Summary of Probability Theory

- States in a Probability Theory are unit vectors in 1-norm that lie in the nonnegative orthant of a vector space with dimension equal to the number of possible events.

- The event space of a composite system is the cartesian product of event spaces for the component subsystem. The state of a composite system is a tensor product of states on the component subsystems iff if the component subsystems are independent / uncorrelated.

- The observables in a probability theory can only be measured in expectation. We can use independent samples to quickly estimate expectation values using Hoeffding's theorem.

- There are many notions of distance between states in a probability theory. Arguably the most operationally meaningful is the total variation distance, which bounds the difference in expectations of observables.

Generalized Probability Theory

- In wave theory, quantities such as energy are proportional to the square of the amplitude of a wave, which is called the magnitude.
- Interference occurs because we add wave forms with real-valued amplitudes, and then square the result:

$$\left[\sin\left(\frac{\pi x}{2}\right) + \sin\left(\frac{\pi x}{4}\right)\right]^2 \neq \sin\left(\frac{\pi x}{2}\right)^2 + \sin\left(\frac{\pi x}{4}\right)^2$$

- As in a probability theory, quantum theory considers a vector space of dimension equal to the number of possible events. Each event is now associated with a real or complex valued amplitude. The squared magnitude of the amplitudes represent probabilities, and so the states are unit vectors in the 2-norm.

Quantum states: $\boldsymbol{\alpha} = (\alpha_0, \ldots, \alpha_{2^n-1}) \in \mathbb{C}^{2^n}$ $\quad \displaystyle\sum_{i=0}^{2^n-1} |\alpha_i|^2 = 1$

e.g. quantum state on the space of *n* “quantum coins” or up/down quantum spins

Closed under any linear combinations that are normalized in the 2-norm,

$$z_1\boldsymbol{\alpha}_1 + z_2\boldsymbol{\alpha}_2\ ,\ z_1, z_2 \in \mathbb{C}\ ,\ |z_1|^2 + |z_2|^2 = 1$$

Generalized Probability Theory

- As before, we may associate outcomes of flipping our n quantum coins with the events space of n bit strings.

- We may form observables just as in a probability theory, by considering a function *f* defined on the space of events corresponding to outcomes of these n bit strings,

$$f : \{0,1\}^n \to \mathbb{R} \quad , \quad \langle f \rangle_{\boldsymbol{\alpha}} = \sum_{i=0}^{2^n-1} f(i)|\alpha_i|^2$$

- However, a natural notion for the microscopic world is that the act of observation itself becomes likely to disturb the system. A growing sense developed that certain observable quantities in quantum systems were incompatible, like position and momentum of particles in space, or spin along X , Y , Z directions.

- From the perspective of probability theory, we may assign a distribution to position outcomes, and to momentum outcomes, but the corresponding event spaces are fundamentally distinct and incompatible.

- Heisenberg realized that this could be accounted for in the generalized probability theory by associating an incompatible event space with a different basis!

Generalized Probability Theory

- From this notion of a basis as an event space, we can now think of any physically observable quantity as a function on the appropriate event space (position, momentum, spin direction, etc).
- If an observable is a basis together with a real-valued function, then we could say the observable is a Hermitian operator. The eigenvectors of the operator form the basis of events, and the eigenvalues of the operator associate the value of the observable that we associate with each event.
- If we have a quantum state and want to measure an observable then we project our state onto the appropriate basis, yielding complex amplitudes given by inner products, and then compute the expectation with respect to the resulting (squared magnitude) probability distribution.
- The idea that different observables have different event spaces (bases) motivates a basis-independent perspective of the state itself, we can view it as an abstract member of a vector space. Once we choose a basis, then the coordinates of the vector in that basis are called the **wave function**.

$$|\psi\rangle \in \mathcal{V}$$

Abstract Vector

$$\psi_i = \langle i|\psi\rangle$$

Inner Product

$$|\psi\rangle = \sum_{i\in\{0,1\}^n} \psi_i|i\rangle$$

Decomposition in particular basis

$$\boldsymbol{\psi} = (\psi_0, \ldots, \psi_{2^n-1}) \quad ,$$

Wave function

Generalized Probability Theory

- Unit vectors in the 2-norm, projections of vectors defined by inner products, and observables represented by eigenvalues of Hermitian operators. All this structure belongs to a **complex inner product space**.
- An inner product space with the additional property of being complete is called a **Hilbert space**.

(A) A vector space $\mathcal{V}$ over the field of complex numbers $\mathbb{C}$. Vectors denote $|\psi\rangle \in \mathcal{V}$ (Dirac's ket notation)

$$|\psi_1\rangle, |\psi_2\rangle \in \mathcal{V}, \alpha, \beta \in \mathbb{C} \implies \alpha|\psi_1\rangle + \beta|\psi_2\rangle \in \mathcal{V}$$

(B) An inner product $\langle\psi_1|\psi_2\rangle$ defined for all vectors $|\psi_1\rangle, |\psi_2\rangle \in \mathcal{V}$ which satisfies

Positivity: $\langle\psi_1|\psi_1\rangle > 0$

Linearity: $\langle\phi|\left(\alpha|\psi_1\rangle + \beta|\psi_2\rangle\right) = \alpha\langle\phi|\psi_1\rangle + \beta\langle\phi|\psi_2\rangle$

Skew-Symmetry: $\langle\psi_1|\psi_2\rangle = \langle\psi_2|\psi_1\rangle^*$

Complete in the 2-norm: $\||\psi\rangle\|_2 = \sqrt{\langle\psi|\psi\rangle}$

- **Axiom of States**. Quantum states are rays in Hilbert space. A ray is an equivalence class $\{\alpha|\psi\rangle : \alpha \in \mathbb{C}\}$ of scalar multiples of a state. This notion appears because of normalization $\langle\psi|\psi\rangle = 1$, which is equivalent to the squared magnitude of the components in any choice of basis summing to 1.

Generalized Probability Theory

- **Axiom of States**. Quantum states are rays in Hilbert space. A ray is an equivalence class $\{\alpha|\psi\rangle : \alpha \in \mathbb{C}\}$ of scalar multiples of a state. This notion appears because of normalization $\langle\psi|\psi\rangle = 1$, which is equivalent to the squared magnitude of the components in any choice of basis summing to 1.

- Every N-dimensional Hilbert space is isomorphic to $\mathbb{C}^N$. We are free to choose the representation of a vectors, and so we choose the basis vector $|i\rangle$ to be a 1 in the i-th position, $|i\rangle = (0, 0, ..., 1, ..., 0)$

- With this choice, the abstract vectors $|\phi\rangle, |\psi\rangle$ can be represented in components

$$\boldsymbol{\phi} = (\phi_0, ..., \phi_{2^n-1}) \quad , \quad \phi_i = \langle i|\phi\rangle \quad , \quad \boldsymbol{\psi} = (\psi_0, ..., \psi_{2^n-1}) \quad , \quad \psi_i = \langle i|\psi\rangle$$

- And the inner product takes the familiar form of a complex dot product:

$$\langle\phi|\psi\rangle = \sum_{i=0}^{2^n-1} \phi_i^* \psi_i$$

Generalized Probability Theory

- If $|\psi\rangle$ is a vector, then what is $\langle\psi|$? The relationship of $|\psi\rangle$(a "ket") to $\langle\psi|$ (a "bra") is effectively the same as the relationship between a column vector and a row vector.

- But since QM takes a basis-independent point of view (a vector is not just thought of as a list of coordinates , because that implies a choice of basis) it is useful to be aware of the abstract definition of "bras" like $\langle\psi|$.

- Given any vector space $\mathcal{V}$, one can consider the dual vector space $\mathcal{V}^*$ consisting of **linear functionals** that map vectors in $\mathcal{V}$ to real numbers, $\phi : \mathcal{V} \to \mathbb{C}$,

$$\phi\left(\alpha_1|\psi_1\rangle + \alpha_2|\psi_2\rangle\right) = \alpha_1\phi\left(|\psi_1\rangle\right) + \alpha_2\phi\left(|\psi_2\rangle\right) \quad , \quad \forall\alpha_1, \alpha_2 \in \mathbb{C} \ , \ |\psi_1\rangle, |\psi_2\rangle \in \mathcal{V}$$

- In mathematics, "functional" is a somewhat vague redundant term that is used mostly for historical reasons. A functional is a function that maps some higher dimensional object (like a function or vector) to a number.

- It turns out that for finite dimensional vector spaces, the dual space of linear functionals is always isomorphic to the original space, with every functional represented by the inner product with some vector:

$$\phi \mapsto \langle\phi, \cdot\rangle$$

- In QM, we represent the functional that is dual to $|\psi\rangle$ under this isomorphism by $\langle\psi|$. Therefore the "bras" are also called "dual vectors" or "co-vectors."

Generalized Probability Theory

- We have seen that an observable is a function that assigns real numbers to events, so that we can compute expected values of the observable by averaging this function over the probability distribution on events. We can think of an observable A as a collection of events $\{|a_i\rangle\}$ and values $\{a_i\}$ associated with those events.

- To compute an expectation value, we project onto the basis of events for the observable and take the average with respect to the squared magnitudes of the components. Define the Hermitian operator:

$$A = \sum_{i=1}^{m} a_i |a_i\rangle\langle a_i|$$

(m is the number of outcomes)

($|a_i\rangle\langle a_i|$ is an outer product ,

compare with $\mathbf{a}\mathbf{a}^T$)

- Now the change of basis and averaging for expectation values is simply an inner product:

$$\langle\psi|A|\psi\rangle = \sum_{i=1}^{m} a_i \langle\psi|a_i\rangle\langle a_i|\psi\rangle = \sum_{i=1}^{m} a_i |\langle a_i|\psi\rangle|^2 = \sum_{i=1}^{m} a_i |\psi(a_i)|^2$$

Generalized Probability Theory

- **Axiom of Observables.** An observable A as a collection of events $\{|a_i\rangle\}$ and values $\{a_i\}$associated with those events, conveniently represented as a Hermitian operator:

$$A = \sum_{i=1}^{m} a_i |a_i\rangle\langle a_i|$$

- So that the change of basis and averaging for expectation values is an inner product:

$$\langle\psi|A|\psi\rangle = \sum_{i=1}^{m} a_i \langle\psi|a_i\rangle\langle a_i|\psi\rangle = \sum_{i=1}^{m} a_i |\langle a_i|\psi\rangle|^2 = \sum_{i=1}^{m} a_i |\psi(a_i)|^2$$

Generalized Probability Theory

- Depending on the context of the model, it may be appropriate to update the state / probability vector describing the system after making a measurement.

- For example, if all I know is that I flipped n fair coins, then I would describe the system by a uniform distribution.   But if I learn the parity of the associated bit string (the number of tails mod 2) then I would "update my prior" and describe the system instead by a uniform distribution over strings of the correct parity.

- An exactly similar update occurs after measurements in QM.   From the perspective of GPT, there is no mystery in "collapsing the wave function". In classical theory, measurement identifies the state of the system with a particular event. The only new thing is that QM allows for incompatible sets of events.

Generalized Probability Theory

- In QM, an observable is a set of events $\{|\psi_a\rangle\}$ together we associated real values $\{a\}$. Therefore it is natural for a measurement of the observable that yields "a" to inform us that the system is in the state $|\psi_a\rangle$.

Axiom of Measurement. Measurement of an observable $A = \sum_{i=1}^{m} a_i |a_i\rangle\langle a_i|$ on the state $|\psi\rangle$ yields one of possible outcomes $a_1, \ldots, a_m$ with corresponding probabilities $|\langle a_1|\psi\rangle|^2, \ldots, |\langle a_m|\psi\rangle|^2$. If the outcome is a_i then the state of the system immediately after the measurement is:

$$|\psi\rangle \mapsto \frac{\langle a_i|\psi\rangle}{\sqrt{|\langle a_i|\psi\rangle|}} |a_i\rangle$$

Generalized Probability Theory

- Composition of subsystems in QM is a direct generalization of composition in probability theories. The Hilbert space of a composite system is the tensor product of the component Hilbert spaces.

$$\mathcal{H}_{12} = \mathcal{H}_1 \otimes \mathcal{H}_{\in}$$

- Just as in classical probability theories, the joint state of a composite system consisting of independent and uncorrelated states on the component subsystems is given by the tensor product

$$|\psi_{12}\rangle = |\psi_1\rangle \otimes |\psi_2\rangle$$

- These states which contain no quantum correlations ("entanglement") are called product states. They are described by a number of parameters that is linear in the number of subsystems (instead of exponential). **Exercise**: write out the resulting prob distributions for a product state, and see the product rule for probs.

Axiom of Composition. The joint state of a composite system is a vector in the tensor product Hilbert space of the component subsystem Hilbert spaces. The joint state of two subsystems is a tensor product of states on the component subsystems iff the resulting states are uncorrelated.

Summary of (static) Axioms of QM

- **Axiom of states.** States are rays in Hilbert space, or unit vectors in the 2-norm.

- **Axiom of composition.** The state of a composite system is a vector in the Hilbert space given by the tensor product of the Hilbert spaces of the component subsystems.

- **Axiom of observables.** An observable is a set of events that are associated with real values, formally represented as a Hermitian operator. Expectation values are given by inner products with this operator.

- **Axiom of Measurement.** Measuring a particular observable on a particular state returns a value associated with a particular event, after which the state is updated to the vector corresponding to that event.

- The remaining axiom we will need before moving onto the content of the theory is related to **dynamics**: the time evolution of quantum states. To motivate this we will return to classical probability theories and describe a dynamics that maps probability vectors to other probability vectors.

A closer look at probability theory

The notion of stochastic dynamics that describes the evolution of probability distributions can also be motivated from an algorithmic perspective.

Suppose we want to simulate a biased coin $\boldsymbol{\mu} = (1-\epsilon, \epsilon)$. We flip our fair coin enough times to make at least $1/\epsilon$ many events, and call one of these "tails."

This simple method of simulation could be called "binning", because in general we flip our coin many times and group together ("bin") sets of coin flip events to match the event probabilities in the system being simulated.

A closer look at probability theory

In general it will be computationally complex to determine the bins that correspond to a given probability distribution.

For example, consider the uniform distribution over the set of all combinatorial graphs (defined by vertices and edges) that have 40 vertices.   This probability distribution is well-defined, but how can we sample from it? We don’t even know the normalization constant?

A closer look at probability theory

Markov chains: powerful algorithmic technique for sampling desired distributions. Evolve with stochastic matrices that conserve probability.

N x N Stochastic Matrix *P:*

$$P_{ij} \geq 0 \;\forall i, j = 1, .., N$$

$$\sum_{j=1}^{N} P_{ij} = 1 \;\forall i = 1, ..., N$$

P maps probability distributions to probability distributions.

Terminology: a Markov chain is a stochastic process, a stochastic matrix is an algebraic object.

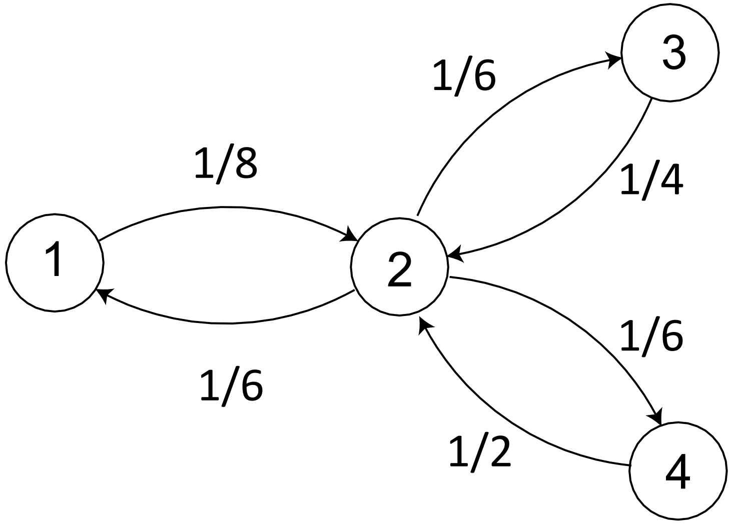
A closer look at probability theory

N x N Stochastic Matrix *P:*

$$P_{ij} \geq 0 \; \forall i, j = 1, .., N$$

$$\sum_{j=1}^{N} P_{ij} = 1 \; \forall i = 1, ..., N$$

Represent *P* as a graph: the indices *i,j* are vertices, and the *P*$_{ij}$ are edges between them.



$$P = \begin{bmatrix} 7/8 & 1/8 & 0 & 0 \\ 1/6 & 1/2 & 1/6 & 1/6 \\ 0 & 1/4 & 3/4 & 0 \\ 0 & 1/2 & 0 & 1/2 \end{bmatrix}$$

A closer look at probability theory

N x N Stochastic Matrix *P:*

$$P_{ij} \geq 0 \; \forall i, j = 1, .., N$$

$$\sum_{j=1}^{N} P_{ij} = 1 \; \forall i = 1, ..., N$$

P has one eigenvalue equal to 1, and all eigenvalues less than or equal to 1 in magnitude.

Therefore if P is designed so that the desired distribution is an eigenvector with eigenvalue 1, and all other eigenvectors are less than 1 in absolute value, then repeatedly applying P will approximately prepare the desired distribution.

A closer look at probability theory

$$P_{ij} \geq 0 \; \forall i, j = 1, .., N$$

N x N Stochastic Matrix *P:*

$$\sum_{j=1}^{N} P_{ij} = 1 \; \forall i = 1, ..., N$$

Stationary Distribution: $\boldsymbol{\pi} P = \boldsymbol{\pi}$

$$\mathrm{spec}(P) = \{1, 0.83, 0.64, 016\}$$

$$\boldsymbol{\pi} = \begin{bmatrix} 0.4 \\ 0.3 \\ 0.2 \\ 0.1 \end{bmatrix}$$

$$P = \begin{bmatrix} 7/8 & 1/8 & 0 & 0 \\ 1/6 & 1/2 & 1/6 & 1/6 \\ 0 & 1/4 & 3/4 & 0 \\ 0 & 1/2 & 0 & 1/2 \end{bmatrix}$$

A closer look at probability theory

While Markov chains are most commonly taken to be “homogeneous” (the same stochastic matrix applied at every time step), we can also consider inhomogeneous (time dependent) stochastic evolutions.

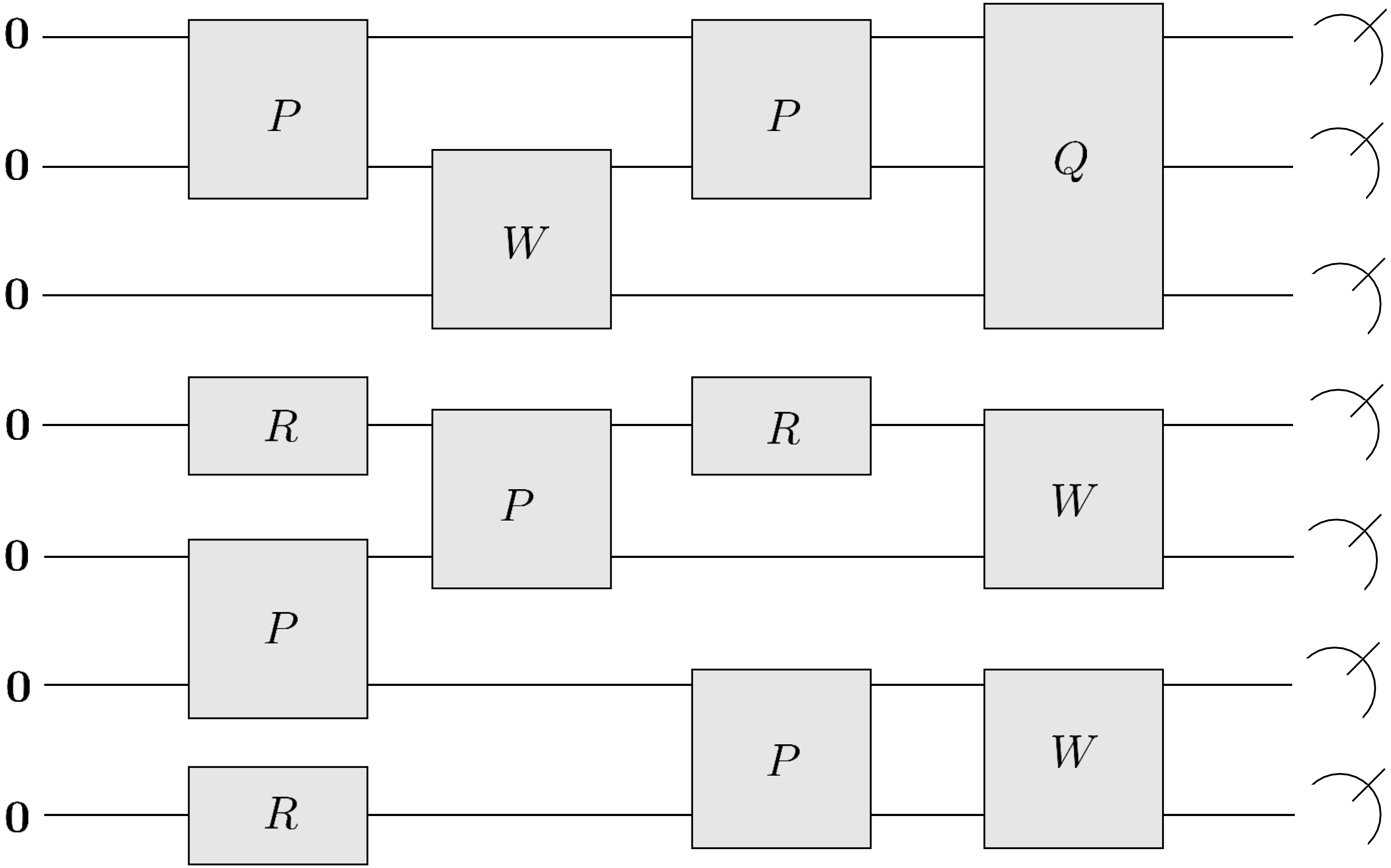
Another possibility is to consider composite systems, with stochastic matrices on the full system defined in terms of tensor products of stochastic matrices on subsystems. If $\Omega_{12} = \Omega_1 \times \Omega_2$, and P_1, P_2 are stochastic matrices defined on Ω_1, Ω_2 respectively, then a joint stochastic evolutions we can consider is:

$$P_1 \otimes P_2$$

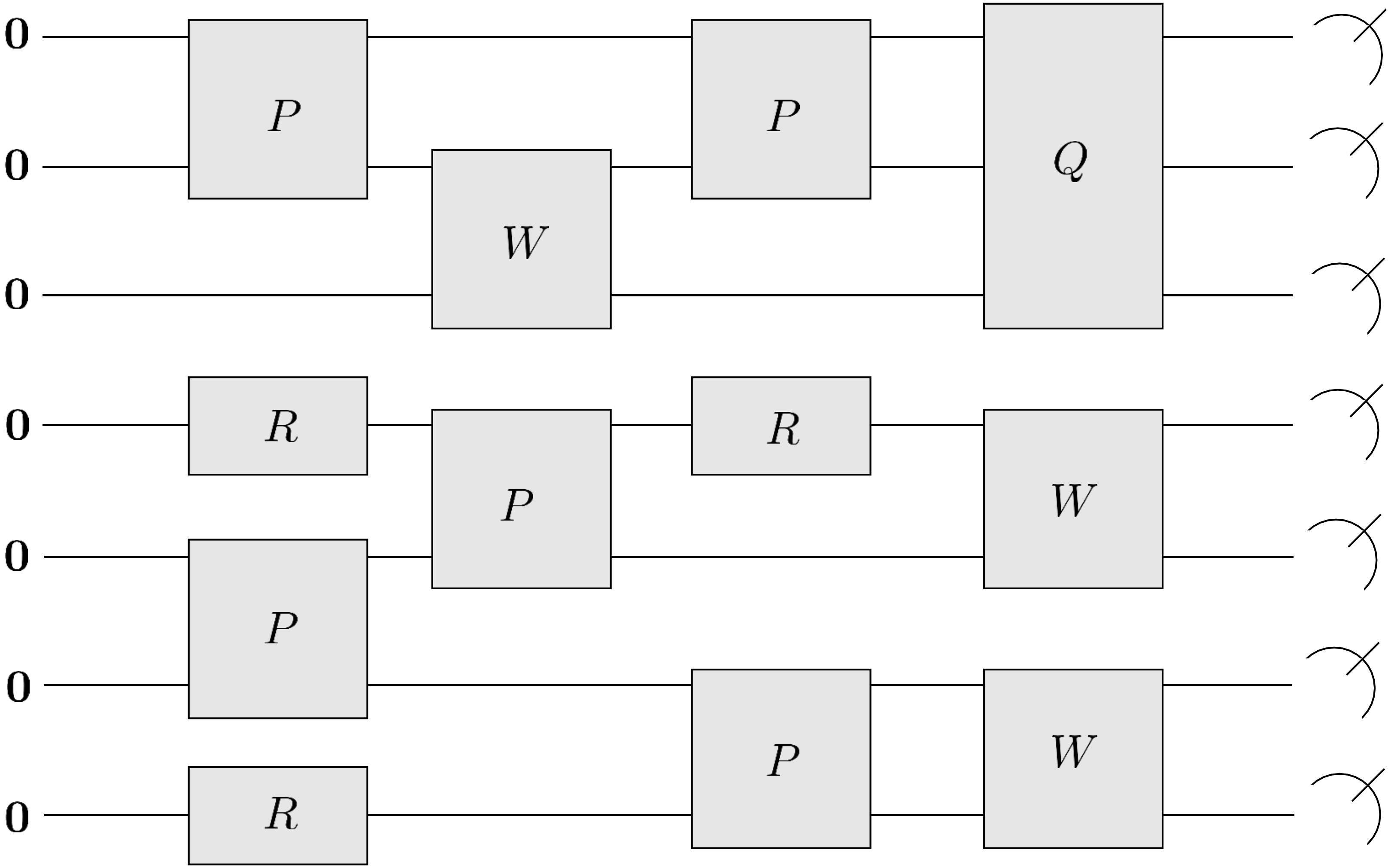
This joint Markov chain independently evolves each of the component subsystems. Therefore it is incapable of generating correlations across subsystems.

To create correlations we must include stochastic matrices that act across two or more subsystems.

By combining the notion of inhomogenous Markov chains, and stochastic matrices formed by tensor products, we can consider a model of stochastic circuits (P, W, R, Q are stochastic matrices = “gates”):

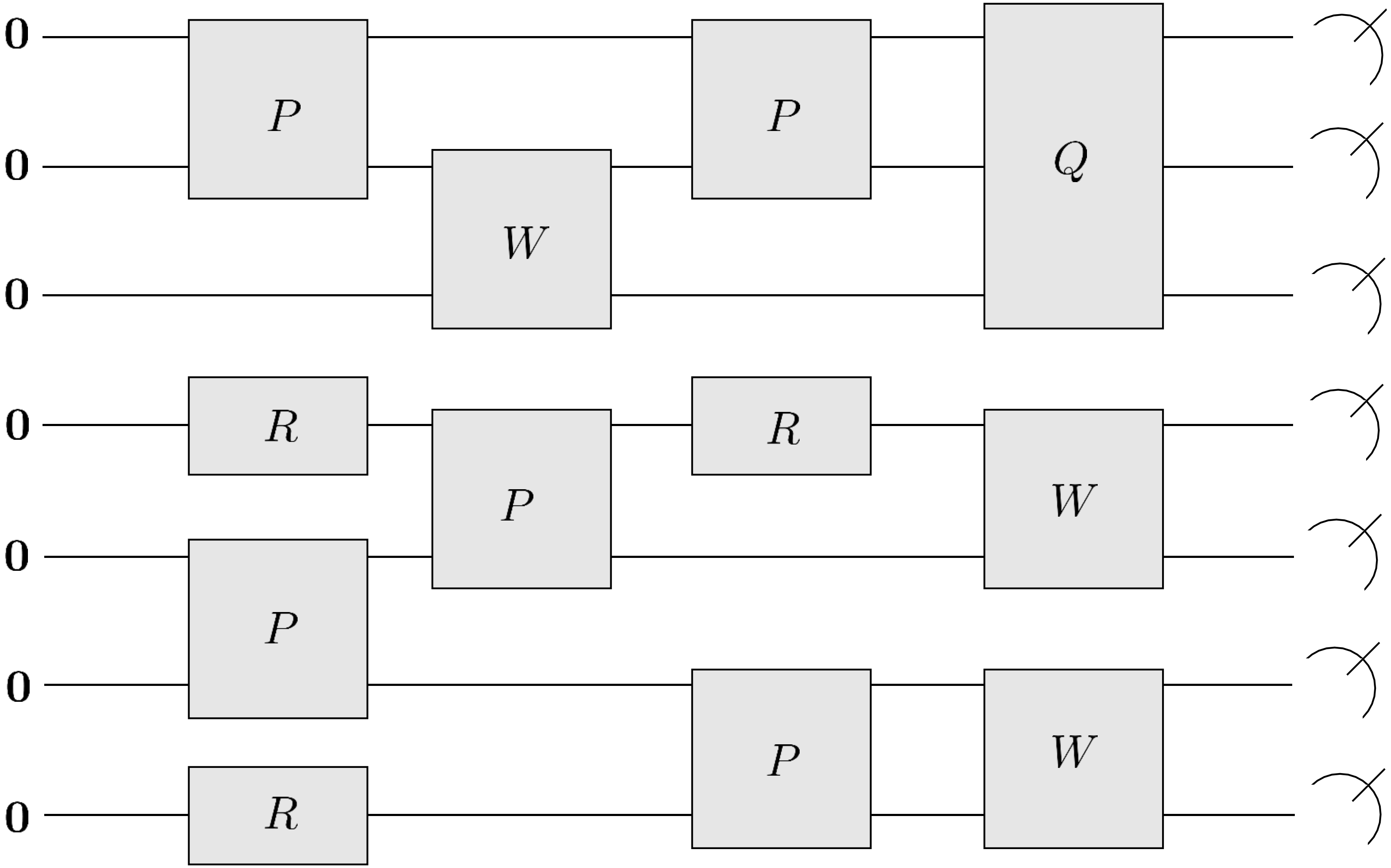


Simulate the model on a computer: brute-force exponential-time simulation of the probability vector?

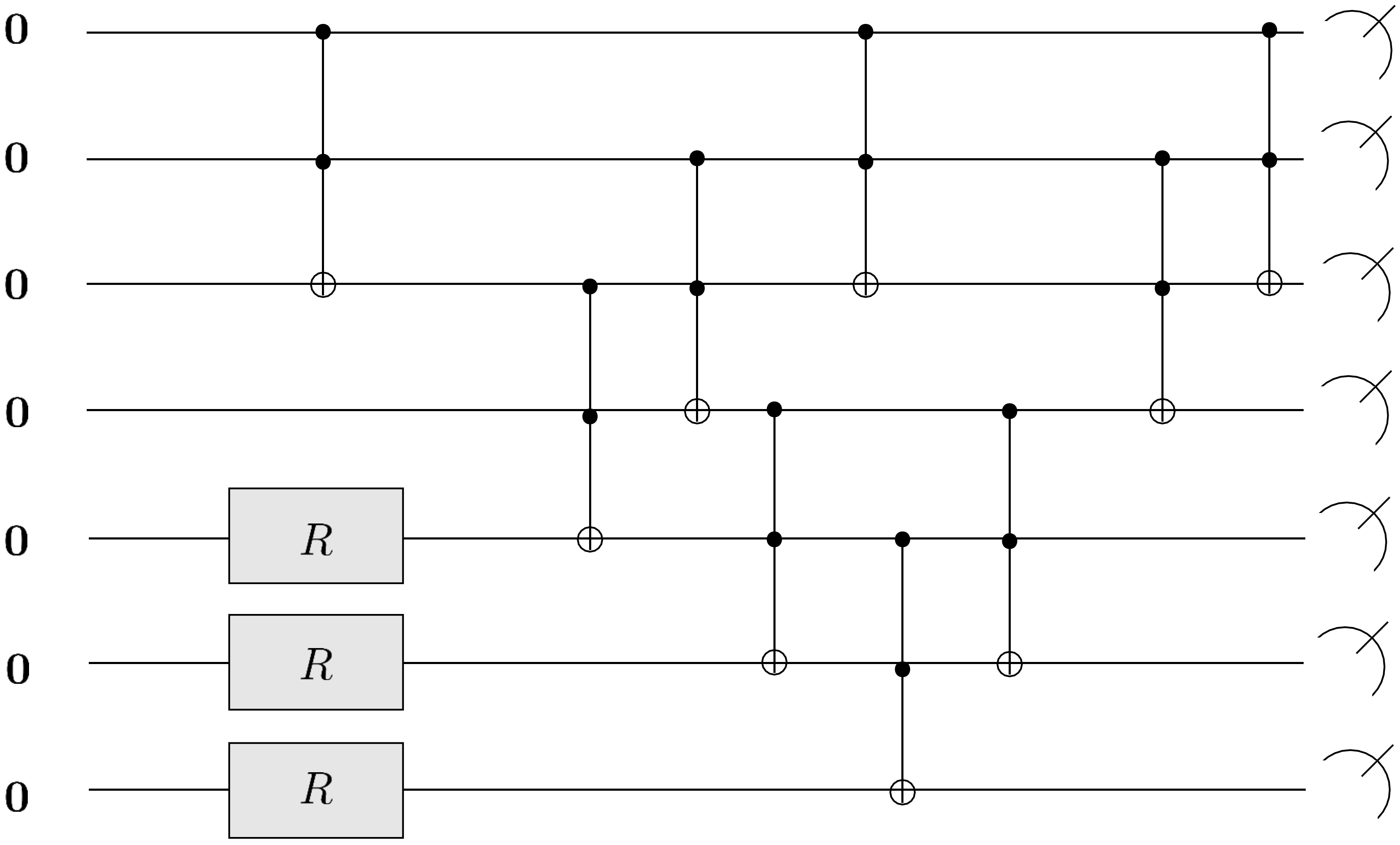


Simulate the model on a computer: brute-force exponential-time simulation of the probability vector?

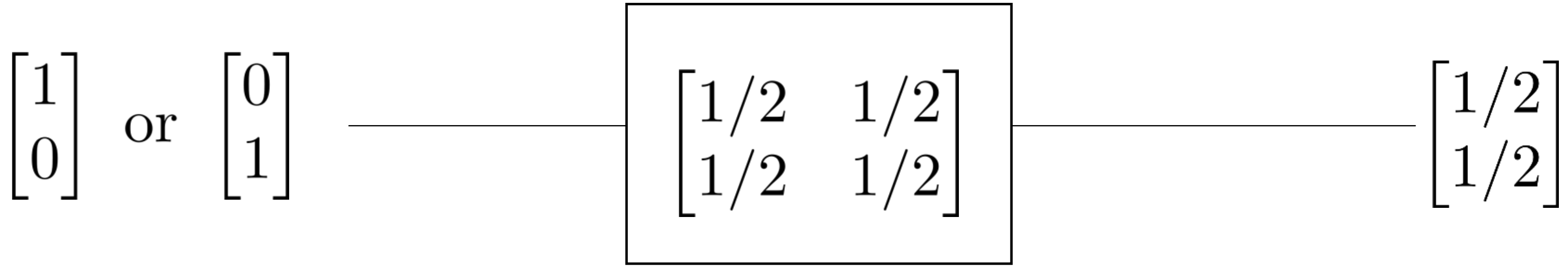
Efficient simulation by “random walk”!



Universality: what gates do we need to include? It suffices to have a classical reversible circuit together with ancillary fair coins as input. (for those who know: analogous to “Clifford + magic state QC”)



Irreversibility: When are stochastic gates invertible? Consider the simple case of an input that is either 0 or 1, passing through a fair coin flip gate:



$$\begin{bmatrix}1\\0\end{bmatrix} \text{ or } \begin{bmatrix}0\\1\end{bmatrix} \quad \begin{bmatrix}1/2 & 1/2\\1/2 & 1/2\end{bmatrix} \quad \begin{bmatrix}1/2\\1/2\end{bmatrix}$$

Nielson and Chuang mention in a comparison of classic logic gates and “quantum gates”, that the former are discrete and the latter continuous. One may disagree with this because stochastic gates form a continuous set, but the price that must be paid for consider these gates is **time-irreversibility**.

Dynamics for Probability Theories

- Stochastic matrices map probability distributions to probability distributions (i.e. they preserve normalization and nonnegativity).

- We can evolve probability distributions according to either homogeneous (time-independent) or inhomogeneous (time-dependent) Markov processes / stochastic matrices.

- For composite systems, one can form stochastic matrices by taking tensor products or stochastic matrices that act “locally” on a constant number of component subsystems.

- A fundamental property of stochastic evolutions is time-irreversibility. The very notion of a stationary distribution implies the time-irreversibility of stochastic process.

Quantum Theory: Unitary Dynamics

- **Quantum dynamics** should map quantum states to quantum states, which means they should preserve the 2-norm. These are called **orthogonal transformations** (over $\mathbb{R}$) and **unitary transformations** (over $\mathbb{C}$).

- Orthogonal and unitary transformations are defined as those linear transformations that preserve inner products. Given a linear operator U on a vector space $\mathcal{V}$, there is a corresponding adjoint operator $U^\dagger$ which acts on the dual space, so that U the co-vector of $U|\psi\rangle$ is $\langle\psi|U^\dagger$. In finite dimensional spaces, the matrix corresponding to $U^\dagger$ is the transpose conjugate of U .

- The property of preserving inner products means $\langle\phi|U^\dagger U|\psi\rangle = \langle\phi|\psi\rangle$ for all $|\psi\rangle, |\phi\rangle$, which implies

$$\boxed{U^\dagger U = I}$$

This equation implies $U^{-1} = U^\dagger$, in other words unitary evolution is always invertible.

Quantum Theory: Unitary Dynamics

An immediate intuition for unitary transformations can be gained by the fact that they preserve inner products (and hence the 2-norm). They preserve Euclidean distance, and are therefore **rotations**.

Invertibility is just the tip of the iceberg. This property implies that unitary transformations (or unitary matrices) form a group under multiplication. A continuous group that is also a compact manifold. The group U(d) of unitary transformations on a d-dimensional space is a classically studied **Lie Group**.

Consider a continuous one-parameter group of unitary transformations $\{U(t) : t \in \mathbb{R}\}$ satisfying

$$U(t)U(t') = U(t+t')$$

Representing the homogeneous time-evolution of a quantum syste. Stone's theorem says that

$$U(t) = e^{itA}$$

For some self-adjoint operator A (self-adjoint means $A = A^\dagger$, which is equivalently called a Hermitian operator i.e. an observable…)

Quantum Theory: Unitary Dynamics

More generally, the instantaneous unitary time evolution of a quantum state $|\psi(t)\rangle$ is *generated* by a (potentially time-dependent) Hermitian operator $H(t)$, yielding the famous **Schrodinger equation**:

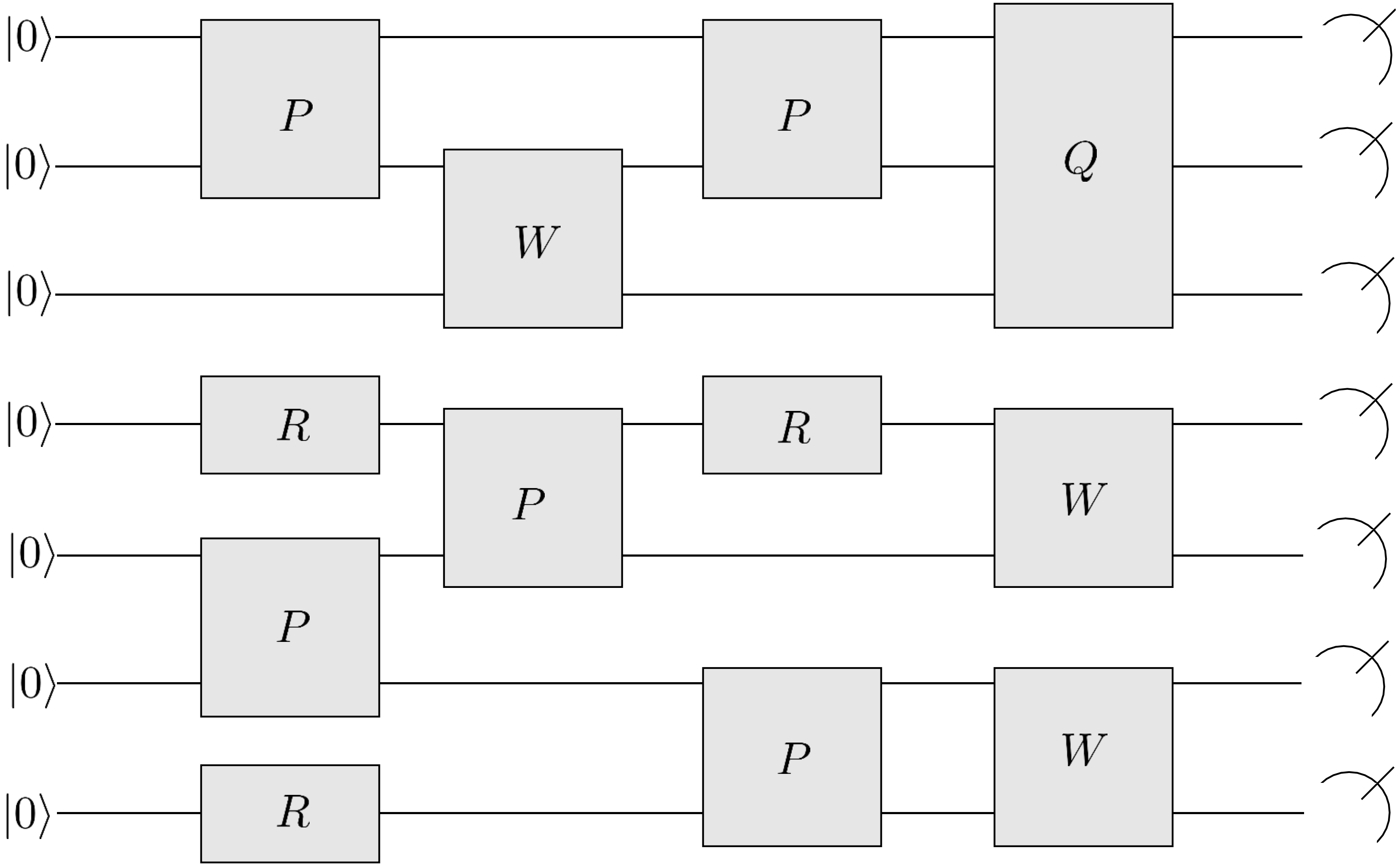
$$i\frac{d}{dt}|\psi(t)\rangle = H(t)|\psi(t)\rangle$$

The operator $H(t)$ is called the **Hamiltonian**, and it turns out to correspond to an observable of fundamental importance called the **energy** of the quantum system.

The point of "energy" is that it is conserved. It is invariant under time relations. In physics, symmetries correspond to conserved quantities, and energy is the conserved quantity corresponding to time translation symmetry.

This relationship of energy and time translation symmetry is what lies behind the Schrodinger equation. A beautiful treatment can be found in Sakurai's Graduate Physics QM book. For us it is too much of a tangent and so we will accept it as the **axiom of dynamics**.

Quantum Computer: just like the stochastic gate model, but now all gates (P, W, R, Q) are unitary.



Axioms of QM

- **Axiom of states.** States are rays in Hilbert space, or unit vectors in the 2-norm.

- **Axiom of composition.** The state of a composite system is a vector in the Hilbert space given by the tensor product of the Hilbert spaces of the component subsystems.

- **Axiom of observables.** An observable is a set of events that are associated with real values, formally represented as a Hermitian operator. Expectation values are given by inner products with this operator.

- **Axiom of Measurement.** Measuring a particular observable on a particular state returns a value associated with a particular event, after which the state is updated to the vector corresponding to that event.

- **Axiom of Dynamics**. Quantum time-evolution is a unitary transformation. This evolution is generated at each instant in time by a Hermitian operator called the Hamiltonian, which is the observable corresponding to energy, according to the Schrodinger equation.