network consumes one distributed pair and those two bits to move one state between nodes.
- Grover: $\sin\theta = \sqrt{M/N}$, one iteration is a rotation by $2\theta$, and $P(r) = \sin^2((2r+1)\theta)$ with $r_* = \frac{\pi}{4\theta} - \frac{1}{2}$. Past the optimum the probability falls. Quadratic in queries, optimal in that model, and silent about time.

> FOUR ERRORS THAT COST A WHOLE QUESTION
>
> Reporting a gate count where a depth was asked for. Reading a printed bit string in the wrong direction. Saying that teleportation finished before the classical bits arrived. And running more Grover iterations than the optimum, which lowers the success probability rather than raising it.

> WHAT COMES NEXT
>
> Chapter 6 keeps the oracle and the kickback and changes what is done between them: one mechanism, four algorithms, and the resource discipline of this chapter applied to every one of them.

6

# Quantum algorithms

*A quantum computer does not try every answer at once. It holds many amplitudes, and a measurement returns one string of bits. An algorithm earns something only if it can arrange for the amplitudes of the answers it does not want to cancel before that measurement happens. This chapter is four ways of arranging exactly that, and they are all the same arrangement.*

The mechanism is called **phase kickback**. An operation is written to change one register; it is pointed instead at a register already prepared in a state that operation cannot move; and so the only thing it can do is write a phase onto the register that controlled it. That is the whole of Deutsch, of Deutsch–Jozsa, of phase estimation and of the order finding inside Shor's algorithm. What differs between them is which question was written into the phase, and which transform was used to make the wrong answers cancel afterwards.

THE SENTENCE EVERY SECTION HERE ANSWERS

Preparing a superposition of $2^n$ inputs is one layer of Hadamards, it is free, and on its own it is worth nothing: a measurement of that state returns a uniformly random string, which is what a coin does. Every gain in this chapter comes from the interference after the query, never from the superposition before it.

## 6.1 What a query model counts

Most of the results here are stated in a **query model**. The algorithm is given a black box that computes an unknown function $f$, and it is charged one **query** each time it uses that box, whatever is inside. The exclusive-or in the second register is what makes the box reversible, and chapter 4 built exactly this embedding.

THE ORACLE, AND WHAT ONE QUERY MEANS

$$U_f\,|x\rangle|y\rangle = |x\rangle\,|y \oplus f(x)\rangle$$

The model is useful because it can be reasoned about: how many questions must be asked before the answer is forced is a question with a proof attached to it. It is also a model that hides things on purpose. Building $U_f$ as a real circuit costs gates, and those gates are not counted. Loading data into the box, if the function is a lookup rather than a formula, is not counted. Error correction is not counted. A query separation is a theorem about one column of a table with three columns in it.

Three costs, then, and they are not interchangeable. The **query count** is how many times the box is opened. The **gate count** is what it cost to build the box and everything around it. The **end-to-end cost** is both of those plus input preparation, fault tolerance, measurement and classical post-processing. An algorithm that needs $25$ queries where a classical method needs $512$ has improved the first column by a factor of twenty; if one quantum query is a circuit of four thousand gates and one classical evaluation is thirty operations, the second column has got six times worse.

A word about what is being counted as an input. A **decision problem** asks a yes-or-no question, and its size is the number of bits the input is written in. The number $N$ has about $\log_2 N$ bits, so an algorithm that runs in time $N$ is exponential in the size of its input and not linear. That single point decides most of the arguments in this subject.

- **P** — decision problems an ordinary algorithm settles in polynomial time.
- **BPP** — the same, for an algorithm allowed to toss coins and be wrong with probability below one third.
- **BQP** — the same, for a quantum algorithm with the same allowance.
- **NP** — problems whose yes answers carry a certificate that can be checked quickly, which says nothing about finding it.

WHAT IS KNOWN, AND WHAT IS NOT

$\mathrm{P} \subseteq \mathrm{BPP} \subseteq \mathrm{BQP}$ and $\mathrm{P} \subseteq \mathrm{NP}$. Whether $\mathrm{NP} \subseteq \mathrm{BQP}$ is open, and no efficient quantum algorithm is known for any NP-complete problem. Grover gives a square root on the search, and a square root of an exponential is still an exponential. Factoring is in NP and in BQP and is not known to be NP-complete — which is exactly why it can have the structure the rest of this chapter exploits.

## 6.2 One mechanism: phase kickback

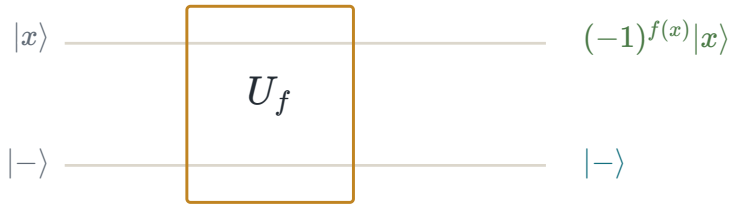
The oracle writes $f(x)$ into the second register by flipping it, and a flip is the gate $X$. So choose for the second register the one state that $X$ leaves alone up to a sign. When $f(x) = 0$ nothing happens; when $f(x) = 1$ the flip happens and produces a minus sign; and both cases are one line.

PHASE KICKBACK

$$X\,|-\rangle = -\,|-\rangle$$

$$U_f\,|x\rangle|-\rangle = (-1)^{f(x)}\,|x\rangle|-\rangle$$

The second register never changes, so it can be prepared once and ignored. The value of $f$ has become a **relative phase** among the terms of the first register, and chapter 1 said what that is worth: a relative phase is observable, and it is observable through interference and through nothing else. The step readers re-derive three times before believing it is that the sign belongs to the first register even though the gate was written on the second. It multiplies a term of the whole state, and because the second factor is the same $|-\rangle$ in every term, the only place the sign can make a difference is between the terms of the first.




the target leaves unchanged

The oracle acts, the lower wire comes out exactly as it went in, and the answer is a sign attached to the upper one. Nothing was measured and nothing was copied.

Nothing in that argument used the fact that the operation was a bit flip. It used only that the target was prepared in an **eigenstate** of it. Replace $X$ by any unitary $U$ and $|-\rangle$ by an eigenstate $|u\rangle$, and the same thing happens with a general phase in place of the sign. Because $U$ is unitary its eigenvalues have modulus one, so they are pure phases and can always be written this way.

THE GENERAL FORM

$$U|u\rangle = e^{2\pi i\varphi}|u\rangle$$

$$\mathrm{c}U\,\tfrac{1}{\sqrt{2}}\big(|0\rangle+|1\rangle\big)|u\rangle = \tfrac{1}{\sqrt{2}}\big(|0\rangle+e^{2\pi i\varphi}|1\rangle\big)|u\rangle$$

The number wanted is now a relative phase on one qubit, which is the one kind of number this course can read. If the third register holds a superposition $\sum_k c_k|u_k\rangle$ rather than a single eigenstate, the procedure returns an estimate of $\varphi_k$ with probability $|c_k|^2$: it samples one eigenphase and does not list the spectrum. That is a limitation everywhere else in the chapter and it is the whole mechanism of order finding.

Now the rule the rest of the chapter answers to. One layer of Hadamards on $n$ qubits produces every input at once, and one query produces every output at once. Measure at that point and a uniformly random $x$ comes out: every amplitude has the same modulus, so every string is equally likely, and the signs — which are the entire content of what the query returned — are invisible in this basis. The run has cost a query and returned a coin toss.

WHAT THE LAST LAYER OF HADAMARDS IS FOR

$$\langle 0^n|\,H^{\otimes n}\Big(\tfrac{1}{2^{n/2}}\textstyle\sum_x(-1)^{f(x)}|x\rangle\Big) = \tfrac{1}{2^n}\sum_x(-1)^{f(x)}$$

That single number is what the algorithm is for. It is not any one value of $f$; it is a global property of the whole table, and it arrived because the unwanted terms cancelled. A quantum algorithm has to make the amplitudes of every answer it does not want cancel, and it has to do that before anything is measured. The readout returns $n$ bits, never $2^n$ numbers, so a design that needs to see many amplitudes has already failed.

"IT COMPUTES ALL THE VALUES IN PARALLEL" IS TRUE AND USELESS

The superposition really does contain all $2^n$ values. It is useless because the only way to look at it is to measure it, and a measurement of a state with $2^n$ equal amplitudes is a random number generator. The interesting question is never how many values are held; it is which arrangement of phases lets one number survive the sum.

## 6.3 Deutsch and Deutsch–Jozsa

A black box holds one of the four functions $f:\{0,1\}\to\{0,1\}$. Two of them are **constant** and two are **balanced**, returning $0$ once and $1$ once. The task is to say which class it is in, and not to find the function. Classically this needs both values, because knowing $f(0)$ alone rules nothing out. The quantum circuit uses one query: prepare $|+\rangle|-\rangle$, query once, and apply a Hadamard to the query qubit.

DEUTSCH'S ALGORITHM, IN ONE LINE

$$\tfrac{1}{\sqrt{2}}\Big((-1)^{f(0)}|0\rangle + (-1)^{f(1)}|1\rangle\Big) \xrightarrow{H} \pm|\, f(0)\oplus f(1)\,\rangle$$

The reading is $f(0)\oplus f(1)$, which is $0$ for both constant functions and $1$ for both balanced ones. It is exactly the one bit that was asked for, and the two bits that were not asked for never appear. A saving of one query on a problem with four instances is not a useful computation and this circuit is not a benchmark; its importance is that it is the smallest complete example of the mechanism.

The same circuit works on $n$ query qubits, for a function promised to be constant or to return $1$ on exactly half of its $2^n$ inputs. The only quantity that has to be computed is the amplitude of the string $0^n$, and the Hadamard layer makes that a plain average of the signs. A constant function makes every sign the same, so the average is $\pm 1$ and the probability of reading $0^n$ is one. A balanced function has exactly $2^{n-1}$ plus signs and $2^{n-1}$ minus signs, so the average is exactly zero. Nothing is approximate.

| | $(-1)^{f(x)}$ for the eight inputs | | | | | | | | mean |
|---|---|---|---|---|---|---|---|---|---|
| constant, value 0 | + | + | + | + | + | + | + | + | $+1$ |
| constant, value 1 | - | - | - | - | - | - | - | - | $-1$ |
| balanced, parity | + | - | - | + | - | + | + | - | $0$ |
| balanced, top bit | + | + | + | + | - | - | - | - | $0$ |

Four promised functions on three bits and the average of their signs. The two balanced rows cancel term by term; that cancellation is the algorithm, and the promise is what guarantees it.

THE PROMISE IS NOT DECORATION

Drop it and the algorithm is worthless. For a function that is neither constant nor balanced the amplitude of $0^n$ is some number between $-1$ and $1$, so a reading of $0^n$ no longer proves anything and a reading of something else no longer proves anything either. A function on four bits returning $1$ on six of its sixteen inputs gives an amplitude of $0.25$ and a probability of $0.0625$, and neither reading licenses a conclusion.

What the separation is, stated carefully. A classical algorithm that must never be wrong has to keep asking until the answer is forced: in the worst case it sees $2^{n-1}$ equal values and still cannot tell, so it needs $2^{n-1}+1$ queries against the quantum circuit's one. That is a genuine exponential separation, and it is a separation between **exact** query complexities **under a promise**.

EXAMPLE 6.1 · THE GAP COLLAPSES AGAINST A RANDOMISED ALGORITHM

GIVEN A promised function on $n = 10$ bits, so $1024$ inputs.

FIND The exact classical count, the quantum count, and the count for an error probability below $10^{-6}$.

SOLUTION Exact: $2^9 + 1 = 513$. Quantum: $1$. Randomised: $k$ distinct queries all agree on a balanced function with probability below $2^{-(k-1)}$, so $k = 21$ suffices.

CHECK Push $n$ to $30$. The exact count becomes $5.4 \times 10^8$ and the randomised count is still $21$: the first grows and the second does not. The exponential separation is the distance to the first count only.

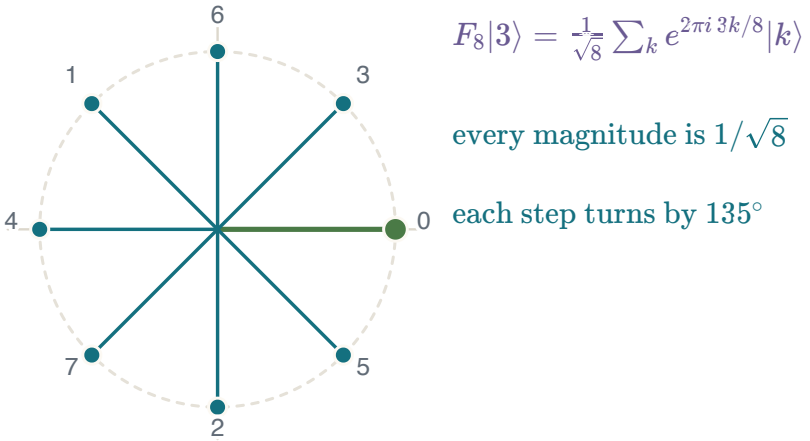
## 6.4 The quantum Fourier transform

On $n$ qubits there are $Q = 2^n$ basis states, and the transform is defined on each of them. Every output amplitude has the same modulus $1/\sqrt{Q}$; what carries the input is the **rate** at which the phase turns as the output index advances.

THE TRANSFORM

$$F_Q\,|x\rangle = \frac{1}{\sqrt{Q}} \sum_{k=0}^{Q-1} e^{2\pi i\, xk/Q}\,|k\rangle$$

It is unitary, so it is a change of basis and nothing more: no information is created and none is destroyed, and the inverse takes the opposite sign in the exponent. On a general input it acts by linearity. The one input worth memorising is $F_Q|0\rangle$, which is the uniform superposition because every phase is one — so the layer of Hadamards that starts every algorithm in this chapter **is** the transform of the all-zero state, and on one qubit the Hadamard is exactly $F_2$.




The eight amplitudes of the transform of the state three, drawn in the complex plane. All eight have the same length, and each step turns by $3 \times 45° = 135°$. Measuring at this point returns a uniformly random index, whatever the input was.

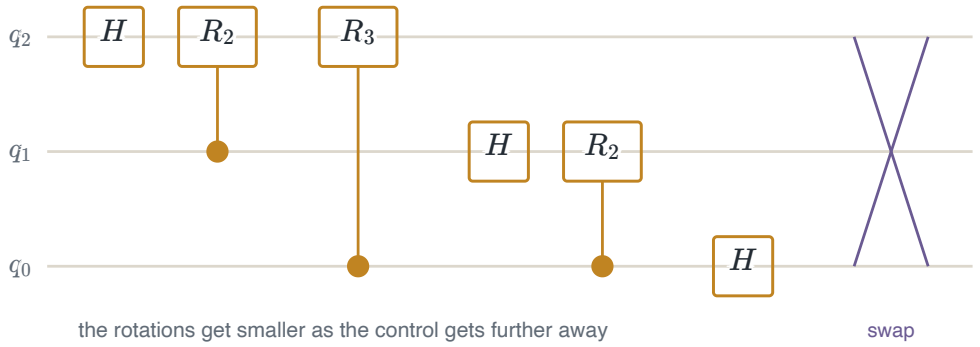
The transform looks like a $Q \times Q$ matrix and would take $Q^2$ multiplications to apply as one. It factors instead into a short circuit, because the phase separates into one factor per bit of the input. Each qubit gets one Hadamard and then one controlled rotation from every qubit below it, with the rotation getting smaller as the control gets further away.

THE CIRCUIT, AND ITS SIZE

$$R_k = \begin{bmatrix} 1 & 0 \\ 0 & e^{2\pi i/2^k} \end{bmatrix}$$

$$n \text{ Hadamards} + \tfrac{1}{2}n(n-1) \text{ rotations} = \tfrac{1}{2}n(n+1) \text{ gates}$$

Ten qubits means fifty-five gates and five output swaps; the smallest rotation there is an angle of $2\pi/1024$, about a third of a degree. Rotations below a threshold are usually dropped, which turns a hardware problem into a bounded algorithmic error and takes the gate count from $O(n^2)$ to $O(n \log n)$. That approximate transform is the standard choice in every resource estimate for factoring.




Three qubits: three Hadamards, three controlled rotations and one swap. Six gates for a transform whose matrix has sixty-four entries.

THE TRANSFORM DOES NOT RETURN A SPECTRUM

Hand it a state $\sum_x a_x |x\rangle$ and it produces $\sum_k \tilde{a}_k |k\rangle$. It does not print the $Q$ numbers $\tilde{a}_k$, and it cannot: the readout returns $n$ bits, so one run returns one index $k$ drawn with probability $|\tilde{a}_k|^2$. It is useful only where the thing wanted can be inferred from samples of that distribution — a period, an order, an eigenphase. A classical fast Fourier transform takes a stored vector of $Q$ numbers and returns $Q$ numbers in about $Q \log_2 Q$ operations; the two solve different problems and their costs should not be compared as though they did not.

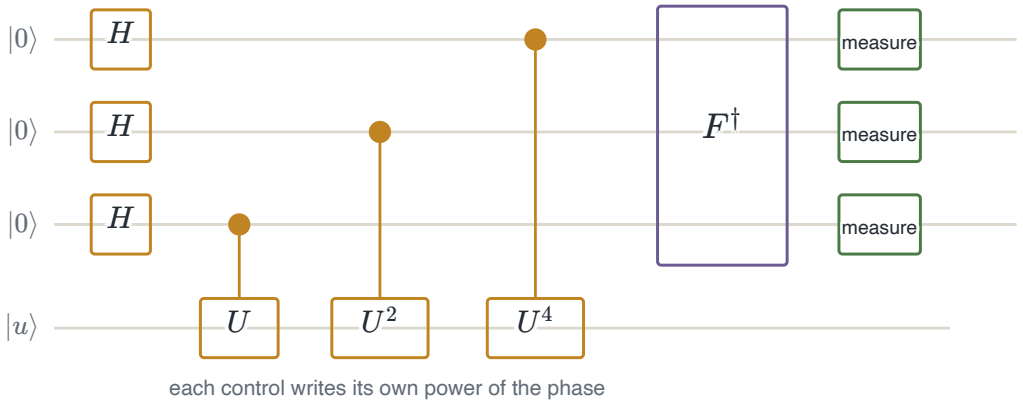
## 6.5 Phase estimation

One controlled $U$ put a phase onto one control qubit. Use $t$ control qubits, and let control $j$ apply $U$ raised to the power $2^j$, so that the phase it collects is doubled each time. The counting register is then exactly the state the Fourier transform produces from the number $2^t\varphi$, so applying the **inverse** transform collapses the phase ramp onto the number that produced it.

WHAT THE COUNTING REGISTER HOLDS BEFORE THE INVERSE TRANSFORM

$$\frac{1}{\sqrt{2^t}} \sum_{k=0}^{2^t-1} e^{2\pi i k\varphi} \, |k\rangle$$

Measuring the $t$ qubits gives an integer $y$, and $y/2^t$ estimates $\varphi$. The powers are powers of two because a $t$-bit binary fraction is $\varphi \approx 0.b_1 b_2 \ldots b_t$ and multiplying by $2^j$ shifts the binary point $j$ places: the control that applies $U^{2^j}$ is reading one bit of the answer. The circuit is a binary expansion written in phases, and the inverse transform is the machine that reads a binary expansion out of phases.



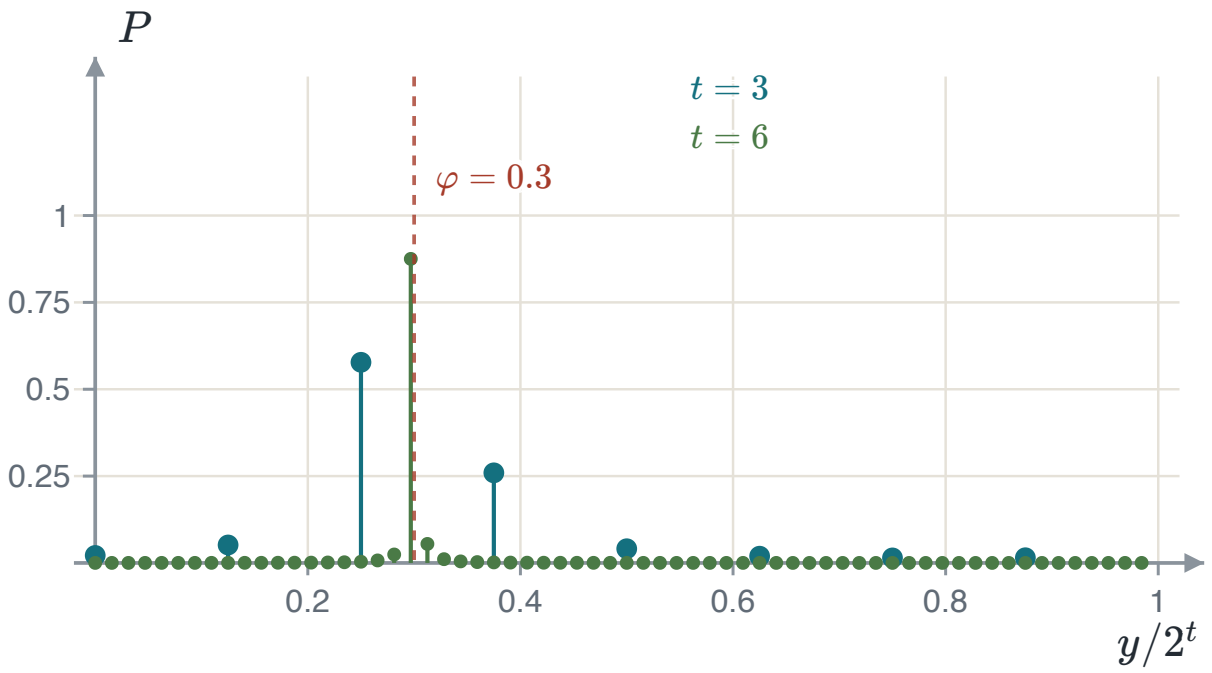

Three counting qubits, three controlled powers and the inverse transform. The eigenstate on the bottom wire is unchanged from beginning to end.

Take first the case where the phase happens to be a $t$-bit binary fraction, $\varphi = m/2^t$. Then the counting register holds exactly the transform of $|m\rangle$, the inverse transform returns $|m\rangle$, and the other $2^t - 1$ outcomes have amplitude exactly zero. It is worth seeing why: the amplitude of outcome $y$ is a geometric sum, and away from $y = m$ its terms are $2^t$ unit vectors spread evenly round the circle, which add to nothing. That is the cancellation of section 6.2, complete rather than approximate — and it is the special case, because almost every real number is not a $t$-bit fraction.

THE GENERAL CASE

$$P(y) = \frac{1}{2^{2t}} \left| \frac{\sin\!\left(\pi \, 2^t \delta\right)}{\sin(\pi\delta)} \right|^2, \qquad \delta = \varphi - \frac{y}{2^t}$$

Two guarantees follow and both are worth remembering: the single nearest outcome carries at least $4/\pi^2 \approx 0.405$, and the two nearest together at least $8/\pi^2 \approx 0.811$. Neither floor moves as the register grows. To get $n$ correct bits with failure probability at most $\varepsilon$, take $t = n + \lceil \log_2(2 + 1/2\varepsilon) \rceil$ counting qubits: the extra qubits are cheap, since halving the failure probability adds about one, while accuracy is expensive, because each extra bit of $n$ doubles the applications of $U$.




The same phase read with three counting qubits and with six. The peak sharpens onto a finer grid and the tails shrink, and at no register size does one outcome carry all the probability.

---

EXAMPLE 6.2 · A PHASE THAT FITS NOTHING

GIVEN $\varphi = 0.3$ and $t = 3$, so $2^t\varphi = 2.4$.

FIND The two most likely readings and their probabilities.

SOLUTION The bracketing outcomes are $y = 2$ and $y = 3$. Putting $\delta = 0.05$ and $\delta = -0.075$ into the formula gives $P(2) = 0.577$ and $P(3) = 0.259$.

CHECK Together $0.836$, above the guaranteed $0.811$ as it must be. The remaining $0.164$ is spread over the other six outcomes, and a run that lands there returns a badly wrong phase with no warning attached.

---

Now add up what the circuit asks for. The inverse transform is $\frac{1}{2}t(t+1)$ gates. The controlled powers are $U$ applied $1 + 2 + \cdots + 2^{t-1} = 2^t - 1$ times, which is exponential in the register size and has to be: reading $t$ bits of a phase means evolving for a time proportional to $2^t$, however the circuit is arranged. At ten counting qubits that is $1023$ applications against $55$ gates, and at a $U$ costing $500$ gates the transform is about one part in ten thousand of the circuit.

EFFICIENT ONLY WHEN THE POWERS CAN BE BUILT DIRECTLY

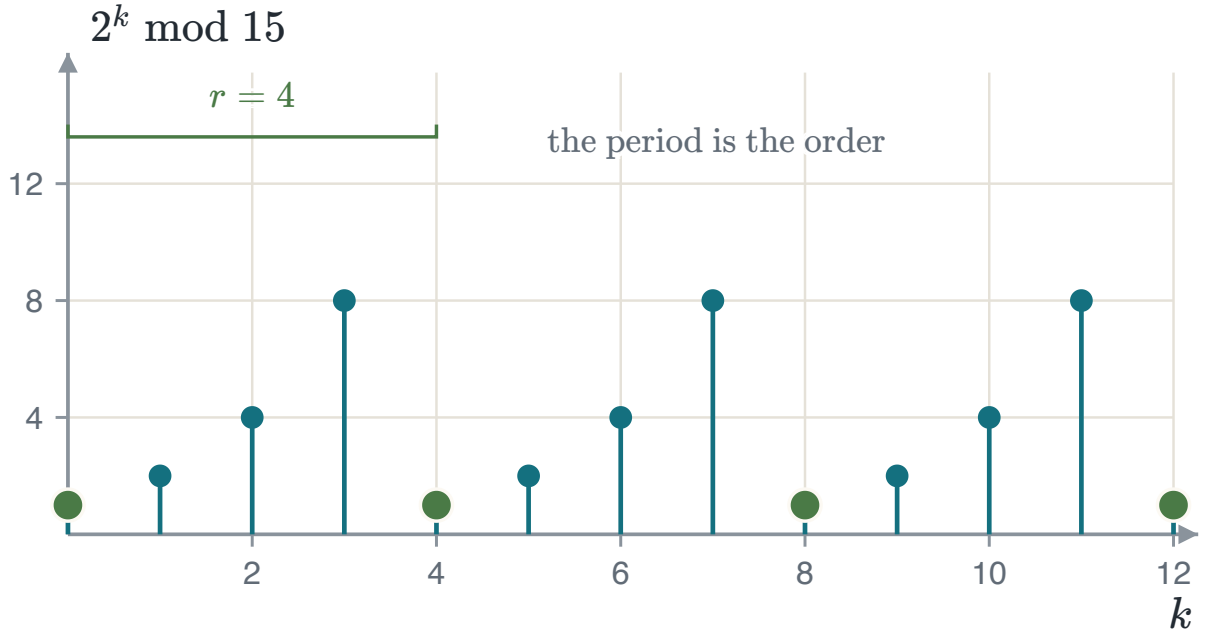
Phase estimation is efficient when $U^{2^j}$ can be built in a circuit of size polynomial in $j$, rather than by repeating $U$ that many times. Order finding can do it, because squaring a modular multiplier is another modular multiplier. For an arbitrary $U$ it cannot be done, and calling the inverse transform efficient does not make the algorithm efficient. Replacing the counting register by one qubit measured and reused, with the earlier results fed forward, saves $t - 1$ qubits and saves nothing at all in the controlled evolution.

One application closes a hole chapter 5 left open. The best number of Grover iterations needs the number of marked candidates, which is usually exactly what is not known. The Grover step is a rotation of a plane by $2\theta$ and a rotation of a plane has eigenvalues $e^{\pm 2i\theta}$, so it is a unitary with the wanted number in its eigenphase. Estimating $\theta$ gives $M = N \sin^2 \theta$, and the register does not even need an eigenstate: the uniform superposition is a combination of the two eigenvectors, and $\pm\theta$ give the same $\sin^2 \theta$. The counting costs about $\sqrt{N}$ applications itself, so counting first and searching after is still a square-root method overall.

## 6.6 Order finding

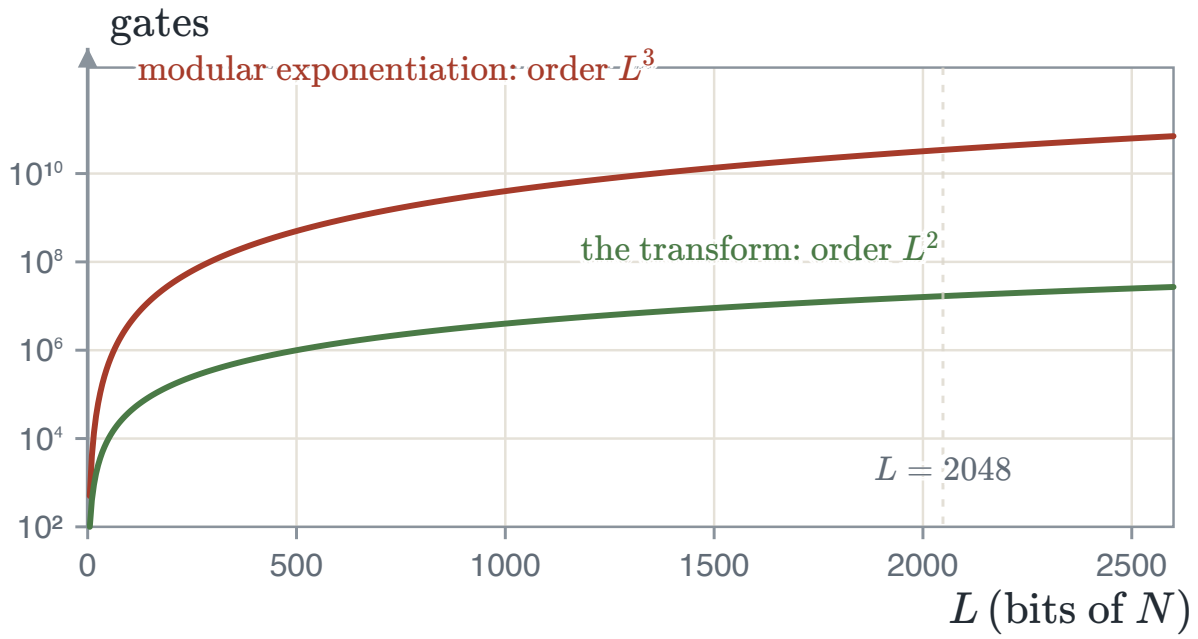
Fix a composite $N$ and a number $a$ with no factor in common with it. The **order** of $a$ modulo $N$ is the smallest positive $r$ with $a^r \equiv 1$. Because $a$ and $N$ share no factor, multiplying by $a$ is a permutation of $0, 1, \ldots, N-1$ — nothing is lost and nothing collides — and a permutation is a unitary, so it is a legal gate. The work register holds $m$ qubits with $2^m \geq N$, and the basis states above $N$ are completed as any permutation, usually the identity.

> THE MULTIPLIER, AND WHAT ITS EIGENPHASES CARRY
>
> $$U_a \,|y\rangle = |\, a\,y \bmod N \,\rangle$$
>
> $$U_a \,|u_s\rangle = e^{2\pi i s/r}\,|u_s\rangle, \qquad |u_s\rangle = \frac{1}{\sqrt{r}} \sum_{k=0}^{r-1} e^{-2\pi i s k/r}\,|a^k \bmod N\rangle$$
>
> The order sits in the denominators of the eigenphases, and phase estimation reads eigenphases: that is the whole reduction. One obstacle remains — preparing $|u_s\rangle$ needs $r$, which is what is being looked for — and the way round it is that the $r$ eigenvectors add up to something trivial. Since $r^{-1/2} \sum_s |u_s\rangle = |1\rangle$, starting the work register in the state one costs a single gate, and the circuit then samples one eigenphase $s/r$ with $s$ uniform on $0, \ldots, r-1$. The limitation of section 6.2 has become the preparation.




Powers of two modulo fifteen. The sequence returns to one after four steps and then repeats forever, so the order is four. Finding this length is the only quantum step in factoring.

Phase estimation needs $U_a^{2^j}$ for every counting qubit, and applying $U_a$ that many times would be exponential. It is not done that way: squaring the number instead of repeating the gate gives $U_a^{2^j} = U_{a^{2^j} \bmod N}$, and the constant $a^{2^j} \bmod N$ is computed classically before the circuit is built. Each counting qubit therefore controls one modular multiplication by a precomputed constant, and the whole middle of the circuit is a reversible modular exponentiation.




The two parts of the circuit against the size of the number being factored. A reversible multiplier on $L$-bit numbers costs about $L^2$ gates and there are about $2L$ of them, so the arithmetic is order $L^3$ where the transform is order $L^2$.

> "SHOR'S ALGORITHM IS THE QUANTUM FOURIER TRANSFORM" IS THE WRONG SUMMARY
>
> At two thousand bits the arithmetic is about two thousand times the transform. What makes the algorithm work is that modular exponentiation can be done coherently and reversibly, so the phase information survives it; what makes it expensive is that the same modular exponentiation has to be done coherently and reversibly. Chapter 4's ancillas and uncomputing are where that cost comes from.

The measurement returns an integer $y$, and $y/Q$ is close to some $s/r$ — but $Q$ is a power of two and $r$ is not, so it never equals it. What is needed is a way to find a fraction with a small denominator near a given number, and that is an old classical algorithm: repeatedly take the whole part and invert the remainder. The fractions this builds, the **convergents**, are the best approximations with small denominators, and the one wanted is the last whose denominator is below $N$.

WHY THE READING IS ENOUGH

$$\left|\frac{y}{Q}-\frac{s}{r}\right| \le \frac{1}{2Q}, \qquad Q > N^2 \implies \frac{s}{r} \text{ is the only such fraction}$$

The uniqueness needs $Q > N^2$, so the counting register carries about $2L$ qubits for an $L$-bit $N$. That is not a detail: it doubles the counting register and doubles the number of controlled modular multiplications, and it is where the qubit counts in published resource estimates come from. The candidate is then confirmed by computing $a^r \bmod N$ once, which is fast, and rejected if the answer is not one.

---

EXAMPLE 6.3 · ONE READING, ONE ORDER

GIVEN $N = 21$, $a = 2$, $t = 9$ so $Q = 512$, and a reading of $y = 427$.

FIND The order.

SOLUTION $427/512 = 0.833984$, whose expansion is $[0; 1, 5, 42, 2]$ with convergents $0/1$, $1/1$, $5/6$, $211/253$ and $427/512$. The last denominator below $21$ is $6$, so $r = 6$; and $2^6 = 64 = 3 \times 21 + 1$ confirms it.

CHECK The true fraction is $5/6 = 0.833333$, and the reading missed it by $0.00065$ — below $1/2Q = 0.00098$, exactly as the bound promises. The reading was never exact and did not need to be.

---

A run can fail in four ways, and all four are detected by arithmetic that costs nothing. The measurement can land on $s = 0$, which says nothing about $r$. The values $s$ and $r$ can share a factor, so the convergent is a proper divisor of $r$ and the check fails. The order can turn out odd, so there is no square root to take. And $a^{r/2}$ can be $-1$ modulo $N$, in which case both greatest common divisors come out trivial. The first two are repaired by running the circuit again; the last two by choosing a different $a$.

A FAILURE THAT ANNOUNCES ITSELF IS NOT A REAL PROBLEM

For $N = pq$ the probability that the order is even and its half power is not $-1$ is at least a half, so a constant number of attempts suffices and the repetition costs a factor rather than an exponent. What makes this tolerable is that success is **checkable**: multiply the candidate factors and see whether they give $N$. Compare Deutsch–Jozsa or Grover, where a wrong answer looks exactly like a right one and the confidence has to come from the analysis instead.

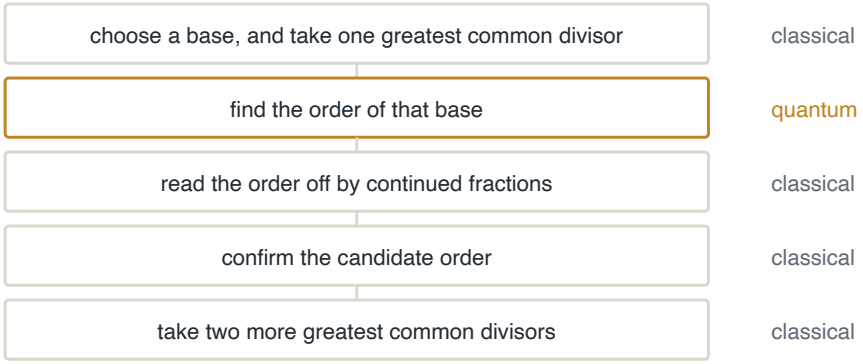
## 6.7 Factoring

The reduction from factoring to order finding is a page of school algebra. Suppose the order $r$ of $a$ is even. Then $a^r - 1 \equiv 0$, so the product of $a^{r/2} - 1$ and $a^{r/2} + 1$ is divisible by $N$ — and provided $a^{r/2} \not\equiv \pm 1$, neither factor is divisible by $N$ on its own. The prime factors of $N$ must therefore be split between the two.

THE FACTORS, FROM THE ORDER

$$\gcd\left(a^{r/2}-1,\,N\right) \quad \text{and} \quad \gcd\left(a^{r/2}+1,\,N\right)$$

Euclid's algorithm finds a greatest common divisor in about $\log N$ steps and it is the cheapest thing in the whole procedure. The algebra above turns "the number $N$ is a product" into "these two specific numbers share a factor with $N$", and the second question is easy. The quantum step exists only to supply the exponent that makes the algebra apply.




The five steps, with the quantum one marked. Four of them run on a laptop in microseconds; the whole cost of the algorithm is inside the one box that does not.

---

EXAMPLE 6.4 · FIFTEEN, AND THE BASE THAT FAILS

GIVEN $N = 15$ with $a = 2$, and then the same $N$ with $a = 14$.

FIND The factors in each case.

SOLUTION For $a = 2$: the powers are $1, 2, 4, 8, 1$, so $r = 4$; it is even and $2^2 = 4$ is not $-1$ modulo fifteen, so $\gcd(3, 15) = 3$ and $\gcd(5, 15) = 5$. For $a = 14$: $14^2 = 196 = 13 \times 15 + 1$, so $r = 2$, but $14 \equiv -1$, and both divisors come out trivial.

CHECK $3 \times 5 = 15$. And $a = 14$ is not a rare accident: of the eight numbers below fifteen coprime to it, two are useless — one because its order is odd and one for the reason above — so the procedure has to be ready to discard a base and pick another.

---

WHAT A SMALL DEMONSTRATION SHOWS

Fifteen has been factored on hardware many times, and the circuits used are almost always simplified using knowledge of the answer: a multiplier built for a known order is not the general circuit. A demonstration at this size shows that the pieces fit together. It is not evidence about what a large number would cost, and the difference between the two is about ten orders of magnitude.

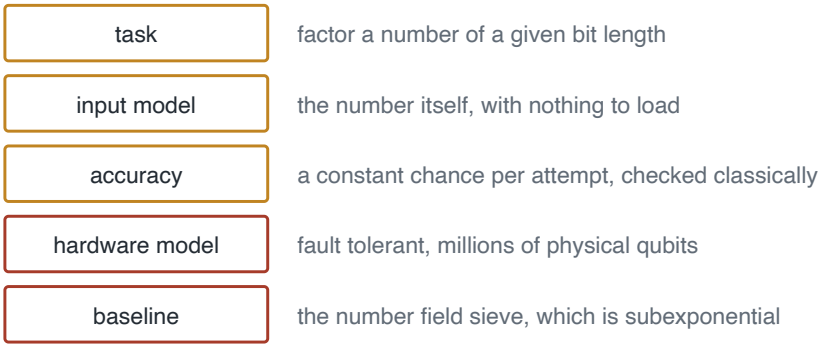
RSA publishes a modulus $N = pq$ and keeps the factorisation secret; anyone who can factor $N$ has the private key. Diffie–Hellman and elliptic-curve schemes rest on the discrete logarithm, which the same machinery solves. Symmetric ciphers and hash functions are in a different position: the best known quantum attack is Grover, a square root, and doubling the key length restores the margin. Their security was never a proof in either case — it is the absence of a known efficient attack at the sizes in use, and that is what changed when the attack became known and only the machine was missing.

THE MIGRATION DEADLINE IS SET BY THE DATA, NOT BY THE MACHINE

Traffic can be recorded today at negligible cost and decrypted whenever a capable machine exists. So the question for any particular secret is how long it has to hold, and a secret that must last twenty years is already exposed. A **post-quantum** scheme is a classical algorithm — it runs on ordinary computers — chosen because it rests on a problem for which no efficient quantum attack is known. It is not quantum cryptography, which is a different subject about distributing keys over physical channels.

## 6.8 What the claim says

Chapter 5 said a resource claim names five things. Written out for factoring, two of the five carry almost all the weight.




The claim against the five. The two in the error tone are the ones usually left unstated, and they are the ones that decide whether any of this happens on a real machine.

- **Hardware model.** The logical circuit is about $L^3$ gates, but every logical gate is a code block of many physical qubits with rounds of error correction. Published estimates for a two-thousand-bit modulus run to millions of physical qubits and hours of running time.
- **Baseline.** The general number field sieve takes about $e^{c\,L^{1/3}(\log L)^{2/3}}$ operations, which is subexponential and not exponential. So the gap is superpolynomial rather than exponential, and it is a gap against the best **known** classical algorithm.

WHAT THE RESULT DOES SAY, STATED FAIRLY

On an ideal fault-tolerant quantum computer, factoring takes a number of operations polynomial in the bit length, where every known classical method takes subexponentially many. That is a genuine and important separation and the strongest known statement of its kind. It is a statement about two algorithms, not about the difficulty of the problem: no theorem says factoring is hard, and a fast classical factoring algorithm would remove the separation without making any of the quantum mechanics wrong.

One last placing. Nothing in the order-finding circuit used the fact that the operation was multiplication; it used that some function repeats, with distinct values within one period. Given a circuit for such a function, the same construction returns the period. Order finding is the case $f(x) = a^x \bmod N$, the discrete logarithm is another case, and all of them are instances of one statement about hidden structure in a commutative group. That is where nearly all the known superpolynomial quantum speedups live. Grover is

the conspicuous exception, and it is an exception because unstructured search has no such structure to find — which is exactly why its saving is only a square root.

## 6.9 Summary

- Phase kickback is the one mechanism: $U_f|x\rangle|-\rangle = (-1)^{f(x)}|x\rangle|-\rangle$, and with any unitary, a controlled $U$ acting on one of its eigenstates puts $e^{2\pi i\varphi}$ on the control.
- A superposition costs one layer of Hadamards and buys nothing on its own. Every gain comes from making the unwanted amplitudes cancel before the measurement.
- Deutsch–Jozsa: the amplitude of $0^n$ is $2^{-n}\sum_x(-1)^{f(x)}$, which is $\pm1$ for a constant function and exactly $0$ for a balanced one. One query against $2^{n-1}+1$ exact classical queries, and against about $21$ randomised ones whatever $n$ is.
- The transform $F_Q|x\rangle = Q^{-1/2}\sum_k e^{2\pi ixk/Q}|k\rangle$ costs $\frac{1}{2}n(n+1)$ gates and returns one index, never a spectrum.
- Phase estimation costs $2^t-1$ applications of $U$ against $\frac{1}{2}t(t+1)$ gates in the inverse transform. A phase that is a $t$-bit fraction is read with certainty; otherwise the two nearest readings carry at least $8/\pi^2$, and $t = n + \lceil\log_2(2+1/2\varepsilon)\rceil$ buys $n$ bits at confidence $1-\varepsilon$.
- Order finding: $U_a|y\rangle = |ay \bmod N\rangle$ has eigenphases $s/r$, and the state one is their even mixture. Continued fractions recover $r$ from a reading, one modular exponentiation confirms it, and $\gcd(a^{r/2}\pm1, N)$ gives the factors.
- Modular exponentiation is order $L^3$ and the transform is order $L^2$, so the arithmetic is the algorithm's cost. Everything around order finding — the base, the continued fractions, the checks, the repeats — is classical.

> FOUR ERRORS THAT COST A WHOLE QUESTION
>
> Calling a query count a runtime. Quoting the Deutsch–Jozsa separation without the words "exact" and "promised". Saying that the Fourier transform returns the spectrum. And treating the classical work around order finding as the expensive part of factoring.

> WHERE THIS LEAVES THE THREE SENTENCES OF CHAPTER 0
>
> The state is large and the readout is small: every algorithm here reads $n$ bits and had to make everything else cancel first. A relative phase is everything: kickback writes the answer as one and the transform reads it back. And a resource claim names five things: written out in full, Deutsch–Jozsa is a promise problem, Grover is quadratic, and Shor is a gap against one classical algorithm rather than against a proof.

APPENDIX A

# Formula summary

*Every formula the course establishes, in the order the course establishes it. Nothing here is derived and nothing here is new: each entry names the section that develops it, and the argument is in that section. Read this page to check a result you already understand, never to meet one for the first time.*

Four conventions are in force throughout and are not repeated under each entry. A register of $n$ qubits is written $|q_{n-1}\dots q_1 q_0\rangle$, and entry $x$ of its column of amplitudes is the amplitude of $|x\rangle$ with $x$ read as a binary number. A phase on the whole state is not physical; a phase between two terms is. The inner product conjugates its first argument. And $\hbar = 1$, so a Hamiltonian is in angular frequency, evolution is $U(t) = e^{-iHt}$, and a logarithm of a probability is taken base two.

## A.1 Chapter 1 · The mathematics of quantum states

A QUBIT STATE

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle = \begin{bmatrix}\alpha\\ \beta\end{bmatrix}, \qquad |\alpha|^2 + |\beta|^2 = 1$$

The object every other formula acts on. Chapter 1, section 1.1.

THE INNER PRODUCT

$$\langle a| = |a\rangle^\dagger = \begin{bmatrix}a_1^* & \cdots & a_n^*\end{bmatrix}$$

$$\langle a|b\rangle = \sum_k a_k^*\, b_k$$

The conjugate sits on the first argument and on nothing else, so $\langle b|a\rangle = \langle a|b\rangle^*$. Chapter 1, section 1.1.

A COEFFICIENT IS AN INNER PRODUCT

$$|v\rangle = \sum_i v_i|e_i\rangle \qquad \Longrightarrow \qquad v_j = \langle e_j|v\rangle$$

True only for an orthonormal basis, and it is what an orthonormal basis is chosen for. Chapter 1, section 1.1.

CAUCHY–SCHWARZ, AND THE TWO ENDS OF IT

$$|\langle a|b\rangle| \le \|a\|\,\|b\|, \qquad 0 \le |\langle a|b\rangle|^2 \le 1 \ \text{ for normalised states}$$

Zero means orthogonal and one measurement separates the two with certainty; one means the same state up to a global phase. Chapter 1, section 1.1.

MODULUS AND PHASE

$$z = x + iy = r\,e^{i\varphi}, \qquad zz^{*} = x^{2} + y^{2} = |z|^{2}$$

Every probability in the course is a $|z|^2$ and every interference effect is a difference of two $\varphi$. Chapter 1, section 1.2.

A GLOBAL PHASE IS NOT PHYSICAL, A RELATIVE PHASE IS

$$e^{i\gamma}|\psi\rangle \;\equiv\; |\psi\rangle$$

$$\alpha|0\rangle + e^{i\varphi}\beta|1\rangle \;\not\equiv\; \alpha|0\rangle + \beta|1\rangle$$

The second line is the mechanism every algorithm in the course runs on. Chapter 1, section 1.2.

THE OUTER PRODUCT AND THE PROJECTOR

$$|a\rangle\langle b| = |a\rangle\,|b\rangle^{\dagger}, \qquad \big(|a\rangle\langle b|\big)_{jk} = a_j\,b_k^{*}$$

$$P_u = |u\rangle\langle u|, \qquad P_u^{\dagger} = P_u, \qquad P_u^{2} = P_u$$

A bra eats the ket to its right first, so $\big(|a\rangle\langle b|\big)\,|v\rangle = \langle b|v\rangle\,|a\rangle$. Chapter 1, section 1.3.

THE RESOLUTION OF THE IDENTITY

$$\sum_{k} |e_k\rangle\langle e_k| = I$$

Used as a move rather than as a fact: inserting it turns a statement about a vector into a statement about its coefficients. Chapter 1, section 1.3.

GRAM–SCHMIDT

$$u_j = v_j - \sum_{i<j} \langle e_i|v_j\rangle\, e_i, \qquad e_j = \frac{u_j}{\|u_j\|}$$

The modified recursion subtracts against the partly built $u_j$ rather than against $v_j$, and keeps far more digits. Chapter 1, section 1.4.

THE TENSOR PRODUCT

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \otimes \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} a_1 b_1 \\ a_1 b_2 \\ a_2 b_1 \\ a_2 b_2 \end{bmatrix}$$

$$(A \otimes B)_{(i,j)} = A_{ij} B, \qquad \dim = 2^{n} \;\text{ for } n \text{ qubits}$$

Two qubits are four numbers and not two plus two, and that is where the exponential comes from. Chapter 1, section 1.5.

ADJOINT, HERMITIAN, UNITARY

$$A^\dagger = (A^*)^T, \qquad (AB)^\dagger = B^\dagger A^\dagger$$

$$A = A^\dagger \implies \text{real eigenvalues}, \qquad U^\dagger U = I \iff U^{-1} = U^\dagger$$

A Hermitian operator is a measurable quantity and a unitary one is a reversible motion. Chapter 1, section 1.6.

THE EXPONENTIAL THAT JOINS THEM

$$G = G^\dagger,\ \theta \in \mathbb{R} \qquad \implies \qquad U(\theta) = e^{-i\theta G} \ \text{is unitary}$$

$$e^{-i\theta\sigma/2} = \cos\!\left(\tfrac{\theta}{2}\right) I \ - \ i \sin\!\left(\tfrac{\theta}{2}\right) \sigma$$

The second line holds for any operator whose square is the identity, which is every Pauli and every $\mathbf{n}\cdot\boldsymbol{\sigma}$. Chapter 1, section 1.6.

THE SPECTRAL DECOMPOSITION, AND A FUNCTION OF AN OPERATOR

$$A = \sum_k \lambda_k P_k, \qquad P_j P_k = \delta_{jk} P_k, \qquad \sum_k P_k = I$$

$$f(A) = \sum_k f(\lambda_k)\, P_k, \qquad e^{-iAt} = \sum_k e^{-i\lambda_k t}\, P_k$$

The second line is what makes an evolution operator computable at all. Chapter 1, section 1.7.

FUNCTIONS AS VECTORS, PARSEVAL AND TRUNCATION

$$\langle f|g\rangle = \int_a^b f^*(x) g(x)\,\mathrm{d}x, \qquad |f\rangle = \sum_{n=1}^{\infty} c_n |u_n\rangle, \qquad c_n = \langle u_n|f\rangle$$

$$\|f\|^2 = \sum_{n=1}^{\infty} |c_n|^2, \qquad \left\| f - \sum_{n=1}^{N} c_n u_n \right\|^2 = \sum_{n>N} |c_n|^2$$

The coefficient tail is the exact squared norm error of a truncated complete orthonormal expansion. Chapter 1, section 1.9.

## A.2 Chapter 2 · States, measurement and dynamics

THE BORN RULE

$$p(n) = |\langle n|\psi\rangle|^2$$

The one postulate that connects a state to a number an instrument prints. Chapter 2, section 2.1.

PASSIVE COORDINATES AGAINST AN ACTIVE TRANSFORMATION

$$|\psi\rangle' = V^\dagger|\psi\rangle, \qquad A' = V^\dagger A V, \qquad \langle\psi|A|\psi\rangle = \langle\psi'|A'|\psi'\rangle$$

$$|\psi\rangle \longmapsto U|\psi\rangle \qquad \text{with } A \text{ fixed}$$

Passive coordinates preserve every prediction; an active operation generally changes them. Choosing new measurement projectors is a third operation. Chapter 2, section 2.1.

A PROJECTIVE MEASUREMENT, AND THE STATE IT LEAVES

$$p(a) = \langle\psi|P_a|\psi\rangle, \qquad P_j P_k = \delta_{jk} P_k, \qquad \sum_a P_a = I$$

$$|\psi_a\rangle = \frac{P_a|\psi\rangle}{\sqrt{p(a)}}$$

Orthogonality makes the outcomes exclusive and completeness makes them add to one. Chapter 2, section 2.2.

A POVM AND ITS MEASUREMENT INSTRUMENT

$$E_m = \sum_\alpha M_{m\alpha}^\dagger M_{m\alpha}, \qquad p(m) = \mathrm{Tr}(\rho E_m), \qquad \sum_m E_m = I$$

$$\rho_m = \frac{\sum_\alpha M_{m\alpha}\rho M_{m\alpha}^\dagger}{p(m)}$$

The effects fix the probabilities; the instrument also fixes the conditional state. The projective case is $M_m = P_m$. Chapter 2, section 2.2.

EXPECTATION AND VARIANCE

$$\langle A\rangle = \langle\psi|A|\psi\rangle = \sum_a a\, p(a)$$

$$\mathrm{Var}(A) = \langle A^2\rangle - \langle A\rangle^2, \qquad \Delta A = \sqrt{\mathrm{Var}(A)}$$

For any Pauli $A^2 = I$, so $\mathrm{Var}(A) = 1 - \langle A\rangle^2$ and no second sandwich is needed. Chapter 2, section 2.3.

THE ROBERTSON RELATION

$$\Delta A\,\Delta B \;\geq\; \tfrac{1}{2}\,|\langle [A,B]\rangle|\,, \qquad [A,B] = AB - BA$$

The bound depends on the state, so a vanishing right-hand side says nothing rather than saying both spreads are zero. Chapter 2, section 2.4.

| OBSERVABLE | $+1$ EIGENSTATE | $-1$ EIGENSTATE |
|---|---|---|
| $X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ | $\lvert+\rangle = \frac{1}{\sqrt{2}}(1,1)$ | $\lvert-\rangle = \frac{1}{\sqrt{2}}(1,-1)$ |
| $Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$ | $\lvert+i\rangle = \frac{1}{\sqrt{2}}(1,i)$ | $\lvert-i\rangle = \frac{1}{\sqrt{2}}(1,-i)$ |
| $Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ | $\lvert 0\rangle = (1,0)$ | $\lvert 1\rangle = (0,1)$ |

THE PAULI ALGEBRA

$$\sigma_i\sigma_j = \delta_{ij}\,I + i\sum_k \varepsilon_{ijk}\,\sigma_k$$

$$[\sigma_i,\sigma_j] = 2i\sum_k \varepsilon_{ijk}\sigma_k, \qquad \{\sigma_i,\sigma_j\} = 2\delta_{ij}I$$

Cyclic order $X \to Y \to Z \to X$ carries the plus sign. Two different Pauli operators anticommute. Chapter 2, section 2.5.

MEASURING ALONG A DIRECTION

$$\mathbf{n}\cdot\boldsymbol{\sigma}, \qquad P_\pm = \tfrac{1}{2}\left(I \pm \mathbf{n}\cdot\boldsymbol{\sigma}\right), \qquad r_a = \langle\sigma_a\rangle$$

$$p(\pm) = \tfrac{1}{2}\left(1 \pm \mathbf{n}\cdot\mathbf{r}\right), \qquad p(+) = \cos^2\frac{\alpha}{2}$$

$\alpha$ is the angle between two vectors, which is twice the angle between the two states. Chapter 2, section 2.5.

COORDINATE OPERATORS AND A FREE PARTICLE

$$(\hat{x}\psi)(x) = x\psi(x), \qquad (\hat{p}\psi)(x) = -i\hbar\frac{\mathrm{d}\psi}{\mathrm{d}x}$$

$$\hat{H}_0 = \frac{\hat{p}^2}{2m} = -\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2}, \qquad p = \hbar k, \qquad E = \frac{\hbar^2 k^2}{2m}$$

A plane wave is a generalised eigenfunction; a physical free-particle state is a normalisable wave packet. Chapter 2, section 2.6.

THE INFINITE SQUARE WELL

$$\phi_n(x) = \sqrt{\frac{2}{a}} \sin\!\left(\frac{n\pi x}{a}\right), \qquad n = 1, 2, \dots$$

$$E_n = \frac{\hbar^2 \pi^2 n^2}{2ma^2}$$

The boundary conditions select discrete wave numbers. There is no $n = 0$ state, and $E_2 = 4E_1$. Chapter 2, section 2.6.

CLOSED-SYSTEM EVOLUTION

$$i\,\frac{\mathrm{d}}{\mathrm{d}t}|\psi(t)\rangle = H\,|\psi(t)\rangle$$

$$|\psi(t)\rangle = U(t)|\psi(0)\rangle, \qquad U(t) = e^{-iHt}$$

Replacing $H$ by $H + cI$ multiplies $U$ by a global phase, so only energy differences are physical. Chapter 2, section 2.6.

A DRIVEN QUBIT

$$H = \tfrac{1}{2}\left(\Omega_x X + \Omega_y Y + \Delta Z\right) = \tfrac{\Omega}{2}\,\mathbf{n}\cdot\boldsymbol{\sigma}, \qquad \Omega = \sqrt{\Omega_x^2 + \Omega_y^2 + \Delta^2}$$

$$U(t) = \cos\!\left(\tfrac{\Omega t}{2}\right) I - i \sin\!\left(\tfrac{\Omega t}{2}\right) \mathbf{n}\cdot\boldsymbol{\sigma}, \qquad p_{\max} = \left(\frac{\Omega_x}{\Omega}\right)^2$$

Pulse area sets the angle, drive phase sets the axis in the equator, and detuning tilts that axis towards $z$. Chapter 2, section 2.6.

WHAT A FINITE RUN IS WORTH

$$K \sim \mathrm{Binomial}(N, p), \qquad \mathrm{SE}\!\left(\tfrac{K}{N}\right) = \sqrt{\frac{p(1-p)}{N}} \;\le\; \frac{1}{2\sqrt{N}}$$

One more decimal place costs a hundred times the shots. This is counting and not physics, so no hardware improvement changes it. Chapter 2, section 2.7.

## A.3 Chapter 3 · Mixed states and entanglement

THE DENSITY OPERATOR

$$\rho_\psi = |\psi\rangle\langle\psi|$$

$$\rho = \sum_i p_i \, |\psi_i\rangle\langle\psi_i|, \qquad p_i \geq 0, \qquad \sum_i p_i = 1$$

It does not remember which preparation made it: two different mixtures with the same $\rho$ are the same state. Chapter 3, section 3.1.

WHICH MATRICES ARE STATES, AND EVERY PREDICTION FROM ONE TRACE

$$\rho = \rho^\dagger, \qquad \rho \succeq 0, \qquad \operatorname{Tr}\rho = 1$$

$$\langle A \rangle = \operatorname{Tr}\left(\rho A\right), \qquad p(m) = \operatorname{Tr}\left(\rho E_m\right)$$

For one qubit positivity is the single inequality $|\rho_{01}|^2 \leq \rho_{00}\rho_{11}$. Chapter 3, section 3.1.

PURITY

$$\gamma = \operatorname{Tr}\left(\rho^2\right) = \sum_k \lambda_k^2, \qquad \frac{1}{d} \leq \gamma \leq 1$$

It says how mixed and never which state: one number where a qubit state needs three. Chapter 3, section 3.2.

A QUBIT STATE AS A VECTOR

$$\rho = \tfrac{1}{2}\left(I + \mathbf{r}\cdot\boldsymbol{\sigma}\right), \qquad r_a = \operatorname{Tr}\left(\rho\,\sigma_a\right), \qquad |\mathbf{r}| \leq 1$$

$$\lambda_\pm = \tfrac{1}{2}\left(1 \pm |\mathbf{r}|\right), \qquad \operatorname{Tr}\rho^2 = \tfrac{1}{2}\left(1 + |\mathbf{r}|^2\right)$$

From a state vector $(a, b)$ the $y$ component reads $r_y = 2\,\mathrm{Im}(a^* b)$; from $\rho$ it reads $r_y = -2\,\mathrm{Im}\,\rho_{01}$. The two signs are not a contradiction, and both give $r_y = +1$ on $|{+i}\rangle$. Chapter 3, section 3.2.

A CHANNEL IN KRAUS FORM

$$\mathcal{E}(\rho) = \sum_k K_k \, \rho \, K_k^\dagger, \qquad \sum_k K_k^\dagger K_k = I$$

Every channel arises this way, and two different sets of Kraus operators can describe one channel. Chapter 3, section 3.3.

AMPLITUDE DAMPING

$$K_0 = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{bmatrix}, \qquad K_1 = \begin{bmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{bmatrix}$$

$$\rho_{11} \mapsto (1-\gamma)\rho_{11}, \qquad \rho_{01} \mapsto \sqrt{1-\gamma}\,\rho_{01}$$

At $\gamma = 1$ every state becomes $|0\rangle$, which is why this is the model of relaxation. Chapter 3, section 3.3.

THE PHASE-FLIP CHANNEL

$$\mathcal{E}_Z(\rho) = (1-p)\,\rho + p\,Z\rho Z$$

$$\rho_{00}, \rho_{11} \text{ unchanged}, \qquad \rho_{01} \mapsto (1-2p)\,\rho_{01}$$

The damage peaks at $p = \frac{1}{2}$ and vanishes at $p = 1$, where the channel is the gate $Z$. Chapter 3, section 3.3.

THE TWO DECAY TIMES

$$\rho_{11}(t) = \rho_{11}^{\text{eq}} + [\rho_{11}(0) - \rho_{11}^{\text{eq}}]\,e^{-t/T_1}, \qquad \rho_{01}(t) = e^{-t/T_2}\rho_{01}(0)$$

$$\frac{1}{T_2} = \frac{1}{2T_1} + \frac{1}{T_\phi} \qquad \Longrightarrow \qquad T_2 \le 2T_1$$

Populations decay on $T_1$ and the coherence on $T_2$, and the inequality is a theorem rather than a measurement. Chapter 3, section 3.4.

THE ORDERING THIS COURSE FIXES

$$\begin{bmatrix} a \\ b \end{bmatrix} \otimes \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} ac \\ ad \\ bc \\ bd \end{bmatrix}, \qquad |q_1 q_0\rangle$$

Read under the other convention this names a different state, and every number downstream is quietly wrong. Chapter 3, section 3.5.

THE PARTIAL TRACE

$$\rho_A = \mathrm{Tr}_B\left(\rho_{AB}\right) = \sum_j \left(I_A \otimes \langle j|\right) \rho_{AB} \left(I_A \otimes |j\rangle\right)$$

$$\mathrm{Tr}\left(\rho_A A\right) = \mathrm{Tr}\left[\rho_{AB}\left(A \otimes I_B\right)\right] \quad \text{for every } A$$

The second line is the property that makes it the right operation: it is the state that answers every question about $A$ alone. Chapter 3, section 3.5.

SEPARABILITY, AND THE SCHMIDT FORM

$$|\psi\rangle \text{ is a product } \iff c_0c_3 - c_1c_2 = 0$$

$$|\psi\rangle_{AB} = \sum_{k=1}^{r} \sqrt{\lambda_k}\,|u_k\rangle_A |v_k\rangle_B, \qquad \lambda_k > 0, \quad \sum_k \lambda_k = 1$$

The Schmidt rank $r$ is one exactly when the pair is a product. Chapter 3, section 3.6.

ENTANGLEMENT ENTROPY

$$S(\rho) = -\operatorname{Tr}(\rho \log_2 \rho) = -\sum_k \lambda_k \log_2 \lambda_k, \qquad S(\rho_A) = S(\rho_B)$$

Zero for a product and one bit for a Bell state. Chapter 3, section 3.6.

THE BELL STATES

$$|\Phi^{\pm}\rangle = \frac{|00\rangle \pm |11\rangle}{\sqrt{2}}, \qquad |\Psi^{\pm}\rangle = \frac{|01\rangle \pm |10\rangle}{\sqrt{2}}$$

$$\text{for } |\Phi^{+}\rangle: \quad \langle X \otimes X\rangle = 1, \qquad \langle Y \otimes Y\rangle = -1, \qquad \langle Z \otimes Z\rangle = 1$$

All four share the same reduced state $I/2$ and differ only in their joint correlations. Chapter 3, section 3.7.

CHSH, AND THE TWO CEILINGS

$$S = \langle A_0 B_0\rangle + \langle A_0 B_1\rangle + \langle A_1 B_0\rangle - \langle A_1 B_1\rangle$$

$$|S| \le 2 \text{ for pre-existing values}, \qquad |S| \le 2\sqrt{2} \text{ for quantum states}$$

Both bounds are theorems. A measured value above two is what rules out the first model. Chapter 3, section 3.7.

## A.4 Chapter 4 · The Bloch sphere and quantum gates

A PURE QUBIT STATE, AND THE POINT IT NAMES

$$|\psi(\theta,\varphi)\rangle = \cos\frac{\theta}{2}\,|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}\,|1\rangle, \qquad 0 \le \theta \le \pi, \quad 0 \le \varphi < 2\pi$$

$$\mathbf{r} = (\sin\theta\cos\varphi,\ \sin\theta\sin\varphi,\ \cos\theta)\,, \qquad |\mathbf{r}| = 1$$

At a pole the azimuth means nothing, because $\sin 0 = 0$ leaves nothing for the phase to multiply. Chapter 4, section 4.1.

THE OVERLAP OF TWO PURE STATES

$$|\langle\chi|\psi\rangle|^2 = \frac{1+\mathbf{r}\cdot\mathbf{s}}{2} = \cos^2\frac{\Theta}{2}$$

Orthogonal states sit at opposite points; points at a right angle overlap with probability one half. Chapter 4, section 4.1.

A FULL TURN IS NOT THE IDENTITY

$$R_{\mathbf{n}}(2\pi) = -I, \qquad R_{\mathbf{n}}(4\pi) = +I$$

$U$ and $-U$ give the same rotation of the sphere, so the map from gates to rotations is two to one. Chapter 4, section 4.2.

THE ROTATION OPERATOR

$$R_{\mathbf{n}}(\alpha) = e^{-i\alpha\,\mathbf{n}\cdot\boldsymbol{\sigma}/2} = \cos\frac{\alpha}{2}\,I - i\sin\frac{\alpha}{2}\,\mathbf{n}\cdot\boldsymbol{\sigma}$$

$$\text{every } 2\times 2 \text{ unitary} = e^{i\gamma}R_{\mathbf{n}}(\alpha) \ \text{ for some } \gamma,\ \mathbf{n},\ \alpha$$

Rotation is not one family of one-qubit gates among many; it is all of them. Chapter 4, section 4.3.

THE PAULI GATES AS ROTATIONS

$$R_x(\pi) = -iX, \qquad R_y(\pi) = -iY, \qquad R_z(\pi) = -iZ$$

$$\begin{aligned} X:&\ |0\rangle \leftrightarrow |1\rangle, && |+\rangle \text{ and } |-\rangle \text{ fixed} \\ Z:&\ |+\rangle \leftrightarrow |-\rangle, && |0\rangle \text{ and } |1\rangle \text{ fixed} \end{aligned}$$

Which gate looks like a flip depends entirely on the basis the state is written in. Chapter 4, section 4.3.

THE HADAMARD

$$H = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix} = \frac{X+Z}{\sqrt{2}}, \qquad H^2 = I$$

$$HXH = Z, \qquad HZH = X, \qquad HYH = -Y$$

Measuring $X$ is applying $H$ and then measuring $Z$, which is the only reading a machine offers. Chapter 4, section 4.3.

THE PHASE FAMILY

$$P(\varphi) = \begin{bmatrix}1 & 0\\ 0 & e^{i\varphi}\end{bmatrix} = e^{i\varphi/2}\,R_z(\varphi)$$

$$S = P\left(\tfrac{\pi}{2}\right), \qquad T = P\left(\tfrac{\pi}{4}\right), \qquad S^2 = Z, \quad T^2 = S$$

$S$ is a Clifford gate and cheap to correct for; $T$ is not, and that difference sets the cost of fault tolerance. Chapter 4, section 4.3.

THE ORDER A CIRCUIT MULTIPLIES IN

$$U_1 \text{ first, } U_2 \text{ next, } U_3 \text{ last} \qquad \Longrightarrow \qquad U = U_3\,U_2\,U_1$$

$$(A \otimes I)\,(I \otimes B) = A \otimes B = (I \otimes B)\,(A \otimes I)$$

Gates on different qubits always commute; gates sharing a qubit usually do not. Chapter 4, section 4.4.

THE EULER DECOMPOSITION, AND THE GATE WITH THREE DIALS

$$U = e^{i\alpha}\,R_z(\phi)\,R_y(\theta)\,R_z(\lambda)$$

$$U(\theta,\phi,\lambda) = \begin{bmatrix}\cos\frac{\theta}{2} & -e^{i\lambda}\sin\frac{\theta}{2}\\ e^{i\phi}\sin\frac{\theta}{2} & e^{i(\phi+\lambda)}\cos\frac{\theta}{2}\end{bmatrix}$$

$H = U(\frac{\pi}{2}, 0, \pi)$, $X = U(\pi, 0, \pi)$ and $P(\varphi) = U(0, \varphi, 0)$. Chapter 4, section 4.4.

REVERSIBLE EMBEDDINGS AND THE ORACLE PATTERN

$$(a,\ b) \longmapsto (a,\ a \oplus b) \qquad \text{is exactly CNOT}$$

$$|x\rangle|y\rangle \longmapsto |x\rangle\,|y \oplus f(x)\rangle \qquad \text{for any } f$$

AND needs a third wire, $(a, b, c) \mapsto (a, b, c \oplus ab)$, which is the Toffoli gate. Chapter 4, section 4.5.

A REVERSIBLE HALF ADDER

$$(a, b, 0, 0) \longmapsto (a, b, a \oplus b, ab)$$

Two CNOTs write the sum and one Toffoli writes the carry. Preserving the inputs makes the map invertible. Chapter 4, section 4.5.

COMPUTE, COPY OUT, UNCOMPUTE

$$V_f^\dagger \ (\text{copy}) \ V_f \ : \ |x\rangle|0\rangle|y\rangle \ \longmapsto \ |x\rangle|0\rangle|y \oplus f(x)\rangle$$

An ancilla that leaves in a state depending on $x$ stays entangled with the register, and the interference at the end does not happen. Chapter 4, section 4.5.

A ONE-QUBIT GATE ON A NAMED QUBIT

$$\text{on } q_0 : \ I \otimes A, \qquad \text{on } q_1 : \ A \otimes I$$

$$(I \otimes X)\,|10\rangle = |11\rangle, \qquad (X \otimes I)\,|10\rangle = |00\rangle$$

Both are valid gates and they do different things. Nothing in the mathematics protects a reader who picks the wrong one. Chapter 4, section 4.6.

THE TWO CNOT MATRICES, IN THE BASIS ORDER $|00\rangle, |01\rangle, |10\rangle, |11\rangle$

$$\text{CNOT} : \ |c\rangle|t\rangle \longmapsto |c\rangle|c \oplus t\rangle$$

$$\text{CNOT}_{0\to1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \qquad \text{CNOT}_{1\to0} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

The subscript reads control to target. Chapter 4, section 4.6.

CZ AND SWAP

$$\text{CZ} = \text{diag}(1, 1, 1, -1) = (H \otimes I)\ \text{CNOT}_{0\to1}\ (H \otimes I)$$

$$\text{SWAP} = \text{CNOT}_{0\to1}\ \text{CNOT}_{1\to0}\ \text{CNOT}_{0\to1}$$

Three two-qubit gates to move a state one step across a chip, which is what routing costs. Chapter 4, section 4.6.

MAKING A BELL PAIR, AND TWO UNIVERSAL SETS

$$|00\rangle \xrightarrow{I\otimes R_y(\theta)} \cos\tfrac{\theta}{2}|00\rangle + \sin\tfrac{\theta}{2}|01\rangle \xrightarrow{\text{CNOT}_{0\to1}} \cos\tfrac{\theta}{2}|00\rangle + \sin\tfrac{\theta}{2}|11\rangle$$

$$\{\text{every one-qubit gate}\} \cup \{\text{CNOT}\} \ \text{exactly}, \qquad \{H,\ S,\ T,\ \text{CNOT}\} \ \text{to any } \varepsilon > 0$$

A gate is entangling if it entangles some input, not every input. The price of $\varepsilon$ is a length growing like a power of $\log(1/\varepsilon)$. Chapter 4, section 4.7.

## A.5 Chapter 5 · Circuits and protocols

THE CIRCUIT MODEL, IN ONE LINE

$$|0\rangle^{\otimes n} \xrightarrow{U_d \cdots U_2 U_1} |\psi\rangle \xrightarrow{\text{measure}} x \in \{0,1\}^n$$

Fixed input, fixed gate set, fixed measurement basis, and one string per run. Chapter 5, section 5.1.

ONE STATE, NAMED FOUR WAYS

$$|q_{n-1} \ \cdots \ q_1 q_0\rangle, \qquad x = \sum_k 2^k q_k$$

A run that prints $101$ read $q_2 = 1$, $q_1 = 0$, $q_0 = 1$, which is entry $5$ of the state vector. Chapter 5, section 5.1.

DEPTH, GATE COUNT AND SIMULATION COST

$$\text{depth} = \text{layers of gates that must run in order}, \qquad \text{count} = \text{gates}$$

$$\text{bytes} = 16 \cdot 2^n, \qquad n = 30 \Rightarrow 17\,\text{GB}, \qquad n = 50 \Rightarrow 18\,\text{PB}$$

A coherence time is spent against depth and an error budget against the two-qubit count. The size of the vector is an upper bound on the difficulty of simulation and never a lower one. Chapter 5, sections 5.1 and 5.2.

THE ERROR BAR ON A PROBABILITY READ FROM A FINITE RUN

$$K \sim \text{Binomial}(N, p), \qquad \text{SE}(\hat{p}) = \sqrt{\frac{p(1-p)}{N}}$$

An estimate near one half to within $\pm 0.005$ needs ten thousand shots. Chapter 5, section 5.2.

DEFERRED MEASUREMENT

$$\text{measure } q_0, \text{ then } X^m \text{ on } q_1 \ \equiv \ \text{CNOT}_{0\to 1}, \text{ then measure } q_0$$

The rule is about a measured wire used as a control and about nothing else. Chapter 5, section 5.2.

TWO EXACT REWRITES A COMPILER USES

$$H = e^{i\pi/2} R_z\left(\tfrac{\pi}{2}\right) R_x\left(\tfrac{\pi}{2}\right) R_z\left(\tfrac{\pi}{2}\right)$$

$$\text{CNOT}_{0\to 1} = (H \text{ on } q_1) \ \text{CZ} \ (H \text{ on } q_1)$$

What is lost is only length; the two-qubit count does not move. The phase in the first line becomes observable the moment the gate is controlled. Chapter 5, section 5.3.

THE FAULT-TOLERANCE THRESHOLD CONDITION

$$p < p_{\text{th}}$$

For a stated code, noise model and control system, increasing the code distance can then reduce the logical error rate. The threshold is not one universal number. Chapter 5, section 5.3.

TURNING A PHASE INTO COUNTS

$$|0\rangle \xrightarrow{H} \tfrac{1}{\sqrt{2}}\left(|0\rangle + |1\rangle\right) \xrightarrow{P(\varphi)} \tfrac{1}{\sqrt{2}}\left(|0\rangle + e^{i\varphi}|1\rangle\right)$$

$$\xrightarrow{H} \tfrac{1}{2}\left(1 + e^{i\varphi}\right)|0\rangle + \tfrac{1}{2}\left(1 - e^{i\varphi}\right)|1\rangle, \qquad p(0) = \cos^2\frac{\varphi}{2}$$

Measuring in the middle learns nothing; the second Hadamard is what makes the phase readable. Chapter 5, section 5.4.

TELEPORTATION, AS ONE IDENTITY

$$|\Psi_2\rangle = \tfrac{1}{2}\sum_{m_1,m_0\in\{0,1\}} |m_1 m_0\rangle_{10} \otimes X^{m_1}Z^{m_0}|\psi\rangle_2$$

Exact, and independent of $\alpha$ and $\beta$. Bob applies $Z^{m_0}$ and then $X^{m_1}$ to undo it. Chapter 5, section 5.5.

WHAT BOB HOLDS BEFORE THE BITS ARRIVE

$$\rho_B = \tfrac{1}{4}\sum_{m_1,m_0} X^{m_1}Z^{m_0}\,\rho\,Z^{m_0}X^{m_1} = \frac{I}{2}$$

No $\rho$ on the right-hand side, which is why the protocol sends nothing ahead of the two classical bits. Chapter 5, section 5.5.

PHASE KICKBACK

$$X|-\rangle = -|-\rangle$$

$$U_f\,|x\rangle|-\rangle = (-1)^{f(x)}\,|x\rangle|-\rangle$$

The target comes out unchanged, so the answer has landed on the first register as a sign. Chapter 5, section 5.6, and again in chapter 6.

THE PLANE GROVER SEARCH LIVES IN

$$|G\rangle = \frac{1}{\sqrt{M}} \sum_{f(x)=1} |x\rangle, \qquad |B\rangle = \frac{1}{\sqrt{N-M}} \sum_{f(x)=0} |x\rangle$$

$$|s\rangle = \sin\theta\,|G\rangle + \cos\theta\,|B\rangle, \qquad \sin\theta = \sqrt{\frac{M}{N}}$$

A real two-dimensional plane inside a complex space of $2^n$ dimensions. It is not a Bloch sphere. Chapter 5, section 5.6.

THE GROVER SUCCESS PROBABILITY, AND WHERE IT PEAKS

$$P_{\text{good}}(r) = \sin^2\big((2r+1)\theta\big)$$

$$r_* = \frac{\pi}{4\theta} - \frac{1}{2}, \qquad r_* \approx \frac{\pi}{4}\sqrt{\frac{N}{M}} \ \text{ when } M \ll N$$

Each iteration adds a fixed angle $2\theta$, and a right angle has to be reached. Iterating past $r_*$ makes the answer less likely. Chapter 5, section 5.6.

WHAT A RESOURCE CLAIM MUST NAME

The task, the input model, the accuracy, the hardware model, and the classical baseline it is measured against. A claim missing one of the five is not yet a claim, and a query count is not a runtime. Chapter 5, section 5.7.

## A.6 Chapter 6 · Quantum algorithms

THE ORACLE, AND WHAT ONE QUERY MEANS

$$U_f\,|x\rangle|y\rangle = |x\rangle\,|y \oplus f(x)\rangle$$

Building $U_f$ costs gates and those gates are not counted. A query separation is a theorem about one column of the table. Chapter 6, section 6.1.

KICKBACK WITH AN ARBITRARY UNITARY

$$U\,|u\rangle = e^{2\pi i\varphi}\,|u\rangle$$

$$\mathrm{c}U\,\tfrac{1}{\sqrt{2}}\big(|0\rangle + |1\rangle\big)|u\rangle = \tfrac{1}{\sqrt{2}}\big(|0\rangle + e^{2\pi i\varphi}|1\rangle\big)|u\rangle$$

On a superposition of eigenstates the procedure samples one eigenphase with probability $|c_k|^2$; it does not list the spectrum. Chapter 6, section 6.2.

WHAT THE LAST LAYER OF HADAMARDS IS FOR

$$\langle 0^n|\,H^{\otimes n}\Big(\tfrac{1}{2^{n/2}}\textstyle\sum_x (-1)^{f(x)}|x\rangle\Big) = \tfrac{1}{2^n}\textstyle\sum_x (-1)^{f(x)}$$

A superposition that never interferes has bought nothing. The readout returns $n$ bits, never $2^n$ numbers. Chapter 6, section 6.2.

DEUTSCH, AND DEUTSCH–JOZSA

$$\tfrac{1}{\sqrt{2}}\Big((-1)^{f(0)}|0\rangle + (-1)^{f(1)}|1\rangle\Big) \xrightarrow{\;H\;} \pm|\,f(0)\oplus f(1)\,\rangle$$

$$\text{amplitude of } |0^n\rangle = \frac{1}{2^n}\sum_x (-1)^{f(x)} = \begin{cases} \pm 1 & \text{constant} \\ 0 & \text{balanced} \end{cases}$$

One query against $2^{n-1}+1$ exact classical queries, on a promised input. Chapter 6, section 6.3.

THE QUANTUM FOURIER TRANSFORM, AND ITS CIRCUIT

$$F_Q\,|x\rangle = \frac{1}{\sqrt{Q}}\sum_{k=0}^{Q-1} e^{2\pi i\,xk/Q}\,|k\rangle, \qquad Q = 2^n$$

$$R_k = \begin{bmatrix} 1 & 0 \\ 0 & e^{2\pi i/2^k} \end{bmatrix}, \qquad n + \tfrac{1}{2}n(n-1) = \tfrac{1}{2}n(n+1) \text{ gates}$$

It is a change of basis and returns one index, never a spectrum. $F_2$ is the Hadamard. Chapter 6, section 6.4.

PHASE ESTIMATION

$$\frac{1}{\sqrt{2^t}}\sum_{k=0}^{2^t-1} e^{2\pi i k\varphi}\,|k\rangle \xrightarrow{F^\dagger} y, \qquad \hat{\varphi}=\frac{y}{2^t}$$

$$P(y)=\frac{1}{2^{2t}}\left|\frac{\sin\!\left(\pi\,2^t\delta\right)}{\sin(\pi\delta)}\right|^2, \qquad \delta=\varphi-\frac{y}{2^t}$$

Costs $2^t-1$ applications of $U$ against $\frac{1}{2}t(t+1)$ gates in the inverse transform. Chapter 6, section 6.5.

THE TWO FLOORS, AND HOW MANY COUNTING QUBITS TO TAKE

$$P(\text{nearest})\geq\frac{4}{\pi^2}\approx 0.405, \qquad P(\text{two nearest})\geq\frac{8}{\pi^2}\approx 0.811$$

$$t=n+\left\lceil\log_2\!\left(2+\frac{1}{2\varepsilon}\right)\right\rceil$$

Neither floor moves as the register grows: extra qubits buy accuracy and never confidence. Chapter 6, section 6.5.

ORDER FINDING

$$U_a\,|y\rangle=|\,a\,y \bmod N\,\rangle, \qquad U_a\,|u_s\rangle=e^{2\pi i s/r}\,|u_s\rangle$$

$$\frac{1}{\sqrt{r}}\sum_{s=0}^{r-1}|u_s\rangle=|1\rangle, \qquad |u_s\rangle=\frac{1}{\sqrt{r}}\sum_{k=0}^{r-1}e^{-2\pi i s k/r}\,|a^k \bmod N\rangle$$

The order sits in the denominators of the eigenphases, and starting the work register in the state one costs a single gate. Chapter 6, section 6.6.

WHY THE READING IS ENOUGH

$$\left|\frac{y}{Q}-\frac{s}{r}\right|\leq\frac{1}{2Q}, \qquad Q>N^2 \implies \frac{s}{r}\ \text{is the only such fraction}$$

Continued fractions recover $s/r$ from the reading, and one modular exponentiation confirms $r$. The condition $Q>N^2$ is what doubles the counting register. Chapter 6, section 6.6.

THE FACTORS, FROM THE ORDER

$$\gcd\left(a^{r/2}-1,\,N\right) \quad\text{and}\quad \gcd\left(a^{r/2}+1,\,N\right)$$

Needs $r$ even and $a^{r/2}\not\equiv -1 \bmod N$; a random $a$ satisfies both at least half the time. Chapter 6, section 6.7.

WHERE THE WORK IS

$$\text{modular exponentiation } O(L^3), \qquad \text{transform } O(L^2), \qquad \text{everything else classical}$$

The arithmetic is the algorithm, and the quantum step exists only to supply the exponent that makes the algebra apply. Chapter 6, section 6.7.