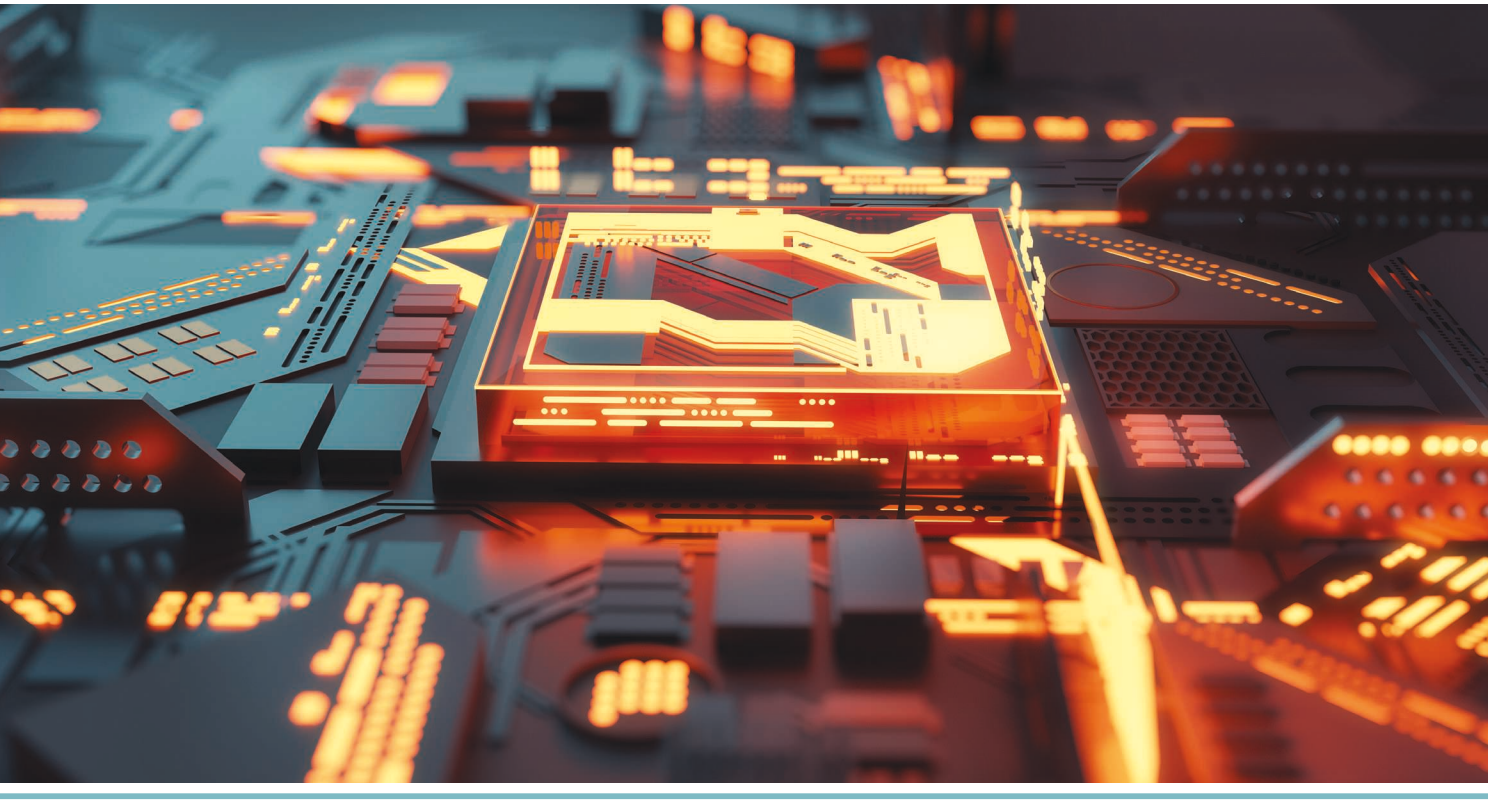
Quantum Computing

Joseph C. Bardin, Daniel Sank, Ofer Naaman, and Evan Jeffrey



During the past decade, quantum computing has grown from a field known mostly for generating scientific papers to one that is poised to reshape computing as we know it [1]. Major industrial research efforts in quantum computing are currently underway at many companies, including IBM [2], Microsoft [3], Google [4], [5], Alibaba [6], and Intel [7], to name a few. The European Union [8], Australia [9], China [10], Japan [11], Canada [12], Russia [13], and the United States [14] are each funding large national research initiatives focused on the quantum information

Joseph C. Bardin (jbardin@engin.umass.edu) is with the University of Massachusetts Amherst and Google, Goleta, California. Daniel Sank (danielsank@google.com), Ofer Naaman (ofernaaman@google.com), and Evan Jeffrey (ejeffrey@google.com) are with Google, Goleta, California.

Digital Object Identifier 10.1109/MMM.2020.2993475

Date of current version: 8 July 2020

sciences. And, recently, tens of start-up companies have emerged with goals ranging from the development of software for use on quantum computers [15] to the implementation of full-fledged quantum computers (e.g., Rigetti [16], ION-Q [17], Psi-Quantum [18], and so on). However, despite this rapid growth, because quantum computing as a field brings together many different disciplines, there is currently a shortage of engineers who understand both the engineering aspects (e.g., microwave design) and the quantum aspects required to build a quantum computer [19]. The aim of this article is to introduce microwave engineers to quantum computing and demonstrate how the microwave community's expertise could contribute to that field.

To begin, we pose the question, What's the big deal about quantum computing anyway? To put things in perspective, let's start by considering a regular digital computer—or, in the language of quantum engineers, a classical computer. The fundamental unit of information in a computer is the bit. In a conventional computer, one bit of memory is stored in the charge on a dynamic random-access memory (DRAM) cell. A memory register with N bits has 2^N possible states in the range $\{0\ldots00, 0\ldots01, 0\ldots10, \ldots\ldots, 1\ldots11\}$. Thus, a 300-b memory can store any of 2^{300} possible bit strings, a set so large that it is physically impossible to print it out since there is not enough matter in the universe to make the ink, not to mention the paper (the information capacity of the observable universe, if every degree of freedom of every particle was used to encode a bit, is roughly 2^{300} b [20]). However, while a 300-b register can assume any of its 2^{300} possible states, we know that each bit is either zero or one, so at any given time the register can store only one of the bit strings.

In a classical computer, computations are carried out by moving bits into a CPU, where logical operations are implemented in physical circuitry. The essential function of the CPU is to provide a prescribed output bit string for each possible input bit string so as to carry out a desired logical operation. Since this is digital circuitry, we know that for each input bit string the processor will produce exactly one output bit string and that this input-to-output mapping is defined in hardware, perhaps with some subset of the input bits determining the logical operation applied to the remainder of the bits.

In a quantum computer, we still have physical objects representing bits, but the difference is that the memory register exists as a quantum state. The basic building block in a quantum register is the quantum bit, or qubit. Just like a regular bit, a qubit has a zero (ground) state and a one (excited) state. However, what distinguishes a qubit from a classical bit is that it can be in a superposition of its zero and one states. Mathematically, the state of a qubit is described by a 2D complex state vector of unit amplitude,

Quantum computing has grown from a field known mostly for generating scientific papers to one that is poised to reshape computing as we know it.

$$|\psi\rangle = \begin{bmatrix}\alpha_0\\ \alpha_1\end{bmatrix} = \alpha_0\,|0\rangle + \alpha_1|1\rangle, \tag{1}$$

where $|0\rangle$ and $|1\rangle$ are the zero and one computational-basis vectors of the qubit, respectively, and the complex coefficients α_0 and α_1 are referred to as *probability amplitudes*. Physically, the probability amplitudes describe the likelihood that one would find the qubit in the associated state were a measurement to be carried out: $P\{|0\rangle\} = |\alpha_0|^2$ and $P\{|1\rangle\} = |\alpha_1|^2$. Note that in (1) we have implicitly introduced Dirac notation, where each of our computational-basis vectors is represented by a ket (i.e., $|0\rangle$ and $|1\rangle$).

A quantum register comprises an ensemble of qubits, and, just as with the digital register, an N-qubit register has 2^N possible states. However, since qubits can exist in superposition states, the instantaneous state of the quantum register can be in any superposition of its 2^N basis states; that is, $|\psi\rangle = [\alpha_{0\ldots00}\ \ \alpha_{0\ldots01}\ \ \ldots\ \ \alpha_{1\ldots11}]^{\mathrm{T}} = \alpha_{0\ldots00}|0\ldots00\rangle + \alpha_{0\ldots01}|0\ldots01\rangle + \cdots + \alpha_{1\ldots11}|1\ldots11\rangle$, where the probability amplitudes, α_i, are constrained so that $|\alpha_{0\ldots00}|^2 + |\alpha_{0\ldots01}|^2 + \cdots + |\alpha_{1\ldots11}|^2 = 1$. Similar to the case of a single qubit, each probability amplitude describes the likelihood that the associated bit string would be measured the associated bit string if the state of the quantum register were measured.

Computation on a quantum computer is typically carried out as an in-memory operation. That is, rather than moving data from the memory to a CPU and back, the quantum register is operated upon directly. An operator that can be applied to a quantum register is described by a unitary matrix $\hat{\mathbf{U}}$, which preserves the length of the unit state vector ($\hat{\mathbf{U}}\hat{\mathbf{U}}^{\dagger} = \mathbb{I}$). A universal quantum computer permits the application of any arbitrary unitary operation to the quantum register; i.e., a universal quantum computer can perform any transformation of the form

$$|\psi'\rangle = \hat{\mathbf{U}}|\psi\rangle, \tag{2}$$

where $\hat{\mathbf{U}}$ is an arbitrary unitary matrix.

Unitary operators should be nothing new to a microwave engineer; the scattering matrix of a lossless passive network is unitary ($\mathbf{SS}^{\dagger} = \mathbb{I}$). In fact, from a mathematical perspective, the way a quantum computation acts upon the 2^N states of a quantum register is identical to the operation that occurs when N incident waves scatter

from a lossless N-port network (see Figure 1). Therefore, we can see that familiar microwave concepts, such as superposition and interference, are also useful in providing intuition with respect to quantum computing; complex microwave signal amplitudes are analogous to complex quantum probability amplitudes, and, whereas the signals in Figure 1(a) carry power, those in Figure 1(b) carry probabilities.

There is an important distinction between the physical N-port model and the conceptual model of quantum computation: while the microwave N-port has one physical port associated with each input/output wave, the equivalent diagram describing quantum computation has a port associated with each of its 2^N computational states. Thus, the computational space of a universal quantum computer grows exponentially with the number of qubits. This massive computational capacity explains why quantum computing is so promising: even at roughly 50 qubits, the computational space is too large to simulate using the best supercomputers, while an ideal quantum computer can easily operate across the entire expanse [5].

Although superposition and interference make qubits a little more complex than classical bits, those features alone do not provide the exponential computational space described previously. For instance, a set of N isolated qubits is only about as complex as a set of N classical bits because, when the qubits operate in isolation, only a small subset of possible superposition states is accessed; practically speaking, we really still have $2N$ degrees of freedom here. However, when the individual qubits are allowed to interact, the total register of qubits accesses the full set of available superposition states. For example, the two-qubit state $|\psi\rangle = 0.5(|00\rangle + |01\rangle + |10\rangle + |11\rangle)$, which can be realized by two isolated qubits, may be "factored" into two single-qubit states $|\psi\rangle = 0.5(|0\rangle + |1\rangle) \times (|0\rangle + |1\rangle)$. On the other hand, the Bell state $|\psi\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ cannot be factored (try it!), so it cannot be thought of as representing the information of two independent bits. States that cannot be thought of as combinations of single-qubit states are called *entangled*, and the generation of such states is essential in accessing the power of quantum computing.

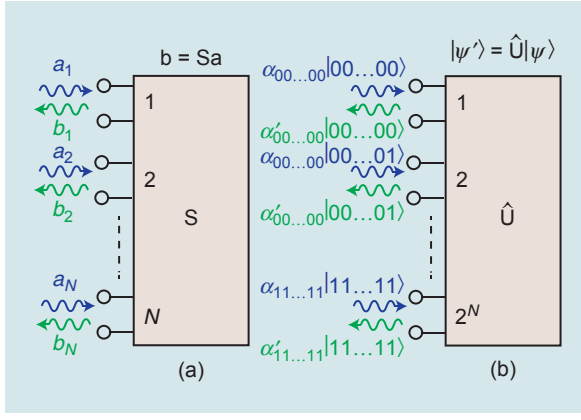


Figure 1. *The similarity between (a) scattering from a lossless microwave N-port and (b) quantum computation using an N-qubit quantum computer. While the mathematical operations are identical, the number of ports for the microwave network is a physical number of terminals, whereas the number of ports for the quantum operator scale exponentially in the number of qubits.*

While universal quantum computing permits operation on all 2^N computational states of a quantum register, we cannot perform a measurement of a quantum superposition state. Instead, when we measure the state of the quantum register at the end of a computation, the computer must make a decision. As dictated by the postulates of quantum mechanics, such a measurement returns a single classical bit of information per measured qubit, meaning that, if one measures the state of an N-qubit quantum register, one will find the register in one of its 2^N basis states, with the likelihood of each outcome equal to the modulus of the associated probability amplitude (i.e., $P\{|0\ldots00\rangle\} = |\alpha_{0\ldots00}|^2$, $P\{|0\ldots01\rangle\} = |\alpha_{0\ldots01}|^2$, and so on). Additionally, the act of measuring the state of a quantum register will cause any superposition state to collapse, and the best one can do is that the postmeasurement state of the register agrees with the measurement result. It is the job of the quantum algorithm designer to manipulate the quantum register into a state where, if measured, a certain bit string that corresponds to the desired solution will be returned with high probability.

So, from a high level, that's it. Quantum computing provides access to an exponential computational space, with the caveat that at the end of a computation you get only one answer. The remainder of this article is aimed at digging deeper. The following questions are addressed: What are the requirements of a technology to be used in a universal quantum computer? How can these requirements be met with practical components? How is the state of a quantum processor controlled? How is it measured? What is the state of the art today, and how must this be improved in the future if the full promise of quantum computing is to be realized? And, finally, as a microwave engineer, why should you care? Throughout the article, superconducting qubit technology is used to illustrate the key concepts.

Items beyond the scope of the article include quantum algorithms and derivations of the quantum mechanical properties of superconducting circuits. For review articles describing these items in detail, the curious reader is referred to [21] and [22]. In addition, we refer the reader to [23] for a detailed review of superconducting qubits. A sidebar, "Key Concepts,"

Key Concepts

In quantum computing, we manipulate the state of a quantum system through the Hamiltonian operator. But what exactly do we mean by "the state of a quantum system"? What is a quantum operator, and what is the significance of the Hamiltonian operator? Here, we answer each of these questions, using the quantum mechanical *LC* circuit as an example. Interested readers can find a more thorough background in one of the classic textbooks in the field, including [35], [36], and [46].

What Is a Quantum State?

In classical physics, we imagine an atom as a planetary system, with the electron orbiting the nucleus, i.e., describing the position of the electron with a function $\mathbf{r}(t)$. Quantum mechanics replaces this picture of a solid point-like electron with an "electron cloud" described by a complex-valued function $\psi(r,t)$ whose physical significance is that $|\psi(r,t)|^2$ is the spatial probability density of the electron's position at time t. The function $\psi(r,t)$ is called a *wave function*.

Similarly, in classical electrical engineering, we can describe the state of a circuit—in this case, the *LC* resonator—by its time-dependent flux $\Phi(t)$ and charge $Q(t)$, but quantum mechanics replaces that picture with the wave function $\psi(\Phi,t)$. Analogous to the atomic system, $|\psi(\Phi,t)|^2$ is the probability density for the flux Φ. What this means is that a measurement of the flux through the inductor at time t yields a random result, and the probability that the result is between Φ_a and Φ_b is $\int_{\Phi_a}^{\Phi_b}|\psi(\Phi,t)|^2 d\Phi$.

A wave function provides a complete representation of the state of a quantum system. While we have described the wave function of the quantum *LC* resonator in terms of flux, it is also possible to express the wave function with respect to a different observable. For instance, one could write the wave function in terms of charge, $\psi(Q,t)$, where, interestingly, $\psi(\Phi,t)$ and $\psi(Q,t)$ are related via a Fourier transform,

$$\psi(\Phi,t)=\frac{1}{\sqrt{2\pi\hbar}}\int_{-\infty}^{\infty}\psi(Q,t)e^{-jQ\Phi/\hbar}dQ, \tag{S1}$$

where $\hbar$ is the reduced Planck constant. Just as the time and frequency domain pictures of a signal carry the same underlying information, both $\psi(Q,t)$ and $\psi(\Phi,t)$ are equally valid wave functions that fully describe the same underlying quantum state. However, while $\psi(\Phi,t)$ describes the probability distribution for a measurement of flux in the *LC* resonator, $\psi(Q,t)$ describes the probability distribution for a measurement of charge across the capacitor.

What Is a Quantum Operator?

In quantum mechanics, all physical observables are represented by linear transformations—or, as physicists often call them, linear operators—which are applied to the wave function. In fact, for every observable in classical mechanics, there is a corresponding operator in quantum mechanics. For instance, one can write operators for the voltage, current, flux, charge, or other desired quantities. While, as with a wave function, the exact form of an operator depends on which basis we write it in, it turns out that the eigenvalues of an operator are independent of basis. Moreover, in quantum mechanics, eigenvalues of operators have special meaning: for a measurement outcome to be possible, it must be an eigenvalue of the associated operator. Moving forward, we focus here on the total energy operator, which has particular significance to quantum computing.

What Is the Importance of the Hamiltonian Operator?

The linear operator corresponding to the system's total energy is called, for historical reasons, the *Hamiltonian*, and it is denoted by $\hat{\mathrm{H}}$. The Hamiltonian is of particular importance to our study of quantum computing for two primary reasons. First, its eigenvalues are the discrete energies, E_n, that can be measured in a quantum system. Physicists refer to these energies as *quantum levels* or *eigenenergies*. For the undriven *LC* resonator of Figure S1(a), $E_n=\hbar\omega_0(n+1/2)$, where $\omega_0=1/\sqrt{LC}$ is the resonant frequency of the circuit and n is an integer that is greater than or equal to zero. When determining the state of a quantum processor, we typically make an indirect measurement of the qubits' energies. In the case of transmon qubits, we measure impedance, which depends on energy because of the nonlinearity of a transmon qubit (see the "Readout" section). Since the energies are discrete, it is also helpful to expand the wave function in terms of the eigenfunctions $[\psi_n(\Phi)]$ of the Hamiltonian operator, which are referred to as *energy eigenstates*:

$$\psi(\Phi,t)=\sum_{n=0}^{\infty}\alpha_\mathrm{n}(t)\psi_n(\Phi), \tag{S2}$$

where the energy eigenstates satisfy the typical eigenvalue equation,

$$\hat{\mathrm{H}}\psi_\mathrm{n}(\Phi)=E_\mathrm{n}\psi_\mathrm{n}(\Phi). \tag{S3}$$

Since (S3) is really a set of n independent equations, this choice of expansion leads us to a matrix representation for the Hamiltonian operator and

a vector representation for the wave function in the eigenenergy basis:

$$\hat{H} = \begin{bmatrix} E_0 & 0 & 0 & \cdots \\ 0 & E_1 & 0 & \cdots \\ 0 & 0 & E_2 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \tag{S4}$$

and

$$|\psi(t)\rangle = \begin{bmatrix} \alpha_0(t) \\ \alpha_1(t) \\ \alpha_2(t) \\ \vdots \end{bmatrix}. \tag{S5}$$

Note that $|\psi(t)\rangle$ is simply the wave function expressed as a vector in the energy eigenstate basis. There is something pretty deep here. Each entry in the Hamiltonian matrix corresponds to one of the possible measurement outcomes; that is, if a measurement of the energy in the system were made, then the outcome must be one of the diagonal entries of (S4). Since there is a 1:1 correspondence between the eigenvalues, E_n, and the energy eigenstates, $\psi_n(\Phi)$, each of the energy eigenstates is associated with a definite energy; in other words, one can find the probability of measuring E_i as $|\alpha_i|^2$. Interestingly, this would not be true if we used a different basis for our expansion of the wave function since the Hamiltonian matrix would no longer be diagonal.

Note, that the left-hand side of (S5) represents the state without giving its decomposition in any particular basis. This is common practice in the quantum physics community. For an n-level system, one typically writes $|0\rangle$, $|1\rangle, \ldots, |n\rangle$ rather than, for instance, $\psi_0(\Phi)$, $\psi_1(\Phi), \ldots, \psi_n(\Phi)$, and, when we say in the main text that we use the $|0\rangle$ and $|1\rangle$ states as computation-basis states, it means that our computational bases are the $\psi_0(\Phi)$ and $\psi_1(\Phi)$ components, and we are operating on α_0 and α_1.

But what if we wanted to determine expressions for the energy eigenstates in the flux basis? Each of the energy eigenstates can be found by solving the eigenvalue (S3), plugging in the appropriate value for E_n. Since the Hamiltonian operator in the flux basis is [22] $\hat{H} = ((-\hbar/2C)(d^2/d\Phi^2) + \Phi^2/2L)$, this corresponds to solving a second-order differential equation of the form

$$\left(-\frac{\hbar^2}{2C}\frac{d^2}{d\Phi^2} + \frac{\Phi^2}{2L} - E_n\right)\psi_n(\Phi) = 0. \tag{S6}$$

The first three energy eigenstates appear in Figure S1(b). Note that, if the wave function were expressed with respect to a different basis, then the general procedure would be exactly the same, but the form of the Hamiltonian operator would change in the new basis.

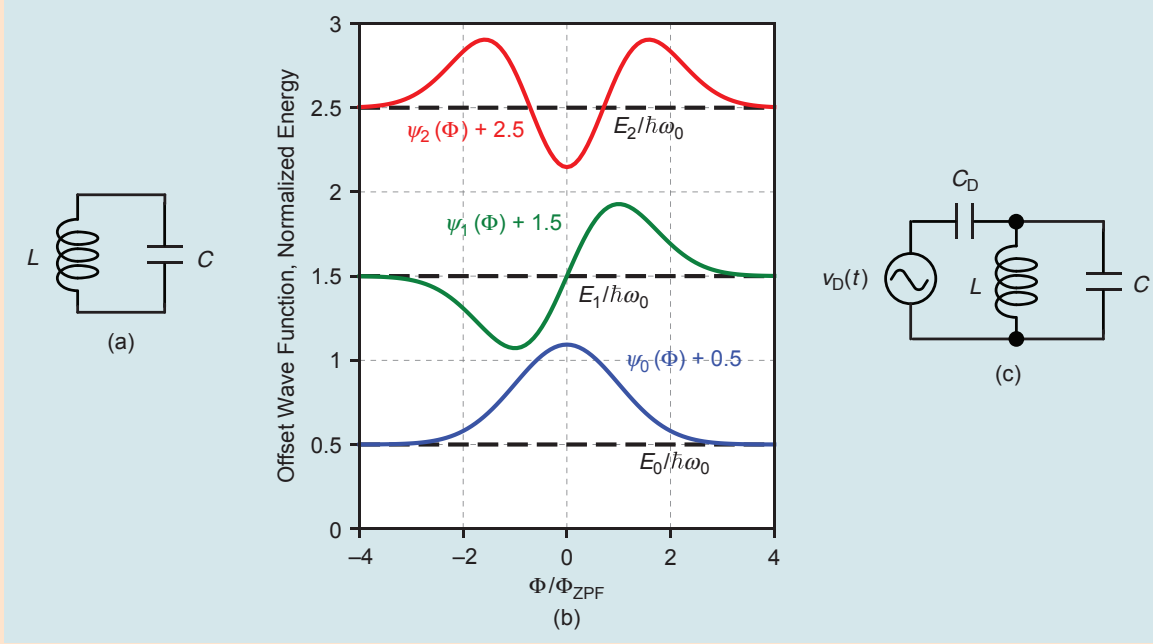


Figure S1. *The quantum LC resonator. (a) An undriven LC resonator. (b) The first three wave functions (solid lines) and energy levels (dashed lines) for the undriven quantum LC resonator, plotted as a function of flux, normalized to* $\Phi_{ZPF} = C\sqrt{\hbar\omega_0/2}$. *The wave functions have been offset to align with the associated energy level. (c) A quantum LC resonator with a drive source. It is assumed that* $C_D \ll C$ *so that* ω_0 *is not affected by this additional capacitance.*

The second reason that the Hamiltonian operator is significant to our current discussion is that the time evolution of a quantum state is governed by the so-called Schrödinger equation:

$$j\hbar\frac{\mathrm{d}}{\mathrm{d}t}|\psi(t)\rangle=\hat{\mathrm{H}}|\psi(t)\rangle. \tag{S7}$$

In quantum computing, our goal is to control the time evolution of a quantum system, so intuition into the dynamics of the Schrödinger equation is essential. It turns out that expanding the wave function in terms of the energy eigenstates as done earlier greatly simplifies the solution to the Schrödinger equation, which collapses to

$$\frac{\mathrm{d}\alpha_n(t)}{\mathrm{d}t}=-\frac{j}{\hbar}E_n\alpha_n(t), \tag{S8}$$

with the solution $\alpha_n(t)=\alpha_n(0)\exp\{-jE_nt/\hbar\}$. Thus, one can write the full wave function in the energy eigenstate basis as

$$|\psi(t)\rangle=\sum_{n=0}^{\infty}\alpha_n(0)e^{-j\omega_n t}|n\rangle, \tag{S9}$$

where $\omega_n\equiv E_n/\hbar=\omega_0(n+1/2)$. So, the expansion coefficients experience a phase progression that varies with index n, but the overall energy in the system stays the same. That the energy does not change through time is exactly what one would expect for an undriven lossless *LC* resonator.

So how do we cause the circuit to switch between states as we want to do when we control a qubit? This can be accomplished by adding a drive source to the circuit as in Figure S1(c). If we keep our basis states as the energy eigenstates of the undriven circuit [i.e., those of the circuit in Figure S1(a)], then the effect of the microwave drive is to add extra off-diagonal terms to the Hamiltonian:

$$\hat{\mathrm{H}}=\begin{bmatrix} E_0 & -j\gamma v_{\mathrm{D}}(t) & 0 & \cdots \\ j\gamma v_{\mathrm{D}}(t) & E_1 & -j\sqrt{2}\,\gamma v_{\mathrm{D}}(t) & \cdots \\ 0 & j\sqrt{2}\,\gamma v_{\mathrm{D}}(t) & E_2 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \tag{S10}$$

where γ is a coupling parameter related to the microwave design. When the new Hamiltonian is plugged into the Schrödinger equation, the extra off-diagonal terms cause mixing between adjacent levels:

$$\frac{\mathrm{d}\alpha_0(t)}{\mathrm{d}t}=-\frac{j}{\hbar}(E_0\alpha_0(t)-j\gamma v_{\mathrm{D}}(t)\alpha_1(t)), \tag{S11}$$

and, for $n\geq 1$,

$$\frac{\mathrm{d}\alpha_n(t)}{\mathrm{d}t}=-\frac{j}{\hbar}(j\sqrt{n}\,\gamma v_{\mathrm{D}}(t)\alpha_{n-1}+E_n\alpha_n(t) - j\sqrt{n+1}\,\gamma v_{\mathrm{D}}(t)\alpha_{n+1}(t)). \tag{S12}$$

Therefore, by engineering the Hamiltonian to add drive terms at appropriate frequencies, amplitudes, phases, and durations, we can control the evolution of the circuit's quantum state.

explains some quantum mechanics fundamentals that are relevant to the remainder of the article. It has been provided for those who are new to the field, and it may be helpful to read it before proceeding.

What Is Required to Build a Quantum Computer?

Just as the construction of a classical computer has well-defined technological requirements, so does the design of a universal quantum computer. These requirements can be summarized as five items known as the *DiVincenzo criteria* [24]:

1) *A well-understood and scalable qubit technology*: If we are going to control a large array of qubits, we will need a model that describes how the qubits interact with external controls and with one another. Beyond simply having a model, the technology should also be amenable for use in large arrays of qubits without a loss of performance.
2) *A mechanism for initializing each qubit into a known state*: It is essential that we can reliably initialize the state of a quantum processor to some well-known state. While the exact state (e.g., $|00\ldots0\rangle$ or $|11\ldots1\rangle$) is not important, it is essential that any initialization procedure be robust for all possible preinitialization states of the quantum processor.
3) *Fast gate operations with respect to the qubit coherence time*: A qubit behaves as an analog object during computation, and its coherence time is a metric that quantifies the characteristic timescale on which its state randomizes. Since a quantum computation consists of a series of gate operations, it is critical that the individual processes be carried out on a timescale orders of magnitude shorter than the coherence time of the individual qubits.
4) *A universal set of quantum gates*: A universal quantum computer allows for arbitrary unitary computations, which require a universal gate set.

Fortunately, this can be constructed from a set of single qubit gates (i.e., operations applied to just one qubit) and a suitable choice of a single two-qubit gate [25]. There are many different options for the constituents of a universal gate set.

5) *A method for measuring the state of individual qubits*: Finally, we must be able to determine the solution of a computation in a manner that is faithful to the probability amplitudes of the quantum state. What this really means is that we need a method to measure the state of each qubit.

How Can We Build a Quantum Computer?

How can we actually go about building a quantum computer that satisfies the requirements described previously? Let us consider the constraints one by one.

A Well-Understood Qubit Technology

First, we need a technology that provides scalable and well-understood qubits. A variety of candidate qubit technologies exist, including trapped ions [26], bulk and integrated photonics [27], cold atoms [28], semiconductor spin qubits [29], and superconducting circuits [23]. Here, we focus on superconducting circuits, which are monolithically fabricated devices that are engineered, controlled, and measured using familiar microwave circuit techniques. In particular, we discuss one type of superconducting qubit called the *transmon*, which is a nonlinear microwave resonator [30].

To understand the transmon, we start by describing the behavior in the quantum mechanical limit of a simpler circuit, a parallel LC resonator [Figure 2(a)] characterized by a resonant frequency, $\omega_0 = 1/\sqrt{LC}$, and a quality factor (Q-factor), $Q = R\sqrt{C/L}$. Normally, we assume that the amount of energy stored in an LC resonator can adopt any positive value. However, if we were to measure the energy in a high Q-factor resonator with very high sensitivity, we would find that the result is quantized to a set of allowed energies known as *quantum energy levels*,

$$E_{n,\mathrm{LC}} = \hbar\omega_0\left(n + \frac{1}{2}\right), \tag{3}$$

where $\hbar \approx 10^{-34}$ J·s is the reduced Planck constant and the index n is a nonnegative integer. The corresponding energy diagram appears in Figure 2(b). The equal spacing of the energy levels makes a lot of sense; the electromagnetic quantum of energy is $\hbar\omega$, and here we see that the step between allowed energies is equal to $\hbar\omega_0$. Thus, moving between the various energy levels corresponds to adding or removing an integer number of microwave quanta (photons) at the resonant frequency of the circuit. If we were to use this device as a qubit, we could employ the $|0\rangle$ and $|1\rangle$ states for our computational basis.

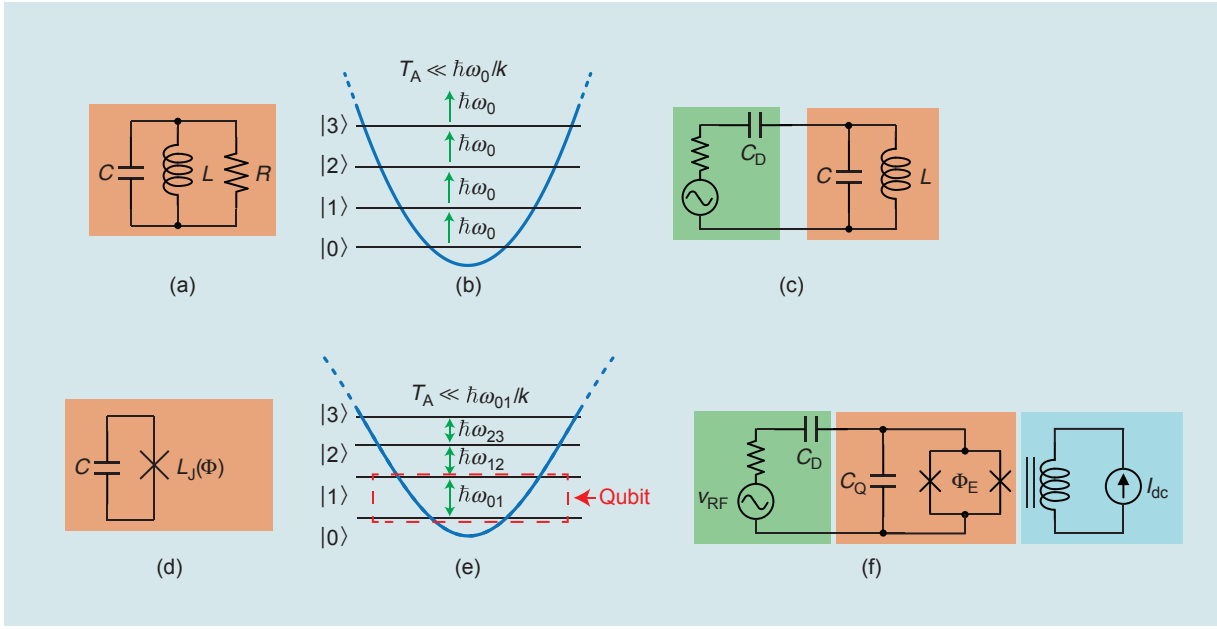


Figure 2. *From a linear resonator to the transmon. (a) The linear resonator circuit diagram. (b) The energy diagram for a high Q-factor LC linear resonator. (c) The drive circuit (left) and the linear resonator (right). (d) The fixed-frequency transmon (nonlinear resonator). The transmon is an LC resonator with a nonlinear inductance, realized with a Josephson junction (JJ) (represented by the X symbol). (e) The transmon energy diagram. The energy spacing between each level is unique due to the transmon nonlinearity. (f) The frequency-tunable transmon with RF and dc control ports: the XY drive (left), transmon qubit (nonlinear resonator) (center), and flux bias (Z drive) (right).*

Why do we usually ignore these quantum energy levels and just assume things are continuous? Microwave engineers rarely work in a situation where the electromagnetic quantum, $\hbar\omega$, is anywhere near the energy scale associated with the thermal energy of the environment, kT_A, where $k \approx 1.38 \times 10^{-23}$ J/K is Boltzmann's constant and T_A is the ambient temperature to which the resonator is thermalized. In the usual limit (e.g., at room temperature), Johnson noise from the resistor drives the resonator into a random distribution across many energy levels. In this case, the energy levels are literally in the noise. However, once we enter the regime where $T_A \sim \hbar\omega/k \approx 50\,\text{mK} \times (f/1\,\text{GHz})$, quantum mechanical effects start to manifest. The thermal excitation probability goes as $\exp\{-\hbar\omega/kT_A\}$; so, to ensure that thermal excitations of the resonator are lower than 1%, we need to reduce T_A to five times below $\hbar\omega/k$. A commercial dilution refrigerator can get down to 10 mK, so we expect to observe quantum mechanical behavior as long as $\omega_0 \geq 1$ GHz.

So far, things look promising. We have a device that behaves quantum mechanically if cooled to low-enough temperatures, and we understand its properties (e.g., ω_0) well enough that we can engineer them through the choice of circuit parameters. However, if we are going to use it as a qubit, we will need a way to deterministically transition between the $|0\rangle$ and $|1\rangle$ states. To do this, a drive at ω_0 can be capacitively coupled to the circuit, as in Figure 2(c). The drive causes the circuit to move between the various energy levels, and it is the basis for controlling the quantum state of the resonator. However, there is a problem. We want a system that behaves as if it has only two levels (i.e., $|0\rangle$ and $|1\rangle$). But when we excite the resonator at ω_0, we end up driving all the transitions, which leads to a superposition state that includes the higher levels. Therefore, instead of the two-level qubit described earlier, we have a system with many active levels.

Fortunately there is a fairly straightforward change we can make to the *LC* resonator that enables us to independently address the $|0\rangle \rightarrow |1\rangle$ transition: if we add nonlinearity to the resonator, the energy level spacing becomes nonuniform, a property that quantum engineers refer to as *anharmonicity*. In a transmon qubit, this nonlinearity is realized by replacing the inductor we had in the *LC* resonator by a Josephson junction (JJ), which is a superconducting tunnel junction that behaves as a nonlinear inductor of inductance $L_J = L_{J0}/\sqrt{1 - I_J^2/I_C^2}$, where I_J is the current through the JJ, I_C is the critical current of the JJ, $L_{J0} = \Phi_0/2\pi I_C$ is the zero-flux inductance of the JJ, $\Phi_0 = \pi\hbar/q \approx 2.07$ mV·ps is the magnetic flux quantum, and $q = 0.16$ aC is the charge of an electron. The transmon circuit is shown in Figure 2(d), and its energy levels are illustrated in Figure 2(e).

Defining Josephson and capacitive energies, $E_J = \Phi_0 I_C/(2\pi)$ and $E_C = q^2/(2C_Q)$, respectively, and in the usual case that $E_J \gg E_C$, the quantum energies of the transmon qubit are [30]

$$E_n \approx \underbrace{\hbar\omega_0\left(n + \frac{1}{2}\right)}_{\substack{\text{Linear resonator}\\ \text{energies}}} - \underbrace{\frac{E_C}{12}(6n^2 + 6n + 3)}_{\text{Anharmonicity}}, \tag{4}$$

where $\omega_0 = 1/\sqrt{L_{J0}C_Q}$ and $\eta = -E_C/\hbar$. For typical transmon parameters ($L_{J0} = 8$ nH and $C_Q = 80$ fF [31]), $\omega_0/2\pi$ and $\eta = -E_C/2\pi\hbar$ are roughly 6.2 GHz and –240 MHz, respectively, (for these same parameters, $E_J/2\pi\hbar$ is approximately 20 GHz, so our assumption that $E_J \gg E_C$ is valid). From (4), we can write the various transition frequencies

$$\omega_{01} = \frac{E_1 - E_0}{\hbar} \approx \omega_0 + \eta, \tag{5}$$

$$\omega_{12} = \frac{E_2 - E_1}{\hbar} \approx \omega_0 + 2\eta, \tag{6}$$

$$\omega_{23} = \frac{E_3 - E_2}{\hbar} \approx \omega_0 + 3\eta, \tag{7}$$

and so on. As depicted in Figure 2(e), the effect of the JJ nonlinearity is to create anharmonic energy levels ($\omega_{01} \neq \omega_{12} \neq \omega_{23} \ldots$). If we restrict the bandwidth of the microwave drive signal such that only ω_{01} is excited, then it is possible to treat this device as if it were an ideal two-level qubit.

So how nonlinear is a transmon? From (4), we see that each time a photon is added to the resonator, the resonant frequency drops by $|\eta|$. Moreover, η is a parameter that can be engineered through the choice of C_Q. However, one cannot arbitrarily increase the anharmonicity without paying a price in other performance metrics, and a typical value for $|\eta|/2\pi$ is roughly 200 MHz. Nonetheless, this is a very strong nonlinearity: one photon of energy (e.g., 3.3×10^{-24} J at 5 GHz) causes a 200-MHz shift in the transmon's resonant frequency. However, despite this, the transmon is still referred to by quantum engineers as *weakly anharmonic*, since the relative difference between the ω_{01} and ω_{12} transitions is much smaller than ω_{01}. This weak anharmonicity means that the minimum duration of the microwave pulses used to drive the ω_{01} transition will be limited by the necessity to avoid exciting the ω_{12} transition.

In many architectures, one must be able to dynamically tune the $|0\rangle \rightarrow |1\rangle$ transition frequency (ω_{01}) of a transmon. This is facilitated by replacing the JJ with a superconducting loop interrupted by two JJs. Such a loop, known as a *dc superconducting quantum interference*

device (*dc SQUID*), behaves like a single JJ where L_{J0} depends on the magnetic flux bias through the loop. The flux bias is typically provided by an on-chip bias coil near the SQUID. A complete schematic of a flux-tunable transmon qubit with microwave (XY) and dc-SQUID flux (Z) drivelines is shown in Figure 2(f), and photographs of a junction and complete qubit appear in Figure 3(a)–(c). The reason for the terms *XY* and *Z* will be become apparent shortly.

A critical property of a qubit is its coherence time: the timescale on which the state of the qubit is randomized by noise or loss. For a quantum technology to be useful in a quantum computer, its coherence time should be much greater than the timescale on which its quantum state can be manipulated. Quantum engineers refer to two quantities to describe decoherence: T_1, which is the energy relaxation time, and T_2, which is the dephasing time (in the physics community, it is customary to use the notation T_X to denote time constant *X* related to decoherence; here, we maintain this notation, with the caveat that we reserve the italicized *T* to refer to noise temperatures.). T_1 is related to the loaded Q-factor of the qubit (Q_L). A qubit initially in its $|1\rangle$ state will randomly flip to the $|0\rangle$ state due to electrical losses on a characteristic timescale of $T_1 = Q_L/\omega_{01}$. These losses can be resistances associated with the qubit materials or losses due to the coupling of the qubit to its control wiring. Achieving a typical value for T_1 of 50 μs for $\omega_{01}/2\pi = 5$ GHz requires $Q_L \approx 1.6 \times 10^6$; many years of materials research went into achieving this specification.

The second metric, T_2, is related to the frequency jitter of the qubit. For example, fluctuations in the JJ critical current cause the qubit frequency to fluctuate, leading to the accumulation of random phase. In tunable qubit architectures, another important source of noise is the flux-bias (frequency-control) line. Fixed-frequency (i.e., nontunable) transmons can usually achieve a better T_2 than tunable transmons.

Based on the preceding discussion, this technology appears to meet our first requirement; it is monolithic, so the qubits can be mass produced, and we can engineer the properties of the devices at the circuit level, so we have a pretty good

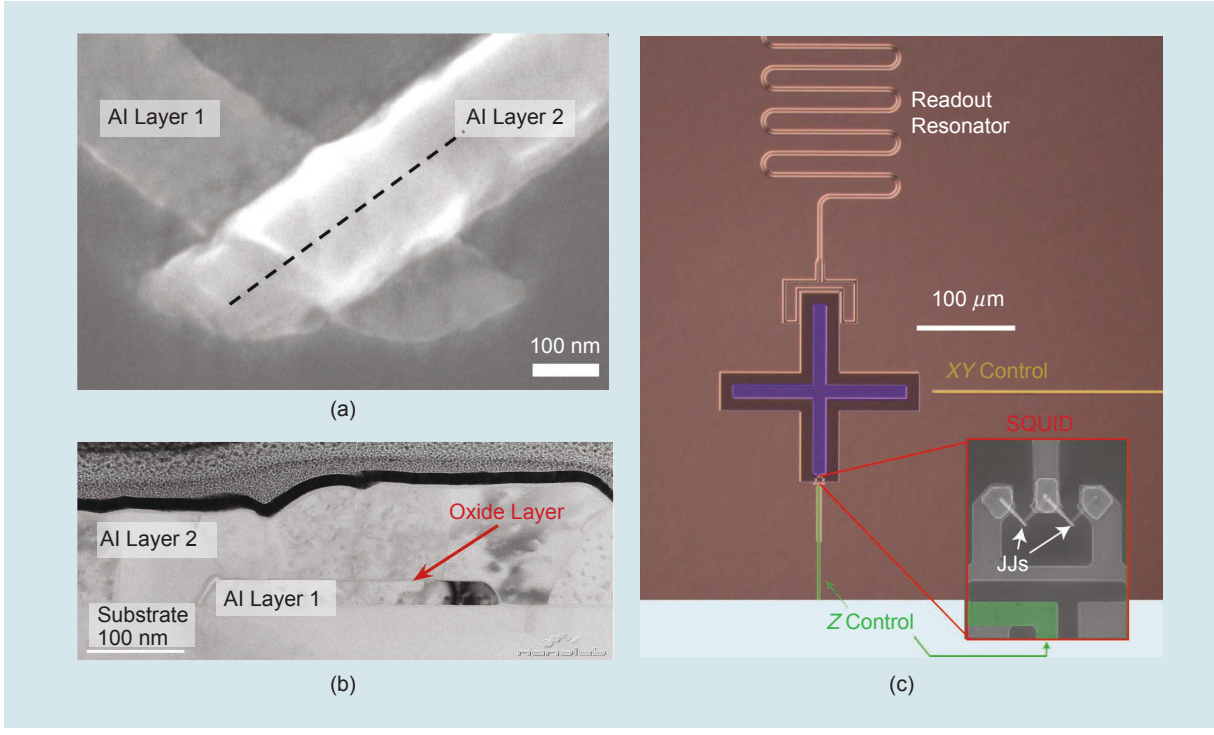


Figure 3. *A typical transmon qubit as seen through imaging. The (a) closeup view and (b) cross section of a JJ. (c) A complete qubit. The distributed qubit capacitance, XY coplanar waveguide (CPW) line, and Z CPW line are highlighted in blue, yellow, and green, respectively. A blow-up of the SQUID and flux-bias coupling is also shown. The JJs are typically patterned with electron beam lithography and made by a two-step aluminum (Al) deposition, interrupted by an oxidation step in a liftoff process, on sapphire or high-resistivity silicon wafers. Larger features in the circuit are typically patterned with photolithography and made with Al or niobium metallization in a liftoff or etching process. State-of-the-art JJ uniformity is typically in the 1% range, contributing a ~30-MHz variation in the qubit frequency. In most cases, superconducting qubits are made in low-volume research foundries on 100–200-mm tools, so, in practice, device parameter variation can significantly exceed 1%.*

understanding of the quantum behavior. But what about the other criteria?

Initialization

Initialization is relatively easy. Qubits naturally relax on the timescale T_1, so initialization into $|0\rangle$ can be done simply by waiting. Of course, since we want our qubit operations to be faster than T_1, we may need a way to initialize faster than T_1. A variety of techniques have been discovered and are in active use (e.g., [32] and [33]).

Fast and Universal Gates

Our next two criteria—fast gates with respect to the coherence time of our qubits and a universal gate set—are related, so we treat them together. What we are really after here is to develop an understanding of how the quantum state of a transmon-based processor is controlled via electrical signals. To develop this understanding, we begin by first considering how quantum states can be controlled in an abstract sense. We will see that the total energy operator, known as the *Hamiltonian operator*, is the handle that enables control of the quantum state. We then consider how quantum gates can be implemented, first in the abstract sense, and then we apply this knowledge to see how single- and two-qubit gates can be carried out on transmon qubits.

To begin, let's consider how physics permits the control of a quantum state in the first place. Recall that, when we perform a quantum computation, our goal is to apply a unitary operation, $\hat{\mathrm{U}}(t, t_0)$, that takes a state $|\psi(t_0)\rangle$ to a new state $|\psi(t)\rangle$ in time $\Delta t = t - t_0$, i.e.,

$$|\psi(t)\rangle = \hat{\mathrm{U}}(t, t_0)|\psi(t_0)\rangle. \tag{8}$$

So what physics can we leverage to implement the unitary operator $\hat{\mathrm{U}}(t, t_0)$? The key here is understanding how a quantum state evolves as a function of time and how to leverage this intuition to engineer quantum gate operations. The time evolution of a quantum state is dictated by the time-dependent Schrödinger equation,

$$\frac{d}{dt}|\psi(t)\rangle = -\frac{j}{\hbar}\hat{\mathrm{H}}(t)|\psi(t)\rangle, \tag{9}$$

which says that the time evolution of a quantum state is entirely determined by the Hamiltonian (total energy) operator of the system, $\hat{\mathrm{H}}(t)$. Since the state of an N-qubit register is described by a state vector of length 2^N, the Hamiltonian operator for an N-qubit register is a $2^N \times 2^N$ matrix. Hence, a single qubit has a 2×2 Hamiltonian, a pair of qubits share a 4×4 Hamiltonian, and so on. Solving (9), given an initial state at $t = t_0$ under the assumption that the Hamiltonian operator commutes with itself throughout the operation,

$$|\psi(t)\rangle = \underbrace{\exp\left\{-\frac{j}{\hbar}\int_{t_0}^{t}\hat{\mathrm{H}}(t')\mathrm{d}t'\right\}}_{\hat{\mathrm{U}}(t,t_0)}|\psi(t_0)\rangle. \tag{10}$$

Thus, we can control the state evolution of a quantum register if we can program its Hamiltonian; therefore, the name of the game in quantum control is to engineer knobs that permit doing so deterministically. These knobs may be microwave drive signals (superconducting [23] and semiconductor spin qubits [34]), flux biases (superconducting qubits [23]), laser drive signals (trapped ion qubits [26]), phase shifters (photonic qubits [27]), and so on. Shortly, we will dive into the details of how this is accomplished for transmon qubits. However, let us first consider what kind of gate operations we need to develop.

In quantum computing, where individual bits can exist in a complex superposition state, there is a full continuum of gate operations that can be carried out.

For universal quantum computing, we need to apply an arbitrary $2^N \times 2^N$ unitary to a quantum register. Producing a fully controllable $2^N \times 2^N$ Hamiltonian would be intractable. Fortunately, any arbitrary quantum operator $\hat{\mathrm{U}}$ can be decomposed into a sequence of elementary operations applied to just a single qubit or a pair of qubits. These elementary operations are referred to as *quantum gates*, and the quantum algorithms used to implement a desired $2^N \times 2^N$ unitary operator are built up using sequences of them. Thus, we can focus our attention on one- or two-qubit subsystems when considering how to control a quantum computer.

Single-Qubit Gate Operations

First, we consider single-qubit gate operations in general before seeing how one carries out such operations on transmon qubits. In regular digital logic, there are really only two useful cells that operate on a single bit: an inverter and a buffer. However, in quantum computing, where individual bits can exist in a complex superposition state, there is a full continuum of gate operations that can be carried out. Single-qubit states are easy to understand using a visual representation of the qubit state. After removing an unobservable global phase term from (1), we can write the state in terms of two variables, θ and ϕ:

$$|\psi\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + e^{j\phi}\sin\left(\frac{\theta}{2}\right)|1\rangle. \tag{11}$$

The continuum of values that can be assigned to these two variables maps to the surface of a sphere, which is referred to as the *Bloch sphere*.

The Bloch sphere representation of a qubit is shown in Figure 4(a). Any single-qubit state can be represented by a unit vector, known as a *Bloch vector*, on the surface of the Bloch sphere. The ground state $|0\rangle$ is at the north pole, the excited state $|1\rangle$ is at the south pole, and all other points on the surface of the sphere represent unique superposition states. Since all single-qubit states lie on the surface of the Bloch sphere, any single-qubit gate operation corresponds to a rotation of the Bloch vector around some axis. When a measurement of the qubit state is carried out, the result depends solely on the elevation angle, θ; that is, $P\{|0\rangle\} = \cos^2(\theta/2)$ and $P\{|1\rangle\} = \sin^2(\theta/2)$. So one can easily estimate the measurement probabilities from this graphical representation. On the other hand, since any single-qubit gate can be viewed as a rotation operation, its impact will depend on both θ and ϕ.

When we carry out single-qubit gate operations, we are moving the Bloch vector of a qubit around the surface of the Bloch sphere, and the handle that enables us to achieve this goal is the Hamiltonian operator of the qubit. But what exact Hamiltonian operator(s) would we like to engineer? That is, what form would we like our single-qubit Hamiltonian to take if we wish to rotate the Bloch vector around an arbitrary axis? To answer this question, let's consider an expression for a generic single-qubit Hamiltonian,

$$\hat{\mathrm{H}} = \frac{\hbar\omega_{\mathrm{R}}}{2}(w_{\mathrm{X}}\hat{\sigma}_{\mathrm{X}} + w_{\mathrm{Y}}\hat{\sigma}_{\mathrm{Y}} + w_{\mathrm{Z}}\hat{\sigma}_{\mathrm{Z}}), \tag{12}$$

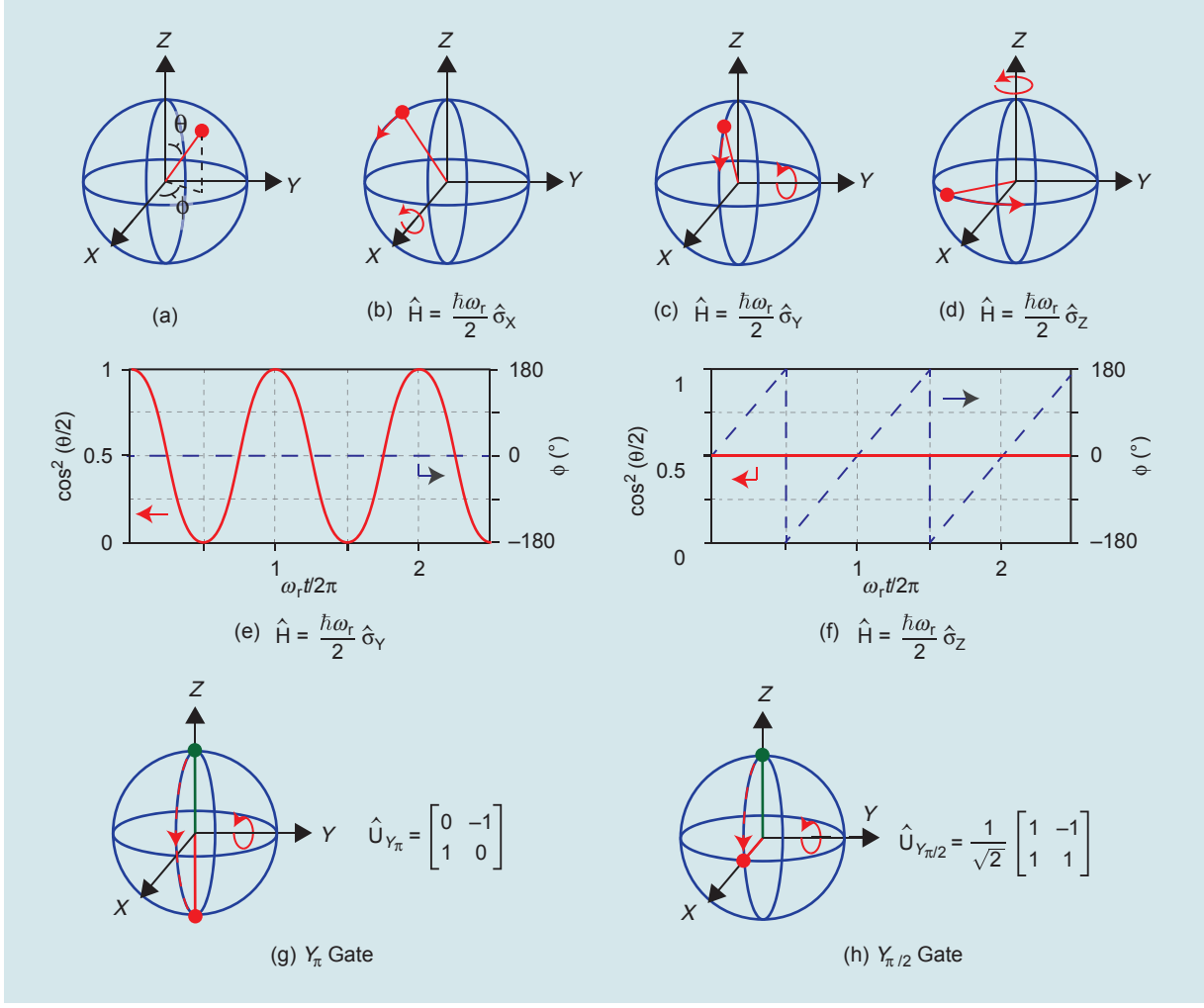


Figure 4. *The Bloch sphere is a valuable tool to understand single-qubit states and control operations. (a) The Bloch sphere representation of a qubit state* $|\psi\rangle = \cos(\theta/2)|0\rangle + \exp\{j\phi\}\sin(\theta/2)|1\rangle$. *The effect of (b) Pauli matrices* $\hat{\sigma}_{\mathrm{X}}$, *(c)* $\hat{\sigma}_{\mathrm{Y}}$, *and (d)* $\hat{\sigma}_{\mathrm{Z}}$ *is to rotate the qubit state around the x-, y-, and z-axes, respectively. The effect of (e)* $\hat{\sigma}_{\mathrm{Y}}$ *on an initial state* $|\psi(0)\rangle = |0\rangle$ *and (f)* $\hat{\sigma}_{\mathrm{Z}}$ *on an initial state* $|\psi(0)\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$. *The action of the (g)* Y_π *and (h)* $Y_{\pi/2}$ *gates. The Bloch sphere pictures show the effect of the gates on a qubit that is initially in the* $|0\rangle$ *state, whereas the matrices describe the effect on an arbitrary state.*

where ω_R is an angular frequency of rotation; w_X, w_Y, and w_Z are dimensionless weights ($|w_X|^2 + |w_Y|^2 + |w_Z|^2 = 1$); and

$$\hat{\sigma}_X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \hat{\sigma}_Y = \begin{bmatrix} 0 & -j \\ j & 0 \end{bmatrix}, \quad \text{and} \quad \hat{\sigma}_Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{13}$$

are the Pauli spin matrices [37]. What we have done here is write the general Hamiltonian of our qubit in terms of three independent operators ($\hat{\sigma}_X$, $\hat{\sigma}_Y$, and $\hat{\sigma}_Z$). It turns out that breaking things up in this way is particularly useful when trying to visualize the effect of the Hamiltonian on the qubit dynamics.

As illustrated in Figure 4(b)–(d), the presence of each of the Pauli spin matrices in a Hamiltonian corresponds to rotation of the Bloch vector around one of the major axes; i.e., $\hat{\sigma}_X$, $\hat{\sigma}_Y$, and $\hat{\sigma}_Z$ describe rotation about the *x*-, *y*-, and *z*-axes, respectively. Example time evolutions for $\hat{\sigma}_Y$ and $\hat{\sigma}_Z$ operators appear in Figure 4(e) and (f), respectively. Rotation around the *y*-axis causes the measurement probabilities to oscillate with time; these probability oscillations are referred to by the quantum community as *Rabi oscillations*. On the other hand, rotation around the *z*-axis leaves the measurement probabilities unaffected and influences only the qubit phase (ϕ). By controlling the duration through which the $\hat{\sigma}_Y$ and $\hat{\sigma}_Z$ operators are enabled, we can carry out deterministic rotations around the *y*- and *z*-axes, respectively. For instance, as shown in Figure 4(g), if we enable $\hat{\sigma}_Y$ for a duration $t = \pi/\omega_R$, then we can execute a Y_π gate, which initially flips a qubit in state $|0\rangle$ to state $|1\rangle$ and vice versa, analogous to a classical NOT gate. Similarly, as shown in Figure 4(h), if we turn on $\hat{\sigma}_Y$ for a duration $t = \pi/2\omega_R$, then we can execute a $Y_{\pi/2}$-pulse, which creates a superposition state from a qubit initialized to the $|0\rangle$ or $|1\rangle$ states. If we had control knobs that permitted us to enable and disable each of the Pauli operators at will, we would be able to perform any single-qubit gate operation we wanted. Now, let's see how this is accomplished for the transmon qubit.

Control of Transmon Qubit

Let's first write the Hamiltonian of the isolated qubit of Figure 2(d) and consider the drive later. Throughout, we make the assumption that the control operations do not take us out of the two-level computational space, and we will deal with the constraints this imposes when we describe the practical requirements for single-qubit control. The Hamiltonian of the isolated transmon can, then, be determined by considering the qubit's two lowest energies [see (4)]:

$$\hat{H}_0 = -\frac{\hbar\omega_{01}}{2}\hat{\sigma}_Z. \tag{14}$$

What this means is that the state of a transmon qubit naturally rotates about the –*z*-axis of the Bloch sphere at a frequency ω_{01}, i.e., $|\psi(t)\rangle = \cos(\theta_0/2)|0\rangle + \sin(\theta_0/2)\exp\{j(\phi_0 - \omega_{01}t)\}|1\rangle$, where θ_0 and ϕ_0 are the initial values of θ and ϕ, as defined in Figure 4(a). For this reason, we typically refer to ω_1 as the *qubit frequency*.

To understand the action of quantum gate operations, it is helpful to first remove the natural rotation of the qubit, analogous to working in the baseband envelope domain. After making this change of coordinates to what we call the *rotating frame*, the isolated qubit Hamiltonian simplifies as $\hat{H}_{0,R} = 0$ since there is no relative motion in the new frame. Note that the subscript *R* has been added to signify that the expression is valid in the rotating frame. Moving forward, we will work in the rotating frame.

Now we are ready to talk about how one can apply gates to the transmon. Let us assume that we want to apply a phase gate, which rotates the single qubit state around the *z*-axis by some desired angle. To see how one might do that, we can rewrite (14) in the rotating frame, while allowing for a frequency detuning, $\Delta\omega_{01}$, from the idling value of ω_{01},

$$\hat{H}_{0,R}(\Delta\omega_{01}) = -\frac{\hbar\Delta\omega_{01}}{2}\hat{\sigma}_Z. \tag{15}$$

So we can enable a rotation around the *z*-axis by controlling the frequency of the qubit through the flux-tuning line, which, accordingly, is referred to as *Z control*. This mechanism is similar to the way a phase-locked loop adjusts the phase of a voltage-controlled oscillator (VCO) using the frequency-tuning voltage.

What about rotations around the *x*- and *y*-axes? Since these are transverse rotations (i.e., from $|0\rangle$ to $|1\rangle$ and vice versa), a method to couple energy into the circuit is required. This can be achieved through capacitive [as in Figure 2(e)] or inductive coupling of a microwave generator to the transmon, and it adds a drive term to the Hamiltonian. If $v_{RF}(t) = e(t)\sin(\omega_{01}t + (\pi - \phi_D))$, then the drive term corresponding to capacitive coupling is

$$\hat{H}_{D,R} \approx \frac{e(t)}{2}\frac{C_D}{C_D + C_Q}\sqrt{\frac{\hbar}{2Z_Q}}(\cos(\phi_D)\hat{\sigma}_X + \sin(\phi_D)\hat{\sigma}_Y), \tag{16}$$

where $Z_Q \equiv \sqrt{L_{J0}/C_Q}$ (typically, Z_Q is roughly 300 Ω [31]). The overall Hamiltonian is now $\hat{H}_{0,R} + \hat{H}_{D,R}$. So the microwave drive induces a rotation of the Bloch vector around an axis in the *xy*-plane whose angle is determined by the carrier phase (ϕ_D). The rate of this rotation is determined by the amplitude of the microwave signal. For a universal gate set, we really only require rotations around the *x*- and *y*-axes, so the key

requirements are that we can shift the carrier by 90° and that we have accurate control of the integrated envelope amplitude, which sets the amount of rotation.

Now, let's consider some practical constraints. The required amplitude of the drive signal depends on C_D, which sets the drive-coupling strength. However, the driveline presents a resistive loss channel to the qubit, limiting the qubit relaxation time to $T_1 \leq \gamma_{\text{loading}}^{-1} = Q_D/\omega_{10}$, where Q_D is the Q-factor imposed on the qubit due to loading by the driveline. The loading is related to the drive capacitance via $Q_D \approx (C_Q/C_D)^2 Z_Q/Z_0$, where Z_0 is the generator impedance. We typically engineer Q_D to be on the order of 5×10^7 or so, which means that we need to keep our coupling very weak: C_Q and C_D are usually on the order of 100 fF and 25 aF, respectively [31]. Referring to Figure 3(c), we see that the small drive capacitance has been realized as coupling from an open-circuited coplanar waveguide (CPW) line to the qubit across a region of the ground plane, which can be designed using modern electromagnetic simulation tools.

Care also must be taken in coupling the dc flux bias to the qubit. While noise on the XY line drives XY rotations of random phase, noise on the Z driveline results in fluctuations of the qubit frequency, analogous to noise on the control line of a VCO. To keep these fluctuations in a range where they do not limit the coherence of the qubit, the magnetic coupling is very weak, with typical mutual inductances on the order of 2 pH. Such small inductances can be realized by keeping both the self-inductance of the flux line and the geometry of the SQUID small, as illustrated in Figure 3(c).

It is also important to consider which waveforms we want to apply to the microwave control port. We would like to use microwave pulses to drive the qubit state around the surface of the Bloch sphere as quickly as possible without introducing errors. But, since a transmon qubit is a weakly anharmonic device, we need to be especially careful that our microwave pulses do not excite the ω_{12} transition. When engineering the qubit, there is a tradeoff between the dephasing time and anharmonicity, with a larger anharmonicity (a smaller C_Q) corresponding to more dephasing [30]. In practice, a typical value of $-\eta/2\pi$ is roughly 200 MHz. So carrying out fast XY pulses is quite challenging, since the relative energy at this offset frequency must be suppressed by more than 40 dB to prevent errors due to $|2\rangle$ population from being the dominant source of gate error. On the other hand, gates with a duration of approximately 10 ns are necessary to limit errors due to decoherence. To balance this tradeoff, XY pulses typically employ Gaussian or raised-cosine envelopes in conjunction with spectral-shaping techniques, such as Derivative Removal by Adiabatic Gate (DRAG) [38], [39], which can be employed to place a notch in the pulse's spectral content at f_{12}.

How strong should these pulses be? If we want to carry out an XY gate of duration Δt that rotates the qubit state by an angle, $\Delta\theta$, while using a raised-cosine envelope, the required peak power available from the generator at the reference plane shown in Figure 2(e) is

$$P_{AV} = (\hbar Q_D \Delta\theta^2)/(\Delta t^2). \tag{17}$$

It is important to note here that the rotation angle scales with amplitude as opposed to power since a qubit is a coherent device. Taking the value of Q_D given previously, we find that carrying out a π-pulse (corresponding to a 180° rotation) within 10 ns requires a peak available power of roughly -63 dBm, which is only a tiny fraction of the energy makes it to the qubit.

The signal-to-noise ratio (SNR) of the XY control signal is also an important consideration because noise on the XY driveline causes decoherence by driving small rotations about random axes on the Bloch sphere. The weak coupling of the XY line to the qubit helps reject noise, but there's still a limit to how much noise can be tolerated. The characteristic timescale at which the qubit state randomizes due to noise on the driveline is

$$\gamma_{\text{noise}}^{-1} = \frac{\hbar Q_D}{S_a}, \tag{18}$$

where S_a is the spectral density of available noise power at ω_{01}. This formula makes sense because more noise should cause more decoherence, and a higher Q_D means that the qubit is more weakly coupled to the XY driveline. How much attenuation do we need between our pulse generator and the qubit to sufficiently suppress generator noise? Recall that Q_D sets the decoherence rate due to resistive loading from the XY line. If we require $\gamma_{\text{noise}}^{-1}$ to be at least as large as $\gamma_{\text{loading}}^{-1}$, then we need $S_a < \hbar\omega_{10}$. Rephrasing in terms of the effective noise temperature looking back toward the generator from the drive port (T_{drive}), we need $T_{\text{drive}} < \hbar\omega_{10}/k = 48\ \text{mK} \times (\omega_{10}/2\pi)/\text{GHz}$. So, for a qubit frequency of 5 GHz, we need $T_{\text{drive}} < 240$ mK, which is roughly 30 dB lower than room temperature. A typical conservative approach is to attenuate the drive signal by a ratio of $T_{\text{GEN}}/T_{\text{BASE}}$, where T_{GEN} and T_{BASE} are the temperatures to which the generator and quantum integrated circuit are thermalized, respectively, with the attenuation distributed along the thermal gradient to balance the thermal noise of the attenuators with heating due to dissipation. This approach assumes that the noise floor of the generator is thermally limited, and additional attenuation may be required if the pulse generator has a higher

equivalent output noise floor, as is typically the case. Assuming a 30-mK base temperature and a thermally limited pulse generator at room temperature, this approach would require 40 dB of attenuation, meaning that pulses with a peak power of roughly –23 dBm must be generated at room temperature. This number grows proportionally as the generator noise floor increases.

Two-Qubit Gates

In addition to controlling the state of individual qubits, we also need to interact pairs of qubits in a well-controlled manner to implement two-qubit gates. There are many possible two-qubit gate operations; to understand how one can be implemented, we consider the circuit shown in Figure 5, which consists of a pair of frequency-tunable qubits coupled through a capacitance. Imagine that the qubits are tuned to be at the same frequency and that both are initially in the same state (e.g., $|\psi\rangle = |00\rangle$ or $|\psi\rangle = |11\rangle$). This is analogous to a common mode in classical electronics, so no energy is exchanged between the qubits. On the other hand, if only one of the qubits is initially in the excited state, energy will periodically swap back and forth between the qubits. The rate of the energy exchange will depend on the coupling strength, which is set by the relative size of C_C with respect to the qubit capacitances. This interaction can be turned on and off by tuning the relative qubit frequencies using the Z control line so that the qubits are on- and off-resonance, respectively. Thus, we can perform deterministic operations by enabling the interaction for a controlled duration. Gates leveraging this form of interaction are referred to as i*SWAP* gates and are characterized by a time evolution operator (recall, a two-qubit gate operates on a state vector of the form $|\psi\rangle = [\alpha_{00} \;\; \alpha_{01} \;\; \alpha_{10} \;\; \alpha_{11}]^{\mathrm{T}}$),

$$\hat{\mathrm{U}}_{i\mathrm{SWAP}}(\omega_s \Delta t) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\omega_s \Delta t/2) & -j\sin(\omega_s \Delta t/2) & 0 \\ 0 & -j\sin(\omega_s \Delta t/2) & \cos(\omega_s \Delta t/2) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{19}$$

where the parameter ω_s describes the rate at which energy moves back and forth between the two qubits.

If we enable the interaction for exactly half of the exchange cycle (i.e., $\omega_s \Delta t = \pi$), the time evolution operator becomes

$$\hat{\mathrm{U}}_{i\mathrm{SWAP}}(\pi) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -j & 0 \\ 0 & -j & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{20}$$

This swaps the $|10\rangle$ and $|01\rangle$ amplitudes while leaving $|00\rangle$ and $|11\rangle$ unchanged, as swapping does nothing if the two qubits are in the same initial state. The $|01\rangle$ and $|10\rangle$ amplitudes also experience a –90° phase shift, which creates entanglement. It can be shown that this gate satisfies the requirements for universal quantum computing [40].

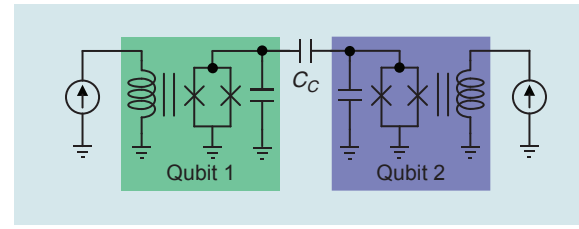


Figure 5. *Capacitively coupled qubits.*

While just one two-qubit gate is required for universal quantum computing, it is desirable to minimize the steps in an algorithm. In this respect, it is often advantageous to employ other gates in the family described by (19) as well as controlled Z [41] and controlled NOT [42] gates, which can be realized through flux tuning and a cross-resonant microwave drive, respectively.

Putting It All Together

Now that we understand how to implement the gates in a quantum algorithm, we can consider the simple gate sequence shown in Figure 6, which is used to generate a maximally entangled two-qubit state. To begin, the qubits are initialized to the $|00\rangle$ state and detuned in frequency. Next the XY line of each qubit is excited by a $\pi/2$-pulse (90° rotation), with qubit 1 rotating about the $-y$-axis and qubit 2 rotating about the $+y$-axis, resulting in a superposition state $|\psi\rangle = 0.5(|00\rangle - j|01\rangle + j|10\rangle + |11\rangle)$. This state is the equivalent of uniform noise in that, if a measurement were carried out, any of the four possible outcomes would be equally likely.

Next, the qubits are momentarily brought on-resonance to perform an $i\mathrm{SWAP}(\pi)$ gate, resulting in state $0.5(|00\rangle - |01\rangle + |10\rangle + |11\rangle)$. At this point, if we were to measure qubit 1, the state of qubit 2 would depend upon the measurement outcome. That is, if we found qubit 1 in the $|0\rangle$ state, qubit 2 would collapse to $(1/\sqrt{2})(|0\rangle + |1\rangle)$, whereas, if we found qubit 1 in the $|1\rangle$ state, qubit 2 would collapse to $-(1/\sqrt{2})(|0\rangle - |1\rangle)$. Therefore, the *i*SWAP gate has entangled the two qubits.

Finally, we can apply a $\pi/2$-pulse about the x-axis to qubit 1 to interfere with the $|01\rangle$ and $|10\rangle$ states, resulting in the final state $|\psi\rangle = (1/\sqrt{2})(|00\rangle + |11\rangle)$. This is a pretty interesting state. If we were to measure either qubit 1 or 2, we would find $|0\rangle$ with a 50% probability and $|1\rangle$ with a 50% probability. However, once we measured the first qubit, we would then know the state of the other qubit with a 100% probability. This entanglement, or correlation, is unique to quantum systems, and it is the essential feature that enables one to build

up computational spaces that are exponential with respect to the number of qubits.

Readout

Our fifth and final requirement is that we must be able to perform measurements of individual qubits. As mentioned, such measurements are nominally projective, meaning that at the end of a one, the state of a qubit that was initially $\alpha|0\rangle + \beta|1\rangle$ becomes $|0\rangle$ with probability $|\alpha|^2$ and $|1\rangle$ with probability $|\beta|^2$, and the evaluation tells us which of these outcomes has occurred. For practical quantum algorithms, we would like our readout technique to provide error rates of 1% or lower. How can we do this?

Let's consider the requirements. First, we need a physical parameter of the qubit whose value in the $|0\rangle$ and $|1\rangle$ states differs enough that measuring that property distinguishes the states with a high probability. In conventional DRAM, this parameter is the charge of a storage capacitance. Similarly, each superconducting qubit type has different electric and magnetic properties available for measurement. Second, and critically, we must be able to selectively activate the measurement: qubits do their thing (superposition, entanglement, and so on) only in private. Therefore, our measurement apparatus must not be extracting information, either intentionally or through parasitic losses, while the qubits are being controlled for logic gates.

What characteristic of the qubit $|0\rangle$ or $|1\rangle$ states can be used to distinguish them during measurement? For the transmon qubit, one choice of a physical parameter is the qubit's energy, since the states $|0\rangle$ and $|1\rangle$ are precisely those where the qubit has definite and different energies. However, this amounts to detecting the presence (if the qubit is in $|1\rangle$) or absence (if the qubit is in $|0\rangle$) of a single ~5-GHz microwave photon carrying an energy of about 20 μeV, and conventional photodetection technology is far too insensitive. There is another parameter that does differ between the two qubit states. Because the transmon is a nonlinear resonator its admittance depends on its oscillation amplitude. So one can determine the state of a qubit by measuring its reflection coefficient.

A generic approach is shown in Figure 7(a). A nonlinear resonator (the qubit) whose admittance is state-dependent is connected via a passive network, $Y(\omega)$, to a load that represents the measuring device. The goal is to design $Y(\omega)$ such that a probe signal scattering off the network can robustly distinguish between the two qubit states while preserving the qubit's internal Q-factor at ω_{01}. Naively, one could imagine connecting the qubit through a capacitor C_κ to a transmission line and using the qubit to scatter a microwave probe pulse, as illustrated in Figure 7(b). Note that a circulator has been employed to separate the forward (a_1) and reflected (b_1) waves. The admittance difference between the two quantum states could, then, be detected as a phase shift in the reflected signal (while we demonstrate the essential concepts here using a reflection measurement setup, in many cases state readout can also be performed with a transmission measurement).

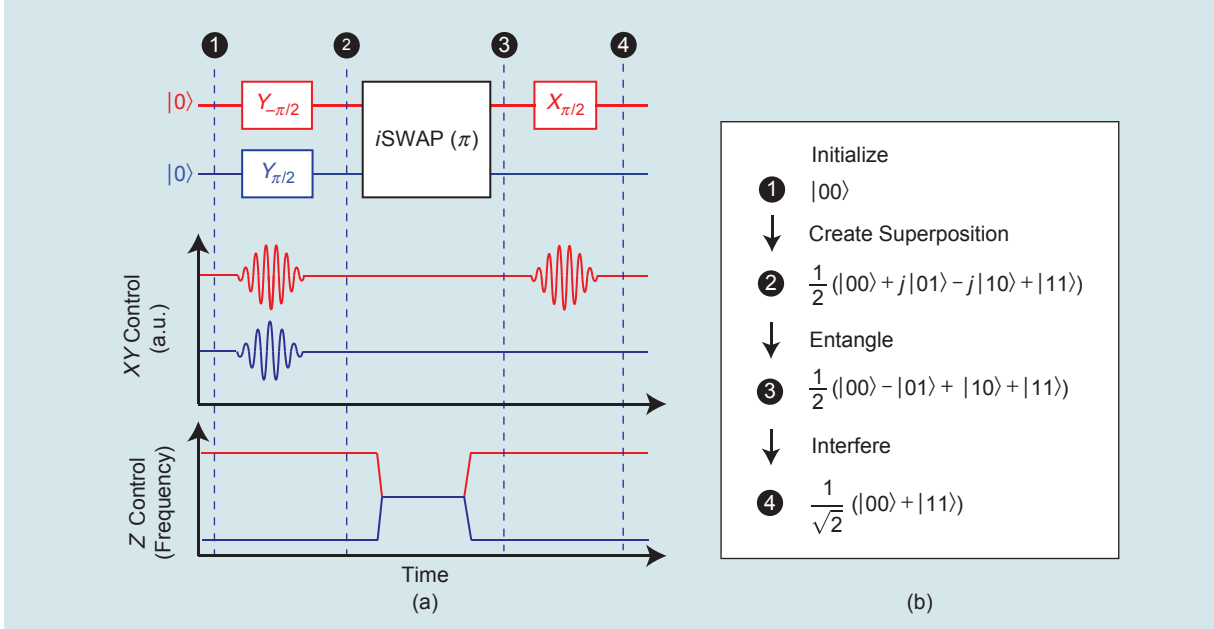


Figure 6. *An example quantum algorithm used to generate a maximally entangled pair. (a) The quantum circuit diagram and associated waveforms. (b) The effect of each step of the quantum algorithm. a.u.: arbitrary units.*

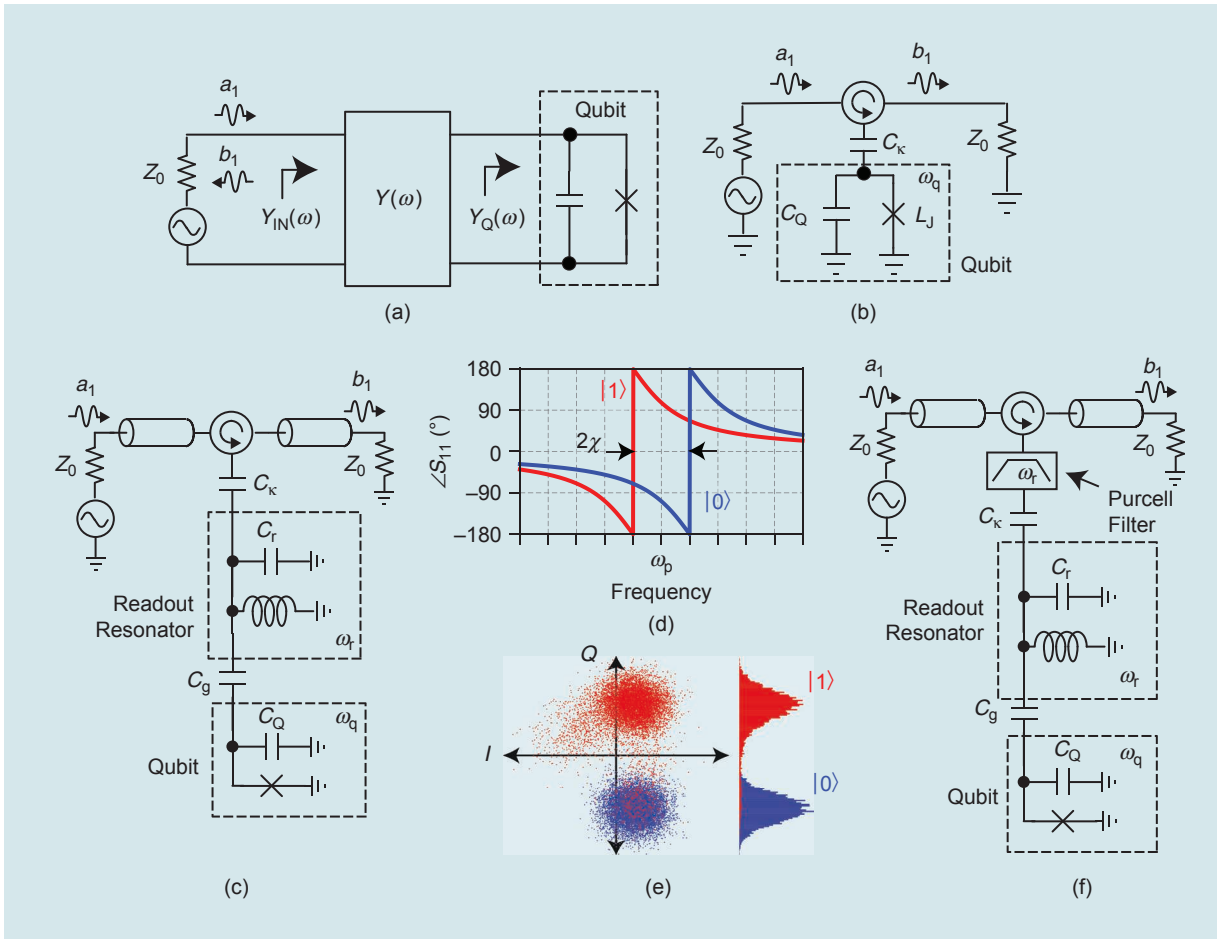


Figure 7. *Qubit readout as a microwave design problem. (a) A qubit is generically coupled to a load via an admittance $Y(\omega)$. (b) Directly coupled measurement in reflection. (c) Dispersive readout using a detuned resonator. (d) Example phase shifts associated with the circuit in (c). (e) Typical experimentally measured constellations. Each data point was acquired by initializing the qubit to the $|0\rangle$ (blue) or $|1\rangle$ state (red) and then performing a state measurement. (f) A typical readout system, including a Purcell filter.*

However, the circuit shown in Figure 7(b) has a fundamental limitation. The time it takes for the probe pulse to interact with the qubit and acquire an amplitude and phase shift is characterized by a ring-up rate, $\kappa = \omega_q/Q_m$, where $Q_m = (C_Q/C_\kappa)^2(Z_Q/Z_0)$ is the contribution to the loaded Q-factor of the qubit associated with its coupling to the transmission line. To make the measurement fast, we want a high κ, corresponding to a strong coupling between the qubit and transmission line, i.e., a low Q_m. However, in this circuit, the qubit's energy can leak through C_κ into the transmission line; in fact, it does so with precisely the same timescale, so $T_1 \approx \kappa^{-1}$. In other words, the measurement ring-up time is also the ring-down time of the qubit itself. Therefore, in this circuit, the qubit lifetime and measurement time are necessarily equal, which precludes any possibility of a high-accuracy readout because the qubit has a high probability to decay while we're measuring it.

To circumvent this problem, a typical readout system is configured as in Figure 7(c) and contains an additional linear readout resonator that is coupled to the readout line through capacitor C_κ. The readout resonator has resonance frequency $\omega_r = \omega_{01} \pm \Delta$, where $\Delta/2\pi$ is typically between 500 MHz and 1.5 GHz. The qubit and the readout resonator interact dispersively, meaning that the qubit pulls the readout resonator frequency. Because of the qubit nonlinearity, this frequency pull is different in the two states by an amount called the *dispersive shift*, 2χ, where

$$\chi = \frac{g^2}{\Delta}\left(\frac{\eta}{\Delta+\eta}\right)\left(\frac{\omega_r}{\omega_{10}}\right) \tag{21}$$

and $g \approx (C_g\sqrt{\omega_{01}\omega_r})/(2\sqrt{C_Q C_r})$ is the strength of the coupling between the qubit and the resonator through C_g [30], [43], [44].

The readout resonator can now be coupled relatively strongly to the measurement line through C_κ for fast ring-up and measurement times. Since the qubit frequency is detuned from the readout resonator by many linewidths, the readout resonator effectively shorts the resistive load presented by the readout line, and T_1 can remain much larger than T_m. In this circuit, the ring-up rate, κ, and the qubit lifetime, T_1, are constrained by $\kappa T_1 \lesssim \Delta/\chi$. In addition to speed, the measurement result also depends on the phase contrast between the $|0\rangle$ and $|1\rangle$ states. If we choose the dispersive shift to be roughly equal to the readout resonator linewidth, $2\chi \approx \kappa$ [30], and probe between the $|0\rangle$ and $|1\rangle$ state readout resonator frequencies, a phase contrast as high as 180° can be realized, as shown in Figure 7(d) and (e), and the state can be determined using demodulation techniques similar to those used for binary phase-shift keying.

For an optimal design achieving a 180° phase contrast, $\kappa^2 T_1 \lesssim \Delta$, detuning the readout resonator from the qubit eases the tradeoff between T_1 and κ [45]. To get a feel for the numbers, suppose we want to measure the transmon in 500 ns. Assuming that we need the ring-up time to be, at most, half the total measurement time, the readout circuit would limit the qubit's lifetime to $T_1 \lesssim (1\text{ GHz}) \times (250\text{ ns})^2 = 62.5\ \mu\text{s}$. So we have made the qubit lifetime 125 times larger than the measurement time, which is sufficient for research experiments, but we will still have a ~1% readout error due to the qubit relaxing during measurement. Additionally, the limit of T_1 due to loading by the readout circuit is shorter than the state-of-the-art limits due to material losses, so the setup in Figure 7(c) leaves performance on the table.

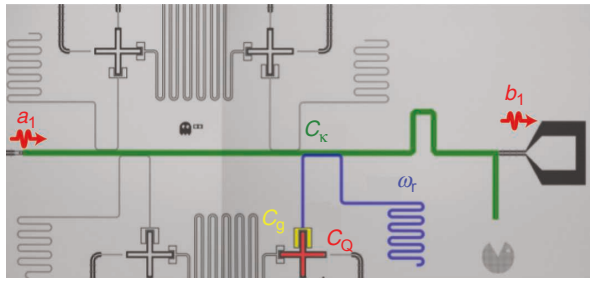


Figure 8. *A four-transmon device with individual readout resonators (blue) coupled to a common Purcell filter (green). (Source: [48]; used with permission.) In this case, the readout is accomplished via a transmission measurement. The input probe signal (a_1) is injected on the left, and the output signal (b_1) is taken from the right.*

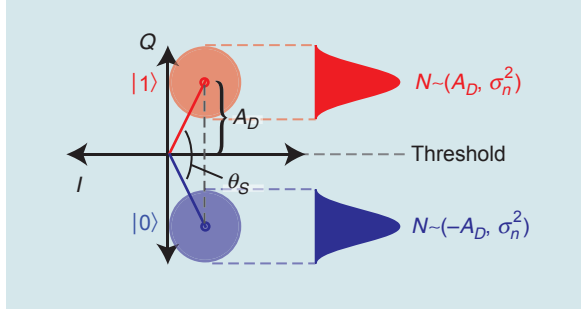


Figure 9. *The geometry for readout SNR calculation.*

Dispersive measurement with the architecture of Figure 7(c) was standard for transmon qubits between 2005 and 2010. However, as superconducting qubit coherence times increased and gate error rates decreased, the readout accuracy had to keep up. To decrease the number of errors stemming from qubit relaxation due to loading by the readout circuit and enable the readout of multiple qubits using frequency domain multiplexing, it is now common to further shape the admittance function $Y(\omega)$ in Figure 7(a). This can be done by adding additional transmission zeros at the qubit frequency [33], incorporating secondary readout resonators [47], or including a bandpass filter near the readout resonator frequency, with strong rejection at the qubit frequency, as in Figure 7(f) [48], [49]. Such a filter is referred to as a *Purcell filter* in the superconducting qubit literature. Figure 8 shows a four-transmon device [48], with each transmon (one of whose capacitance, C_q, is highlighted in red) connected to a CPW readout resonator (blue) via capacitance C_g (yellow). The four readout resonators are coupled to a common quarter-wave resonator that serves as a Purcell filter (green). In this example, the readout probe tone is injected via a small capacitor at the left of the photograph and transmitted to a bond pad at the right.

Having explained how the state of a qubit can be read out through a dispersive measurement of an ancillary linear resonator, we ask the question, Just how easy or hard is it to actually do this? Similar to the design of a communication system, we need to understand how the error probability depends on design parameters. Assume that the reflected signal we are trying to measure has amplitude A_S and root-mean-square noise voltage σ_n within some measurement bandwidth and that this signal takes on two distinct phases, $\pm\theta_S/2$, corresponding to the $|0\rangle$ and $|1\rangle$ states. The measurement will consist of distinguishing between two points in the in-phase/quadrature (IQ) plane, as shown in Figure 9. The error probability given this geometry is [50]

$$P_{\text{error}} = \frac{1}{2}\text{erfc}\left(\frac{A_D}{\sqrt{2}\,\sigma_n}\right) = \frac{1}{2}\text{erfc}\left(\sin\left(\frac{\theta}{2}\right)\sqrt{\text{SNR}}\right), \quad (22)$$

where SNR is the RF SNR in the measurement bandwidth.

We can now express P_{error} in terms of physical parameters. Assuming an ideal measurement, the RF

signal power referenced to the readout resonator port can be related to the energy stored in the resonator as $P_{RF} = E\kappa/2$, so $P_{RF} = \bar{n}\hbar\omega_P\kappa/2$, where ω_P is the measurement frequency and $\bar{n}$ is the average number of photons populating the readout resonator. In practice, this is a dynamic measurement that is carried out before the resonator has a chance to settle. Additionally, photons are lost since the measurement has finite quantum efficiency. To compensate for these facts, we can introduce a loss factor, ϵ, giving $P_{RF} = \epsilon\bar{n}\hbar\omega_P\kappa/2$. Typically, ϵ is in the range of 10–20%. The system noise is $P_{noise} = kT_{SYS}/\tau = (\hbar\omega_P/2\tau)(1 + T_{RX}/T_Q)$, where T_{RX} is the input-referred noise temperature of the measurement receiver and $T_Q = \hbar\omega_P/2k$ is the quantum limit for noise added by a phase-preserving amplifier. The error probability is thus

$$P_{error} = \frac{1}{2}\text{erfc}\left(\sin\left(\frac{\theta_S}{2}\right)\sqrt{\frac{\epsilon\bar{n}\kappa\tau}{1 + T_{RX}/T_Q}}\right). \tag{23}$$

As such, error rates improve if we use more photons, integrate longer, improve the phase contrast, or decrease T_{RX}. However, in optimizing the measurement, the noise temperature is constrained by the quantum limit $T_{RX} \geq T_Q$, and the integration time must be kept low enough such that $|1\rangle$-state readout errors due to relaxation are insignificant ($P_{error,relaxation} \approx \tau/2T_1$).

An obvious approach to improving the readout SNR is to drive the readout resonator harder. However, the interaction goes both ways, and the average number of microwave photons $\bar{n}$ "stored" in the resonator affects the qubit frequency as well. When the nonlinear frequency pull on the qubit due to the field intensity in the resonator (also known as the *ac Stark shift* $\Delta\omega_{01} = 2\chi\,\bar{n}_r$) becomes comparable to the qubit's nonlinearity (η), we start breaking down the distinct energy-level structure that defines the qubit, and the qubit becomes essentially useless. For typical transmon parameters, this can already happen when the resonator stores just $\bar{n} \sim 25$ photons. Even worse, at lower powers, the drive can already cause transitions between the qubit states [51]. For this reason, the readout resonator drive power is typically limited such that the resonator is populated with only a handful of photons ($\bar{n} \lesssim 10$). The power associated with populating the resonator with $\bar{n}$ photons is $P = \hbar\omega_P\bar{n}\kappa/2$; for typical design parameters, this is on the order of just 30 aW per photon. Since the energy in the resonator can be, at most, roughly 10 photons, our peak drive power can be no more than approximately –125 dBm, referenced to the resonator drive port.

Our final degree of freedom is T_{RX}, so let's try to understand the requirements to achieve error rates of 1%. Let's assume that we want to limit τ to 300 ns to avoid excessive errors due to relaxation during the measurement.

Superconductive processors have advanced tremendously, culminating in the recent demonstration of quantum supremacy.

Taking $\epsilon = 0.2$, $\kappa = 2\pi \times 3$ MHz, $\bar{n} = 8$ photons, $\theta_S = 180^\circ$ and $\omega_P = 5$ GHz, we require a receiver noise temperature of roughly 0.56 K, which is approximately twice the quantum limit. Needless to say, achieving this level of performance is a very challenging task.

State-of-the art cryogenic semiconductor amplifiers still have approximately four times too much noise for this application [52]. Moreover, even if the noise of semiconductor low-noise amplifiers could be brought down to T_Q, the devices still dissipate far too much power to be heatsunk to the 10-mK stage of a dilution refrigeration system. Therefore, we need to use something else if we are going to achieve the system sensitivity required to perform one-shot dispersive readout of transmon qubits. In fact, it turns out that the performance we want for transmon qubit readout became possible only during the past 10–15 years through the development of near-quantum-limited Josephson parametric amplifiers [53]–[56], which are currently an essential element that enables high-speed, one-shot dispersive readout of transmon qubits. For more details related to parametric amplifiers, we refer the reader to a review article on this topic, appearing elsewhere in this focus issue [57]. Of course, semiconductor low-noise amplifiers are also required in the readout chain as second-stage amplifiers, since nearly 60 dB of cryogenic gain is required to overcome practical room-temperature noise floors and parametric amplifiers with the requisite linearity are not available.

The Road to Practical Applications

To review, a quantum processor can be built using nonlinear microwave resonators whose quantum state is controlled using a combination of microwave pulses and time-varying bias signals and measured using microwave reflectometry. During the past decade, these superconductive processors have advanced tremendously, culminating in the recent demonstration of quantum supremacy: the ability to perform a computation that is impractical even for today's most powerful of supercomputers [5]. However, today's quantum processors are still limited.

With a regular digital computer, error rates are so low that we can write software while assuming that the bits never suffer physical errors. Conversely, superconducting qubits suffer relatively high error rates, meaning that they can perform only a finite number of logic operations before the quantum state randomizes.

For example, a transmon with an energy decay constant of $T_1 = 50\ \mu$s and a two-qubit gate time of 20 ns can undergo only 2,500 gates before it has a high probability of suffering an error. With such high error rates, how can a quantum computer actually complete a full algorithm? There are currently two main approaches.

In the near term, it is assumed that errors are unavoidable, and researchers are currently considering how to conduct useful quantum computation with a modest number of imperfect qubits. The key thing here is that the algorithms must be short compared to the qubits' coherence times. This is tricky, though: if the quantum algorithm is too short, it can be simulated on a conventional computer, obviating the need for the quantum computer. The most promising uses seem to be in simulating chemistry [58]–[62]. Applications in this regime of so-called noisy intermediate-scale quantum (NISQ) computation [63] constitute an area of active research.

However, the long-term goal of the community is to develop fault-tolerant quantum computers. Much in the same way that redundancy is used in conventional error-correcting code memory to lower error rates, it is theoretically possible to combine many imperfect physical qubits into a single logical qubit whose error rate is lowered exponentially in the number of constituent physical qubits by using quantum error correction. Quantum error correction is somewhat more subtle than conventional error correction because quantum states cannot be copied (this fact is called the *no-cloning theorem*). This limitation makes the subject of quantum error-correcting codes somewhat more delicate and richer than conventional error-correction codes. A computer operating with logical qubits is called *fault-tolerant* because it continues to work without logical errors, even when individual physical qubits suffer faults. According to current estimates based on today's best physical qubit systems, a single logical qubit with low-enough error rates to be used in a fault-tolerant computer would need roughly 1,000 constituent physical qubits, and a fault-tolerant computer would need approximately 1,000 logical qubits, necessitating roughly 1 million physical qubits [64].

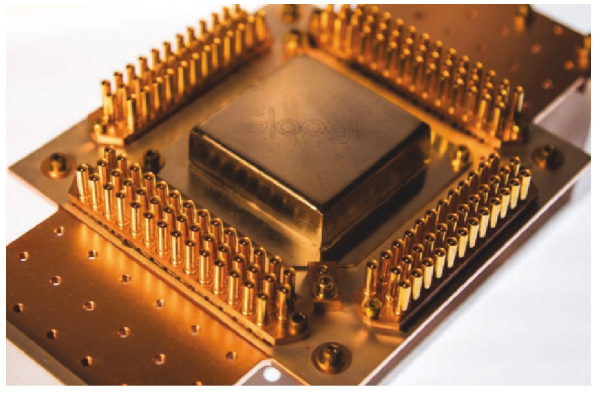


Figure 10. *The package used in a recent quantum computing experiment at Google. (Source: [5]; used with permission.) Vertical launch connectors around the perimeter connect to coaxial cabling and a printed circuit board (PCB). The PCB is covered by a magnetic shield (the square box in the center).*

Let's take stock of what that means in terms of control electronics, wiring, and packaging. In today's state-of-the-art technology [5], each qubit is driven by dedicated on-chip lines that carry microwave signals for XY rotations and 500-MHz dc signals for Z rotations. The XY pulses are generated by a pair of high-speed (1 gigasample per second) digital-to-analog converters (DACs) driving an IQ mixer, while the Z pulses are generated by another high-speed DAC, so we need three DACs per qubit. The signals are carried from the pulse generators to individual qubits through dedicated coaxial cables and CPW interconnects. The cables interface to the package via blind-mate connectors, and the package connects to the chip via wire bonds (see Figure 10). Can these technologies be scaled to a 1-million-qubit computer? For the control system, this would mean at least 3 million DACs and millions of coaxial cables running into a cryostat holding a chip package with millions of blind-mate connectors. This is where microwave engineering really comes in: the cost, size, thermal, assembly, reliability, and signal integrity concerns arising in a system with that much microwave hardware demand innovations in the microwave domain.

Let us consider the optimization of the microwave interface required to control a million-qubit system. If the control system used in the recent quantum supremacy experiment at Google [5] were scaled to control a 1-million-qubit system, it would occupy three football fields of floor space and consume roughly 40 MW of dc power, not to mention the significant power that would be required to compensate for losses associated with signal distribution. Development of integrated control systems specifically tailored toward this application [65]–[68] is an important area for future research that could impact quantum computing in the way that integrated beamformers have enabled more scalable phased arrays [69].

Innovation is also required in the interconnections used in quantum computing systems. Control signals are currently brought from the electronics into the cryostat and down to the quantum chip through coaxial cable, which raises several important questions. How do we connect the cables to the package? The connection has to have a low loss to avoid thermally loading the 20-mK stage of the cryostat. Dilution refrigerators can cool to approximately 10 μW at 20 mK, so, assuming that we have one dilution cooler per 65,000 qubits and that we need to source roughly

300 μA of dc current per qubit, the connector at the 20-mK stage must have a contact resistance of $\lesssim 2\ \text{m}\Omega$ per line. The microwave connector also must have low crosstalk despite its high density. Developing a cost-effective connector with a low-enough profile while maintaining acceptable channel-to-channel crosstalk is another critical step toward a fault-tolerant quantum computer. Another important concern for the cables themselves is the thermal load they bring into the cryostat. Can we find a wiring solution that is dense and cost effective and that will not overwhelm the cryostat's cooling-power budget? As we go from systems with 100 to 1,000 qubits, these questions will need to be answered.

Next, we can consider the quantum chip itself. The typical pitch of 2D grids of transmon qubits is approximately 1 mm, so a notional grid of 1 million qubits might therefore cover a square measuring 1 m on a side. Such a device will not be composed of a single silicon chip. Instead, many smaller chips will need to be stitched together. Multichip assembly and 3D integration technologies will have to be adapted to the material sets, loss and crosstalk requirements, cryogenic operation, and layout constraints imposed by this application. The chip-to-chip interconnect will need to have exquisitely low loss and low crosstalk while maintaining the coupling strength needed to do fast qubit–qubit logic gates. Electromagnetic, thermal, and mechanical simulations of the packaged assembly will become crucial. Parasitic resonances and package modes in assemblies of the envisioned size could overlap the qubit frequencies, and their coupling to the qubits will have to be understood and mitigated. These are just a few of the many areas in which microwave engineering will be required in the quest to implement a fault-tolerant quantum computer. As the systems progress, it will be essential that more microwave engineers engage in these efforts.

Acknowledgment

This work was partially supported by National Science Foundation grant 1809114.

References

[1] E. Grumbling and M. Horowitz, Eds., *Quantum Computing: Progress and Prospects*. Washington, D.C.: The National Academies Press, 2019. [Online]. Available: https://www.nap.edu/catalog/25196/quantum-computing-progress-and-prospects

[2] M. Berboucha, "IBM's first commercial quantum computer," *Forbes*, Jan. 23, 2019. [Online]. Available: https://www.forbes.com/sites/meriameberboucha/2019/01/23/ibms-first-commercial-quantum-computer/#cf020a441b2a

[3] F. Lardinois, "Microsoft open-sources its quantum computing development tools," TechCrunch, San Francisco, May 6, 2019. [Online]. Available: https://techcrunch.com/2019/05/06/microsoft-open-sources-its-quantum-computing-development-tools/

[4] J. Hsu, "Google's quantum tech milestone excites scientists and spurs rivals," *IEEE Spectrum*, 2019. [Online]. Available: https://spectrum.ieee.org/tech-talk/computing/hardware/googles-quantum-tech-milestone-excites-scientists-and-spurs-rivals

[5] F. Arute et al., "Quantum supremacy using a programmable superconducting processor," *Nature*, vol. 574, pp. 505–510, Oct. 2019. doi: 10.1038/s41586-019-1666-5.

[6] Y. Sun, "Why Alibaba is betting big on AI chips and quantum computing," *MIT Technology Review*, Cambridge, MA, Sept. 25, 2018. [Online]. Available: https://www.technologyreview.com/s/612190/why-alibaba-is-investing-in-ai-chips-and-quantum-computing/

[7] S. K. Moore and A. Nordrum, "Intel's new path to quantum computing," *IEEE Spectrum*, New York, June 8, 2018. [Online]. Available: https://spectrum.ieee.org/nanoclast/computing/hardware/intels-new-path-to-quantum-computing

[8] M. F. Riedel, D. Binosi, R. Thew, and T. Calarco, "The European quantum technologies flagship programme," *Quantum Sci. Technol.*, vol. 2, no. 3, p. 030501, 2017. doi: 10.1088/2058-9565/aa6aca.

[9] T. Roberson and A. White, "Charting the Australian quantum landscape," *Quantum Sci. Technol.*, vol. 4, no. 2, p. 020505, 2019. doi: 10.1088/2058-9565/ab02b4.

[10] M. Giles, "The man turning China into a quantum superpower," *MIT Technology Review*, Cambridge, MA, Dec. 19, 2018. [Online]. Available: https://www.technologyreview.com/s/612596/the-man-turning-china-into-a-quantum-superpower/

[11] Y. Yamamoto, M. Sasaki, and H. Takesue, "Quantum information science and technology in japan," *Quantum Sci. Technol.*, vol. 4, no. 2, p. 020502, 2019. doi: 10.1088/2058-9565/ab0077.

[12] B. Sussman, P. Corkum, A. Blais, D. Cory, and A. Damascelli, "Quantum Canada," *Quantum Sci. Technol.*, vol. 4, no. 2, p. 020503, 2019. doi: 10.1088/2058-9565/ab029d.

[13] Q. Schiermeier, "Russia joins race to make quantum dreams a reality," *Nature*, vol. 577, no. 7788, p. 14, 2020. doi: 10.1038/d41586-019-03855-z. [Online]. Available: https://www.nature.com/articles/d41586-019-03855-z

[14] M. G. Raymer and C. Monroe, "The US national quantum initiative," *Quantum Sci. Technol.*, vol. 4, no. 2, p. 020504, 2019. doi: 10.1088/2058-9565/ab0441.

[15] "Private/startup companies," *Quantum Computing Report*, Tustin, CA. [Online]. Available: https://quantumcomputingreport.com/players/privatestartup/

[16] F. Lardinois, "Rigetti announces its hybrid quantum computing platform - and a $1m prize," *TechCrunch*, San Francisco, 2018. [Online]. Available: https://techcrunch.com/2018/09/07/rigetti-announces-its-hybrid-quantum-computing-platform-and-a-1m-prize/

[17] J. Russel, "Quantum upstart: IonQ sets sights on challenging IBM, Rigetti, others," *HPCwire*, San Diego, CA, 2019. [Online]. Available: https://www.hpcwire.com/2019/06/05/quantum-upstart-ionq-sets-sights-on-challenging-ibm-rigetti-others/

[18] M. Murgia, "British quantum computing experts leave for silicon valley," *Financial Times*, 2019. [Online]. Available: https://www.ft.com/content/afc27836-9383-11e9-aea1-2b1d33ac3271

[19] C. Leddy, "Q&A: The talent shortage in quantum computing," *MIT News*, Cambridge, MA, 2019. [Online]. Available: http://news.mit.edu/2019/mit-william-oliver-qanda-talent-shortage-quantum-computing-0123

[20] S. Lloyd, "Computational capacity of the universe," *Physical Rev. Lett.*, vol. 88, no. 23, p. 237901, 2002. doi: 10.1103/PhysRevLett.88.237901.

[21] A. Montanaro, "Quantum algorithms: An overview," *npj Quantum Inform.*, vol. 2, p. 15023, Jan. 2016. doi: 10.1038/npjqi.2015.23.

[22] U. Vool and M. Devoret, "Introduction to quantum electromagnetic circuits," *Int. J. Circuit Theory Appl.*, vol. 45, no. 7, pp. 897–934, 2017. doi: 10.1002/cta.2359.

[23] P. Krantz, M. Kjaergaard, F. Yan, T. P. Orlando, S. Gustavsson, and W. D. Oliver, "A quantum engineer's guide to superconducting qubits," *Appl. Phys. Rev.*, vol. 6, no. 2, p. 021318, 2019. doi: 10.1063/1.5089550.

[24] D. P. DiVincenzo, "The physical implementation of quantum computation," *Fortschritte der Physik: Progr. Phys.*, vol. 48, no. 9–11, pp. 771–783, 2000. doi: 10.1002/1521-3978(200009)48:9/11<771::AID-PROP771>3.0.CO;2-E.

[25] D. P. DiVincenzo, "Two-bit gates are universal for quantum computation," *Phys. Rev. A*, vol. 51, no. 2, p. 1015, 1995. doi: 10.1103/PhysRevA.51.1015.

[26] C. D. Bruzewicz, J. Chiaverini, R. McConnell, and J. M. Sage, "Trapped-ion quantum computing: Progress and challenges," *Appl. Phys. Rev.*, vol. 6, no. 2, p. 021314, 2019. doi: 10.1063/1.5088164.
[27] J. L. O'brien, A. Furusawa, and J. Vučković, "Photonic quantum technologies," *Nat. Photon.*, vol. 3, no. 12, p. 687, 2009. doi: 10.1038/nphoton.2009.229.
[28] M. Saffman, "Quantum computing with atomic qubits and Rydberg interactions: Progress and challenges," *J. Phys. B, Atom Mol. Opt. Phys.*, vol. 49, no. 20, p. 202001, Oct. 2016. doi: 10.1088/0953-4075/49/20/202001.
[29] A. S. Dzurak, "Spin-based quantum computing in silicon CMOS-compatible platforms," in *Proc. IEEE Int. Electron Devices Meeting (IEDM)*, Dec. 2016, pp. 13.2.1–13.2.2. doi: 10.1109/IEDM.2016.7838407.
[30] J. Koch et al., "Charge-insensitive qubit design derived from the cooper pair box," *Phys. Rev. A*, vol. 76, no. 4, p. 042319, 2007. doi: 10.1103/PhysRevA.76.042319.
[31] Z. Chen, "Metrology of quantum control and measurement in superconducting qubits," Ph.D. dissertation, UC Santa Barbara, 2018.
[32] P. Magnard et al., "Fast and unconditional all-microwave reset of a superconducting qubit," *Phys. Rev. Lett.*, vol. 121, no. 6, p. 060502, Aug. 2018. doi: 10.1103/PhysRevLett.121.060502.
[33] M. D. Reed et al., "Fast reset and suppressing spontaneous emission of a superconducting qubit," *Appl. Phys. Lett.*, vol. 96, no. 20, p. 203110, 2010. doi: 10.1063/1.3435463.
[34] M. Veldhorst et al., "An addressable quantum dot qubit with fault-tolerant control-fidelity," *Nat. Nanotechnol.*, vol. 9, no. 12, p. 981, 2014. doi: 10.1038/nnano.2014.216.
[35] R. Shankar, *Principles of Quantum Mechanics*, New York: Springer-Verlag, 2012.
[36] E. Merzbacher, *Quantum Mechanics*, 3rd ed. New York: Wiley, 1998.
[37] C. Cohen-Tannoudji, B. Diu, and F. Laloë, *Quantum Mechanics*, vol. 1, Paris: Hermann and John Wiley & Sons, 1977.
[38] F. Motzoi, J. M. Gambetta, P. Rebentrost, and F. K. Wilhelm, "Simple pulses for elimination of leakage in weakly nonlinear qubits," *Phys. Rev. Lett.*, vol. 103, no. 11, p. 110501, 2009. doi: 10.1103/PhysRevLett.103.110501.
[39] J. M. Gambetta, F. Motzoi, S. Merkel, and F. K. Wilhelm, "Analytic control methods for high-fidelity unitary operations in a weakly nonlinear oscillator," *Phys. Rev. A*, vol. 83, no. 1, p. 012308, 2011. doi: 10.1103/PhysRevA.83.012308.
[40] P. Echternach, C. Williams, J. Dowling, S. Dultz, and S. Braunstein, "Universal quantum gates for single cooper pair box based quantum computing," *Quantum Inf. Comput.*, vol. 1, no. 4, pp. 143–150, 2001. doi: 10.5555/2016994.2017010.
[41] R. Barends et al., "Superconducting quantum circuits at the surface code threshold for fault tolerance," *Nature*, vol. 508, no. 7497, p. 500, 2014. doi: 10.1038/nature13171.
[42] G. Paraoanu, "Microwave-induced coupling of superconducting qubits," *Phys. Rev. B*, vol. 74, no. 14, p. 140504, 2006. doi: 10.1103/PhysRevB.74.140504.
[43] M. Khezri, "Dispersive measurement of superconducting qubits," Ph.D. dissertation, Univ. California, Riverside, 2018.
[44] T. White, "Preserving entanglement during weak measurement demonstrated with a violation of the Bell-Leggett-Garg inequality," Ph.D. dissertation, Univ. California, Santa Barbara, 2015.
[45] A. A. Houck et al., "Controlling the spontaneous emission of a superconducting transmon qubit," *Phys. Rev. Lett.*, vol. 101, no. 8, p. 080502, Aug. 2008. doi: 10.1103/PhysRevLett.101.080502.
[46] J. J. Sakuari, *Modern Quantum Mechanics*, 2nd ed. Pearson, 2010.
[47] J. Heinsoo et al., "Rapid high-fidelity multiplexed readout of superconducting qubits," *Phys. Rev. APPl.*, vol. 10, no. 3, p. 034040, Sept. 2018. doi: 10.1103/PhysRevApplied.10.034040.
[48] E. Jeffrey et al., "Fast accurate state measurement with superconducting qubits," *Phys. Rev. Lett.*, vol. 112, no. 19, p. 190504, May 2014. doi: 10.1103/PhysRevLett.112.190504.
[49] N. T. Bronn et al., "Broadband filters for abatement of spontaneous emission in circuit quantum electrodynamics," *Appl. Phys. Lett.*, vol. 107, no. 17, p. 172601, 2015. doi: 10.1063/1.4934867.
[50] B. Razavi, *Design of Integrated Circuits for Optical Communications*. Hoboken, NJ: Wiley, 2003.
[51] D. Sank et al., "Measurement-induced state transitions in a superconducting qubit: Beyond the rotating wave approximation," *Phys. Rev. Lett.*, vol. 117, no. 19, p. 190503, Nov. 2016. doi: 10.1103/PhysRevLett.117.190503.
[52] "4-8 GHz cryogenic low noise amplifier," Low Noise Factory, Gothenburg, Sweden, Jan. 2020. [Online]. Available: https://www.lownoisefactory.com/files/7015/7825/6000/LNF-LNC4_8C.pdf
[53] N. Bergeal et al., "Phase-preserving amplification near the quantum limit with a Josephson ring modulator," *Nature*, vol. 465, no. 7294, p. 64, 2010. doi: 10.1038/nature09035.
[54] B. Abdo, F. Schackert, M. Hatridge, C. Rigetti, and M. Devoret, "Josephson amplifier for qubit readout," *Appl. Phys. Lett.*, vol. 99, no. 16, p. 162506, 2011. doi: 10.1063/1.3653473.
[55] J. Y. Mutus et al., "Strong environmental coupling in a Josephson parametric amplifier," *Appl. Phys. Lett.*, vol. 104, no. 26, p. 263513, 2014. doi: 10.1063/1.4886408.
[56] C. Macklin et al., "A near–quantum-limited Josephson traveling-wave parametric amplifier," *Science*, vol. 350, no. 6258, pp. 307–310, 2015. doi: 10.1126/science.aaa8525. [Online]. Available: https://science.sciencemag.org/content/350/6258/307
[57] J. Aumentado, "Superconducting parametric amplifiers," *IEEE Microw. Mag.*, vol. 21, no. 8, pp. 45–59, Aug. 2020.
[58] Y. Cao et al., "Quantum chemistry in the age of quantum computing," *Chem. Rev.*, vol. 119, no. 19, pp. 10,856–10,915, 2019. doi: 10.1021/acs.chemrev.8b00803.
[59] P. J. J. O'Malley et al., "Scalable quantum simulation of molecular energies," *Phys. Rev. X*, vol. 6, no. 3, p. 031007, July 2016. doi: 10.1103/PhysRevX.6.031007. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevX.6.031007
[60] R. Babbush, N. Wiebe, J. McClean, J. McClain, H. Neven, and G. K.-L. Chan, "Low-depth quantum simulation of materials," *Phys. Rev. X*, vol. 8, no. 1, p. 011044, Mar. 2018. doi: 10.1103/PhysRevX.8.011044. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevX.8.011044
[61] I. D. Kivlichan et al., "Quantum simulation of electronic structure with linear depth and connectivity," *Phys. Rev. Lett.*, vol. 120, no. 11, p. 110501, Mar. 2018. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.120.110501 doi: 10.1103/PhysRevLett.120.110501.
[62] A. Kandala et al., "Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets," *Nature*, vol. 549, no. 7671, p. 242, 2017. doi: 10.1038/nature23879.
[63] J. Preskill, "Quantum Computing in the NISQ era and beyond," *Quantum*, vol. 2, p. 79, 2018. doi: 10.22331/q-2018-08-06-79.
[64] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, "Surface codes: Towards practical large-scale quantum computation," *Phys. Rev. A*, vol. 86, no. 3, p. 032324, 2012. doi: 10.1103/PhysRevA.86.032324.
[65] J. C. Bardin et al., "29.1 a 28nm bulk-CMOS 4-to-8GHz 2mW cryogenic pulse modulator for scalable quantum computing," in *Proc. 2019 IEEE Int. Solid-State Circuits Conf. (ISSCC)*, 2019, pp. 456–458. doi: 10.1109/ISSCC.2019.8662480.
[66] J. C. Bardin et al., "Design and characterization of a 28-nm bulk-CMOS cryogenic quantum controller dissipating less than 2 mW at 3 K," *IEEE J. Solid-State Circuits*, vol. 54, no. 11, pp. 3043–3060, 2019. doi: 10.1109/JSSC.2019.2937234.
[67] E. Charbon et al., "Cryo-CMOS circuits and systems for scalable quantum computing," in *Proc. 2017 IEEE Int. Solid-State Circuits Conf. (ISSCC)*, 2017, pp. 264–265. doi: 10.1109/ISSCC.2017.7870362.
[68] B. Patra et al., "Cryo-CMOS circuits and systems for quantum computing applications," *IEEE J. Solid-State Circuits*, vol. 53, no. 1, pp. 309–321, 2017. doi: 10.1109/JSSC.2017.2737549.
[69] B. Sadhu, X. Gu, and A. Valdes-Garcia, "The more (antennas), the merrier: A survey of silicon-based mm-wave phased arrays using multi-IC scaling," *IEEE Microw. Mag.*, vol. 20, no. 12, pp. 32–50, 2019. doi: 10.1109/MMM.2019.2941632.