A quantum engineer's guide to superconducting qubits

Downloaded from: https://research.chalmers.se, 2022-08-27 12:09 UTC

Citation for the original published paper (version of record):

N.B. When citing this work, cite the original published paper.
A quantum engineer's guide to superconducting qubits

Cite as: Appl. Phys. Rev. 6, 021318 (2019); https://doi.org/10.1063/1.5089550
Submitted: 20 January 2019. Accepted: 03 May 2019. Published Online: 17 June 2019

P. Krantz, M. Kjaergaard, F. Yan, T. P. Orlando, S. Gustavsson, and W. D. Oliver

COLLECTIONS

Note: This paper is part of the Special Topic on Quantum Computing.

ARTICLES YOU MAY BE INTERESTED IN

Trapped-ion quantum computing: Progress and challenges
Applied Physics Reviews 6, 021314 (2019); https://doi.org/10.1063/1.5088164

Machine learning meets quantum physics
Physics Today 72, 48 (2019); https://doi.org/10.1063/PT.3.4164

Planar superconducting resonators with internal quality factors above one million
Applied Physics Letters 100, 113510 (2012); https://doi.org/10.1063/1.3693409
A quantum engineer’s guide to superconducting qubits

Cite as: Appl. Phys. Rev. 6, 021318 (2019); doi: 10.1063/1.5089550
Submitted: 20 January 2019 · Accepted: 3 May 2019 ·
Published Online: 17 June 2019

P. Krantz,1,2,a) M. Kjaergaard,1 F. Yan,1 T. P. Orlando,1 S. Gustavsson,1 and W. D. Oliver1,3,b)

AFFILIATIONS
1Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
2Wallenberg Centre for Quantum Technology (WACQT), Chalmers University of Technology, Gothenburg SE-41296, Sweden
3MIT Lincoln Laboratory, 244 Wood Street, Lexington, Massachusetts 02420, USA

Note: This paper is part of the Special Topic on Quantum Computing.
a)Author to whom correspondence should be addressed: philipk@mit.edu
b)william.oliver@mit.edu

ABSTRACT
The aim of this review is to provide quantum engineers with an introductory guide to the central concepts and challenges in the rapidly accelerating field of superconducting quantum circuits. Over the past twenty years, the field has matured from a predominantly basic research endeavor to one that increasingly explores the engineering of larger-scale superconducting quantum systems. Here, we review several foundational elements—qubit design, noise properties, qubit control, and readout techniques—developed during this period, bridging fundamental concepts in circuit quantum electrodynamics and contemporary, state-of-the-art applications in gate-model quantum computation.

Published under license by AIP Publishing. https://doi.org/10.1063/1.5089550

TABLE OF CONTENTS
I. INTRODUCTION ............................................. 2
A. Organization of this article ............................... 2
II. ENGINEERING QUANTUM CIRCUITS .............. 3
A. From quantum harmonic oscillator to the transmon qubit ............................................. 3
B. Qubit Hamiltonian engineering ....................... 5
1. Tunable qubit: Split transmon ......................... 5
2. Toward larger anharmonicity: Flux qubit and fluxonium ............................................. 7
C. Interaction Hamiltonian engineering ............... 8
1. Physical coupling: Capacitive and inductive ...... 8
2. Coupling axis: Transverse and longitudinal ...... 9
III. NOISE, DECOHERENCE, AND ERROR MITIGATION ............................................. 10
A. Types of noise .......................................... 10
1. Systematic noise ...................................... 10
2. Stochastic noise ...................................... 10
3. Noise strength and qubit susceptibility .......... 10
B. Modeling noise and decoherence .................... 10
1. Bloch sphere representation ......................... 10
2. Bloch-Redfield model of decoherence .......... 11
3. Modification due to 1/f-type noise .............. 14
4. Noise power spectral density (PSD) .............. 14
C. Common examples of noise ......................... 15
1. Charge noise ........................................ 15
2. Magnetic flux noise ................................ 15
3. Photon number fluctuations ...................... 15
4. Quasiparticles ....................................... 16
D. Operator form of qubit-environment interaction 16
1. Connecting T1 to S(x) .............................. 16
2. Connecting T2 to S(x) .............................. 16
3. Noise spectroscopy ................................ 17
E. Engineering noise mitigation ....................... 18
1. Materials and fabrication improvements ....... 18
2. Design improvements ............................... 18
3. Dynamical error suppression ...................... 18
4. Cryogenic engineering .............................. 18
IV. QUBIT CONTROL .......................................... 18
A. Boolean logic gates used in classical computers 19
B. Quantum logic gates used in quantum computers 19
C. Comparing classical and quantum gates ....... 20
1. Gate sets and gate synthesis ...................... 22
2. Addressing superconducting qubits .......... 23
D. Single-qubit gates ................................... 23
1. Capacitive coupling for X, Y control .......... 23
I. INTRODUCTION

Quantum processors harness the intrinsic properties of quantum mechanical systems—such as quantum parallelism and quantum interference—to solve certain problems where classical computers fall short. Over the past two decades, rapid developments in the science and engineering of quantum systems have advanced the frontier in single isolated quantum systems toward the creation and manipulation of multiqubit processors. In particular, the requirements imposed by larger quantum processors have shifted the mindset within the community, from solely scientific discovery to the development of new, foundational engineering abstractions associated with the design, control, and readout of multiqubit quantum systems. The result is the emergence of a new discipline termed "quantum engineering," which serves to bridge basic sciences, mathematics, and computer science with fields generally associated with traditional engineering.

One prominent platform for constructing a multiqubit quantum processor involves superconducting qubits, in which information is stored in quantum degrees of freedom (DOFs) of nanofabricated, anharmonic oscillators (AHOs) constructed from superconducting circuit elements. In contrast to other platforms, e.g., electron spins in silicon and quantum dots, trapped ions, ultracold atoms, nitrogen-vacancy diamonds, and polarized photons where the quantum information is encoded in natural microscopic quantum systems, superconducting qubits are macroscopic in size and lithographically defined.

One remarkable feature of superconducting qubits is that their energy-level spectra are governed by circuit element parameters and thus are configurable; they can be designed to exhibit "atomlike" energy spectra with the desired properties. Therefore, superconducting qubits are also often referred to as "artificial atoms," offering a rich parameter space of possible qubit properties and operation regimes, with predictable performance in terms of transition frequencies, anharmonicity, and complexity.

While there are many other excellent reviews on superconducting qubits, see, e.g., Refs. 34–43, this work specifically aims to introduce new quantum engineers (academic and industrial alike) to the terminology and state-of-the-art practices used in the rapidly accelerating field of superconducting quantum computing. The reader is assumed to be familiar with the basic concepts that span classical physics, quantum mechanics, and electrical engineering. In particular, readers will find it useful to have had previous exposure to classical mechanics, the Schrödinger equation, the Bloch sphere representation of qubits states, second quantization, basic concepts of superconductivity, electromagnetism, introductory circuit analysis, classical Boolean logic, linear dynamical systems, analog and digital signal processing, and familiarity with microwave components such as transmission lines and mixers. These topics will be introduced as they arise, but having basic prior knowledge will be helpful.

A. Organization of this article

This review is organized in the following four sections: first, in Sec. II, we explore the parameter space available when designing superconducting circuits. In particular, we look at the promising capacitively shunted planar qubit modalities and how these can be engineered with the desired properties, such as transition frequency, anharmonicity, and reduced susceptibility to various sources of noise. In this section, we also introduce several ways in which interactions between qubits can be engineered, in order to implement two-qubit entangling operations, needed for a universal gate set.

In Sec. III, we discuss systematic and stochastic noise, the concepts of noise strength and qubit noise susceptibility, and the common sources of noise which lead to decoherence in superconducting circuits. We introduce the Bloch–Redfield model of decoherence, characterized by longitudinal and transverse relaxation times \( T_1 \) and \( T_2 \), and discuss the implications of 1/f noise. We then define the noise power spectral density (PSD), which is commonly used to characterize noise processes, and describe how it drives decoherence. Finally, we close the section with a review of coherent control methods used to mitigate certain types of coherence and reversible noise.

---

2. Virtual Z gate ........................................ 25
3. The DRAG scheme .................................. 25
E. The \( i\text{SWAP} \) two-qubit gate in tunable qubits ... 26
1. Deriving the \( i\text{SWAP} \) unitary .................... 26
2. Applications of the \( i\text{SWAP} \) gate .................. 28
F. The CPHASE two-qubit gate in tunable qubits ... 28
1. Trajectory design for the CPHASE gate ...... 29
2. The CPHASE gate for quantum error correction ........................................ 30
3. Quantum simulation and algorithm demonstrations using CPHASE ............. 30
G. Two-qubit gates using only microwaves .......... 31
1. The operational principle of the CR gate ... 31
2. Improvements to the CR gate and quantum error correction experiments using CR ..... 32
3. Quantum simulation and algorithms demonstrations with the CR gate ............ 32
4. Other microwave-only gates: hSWAP, MAP, and RIF ........................................ 33
H. Gate implementations with tunable coupling .. 34
V. QUBIT READOUT .................................. 34
A. Dispersive readout .................................. 35
B. Measuring the resonator amplitude and phase ........................................ 36
1. Representation of the readout signal ....... 37
2. I-Q mixing .............................................. 37
3. Homodyne demodulation .......................... 37
4. Heterodyne demodulation .......................... 38
C. Weak and strong qubit measurements: Impact of noise .............................. 39
D. "Purcell filters" for faster readout ................ 40
E. Improve signal-to-noise ratio: Parametric amplification ................................ 42
1. Quantum-limited amplification processes ... 42
2. Operation of Josephson parametric amplifiers ........................................ 43
3. The traveling wave parametric amplifier ........ 45
VI. SUMMARY AND OUTLOOK ..................... 46
A. Quantum annealing ................................. 46
B. Cavity based QIP ...................................... 46
C. Cryogenics and software development ....... 46
D. Quantum error correction .......................... 46
E. Quantum computational supremacy ............ 47
In Sec. IV, we provide a review of how single- and two-qubit operations are typically implemented in superconducting circuits, by using a combination of local magnetic flux control and microwave drives. In particular, we discuss the family of two-qubit gates arising from a capacitive coupling between qubits, and introduce several recent advances that have been demonstrated to achieve high-fidelity gates, as well as applications in quantum information processing that use these gates. The continued development of high-fidelity two-qubit gates in superconducting qubits is a highly active research area. For this reason, we include sufficient technical details that a reader may use this review as a starting point to critically assess the pros and cons of the various gates, as well as develop an appreciation for the types of gate-engineering already implemented in state-of-the-art superconducting quantum processors.

Finally, in Sec. V, we discuss the physics and engineering associated with the dispersive readout technique, typically used to measure the individual qubit states in modern quantum processors. After a discussion of the theory behind dispersive coupling, we give an introduction to design of Purcell filters and the development of quantum-limited parametric amplifiers (PAs).

II. ENGINEERING QUANTUM CIRCUITS

In this section, we will demonstrate how quantum systems based on superconducting circuits can be engineered to achieve certain desired properties. Using the most common qubit modalities, we discuss how properties such as the qubit transition frequency, anharmonicity, and noise susceptibility can be tailored by the choice of circuit topology and element parameter values. We also discuss how to engineer the interactions between different quantum systems, in particular, the cases of qubit-qubit and qubit-resonator couplings.

A. From quantum harmonic oscillator to the transmon qubit

A quantum mechanical system is governed by the time-dependent Schrödinger equation

\[ \hat{H}(t) \psi(t) = i \hbar \frac{d}{dt} \psi(t), \]  

where \( \psi(t) \) is the state of the quantum system at time \( t \), \( \hbar \) is the reduced Planck’s constant, and \( \hat{H} \) is the “Hamiltonian” that describes the total energy of the system. The “hat” is used to indicate that \( \hat{H} \) is a quantum operator. As the Schrödinger equation is a first-order linear differential equation, the temporal dynamics of the quantum system may be viewed as a straightforward example of a linear dynamical system with a formal solution

\[ \psi(t) = e^{-i \hat{H} t / \hbar} \psi(0). \]  

The time-independent Hamiltonian \( \hat{H} \) governs the time evolution of the system through the operator \( e^{-i \hat{H} t / \hbar} \). Thus, just as with classical systems, determining the Hamiltonian of a system—whether the classical Hamiltonian \( H \) or its quantum counterpart \( \hat{H} \)—is the first step to deriving its dynamical behavior. In Sec. IV, we consider the case when the Hamiltonian is time-dependent in the context of qubit control.

To understand the dynamics of a superconducting qubit circuit, it is natural to start with the classical description of a linear LC resonant circuit [Fig. 1(a)]. In this system, energy oscillates between electrical energy in the capacitor \( C \) and magnetic energy in the inductor \( L \). In the following, we will arbitrarily associate the electrical energy with the “kinetic energy” and the magnetic energy with the “potential energy” of the oscillator. The instantaneous, time-dependent energy in each element is derived from its current and voltage

\[ E(t) = \int_{-\infty}^{t} V(t') I(t') dt', \]  

where \( V(t') \) and \( I(t') \) denote the voltage and current of the capacitor or inductor.

To derive the classical Hamiltonian, we follow the standard approach used in classical mechanics: the Lagrange-Hamilton formulation. Here, we represent the circuit elements in terms of one of its generalized circuit coordinates, charge or flux. In the following, we pick flux, defined as the time integral of the voltage

\[ \Phi(t) = \int_{-\infty}^{t} V(t') dt'. \]  

In this example, the voltage at the node is also the branch voltage across the element. In this section, we will simply refer to these as node voltages and fluxes for convenience. For a more detailed discussion of nodes and branches in this context, we refer the reader to Ref. 44.

Note that in the following, we could have exchanged our associations with kinetic energy (momentum coordinate) and potential energy (position coordinate), and instead start with the charge variable \( Q(t) \), which is the time integral of the current \( I(t) \).

By combining Eqs. (3) and (4), using the relations \( V = L dI / dt \) and \( I = C dV / dt \), and applying the integration by parts formula, we...
can write down energy terms for the capacitor and inductor in terms of the node flux
\[ T_C = \frac{1}{2} C \Phi^2, \]
\[ U_L = \frac{1}{2L} \Phi^2. \]  

The Lagrangian is defined as the difference between the kinetic and potential energy terms and can thus be expressed in terms of Eqs. (5) and (6)
\[ L = T_C - U_L = \frac{1}{2} C \Phi^2 - \frac{1}{2L} \Phi^2. \]  

From the Lagrangian in Eq. (7), we can further derive the Hamiltonian using the Legendre transformation, for which we need to calculate the momentum conjugate to the flux, which in this case, is the charge on the capacitor
\[ Q = \frac{\partial L}{\partial \Phi} = C \Phi. \]  

The Hamiltonian of the system is now defined as
\[ H = \Phi \dot{\Phi} - L = \frac{\Phi^2}{2C} + \frac{1}{2} CV^2 + \frac{1}{2} L I^2, \]  

as one would expect for an electrical LC circuit. Note that this Hamiltonian is analogous to that of a mechanical harmonic oscillator, with mass \( m = C \) and resonant frequency \( \omega = 1/\sqrt{LC} \), which expressed in position, \( x \), and momentum, \( p \), coordinates takes the form \( H = p^2/2m + m \omega^2 x^2/2 \).

The Hamiltonian described above is classical. In order to proceed to a quantum-mechanical description of the system, we need to promote the charge and flux coordinates to quantum operators, whereas the classical coordinates satisfy the Poisson bracket
\[ \{f, g\} = \frac{\partial f}{\partial \Phi} \frac{\partial g}{\partial \dot{\Phi}} - \frac{\partial f}{\partial \dot{\Phi}} \frac{\partial g}{\partial \Phi} \]
\[ \rightarrow \{\Phi, Q\} = \frac{\partial \Phi}{\partial \Phi} \frac{\partial Q}{\partial \dot{\Phi}} - \frac{\partial \Phi}{\partial \dot{\Phi}} \frac{\partial Q}{\partial \Phi} = 1 - 0 = 1, \]  

the quantum operators similarly satisfy a “commutation relation”
\[ [\Phi, \dot{Q}] = \dot{\Phi} Q - \dot{Q} \Phi = i\hbar, \]  

where the operators are indicated by hats. From this point forward, however, the hats on operators will be omitted for simplicity.

In a simple LC resonant circuit [Fig. 1(a)], both the inductor \( L \) and the capacitor \( C \) are linear circuit elements. Defining the reduced flux \( \phi \equiv 2\pi \Phi/\Phi_0 \) and the reduced charge \( n = Q/2e \), we can write down the following quantum-mechanical Hamiltonian for the circuit
\[ H = 4E_C n^2 + \frac{1}{2} E_L \phi^2, \]  

where \( E_c = e^2/(2C) \) is the charging energy required to add “each” electron of the Cooper-pair to the island and \( E_L = (\Phi_0/2\pi)^2/L \) is the inductive energy, where \( \Phi_0 = h/2e \) is the superconducting magnetic flux quantum. Moreover, the quantum operator \( n \) is the excess number of Cooper-pairs on the island, and \( \phi \)—the reduced flux—is denoted the “gauge-invariant phase” across the inductor. These two operators form a canonical conjugate pair, obeying the commutation relation \([\phi, n] = i\). We note that the factor 4 in front of the charging energy \( E_c \) is solely a historical artifact, namely, that this energy scale was first defined for single-electron systems and then adopted to two-electron Cooper-pair systems.

The Hamiltonian in Eq. (13) is identical to the one describing a particle in a one-dimensional quadratic potential, a quantum harmonic oscillator (QHO). We can treat \( \phi \) as the generalized position coordinate, so that the first term is the kinetic energy and the second term is the potential energy. We emphasize that the functional form of the potential energy influences the eigensolutions. For example, the fact that this term is quadratic \((\hat{U}_L \propto \phi^2)\) in Eq. (13) gives rise to the shape of the potential in Fig. 1(b). The solution to this eigenvalue problem gives an infinite series of eigenstates \( |k\rangle \), \( (k = 0, 1, 2, \ldots) \), whose corresponding eigenenergies \( E_k \) are all equidistantly spaced, i.e., \( E_{k+1} - E_k = \hbar \omega_k \), where \( \omega_k = \sqrt{8E_k E_C}/\hbar = 1/\sqrt{LC} \) denotes the resonant frequency of the system, see Fig. 1(b). We may represent these results in a more compact form (second quantization) for the quantum harmonic oscillator (QHO) Hamiltonian
\[ H = \hbar \omega (\hat{a}^\dagger \hat{a} + 1/2), \]  

where \( \hat{a}^\dagger (a) \) is the creation (annihilation) operator of a single excitation of the resonator. The Hamiltonian in Eq. (14) is written as energy. It is, however, often preferred to divide by \( \hbar \) so that the expression has units of radian frequency, since we will later resonate drive transitions at a particular frequency or reference the rate at which two systems interact with one another. Therefore, from here on, \( \hbar \) will be omitted.

The original charge number and phase operators can be expressed as \( n = n_{qd} \times \hat{i}(a - a^\dagger) \) and \( \phi = \phi_{qd} \times \hat{i}(a + a^\dagger) \), where \( n_{qd} = [E_k/(32E_C)]^{1/4} \) and \( \phi_{qd} = (2E_k/E_0)^{1/4} \) are the “zero-point fluctuations” of the charge and phase variables, respectively. Quantum mechanically, the quantum states are represented as wavefunctions that are generally distributed over a range of values of \( n \) and \( \phi \) and, consequently, the wavefunctions have nonzero standard deviations. Such wavefunction distributions are referred to as “quantum fluctuations,” and they exist, even in the ground state, where they are called zero-point fluctuations.

The linear characteristics of the QHO have a natural limitation in its applications for processing quantum information. Before the system can be used as a qubit, we need to be able to define a computational subspace consisting of only two energy states (usually the two-lowest energy eigenstates) in between which transitions can be driven without also exciting other levels in the system. Since many gate operations, such as single-qubit gates (Sec. IV), depend on frequency selectivity, the equidistant level-spacing of the QHO, illustrated in Fig. 1(b), poses a practical limitation. To mitigate the problem of unwanted dynamics involving non-computational states, we need to add anharmonicity (or nonlinearity) into our system. In short, we require the transition frequencies \( \omega_q^{-1} \) and \( \omega_{q+1}^{-1} \) be sufficiently different to be individually addressable. In general, the larger the anharmonicity the better it is. In practice, the amount of anharmonicity sets a limit on how short the pulses used to drive the qubit can be. This is discussed in detail in Sec. IV D 3.

To introduce the nonlinearity required to modify the harmonic potential, we use the Josephson junction—a nonlinear, dissipationless
circuit element that forms the backbone in superconducting circuits.\textsuperscript{66-67} By replacing the linear inductor of the QHO with a Josephson junction, playing the role of a nonlinear inductor, we can modify the functional form of the potential energy. The potential energy of the Josephson junction can be derived from Eq. (3) and the two Josephson relations
\begin{equation}
I = I_0 \sin (\phi), \quad V = \frac{h}{2e} \frac{d\phi}{dt},
\end{equation}
resulting in a modified Hamiltonian
\begin{equation}
H = 4E_C n^2 - E_J \cos (\phi),
\end{equation}
where $E_C = e^2/(2C_0)$, $C_0 = C_s + C_j$ is the total capacitance, including both shunt capacitance $C_s$ and the self-capacitance of the junction $C_j$ and $E_J = I_0 \Phi_0/2\pi$ is the Josephson energy, with $I_0$ being the critical current of the junction.\textsuperscript{489} After introducing the Josephson junction in the circuit, the potential energy no longer takes a manifestly parabolic form (from which the harmonic spectrum originates), but rather features a cosinusoidal form, see the second term in Eq. (16), which makes the energy spectrum nondegenerate. Therefore, the Josephson junction is the key ingredient that makes the oscillator anharmonic and thus allows us to identify a uniquely addressable quantum two-level system, see Fig. 1(d).

Once the nonlinearity has been added, the system dynamics is governed by the dominant energy in Eq. (16), reflected in the $E_J/E_C$ ratio. Over time, the superconducting qubit community has converged toward circuit designs with $E_J \gg E_C$. In the opposite case when $E_J \ll E_C$, the qubit becomes highly sensitive to charge noise, which has proven more challenging to mitigate than flux noise, making it very hard to achieve high coherence. Another motivation is that current technologies allow for more flexibility in engineering the inductive (or potential) part of the Hamiltonian. Therefore, working in the $E_J \ll E_C$ limit, makes the system more sensitive to the change in the potential Hamiltonian. Therefore, we will focus here on the state-of-the-art qubit modalities that fall in the regime $E_J \gg E_C$. For readers who are interested in the physics in the $E_J \ll E_C$ regime, such as the earlier Cooper-pair box charge qubit, we refer to Refs. 48–51.

To access the $E_J \gg E_C$ regime, one preferred approach is to make the charging $E_C$ small by shunting the junction with a large capacitor, $C_s \gg C_j$, effectively making the qubit less sensitive to charge noise—a circuit commonly known as the transmon qubit.\textsuperscript{21} In this limit, the superconducting phase $\phi$ is a good quantum number, i.e., the spread (or quantum fluctuation) of $\phi$ values represented by the quantum wavefunction is small. The low-energy eigenstates are therefore, to a good approximation, localized states in the potential well, see Fig. 1(d). We may gain more insight by expanding the potential term of Eq. (16) into a power series (since $\phi$ is small), that is
\begin{equation}
E_J \cos (\phi) = \frac{1}{2} E_J \phi^2 + \frac{1}{24} E_J \phi^4 + O(\phi^6).
\end{equation}

The leading quadratic term in Eq. (17) alone will result in a QHO, recall Eq. (13). The second term, however, is quartic which modifies the eigenstructure and disrupts the otherwise harmonic energy structure. Note that, the negative coefficient of the quartic term indicates that the anharmonicity $\kappa = \omega_0^4 - 2\omega_0^2 \Omega^2$ is negative and its limit in magnitude thus cannot be made arbitrarily large. For the case of the transmon, $\kappa = -E_C$ is usually designed to be 100–300 MHz, as required to maintain a desirable qubit frequency $\omega_0 = (\sqrt{8E_C E_0} - E_C)/\hbar = 3-6$ GHz, while keeping an energy ratio sufficiently large ($E_J/E_C \geq 50$) to suppress charge sensitivity.\textsuperscript{22} Fortunately, the charge sensitivity is exponentially suppressed for an increased $E_J/E_C$, while the reduction in anharmonicity only scales as a weak power law, leading to a workable device.

Including terms up to fourth order and using the QHO eigenbases, the system Hamiltonian resembles that of a Duffing oscillator
\begin{equation}
H = \omega_0 a^\dagger a + \kappa a^4 + \omega_0^2 a a^\dagger a.
\end{equation}

Since $|\kappa| \ll \omega_0$, we can see that the transmon qubit is basically a weakly anharmonic oscillator (AHO). If excitation to higher noncomputational states is suppressed over any gate operations, either due to a large enough $|\kappa|$ or due to robust control techniques such as the derivative reduction by adiabatic gate (DRAG) pulse, see Sec. IV D 3, we may effectively treat the AHO as a quantum two-level system, simplifying the Hamiltonian to
\begin{equation}
H = \omega_0 \sigma_x^2.
\end{equation}

where $\sigma_x$ is the Pauli-$x$ operator. However, one should always keep in mind that higher levels physically exist.\textsuperscript{53} Their influence on the system dynamics should be taken into account when designing the system and its control processes. In fact, there are many cases where the higher levels have proven useful to implement more efficient gate operations.\textsuperscript{54}

In addition to reducing the charge dispersion, the use of a large shunt capacitor also enables us to engineer the electric field distribution of the quantum system, and thus the participation of surface loss mechanisms. In the development of the 3D transmon,\textsuperscript{65} e.g., a 2D transmon coupled to a 3D cavity, it was demonstrated that by making the gap between the two lateral capacitor plates large (compared to the film thickness) the coherence time increases since a smaller portion of the electric field interacts with the lossy interfaces, e.g., metal-substrate and substrate-vacuum interfaces, which has been studied extensively.\textsuperscript{56–61}

B. Qubit Hamiltonian engineering

1. Tunable qubit: Split transmon

To implement fast gate operations with high-fidelity, as needed to implement quantum logic, many (though not all\textsuperscript{62}) of the quantum processor architectures implemented today feature tunable qubit frequencies.\textsuperscript{64–67} For instance, in some cases, we need to bring two qubits into resonance to exchange (swap) energy, while we also need the capability of separating them during idling periods to minimize their interactions. To do this, we need an external parameter which allows us to access one of the degrees of freedom of the system in a controllable fashion.

One widely used technique is to replace the single Josephson junction with a loop interrupted by two identical junctions—forming a DC superconducting quantum interference device (DC-SQUID).\textsuperscript{23} Due to the interference between the two arms of the SQUID, the effective critical current of the two parallel junctions can be decreased by applying a magnetic flux threading the loop, see Fig. 2(a). Due to the fluxoid quantization condition, the algebraic sum of branch flux of all
of the inductive elements along the loop plus the externally applied flux equal an integer number of superconducting flux quanta, that is
\[ \phi_1 - \phi_2 + 2\phi_p = 2\pi k, \tag{20} \]

where \( \phi_p = \pi \Phi_{\text{ext}} / \Phi_0 \). Using this condition, we can eliminate one degree of freedom and treat the SQUID-loop as a single junction, but with the important modification that \( E_J \) is tunable (via the SQUID critical current) by means of the external flux \( \Phi_{\text{ext}} \). The effective Hamiltonian of the so-called split transmon (ignoring the constant) is
\[ H = 4E_J n^2 - E_J \cos(\phi_q) \cos(\phi). \tag{21} \]

We can see that Eq. (21) is analogous to Eq. (16), with \( E_J \) replaced by \( E_J(\phi_q) = 2E_J(\cos(\phi_q)) \). The magnitude of the net, effective Josephson energy \( E_J \) has a period of \( \Phi_0 \) in applied flux and spans from 0 to its maximum value \( 2E_J \). Therefore, the qubit frequency can be tuned periodically with \( \Phi_{\text{ext}} \) see Fig. 2(b).

While the split transmon enables frequency tunability by the externally applied magnetic field, it also introduces sensitivity to random flux fluctuations, known as flux noise. At any working point, the slope of the qubit spectrum, \( \partial \omega_q / \partial \Phi_{\text{ext}} \), indicates to first order how strongly this flux noise affects the qubit frequency. The sensitivity is generally nonzero, except at multiples of the flux quantum, \( \Phi_{\text{ext}} = k \Phi_0 \), where \( k \) is an integer, where \( \partial \omega_q / \partial \Phi_{\text{ext}} = 0 \).

One recent development has focused on reducing the qubit sensitivity to flux noise, while maintaining sufficient tunability to operate our quantum gates. The idea is to make the two junctions in the split transmon asymmetric, see Fig. 2(c). This yields the following Hamiltonian
\[ H = 4E_J n^2 - E_J \sqrt{\cos^2(\phi_q) + d^2 \sin^2(\phi_q) \cos(\phi)}, \tag{22} \]

where \( E_J = E_{J1} + E_{J2} \) and \( d = (\gamma - 1)/(\gamma + 1) \) is the junction asymmetry parameter, with \( \gamma = E_{J2} / E_{J1} \). Again, we can treat the two junctions as a single-junction transmon, with an effective Josephson energy \( E_J(\phi_q) \). In particular, we can recognize the two special cases; for \( d = 0 \), the Hamiltonian in Eq. (22) reduces to the symmetric case with \( E_J(\phi_q) = E_{J2} \cos(\phi_q) \), as in Eq. (21) with \( E_{J2} = 2E_J \). In the other limit, when \( |d| \to 1 \), \( E_J(\phi_q) \to E_{J2} \) and the flux-tunability of the Josephson energy vanishes, which is equivalent to the single junction case, recall Eq. (16).

From the discussion above we see that going from symmetric to asymmetric transmons does not change the circuit topology. This seemingly trivial modification, however, has a profound impact for practical applications. As we can see from the qubit spectra, Fig. 2(d), the flux sensitivity is suppressed across the entire tunable frequency range. For example, the performance of the cross-resonance gate is optimized with a certain frequency detuning between two qubits. Therefore, by using an asymmetric transmon, a small frequency-tuning range is introduced that is sufficient to compensate for fabrication variations, without introducing unnecessary large susceptibility to flux noise and thus maintaining high coherence. For another example, a surface code scheme based on the adiabatic controlled phase (CPHASE)-gate requires specific frequency configuration among qubits in order to avoid frequency crowding issues, and asymmetric transmons fit well with its well-defined frequency range. In general, as the quantum processors scale up and fabrication improves, asymmetric transmons are likely to be found in wider applications in the future.
2. Toward larger anharmonicity: Flux qubit and fluxonium

We see that split transmon qubits, be it symmetric or not, still share the same topology as the single junction version, yielding a sinusoidal potential. Therefore, the degree to which the properties of these qubits can be engineered has not fundamentally changed. In particular, the limited anharmonicity in transmon-type qubits intrinsically causes significant residual excitation to higher-energy states, undermining the performance of gate operations. To go beyond this, it is necessary to introduce additional complexity into the circuit.

One outstanding development in this regard is the invention of the flux qubit,\textsuperscript{22,23} where the qubit loop is interrupted by three (or four) junctions, see Fig. 2(e). On one branch is one smaller junction; on the other branch are two identical junctions, both a factor $\gamma$ larger in size compared to the small junction. The addition of one more junction as compared to the split transmon is nontrivial, as it changes the circuit topology and reshapes the potential energy profile.

Each junction is associated with a phase variable, and the fluxoid quantization condition again allows us to eliminate one degree of freedom. Consequently, we have a two-dimensional potential landscape, which in comparison to the simpler topology of the transmon, complicates the problem both conceptually and computationally. Fortunately, under the assumed setting that the array junctions are larger in size ($\gamma > 1$), it is usually a good approximation to treat the problem as a particle moving in a quasi-1D potential, which also helps us gain more insight and intuition about the system and draw qualitative conclusions. The Hamiltonian under this "quasi-1D approximation" reads

$$H \approx 4E_C n^2 - E_J \cos (2\phi + \varphi_9) - 2\gamma E_J \cos (\phi). \quad (23)$$

Note that the phase variable in Eq. (23) is the sum of the branch phases across the two array junctions, $\phi = (\varphi_1 + \varphi_2)/2$, assuming the same current direction across $\varphi_1$ and $\varphi_2$. The external magnetic flux is denoted $\varphi_9 = 2\pi \Phi_{ext} / \Phi_0$. The second term in Eq. (23) is contributed by the small junction with Josephson energy $E_J$, whereas the third term takes into account the two array junctions, together with Josephson energy $2\gamma E_J$. Clearly, the sum of these two terms no longer has the characteristics of a simple cosineoid, and the final potential profile as well as the corresponding eigenstates depend on both the external flux $\varphi_9$ and the junction area ratio $\gamma$.

The most common working point for this system is when $\varphi_9 = \pi + 2\pi k$, where $k$ is an integer—that is when half a superconducting flux quantum threads the qubit loop. At this flux bias point, the qubit spectrum reaches its minimum, and the qubit frequency is first-order insensitive to flux noise, see Fig. 2(f). This point is often referred to as "the flux degeneracy point," where flux qubits tend to have the optimal coherence time.

At this operation point, the potential energy may assume a single-well ($\gamma > 2$) or a double-well ($\gamma < 2$) profile. The single-well case shares some similarities with the transmon qubit, where the quadratic and quartic terms of the Hamiltonian determines the harmonic-ity and anharmonicity, respectively. The capacitively shunted flux qubit (CSFQ)\textsuperscript{24,27} was explored in this regime, demonstrating long coherence and decently high anharmonicity. Note that as opposed to the transmon qubit, the anharmonicity of the CSFQ is "positive" ($\alpha > 0$). While the improvement in anharmonicity can be associated with reshaping the energy potential, the improved coherence over the first flux qubits can be attributed to the introduction of the capacitive shunt, similar to the modified Cooper-pair box leading to the transmon qubit.

The double-well case obtained for $\gamma < 2$ was demonstrated and investigated much earlier.\textsuperscript{22,27} The intuitive picture based on circulating current states—so it gets the name persisting-current flux qubit (PCFQ)—gives a satisfying physical description of the qubit degrees of freedom. However, from the perspective of a quantum engineer, the qubit properties are of more interest, even if sometimes we may lose physical intuition about the system in certain regimes; such as when $\gamma \approx 2$ and there are no clear circulating current states. The most important feature of the PCFQ is that its anharmonicity can be much greater than the transmon and CSFQ and the transition matrix elements $\langle \langle 1 \mid \overline{0} \rangle \rangle$, $\langle 1 \mid \overline{\phi} \rangle \rangle$ become considerably smaller given equivalent $E_J/E_C$. Therefore, a longer relaxation time can be expected. These features have been demonstrated even more prominently in its close relative, the fluxonium qubit.\textsuperscript{75}

The flux qubit is a striking example that illustrates how one dramatically can engineer the qubit properties through the choice of various circuit parameters. The introduction of array junctions and consequent biharmonic profile generates rich dynamics as well as broad applications. An extension of this idea is the fluxonium qubit, which generated substantial interest recently, due partly to its capability of engineering the transition matrix elements to achieve millisecond $T_1$ time, and due partly to the invention of novel gate schemes applicable to such well-protected qubits.\textsuperscript{76,77}

Compared to flux qubits, which usually contain two or three array junctions,\textsuperscript{78} the number of array junctions in the fluxonium qubit is dramatically increased,\textsuperscript{79,80} in some cases, to the order of 100, see Fig. 2(g). Following the same quasi-1D approximation as for the flux qubit, the last term in Eq. (23) becomes $-N^2 E_J \cos (\phi/N)$, where $N$ denotes the number of array junctions. For large $N$, the argument in the cosine term $\phi/N$ becomes sufficiently small that a second order expansion is a good approximation. This results in the fluxonium Hamiltonian

$$H \approx 4E_C n^2 - E_J \cos (\phi + \varphi_9) + \frac{1}{2} E_J \phi^2, \quad (24)$$

where $E_L = (\gamma/N)E_J$ is the inductive energy of the effective inductance contributed by the junction array—often known as superinductance due to its large value.\textsuperscript{73,83,84} Therefore, we can treat the potential energy as a quadratic term modulated by a sinusoidal term, similar to that of an rf-SQUID type flux qubit.\textsuperscript{83} However, the kinetic inductance of the Josephson junction array is in general much larger than the geometric inductance of the wire in an rf-SQUID.

Depending on the relative magnitude of $E_J$ and $E_L$, the fluxonium system could involve plasmon states (in the same well) and fluxon states (in different wells). There are a variety of schemes to utilize them for quantum information processing. Generally, the spectrum of the transition between the lowest energy states is similar to that of the flux qubit, see Fig. 2(h). Both long coherence and high anharmonicity can be expected at the flux sweet spot.

Lastly, we want to point out a further extension—the 0–π qubit—which has even stronger topological protection from noise.\textsuperscript{83,84} However, the strongly suppressed sensitivity to external fluctuations also makes it hard to manipulate.
C. Interaction Hamiltonian engineering

To generate entanglement between individual quantum systems—it is necessary to engineer an interaction Hamiltonian that connects degrees of freedom in those individual systems. In this section, we discuss the physical coupling mechanism and its representation in the qubit eigenbasis. The use of coupling to form 2-qubit gates is discussed in Sec. IV.

1. Physical coupling: Capacitive and inductive

The Hamiltonian of two coupled systems takes a generic form

\[ H = H_1 + H_2 + H_{int}, \]

(25)

where \( H_1 \) and \( H_2 \) denote the Hamiltonians of the individual quantum systems, which could be any combination of the qubit circuits mentioned in Secs. II A and II B. The last term, \( H_{int} \), is the interaction Hamiltonian, which couples the variables of both systems. In superconducting circuits, the physical form of the coupling energy is either an electric or magnetic field (or a combination thereof).

To achieve capacitive coupling, a capacitor is placed between the voltage nodes of the two participating circuits, yielding an interaction Hamiltonian \( H_{int} \) of the form

\[ H_{int} = C_g V_1 V_2, \]

(26)

where \( C_g \) is the coupling capacitance and \( V_1(V_2) \) is the voltage operator of the corresponding voltage node being connected. Figure 3(a) illustrates a realistic example of a direct capacitive coupling between the top nodes of two transmon qubits. Circuit quantization in the limit of \( C_g \ll C_1, C_2 \) yields

\[ H = \sum_{i=1,2} \left[ 4E_C n_i^2 - E_{1,i} \cos(\phi_i) \right] + 4\varepsilon^2 \frac{C_1}{C_1 + C_2} n_1 n_2, \]

(27)

where the expressions in brackets are the two Hamiltonians of the individual qubits, [see Eq. (16)], and we take \( V_i = (2e/C_i)n_i \) in Eq. (26). From Eq. (27), we see that the coupling energy depends on the coupling capacitance as well as the matrix elements of the voltage operators. The dependencies are bilinear in the perturbative limit \( C_g \ll C_1, C_2 \).

To implement the coupling capacitance, one only need bring the edges of the capacitor pads into close proximity, as has been demonstrated in-state-of-the-art planar designs.\(^8,9\) The coupling capacitance is determined by the planar capacitor geometry as well as the surrounding environment, such as the dielectric constant of the substrate and the ground plane proximity.

In the case of inductive coupling, a mutual inductance shared by two loops is the coupling mechanism, yielding an interaction Hamiltonian of the form

\[ H_{int} = M_{12} I_1 I_2, \]

(28)

where \( M_{12} \) denotes the mutual inductance and \( I_1(I_2) \) is the current operator of the loop current. A typical example is two closely positioned (rf-SQUID type) flux qubits, as illustrated in Fig. 3(c). The system Hamiltonian can be expressed as

\[
\begin{align*}
H &= \sum_{i=1,2} \left[ 4E_C n_i^2 + \frac{1}{2} E_{L,i} \phi_i^2 - E_{1,i} \cos(\phi_i) \right] \\
&\quad + M_{12} I_1 \sin(\phi_1) I_2 \sin(\phi_2),
\end{align*}
\]

(29)

where the individual qubit Hamiltonians are identical to that of the fluxonium in Eq. (24), and the current operators, \( I_i = I_{0,i} \sin(\phi_i) \) with \( i \in 1, 2 \), is the familiar DC-Josephson relation for each junction, see Eq. (15). In this case, the strength of the inductive coupling energy depends on the mutual inductance as well as the matrix element of the current operators.

To realize a mutual inductance, two looped circuits are brought into close proximity to one another, or, to make them stronger, overlapping with each other,\(^8,9\) and even may share the same wire or Josephson junction inductor.\(^8,9\) In the case of a Josephson junction, and for certain metals, the inductance is dominated by "kinetic inductance" contributions, rather than solely geometric inductance.\(^8,9\) Kinetic inductance arises from the mechanical, inertial mass of the charge carriers, but is only practically witnessed in very high-conductance materials like superconductors. A primary feature of kinetic inductance is that its values can vastly exceed those of conventional geometric inductances, which are generally limited by electromagnetic considerations.\(^8,9\)
2. Coupling axis: Transverse and longitudinal

Regardless of its physical realization, the effect of a coupling on system dynamics is determined by its form as represented in the eigenbasis of the individual systems. That is, how $H_{\text{int}}$ appears in the representation spanned by the eigenbasis of $H_1 \otimes H_2$.

Let us start with the previous example of two capacitively coupled transmon qubits [Fig. 3(a)]. Using second quantization, the system Hamiltonian in Eq. (27) can be expressed as

$$H = \sum_{i=1,2} \left( \omega_0 a_i^\dagger a_i + \frac{2g}{3} a_i^\dagger a_i^\dagger a_i a_i a_i \right) - g (a_i - a_i^\dagger) (a_2 - a_2^\dagger),$$

where the expression within brackets represent the Duffing oscillator Hamiltonian for the qubits and $g$ is the coupling energy. Since we define $V \propto n \propto i(a - a^\dagger)$, and consequently $I \propto \phi \propto (a + a^\dagger)$, the original $n_1n_2$-term becomes what is shown in Eq. (30). Such a coupling is called “transverse,” because the coupling Hamiltonian non-zero matrix elements only at off-diagonal positions with respect to both oscillators, i.e., $\langle k|a_i - a_i^\dagger|k\rangle = 0$ for any integer $k$ and for $i \in 1,2$, and in this case, $\langle k|\pm 1|a_i - a_i^\dagger|k\rangle \neq 0$.

If we can ignore higher energy levels ($k \geq 2$) either because of sufficient anharmonicity or through careful control protocols that ensure these levels never have influence, we may truncate the Hamiltonian in Eq. (30) to

$$H = \sum_{i=1,2} \frac{1}{2} \omega_0 a_i^\dagger a_i + g \sigma_i^x \sigma_i^z.$$  

This is a Hamiltonian of two spins, coupled by an exchange interaction. As we will see in Sec. IV D 1, such a Hamiltonian is most commonly used in contemporary implementations and can generate various types of two-qubit entangling gates. Note that, more often, we see that the interaction term is expressed in $\sigma_1 \sigma_2$ instead of $\sigma_x \sigma_y$. The choice in the context here is arbitrary and does not change the dynamics. However, when both capacitive and inductive couplings are present in the system, both $\sigma_1 \sigma_2$ and $\sigma_x \sigma_y$ may be needed. In this case, the voltage operator $V \propto i(a - a^\dagger)$ (reduced to $\sigma_z$ when two-level approximation in the lab frame) is transversal to the current operator $I \propto (a + a^\dagger)$ (reduced to $\sigma_0$) and both of them may be transverse to the qubit. A similar example is demonstrated between a qubit and a resonator by Lu et al.\textsuperscript{11}

Transverse coupling can be engineered between a qubit and a harmonic oscillator, see Fig. 3(b). In this case, the Hamiltonian becomes

$$H = \frac{1}{2} \omega_0 \sigma_0^x + \omega_0 a_i^\dagger a_i + g (\sigma_0^x a_0 + \sigma_0^x a_i^\dagger),$$

where $\omega_0$ and $\omega_0$ denote the qubit and resonator frequencies, and $\sigma_0^x = |0\rangle\langle 1|$ and $\sigma_0^y = |1\rangle\langle 0|$ describes the processes of exciting and de-exciting the qubit, respectively. Here, we have assumed that the coupling is in the dispersive limit, i.e., $g \ll \omega_0, \omega_0$; hence ignoring the double (de)excitation terms proportional to $\sigma_0^y a_i^\dagger$ and $\sigma_0^y a_i$, which under typical operation regimes oscillate sufficiently fast to average to zero. The Hamiltonian in Eq. (32), is the standard model used for describing how a two-level atom interacts with a resonant cavity that houses it. Such a structure is also known as cavity quantum electrodynamics (cQED), and it is extended to the circuit version here. It has many useful applications in superconducting quantum information architectures, such as high-fidelity readout,\textsuperscript{13} see Sec. V, cavity buses,\textsuperscript{3,14} quantum memory,\textsuperscript{36,67} quantum computation with cat states,\textsuperscript{38–100} etc.

Here, we briefly mention the use of a cavity or resonator to mediate coupling between two qubits, which may be physically well-separated ($\approx 1$ cm). Since most superconducting resonators are in the GHz frequency range, they can be made much longer than any dimension of a qubit circuit ($\approx 1$ mm). One can use such a resonator to mediate coupling between two or more otherwise noninteracting qubits. An example is shown in Fig. 3(b), where two transmon qubits are both capacitively coupled to the center resonator. The two-level system Hamiltonian is:

$$H = \sum_{i=1,2} \left( \omega_0 a_i^\dagger a_i + \frac{2g}{3} a_i^\dagger a_i^\dagger a_i a_i a_i \right) + \omega_0 a_i^\dagger a_i + g \sigma_0^y a_0 a_i a_i^\dagger + g \sigma_0^y a_i a_i^\dagger a_i a_i^\dagger.$$  

It can be shown that in the dispersive limit, i.e., $g_0 \ll |\omega_0 - \omega_0|$, the resonator can—after proper transformation and approximation—be treated as an isolated system, and the composite system simplified to two transversely coupled qubits, see Eq. (31).

We now turn to the previous example of two inductively coupled flux qubits, see Fig. 3(c). Assume that the double-well potential [Fig. 2(g)] has a relatively high interwell barrier, which leads to an exponentially small qubit transition frequency at the energy degeneracy point, ($\Phi_0 = \pi$). Around this degeneracy point, the off-diagonal matrix element of $sin\phi$ is zero, i.e., the ground and excited states are localized in different wells and $\langle 1|sin\phi|1\rangle - \langle 0|sin\phi|0\rangle \neq 0$. We can then rewrite the Hamiltonian in Eq. (29) as

$$H = \sum_{i=1,2} \frac{1}{2} \omega_0 \sigma_i^x + g \sigma_i^z \sigma_i^z.$$  

Now, the coupling axis is the same as the qubit quantization axes and therefore termed “longitudinal coupling.” Note, however, that the physical $\sigma_x \sigma_x$ and $\sigma_x \sigma_y$ couplings can change in the qubit frame.

Longitudinal coupling is an important type of interaction, because it can generate entanglement without energy exchange. Moreover, it is found a necessary ingredient in the application of quantum annealing, where certain hard combinatorial optimization problems can be modeled by the Ising Hamiltonian in Eq. (34) and finding its ground state would solve this problem.

An intermediate qubit mode may also be used as a coupler in the longitudinal case. In Fig. 3(d), an additional rf-SQUID is used to mediate the coupling. The coupling strength can be tuned by the flux bias of the coupler SQUID.\textsuperscript{101} Note that a tunable coupler may also be realized in a structure with capacitive couplings.\textsuperscript{65} A tunable coupler is useful because it provides a wide range of coupling strengths,\textsuperscript{102} a high on-off ratio\textsuperscript{103} for reducing gate error-rates, and more ways of achieving high-fidelity entangling gates.\textsuperscript{67,104–106} The trade-off is an additional control line.

In addition to the pure transversal and longitudinal qubit-qubit interactions, there are also examples of mixed types of interaction Hamiltonians\textsuperscript{105}

$$H = \frac{1}{2} \omega_0 \sigma_z + \omega_0 a_i^\dagger a_i + g \sigma_z (a + a^\dagger).$$

\textsuperscript{105}
which are longitudinal with respect to a qubit, but transverse with respect to a harmonic oscillator in a qubit-resonator system. Such a model is called longitudinal but one should note that it is only longitudinal to one participating system. It is hard to engineer physically longitudinal coupling with respect to a harmonic oscillator, since either the $E$-field ($V$) or the $B$-field ($I$) is transverse with respect to the eigen field of the harmonic oscillator. Note, however, that a transversal model such as in Eq. (32) may be transformed into a longitudinal one in certain operating regimes, see Sec. V.

In some applications, such as for quantum annealing, both longitudinal and transverse couplings are desired ($\sigma_0\sigma_z$ coupling for mapping the problem and $\sigma_0\sigma_x$ coupling for enhancing the annealing performance) and require independent control.

III. NOISE, DECOHERENCE, AND ERROR MITIGATION

Random, uncontrollable physical processes in the qubit control and measurement equipment or in the local environment surrounding the quantum processor are sources of noise that lead to decoherence and reduce the operational fidelity of the qubits. In this section, we introduce the basics of noise leading to decoherence in superconducting circuits, and we discuss coherent control methods to mitigate certain types of noise.

A. Types of noise

In a closed system, the dynamical evolution of a qubit state is deterministic. That is, if we know the starting state of the qubit and its Hamiltonian, then we can predict the state of the qubit at any time in the future. However, in open systems, the situation changes. The qubit now interacts with uncontrolled degrees of freedom in its environment, which we refer to as fluctuations or noise. In the presence of noise, as time progresses, the qubit state looks less and less like the state we would have predicted and, eventually, the state is lost. There are many different sources of noise that affect quantum systems, and they can be categorized into two primary types: systematic noise and stochastic noise.

1. Systematic noise

Systematic noise arises from a process that is traceable to a fixed control or readout error. For example, we apply a microwave pulse to the qubit that we believe will impart a $180^\circ$ rotation. However, the control field is not tuned properly and, rather than rotating the qubit $180^\circ$, the pulse slightly over-rotates or under-rotates the qubit by a fixed amount. The underlying error is “systematic,” and it therefore leads to the same rotation error each time it is applied. However, when such erroneous pulses are used in practice in a variety of control sequences, the observed results may appear to be influenced by random noise. This is because the pulse is generally not applied in the same way for each experiment: it could be applied a different number of times, interspersed with different pulses in different orders, and therefore generally differs from experiment to experiment. However, once systematic errors are identified, they can generally be corrected through proper calibration or the use of improved hardware.

2. Stochastic noise

The second type of noise is stochastic noise, arising from random fluctuations of parameters that are coupled to our qubit. For example, thermal noise of a 50 $\Omega$ resistor in the control lines leading to the qubit will have voltage and current fluctuations—Johnson noise—with a noise power that is proportional to both temperature and bandwidth. Or, the oscillator that provides the carrier for a qubit control pulse may have amplitude or phase fluctuations. Additionally, randomly fluctuating electric and magnetic fields in the local qubit environment—e.g., on the metal surface, on the substrate surface, at the metal-substrate interface, or inside the substrate—can couple to the qubit. This creates unknown and uncontrolled fluctuations of one or more qubit parameters, and this leads to qubit decoherence.

3. Noise strength and qubit susceptibility

The degree to which a qubit is affected by noise is related to the amount of noise impinging on the qubit, and the qubit’s susceptibility to that noise. The former is often a question of materials science and fabrication; that is, can we make devices with lower levels of noise. Or, it may be related to the quality of the control electronics and cryogenic engineering to limit the levels of noise on the control lines that necessarily connect to the qubits to control them. The latter—qubit susceptibility—is a question of qubit design. Qubits can be designed to trade-off sensitivity to one type of noise at the expense of increased sensitivity to other types of noise. Therefore, materials science, fabrication engineering, electronics design, cryogenic engineering, and qubit design all play a role in creating devices with high coherence. In general, one should strive to eliminate the sources of noise, and then design qubits that are insensitive to the residual noise.

The qubit response to noise depends on how the noise couples to it—either through a longitudinal or a transverse coupling as referenced to the qubit quantization axis. This can be visualized using a Bloch Sphere picture of the qubit state, as illustrated in Fig. 4 and discussed in detail in Sec. III B.

B. Modeling noise and decoherence

1. Bloch sphere representation

The “Bloch sphere” is a unit sphere used to represent the quantum state of a two-level system (qubit). Figure 4(a) shows a Bloch sphere with a “Bloch vector” representing the state $|\psi\rangle = \alpha |0\rangle + \beta |1\rangle$. If we visualize the Bloch sphere as the planet Earth, then by convention, the north pole represents state $|0\rangle$ and the south pole state $|1\rangle$. For pure quantum states such as $|\psi\rangle$, the Bloch vector is of unit length, $|\alpha|^2 + |\beta|^2 = 1$, connecting the center of the sphere to any point on its surface.

The $z$-axis connects the north and south poles. It is called the “longitudinal axis,” since it represents the “qubit quantization axis” for the states $|0\rangle$ and $|1\rangle$ in the qubit eigenbasis. In turn, the $x$-$y$ plane is the “transverse plane” with “transverse axes” $x$ and $y$. In this Cartesian coordinate system, the unit Bloch vector $\vec{d} = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)$ is represented using the polar angle $0 \leq \theta \leq \pi$ and the azimuthal angle $0 \leq \phi < 2\pi$, as illustrated in Fig. 4(a). Following our convention, state $|0\rangle$ at the north pole is associated with $\theta = 1$, and state $|1\rangle$ (the south pole) with $\theta = 1$. We can similarly represent the quantum state using the angles $\theta$ and $\phi$. 

---

---
The Bloch vector is stationary on the Bloch sphere in the "rotating frame picture." If state $|1\rangle$ has a higher energy than state $|0\rangle$ (as it generally does in superconducting qubits), then in a stationary frame, the Bloch vector would precess around the $z$-axis at the qubit frequency $(E_1 - E_0)/\hbar$. Without loss of generality (and much easier to visualize), we instead "choose" to view the Bloch sphere in a reference frame where the $x$ and $y$-axes also rotate around the $z$-axis at the qubit frequency. In this "rotating frame," the Bloch vector appears stationary as written in Eq. (36). The rotating frame will be described in detail in Sec. IV D 1 in the context of single-qubit gates.

For completeness, we note that the density matrix $\rho = \langle \psi | \psi \rangle$ for a pure state $|\psi\rangle$ is equivalently

$$\rho = \frac{1}{2} (I + \hat{a} \cdot \hat{\sigma}) = \frac{1}{2} \begin{pmatrix} 1 + \cos \theta e^{i\phi} & \sin \theta e^{i\phi} \\ \sin \theta e^{-i\phi} & 1 - \cos \theta e^{-i\phi} \end{pmatrix},$$

(37)

$$= \begin{pmatrix} \cos^2 \frac{\theta}{2} e^{i\phi} & \sin^2 \frac{\theta}{2} e^{i\phi} \\ \sin^2 \frac{\theta}{2} e^{-i\phi} & \cos^2 \frac{\theta}{2} e^{-i\phi} \end{pmatrix},$$

(38)

$$= \frac{|\psi|^2}{2} \begin{pmatrix} 1 & \pm \beta \sqrt{\frac{\theta}{2}} \\ \beta^* \sqrt{\frac{\theta}{2}} & |\beta|^2 \end{pmatrix},$$

(39)

where $I$ is the identity matrix, and $\hat{\sigma} = [\sigma_x, \sigma_y, \sigma_z]$ is a vector of Pauli matrices. If the Bloch vector $\hat{a}$ is a unit vector, then $\rho$ represents a pure state $\psi$ and $\text{Tr}(\rho^2) = 1$. More generally, the Bloch sphere can be used to represent "mixed states," for which $|\hat{a}| < 1$; in this case, the Bloch vector terminates at points "inside" the unit sphere, and $0 \leq \text{Tr}(\rho^2) < 1$. To summarize, the surface of the unit sphere represents pure states, and its interior represents mixed states.

2. Bloch-Redfield model of decoherence

Within the standard Bloch-Redfield picture of two-level system dynamics, noise sources weakly coupled to the qubits have short correlation times with respect to the system dynamics. In this case, the relaxation processes are characterized by two rates (see Fig. 4),

longitudinal relaxation rate: $\Gamma_1 \equiv \frac{1}{T_1}$,

(40)

transverse relaxation rate: $\Gamma_2 \equiv \frac{1}{T_2} = \frac{\Gamma_1}{2} + \Gamma_\phi$,

(41)

which contains the pure dephasing rate $\Gamma_\phi$. We note that the definition of $\Gamma_2$ as a sum of rates presumes that the individual decay functions are exponential, which occurs for Lorentzian noise spectra (centered at $\omega = 0$) such as white noise (short correlation times) with a high-frequency cutoff.

The impact of noise on the qubit can be visualized on the Bloch sphere (Fig. 4(a)). For an initial state ($t = 0$)

$$|\psi\rangle = \alpha |0\rangle + \beta |1\rangle,$$

(42)

the Bloch-Redfield density matrix $\rho_{BR}$ for the qubit is written

$$\rho_{BR} = \begin{pmatrix} 1 + \left(\frac{|\psi|^2}{2}\right) & z\beta \sqrt{\frac{\theta}{2}} e^{-i\phi} e^{-\Gamma_1 t} \\ z^* \beta^* \sqrt{\frac{\theta}{2}} e^{i\phi} e^{-\Gamma_2 t} & |\beta|^2 e^{-\Gamma_1 t} \end{pmatrix}.$$  

(43)

There are a few important distinctions between Eqs. (43) and (39), which we list here and then describe in more detail in Secs. III B 2 a–III B 2 c.

- First, we have introduced the "longitudinal decay function" $e^{-\Gamma_1 t}$, which accounts for longitudinal relaxation of the qubit.
- Second, we introduced the "transverse decay function" $e^{-\Gamma_2 t}$, which accounts for transverse decay of the qubit.
- Third, we have introduced an explicit phase accrual $e^{i\phi t}$, where $\omega = \omega_q - \omega_n$, which generalizes the Bloch sphere picture to account for cases where the qubit frequency $\omega_q$ differs from the rotating-frame frequency $\omega_n$ as we will see later when discussing measurements of $T_2$ using Ramsey interferometry, and in Sec. IV D 1, in the context of single-qubit gates.
- Fourth, we have constructed the matrix such that for $t \gg (T_1, T_2)$, the upper-left matrix element will approach a unit value, indicating...
that all populations relax to the ground state, while the other three matrix elements decay to zero. This is related to the assumption that the environmental temperature is low enough that thermal excitations of the qubit from the ground to the excited state rarely occur.

a. Longitudinal relaxation. The longitudinal relaxation rate $\Gamma_1$ describes depolarization along the qubit quantization axis, often referred to as "energy decay" or "energy relaxation." In this language, a qubit with polarization $p = 1$ is entirely in the ground state ($|0\rangle$) at the north pole, $p = -1$ is entirely in the excited state ($|1\rangle$) at the south pole, and $p = 0$ is a completely depolarized mixed state at the center of the Bloch sphere.

As illustrated in Fig. 4(b), longitudinal relaxation is caused by "transverse noise," via the $x$- or $y$-axis, with the intuition that off-diagonal elements of an interaction Hamiltonian are needed to connect and drive transitions between states $|0\rangle$ and $|1\rangle$.

Depolarization occurs due to energy exchange with an environment, generally leading to both an "up transition rate" $\Gamma_{1u}$ (excitation from $|0\rangle$ to $|1\rangle$), and a "down transition rate" $\Gamma_{1d}$ (relaxation from $|1\rangle$ to $|0\rangle$). Together, these form the longitudinal relaxation rate $\Gamma_1$:

$$\Gamma_1 = \frac{1}{T_1} = \Gamma_{1u} + \Gamma_{1d}; \quad (44)$$

$T_1$ is the $1/e$ decay time in the exponential decay function in Eq. (43), and it is the characteristic time scale over which the qubit population will relax to its steady-state value. For superconducting qubits, this steady-state value is generally the ground state, due to Boltzmann statistics and typical operating conditions. Boltzmann equilibrium statistics lead to the "detailed balance" relationship $\Gamma_{1u} = \exp(-\hbar\omega_d/k_BT)\Gamma_{1d}$, where $T$ is the temperature and $k_B$ is the Boltzmann constant, with an equilibrium qubit polarization approaching $p = \tanh(\hbar\omega_d/(2k_BT))$. Typical qubits are designed at frequency $\omega_d/2\pi \approx 5$ GHz and are operated at dilution refrigerator temperatures $T \approx 20$ mK. In this limit, the up-rate $\Gamma_{1u}$ is exponentially suppressed by the Boltzmann factor $\exp(-\hbar\omega_d/k_BT)$, and so only the down-rate $\Gamma_{1d}$ contributes significantly, relaxing the population to the ground state. Thus, qubits generally spontaneously lose energy to their cold environment, but the environment rarely introduces a qubit excitation. As a result, the equilibrium polarization approaches unity [see Eq. (43)]..

Only noise at the qubit frequency mediates qubit transitions, whether absorption or emission, and this noise is generally "well behaved" (short correlation time, many modes weakly coupled to qubit, no divergences) around the qubit frequency for superconducting qubits. The intuition is that qubit-transition linewidths are relatively narrow in frequency, and so the noise generally does not vary much over this narrow frequency range. Although there are a few notable exceptions, for example, qubit decay in the presence of hot quasiparticles, which can lead to nonexponential decay functions, longitudinal depolarization measurements generally present exponential decay functions consistent with the Bloch-Redfield picture.

An example of a $T_1$ measurement is shown in Fig. 5(a). The qubit is prepared in its excited state using an $X_u$-pulse, and then left to spontaneously decay to the ground state for a time $\tau$, after which the qubit is measured. A single measurement will project the quantum state into either state $|0\rangle$ or state $|1\rangle$, with probabilities that correspond to the qubit polarization. To make an estimate of this polarization, one needs to independently prepare the qubit and repeat the experiment many times. This is analogous to flipping a coin: any single flip will yield heads or tails, but the probability of obtaining a heads or tails can be estimated by flipping the coin many times and taking the ensemble average. The resulting exponential decay has a characteristic time $T_1 = 85 \mu s$.

b. Pure dephasing. The "pure dephasing" rate $\Gamma_\phi$ describes depolarization in the $x$-$y$ plane of the Bloch sphere. It is referred to as "pure dephasing," to distinguish it from other phase-breaking processes such as energy excitation or relaxation.

As illustrated in Fig. 4(c), pure dephasing is caused by "longitudinal noise" that couples to the qubit via the $z$-axis. Such longitudinal noise causes the qubit frequency $\omega_0$ to fluctuate, such that it is no longer equal to the rotating frame frequency $\omega_R$ and causes the Bloch vector to precess forward or backward in the rotating frame. Intuitively, we can imagine independently preparing several instances of the Bloch vector along the $x$-axis. For each instance, the stochastic fluctuations of qubit frequency will result in a different precession frequency, resulting in a net fanout of the Bloch vector in the $x$-$y$ plane. This eventually leads to a complete depolarization of the azimuthal angle $\phi$. Note that this stochastic effect will be captured in the transverse relaxation rate $\Gamma_2$ (Sec. III B 2 c), it is "not" the deterministic term $\exp(\pm i\delta t)$ that appears in Eq. (43), which represents intentional detuning of the qubit reference frame.

There are a few important distinctions between pure dephasing and energy relaxation. First, in contrast to energy relaxation, pure dephasing is not a resonant phenomenon; noise at any frequency can modify the qubit frequency and cause dephasing. Thus, qubit dephasing is subject to broadband noise. Second, since pure dephasing is elastic (there is no energy exchange with the environment), it is in principle "reversible." That is, the dephasing can be "undone"—with quantum information being preserved—through the application of unitary operations, e.g., dynamical decoupling pulses, see Sec. III D 2.

The degree to which the quantum information can be retained depends on many factors, including the bandwidth of the noise, the rate of dephasing, the rate at which unitary operations can be performed, etc. This should be contrasted with spontaneous energy relaxation, which is an "irreversible" process. Intuitively, once the qubit emits energy to the environment and its myriad uncontrollable modes, the quantum information is essentially lost with no hope for its recovery and reconstitution back into the qubit.

c. Transverse relaxation. The transverse relaxation rate $\Gamma_2 = \Gamma_{1u}/2 + \Gamma_{1d}$ describes the loss of coherence of a superposition state, for example $(1/\sqrt{2})(|0\rangle + |1\rangle)$, pointed along the $x$-axis on the equator of the Bloch sphere as illustrated in Fig. 4(d). Decoherence is caused in part by longitudinal noise, which fluctuates the qubit frequency and leads to pure dephasing $\Gamma_\phi$ (red). It is also caused by transverse noise, which leads to energy relaxation of the excited-state component of the superposition state at a rate $\Gamma_1$ (blue). Such a relaxation event is also a phase-breaking process, because once it occurs, the Bloch vector points to the north pole, $|0\rangle$, and there is no longer any knowledge of which direction the Bloch vector had been pointing along the equator; the relative phase of the superposition state is lost.

Transverse relaxation $T_2$ can be measured using Ramsey interferometry, as shown and described in Fig. 5(b). The protocol positions

Appl. Phys. Rev. 6, 021318 (2019); doi: 10.1063/1.5089550
Published under license by AIP Publishing
the Bloch vector on the equator using a $X_p = 2$-pulse. Typically, the carrier frequency of this pulse is slightly detuned from the qubit frequency by an amount $\delta \nu$. As a result, the Bloch vector will precess around the $z$-axis at a rate $\delta \nu$. This is done for convenience sake, so that the resulting Ramsey measurement will oscillate, making it easier to analyze. After precessing for a time $\tau$, a second $X_p = 2$-pulse projects the Bloch vector back on to the $z$-axis. Repeated measurements are made to take an ensemble averaged estimate of the qubit polarization, +1 corresponding to $|0\rangle$ and −1 corresponding to $|1\rangle$. The resulting exponential decay function has a characteristic time $T_1 = 85 \mu s$. (b) Transverse relaxation (decoherence) measurement via Ramsey interferometry. The qubit is prepared on the equator using an $X_p = 2$-pulse, intentionally detuned from the qubit frequency by $\delta \nu$, causing the Bloch vector to precess in the rotating frame at a rate $\delta \nu$ around the $z$-axis. After a time $\tau$, a second $X_p = 2$ pulse then projects the Bloch vector back on to the $z$ axis, effectively mapping its former position on the equator to a position on the $z$ axis. The oscillations decay with an approximately (but not exactly) exponential decay function, with a characteristic time $T_2 = 120 \mu s$. The coherence improvement using the Hahn echo over panel (b) indicates that some low-frequency dephasing noise has been mitigated; however, a small amount remains since $T_2E$ has not yet reached the $2T_1$ limit. (d) Coherence function incorporating $T_1$ loss and Gaussian dephasing components of the Ramsey interferometry data in panel (b). The Gaussian-distributed $1/f$ noise spectrum of magnetic flux noise leads to a decay function $\exp \left( -\frac{t}{T_1} \right) \times \exp \left( -\frac{t^2}{T_{2E}^2} \right)$ in Eq. (45). These two decay functions together match well with the Ramsey data in panel (b).
With the known $T_1$ and $T_2$ times, one can infer the pure dephasing time $T_\nu$ from Eq. (41), provided the decay functions are exponential. In superconducting qubits, however, the broadband dephasing noise (e.g., flux noise, charge noise, critical-current noise, ... ) tends to exhibit a 1/f-like power spectrum. Such noise is singular near $\omega = 0$, has long correlation times, and generally does not fall within the Bloch-Redfield description. The decay function of the off-diagonal terms in Eq. (43) is generally nonexponential, and for such cases, the simple expression in Eq. (41) is not applicable.

3. Modification due to 1/f-type noise

If we assume that the qubit is coupled to many independent fluctuators, then, regardless of their individual statistics, they will in concert generate noise with a Gaussian distribution due to the central limit theorem. We therefore say that the longitudinal fluctuations exhibit Gaussian-distributed 1/f noise. For 1/f noise spectra, the phase decay function is itself a Gaussian $\exp\left(-\frac{t}{T_\nu G f}\right)$, where we write $T_{\nu G f}$ to distinguish it from $T_\nu$ used in Eq. (41). Furthermore, this function is separable from the $T_1$-type exponential decay, because the $T_1$-noise remains regular at the qubit frequency. The density matrix in Eq. (43) becomes, following Refs. 78 and 112,

$$\rho = \left( \begin{array}{cc} 1 + (|\gamma|^2 - 1)e^{-1/T_1} & \gamma^* \beta^* e^{i\delta(t)} e^{-\frac{t}{T_1}} e^{-\frac{t^2}{2T_1}} \gamma^* \beta^* e^{i\delta(t)} e^{-\frac{t}{T_1}} e^{-\frac{t^2}{2T_1}} \\ \beta e^{i\delta(t)} e^{-\frac{t}{T_1}} e^{-\frac{t^2}{2T_1}} \gamma & |\gamma|^2 e^{-2/T_1} \end{array} \right),$$

(45)

where the decay function (with $-\gamma_0(t)$) contains the “coherence function” $\gamma_0(t)$, which generalizes pure dephasing to include nonexponential decay functions. As we shall see later, the subscript $N$ labeling the decay function refers to the number of $\pi$-pulses used to refocus the low-frequency noise, which impacts the form of the decay function. Because the function is no longer purely exponential, we cannot formally write the transverse relaxation decay function as $\exp(-t/T_2)$. However, an exponential decay remains a practically reasonable approximation for $T_\nu \gg T_1$. We also note that the energy decay component of the transverse relaxation is $\exp(-t/2T_1)$, and so $T_2$ cannot be smaller than $2T_1$. In the absence of pure dephasing, the maximum $T_2 = 2T_1$ is reached.

As an example, consider the Ramsey interferometry data in Fig. 5(b). Since the dephasing is relatively weak, the transverse relaxation function as $\exp(-t/T_2)$ is a reasonable fit and yields $T_2 = 95 \mu$s. However, using the value $T_1 = 85 \mu$s from Fig. 5(a) and dividing out $\exp(-t/2T_1)$ from the data in Fig. 5(b), the remaining pure dephasing decay function is shown in Fig. 5(d) and assumes a Gaussian envelope (with $-\gamma_0(t)$) $\exp\left(-\frac{t}{T_{\nu G f}}\right)$, with $T_{\nu G f} = 98 \mu$s. The Hahn echo data in Fig. 5(c) may be treated similarly.

For completeness, in addition to 1/f dephasing mechanisms, we note that there are also “white” pure dephasing mechanisms, which give rise to an exponential decay function for the dephasing component of $T_2$. One common example is dephasing due to the shot noise of residual photons in the readout resonator coupled to superconducting qubits, as we discuss in Sec. III C 3.

4. Noise power spectral density (PSD)

The frequency distribution of the noise power for a stationary noise source $\lambda$ is characterized by its PSD $S_\lambda(\omega)$.

$$S_\lambda(\omega) = \int_{-\infty}^{\infty} dt \langle \hat{\lambda}(\tau) \hat{\lambda}(0) \rangle e^{i\omega \tau}. \quad \text{(46)}$$

The Wiener-Khintchine theorem states that the PSD is the Fourier transform of the autocorrelation function $c_\lambda(\tau) = \langle \hat{\lambda}(\tau) \hat{\lambda}(0) \rangle$ of the noise source $\lambda$. Since the integration limits are $(-\infty, \infty)$, this is the bilateral PSD. Symmetrizing the PSD allows one to consider only positive frequencies, which is termed a unilateral PSD. Both unilateral and bilateral PSDs are used, often with the same notation, and so one needs to know how the PSD is defined, to keep track of the factors 2 and $\pi$, and also be aware of the implications for quantum systems.

For classical systems, the noise power spectral density is symmetric. This is because the autocorrelation function of real signals is itself a real function, and the Fourier transform of a real temporal function is symmetric in the frequency domain. Dephasing noise is caused by real, fluctuating fields, and so its PSD is generally symmetric. Examples of such classical noise include thermal (Johnson) noise and 1/f noise [35] (see Fig. 6).

In turn, the inverse Fourier transform of the PSD will yield the autocorrelation function

$$c_\lambda(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega S_\lambda(\omega)e^{-i\omega \tau}. \quad \text{(47)}$$

This implies that integrating the noise power spectral density with $\tau = 0$ yields the second moment of the noise, or, for zero-mean fluctuations, the variance.

However, the autocorrelation function for a quantum system may be complex-valued due to the fact that quantum operators generally do not commute at different times. This means that time-ordering of the operators matters, and the PSD need not be symmetric in frequency. This is generally the case for transverse noise causing...
longitudinal energy relaxation. Noise at a positive frequency $S(\omega_0)$ corresponds to energy transfer from the qubit to the environment, including both stimulated and spontaneous emission, associated with the down-rate $\Gamma_1$. Noise at a negative frequency $S(-\omega_0)$ corresponds to energy transfer to the qubit from the environment, associated with the up-rate $\Gamma_1$. This energy transfer becomes exponentially suppressed when the qubit frequency is larger than thermal energy ($k_B T$), as shown in Fig. 6. For a detailed discussion, see Refs. 123 and 124. Spontaneous emission to a cold environment or electromagnetic vacuum, represented by Nyquist noise in Fig. 6, is an example of an asymmetric noise PSD.

In general, making a connection between $S_i(\omega)$ and the measured qubit decay functions is the basis for noise spectroscopy up to second-order statistics. The search for higher-order spectra related to non-Gaussian noise is a current topic of active research.

C. Common examples of noise

There are many sources of stochastic noise in superconducting qubits, and we refer the reader to Ref. 40 for a review. Here, we briefly present several of the most common types of noise, their effect on coherence, and refer the reader to the references for a more detailed discussion.

1. Charge noise

"Charge noise" is ubiquitous in solid-state devices. It arises from charged fluctuators present in the defects or charge traps that reside in interfacial dielectrics, the junction tunnel barrier, and in the substrate itself. These are often modeled as an ensemble of fluctuating two-level systems or as bulk dielectric loss. For example, in the case of a transmon qubit, the electric field between the capacitor plates traverses and couples to dielectric defects residing on the metal surfaces of the plates (for lateral-plate-type capacitors) or the capacitor dielectric between the plates (for parallel-plate-type capacitors). The electric field variable is transverse with respect to the quantization axis of the transmon qubit, which means that this noise is mainly responsible for energy relaxation ($T_1$). Additionally, if the $E_0/E_C$ ratio of the transmon is not made sufficiently large (smaller than around 60), the qubit frequency itself will also be sensitive to broadband charge fluctuations. In this case, low-frequency charge noise couples longitudinally to the transmon and causes pure dephasing ($T_\phi$).

Charge noise is modeled primarily as a combination of inverse-frequency noise and Nyquist noise, also referred to as "ohmic" noise. At lower frequencies, the spectral density takes the form

$$S_Q(\omega) = A_Q^2 \left( \frac{2\pi \times 1\text{Hz}}{\omega} \right) \gamma_Q \nu, \quad (48)$$

with quasiuniversal values $A_Q = (10^{-6} \text{pF}^2)/\text{Hz}$ at 1 Hz, and $\gamma_Q \approx 1$. In addition to large $1/f$ fluctuations, early charge qubits often witnessed discrete, charge offsets reminiscent of random telegraph noise. Together, these two mechanisms severely limited the utility of charge qubits, and served as a strong motivation to move to capacitively shunted charge qubits (transmons), which greatly reduced the qubit longitudinal sensitivity to charge noise. At higher frequencies, the power spectrum takes the form $S_Q(\omega) = B_Q^2 \omega / (2\pi \times 1\text{Hz})$, where the noise strength $B_Q^2$ at 1 Hz can assume a range of values depending on the level of dissipation in the system. Likewise, the cross-over from $1/f$-like behavior to $f$-like behavior generally occurs at around 1 GHz, but will vary higher or lower between samples depending on the degree of dissipation.

2. Magnetic flux noise

Another commonly observed noise in solid-state devices is magnetic "flux noise." The origin of this noise is understood to arise from the stochastic flipping of spins (magnetic dipoles) that reside on the surfaces of the superconducting metals comprising the qubit, resulting in random fluctuations of the effective magnetic field that biases flux-tunable qubits.

For example, in the case of the split transmon, the external magnetic field threading the loop couples longitudinally to the qubit and modulates the transition frequency via the Josephson energy $E_J$ (except at $\phi_0 = 0$, where the qubit is first-order insensitive to magnetic-field fluctuations). Because the flux noise is longitudinal to the transmon, it contributes to pure dephasing ($T_\phi$). However, in the case of the flux qubit, and depending on the flux-bias point, the flux noise may be either longitudinal—causing dephasing $T_\phi$—or it may couple transversely and thus contribute to $T_1$ relaxation. The noise power spectrum of these fluctuations generally exhibits a "quasiuniversal" dependence

$$S_\phi(\omega) = A_\phi^2 \left( \frac{2\pi \times 1\text{Hz}}{\omega} \right)^{\nu}, \quad (49)$$

with $A_\phi \approx 0.8 - 1.0$ and $A_\phi^2 \approx (1 \mu \text{Hz})^2/\text{Hz}$, and has been shown to extend from less than millihertz to beyond gigahertz frequencies.

The large, low-frequency weighting of the $1/f$ power distribution enables the use of engineered error mitigation techniques—such as dynamical decoupling—to achieve better coherence and for improving single and two-qubit gate fidelity. It was recently demonstrated that $1/f$ flux noise is also a $T_1$ mechanism when extended out to the qubit frequency, and one similarly expects a crossover to ohmic flux noise at high enough frequencies.

Although much is known about the statistics and number of the defects presumed responsible for flux noise, their precise physical manifestation remains uncertain. The fact that the $1/f$ noise is quasiuniversal and largely independent of device, strongly suggests a common origin for the noise. Recent studies suggest that adsorbed molecular oxygen may be responsible for flux-noise.

3. Photon number fluctuations

In the circuit QED architecture, resonator "photon number fluctuation" is another major decoherence source. Residual microwave fields in the cavity have photon-number fluctuations that in the dispersive regime impact the qubit through an interaction term $\gamma Q, n$, see Sec. II C 2, leading to a frequency shift $\Delta_{\text{shift}} = 2\eta Q n$, where $n$ is the average photon number, and $\eta = e^2/(\kappa^2 + 4\pi^2)$ effectively scales the photon population seen by the qubit due to the interplay between the qubit-induced dispersive shift of the resonator frequency ($\chi$) and the resonator decay rate ($\kappa$).

In the dispersive limit, the noise is longitudinal coupled to the qubit and leads to pure dephasing at a rate...
The corresponding noise spectral density is of Lorentzian type

\[ S(\omega) = \frac{4\gamma^2}{\omega^2 + \omega_0^2}, \]

which exhibits an essentially white noise spectrum up to a 3 dB cutoff frequency \( \omega = \kappa \) set by the resonator decay rate \( \kappa \), see Ref. 62.

### 4. Quasiparticles

"Quasiparticles," i.e., unpaired electrons, are another important noise source for superconducting devices.\textsuperscript{119,120} The tunneling of quasiparticles through a qubit junction may lead to both relaxation and pure dephasing \( T_1 \), depending on the type of qubit, the bias point, and the junction through which the tunneling event occurs.\textsuperscript{119,120} Quasiparticles are naturally excited due to thermodynamics, and the quasiparticle density in equilibrium superconductors should be exponentially suppressed as the temperature decreases. However, below about 150 mK, the quasiparticle density observed in superconducting devices—generally in the range \( 10^{-8} - 10^{-6} \) per Cooper pair—is much higher than what the BCS theory would predict for a superconductor in equilibrium with its cryogenic environment at 10 mK. The reason for this excess quasiparticle population is unclear, but it is very likely related to the presence of additional, nonthermal mechanisms that increase the generation rates, "bottleneck effects" that occur at millikelvin temperatures to reduce recombination rates, or a combination of both.

It has been shown that the observed \( T_1 \) and excess excited-state population measured in today’s state-of-the-art high-coherence transmon are self-consistent with excess "hot" nonequilibrium quasiparticles at the quasuniversal density of around \( 10^{-7} - 10^{-6} \) per Cooper pair.\textsuperscript{119,120} Although this quasiparticle generation mechanism is not yet well understood, it has been shown that quasiparticles can be transiently pumped away, improving \( T_1 \) times and reducing \( T_1 \) temporal variation.\textsuperscript{120}

### D. Operator form of qubit-environment interaction

Similar to the way two qubits are coupled, a qubit may couple and interact with uncontrolled degrees of freedom (DOF) in its environment (the noise sources). The interaction Hamiltonian between the qubit DOF (\( \hat{O}_q \)) and those of the noise source (\( \hat{\lambda} \)) may be expressed in a general form

\[ \hat{H}_{\text{int}} = \nu \hat{O}_q \hat{\lambda}, \]

where \( \nu \) denotes the coupling strength—which is related to the sensitivity of the qubit to environmental fluctuations \( \partial \hat{H}_q / \partial \hat{\lambda} \)—and we assume that \( \hat{O}_q \) is a qubit operator within the qubit Hamiltonian \( \hat{H}_q \). The noisy environment represented by the operator \( \hat{\lambda} \) produces fluctuations \( \delta \hat{\lambda} \). Note that we retained the hats in this section to remind us that these are quantum operators.

### 1. Connecting \( T_1 \) to \( S(\omega) \)

If the coupling is transverse to the qubit, e.g., \( \hat{O}_q \) is of the type \( \sigma_x \) or \( (a + a^\dagger) \)—see the related case of qubit-qubit coupling treated in Sec. II C—then noise at the qubit frequency can cause transitions between the qubit eigenstates. Since this is a stochastic process, the ensemble-average manifests itself as a decay (usually exponential) of the qubit population toward a certain equilibrium value (usually the qubit ground state \( |0\rangle \) for \( k_B T \ll \hbar \omega_q \)). Again, this process is equivalently referred to as "\( T_1 \) relaxation." "Energy relaxation," or "longitudinal relaxation." As stated above, \( T_1 \) is the characteristic time scale of the decay. Its inverse, \( \Gamma_1 = 1/T_1 \), is called the relaxation rate and depends on the power spectral density of the noise \( S(\omega) \) at the transition frequency of the qubit \( \omega = \omega_q \)

\[ \Gamma_1 = \frac{1}{\hbar} \left| \int \frac{\partial \hat{H}_q}{\partial \hat{\lambda}} \langle 1 \rangle \right|^2 S_q(\omega_q), \]

where \( \partial \hat{H}_q / \partial \hat{\lambda} \) is the qubit transverse susceptibility to fluctuations \( \delta \hat{\lambda} \), such that \( \langle \delta \hat{\lambda} \rangle^2 \) is the ensemble average value of the environmental noise sources as seen by the qubit. Equation (53) is equivalent to Fermi’s Golden Rule, in which the qubit’s transverse susceptibility to noise is driven by the noise power spectral density. The qubit transverse susceptibility can be used to calculate the prefactors; for example, for fluctuations \( \delta \hat{\lambda} = \delta n \), the relevant term in the transmon Hamiltonian in Eq. (16) is \( 4E_C(n - n_q)^2 \), where we allow for an offset charge \( n_w \) and the susceptibility is given by \( 8E_C \delta n \). We refer the reader to Refs. 40, 146, and 147 for more details.

### 2. Connecting \( T_2 \) to \( S(\omega) \)

If the coupling to the qubit is instead longitudinal, e.g., \( \hat{H}_q \) is of the type \( \sigma_z \) or \( a^\dagger a \), the noise will stochastically modulate the transition frequency of the qubit and thereby introduce a stochastic phase evolution of a qubit superposition state. This gradually leads to a loss of phase information, and it is therefore called pure dephasing (time constant \( T_2 \)). Unlike \( T_1 \) relaxation, which is generally an irreversible (incoherent) error, pure dephasing \( T_2 \) is in principle reversible (a coherent error). The degree of pure dephasing depends on the control pulse sequence applied while the qubit is subject to the noise process.

Consider the relative phase \( \phi \) of a superposition state undergoing free evolution

\[ \phi(t) = \int_0^t \omega_q dt' + \langle \omega_q \rangle t + \delta \phi(t) \]

diffuses due to adiabatic fluctuations of the transition frequency,

\[ \delta \phi(t) = -\frac{\partial \omega_q}{\partial \lambda} \int_0^t \delta \lambda(t') dt', \]

where \( \partial \omega_q / \partial \lambda = (1/\hbar)[(\partial \hat{H}_q / \partial \hat{\lambda})] \) is the qubit’s longitudinal sensitivity to \( \lambda \)-noise. For noise generated by a large number of fluctuators that are weakly coupled to the qubit, its statistics are Gaussian. Ensemble averaging over all realizations of the Gaussian-distributed stochastic process \( \delta \lambda(t) \), the dephasing is

\[ \langle e^{i\delta \phi(t)} \rangle = e^{-\frac{1}{2} \langle \delta \phi(t)^2 \rangle} \approx e^{-\langle \ln(\gamma^2 / 2\kappa^2) \rangle}, \]

leading to a coherence decay function.
where \( g(\omega, \tau) \) is a dimensionless weighting function.

The function \( g_N(\omega, \tau) \) can be viewed as a frequency-domain filter of the noise \( S_1(\omega) \) [see Fig. 7(a)]. In general, its filter properties depend on the number \( N \) and distribution of applied pulses. For example, considering sequences of \( \pi \)-pulses,

\[
g_N(\omega, \tau) = \frac{1}{(\omega \tau)^2} \left( 1 + (-1)^{N+1} \exp(i\omega \tau) \right) + 2 \sum_{j=1}^{N} (-1)^{j} \exp(i\omega_j \tau) \cos(\omega \tau_j / 2),
\]

where \( \delta_j \in [0, 1] \) is the normalized position of the center of the \( j \)th \( \pi \)-pulse between the two \( \pi/2 \)-pulses, \( \tau \) is the total free-induction time, and \( \tau_0 \) is the length of each \( \pi \)-pulse, yielding a total sequence length \( \tau + N\tau_0 \). As the number of pulses increases for a fixed \( \tau \), the filter function’s peak shifts to higher frequencies, leading to a reduction in the net integrated noise for \( 1/f^3 \)-type noise spectra with \( x > 0 \). Similarly, for a fixed \( N \), the filter function will shift in frequency with \( \tau \). Additionally, for a fixed time separation \( \tau' = \tau/N \) ([valid for \( N \geq 1 \)], the filter sharpens and asymptotically peaks at \( \omega' / 2\pi = 1/2\tau' \) as more pulses are added. \( g_N(\omega, \tau) \) is thus called the "filter function," \([58,150]\) and it depends on the pulse sequences being applied.

From Eq. (57), the pure dephasing decay arises from a noise spectral density that is "shaped" or "filtered" by the sequence-specific filter function. By choosing the number of pulses, their rotation axes, and their arrangement in time, we can design filter functions that minimize the net noise power for a given noise spectral density within the experimental constraints of the experiment (e.g., pulse-modulation bandwidth of the electronics used to control the qubits).

To give a standard example, we compare the coherence integral for two cases: a Ramsey pulse sequence and a Hahn echo pulse sequence. Both sequences involve two \( \pi/2 \) pulses separated by a time \( \tau \), during which free evolution of the qubit occurs in the presence of low-frequency dephasing noise. The distinction is that the Hahn echo will place a single \( \pi \) pulse \((N = 1)\) in the middle of the free-evolution period, whereas the Ramsey does not use any additional pulses \((N = 0)\). The resulting filter functions are

\[
g_0(\omega, \tau) = \frac{\sin^2(\omega \tau / 2)}{2},
\]

\[
g_1(\omega, \tau) = \frac{\sin^2(\omega \tau / 4)}{4},
\]

where the subscripts \( N = 0 \) and \( N = 1 \) indicate the number of \( \pi \)-pulses applied for the Ramsey and Hahn echo experiments, respectively. The filter function \( g_0(\omega, \tau) \) for the Ramsey case is a sinc-function centered at \( \omega = 0 \). For noise that decreases with frequency, e.g., \( 1/f \) flux noise in superconducting qubits, the Ramsey experiment windows through the noise in \( S(\omega) \) where it has its highest value. This is the worst choice of filter function for \( 1/f \) noise. In contrast, the Hahn echo filter function has a centroid that is peaked at a higher frequency, away from \( \omega = 0 \). In fact, it has zero value at \( \omega = 0 \). For noise that decreases with frequency, such as \( 1/f \) noise, this is advantageous. This concept extends to larger numbers \( N \) of \( \pi \) pulses, and is called a Carr-Purcell-Meiboom-Gill (CPMG) sequence.\([53,154]\) In Fig. 7(b), the \( T_2 \) time of a qubit under the influence of strong dephasing noise is increased toward the \( 2T_1 \) limit using a CPMG dynamical error-suppression pulse sequence with an increasing number of pulses, \( N \). We refer the reader to Refs. 78, 155, and 156, where these experiments were performed with superconducting qubits.

3. Noise spectroscopy

The qubit is highly sensitive to its noisy environment, and this feature can be used to map out the noise power spectral density. In general, one can map the noise PSD during “free evolution”—periods of time for which no control is applied to the qubit, except for very short dynamical decoupling pulses—and during “driven evolution”—periods of time during which the control fields are applied to the qubit.
Both free-evolution and driven-evolution noise is important to characterize, as the noise PSD may differ for these two types of evolution, and both are utilized in the context of universal quantum computation. We refer the reader to Ref. 128 for a summary of noise spectroscopy during both types of evolution.

The Ramsey frequency itself is sensitive to longitudinal noise, and monitoring its fluctuations is one means to map out the noise spectral density over the submillihertz to about 100 Hz range.27,45

At higher frequencies, the CPMG dynamical decoupling sequence can be used to create narrow-band filters that “sample” the noise at different frequencies as a function of the free-evolution time $\tau$ and the number of pulses $N$. This has been used to map out the noise PSD in the range 0.1–300 MHz.78 One must be careful of the additional small peaks at higher-frequencies, which all contribute to the dephasing used to perform the noise spectroscopy.98

In fact, using pulse envelopes such as Slepians150—which are designed to have a concentrated frequency response—to perform noise spectroscopy is one means to reduce such errors.151

At even higher frequencies, measurements of $T_1$ can be used in conjunction with Fermi’s golden rule to map out the transverse noise spectrum above 1 GHz.62,156

The aforementioned are all examples of noise spectroscopy during free evolution. Noise spectroscopy during driven evolution was also demonstrated using a “spin-locking” technique, where a strong drive along $x$ or $y$ axes defines a new qubit quantization axis, whose Rabi frequency is the new qubit frequency in the spin-locking frame. The spin-locking frame is then used to infer the noise spectrum while the qubit is continually subject to a driving field. For more information, we refer the reader to Ref. 128.

E. Engineering noise mitigation

Here, we briefly review a few examples of techniques that have been developed to reduce noise or reduce its impact on decoherence (sensitivity). We stress that improving gate fidelity is a comprehensive optimization task, one that is full of trade-offs. It is thus important to identify what the limiting factors are, what price we have to pay to diminish these limiting factors, and what advantage we can achieve until reaching a better trade-off. These all require an accurate understanding of the gate fidelity, the sources of decoherence, the properties of the noise, and how it affects the system performance.

1. Materials and fabrication improvements

Numerous efforts have been undertaken to reduce noise-induced defects due to materials and fabrication.60,104 In the case of charge noise, significant efforts have been made to reduce the number of defects, such as substrate cleaning,99,105 substrate annealing,106 and trenching.41,107 In the case of flux noise, several groups have performed experiments to characterize the behavior and properties of magnetic-flux defects.13,164,186

More recently, a number of groups have tried optical surface treatments to remove these defects.167,168

In the context of residual quasiparticles, it has been shown that adding quasiparticle traps to the circuit design can reduce the quasiparticle number, particularly in devices that create excess quasiparticles, such as classical digital logic or operation in the presence of thermal radiation.165

2. Design improvements

Another strategy is to reduce qubit sensitivity to the noise by design. A qubit can only lose energy to defects if it couples to them. It has been demonstrated that altering the capacitor geometry to increase the electric-field mode volume reduces the electric field density in the thin dielectric regions that cause loss. This effectively reduces the “participation” of the defects and makes the qubits less sensitive to these noise sources.

In another example, the split transmons built using asymmetric junctions have lower sensitivity to flux noise than their symmetric counterparts at the expense of decreased frequency tunability.69 This is a good trade-off to make, because generally one is interested in tuning the qubit frequency over a somewhat restricted range (typically around 1 GHz) about the qubit frequency. When such asymmetric transmons are used in a gate scheme such as the adiabatic CPHASE gate,65 (see Sec. IV F) the qubit is less sensitive to flux noise, has a lower dephasing rate, and this should improve the gate fidelity in general.

3. Dynamical error suppression

As introduced in Sec. III D 2, it is advantageous to leverage the 1/$\omega$ distribution of flux noise, wherein a considerable amount of the noise power resides at low frequencies, and so the noise is “quasistatic.” The spin-echo technique,115 which disrupts the free evolution by a $\pi$-pulse, is extremely effective in mitigating the pure dephasing by refocusing the coherent phase dispersion due to low-frequency noise. The more advanced versions, such as the CPMG-sequence, use multiple $\pi$-pulses to interrupt the system more frequently, pushing the filter band to even higher frequencies—a technique known as “dynamical decoupling.”117

Returning to excess quasiparticles, it has been shown that quasiparticles can be stochastically pumped away from the qubit region, resulting in longer, and more stable $T_1$ times.129 Although the pumping technique uses a series of $\pi$-pulses, this technique differs from dynamical error suppression of coherent errors in that pulses are stochastically applied, and that it addresses incoherent errors ($T_1$).

4. Cryogenic engineering

In the case of photon shot-noise, in addition to applying dynamical decoupling techniques, there have been several recent studies aimed at reducing the thermal photon flux that reaches the device. This include optimizing the attenuation of the cryogenic setup,108,144,145,147 remaking the cryogenic attenuators with more efficient heat sinking,143 adding absorptive “black” material to absorb stray thermal photons,148,149 and adding additional cavity filters for thermalization.170

IV. QUBIT CONTROL

In this section, we will introduce how superconducting qubits are manipulated to implement quantum algorithms. Since the transmon-like variety of superconducting qubits has so far been the most widely deployed modality for implementing quantum programs, the discussion throughout this section will be focused on modern techniques for transmons. Nonetheless, the techniques introduced here are applicable to all types of superconducting qubits.
We start with a brief review of the gates used in classical computing as well as quantum computing, and the concept of universality. Subsequently we discuss the most common technique of driving single qubit gates via a capacitive coupling of a microwave line, coupled to the qubit. We introduce the notion of “virtual Z gates” and “DRAG” pulsing. In the latter part of this section, we review some of the most common implementations of two-qubit gates in both tunable and fixed-frequency transmon qubits. The single-qubit and two-qubit operations together form the basis of many of the medium-scale superconducting quantum processors that exist today.

Throughout this section, we write everything in the computational basis \( |0\rangle, |1\rangle \) where \( |0\rangle \) is the +1 eigenstate of \( \sigma_z \) and \( |1\rangle \) is the –1 eigenstate. We use capitalized serif-fonts to indicate the rotation operator of a qubit state, e.g., rotations around the \( \hat{x} \) axis by an angle \( \theta \) is written as

\[
X_\theta = R_x(\theta) = e^{-i \Phi_\theta} = \cos(\theta/2) \hat{x} - i \sin(\theta/2) \sigma_x
\]

and we use the shorthand notation “\( X \)” for a full \( \pi \) rotation about the \( x \) axis (and similarly for \( Y := Y_x \) and \( Z := Z_x \)).

### A. Boolean logic gates used in classical computers

Universal Boolean logic can be implemented on classical computers using a small set of single-bit and two-bit gates. Several common classical logic gates are shown in Fig. 8 along with their truth tables. In classical Boolean logic, bits can take on one of two values: state 0 or state 1. The state 0 represents logical FALSE, and state 1 represents logical TRUE.

Beyond the trivial “identity operation,” which simply passes a Boolean bit unchanged, the only other possible single-bit Boolean logic gate is the NOT gate. As shown in Fig. 8, the NOT gate flips the bit: 0 \( \rightarrow \) 1 and 1 \( \rightarrow \) 0. This gate is reversible, because it is trivial to determine the input bit value given the output bit values. As we will see, for two-bit gates, this is not the case.

There are several two-bit gates shown in Fig. 8. A two-bit gate takes two bits as inputs, and it passes as an output the result of a Boolean operation. One common example is the AND gate, for which the output is 1 if and only if both inputs are 1; otherwise, the output is 0. The AND gate, and the other two-bit gates shown in Fig. 8, are all examples of irreversible gates; that is, the input bit values cannot be inferred from the output values. For example, for the AND gate, an output of logical 1 uniquely identifies the input 11, but an output of 0 could be associated with 00, 01, or 10. Once the operation is performed, in general, it cannot be “undone” and the input information is lost. There are several variants of two-bit gates, including:

- AND and OR;
- NAND (a combination of NOT and AND) and NOR (a combination of NOT and OR);
- XOR (exclusive OR) and NXOR (NOT XOR).

The XOR gate is interesting, because it is a “parity” gate. That is, it returns a logical 0 if the two inputs are the same values (i.e., they have the same parity), and it returns a logical 1 if the two inputs have different values (i.e., different parity). Still, the XOR and NXOR gates are not reversible, because knowledge of the output does not allow one to uniquely identify the input bit values.

<table>
<thead>
<tr>
<th>GATE</th>
<th>CIRCUIT SYMBOL</th>
<th>TRUTH TABLE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NOT</strong></td>
<td>The output is 1 when the input is 0 and 0 when the input is 1.</td>
<td>![NOT Circuit Diagram]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>AND</strong></td>
<td>The output is 1 only when both inputs are 1, otherwise the output is 0.</td>
<td>![AND Circuit Diagram]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>OR</strong></td>
<td>The output is 0 only when both inputs are 0, otherwise the output is 1.</td>
<td>![OR Circuit Diagram]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>NAND</strong></td>
<td>The output is 0 only when both inputs are 1, otherwise the output is 1.</td>
<td>![NAND Circuit Diagram]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>NOR</strong></td>
<td>The output is 1 only when both inputs are 0, otherwise the output is 0.</td>
<td>![NOR Circuit Diagram]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>XOR</strong></td>
<td>The output is 1 only when the two inputs have different value, otherwise the output is 0.</td>
<td>![XOR Circuit Diagram]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>XNOR</strong></td>
<td>The output is 1 only when the two inputs have the same value, otherwise the output is 0.</td>
<td>![XNOR Circuit Diagram]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

FIG. 8. Classical single-bit and two-bit Boolean logic gates. For each gate, the name, a short description, circuit representation, and input/output truth tables are presented. The numerical values in the truth table correspond to the classical bit values 0 and 1. Adapted from Ref. 171.
The concept of “universality” refers to the ability to perform any Boolean logic algorithm using a small set of single-bit and two-bit gates. A universal gate set can in principle transform any state to any other state in the state space represented by the classical bits. The set of gates which enable universal computation is not unique, and may be represented by a small set of gates. For example, the NOT gate and the AND gate together form a universal gate set. Similarly, the NAND gate itself is universal, as is the NOR gate. The efficiency with which one can implement arbitrary Boolean logic, of course, depends on the choice of the gate set.

B. Quantum logic gates used in quantum computers

Quantum logic can similarly be performed by a small set of single-qubit and two-qubit gates. Qubits can of course assume the classical states $|0\rangle$ and $|1\rangle$, at the north pole and south pole of the Bloch sphere, but they can also assume arbitrary superpositions $\alpha|0\rangle + \beta|1\rangle$, corresponding to any other position on the sphere.

Single-qubit operations translate an arbitrary quantum state from one point on the Bloch sphere to another point by rotating the Bloch vector (spin) a certain angle about a particular axis. As shown in Fig. 9, there are several single-qubit operations, each represented by a matrix that describes the quantum operation in the computational basis represented by the eigenvectors of the $\sigma_x$ operator, i.e., $|0\rangle \equiv [1 0]^T$ and $|1\rangle \equiv [0 1]^T$.

For example, the “identity gate” performs no rotation on the state of the qubit. This is represented by a two-by-two identity matrix. The $X$-gate performs a $\pi$ rotation about the $x$ axis. Similarly, the $Y$-gate and $Z$-gate perform a $\pi$ rotation about the $y$ and $z$ axes, respectively. The $S$-gate performs a $\pi/2$ rotation about the $z$ axis, and the $T$-gate performs a rotation of $\pi/4$ about the $z$ axis. The Hadamard gate $H$ is also a common single-qubit gate that performs a $\pi$ rotation about an axis diagonal in the $x$-$z$ plane, see Fig. 9.

Two-qubit quantum-logic gates are generally “conditional” gates and take two qubits as inputs. Typically, the first qubit is the “control” qubit, and the second is the “target” qubit. A unitary operator is applied to the target qubit, dependent on the state of the control qubit. The two common examples shown in Fig. 10 are the controlled NOT (CNOT-gate) and controlled phase (CZ or CPHASE gate). The CNOT-gate flips the state of the target qubit conditioned on the control qubit being in-state $|1\rangle$. The CPHASE-gate applies a $Z$ gate $\sigma_z$ to the target qubit, conditioned on the control qubit being in-state $|1\rangle$. As we will show later, the $i$SWAP gate—another two-qubit gate—can be built from the CNOT-gate and single-qubit gates. The unitary operator of the CNOT gate can be written in a useful way, highlighting that it applies an $X$-gate (a $\sigma_x$ operator) $X$ depending on the state of the control qubit

$$U_{\text{CNOT}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes \sigma_x,$$  \hspace{1cm} (62)

and similarly for the CPHASE gate

$$U_{\text{CPHASE}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} = |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes \sigma_z.$$  \hspace{1cm} (63)

Comparing the last equality above with the unitary for the CNOT [Eq. (62)], it is clear that the two gates are closely related. Indeed, a CNOT can be generated from a CPHASE by applying two Hadamard gates

$$U_{\text{CNOT}} = (I \otimes H)U_{\text{CPHASE}}(I \otimes H),$$  \hspace{1cm} (64)

since $HZH = X$. Due to the form of Eq. (63). The CPHASE gate is also denoted the CZ gate, since it implements a controlled $Z$ gate (a controlled-$\sigma_z$ operation), by analogy with CNOT (a controlled application of the $X$-gate, i.e., the $\sigma_x$ operation). Inspection of the definition of CPHASE in Fig. 10 makes no distinction between which qubit acts as the target and which as the control and, consequently, the circuit-diagram is sometimes drawn in a symmetric fashion

$$\text{CPHASE}.$$  \hspace{1cm} (65)

The CNOT (with qubit 1 as control and qubit 2 as target) can be realized in terms of the CPHASE operation and single-qubit Hadamard gates,

$$\text{CNOT} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$  \hspace{1cm} (66)

Some two-qubit gates such as CNOT and CPHASE are also called “entangling gates,” because they can take product states as inputs and output entangled states. They are thus an indispensable component of a universal gate set for quantum logic. For example, consider two qubits $A$ and $B$ in the following state:

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)|A\rangle|B\rangle.$$  \hspace{1cm} (67)

If we perform a CNOT gate, $U_{\text{CNOT}}$, on this state, with qubit $A$ the control qubit and qubit $B$ the target qubit, the resulting state is (see the truth table in Fig. 10)

$$U_{\text{CNOT}}|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B) \neq \prod_i |\rangle_i \prod_j |\rangle_j,$$  \hspace{1cm} (68)

which is a state that cannot be factored into an isolated qubit-A component and a qubit-B component. This is one of the two-qubit entangled “Bell states,” a manifestly quantum mechanical state.

A universal set of single-qubit and two-qubit gates is sufficient to implement an arbitrary quantum logic. This means that this gate set can in principle reach “any” state in the multiqubit state-space. How efficiently this is done depends on the choice of quantum gates that comprise the gate set. We also note that each of the single-qubit and two-qubit gates is reversible, that is, given the output state, one can uniquely determine the input state. As we discuss further, this distinction between classical and quantum gates arises, because quantum gates are based on “unitary” operations $U$. If a unitary operation $U$ is a particular gate applied to a qubit, then its Hermitian conjugate $U^\dagger$ can be applied to recover the original state, since $U^\dagger U = I$ resolves an identity operation.

C. Comparing classical and quantum gates

The gate-sequences used to represent quantum algorithms have certain similarities to those used in classical computing, with a few striking differences. As an example, we consider first the classical NOT gate (discussed previously), and the related quantum circuit version, shown in Fig. 11.
FIG. 9. Quantum single-qubit gates. For each gate, the name, a short description, circuit representation, matrix representation, input/output truth tables, and Bloch sphere representation are presented. Matrices are defined in the basis spanned by the state vectors $|0\rangle = \frac{1}{\sqrt{2}} |0\rangle$ and $|1\rangle = \frac{1}{\sqrt{2}} |1\rangle$. The numerical values in the truth table correspond to the quantum states $|0\rangle$ and $|1\rangle$. Adapted from Ref. 171.

<table>
<thead>
<tr>
<th>GATE</th>
<th>CIRCUIT REPRESENTATION</th>
<th>MATRIX REPRESENTATION</th>
<th>TRUTH TABLE</th>
<th>BLOCH SPHERE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>$I$</td>
<td>$I = \begin{pmatrix} 1 &amp; 0 \ 0 &amp; 1 \end{pmatrix}$</td>
<td>Input $</td>
<td>0\rangle$ Output $</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$</td>
<td>1\rangle$ Input $</td>
</tr>
<tr>
<td>X</td>
<td>$X$</td>
<td>$X = \begin{pmatrix} 0 &amp; 1 \ 1 &amp; 0 \end{pmatrix}$</td>
<td>Input $</td>
<td>0\rangle$ Output $</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$</td>
<td>1\rangle$ Input $</td>
</tr>
<tr>
<td>Y</td>
<td>$Y$</td>
<td>$Y = \begin{pmatrix} 0 &amp; -1 \ 1 &amp; 0 \end{pmatrix}$</td>
<td>Input $</td>
<td>0\rangle$ Output $i</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$</td>
<td>1\rangle$ Input $-i</td>
</tr>
<tr>
<td>Z</td>
<td>$Z$</td>
<td>$Z = \begin{pmatrix} 1 &amp; 0 \ 0 &amp; -1 \end{pmatrix}$</td>
<td>Input $</td>
<td>0\rangle$ Output $</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$</td>
<td>1\rangle$ Input $-</td>
</tr>
<tr>
<td>S</td>
<td>$S$</td>
<td>$S = \begin{pmatrix} 0 &amp; e^{i\frac{\pi}{4}} \ e^{-i\frac{\pi}{4}} &amp; 0 \end{pmatrix}$</td>
<td>Input $</td>
<td>0\rangle$ Output $</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$</td>
<td>1\rangle$ Input $e^{i\frac{\pi}{4}}</td>
</tr>
<tr>
<td>T</td>
<td>$T$</td>
<td>$T = \begin{pmatrix} 0 &amp; e^{i\frac{\pi}{4}} \ e^{-i\frac{\pi}{4}} &amp; 0 \end{pmatrix}$</td>
<td>Input $</td>
<td>0\rangle$ Output $</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$</td>
<td>1\rangle$ Input $e^{-i\frac{\pi}{4}}</td>
</tr>
<tr>
<td>H</td>
<td>$H$</td>
<td>$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 &amp; 1 \ 1 &amp; -1 \end{pmatrix}$</td>
<td>Input $</td>
<td>0\rangle$ Output $\frac{</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$</td>
<td>1\rangle$ Input $\frac{</td>
</tr>
</tbody>
</table>
While the classic bit-flip gate inverts any input state, the quantum bit-flip does not in general produce the antipodal state (when viewed on the Bloch sphere), but rather exchange the prefactors of the wavefunction written in the computational basis. The $X$ operator is sometimes referred to as "the quantum NOT" (or "quantum bit-flip"), but we note that $X$ only acts similar to the classical NOT gate in the case of classical data stored in the quantum bit, i.e., $X|g\rangle = |\overline{g}\rangle$ for $g \in \{0, 1\}$.

As briefly mentioned in Sec. IV B, "all" quantum gates are reversible, due to the underlying unitary nature of the operators implementing the logical operations. Certain other processes used in quantum information processing, however, are irreversible, namely, measurements (see Sec. V for detailed discussion) and energy loss to the environment (if the resulting state of the environment is not known). Here, we will not consider how these processes are modeled, but refer the interested reader to, e.g., Ref. 172, and will only consider unitary control operations throughout the rest of this section. Finally, we note that quantum circuits are written left-to-right (in order of application), while the calculation of the result of a gate-sequences, e.g., the circuit

$$|\psi_{\text{in}}\rangle \rightarrow U_0 \rightarrow U_1 \rightarrow \cdots \rightarrow U_n \rightarrow |\psi_{\text{out}}\rangle$$

is performed right-to-left, i.e.,

$$|\psi_{\text{out}}\rangle = U_n \cdots U_1 U_0 |\psi_{\text{in}}\rangle.$$

As discussed in Sec. IV A, the NOR and NAND gates are each individually universal gates for classical computing. Since both of these gates have no direct quantum analog (because they are not reversible), it is natural to ask which gates "are" needed to build a universal quantum computer. It turns out that the ability to rotate about arbitrary axes on the Bloch-sphere (i.e., a complete single-qubit gate set), supplemented with any entangling 2-qubit operation will suffice for universality. By using what is known as the "Krauss-Cirac decomposition," any two-qubit gate can be decomposed into a series of $CNOT$ operations.

1. **Gate sets and gate synthesis**

A common universal quantum gate set is

$$\mathcal{G}_0 = \{X_0, Y_0, Z_0, \Phi_0, \text{CNOT}\},$$

where $\Phi_0 = e^{i \theta}|1\rangle$ applies an overall phase $\theta$ to a single qubit. For completeness we mention another universal gate set which is of particular interest from a theoretical perspective, namely,

\[\text{FIG. 10.} \text{ Quantum two-qubit gates: the controlled NOT (CNOT) gate and the controlled phase (CPHASE or CZ). For each gate, the name, a short description, circuit representation, matrix representation, and input/output truth tables are presented. Matrices are defined in the basis spanned by the two-qubit state vectors $|00\rangle$, $|01\rangle$, $|10\rangle$, and $|11\rangle$, where the first qubit is the control qubit, and the second qubit is the target qubit. The CNOT gate flips the state of the target qubit conditioned on the control qubit being in state $|1\rangle$. The CPHASE gate applies a $Z$ gate to the target qubit conditioned on the control qubit being in state $|1\rangle$. Note that the CPHASE gate is symmetric with respect to control and target qubit. Adapted from Ref. 171.}\]

\[\text{FIG. 11.} \text{ Comparison of the classical inverter (NOT) gate and quantum bit flip (X) gate. (a) The classical NOT gate that inverts the state of a classical bit. (b) The quantum X gate, which flips the amplitudes of the two components of a quantum bit.}\]
A technical aside, we mention that the restriction to a discrete gate set still gives rise to universality. This fact relies on using the so-called Solovay-Kitaev\textsuperscript{37,59,106} theorem, which (roughly) states that any other single-qubit gate can be approximated to an error $\epsilon$ using only $O(\log^6(1/\epsilon))$ (where $\epsilon > 0$) single-qubit gates from $G_1$. The gate-set $G_1$ is typically referred to as the "Clifford + $T$" set, where H, S and CNOT are all Clifford gates.

Each quantum computing architecture will have certain gates that are simpler to implement at the hardware level than others (sometimes referred to as "native" gates of the architecture). These are typically the gates for which the Hamiltonian governing the gate-implementation gives rise to a unitary propagator which corresponds to the gate itself. We will show several examples of this in Secs. IV E, IV F, and IV G. Regardless of which gates are natively available, as long as one has a complete gate set, one can use the Solovay-Kitaev theorem to synthesize any other set efficiently. In general one wants to keep the overall number of time steps in which gates are applied (denoted the "depth" of a circuit) as low as possible, and one wants to use as many of the native gates as possible, to reduce as much as possible of the time spent for the synthesis. Moreover, running a quantum algorithm also depends on the qubit connectivity of the device. The process of designing a quantum gate sequence that efficiently implements a specific algorithm, while taking into account the considerations outlined above is known as "gate synthesis" and "gate compilation," respectively. A full discussion of this large research effort is outside the scope of this review, but the interested reader may consult, e.g., Refs. 177–179 and references therein as a starting point.

As a (concrete (and trivial) example of how gate identities can be used, in Eq. (73) we illustrate how the Hadamard gate from $G_1$ can be generated by two single-qubit gates (from $G_0$) and an overall phase gate

$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & -i \\ 0 & i \end{bmatrix} - \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \quad (73)$$

As we show in Sec. IV D 1, the gates $X_\theta$, $Y_\theta$ and $Z_\theta$ are all natively available in a superconducting quantum processor.

We now address the question of how single qubit rotations and two-qubit operations are implemented in transmon-based superconducting quantum processors.

### 2. Addressing superconducting qubits

The modes of addressing transmonlike superconducting qubits can roughly be split into two main categories: (i) Capacitive coupling between a resonator (or a feedline) and the superconducting qubit dipole-field allows for microwave control to implement single-qubit rotations (see Sec. IV D) as well as certain two-qubit gates (see Secs. IV G and IV G 4). (ii) For flux-tunable qubits, the local magnetic fields can be used to tune the frequency of individual qubits. This allows the implementation of $z$-axis single-qubit rotation as well as multiple two-qubit gates (see Secs. IV E, IV F, and IV H).

#### D. Single-qubit gates

In this section, we will review the steps necessary to demonstrate that capacitive coupling of microwaves to a superconducting circuit can be used to drive single-qubit gates. To this end we consider a superconducting qubit to a microwave source (sometimes referred to as a "qubit drive") as shown in Fig. 12(a). A full circuit analysis of the circuit in Fig. 12(a) is beyond the scope of this review, so here we settle for highlighting the steps that elucidate the physics of the qubit/drive coupling. The interested reader may consult a number of lectures notes and pertinent theses (e.g., Refs. 44, 157, and 180–182). Here we follow Ref. 157.

### 1. Capacitive coupling for X, Y control

We start by modeling the qubit as a harmonic oscillator, for which the (classical) circuit Hamiltonian can be calculated by circuit quantization techniques, starting from Kirchhoff's laws, and is given by

$$H = \frac{q^2}{2C_0} - \frac{\Phi^2}{2L} + \frac{C_2}{C_0} V_q(t)Q, \quad (74)$$

where $C_0 = C + C_d$ is the total capacitance to the ground and $Q = C_0 \Phi - C_d V_q(t)$ is a renormalized charge variable for the circuit. We can now promote the flux and charge variables to quantum operators and assume weak coupling to the drive-line, so that $\dot{Q} \approx \tilde{Q}$, and arrive at

$$H = \frac{q^2}{2C_0} + \frac{C_2}{C_0} V_q(t)Q, \quad (75)$$

where $H_{LC} = \frac{q^2}{2C_0} + \frac{\Phi^2}{2L}$ and we have kept only terms that couple to the dynamic variables. Similar to the momentum operator for a harmonic oscillator in $(x, p)$–space, we can express the charge variable in terms of raising and lowering operators, as done in Sec. II

$$\dot{Q} = -i Q_p \tilde{p}(a - a^\dagger), \quad (76)$$

where $Q_p = \sqrt{\hbar/2Z}$ is the zero-point charge fluctuations and $Z = \sqrt{L/C}$ is the impedance of the circuit to ground. Thus, the LC oscillator capacitively coupled to a drive line can be written as

$$H = \omega \left( a^\dagger a + \frac{1}{2} \right) - \frac{C_d}{C_0} V_q(t) i Q_p (a - a^\dagger). \quad (77)$$

Finally, by truncating to the lowest transition of the oscillator, we can make the replacement $a \rightarrow \sigma^- \text{ and } a^\dagger \rightarrow \sigma^+$ throughout and arrive at\textsuperscript{149}

$$H = -\frac{\omega_0}{2} \sigma^z + Q V_q(t) \sigma_z, \quad \frac{i\hbar}{\hbar/\Omega} \quad (78)$$

where $\Omega = (C_d/C_0) Q_p \omega_0$ and $\omega_0 = (E_1 - E_0)/\hbar$.
To elucidate the role of the drive, we move into a frame rotating with the qubit at frequency \( \omega_q \) (also denoted "the rotating frame" or the "interaction frame"). To see the usefulness of this rotating frame, consider a state \(| \psi_0 \rangle = (1, 1)^T / \sqrt{2} \). By the time-dependent Schrödinger equation this state evolves according to
\[
|\psi(t)\rangle = U_{lt} |\psi_0\rangle = \frac{1}{\sqrt{2}} \left( e^{i\omega_q t/2} + e^{-i\omega_q t/2} \right),
\]
where \( U_{lt} \) is the propagator corresponding to \( H_t \). By calculating, e.g., \( \langle \psi_0 | \sigma_x | \psi_0 \rangle = \cos (\omega_q t) \), it is evident that the phase is winding with a frequency of \( \omega_q \) due to the \( \sigma_x \) term. By going into a frame rotating with the qubit at frequency \( \omega_q \), the action of the drive can be more clearly appreciated. To this end, we define \( U_i(t) = e^{iH_i t} = U_{lt} \) and the new state in the rotating frame is \(| \psi_i(t) \rangle = U_{it} |\psi_0\rangle \). The time-evolution in this new frame is again found from the Schrödinger equation (using the shorthand \( \partial_t = \partial / \partial t \))
\[
\begin{align*}
\dot{i}|\psi_i(t)\rangle &= i(\partial_t U_{it}) |\psi_i\rangle + iU_{it} (\partial_/ \partial t |\psi_i\rangle) \\
&= iU_{it} U_{lt}^\dagger |\psi_0\rangle + iU_{lt} H_0 |\psi_0\rangle, \\
&= \left( iU_{it} U_{lt}^\dagger + U_{lt} H_0 U_{lt}^\dagger \right) |\psi_i\rangle. 
\end{align*}
\]

We can think of the term \( \tilde{H}_0 \) in the parentheses in Eq. (82) as the form of \( H_0 \) in the rotating frame. Simple insertion shows that \( \tilde{H}_0 = 0 \) as expected (the rotating frame should take care of the time-dependence). However, one could also think of the term in brackets in Eq. (82) as a prescription for calculating the form of any Hamiltonian in the rotating frame given by \( U_{lt} \), by replacing \( H_0 \) with some other \( H \). In general, we will not find \( \tilde{H}_0 = 0 \).

Returning to Eq. (78), the form of \( H_0 \) in the rotating frame is found to be
\[
\tilde{H}_0 = \Omega V_0(t) \cos (\omega_q t) \sigma_y - \sin (\omega_q t) \sigma_x. 
\]

We can in general assume that the time-dependent part of the voltage \( V(t) \) has the generic form
\[
\begin{align*}
V(t) &= s(t) \sin (\omega_q t + \phi), \\
&= s(t) \cos (\phi) \sin (\omega_q t) + \sin (\phi) \cos (\omega_q t), 
\end{align*}
\]
where \( s(t) \) is a dimensionless envelope function, so that the amplitude of the drive is set by \( V_0 \). Adopting the definitions
\[
\begin{align*}
I &= \cos (\phi) \text{(the 'in - phase' component)}, \\
Q &= \sin (\phi) \text{(the 'out - of - phase' component)}, \\
\end{align*}
\]
the driving Hamiltonian in the rotating frame takes the form
\[
\tilde{H}_d = \Omega V_0 s(t) (I \sin (\omega_q t) - Q \cos (\omega_q t)) \\
\times \left( \cos (\omega_q t) \sigma_y - \sin (\omega_q t) \sigma_x \right) .
\]

Performing the multiplication and dropping fast rotating terms that will average to zero (i.e., terms with \( \omega_q t \sim \omega_d t \), known as the rotating wave approximation (RWA), we are left with
\[
\tilde{H}_d = \frac{1}{2} \Omega V_0 s(t) \left[ (-I \cos (\delta t) + Q \sin (\delta t)) \sigma_x \\
+ (I \sin (\delta t) - Q \cos (\delta t)) \sigma_y \right],
\]
where \( \delta \omega = \omega_q - \omega_d \). Finally, by reusing the definitions from Eq. (85), the driving Hamiltonian in the rotating frame using the RWA can be written as
\[
\tilde{H}_d = \frac{-\Omega}{2} V_0 s(t) \left( 0 \ e^{-i(\delta t + \phi)} \ e^{i(\delta t + \phi)} \right).
\]
Equation (90) is a powerful tool for understanding single-qubit gates in superconducting qubits. As a concrete example, assume that we apply a pulse at the qubit frequency, so that \( \delta \omega = 0 \), then
\[
\tilde{H}_d = -\frac{\Omega}{2} V_0 s(t)(I \sigma_x + Q \sigma_y),
\]
showing that an "in-phase" pulse (\( \phi = 0 \), i.e., the \( I \)-component) corresponds to rotations around the \( x \)-axis, while an out-of-phase pulse (\( \phi = \pi/2 \), i.e., the \( Q \)-component), corresponds to rotations about the \( y \)-axis. As a concrete example of an in-phase pulse, writing out the unitary operator yields
\[
U_{i,0}^{\phi=0}(t) = \exp \left[ \frac{i}{2} \Omega V_0 \int_0^t s(t') dt' \right] \sigma_x ,
\]
which depends only on the macroscopic design parameters of the circuit as well as the envelope of the baseband pulse \( s(t) \) and amplitude \( V_0 \), which can both be controlled using arbitrary waveform generators (AWGs). Equation (92) is known as "Rabi driving" and can serve as a useful tool for engineering the circuit parameters needed for efficient gate operation (subject to the available output voltage \( V_0 \)). To see this, we define the shorthand
\[
\Theta(t) = -\Omega V_0 \int_0^t s(t') dt',
\]
which is the angle by which a state is rotated given the capacitive couplings, the impedance of the circuit, the magnitude \( V_0 \), and the waveform envelope, \( s(t) \). This means that to implement a \( \pi \)-pulse on the \( x \)-axis, one would solve the equation \( \Theta(t) = \pi \) and output the signal in phase with the qubit drive. In this language, a sequence of pulses [see Fig. 13(a)] \( \Theta_2, \Theta_3, \ldots, \Theta_k \) is converted to a sequence of gates operating on a qubit as
\[
U_k \cdots U_1 U_0 = T \prod_{n=0}^k \exp \left[ -i \Theta_n (I \sigma_x + Q \sigma_y) \right],
\]
where \( T \) is an operator that ensures the pulses are generated in the time-ordered sequence corresponding to \( U_k \cdots U_0 \).

In Fig. 13, we outline the typical in-phase and quadrature (IQ) modulation setup used to generate the pulses used in Eq. (94). Figure 13(a) shows how a pulse at frequency \( \omega_q \) is generated using a low phase-noise microwave generator [typically denoted "the local oscillator (LO)"] , while the pulse is shaped by combining the LO signal in an IQ mixer with pulses generated in an AWG. To allow for frequency multiplexing, the AWG signal will typically be generated with a low-frequency component, \( \omega_{AWG} \), and the LO signal will be offset, so that \( \omega_{LO} + \omega_{AWG} = \omega_q \). By mixing in more than one frequency \( \omega_{AWG1}, \omega_{AWG2}, \ldots \), it is possible to address multiple qubits (or readout resonators) simultaneously, via the superposition of individual drives.

The \( I \) (Q) input of the IQ mixer will multiply the baseband signal to the in-phase (out-of-phase) component of the LO. In Fig. 13(b), we schematically show the comparison between XY gates in a quantum
2. Virtual Z Gate

As we saw in Sec. IV D, the distinction between x- and y-rotations was merely a choice of phase on the microwave signals, and the angle to be rotated is given by Θ(τ), both of which are generated using an AWG. Since the choice of phase φ has an arbitrary starting point, we could consider φ → φ + π/2. This would lead to I → Q and Q → −I. Therefore, changing the phase effectively changes rotations around x to rotations around y (and vice-versa, with a change of sign). This is reminiscent of the result of applying a Z rotation to x- and y-rotations, where ZxXy = iYx and ZxYy = −iXy. This analogy between shifting a phase of an AWG-generated signal and applying Z rotations can be utilized to implement “virtual” Z gates.\(^{186}\) As shown by McKay et al., this intuition can be formalized via the following example: consider the case of applying a pulse with an angle θ on the I channel (i.e., a Xθ pulse) followed by another pulse on the I channel, but with a phase φ0 relative to the first pulse (denoted X(φ0)θ, where X indicates we still use the I channel, but the rotation axis is now an angle φ0 away from the x-axis). Using Eq. (94) corresponds to a pulse sequence

\[
X(φ_0)\theta_X_0 = e^{−i\cos(φ_0)\varphi_0^{Q_1} + i\sin(φ_0)\varphi_0^{P_1}}_X_0
\]

\[
= Z_{φ_0}X_{φ_0}Z_{φ_0}X_θ,
\]

from which we see that the effect of the offset phase φ0 is to apply Zφ0. The equality above can be verified with a little trigonometric footwork.

The final Zφ0 is due to the rotation being in the frame of reference of the qubit. However, since the readout is along the z-axis (see Sec. V), a final phase rotation about z will not change the measurement outcome. Thus, if one wants to implement the gate sequence

\[
⋯\rightarrow U_i Z_{θ_0} U_{i+1} Z_{θ_1} U_{i+2}⋯
\]

where Ui’s are arbitrary gates, this can be done by revising the gate sequence (in the control software for the AWG) and changing the phase of subsequent pulses

\[
⋯\rightarrow U_i U(iθ_1)_{i+1} U(iθ_2)_{i+2}⋯
\]

which reduces the number of overall gates. Moreover, the virtual-Z gates are “perfect,” in the sense that no additional pulses are required, and the gate takes “zero time,” and thus the gate fidelity is nominally unity. As we show in Secs. IV E and IV F, operation of two-qubit gates can incur additional single-qubit phases. Using the virtual-Z strategy, these phases can be canceled out, leaving a pure two-qubit interaction.

Finally, we mention one more salient feature of the virtual-Z gates. As shown in Ref. 63, any single-qubit operation (up to a global phase) can be written as

\[
U(θ, φ, λ) = Z_{φ_0} X_{λ} Z_{φ_0} X_{λ+π} Z_{φ_0} X_{λ+2π},
\]

for appropriate choice of angles θ, φ, λ. This means that access to a single physical Xg combined with the virtual-Z gives access to a complete single qubit gate set! An explicit example of Eq. (99) in action is the Hadamard gate, which can be written as H = Z* Xg Zg, but since the Z’s can be virtual, it is possible to implement Hadamards as an effective single pulse operation in superconducting qubits.

3. The DRAG scheme

In going from Eq. (77) to Eq. (78), we assumed we could ignore the higher levels of the qubit. However, for weakly anharmonic qubits, such as the transmon (see Sec. II), this may not be a justified assumption, since ωq \(ω_{q_1}^{-1}\) only differs from \(ω_{q_1}^{-1}\) by the anharmonicity, \(z = ω_{q_1}^{-1} − ω_0\), which is negative and typically around 200 to 300 MHz. This situation is sketched in Figs. 14(a)–14(c), where we illustrate how Gaussian pulses with standard deviations \(σ = \{1, 2, 3\}\) ns have spectral content that leads to nonzero overlaps with the \(ω_{q_1}^{-1}\) frequency. This leads to two deleterious effects: (1) leakage errors which take the qubit out of the computational subspace, and (2) phase errors. Effect 1 can occur because a qubit in the state |1\> may be excited to |2\> as a π pulse is applied, or be excited directly from the |0\> since the qubit spends some amount of time in the |1\> state during the π pulse. Effect 2 occurs because the presence of the drive results in a repulsion between the |1\> and |2\> levels, in turn changing \(ω_{q_1}^{-1}\) as the pulse is applied. This leads to the accumulation of a relative phase between |0\> and |1\> \(188\) The so-called DRAG procedure \(189, 190\) (Derivative Reduction by Adiabatic Gate) seeks to combat these two effects by applying an extra signal in the out-of-phase component. The trick is to modify the waveform envelope s(t) according to
In practice, there can be a deviation from these two optimal values, often due to pulse distortions in the lines leading to the qubits. Typically, randomized benchmarking experiments combined with single-shot measurements (see Sec. V) of the |2⟩ state are used to determine the optimal value of \( \lambda \). The \( \lambda = \{0.5, 1\} \) trade-off was demonstrated explicitly in Refs. \( 186 \) and \( 193 \). However, by extending the original DRAG pulse implementation \( 194,196 \) it is possible to reduce “both” errors “simultaneously.” By introducing a frequency detuning parameter \( \delta f \) to the waveform \( 190 \) (defined such that \( \delta f = 0 \) corresponds to the qubit frequency), i.e.,

\[
s'_{\delta f}(t) = s(t)e^{i2\pi \delta f t},
\]

and choosing \( \lambda \) to minimize leakage errors, then phase errors can be reduced simultaneously \( 195 \). Similarly, by a judicious use of the virtual-Z gate, it is also possible to reduce phase errors in combination with DRAG pulsing to reduce leakage \( 196 \). Modern single-qubit gates using DRAG pulsing now routinely reach fidelities \( F_{\text{U}} \approx 0.99 \) \( 95,97,193,196-199 \) Other techniques also exist for operating single-qubit gates in a spectrally crowded device \( 200,201 \).

### E. The iSWAP two-qubit gate in tunable qubits

As briefly mentioned in Sec. IV C, single-qubit gates supplemented with an entangling two-qubit gate can form the gate set required for universal quantum computation. The two-qubit gates available in the transmonlike superconducting qubit architecture can roughly be split into two broad families as outlined previously: one group requiring local magnetic fields to tune the transition frequency of qubits and one group consisting of all-microwave control. There exist several hybrid schemes that combine various aspects of these two categories and, in particular, the notions of tunable coupling and parametric driving are proving to be important ingredients in modern superconducting qubit processors \( 83,87,89,103,105,106,202-207 \). In this section, however, we start by introducing the iSWAP gate, and then review the CPHASE (controlled-phase) in Sec. IV F and the CR (cross-resonance) in Sec. IV G. We briefly review a few other two-qubit gates and discuss their merits in Secs. IV G 4 and IV H.

#### 1. Deriving the iSWAP unitary

As we saw in Sec. II, Eq. 31 the interaction term between two capacitively coupled qubits (in the two-level approximation) is given by

\[
H_{\text{int}} = g \sigma_1 \otimes \sigma_2,
\]

where \( g \) is the coupling strength and \( \otimes \) is used to emphasize the tensor product. If the capacitive coupling is mediated through a bus resonator, then \( 208,209 \)

\[
g \rightarrow g_{0q} \rightarrow q = \frac{g_1g_2(\Delta_1 + \Delta_2)}{2\Delta_1\Delta_2}.
\]

where \( g_i \) is the resonator coupling to qubit \( i \) (dependent on the qubit-resonator coupling capacitance \( C_{qi} \)) and \( \Delta_i = \omega_{qi} - \omega_q \) is the detuning of qubit \( i \) to the resonator. In the simpler case where the qubits are directly coupled \( 210 \).
The unitary corresponding to a gate that acts to swap an excitation between the two qubits is given by

$$U_{\text{SWAP}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{i\phi} & 0 & 0 \\ 0 & 0 & e^{-i\phi} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

(104)

where $C_{q-q}$ is the qubit-qubit coupling capacitance and $C_i$ is the capacitance of qubit $i$. Throughout this section, we will assume a direct capacitance coupling between qubits of the flux-tunable transmon type, so that $g = g_{q-q}$ and $\omega_{q-q} = \omega_{q-q}(\Phi_i)$. For simplicity, we suppress the explicit flux dependence of the $\omega_{q-q}$'s and simply refer to the coupling as $g$. Equation (102) can be rewritten as

$$H_{qq} = -g\left(\sigma^+ - \sigma^-\right) \otimes \left(\sigma^+ - \sigma^-\right),$$

(105)

and then using the rotating wave approximation again (i.e., dropping fast rotating terms) we arrive at

$$H_{qq} = g\left(e^{i\phi_1 t}\sigma^+ \sigma^- + e^{-i\phi_2 t}\sigma^- \sigma^+ \right),$$

(106)

where we have introduced the notation $\phi_{12} = \phi_1 - \phi_2$ and suppressed the explicit tensor product between qubit subspaces. If we now change the flux of qubit 1 to bring it into resonance with qubit 2 ($\omega_{q1} = \omega_{q2}$), then

$$H_{qq} = g\left(\sigma^+ \sigma^- + \sigma^- \sigma^+ \right) = \frac{g}{2} (\sigma_x \sigma_x + \sigma_y \sigma_y).$$

(107)

The first part of Eq. (107) shows that a capacitive interaction leads to a swapping of excitations between the two qubits, giving rise to the "swap" gate in iSWAP. Moreover, due to the last part of Eq. (107), this capacitive coupling is also sometimes said to give rise to an "XY" interaction. The unitary corresponding to a XY (swap) interaction is

$$U_{\text{XY}}(t) = e^{-i\theta(t) \sigma_x \sigma_y} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\theta t) & -i \sin(\theta t) & 0 \\ 0 & -i \sin(\theta t) & \cos(\theta t) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$ 

(108)

Since the qubits are tunable in frequency, we can now consider the effect of tuning the qubits into resonance for a time $t' = \frac{\pi}{2g}$

$$U_{\text{XY}}\left(\frac{\pi}{2g}\right) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & i \frac{\pi}{2g} \\ 0 & 0 & 0 & -i \frac{\pi}{2g} \\ 0 & 0 & 0 & 1 \end{pmatrix} \equiv i\text{SWAP}. $$

(109)

From this result, we see that a capacitive coupling between qubits turned on for a time $t'$ (inversely related to the coupling strength in units of radial frequency) leads to implementing a so called "iSWAP" gate, which acts to swap an excitation between the two qubits, and add a phase of $\phi = \exp(\pi/2g)$. For completeness, we note that for $t' = \frac{\pi}{4g}$, the resulting unitary

$$U_{\text{XY}}\left(\frac{\pi}{4g}\right) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/\sqrt{2} & -i/\sqrt{2} & 0 \\ 0 & -i/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \equiv \sqrt{\text{SWAP}}$$

(110)

is typically referred to as the "squareroot-iSWAP" gate. The $\sqrt{\text{SWAP}}$ gate can be used to generate Bell-like superposition states, e.g., $|01\rangle + i|10\rangle$.

To elucidate the operating principle behind an iSWAP implementation, we show the spectrum of a flux-tunable qubit using typical transmonlike parameters in Fig. 15(a). The iSWAP is performed at the avoided crossing, where $\Phi = \Phi_{\text{SWAP}}$. By preparing QB1 in-state $|1\rangle$, moving into the avoided crossing, waiting there for a time $\tau$ [see pulse-sequence in inset in Fig. 15(b)], the excitation is swapped back and forth between the two qubits, as shown in Fig. 15(b). In Fig. 15(c), we plot the linecuts of (b) at $\Phi_{\text{SWAP}}$, showing the excitation oscillating back and forth between $|01\rangle$ and $|10\rangle$ with the predicted time $\tau' = \pi/2g$. In turn, the frequency of the oscillation can be used to extract the strength of the coupling, $g = \frac{\pi}{2\tau' \omega_{\text{res}}}$.

So far, we have ignored the role of the single-qubit phases acquired by tuning the qubit frequency. Referring to the pulse-sequence shown in the top panel of Fig. 15(a), we see that each qubit will acquire a phase given by

$$\theta_q = \int \Phi_q(t) dt,$$

(111)

This phase can be conveniently removed either by subsequent application of virtual-Z gates to all following pulses, or by shaping the

---

**FIG. 15.** (a) Spectrum of two transmon qubits (written in the combined basis as qubit 1 |QB1, QB2\rangle) as the local flux through the loop of qubit 1 is increased. The black/dashed lines with arrows indicate a typical flux trajectory to demonstrate operation of iSWAP gate. (b) Probability of swapping into the |01\rangle state as a function of time and flux. The pulse sequence corresponds to preparing |10\rangle and performing a typical iSWAP operation (for a time $\tau$). (c) Probabilities of |01\rangle (black) and |10\rangle (gray) at $\Phi = \Phi_{\text{SWAP}}$ (while dashed line in (b) as the time spent at the operating point (c) is increased. This simulation does not include any decay effects.
waveform of the excursion such that single-qubit phases are exactly canceled.\(^{216}\)

Equations (104) and (108) together present a useful result from a quantum processor design perspective: The operating regime, frequency and time \(\tau\) of the \(iSWAP\) can be calculated (typically simulated) to a high precision, before any processor fabrication is undertaken. The only “quantum” parts that enter \(\omega_{q}\) and \(\omega_{q-1\&q}\) are the qubit frequencies, \(\omega_{q1}(\Phi_{1})\) and \(\omega_{q2}(\Phi_{2})\). If the Josephson energies of the qubits are known (which they typically are, from fabrication parameters), then by simulating the capacitances in \(\omega_{q}\) or \(\omega_{q-1\&q}\), the time \(\tau\) and the pulse shape needed to implement an \(iSWAP\) can be estimated to high precision. Typical values of the coupling strengths, \(g/(2\pi)\), for architectures using the \(iSWAP\) gate are 5–40 MHz and are often very close to expectations from EM simulations.\(^{215,219–217}\)

2. Applications of the \(iSWAP\) gate

The \(iSWAP\) cannot generate a CNOT gate by itself. Rather, to implement a CNOT gate requires stringing together two \(iSWAP\) s and several single qubit gates\(^{211}\).

\[
\begin{array}{c}
\text{Z}_{\frac{\pi}{2}} \\
\text{X}_{\frac{\pi}{2}} \quad \text{SWAP} \\
\end{array}
\]

(112)

As evident from Eq. (112), the \(iSWAP\) gate in general needs to be used twice to generate a single CNOT, leading to a significant overhead when compiling CNOT-dense circuits from \(iSWAP\) gates. However, depending on the context, the \(iSWAP\) can be used efficiently (i.e., without any two-qubit gate overhead) to mimic the behavior of a CNOT. Typically such circuits will not be completely equivalent, but will share certain salient features for specified input states. As an example of this procedure, Neeley et al.\(^{214}\) demonstrated the generation of a 3-qubit Greenberger-Horne-Zeilinger (GHZ) state (which requires two subsequent CNOTs in the simplest construction), by using only two \(iSWAP\) s in a circuit that correctly generates the 3-qubit GHZ state on the \(|000\rangle\) input. Moreover, the \(XY\)-interaction is a powerful tool for certain types of quantum simulation algorithms.\(^{218}\) If one is interested in digital quantum simulation of spinlike systems, then the \(XY\)-interaction can natively simulate, e.g., a Heisenberg interaction

\[
H_{\text{Heisenberg}} = J_x \sigma_x \sigma_x + J_y \sigma_y \sigma_y + J_z \sigma_z \sigma_z.
\]

(113)

This approach to the \(XY\) interaction was demonstrated by Salathé et al.\(^{218}\), where repeated application of the \(iSWAP\) gate interspersed with single-qubit rotations was used to generate successive \(XY\), \(XZ\) and \(YZ\) interactions that lead to an aggregate \(H_{\text{Heisenberg}}\) Hamiltonian. State-of-the-art operation of the \(iSWAP\) gate has also been used to demonstrate a ten-qubit GHZ state.\(^{219}\)

F. The CPHASE two-qubit gate in tunable qubits

In our discussion of the \(iSWAP\) gates, we assumed that the higher energy levels of the superconducting qubit do not play a role. As we show below, it turns out that for the case of transmon qubits (with negative anharmonicity), the higher levels can in fact be utilized to generate a the CPHASE gate directly.\(^{212,220}\)

Recall from Sec. IV C that the CPHASE gate implements the following unitary:

\[
U_{\text{CPHASE}} = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & -1
\end{bmatrix}.
\]

(114)

Our goal for the remainder of this section is to show that the unitary operator of the CPHASE gate appears naturally for capacitively coupled transmon superconducting qubits and review a few of the modern applications of this gate. We have chosen to include a considerable amount of details for the implementation of this gate, as a means to review some of the issues one has to resolve, to engineer high quality two-qubit gates.

The structure of the matrix in Eq. (63) indicates that we need to apply a phase \((-1 = e^{i\phi})\) to the qubits whenever both are in the excited state \(|11\rangle\). Considering the nature of the \(XY\) interaction, which couples \(|01\rangle \leftrightarrow |10\rangle\) and leads to the \(SWAP\) gate (see Sec. IV E), we expect avoided level crossings to exist between higher levels, e.g., \(|11\rangle \leftrightarrow |20\rangle\) and \(|11\rangle \leftrightarrow |02\rangle\). The flux-tunable implementation of the CPHASE gate relies on this higher-level avoided crossing.

To motivate this intuition, we plot the spectrum for two coupled transmon qubits, in Fig. 16(a), including levels with two excitations, as the local magnetic flux in qubit 1 is being tuned. The Hamiltonian for this spectrum, written in the \(|00\rangle, |01\rangle, |10\rangle, |11\rangle, |02\rangle, |20\rangle\)-basis, is approximately given by

\[
H_{\text{excitations}} = \begin{bmatrix}
E_{00} & 0 & 0 & 0 & 0 & 0 \\
0 & E_{01} & g & 0 & 0 & 0 \\
0 & g & E_{10} & 0 & 0 & 0 \\
0 & 0 & 0 & E_{11} & \sqrt{2}g & \sqrt{2}g \\
0 & 0 & 0 & \sqrt{2}g & E_{02} & 0 \\
0 & 0 & 0 & \sqrt{2}g & 0 & E_{20}
\end{bmatrix}
\]

(115)

where \(E_{\text{answ}} = E_{\Phi_1}^0(\Phi_1) + E_{\Phi_2}^0(\Phi_2)\) and \(E_{\Phi_i}(\Phi_i)\) is the flux-dependent energy of the \(i\)-th level of a transmon,\(^{52}\) and the \(|02\rangle, |20\rangle \leftrightarrow |11\rangle\) transitions are scaled by a factor \(\sqrt{2}\) due to the higher photon number. In Fig. 16, we plot the frequencies \(E_{\text{answ}} = E_{\Phi_1} - E_{\Phi_2}\) calculated from Eq. (115), using standard, symmetric, transmonlike parameters, as the local magnetic field of qubit 1 is increased.

The result of the higher levels on the computational basis can be understood by considering a concrete example. By preparing the combined qubit state \(|11\rangle\) and moving slowly toward the avoided crossing between \(|11\rangle\) and \(|20\rangle\) at \(\Phi_{\text{CPHASE}}\), waiting for some time \(\tau\) and moving back [see black line with arrows in Fig. 16(b)], the resulting unitary operator in the computational basis is given by

\[
U_{\text{ad}} = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & e^{i\phi_1(\tau)} & 0 & 0 \\
0 & 0 & e^{i\phi_2(\tau)} & 0 \\
0 & 0 & 0 & e^{i\phi_1(\tau)}
\end{bmatrix}
\]

(116)
In fact, such excursions near space) of the repulsion of j during time s is the phase acquired by the state j along the trajectory ℓ in (Φ, t)-space during time τ. The movement should be sufficiently slow on a time scale set by g that the moving state never populates the |20⟩ state, i.e., the movement should be adiabatic. In terms of applied flux, the avoided crossing between the |11⟩ ↔ |20⟩ state happens before |10⟩ ↔ |01⟩ (due to the negative anharmonicity of the transmons, α ≈ −Ec) and consequently ℓ does not take the states through the ΦSWAP operating point. As shown in Fig. 16(b), we can define a parameter (typically denoted ζ) quantifying the difference in phase acquired by the |11⟩ relative to the single excitation states

$$\zeta = (\theta_{11} - (\theta_{01} + \theta_{10})).$$

(118)

The parameter ζ can be thought of as the result (in the computational space) of the repulsion of |11⟩ due to the |20⟩ state. If we now choose a trajectory ℓ, designed so that \( \int_0^\tau \zeta(\ell(t))dt = \pi \), then

$$\int_0^\tau \zeta(\ell(t))dt = \pi = \theta_{11}(\ell_s) - (\theta_{01}(\ell_s) + \theta_{10}(\ell_s)).$$

(119)

Inserting this expression into Eq. (116), we see that

$$U_{ad} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{i\theta_{10}} & 0 & 0 \\ 0 & 0 & e^{i\theta_{01}} & 0 \\ 0 & 0 & 0 & e^{i(\pi + \theta_{01}(\ell_s) + \theta_{10}(\ell_s))} \end{bmatrix}.$$ 

(120)

After the adiabatic excursion, one can now apply single-qubit pulses (or use virtual-Z gates) to exactly cancel the single-qubit phases such that \( \theta_{01}(\ell_s) = \theta_{10}(\ell_s) = 0 \). This changes \( U_{ad} \) to

$$U_{ad} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} = UC_{PHASE},$$

(121)

From Eq. (121), it is evident that an adiabatic movement of |11⟩, followed by single-qubit gates (virtual or real) efficiently implements a CPHASE and, through Eq. (66), also efficiently implements a CNOT. The CPHASE gate is one of the workhorses of modern superconducting qubit processors with gate fidelities \( \geq 0.99 \).220–221

One is, of course, free to choose an arbitrary trajectory \( \ell_0 \) that implements the phase \( e^{-i\phi} \) on the |11⟩ state. Assuming that the single-qubit phases are properly canceled, one sees that the arbitrary phase version of the CPHASE gate (typically denoted \( CZ_\phi \)) can be written as

$$CZ_\phi = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{-i\phi} \end{bmatrix} = \exp\left[-i\frac{\phi}{4} (\sigma_x \otimes \sigma_x - \sigma_z \otimes \mathbb{1} - \mathbb{1} \otimes \sigma_z)\right].$$

(122)

Because of the form of Eq. (122), one can think of the avoided crossing with the higher levels outside the computational subspace as giving rise to an effective \( \sigma_x \otimes \sigma_x \) coupling within the computational subspace.220

An alternative to the adiabatic approach outlined above to realize CPHASE is to make a sudden excursion to the ΦPHASE operating point, after waiting a time \( t = \pi/\sqrt{2g} \), the state will have completed a single Larmor-type rotation from |11⟩ to |02⟩ and back again to |11⟩, but in the process, acquired an overall \( \pi \) phase, similar to the iSWAP gate, but in the |{11⟩, |20⟩} subspace.224 In fact, such excursions near or through avoided crossings leading to adiabatic and nonadiabatic transitions have been studied extensively in the context of interferometry, cooling, spectroscopy, and quantum control.177,222–231

The remainder of this subsection is devoted to an overview of some of the recent advances and demonstrations using the CPHASE gate since its first demonstration in 2009 where it was used to generate Bell-states and demonstrate two-qubit algorithms.226

1. Trajectory design for the CPHASE gate

The (adiabatic) implementation of \( UC_{PHASE} \) outlined above assumed that the trajectory \( \ell_0 \) was completely adiabatic and that the
This presents a tension for optimally using the interleaved randomized operation. A relevant question is then: what is the “optimal” trajectory $\ell_0^*$ that implements the necessary phase as fast as possible, with as little leakage as possible, for a given size of the avoided crossing between $|11\rangle$ and $|20\rangle$? Given a typical coupling rate $g/2\pi \approx 20$ MHz (as discussed in Sec. IV E), one expects a heuristic lower time limit to be $2\pi/g \approx 50$ ns (stronger coupling of course leads to shorter gate times, but will limit the on/off ratio of the gate). Traditional optimal control of adiabatic movement assumes the movement is “through” the avoided crossing (see, e.g., Ref. 233), but the trajectory $\ell_0$ moves close to and then back from the avoided crossing. This modification to the adiabatic movement protocol was addressed by Martinis and Geller, specifically in the context of errors for a CPHASE gate implementation. The authors show that nonadiabatic errors can be minimal for gate times only slightly longer than $2\pi/g$ using an optimal waveform (based on a Slepian waveform235) to parametrize the trajectory $\ell_0^*(\tau)$.

2. The CPHASE gate for quantum error correction

Using the approach of Martinis and Geller, Barends et al. were able to demonstrate a two-qubit gate fidelity $F_{\text{CPHASE}} = 0.9944$ (determined via a technique known as “interleaved randomized benchmarking”236–238). This implementation had a gate time $\tau = 43$ ns and was implemented with the $\ell_0^*$ waveform, in an “xmon” device—a transmon with a “+”-shaped capacitor. A two-qubit gate fidelity $F > 0.99$ represents a significant milestone, not just from a technical and engineering perspective, but also from a foundational standpoint: The surface code (a quantum error correcting code) has a lenient fault-tolerance threshold of $\sim 1\%$.240–242 This means, roughly speaking, that if the underlying operations on the qubits have fidelities $F > 0.99$, then by adding more qubits to the circuit (and correctly implementing the fault-tolerant quantum error correction protocol) the overall error-rate can be reduced, and one can in principle perform arbitrarily long quantum computations, without errors spreading uncontrollably and corrupting the calculation. Because of its relatively lenient threshold under circuit noise (compared to, e.g., Steane or Shor codes232,233,235) and its use of solely nearest-neighbor coupling, the surface code is one of the most promising quantum error correction codes for medium-to-large scale quantum computing in solid state systems.240-242 Therefore, surpassing the fault-tolerance threshold using CPHASE represents a significant milestone for the field.243 Moreover, practical blueprints for implementing scalable subcells of the surface code using the CPHASE as the fundamental two-qubit gate have also been proposed244 as well as in-situ calibration protocols for large-scale systems operating with CPHASE.245 For a full review of the pros and cons of various quantum error correcting codes we refer the interested reader to, e.g., an introductory review article Ref. 247, or any of the excellent textbooks and more detailed review articles in Refs. 172, 174, 244, and 247–250.

Returning to the CPHASE gate, numerical optimization of $\ell_0^*$ was demonstrated by Kelly et al.25 using the interleaved randomized benchmarking sequence fidelity as a cost function to push a native implementation of $\ell_0^*$ with a fidelity $F = 0.984$ up to $F = 0.993$, surpassing the surface code fault tolerance threshold. In the same work that demonstrated $F_{\text{CPHASE}} = 0.9944$, Barends et al.25 used the CPHASE gate to generate GHZ states, $|\text{GHZ}\rangle = (|0\rangle^N + |1\rangle^N)\sqrt{2}$, of up to $N = 5$ qubits, with a fidelity for the $N = 5$ state of $F = \text{Tr}(\rho_{\text{ideal}}\rho_{N=5}) = 0.817$. The protocol for generating the GHZ state with $N = 2$ and $N = 3$ from CPHASE was originally demonstrated by DiCarlo et al.234,235 The textbook route to generating the $N = 2$ GHZ state, $|\Phi^+\rangle$ (a Bell state) from the all-zero input is

$$|0\rangle \quad \begin{array}{c} \hline H \\ \end{array} \quad |0\rangle = |\Phi^+\rangle.$$ (123)

An equivalent circuit using CPHASE and native single-qubit gates in superconducting qubits is

$$|0\rangle \quad \begin{array}{c} \hline Y_{\pi/2} \\ \end{array} \quad |0\rangle \quad \begin{array}{c} \hline Y_{-\pi/2} \\ \end{array} \quad |0\rangle \quad \begin{array}{c} \hline Y_\pi/2 \\ \end{array}.$$ (124)

By repeating the operation inside the dashed box on additional qubits, an $N$-qubit GHZ state can be generated.246 Since the demonstration of the $N = 5$ GHZ state using the CPHASE gate, the gate has been deployed to demonstrate several important aspects of quantum information processing using superconducting qubits. A nine-qubit implementation of the five-qubit repetition code (five data qubits + four syndrome qubits247) was demonstrated, and the error suppression factor of a single logical quantum bit was shown to increase as the encoding was changed from three data qubits to five data qubits.248 Similarly, in a five-qubit processor the three-qubit repetition code with artificially injected errors was demonstrated,249 building on earlier results utilizing a combination of iSWAP and CPHASE gates to perform parallelized stabilizer readout.250

3. Quantum simulation and algorithm demonstrations using CPHASE

As an example of the utility of the CPHASE gate, we briefly discuss a particular demonstration of a digital quantum simulation. In this context, the CPHASE gate has been utilized to simulate a two-site Hubbard model with four fermionic modes, using four qubits.253 Using the Jordan-Wigner transformation,254,255 it is possible to map fermionic operators onto Pauli spin matrices.254 As shown in Ref. 253, a Hubbard model with two fermionic modes, whose Hamiltonian is given by

$$H_{\text{Hubbard, two mode}} = -t(b_1^\dagger b_2^\dagger + b_2^\dagger b_1^\dagger) + Ub_1^\dagger b_1 b_2^\dagger b_2$$ (125)

can be written in terms of Pauli operators as

$$H = \frac{t}{2}(\sigma_x \otimes \sigma_x + \sigma_y \otimes \sigma_y) + \frac{U}{4}(\sigma_z \otimes \sigma_z + 1 \otimes 1 \otimes \sigma_z),$$ (126)

where $U$ is the repulsion energy and $t$ is the hopping strength. Similar to the Heisenberg interaction discussed briefly in Sec. IV E, it is now a question of producing $\sigma_x \otimes \sigma_z$-type interactions, where the prefactor

---

**Applied Physics Reviews**

**REVIEW**

scitation.org/journal/are
\( \alpha \) can be tuned. Using the CZ\(_{\phi} \) version of CPHASE, a \( U_{ZZ}(\phi) = \exp \left(-i\frac{\phi}{2} \sigma_x \otimes \sigma_x \right) \) unitary can be generated via

\[
U_{ZZ}(\phi) = \begin{array}{cc}
X_x & CZ_x \\hline
CZ_x & X_x
\end{array}
\]

(128)

where \( A_x \in \{x, \ Y_x\} \) is used to allow for small and negative angles. Finally, for completeness, we mention an alternative approach to creating \( U_{ZZ} \) given by\(^{23,266}\)

\[
U'_{ZZ}(\phi) = \begin{array}{cc}
X_{\phi/2} & \\hline
\end{array}
\]

(129)

which has the benefit of relying on CPHASE (through the CNOTs), and the angle can be controlled using the single-qubit Z gates. We refer the interested reader to two reviews on quantum simulations, see, e.g., Refs. 257 and 258.

The CPHASE gate has also been used in a variety of other contexts, e.g., for calculating the dissociation of diatomic hydrogen (\( \text{H}_2 \)) using the variational quantum eigensolver method,\(^{239}\) for feed-forward based teleportation experiments,\(^{260,261}\) as well as initial steps toward demonstrating quantum supremacy and a \( 2 \times 2 \) implementation of the Harrow-Hassidim-Lloyd algorithm.\(^{233,234}\) In the field of hybrid semiconducting nanowire/superconducting qubits (known as the “gatemon” approach\(^{262-265}\)), where the qubit frequency is modified by electrostatically changing the density of carriers in a semiconducting region with proximity-induced superconductivity, the CPHASE gate was also demonstrated between two nanowire qubits.\(^{267}\)

One may worry that operating a qubit by moving its frequency can lead to overlap with frequencies already used by other qubits, in a system with multiple qubits. This issue is known as “frequency crowding.” While the use of asymmetric transmons [with two sweet spots in the range \([-\Phi_0, +\Phi_0]\)], recall that Fig. 2(c)] may help alleviate some frequency crowding issues, a more long-term strategy is needed. One way to circumvent the problem is to utilize on/off tunable coupling schemes, in which qubits can exchange energy only if a coupler activates the interaction.\(^{25,106}\) To address this issue in the context of the CPHASE gate, Chen et al.\(^{103}\) demonstrated a device (named “the gmon”) where the qubit interaction can be tuned with an on/off ratio on the order of 1000, and a CPHASE gate fidelity of \( F = 0.9907 \) was demonstrated.

This concludes the introduction to the physics and operation of the CPHASE gate in its native form. In the remainder of this section, we will introduce a few of the microwave-only gates that have been demonstrated in an effort to sidestep the need for local tunability (and the resulting increased sensitivity to noise) as required by the iSWAP and CPHASE gate.

G. Two-qubit gates using only microwaves

One common (potential) drawback for the iSWAP and CPHASE gates is that their operation requires flux-tunable qubits. Introducing a new control knob, such as flux control, in turn also introduces a new noise channel for the system. Furthermore, the need for flux-tunability increases the sensitivity of the devices to flux noise by tuning the qubits from their “sweet spots,” increases the dephasing rate. From this perspective, one could envision using all-microwave-based gates to remedy these issues. To this end, the cross-resonance (“CR”) gate was developed for operating fixed-frequency superconducting qubits,\(^{269-271}\) which typically feature longer lifetimes and reduced sensitivity to flux noise.

### 1. The operational principle of the CR gate

To elucidate the operation of the CR gate, we briefly revisit the driving Hamiltonian derived in Sec. IV D. There, we considered only a single qubit. However, if one extends this formalism to two qubits, see Fig. 17(a) denoting the frequency difference by \( \Delta_{12} = \omega_{a1} - \omega_{a2} \) and the coupling by \( g \ll \Delta_{12} \), and performing a Schrieffer-Wolff transformation to go to the dressed state picture, the driving Hamiltonians for qubit 1 and 2 become\(^{260,262}\)

\[
H_{d1} = \Omega \nu V_{d1}(t) \left( \sigma_x \otimes 1 + \nu \sigma_y \otimes \sigma_y \right) + \mu \sigma_z \otimes \sigma_z,
\]

(130)

\[
H_{d2} = \Omega \nu V_{d2}(t) \left( 1 \otimes \sigma_x + \nu \sigma_y \otimes \sigma_y \right) + \mu \sigma_z \otimes \sigma_z,
\]

(131)

where

\[
\mu = \pm \frac{g}{\Delta_{12}} \left( \frac{x_1}{x_1 + \Delta_{12}} \right),
\]

(132)

\[
\nu = \pm \frac{g}{\Delta_{12}} \left( \frac{x_2}{x_2 + \Delta_{12}} \right),
\]

(133)

Figure 17. (a) Schematic circuit diagram of two fixed frequency transmons coupled through a resonator yielding an overall coupling coefficient \( g \). Qubit 1 driven at the frequency of qubit 2 leads to the CR gate. (b) Schematic level diagram of the always-on coupling leading to dressed states (01) and (10) with \( \Delta_{12} = \omega_{a1} - \omega_{a2} \). (c) Simulations of the expectation values of \( \langle \sigma_x \rangle \) and \( \langle \sigma_y \rangle \) for qubit 2 as a drive at the frequency of qubit 2 is applied to qubit 1. The upper panel shows regular Rabi oscillations when qubit 1 is in the (0) state. The bottom panel shows a modified Rabi frequency when qubit 1 is in (1) state, in accordance with Eq. [134]. (d) Difference in angle in the \((x, y)\) plane as a function of length of the applied drive to qubit 1. At approximately 200 ns, a \( \pi \)-phase shift has been acquired.
and $\Omega V_{\text{at}}(t)$ is the driving for qubit $i$. From Eq. (130), it is evident that if we drive qubit 1 at the frequency of qubit 2, then to qubit 2, this will look like a combination of $\nu_1^{-1} \otimes \sigma_z$ and $\mu_1 \sigma_z \otimes \sigma_z$. This means that the Rabi oscillations of qubit 2 will have a frequency given by

$$\Omega_{\text{Rabi}} = \Omega V_{\text{at}}(\nu_1^{-1} + z_1 \mu_1),$$  

(134)

where $z_1 = (\sigma_z \frac{1}{2})$, and $z_1$ depends on the state of qubit 1. This effect is demonstrated in Fig. 17(c), where a simulated drive is applied to qubit 1 while the resulting Rabi oscillations in qubit 2 are recorded. We have used typical fixed-frequency transmon parameters from experiments, and we have included a spurious cross-talk term $\eta = 0.03$. In Fig. 17(d), we plot the difference in angle in the $(z, y)$ plane acquired by qubit 2 for different initializations of qubit 1, $\Delta \phi = \phi_{\text{PRE}} - \phi_{\text{PRE}}^0$. For this particular choice of parameters, the cross-resonance gate achieves a $-\pi$-phase shift in $\approx 200$ ns.

This strategy was first demonstrated using flux-tunable transmons in Ref. 274, where a Bell state with fidelity $F_{\text{Bell}} = (\Omega \{\rho | \Omega^\dagger\}) = 0.90$ was achieved. Using quantum process tomography, the gate fidelity was found to be $F_{\text{QFT}} = 0.81$. By moving to fixed-frequency qubits with increased lifetimes, the gate fidelity was increased to $F_{\text{QFT}} = 0.98$ (with subtraction of state initialization and measurement errors). For completeness, we note that due to the form of the last term in Eq. (130), the CR gate is also sometimes denoted the 2X gate. The unitary matrix representation of the CR$_0$ gate is

$$U_{\text{CR}_0} = e^{i \theta \sigma_z} = 
\begin{bmatrix}
\cos \frac{\theta}{2} & -i \sin \frac{\theta}{2} & 0 & 0 \\
-i \sin \frac{\theta}{2} & \cos \frac{\theta}{2} & 0 & 0 \\
0 & 0 & \cos \frac{\theta}{2} & \sin \frac{\theta}{2} \\
0 & 0 & -i \sin \frac{\theta}{2} & \cos \frac{\theta}{2}
\end{bmatrix},$$  

(135)

where $\theta = -\mu_1 \Omega V_{\text{at}}(t)$, which can be used to generate a CNOT with the addition of only single-qubit gates

$$\begin{align*}
\begin{array}{ccc}
\text{Z} & \text{CR} & \text{Z} \\
\text{X} & \text{Z}
\end{array}
\end{align*},$$  

(136)

up to a phase $e^{i \pi/4}$.

2. Improvements to the CR gate and quantum error correction experiments using CR

Since qubit 1 is being driven off-resonance, an ac-Stark shift will add a term $\propto \sigma_z \frac{1}{2}$ to the driving Hamiltonian of qubit 1. The effect of both the spurious ac Stark shift and the direct $\nu_1^{-1} \otimes \sigma_z$ single-qubit rotations was studied in Ref. 239. By modifying the original CR protocol to effectively "echo away" the two unwanted contributions from the $\sigma_z \frac{1}{2}$ and $\sigma_z \sigma_z$ terms, the fidelity of the CR gate was improved to $F_{\text{CR}} = 0.8799$ using quantum process tomography. Using interleaved randomized benchmarking of this improved "echo-CR"-gate (e CR), a gate fidelity of $F_{\text{eCR}} = 0.9347$ was achieved. This gate implementation was used to demonstrate two-qubit parity measurements in a three-qubit device as well as detecting bit-flip and phase-flip errors in a Bell state encoded in a four-qubit device, with gate fidelities from interleaved randomized benchmarking in the range 0.94 to 0.96. Using a similar device, but with five qubits, weight-four parity measurement of the forms $\mathbf{ZZZZ}$ and $\mathbf{XXXX}$ were demonstrated, where the cross-talk to qubits not involved in the CR gates was studied, leading to the development of a four pulse eCR$_4$-pulse scheme.

Based on improvements in the analysis of the Hamiltonian describing the CR drive, Sheldon et al. subsequently demonstrated a version of the CR which reduced the gate time to $\tau = 160$ ns and added an active cancelation tone to the CR previously developed. Using this "active cancelation echo CR" (ace CR), the fidelity was increased to $F_{\text{aceCR}} = 0.991$, measured with interleaved randomized benchmarking. The same sequence without active cancelation on the same qubits yielded $F_{\text{CR}} = 0.948$. The interested reader may consult the follow-up theoretical work with more details on the effective Hamiltonian models. Other approaches to fast, high-fidelity cross-resonance gates have also been proposed.

This series of improvements to the original cross-resonance implementation has increased the gate fidelity to beyond the threshold for fault-tolerance in a surface code, with similar quality to the CPHASE gate. Although improvements should still be made, with the advent of the CR gate, superconducting qubit based quantum computing platforms now offer two entangling two-qubit gates that can be used for implementing surface-code based error correction schemes.

In the initial experiments using CR gates, the gate times were significantly longer than the typical CPHASE gate times ($\tau_{\text{PHASE}} = 30–60$ ns and $\tau_{\text{CR}} = 300–400$ ns), which to a large extent accounts for the observed CR gate fidelities. The time scale for CR operation is set by the frequency detuning, the anharmonicity, and the coupling strength, through Eq. (132). This has the unfortunate drawback that if qubits do not have the intended frequencies (due to fabrication variation), they will be immediately manifested as longer gate times, and in turn, reduced gate fidelity. As fabrication techniques are becoming more sophisticated and reliable, this problem may be of reduced importance. However, since the coupling in the CR scheme is always on, there is an inherent tension between well-isolated qubits for high-fidelity single-qubit operations, and coupling qubits, for fast/high-fidelity two qubit gates.

3. Quantum simulation and algorithm demonstrations with the CR gate

Since the form of the CR Hamiltonian $\sigma_z \otimes \sigma_z$ is not a $(\sigma_z \otimes \sigma_z + \sigma_z \otimes \sigma_z)$-type interaction (leading to iSWAP gate) nor is it an the effective $(\sigma_z \otimes \sigma_z)$-type (leading to CPHASE gate), one could question its applicability to quantum-simulation-type experiments, which often involves terms of the form $\sigma_z \otimes \sigma_z$. However, by developing a variational quantum eigensolver routine that efficiently generates entangled trial states using the CR interaction, Kandala et al. calculated the ground-state energy for $\text{H}_2$, LiH, and $\text{BeH}_2$. This experiment was performed on six fixed-frequency qubits, and it employed a technique for compact encoding of the Hamiltonians corresponding to each molecule. As of this writing, this experiment represents the largest molecule for which the ground state has been found using a purely quantum processing approach.
The CR gate is also the native two-qubit gate available on the IBM Quantum Experience quantum processor, which is accessible online. Using the IBM Quantum Experience processor, Takita et al. demonstrated an implementation of a two-logical-qubit (four physical qubit) error detection code. The implementation was inspired by the proposal of Gottesman, which proposed a minimal experiment to claim observation of fault-tolerant encodings using a four qubit error detection code in a five qubit setup. Due to constraints on the connectivity, the work by Takita et al. demonstrated a modified version of the Gottesman encoding, in which two logical qubits are initialized, but only one of them in a fault-tolerant manner. By artificially injecting an error in the state preparation circuit, the authors demonstrate that the probability of correctly preparing a fault tolerant state is greater than the probability of correctly preparing a non-fault-tolerant qubit. This behavior is consistent with expectations for how fault-tolerant encodings work. Simultaneously, Vuillot also used the IBM Quantum Experience machine to study fault-tolerant schemes encoded in that connectivity.

Beyond the applications to error-correction and error-detection, the cross-resonance gate has also been employed in early demonstrations of quantum advantages in machine learning. Risté et al. studied the so-called "learning parity with noise" problem, in which one attempts to learn a bit-string by querying an oracle function $f(D,k) = D \cdot k \mod 2$ with a user-input bit-string $D$. In a first implementation of this problem, the authors show that for a specific instance of the bit-string $k = 11$, a learner with access to quantum operations needs fewer queries to the function $f$. However, by extending the model of learning parity with noise, the authors demonstrated a consistent advantage of the learner with access to quantum operations.

The CR gate was also used to demonstrate the implementation of a supervised learning algorithm where the feature space is encoded as quantum data on the Bloch sphere. In typical supervised learning, an algorithm is exposed to a training set of labeled data, and is subsequently asked to classify a new, unlabeled set of data. In the support vector machine (SVM) approach to such problems, the data is then mapped nonlinearly onto the so-called "feature space," in which the trained algorithm has constructed a separating hyperplane to classify the data. While a full “quantum Support Vector Machine” proposal exists, the algorithm assumes that the data are already present in a coherent superposition. Instead, Havlíček et al. proposed, and demonstrated, that mapping the classical data nonlinearly onto the Bloch sphere can also be utilized to provide a quantum advantage. For a wider discussion of the important role of quantum data in many quantum machine learning algorithms, the reader is referred to Ref. 290.

### 4. Other microwave-only gates: bSWAP, MAP, and RIP

The CR gate (as outlined above) is not the only all-microwave two-qubit gate available. In particular, the bSWAP gate is an interesting alternative. The bSWAP gate directly drives the $|00\rangle \rightarrow |11\rangle$ transition, made possible by interactions with the higher levels of the qubit, see Fig. 18. Usually, the matrix element for such a transition is small (3rd order in the coupling strength), but if the detuning between the qubits is equal to the anharmonicity, the transition rate is enhanced. Applying a sequence of Schrieffer-Wolff transformations to the coupled-qubit system, and using a carefully chosen drive frequency (close to the midpoint of $\omega_{e1}$ and $\omega_{e2}$), it can be shown that the drive gives rise to a unitary operator

$$U = U_{\text{bSWAP}}(\theta, \phi) U_{\text{ZZ}, \text{ZI}, \text{ZI}}$$

with

$$U_{\text{bSWAP}}(\theta, \phi) = \begin{pmatrix}
\cos \theta & 0 & 0 & -ie^{-i\phi} \sin \theta \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
-ie^{-i\phi} \sin \theta & 0 & 0 & \cos \theta
\end{pmatrix}.$$  (137)

The two unitaries $U_{\text{ZZ}}$ and $U_{\text{ZI}, \text{ZI}}$ only contain terms that commute with $U_{\text{bSWAP}}(\theta, \phi)$, and their effects can be offset in postprocessing. In Eq. (138), $\phi$ is the phase of the drive relative to the single-qubit drive pulses, and $\theta = \Omega \delta t$ with

$$\Omega_b = -2g\Omega^2 \{ -g^2 x_2 + g^2 x_2 z_1 + \Lambda_{12} z_1 z_2 \} / (z_1 + \Lambda_{12} z_2 \Lambda_{12}),$$  (139)

where $\Omega$ is the amplitude of the drive, $g$ is a dimensionless parameter quantifying the coupling coefficient of the drive to qubit 2 in units of coupling strength to qubit 1, and $x_2 = x_1 + \Lambda_{12}$. Explicit derivations leading to Eq. (137) can be found in the supplement of Ref. 291. By applying $U_{\text{bSWAP}}$ for a time that yields $\theta = \pi/2$, and with $\phi = 0$, the resulting gate is denoted $\text{iSWAP}$ and can act as the entangling gate (together with single-qubit gates) that forms a universal gate set. Moreover, the power of the $\text{iSWAP}$ gate becomes apparent when one applies it for the time that yields $\theta = \pi/4$, which from the ground state $|00\rangle$ directly produces the entangled Bell state $|00\rangle + e^{i\phi}|11\rangle$. In line with the definition of $\sqrt{\text{SWAP}}$, this gate is denoted the $\sqrt{\text{iSWAP}}$. In the work by Poletto et al., the fidelity of the $\text{iSWAP}$ gate was $F_{\text{iSWAP}} = 0.9$ (determined from quantum process tomography). The main source of error was the increased dephasing during the relatively long high-power pulse needed to drive the $|00\rangle \rightarrow |11\rangle$ transition. The $\text{iSWAP}$ gate can be viewed as the superconducting qubit analog of the Mølmer-Sørensen gate. In Fig. 18, we outline the level diagram of two coupled qubits, along with the higher levels of
the qubits. The arrows indicate which coupled states are utilized to implement the corresponding gate. As an application of the $\text{bSWAP}$ gate, Collers et al.\textsuperscript{50} used this gate to calculate energies of the excited states of a $H_2$ molecule using a two-qubit transmon processor.\textsuperscript{295}

Another all-microwave gate is the so-called "microwave-activated CPHASE" (or "MAP") gate.\textsuperscript{296} The MAP gate is in a spirit similar to that of the CPHASE gate, where noncomputational states are used to impact a conditional phase inside the computational subspace. In contrast to CPHASE, the MAP gate is implemented without tuning individual qubit frequencies. Rather, the canonical implementation of this gate comprises two fixed-frequency qubits, where the frequencies are carefully designed (and fabricated), such that the $|12\rangle$ and $|03\rangle$ levels are resonant. This leads to a splitting of the otherwise degenerate $|02\rangle \leftrightarrow |01\rangle$ and $|12\rangle \leftrightarrow |11\rangle$ transitions. By driving near resonance with the $|n2\rangle \leftrightarrow |n1\rangle$ transition, an effective $\sigma_z \otimes \sigma_z$ interaction is generated. In a setup comprising two fixed-frequency qubits, the MAP gate was used to implement the unitary

$$U_{\text{MAP}} = \exp \left[ -i \frac{\pi}{4}\sigma_z \otimes \sigma_z \right],$$

with a gate fidelity $F_{\text{MAP}} = 0.87$ (determined via quantum process tomography) in a time $\tau_{\text{MAP}} = 514$ ns.\textsuperscript{297} As the number of qubits in a system increase, one drawback of this gate is the need for a precise matching of higher energy levels across multiple qubits, while simultaneously avoiding spurious couplings to other modes in the system.

The CPHASE, $\text{iSWAP}$ and MAP gates all have quite stringent requirements on the spectral landscape of the qubits in order to obtain fast, efficient gate operation. To address this issue, another all-microwave gate was developed, the so-called "resonator induced phase gate" ("RIP").\textsuperscript{294,295} The RIP gate operates by coupling two fixed-frequency qubits to a bus cavity, from which they are far detuned. By adiabatically applying and removing an off-resonant pulse to the cavity, the system undergoes a closed loop in phase space, after which the cavity is left unchanged, but the qubits acquire a state-dependent phase. By a careful choice of the amplitude and detuning of the pulse, and taking into account the dispersive shift of the cavity, the CPHASE gate can be implemented on the two qubits. This effect was experimentally demonstrated by Paik et al.\textsuperscript{296} in a 3D transmon system,\textsuperscript{297} where four qubits are coupled to the same bus. In this setup, the RIP gate operation results in unitaries with weight on all four qubits simultaneously. In order to isolate just the desired two-qubit coupling terms, Paik et al.\textsuperscript{296} developed a "refocused" RIP (RTRIP) gate that implements

$$U_{\text{RTRIP}} = \exp \left[ -i\hat{n}\sigma_z \otimes \sigma_z \right],$$

where the coupling rate (for an unmodulated drive) scales as

$$\hat{n} \propto \frac{(\Delta V) \omega}{2 \Delta \tilde{\omega}} \frac{Z}{\Lambda_d},$$

where $\hat{n}$ denotes the average number of photons in the bus, $\omega$ is the dispersive shift, and $\Lambda_d$ is the detuning of the drive ($\tilde{d}$) from the cavity ($c$). By choosing $\theta = \pi/4$, it is possible to implement the CPHASE gate. The power of the RIP gate lies in its capability to accommodate large differences in qubit frequencies. To demonstrate this, Paik et al.\textsuperscript{296} performed two-qubit randomized benchmarking between pairs of qubits in a four-qubit device with frequency differences spanning from 0.38 GHz to 1.8 GHz, all with fidelities in the range 0.96–0.98 and gate times in the range 285 to 760 ns.

### H. Gate implementations with tunable coupling

Finally, we briefly review tunable coupling architectures, which have recently emerged as a promising alternative. The idea is to engineer an effective qubit-qubit coupling $\hat{g}$ that is tunable (typically by applying a flux), and such gates are referred to as parametric gates. This can be implemented in two different ways: (i) The coupling strength between two qubits is tuned by a flux, $g \rightarrow g'(\Phi(t))$, or (ii) the resonant frequency of the coupling element is modified $\omega_{\text{coupler}} \rightarrow \omega_{\text{coupler}}(\Phi(t))$, with a fixed $g$, leading to an effective time-dependent coupling parameter. When the tunable coupling element is driven at frequencies corresponding to the detuning of the qubits from the coupler, an entangling gate can be implemented.

In a setup of type (ii), an implementation of the $\text{iSWAP}$ gate was demonstrated by parametrically driving a flux-tunable coupler between two fixed-frequency qubits,\textsuperscript{63} yielding a fidelity $F_{\text{iSWAP}} = 0.9823$ (using interleaved randomized benchmarking) in a time $\tau = 183$ ns. Similarly, the $\text{iSWAP}$ (and $\text{SWAP}$) gates were recently demonstrated, using a flux-tunable transmon connecting two fixed-frequency transmons. Driving the flux through the tunable qubit at the sum frequency of the fixed-frequency transmons results in the $\text{iSWAP}$ gate. This parametrically driven approach is generally significantly faster than implementations relying solely on fixed-frequency qubits.

A hybrid approach, in which a combination of tunable and fixed-frequency qubits is used, was recently demonstrated for both $\text{iSWAP}$ and CPHASE gates.\textsuperscript{60,105,206} This scheme has no added tunable qubits (or resonators) acting as the coupling element, but rather, relies solely on an always-on capacitive coupling between the qubits, and the effective coupling is roughly half that of the always-on coupling. The operational principle here is to modulate the frequency of the tunable qubit (using local flux control) at the transition frequency corresponding to $|01\rangle \leftrightarrow |10\rangle$ for $\text{iSWAP}$ and $|11\rangle \leftrightarrow |02\rangle$ for CPHASE. Using interleaved randomized benchmarking, the authors demonstrated $F_{\text{iSWAP}} = 0.94$ ($\tau = 150$ ns), and $F_{\text{CPHASE}} = 0.93$ ($\tau = 210$ ns) and $F_{\text{CPHASE}} = 0.88$ ($\tau = 290$ ns), showing a slight asymmetry in the direction in which the CPHASE is applied. This hybrid technique was used in Ref. 67 to demonstrate a four-qubit GHZ state with fidelity $F_{\text{4-qubit GHZ}} = 0.79$ (using state tomography). Finally, this gate-architecture was used to demonstrate a hybrid quantum/classical implementation of an unsupervised learning task (determining clustering of data), using nineteen qubits and supplemented by a classical computer as part of the minimization loop.\textsuperscript{65}

### V. QUBIT READOUT

The ability to perform fast and reliable (high fidelity) readout of the qubit states is an important cornerstone of any quantum processor.\textsuperscript{3}

In this section, we give a brief introduction to how readout is performed on superconducting qubits. We start by reviewing the fundamental theory behind "dispersive readout"—the most common readout technique used today in the circuit QED architecture—in which each qubit is coupled to a readout resonator. In the dispersive regime, i.e., when the qubit is detuned from the resonator frequency, the qubit

---

Published under license by AIP Publishing
induces a state-dependent frequency shift of the resonator from which the qubit state can be inferred by interrogating the resonator.

A dispersive readout allows us to map the quantum degree of freedom of the qubit onto the classical response of the linear resonator, thus transforming the readout optimization process into obtaining the best signal-to-noise ratio (SNR) of the microwave signal used to probe the resonator.

We then provide guidance on how to optimize system parameters to perform high-fidelity, single-shot readout. After choosing parameters, such as resonant frequencies and coupling rates, we address the filter and amplifier circuitry positioned in-between the qubit plane and the data acquisition hardware outside of the dilution refrigerator. On this note, we review the basic principles of Purcell filters as well as parametric amplifiers, both of which are necessary to obtain a fast, high-fidelity readout in scaled-up quantum processors.

A. Dispersive readout

A quantum measurement can be described as an entanglement of the qubit degree of freedom with a “pointer variable” of a measurement probe with a quantum Hamiltonian,\(^{307-309}\) followed by classical measurement of the probe. In circuit QED, the qubit (the quantum system) is entangled with an observable of a superconducting resonator (the probe), see Fig. 19(a), allowing us to gain information about the qubit state by interrogating the resonator—rather than directly interacting with the qubit. Therefore, the optimization of the readout performance is translated to maximizing the signal-to-noise ratio of a microwave probe tone sent to the resonator, while minimizing the unwanted “back-action” on the qubit.

The qubit-resonator interaction is described by the Jaynes-Cummings Hamiltonian,\(^{1-3} \text{ Sec. II}\) previously introduced in Sec. II

\[
H_{JC} = \omega_0 \left( a^\dagger a + \frac{1}{2} \right) + \omega_q \left( \frac{1}{2} \sigma_z + g \left( \sigma_z a + \sigma_\alpha a^\dagger \right) \right), \quad (143)
\]

where \(\omega_0\) and \(\omega_q\) denote the resonator and qubit frequencies, respectively, and \(g\) is the transverse qubit-resonator coupling rate. The operators \(\sigma_z\) and \(\sigma_\alpha\) represent the processes of exciting and de-exciting the qubit, respectively.

In the limit when the detuning between the qubit and the resonator is small compared with their coupling rate, i.e., \(\Delta = |\omega_0 - \omega_q| \ll g\), the energy levels of the two systems hybridize and a vacuum Rabi splitting of frequency \(\sqrt{\gamma g}/\pi\) opens up, where \(n = 1, 2, 3...\) denotes the resonator mode. In this regime, excitations are coherently swapped between the two systems. Although useful for certain two-qubit gate operations, recall Sec. IV E, such transverse interactions change the qubit state (since energy is directly exchanged between the resonator and the qubit) and is therefore not desired in the context of “quantum nondemolition” (QND) readout, in which the outcome of the quantum measurement is not altered in the act of reading out the system.

In the dispersive limit, i.e., when the qubit is far detuned from the resonator compared with their coupling rate \(g\) and the resonator linewidth \(\kappa\), \(\Delta \gg g, \kappa\), there is no longer a direct exchange of energy between the two systems. Instead, the qubit and resonator push each others’ frequencies. To see this, the Hamiltonian can be approximated using second-order perturbation theory\(^{26,31}\) in terms of \(g/\Delta\), taken in the limit of few photons in the resonator. This is known as the “dispersive approximation,” after which the Hamiltonian takes the form

\[
H_{\text{disp}} = (\omega_q + \chi \sigma_z) \left( a^\dagger a + \frac{1}{2} \right) + \frac{\gamma g}{2} \sigma_z, \quad (144)
\]

where \(\chi = g^2/\Delta\) is the qubit-state dependent frequency shift, it is called “dispersive shift,” see Fig. 19(b), allowing us to distinguish the two qubit states. This is an asymptotically longitudinal interaction, yielding a QND measurement. Note that, in addition, the qubit frequency also picks up a “Lamb shift,” \(\omega_q^\prime = \omega_q + g^2/\Delta\), induced by the vacuum fluctuations in the resonator. Also note that the dispersive shift

![Fig. 19](https://example.com/fig19.png)

**Fig. 19.** (a) Simplified schematic of a representative experimental setup used for dispersive qubit readout. The resonator probe tone is generated, shaped and timed using an arbitrary waveform generator (AWG), and sent down into the cryostat. The reflected signal \(S_{11}\) is amplified, first in a parametric amplifier and then in a low-noise HEMT amplifier, before it is downconverted using heterodyne mixing and finally sampled in a digitizer. (b) Magnitude \(S_{11}\) and phase \(\theta\) of the resonator with linewidth \(\kappa\), when the qubit is in its ground state \(\ket{0}\) (blue) and excited state \(\ket{1}\) (red), separated with a frequency \(2\pi/\kappa\). (c) Corresponding complex plane representation, where each point is composed of the in-plane \(\Re[S_{11}]\) and quadrature \(\Im[S_{11}]\) components. The highest state discrimination is obtained when probing the resonator just in-between the two resonances, [dashed line in (b)], thus maximizing the distance between the states.
Hamiltonian in Eq. (144) is derived for a two-level atom. Taking the second excited state into account and introducing the anharmonicity \( \chi = \omega q^{\dagger} q^2 - \omega^2 q q^2 \) modifies the expression for the dispersive shift,

\[
\chi = \chi_0 + \frac{\chi_{12}}{2} = -\frac{\chi_0^2}{\Lambda} \left( \frac{1}{1 + \Lambda/\chi} \right),
\]

(145)

which for a transmon qubit with \( \chi < 0 \) implies that the dispersive shift will depend on the detuning. This effect is plotted in Figs. 20(a) and 20(b), where the second energy level manifests itself as a second vertical asymptote at \( \Delta/2\pi = E_c/\hbar \). It is also worth noting that for qubit modalities with positive anharmonicity, e.g., flux qubits, the dispersive shift will also shift the sign.\(^6\)

In the small photon-number limit, the interaction term of the Hamiltonian in Eq. (144) commutes with the qubit observable,\(^{431} \sigma_\mu \), resulting in a QND measurement.\(^{306} \) This is an important condition for many applications in quantum information processing. However, it has been demonstrated that it is still possible to read out the qubit state by applying a very strong resonator drive tone, even though this readout scheme is not QND.\(^{317} \)

In the case when the resonator photon number \( n = a^\dagger a \) exceeds a "critical photon number" \( n_c \equiv E_c^2/(4 \Delta^2) \), the dispersive Hamiltonian in Eq. (144) is no longer a valid approximation.\(^{268,311,312} \)

Therefore, the critical photon number sets an upper bound for the power level of the resonator probe signal to maintain (an approximate) QND measurement.\(^{432} \) This limitation could be lifted by implementing a pure (and not only approximate) QND readout using a manifestly longitudinal coupling between a qubit and the resonator. Several groups are currently pursuing the implementation of "longitudinal readout," in which QND readout could be performed even with a larger number of resonator photons, thus improving the SNR.\(^{105,314,315} \)

We can also interpret the dispersive qubit-resonator interaction in another way; by rearranging the terms in Eq. (144), we can equivalently write

\[
H_{\text{disp}} = \omega_0 \left( a^\dagger a + \frac{1}{2} \right) \left( \frac{h^2}{\Lambda} \right) + \frac{2 \omega^2}{\Lambda} a^\dagger a \sigma_\mu,
\]

(146)

where the bare qubit frequency is shifted by a fixed amount \( \omega^2/\Delta \), known as the Lamb shift,\(^{433} \) as well as an amount proportional to the number of photons populating the resonator.\(^{62,64} \) This effect is known as the "ac-Stark shift." It has the consequence that photon number fluctuations (noise) in the resonator induce small shifts of the qubit frequency, slightly bringing the qubit out of its rotating frame and thus causing dephasing.\(^{442} \)

This means that spurious photon occupation and fluctuation in the resonator, be it thermal or coherent photons, shift the qubit frequency and causing dephasing.\(^{311,316} \) For this reason, it is important to make sure that the processor is properly thermalized,\(^{106} \) and its control lines well filtered\(^{417} \) and attenuated,\(^{443} \) to reduce photon number fluctuation.

### B. Measuring the resonator amplitude and phase

In Sec. V A, we outlined the underlying physics behind the dispersive readout technique, in which we concluded that the qubit induces a state-dependent frequency shift of the resonator. We now focus our attention on how to probe the resonator to "read out the qubit," that is, best distinguish the two classical resonator signatures corresponding to our qubit states, see Figs. 19(b) and 19(c).

The readout circuit can be set up in measuring either reflection or transmission. The best state discrimination is obtained by maximizing the separation between the two states in the \( (\ell, q) \)-plane, i.e., the in-phase and quadrature component of the voltage, see Fig. 19(c). It can be shown that this separation is maximal when the resonator is probed just in-between the two qubit-state dependent resonance frequencies,\(^{147} \) \( \omega_{\ell q} = (\omega_0^0 + \omega_0^1)/2 \). In this case, the reflected magnitude is identical for \( |0\rangle \) and \( |1\rangle \), and all information about the qubit state is encoded in the phase \( \theta \), see dashed line in Fig. 19(b). In turn, the qubit-resonator detuning should be designed to obey the criterion for maximal state visibility, \( \chi = \kappa/2 \), which is maximized for phase measurements while constraining qubit dephasing.

Once we have picked the resonator probe frequency, the quantum dynamics of the qubit can be mapped onto the phase of the classical microwave response. In the following, we discuss how we can use heterodyne detection to measure the phase of the resonator response. We assume that the reader is already somewhat familiar with basic mixer operations, such as modulation and demodulation of signals. For interested readers, we refer to Ref. 318.
1. Representation of the readout signal

A readout event commences with a short microwave tone directed to the resonator at the resonator probe frequency \(\omega_{\text{RO}}\). After interacting with the resonator, the reflected (or transmitted) microwave signal has the form

\[
s(t) = A_{\text{RO}} \cos(\omega_{\text{RO}} t + \theta_{\text{RO}}),
\]

(147)

where \(\omega_{\text{RO}}\) is the “carrier frequency” used to probe the resonator. \(A_{\text{RO}}\) and \(\theta_{\text{RO}}\) are, respectively, the qubit-state-dependent amplitude and phase that we want to measure. One can equivalently use a “complex analytic representation” of the signal,

\[
s(t) = \text{Re} \left\{ A_{\text{RO}} e^{i(\omega_{\text{RO}} t + \theta_{\text{RO}})} \right\} = \text{Re} \left\{ A_{\text{RO}} \cos(\omega_{\text{RO}} t + \theta_{\text{RO}}) + j \sin(\omega_{\text{RO}} t + \theta_{\text{RO}}) \right\}
\]

(148)

where \(\text{Re}\) takes the real part of an expression, e.g., \(\text{Re}\{\exp(jx)\} = \text{Re}\{\cos x + j \sin x\} = \cos x\).

To gain intuition, we can rewrite Eq. (148) in a static “phasor” notation that separates out the time dependence \(e^{j\omega_{\text{RO}} t}\)

\[
s(t) = \text{Re} \left\{ A_{\text{RO}} e^{j\theta_{\text{RO}}} e^{\text{phasor}} \right\},
\]

(149)

where the phasor \(A_{\text{RO}} e^{j\theta_{\text{RO}}} \equiv A_{\text{RO}} \angle \theta_{\text{RO}}\) is a shorthand that fully specifies a harmonic signal \(s(t)\) at a known frequency \(\omega_{\text{RO}}\). To perform qubit readout, we want to measure the “in-phase” component \(I\) and a “quadrature” component \(Q\) of the complex number represented by the phasor

\[
A_{\text{RO}} e^{j\theta_{\text{RO}}} = A_{\text{RO}} \cos \theta_{\text{RO}} + j A_{\text{RO}} \sin \theta_{\text{RO}},
\]

(150)

\(\equiv I + jQ\)

(151)

to determine the amplitude \(A_{\text{RO}}\) and the phase \(\theta_{\text{RO}}\) (Fig. 20).

2. I-Q mixing

One direct means to extract \(I\) and \(Q\) is to perform a “homodyne” or “heterodyne” measurement using an analog “I-Q mixer.” Figure 21 shows a basic electrical schematic of an I-Q mixer. The readout signal \(s(t)\) and a reference local-oscillator signal \(y(t) \equiv A_{\text{LO}} \cos \omega_{\text{LO}} t\) are fed into the mixer via the RF and LO mixer ports. The mixer then equally splits the signal and local oscillator into two branches and multiplies them in the following way: in the I-branch, the signal \(s(t) = s(t)/2\) is multiplied by the local oscillator \(y_I(t) = (A_{\text{LO}}/2) \cos \omega_{\text{LO}} t\); and in the Q-branch, the signal \(s_Q(t) = s(t)/2\) is multiplied by a \(\pi/2\)-phase-shifted version of the local oscillator, \(y_Q(t) = -(A_{\text{LO}}/2) \sin \omega_{\text{LO}} t\).

The “-” sign arises from the choice of using a \(A(\cos \omega_{\text{LO}} t + \phi)\) as the reconstructed real signal. At the I and Q ports, the output signals \(I(t)\) and \(Q(t)\) contain terms at the sum and difference frequencies, generally referred to as an “intermediate frequency,” \(\omega_{\text{IF}} \equiv \omega_{\text{LO}} \pm \omega_{\text{RO}}\). The resulting signals are low-pass filtered, passing only the terms at the difference frequency, \(I_{IF}(t)\) and \(Q_{IF}(t)\), which are then digitized. After digital signal processing, one obtains the static in-phase \(I\) and quadrature \(Q\) components, from which one calculates the amplitude \(A_{\text{RO}}\) and the phase \(\theta_{\text{RO}}\).

Microwave mixers use square-law-type diodes to implement multiplication. The optical analog of a mixer operation is a combination of a balanced (50–50) beamsplitter followed by optical photodetectors, as shown in the inset of Fig. 22(a). The signal and local-oscillator optical fields are first combined at the beamsplitter, yielding superpositions of both fields, and then detected at the photodetectors, which act as square-law devices. To build intuition for how this works, the square of the sum of two electric fields \((E_1^2 + E_2^2 + 2E_1E_2)\) has a cross term that is the multiplication of the two fields. We refer the reader to Ref. 319 for further details.

3. Homodyne demodulation

One direct means to extract \(I\) and \(Q\) is to perform a microwave homodyne measurement using an analog I-Q mixer of the type shown in Fig. 21. In an analog homodyne measurement, the local oscillator (LO) is chosen to be at the carrier frequency \(\omega_{\text{LO}} = \omega_{\text{RO}}\). Upon mixing, \(I(t)\) and \(Q(t)\) contain terms at both DC \((\omega_{\text{IF}} = 0)\) and terms at twice the carrier frequency. Time-averaging (filtering) \(I(t)\) and \(Q(t)\) directly yield the DC terms \(I_{IF}(t) = I\) and \(Q_{IF}(t) = Q\).

\[
I = \frac{1}{T} \int_0^T dt s_I(t) y_I(t) = \frac{A_{\text{RO}} A_{\text{LO}}}{8} \cos(\theta_{\text{RO}}),
\]

(152)

\[
Q = \frac{1}{T} \int_0^T dt s_Q(t) y_Q(t) = -\frac{A_{\text{RO}} A_{\text{LO}}}{8} \sin(\theta_{\text{RO}}),
\]

(153)
where \( T \) is a time interval taken to be an integer number of periods of the readout signal. \( I \) and \( Q \) are then sampled and used to calculate the amplitude and phase,

\[
A_{\text{RO}} \propto \sqrt{I^2 + Q^2},
\]

\[
\theta_{\text{RO}} = \arctan(Q/I).
\]

Note that the global value of \( A_{\text{RO}} \) or \( \theta_{\text{RO}} \) is not what matters; what matters is the “change” in \( A_{\text{RO}} \) and \( \theta_{\text{RO}} \) between the qubit being in-state \( |0\rangle \) and state \( |1\rangle \). For example, the value of \( A \) leaving the resonator and the value \( G \times A \) reaching a measurement stage are different, where \( G \) represents the net gain in the measurement amplifier chain. However, the gain is the same, independent of the qubit state, whereas \( A \) may be different, e.g., \( A_{\text{RO}}^{(0)} = G \times A_{\text{LO}} \) or \( A_{\text{RO}}^{(1)} = G \times A_{\text{LO}} \).

Similarly, the propagation phase \( \phi \) accumulated while a signal travels between the resonator and the measurement stage is also independent of the qubit state, and simply imparts a phase offset to the qubit-induced phase shift, e.g., \( \theta_{\text{RO}}^{(0)} = \theta_{\text{LO}} + \phi \) or \( \theta_{\text{RO}}^{(1)} = \theta_{\text{LO}} + \phi \).

Homodyning works in principle, but there are two drawbacks. First, signals directly demodulated to DC may be subject to lower signal-to-noise ratios, since they fight against 1/f electronics noise, as well as any other noise signals that may have inadvertently been demodulated (e.g., via a square-law detector). The second is that homodyning is not compatible with frequency division multiplexing (FDM), where a single pulse can be used to interrogate \( N \) resonators at different frequencies by applying tones at each resonator frequency using the superposition principle, e.g.,

\[
s(t) = \sum_{i=1}^{N} A_{\text{RO}}^{(i)} \cos(\omega_{\text{RO}}^{(i)} t + \theta^{(i)}).
\]

Homodyning an FDM signal will put all resonator signals at DC, and once downconverted, they cannot be differentiated. To work around this, it is generally advantageous to use “heterodyning,” which uses a two-step demodulation process via an intermediate frequency \( \omega_{\text{IF}} \). Such a scheme is easily compatible with the concept of FDM, because a readout signal is first demodulated to unique intermediate frequencies (IF) frequencies \( \omega_{\text{IF}} \), and then digitally demodulated to extract each \( A_{\text{RO}}^{(i)} \) and \( \theta^{(i)} \).

In the following, we will consider \( N = 1 \) for simplicity, but the process is applicable to larger \( N \) provided the frequencies sufficiently spaced to avoid interference with one another during the demodulation process.

### 4. Heterodyne demodulation

In a heterodyne scheme, a local oscillator at frequency \( \omega_{\text{LO}} \) is offset by an intermediate frequency \( \omega_{\text{IF}} \) to target a unique readout frequency \( \omega_{\text{RO}} \). Up-conversion techniques such as single-sideband modulation with suppressed carrier (SSB-SC) using balanced I-Q mixers (operated in reverse compared with Fig. 21) are commonly used to create such readout signals. We refer the reader to Ref. 318 for more information on how to create such pulses.

Here, we want to extract \( A_{\text{RO}} \) and \( \theta_{\text{RO}} \) (or their scaled and offset versions) from the reflected/transmitted tone using a heterodyning scheme. The first step is to perform analog I-Q mixing, as illustrated in Fig. 22(a). In contrast to the homodyning case, here, the local oscillator and readout tone are at different frequencies, \( \omega_{\text{RO}} = \omega_{\text{RO}} - \omega_{\text{LO}} \). Mixing the LO and RO signals yields the signals \( I(t) \) and \( Q(t) \) with terms at both sum and difference frequencies. Filtering out the sum frequencies using low-pass filtering (time averaging) yields the IF signals.
$$I_{W}(t) = \frac{1}{T} \int_{0}^{T} dt \ s_{I}(t)y_{I}(t)$$
$$= \frac{A_{RO}A_{LO}}{8} \cos (\omega_{RF} + \theta_{RO}), \quad (157)$$
$$Q_{W}(t) = \frac{1}{T} \int_{0}^{T} dt \ s_{Q}(t)y_{Q}(t)$$
$$= \frac{A_{RO}A_{LO}}{8} \sin (\omega_{RF} + \theta_{RO}). \quad (158)$$

As before, we have omitted any offset phases from the LO or from the wave propagation between the resonator and the measurement. Again, these offset values are not what matters; it is the change in $A_{RO}$ and $\theta_{RO}$ with a change in qubit state that allows state discrimination.

The analog-demodulated $I_{W}(t)$ and $Q_{W}(t)$ are now oscillating at a frequency that is generally low enough to be digitized using commonly available analog-to-digital converters (ADCs). The resulting digital signals are now written as $I_{W}[n]$ and $Q_{W}[n]$.

$$I_{W}[n] = \frac{A_{RO}A_{LO}}{8} \cos (\Omega_{RF} n + \theta_{RO}), \quad (159)$$
$$Q_{W}[n] = \frac{A_{RO}A_{LO}}{8} \sin (\Omega_{RF} n + \theta_{RO}), \quad (160)$$

where $n = t/\Delta t$ indexes the sample number of the continuous-time signals $I_{W}(t)$ and $Q_{W}(t)$. $\Omega_{RF} = \omega_{RF} / \Delta t$ is the digital frequency, and $\Delta t$ is the sampling period (typically around 1 ns). Pulsing the resonator is necessarily accompanied by a ring-up time, related to the quality factor of the resonator, and the first few samples may decrease overall signal-to-noise. Consequently, a delayed window of samples $[n_1 : n_2]$ is often used to perform the second digital demodulation of the discrete-time signals $I_{W}[n_1 : n_2]$ and $Q_{W}[n_1 : n_2]$. Note that more complicated windowing functions may also be used to improve state discrimination, but here we use a simple boxcar [see Fig. 22(b)].

Digital demodulation comprises the point-by-point multiplication of $I_{W}[n_1 : n_2]$ and $Q_{W}[n_1 : n_2]$ by $\cos \Omega_{RF} n$ and $\sin \Omega_{RF} n$. Averaging the resulting time series eliminates the $2\Omega_{RF}$ component while retaining the DC component, as in a homodyne measurement, one obtains

$$I = \frac{1}{M} \sum_{n_1} I_{W}[n] \cos (\Omega_{RF} n) = \frac{A_{RO}A_{LO}}{16} \cos \theta_{RO}, \quad (161)$$
$$Q = \frac{1}{M} \sum_{n_1} Q_{W}[n] \sin (\Omega_{RF} n) = \frac{A_{RO}A_{LO}}{16} \sin \theta_{RO}, \quad (162)$$

where $M = n_2 - n_1 + 1$. As before, $I$ and $Q$ can then be used to estimate $A_{RO}$ and $\theta_{RO}$.

The same procedure may be view in the complex $I$–$Q$ plane by the analytic function $z[n]$, as illustrated in Figs. 22(c) and 22(d)

$$z[n] = I_{W}[n] + jQ_{W}[n] \equiv V_{I}[n] + j V_{Q}[n]$$
$$= \frac{A_{RO}A_{LO}}{8} \left[ \cos (\Omega_{RF} n + \theta_{RO}) + j \sin (\Omega_{RF} n + \theta_{RO}) \right] \quad (163)$$
$$= \frac{A_{RO}A_{LO}}{8} e^{j\theta_{RF}} e^{\Omega_{RF} n}, \quad (164)$$

where the digital in-phase and quadrature signals are represented here as the voltages $V_{I}[n]$ and $V_{Q}[n]$ sampled by the ADC, and we have separated the static phasor $A_{RO}A_{LO}/8$ by multiplying by the complex conjugate of the oscillatory exponential,

$$z[n] = z_{W}[n] \ast e^{j\Omega_{RF} n}, \quad (166)$$

where $\ast$ indicates a point-by-point multiplication, and the result is a vector of length $M$ of nominally identical values of the phasor—one for each sample point—with a small amount of additive noise due to noise in the measurement chain, digitization errors, etc. A singular phasor value is then estimated by taking the average,

$$\bar{z}[n] = \frac{1}{M} \sum_{n} z[n]$$
$$= \frac{A_{RO}A_{LO}}{8} e^{j\theta_{RF}}. \quad (168)$$

Such “single-shot measurements” may then be repeated a large number of times to obtain an ensemble average $\langle \bar{z}[n] \rangle$.

C. Weak and strong qubit measurements: Impact of noise

In quantum measurements, noise plays an essential role as it dictates the fidelity of its outcome, recall Fig. 22(c). In the absence of noise, any nonzero dispersive shift (resulting in a resonator field displacement) would suffice to unambiguously separate the qubit states, given a properly chosen resonator linewidth. In practice, however, the outcome of the quantum measurement is generally Gaussian distributed in the $(I, Q)$-plane due to presence of classical and quantum noise. In this section, we review the main sources of noise, as well as how it impedes our ability to extract information from the quantum system. For a rigorous discussion of noise and quantum measurements, the interested reader is referred, e.g., to the work by Clerk et al. and to the textbook by Haus.

The total noise added to the signal has multiple origins. One part of the noise is associated with the microwave signal used to probe the resonator, where each photon has an intrinsic quantum noise power of $h\omega/2$ per unit bandwidth. Another contribution comes from the phase-preserving amplifiers, adding both classical noise and at least $h\omega/2$ of noise as required by Heisenberg’s uncertainty relation. Finally, any attenuation of the signal prior to the first amplifier will appear as added noise. Combined, these noise sources amount to a “system noise temperature,” which can be characterized using a sensitive thermometer, such as a shot-noise tunnel junction or a qubit as a sensor.

The noise results in time-dependent fluctuations of the measured signal, which in turn translates into uncertainty in our demodulated signals, see Fig. 22(c). This can be intuitively understood by considering that our heterodyne detection method requires us to sample for a finite amount of time.

To quantify the impact of the noise on our measurement, we first project the distributed $(I, Q)$ data—corresponding to $|0\rangle$ and $|1\rangle$—onto the axis for which their relative separation in the complex plane is maximized. The line that is used to separate between $|0\rangle$ and $|1\rangle$ is called a “separatrix.”

The noise can now be quantified by comparing the widths of the Gaussian probability distribution surrounding the mean with the peak separation in the $(I, Q)$-plane, thus defining a signal-to-noise ratio
SNR = \delta \theta / (\Delta \theta_1 + \Delta \theta_0), see Fig. 23(a), with \delta \theta = |\theta_1 - \theta_0| representing the signal and \Delta \theta_1, \Delta \theta_0 represent the noise (2\sigma) of each distribution. The SNR allows us to distinguish between a weak and a strong quantum measurement, as illustrated in Figs. 23(b)–23(d).

In a weak measurement, the probabilities are broadly distributed as compared to their relative separation (SNR < 1), which means that only partial information of the quantum state is revealed to the observer, see Fig. 23(b). In a strong measurement, on the other hand, the quantum state is collapsed onto one of the two eigenstates. In this case, the outcome of the measurement can be distinguished unambiguously, which is reflected in two fully separated distributions (SNR > 1), see Fig. 23(d).

In many applications of quantum measurements, it is necessary to unambiguously (and with high fidelity) tell the outcome without repeating the readout measurement. This is known as “single-shot readout” and it often requires the use of a parametric amplifier—a preamplifier used to increase system SNR—which is further discussed in Sec. V E 3.

Assuming that the widths of the two distributions are identical, \Delta \theta_1 = \Delta \theta_0 = \Delta \theta, the separation error can be calculated by deriving the weight of the overlapping region of the Gaussian distributions as

\[ \epsilon_{\text{sep}} = \frac{1}{\sqrt{2\pi(\Delta \theta)^2}} \int_{-\infty}^{\infty} \exp \left[ -\frac{(\theta - \bar{\theta}_1)^2}{2(\Delta \theta)^2} \right] \, d\theta \]

\[ = \frac{1}{2} \text{erfc} \left[ \frac{|\bar{\theta}_0 - \bar{\theta}_1|}{2\sqrt{2}(\Delta \theta)^2} \right] \]

where \text{erfc}(x) denotes the complementary Gaussian error function, defined as

\[ \text{erfc}(x) = 1 - \frac{2}{\sqrt{\pi}} \int_{x}^{\infty} e^{-t^2} \, dt. \]

Using the \text{erfc} in Eq. (170), the separation error in Eq. (169) can be compactly expressed in terms of the signal-to-noise ratio

\[ \epsilon_{\text{sep}} = \frac{1}{2} \text{erfc} \left[ \frac{\text{SNR}}{2} \right]. \]  

\[ (171) \]

Note, however, that the separation error between the two state distributions only tells us the signal-to-noise ratio of our detection scheme. On top of the separation error, fidelity is reduced if the qubit relaxes (or is excited) during the readout. This will result in a count on the “wrong” side of the threshold. This leads to an additional constraint on the readout; the readout cycle needs to be completed on a time scale much shorter than the qubit relaxation time.

In summary, we see that to optimize the qubit readout fidelity, the readout needs to fulfill the following two requirements:

- “Fast readout”: The readout cycle needs to be completed within a time that is short compared with the qubit coherence time. The longer the readout time, the more likely the qubit is to relax, thus reducing the readout fidelity.
- “High signal-to-noise ratio”: The signal-to-noise ratio needs to be sufficiently large to suppress the state separation errors below an acceptable limit where it does not limit the readout fidelity.

In Secs. V D and V E 3, we review how these two conditions are met by carefully engineering the signal path of the readout circuitry.

D. “Purcell filters” for faster readout

To ensure high-fidelity readout performance, it is important to perform single-shot readout at a time scale much shorter than the qubit coherence time, \( \tau_{\text{q}} \ll T_1 \). This motivates us to: (i) make the resonator linewidth wide, thus reducing its ring-up time, \( \tau_{\text{rd}} \), and (ii) keep the integration time \( \tau_r \) as short as possible, see Fig. 22(b). The ability to isolate a quantum system from decohering into its environment while, at the same time, being able to read out its state in a short time represents two contradictory criteria, which must be traded-off.

Even though dispersive readout (in the few-photon limit) has only a small back-action on the qubit state, the qubit will still suffer from \( T_1 \)-relaxation while we are performing a measurement. In fact, this “decay during the readout” often limits the readout fidelity, reducing it to

\[ F(\tau_{\text{rd}}) = 1 - e^{-\tau_{\text{rd}}/T_1}. \]  

\[ (172) \]
sampling time $\tau_r/2$. The fidelity drop in Eq. (172) can be interpreted as a manifestation of the competition between the time scales at which our quantum information reaches our detector or the environment first.

The limitation of qubit coherence originates from an enhanced spontaneous emission of photons, induced by its environment. This is known as the Purcell effect, and is an important consideration when designing qubit-resonator systems. The portion of spontaneous emission that is mediated by the resonator describes how qubit relaxation is enhanced by the resonator Q when on-resonance, and suppressed off-resonance. The aim of this section is twofold: first, we develop an intuition for how the Purcell decay limits qubit coherence, and second, how to properly mitigate this limitation by designing a so-called “Purcell filter,” which modifies the impedance seen by the qubit through the readout resonator. This allows us to maintain a fast readout, while protecting the qubit from relaxing into its environment.

If we would just choose qubit and resonator operation frequencies guided by the resonator linewidth $\kappa$, qubit-resonator coupling $g$, and the amount of dispersive shift $\delta_e$, we would reduce the detuning between the qubit and the resonator, thus maximizing the dispersive shift (recall Fig. 20). However, this presents a trade-off between two important system parameters; on one hand, we want the qubit to be isolated from the resonator environment off-resonant to avoid Purcell-enhanced decay. On the other hand, looking at the dispersive shift, we want the two rates, $g$ and $\kappa$ to be strong, yielding larger dispersive shift as well as short resonator transient and thus a faster readout.

Fortunately, when operating in the dispersive regime, the qubit and resonator are far detuned from each other $\Delta \gg g, \kappa$, which means that their impedance (environment) can be independently engineered through filter design. In essence, one designs a filter to have strong coupling to the readout port at the resonator frequency (large $\kappa$), but isolates the qubit from its environment at the qubit frequency. In other words, an impedance transformation.

Depending on the design of the readout for the quantum processor to which the filter should be coupled, there are different ways to design a Purcell filter; such as quarter-wave stubs, low-Q bandpass filters, and stepped-impedance filters. Which one is optimal depends on the system properties such as qubit-resonator detunings, required bandwidth, and allowed insertion loss.

The most promising Purcell filter designs are the ones that allow for frequency multiplexing, such as the low-Q bandpass filter design, which in addition to Purcell filtering has the function of a quantum bus, connecting several frequency-multiplexed readout resonators sharing the same amplifier chain.

The Purcell effect can be framed in terms of Fermi’s golden rule, where noise in the environment causes the qubit to decay with some probability. We can gain intuition about the Purcell effect (as well as how the qubit can be protected from it) by replacing the Josephson junction in the qubit circuit with an ac-current source, outputting $I(t) = I_0 \sin(\omega t)$, with $I_0 = e\omega_0$ and study the rate at which power is lost into an environmental load resistor $R = Z_0 = 50 \, \Omega$, see Fig. 24(a).

Expressing the power lost in the resistor as $P = I_0^2 (C_g/C_\Sigma) R = (e\omega_0)^2 Z_0$, with $\beta = C_g/C_\Sigma$, the qubit Purcell decay rate into the continuum can be written as

$$\gamma_{\text{Purcell}} = \frac{1}{\tau_{\text{env}}} = \frac{P}{R_0} = \frac{(e\omega_0)^2 Z_0}{R_0} = \frac{g^2}{\omega}. \quad (173)$$

To protect the qubit from decaying into the 50 \, \Omega environment (as well as for deploying our dispersive readout), we can now add a resonator in parallel with the qubit, see Fig. 24(b). The presence of the resonator has the effect of shaping the impedance at the qubit frequency, which in turn modifies the decay rate in Eq. (173) into

$$\gamma_{\text{Purcell}}^{\text{env}} = \frac{g^2}{\omega} \frac{\text{Re}[Z_\epsilon(\omega)]}{Z_0}, \quad (174)$$

where $Z_\epsilon(\omega)$ denotes the impedance of the shunted resonator. We can express the real part of the impedance in terms of the resonator quality factor $Q = \omega_0/\kappa$ and qubit-resonator detuning $\Delta = \omega_0 - \omega_r$:

$$\text{Re}[Z_\epsilon(\omega)] = \frac{QZ_0}{1 + 2(\Delta/\kappa)^2}. \quad (175)$$

Now, by substituting Eq. (175) into Eq. (174), we see that the Purcell decay rate for the qubit depends on the detuning between the resonator and the qubit. This is intuitive, since the resonator can be thought of as a bandpass filter, with center frequency $\omega_c$ and
bandwidth $\kappa$. For the resonant condition, i.e., when $\Delta = 0$, the emission rate into the resonator takes the form

$$\gamma_{\text{res-env}} = \frac{g^2 \text{Re}[Z_i]}{\omega_r} Q = \frac{g^2}{\kappa}$$

(176)

In the dispersive regime $\Delta \gg g, \kappa$, which is also relevant for us in the context of qubit readout, we can make the approximation $\text{Re}[Z_i] \approx Q Z_0 (\kappa/\Lambda)^2$, yielding the familiar expression for the Purcell decay rate in the circuit QED.\(^3\)

$$\gamma_{\text{res-env}} = \frac{g^2 \text{Re}[Z_i]}{\omega_r} Q = \frac{g^2}{\kappa} \left( \frac{\kappa}{\Lambda} \right)^2$$

(177)

The relation for the Purcell limit in Eq. (177) thus provides us with a useful guide on how to design the coupling rates $g$ and $\kappa$, as well as how large qubit-resonator detuning $\Delta$ is necessary to avoid the Purcell limit.

In recent years, however, the intrinsic coherence times for superconducting qubits have reached above 100 $\mu$s, recall Sec. II, imposing practical limitations on how to simultaneously optimize $g$ and $\kappa$, to render fast readout without compromising the qubit coherence. Considering the parameters in Eq. (177), it is not possible to just increase the bound on the relaxation time $T_1$, without at the same time trading off the readout speed and contrast.

We can now introduce the Purcell filter [Figs. 24(c) and 24(d)] in between the readout resonator and the 50 $\Omega$ environment, leading to a reduction of the decay rate according to\(^3\)

$$\gamma_{\text{res-filter-env}} = \kappa \left( \frac{g}{\Lambda} \right)^2 \left( \frac{\omega_b}{\omega_r} \right)$$

(178)

where $Q_0$ denotes the quality factor of the Purcell filter. This is schematically depicted in Fig. 24(d), where the Purcell filter is placed around the resonator frequency, while far detuned from the qubit.

E. Improve signal-to-noise ratio: Parametric amplification

In light of the aforementioned limited signal-to-noise ratio associated with the low photon number of the dispersive qubit readout, and the short sampling time, the noise temperature of the amplifier chain plays a crucial role in determining the fidelity of the measurement.

A useful benchmark for quantum measurements is the "quantum efficiency," defined as

$$\eta_{\text{SQL}} = \frac{\hbar \delta \Omega}{k_B T_{\text{sys}}}, \quad 0 < \eta_{\text{SQL}} < 1,$$

(179)

which quantifies the photon energy to the system noise temperature $T_{\text{sys}}$, thus yielding a measure of how close the signal is to the standard quantum limit (SQL), as imposed by Heisenberg’s uncertainty relation, adding $1/2$ photon of noise when $\eta_{\text{SQL}}$ approaches unity. Since the energy of each microwave photon is much smaller than that of optical photons, it is not easy to build a single-photon detector operating in the microwave domain.\(^3\) Instead, for heterodyne detection in the circuit QED, a set of cascaded microwave amplifiers are used. The system noise temperature for the amplifier chain can be expressed in terms of the individual gain figures $G_n$ and noise temperatures $T_{N_n}$ of each constituent amplifier\(^3\)

$$T_{\text{sys}} = T_{N_1} + \frac{T_{N_2}}{G_1} + \frac{T_{N_3}}{G_1 G_2} + \cdots$$

(180)

where $n = 1, 2, 3, \ldots$ denotes the order of the amplifiers, starting from the qubit chip. From Eq. (180), we see that the noise temperature $T_{\text{sys}}$ is dominated by the noise contribution from the first amplifier, whereas the gain of the first amplifier has the effect of suppressing the noise from the second amplifier, and so on. If the first amplifier is a low-noise high-electron mobility transistor (HEMT) amplifier ($T_R \approx 2$ K), the system noise temperature when implemented in a cryostat is around 7–10 K, corresponding to around 10–20 added photons of noise per signal photon around 5 GHz. In practice, this is generally too much noise to perform a single-shot readout.

This inherently poor signal-to-noise ratio has revived interest in developing quantum-limited parametric amplifiers (PAs)—tailored for the readout of superconducting qubits—featuring the ability to amplify small microwave signals, and adding only approximately the minimum amount of noise allowed by quantum mechanics.\(^3\)

1. Quantum-limited amplification processes

In a linear, phase-insensitive amplifier, an input state $|a_{in}\rangle$ is amplified to an output state $|a_{out}\rangle$, with an amplitude gain factor $\sqrt{G}$. Microwaves are electromagnetic fields and therefore considered to be coherent light comprising microwave photons. As such, they must obey the commutation relations\(^3\)

$$[a_{in}, a_{in}^\dagger] = [a_{out}, a_{out}^\dagger] = 1,$$

(181)

from which it can be shown that it is not possible to simultaneously amplify both quadratures of $a_{in}$ without also adding noise. This is known as "Caves theorem" after the work by Caves,\(^3\) based on an earlier work by Haus and Muller.\(^3\) This can be seen by considering the scattering relation between the input and output microwave fields

$$a_{out} = \sqrt{G} a_{in}.$$  

(182)

The gain relation in Eq. (182) constitutes our ideal scenario for an amplifier process. However, the problem is that that this relation does not satisfy the commutation relation in Eq. (181). To satisfy this relation, we need to also take into account the vacuum fluctuations of another mode $b_{in}, b_{in}^\dagger$—called the "idler" mode $b_{id}$, also satisfying the same commutation relation $[b_{id}, b_{id}^\dagger] = 1$. To satisfy the commutation relation, the idler mode is amplified by the gain factor $\sqrt{G} - 1$. For a large gain, it can be shown that a minimum amount of half a photon of noise $\hbar \omega_{id}/2$ needs to be added to a signal amplified with gain $\sqrt{G}$.

Finally, taking the idler mode into account, the scattering relation for the coherent output field takes the form

$$a_{out} = \sqrt{G} a_{in} + \sqrt{G - 1} b_{id}^\dagger.$$  

(183)

Amplification

Added idler noise

Generally, this process results in a so-called "phase-insensitive" parametric amplification process, in which both quadratures of the input field get equally amplified. This is illustrated in Fig. 25, where
the in-phase (I\text{in}) and quadrature (Q\text{in}) components of the fields are plotted, before and after the parametric amplifier.

Considering the amplification process in Eq. (183), we can find a special case for the idler mode, for which noiseless amplification can be accomplished for one of the two quadratures, but at the expense of adding more noise to the other, thus not violating Heisenberg’s uncertainty relation for the two field quadratures. This mode of operation is known as “phase-sensitive” amplification, and is obtained when the idler mode oscillates at the same frequency as the signal (or a multiple thereof), but can be shifted with an overall phase $\phi \in [0, 2\pi]$. By substituting the idler mode in Eq. (183) with $b_{\text{in}} = e^{i\phi} a_{\text{in}}$, the scattering relation becomes

$$a_{\text{out}} = \sqrt{G} a_{\text{in}} + e^{i\phi} \sqrt{G - 1} a_{\text{in}}.$$  \hspace{1cm} (184)

The overall phase factor allows us to tune the orientation of the amplification (or de-amplification) by means of the pump phase, thus allowing us to choose a quadrature for which we want to reduce the noise, see Fig. 26. Intuitively, this can be understood by considering the interference that occurs when two waves with the same frequency are confined in space, where we obtain constructive or destructive interference, depending on the phase between the two waves. Due to this interference, the noise can be suppressed even below the standard quantum limit (without violating Heisenberg’s uncertainty relation). This is known as “single-mode squeezing” and was first observed in superconducting circuits by Yurke et al. 336

In particular, after the theoretical prediction by Gardiner, 337 Murch et al. showed that the coherence time of a qubit can be enhanced when the qubit is exposed to squeezed vacuum. 338,339 Also “two-mode squeezing” was demonstrated by Eichler et al., 340 where the demodulation setup squeezes both quadratures of the acquired signal. 107

In the context of qubit readout, however, phase-sensitive amplification tends to be experimentally inconvenient. This is mainly due to its phase-dependent gain, which imposes stringent requirements on continuous phase-calibration of the readout signal.

For a detailed theoretical framework developed for quantum limited amplification, the reader is referred to an earlier work by Roy and Devoret, 341 Clerk et al., 124 and Wustmann and Shumeiko. 342

2. Operation of Josephson parametric amplifiers

In this section, we review the basic operation characteristics of parametric amplifiers, and in particular, the Josephson parametric amplifiers (JPAs), that have been exploited for qubit readout. Although many different flavors of parametric systems exist, we here...
focus on the resonant implementations of the Josephson parametric amplifier (JPA), serving as a good system for reviewing the fundamental concepts around parametric amplification.

All parametric amplifiers operate based on one fundamental principle: the incoming “signal” photons are mixed with an applied “pump” tone via an intrinsic nonlinearity, by which energy from the pump is converted into signal photons and thereby providing gain. As we recall from Sec. II, such a nonlinearity can be engineered in the microwave domain using Josephson junctions, and the resonant parametric amplifiers are built from slightly anharmonic oscillators.

The first Josephson parametric amplifiers were built from a coplanar waveguide resonator, made nonlinear by adding a nonlinear Josephson contribution to its total inductance, see Fig. 27(a). The word “parametric” refers to the process of modulating (or “pumping”) one of the parameters of the system’s equation-of-motion (such as frequency or damping) in time. 

The natural way to perform this parametric pumping is to modulate the nonlinear Josephson inductance, which in turn has the effect of modulating the resonator frequency \( \omega_0(t) = 1/\sqrt{L(t)C} \).

Depending on how the pumping is implemented, there are two different mixing processes that can be exploited in Josephson parametric amplifiers, which determines the characteristics of the amplifier. These are illustrated in Figs. 27(b) and 27(c) and are referred to as “current-pumping” and “flux-pumping” respectively. The type of mixing process that takes place depends on the leading order of the nonlinearity of the system, as reflected in its Hamiltonian. In the following, we briefly review the difference between these two pump-schemes.

In the current-pumped case, the dynamics of the system has characteristics of a Duffing oscillator, with a fourth-order nonlinear term in addition to the harmonic oscillator term in its Hamiltonian

\[
H = \omega_0 c^\dagger c + K c^\dagger c^\dagger c c,
\]

where \( c \) denotes the resonator field operator and \( K \) is the “Kerr-nonlinearity.” This process is a so-called “four-wave mixing” process, since it mixes four photons: one signal (\( \omega_s \)), one idler (\( \omega_i \)), and two pump photons (\( \omega_p \)), obeying the energy conservation relation \( \omega_s + \omega_i = 2\omega_p \), as seen in Fig. 27(b). Pioneered by Yurke, this was the first demonstration of microwave amplification using a Josephson parametric amplifier. When the signal and idler modes are at the same frequency, the amplification is said to be “degenerate.” This pumping scheme is the foundation for the Josephson Bifurcation Amplifier (JBA), developed by Siddiqi et al., which has been used to perform single-shot qubit readout, by mapping the quantum states onto the high and low resonator field originating from the sharp bifurcation point of the amplifier.

In the other case, when the system is flux-pumped, the parametric process is driven by threading a magnetic flux \( \Phi_{dc} \) through a SQUID loop, thereby modulating the frequency of the resonator. This results in a “three-wave mixing” process, comprising three photons: one signal, one idler, and one pump photon, with \( \omega_s + \omega_i = \omega_p \), as seen in Fig. 27(c). Therefore, we see that the pump frequency is about twice that of the signal \( \omega_p \approx 2\omega_s \) for \( \omega_s \approx \omega_i \). For degenerate, flux-pumped systems, the leading nonlinearity is a third-order term, yielding a Hamiltonian

\[
H = \omega_p c^\dagger c + K \left( p c^\dagger c^\dagger + p^* c c \right),
\]

where the \( p \) operator denotes the flux-pump mode. This approach to building parametric amplifiers was developed by Yamamoto et al., as well as by Sandberg et al.

The flux-pumping scheme has several practical advantages. First, the large detuning of the pump makes it easier to filter, isolating the readout signal as its passing into the digitizer downstream and preventing the saturation of following amplifier stages. Second, if the resonator is a quarter-wavelength resonator, it has no resonant mode at the pump frequency \( \omega_p \), reducing spurious population or saturation of the system as well as backaction on the qubits in the processor. Third, since the flux pump line is a separate on-chip microwave line, no additional directional coupler is needed.

Due to its rich dynamics, flux-pumping has also proven a useful platform to study the quantum dynamics of Josephson parametric oscillators, both in the context of qubit readout, the dynamical Casimir effect, and to better understand their complex nonlinear dynamics.

In addition to the degenerate parametric interactions described above, parametric gain can be obtained between different resonant modes; either between different modes of the same resonator, or in-between different resonators, as with the Josephson parametric converter (JPC). In addition to the possibility of isolating and amplifying certain frequencies, the JPC can implement frequency mixing.
conversion for which it has some other areas of applications compared with other types of parametric amplifiers.

3. The traveling wave parametric amplifier

In the previously described JPA, parametric amplification is realized using resonators that enhance the parametric interaction between the input signal and the Josephson junction nonlinearity. Essentially, the Q-enhancement of the resonator forces each photon to pass through the junction on average Q times before leaving the resonator, thereby enhancing the nonlinear interaction. Albeit proven to be able to reach near the standard-quantum limit of noise for readout of a small number of qubits, the future direction of the community is heading toward amplifier technologies which are compatible with multiplexed readout of several qubits coupled to the same amplifier chain.\(^{35,66,379-381}\) In this context, resonator-based parametric amplifiers suffer from two major drawbacks: First, the amplifier bandwidth is limited to the resonator linewidth, typically \(\approx 10-50\) MHz, practically limiting the number of multiplexed frequencies that can be amplified. Second, since the Josephson nonlinearity is realized by a small number of junctions, the saturation power is low due to the interplay of higher order nonlinearities, effectively taking the system outside its desired operation regime.\(^{356,367,368,378}\) In practice, this limits how many readout resonators that can be simultaneously read out.

These two bottlenecks can, to a degree, be overcome with microwave engineering. For instance, the linewidth can be made an order of magnitude wider by altering the impedance along the resonator. This is called a stepped-impedance transformer, where the impedance is ramped down from a matched 50 \(\Omega\) at the capacitor down to a small impedance at the SQUID\(^{382}\) shorting the device to the ground. Also, the saturation power can be increased by distributing the nonlinearity across an array consisting of many identical junctions, reducing the Kerr-nonlinearity by a factor \(1/N^2\) with \(N\) representing the number of junctions in the array. This has been demonstrated by using an array of SQUIDs in a resonator, rather than a single one.\(^{352}\)

However, despite the above-mentioned engineering efforts to improve the resonator-based JPAs, the most prominent approach to date is to get rid of the resonator altogether and, instead, construct a microwave analog to optical parametric amplifiers, where kilometers of weakly nonlinear fibers are used. Such a device is called a “traveling wave parametric amplifier” (TWPA) and was developed to surmount the bandwidth and dynamic range limitations of the resonator-based JPAs.

Although operated in similar way, the nonlinearity of TWPAs can be realized in different ways, such as the kinetic inductance of a superconducting film or using an array of Josephson junctions, through which the four-wave mixing process is distributed across a nonlinear lumped element transmission line, see Fig. 28(a).

The Josephson TWPA consists of a few thousand identical unit cells, each comprising a shunt capacitor to ground and a nonlinear Josephson inductor, together yielding a characteristic impedance of \(Z_0 = \sqrt{L_j/C} \approx 50\) \(\Omega\), see Fig. 28(a).

The fact that the nonlinearity is distributed allows for a high saturation power, since each Josephson junction is accessed once. However, even though energy conservation is satisfied, the four-wave mixing process in the device, there is a problem with phase (or momentum) conservation. This is associated with the system nonlinearity as well as the large frequency detuning between the signal and pump photons, yielding a difference in phase-velocity between the two, which in turn gives rise to a nonflat gain profile, as well as an overall reduction in gain.\(^{380}\)

Again, by taking inspiration from the dispersive engineering developed in quantum optics and photonics, where the refractive index can be periodically altered to engineer the momentum of a transferred signal, the solution to this phase-mismatch problem was introduced by O’Brien et al.\(^{386}\) By introducing resonators at periodic
intervals of TWPA unit cells, the pump tone can be given a “momentum kick,” effectively slowing it down and phase-matching the device by means of its wave vector. This technique is called resonant-phase matching (RPM), see Fig. 28(d), and requires that the pump frequency is set on the left side of the dispersion feature (where the wave vector diverges), defined by the resonant frequency of the phase-matching resonators. Note, finally, that broadband parametric amplification with a high dynamic range has been demonstrated in other Josephson-based circuits, e.g., the superconducting nonlinear asymmetric inductive element (SNAIL) parametric amplifier (SPA). 388–390

VI. SUMMARY AND OUTLOOK

In this review, we have discussed the phenomenal progress over the last decade in the engineering of superconducting devices, the development of high-fidelity gate-operations, and quantum nondemolition measurements with a high signal-to-noise ratio. Putting these advances together, we hope that it is clear that the planar superconducting qubit modality is a promising platform for realizing near-term medium scale quantum processors. While we have focused on highlighting the advances made within the fields of realizing, controlling, and reading out planar superconducting qubits specifically used for quantum information processing, there has of course also been tremendous activity in the surrounding fields. In this final section, we brieﬂy mention a few of those ﬁelds, and invite the reader to look into the references, for further details.

A. Quantum annealing

Superconducting qubits also form the basis for certain quantum annealing platforms. 388–390 Quantum annealing operates by ﬁnding the ground state of a given Hamiltonian (typically a classical Ising Hamiltonian), and this state will correspond to the solution of an optimization problem. By utilizing a ﬂux-qubit type design (see Sec. II), the company D-Wave have demonstrated quantum annealing processors 386 which have now reached beyond 2000 qubits. 391 The benchmarking of quantum annealers and attempts to demonstrate a quantum speedup for a general class of problems is a highly active research ﬁeld, and we refer the reader, for example, to recent papers Refs. 392–394 and references therein.

B. Cavity based quantum information processing

A parallel effort to the planar superconducting qubits discussed in this review is the development of 3D cavity-based superconducting qubits. In these systems, quantum information is encoded in superpositions of coherent photonic modes of the cavity. 395 The cat states can be highly coherent due to the inherently high quality factors associated with 3D cavities. 396 This approach has a fairly small hardware overhead to encode a logical qubit, 397 and lends itself to certain implementations of asymmetric error-correcting codes due to the fact that errors due to single-photon loss in the cavity are a tractable observable to decode. Using this architecture, several important advances were recently demonstrated including extending the lifetime of an error-corrected qubit beyond its constituent parts, 398 randomized benchmarking of logical operations, 397 a CNOT gate between two logical qubits 398 as well as Ramsey interference of an encoded quantum error corrected qubit. 399

C. Cryogenics and software development

While the qubit lifetimes and gate fidelity have improved dramatically in the last decades, there remains a need for error correction to reach large-scale processors. While certain strategies exist to extend the computational reach of current state-of-the-art physical qubits, 311 for truly large-scale algorithms addressing practical problems, the quantum data will have to be embedded in an error-correcting scheme. As brieﬂy mentioned in Secs. IV F 2 and IV G 2, certain components of the surface code quantum error correcting scheme have already been demonstrated in superconducting qubits (see, e.g., Refs. 66, 251, and 277). However, the demonstration of a logical qubit with greater lifetime than the underlying physical qubits, remains an outstanding challenge. While the surface code is a promising quantum error correcting code due to its relatively lenient fault tolerance threshold, it cannot implement a “universal” gate set in a fault-tolerant manner. This means that the error-corrected gates in the surface code need to be supplemented, for example, with a T gate, to become universal. Such gates can be implemented by a technique known as “magic state distillation.” 420 The process of “gate-teleportation,” a precursor to magic state distillation, has already been demonstrated using FPGA-based classical feedback with planar superconducting qubits, 185 but showing distillation and injection into a surface code logical state
The basic idea is to demonstrate a calculation, using qubits and algorithmic gates, which is outside the scope of classical computers (assuming some plausible computational complexity conjectures). For a recent review article, the reader is referred to Ref. 424. A first step toward demonstrating quantum supremacy was recently reported, using 9 tunable transmons. It is expected that with somewhere between 50 and 100 qubits, an extension of the protocol from Refs. 262 and 426, will allow researchers to sample from a classically intractable distribution, and thereby demonstrate quantum computational supremacy. The success of this program would constitute a phenomenal result for all of quantum computing.

**ACKNOWLEDGMENTS**

The authors gratefully acknowledge Mollie Kimchi-Schwartz, Jochen Braumüller, and Niels-Jakob Seo Loft for the careful reading of the manuscript and Youngkyu Sung for the use of his time-extension of the protocol from Refs. 262 and 426, will allow researchers to sample from a classically intractable distribution, and thereby demonstrate quantum computational supremacy. The success of this program would constitute a phenomenal result for all of quantum computing.

**REFERENCES**


DiCarlo, “Detecting bit-flip errors in a logical qubit using stabilizer measurements,”


P. Krantz, “Parametrically Pumped Superconducting Circuits (Chalmers University of Technology, 2013).


doi: 10.1063/1.5089550


See https://github.com/Microsoft/Quantum for "The Microsoft Quantum Development kit."


Even though linear resonant systems cannot be addressed properly, their long coherence times have proven them useful as quantum access memories for storing quantum information, where a nonlinear ancilla system is used as a quantum controller for feeding and extracting excitations to/from the resonant cavity modes.

The critical current is the maximum supercurrent that the junction can support before it switches to the resistive state with nonzero voltage.

Starting from a generic qubit Hamiltonian, \( H_0 = E_0 |0\rangle \langle 0| + E_1 |1\rangle \langle 1| \), we can rewrite in terms of Pauli matrices, and get \( H_0 = ((E_0 + E_1)/2)I - (E_1 - E_0)/2\sigma_z \). In the main text, we have ignored the constant offset term.

In reality, superconducting qubits, just like natural atoms, have higher energy levels. These higher levels are outside of the computational subspace, but need to be taken into account for most qubit simulations to get accurate predictions of frequency shifts.

This commutation is approximate and has an asymptotic dependence on the qubit-resonator detuning.

It has been demonstrated that it is still possible to read out the qubit state by applying a very strong resonator drive tone, even though this readout scheme is not QND.

It is worth mentioning that the observed qubit frequency is always the Lamb-shifted frequency and not the bare qubit frequency.

When analyzing the readout data, we have the freedom to choose projections.