This is the accepted manuscript made available via CHORUS. The article has been published as:

Parametrically Activated Entangling Gates Using Transmon Qubits

S. A. Caldwell et al.

Phys. Rev. Applied 10, 034050 — Published 24 September 2018
DOI: 10.1103/PhysRevApplied.10.034050
A central challenge in building a scalable quantum computer with superconducting qubits is the execution of high-fidelity, two-qubit gates within an architecture containing many resonant elements. As more elements are added, or as the multiplicity of couplings between elements is increased, the frequency space of the design becomes crowded and device performance suffers. In architectures composed of transmon qubits [1], there are two main approaches to implementing two-qubit gates. The first utilizes fixed-frequency qubits with static couplings where the two-qubit operations are activated by applying transverse microwave drives [2–8]. While fixed-frequency qubits generally have long coherence times, this architecture requires satisfying stringent constraints on qubit frequencies and anharmonicities [5, 6, 8] which requires some tunability to scale to many qubits [9]. The second approach relies on frequency-tunable transmons, and two-qubit gates are activated by tuning qubits into and out of resonance with a particular transition [10–16]. However, tunability comes at the cost of additional decoherence channels, thus significantly limiting coherence times [17]. In this approach the delivery of shaped unbalanced control signals poses a challenge [15]. Such gates are furthermore sensitive to frequency crowding—avoiding unwanted crossings with neighboring qubit energy levels during gate operations limits the flexibility and connectivity of the architecture.

An alternative to these approaches is to modulate a circuit’s couplings or energy levels at a frequency corresponding to the detuning between particular energy levels of interest [18–26]. This enables an entangling gate between a qubit and a single resonator [21, 22], a qubit and many resonator modes [26], two transmon qubits coupled by a tunable mediating qubit [16, 25], or two tunable transmons coupled to a mediating resonator [23, 24].

Building on these earlier results, we implement two entangling gates, iSWAP and controlled Z (CZ), between a flux-tunable transmon and a fixed-frequency transmon. The gates are activated by modulating the tunable qubit in resonance with particular pairs of transmon states: |10⟩ and |01⟩ for iSWAP and |11⟩ and |20⟩ (or |02⟩) for CZ. This direct qubit-qubit interaction provides each maximally entangling operation in a single step, without sacrificing any qubits to a mediating role. The interaction is a first-order process with an effective strength of roughly half the qubit-qubit coupling, and the resonance conditions depend on both the frequency and amplitude of modulation. The availability of a second dimension in the control space, combined with the selectivity of the resonances, mitigates the effect of frequency crowding in architectures where each qubit has many neighbors. The modulation amplitude can be used to vary the frequency spectrum of the resonances, and up to three entangling gates, per neighbor, can be realized at each amplitude. The ability to apply both iSWAP and CZ between pairs of qubits may reduce circuit complexity, as these operations are not locally equivalent. However, the iSWAP is locally equivalent to a CNOT followed by the interchange of the qubit labels [27]. Both the iSWAP and the CZ gates are universal for quantum computation when combined with local operations on the individual qubits [27, 28].

Our experimental results were obtained with the circuit shown in Fig. 1(a). This circuit comprises a fixed-frequency transmon (F) and a tunable transmon (T), coupled capacitively with a strength g. The tunable transmon, based on an asymmetric SQUID, has a first
transition frequency $\omega_T$ that depends on the externally applied flux bias $\Phi$ as shown in Figure 1(e). We modulate $\omega_T$ by applying a flux bias of the form

$$\Phi(t) = \Phi_0 + \tilde{\Phi} \cos(\omega_p t + \theta_p),$$

with static bias $\Phi_0$, modulation amplitude $\tilde{\Phi}$, frequency $\omega_p$ and phase $\theta_p$. An important special case occurs when the parking flux $\Phi_0$ is set at a turning point in $\omega_T(\Phi)$, where the frequency modulation takes the approximate form

$$\omega_T(t) \approx \tilde{\omega}_T(\tilde{\Phi}) + \omega_T(\tilde{\Phi}) \cos(2\omega_p t + 2\theta_p)$$

for $\tilde{\Phi} \ll \Phi_0/2$. The factors of 2 in Equation 2 arise because, at these turning points, the transmon frequency undergoes two cycles for each cycle of the flux. The nonlinearity of $\omega_T(\tilde{\Phi})$ results in the time-averaged frequency $\tilde{\omega}_T$ being shifted away from $\omega_T(\tilde{\Phi})$ by an amount that depends on $\tilde{\Phi}$. Our experimental results were obtained with the tunable transmon parked at $\tilde{\Phi} = 0$, as depicted in Fig. 1(e). This condition offers first-order suppression of decoherence due to flux noise while idling [29].

The effect of the parametric modulation can be seen in an interaction frame defined by the instantaneous qubit frequencies. Indexing the two-transmon excited states as $|F\rangle$ and approximating the (defined positive) transmon anharmonicities $\eta_F$ and $\eta_T$ as constant, we write the interaction Hamiltonian

$$\dot{H}_{\text{int}} = g \sum_{n=\infty} J_n \left( \frac{\omega_T}{2\omega_p} \right) e^{i(2n\omega_p t + \beta_n)}$$

$$\times \left\{ e^{-i\Delta t} |10\rangle \langle 01| + \sqrt{2} e^{-i(\Delta + \eta_T)t} |20\rangle \langle 11| + \sqrt{2} e^{-i(\Delta - \eta_T)t} |11\rangle \langle 02| + \text{h.c.} \right\},$$

(3)

where $J_n$ is the $n^{th}$ Bessel function of the first kind and where we denote the detuning and phase

$$\Delta = \tilde{\omega}_T(\tilde{\Phi}) - \omega_F$$

$$\beta_n = (\tilde{\omega}_T/2\omega_p) \sin(2\theta_p) + (2\theta_p + \pi)n,$$

(4)

(5)

The interaction Hamiltonian makes plain the resonance conditions

$$2n\omega_p = \Delta(\tilde{\Phi})$$

$$10 \leftrightarrow |01\rangle \text{ iSWAP}$$

$$2n\omega_p = \Delta(\tilde{\Phi}) - \eta_T$$

$$|11\rangle \leftrightarrow |02\rangle \text{ CZ}_{02}$$

$$2n\omega_p = \Delta(\tilde{\Phi}) + \eta_T$$

$$|11\rangle \leftrightarrow |20\rangle \text{ CZ}_{20}$$

(6)

(7)

(8)

with harmonics $n = \pm 1, \pm 2, \pm 3, \ldots$. These three types of entangling interactions are thus available at a series of amplitude-dependent modulation frequencies corresponding to the level spacings in the driven two-qubit system. Each interaction has an effective coupling strength $g_{\text{eff}}^{(n)}$ that determines the Rabi frequency and resonant linewidth of the interaction at the $n^{th}$ harmonic. This is given by the time-independent prefactor of each term in the Hamiltonian, so that

$$g_{\text{eff}}^{(n)} = g J_n \left( \frac{\omega_T}{2\omega_p} \right)$$

$$\text{iSWAP}$$

$$g_{\text{eff}}^{(n)} = \sqrt{2} g J_n \left( \frac{\omega_T}{2\omega_p} \right)$$

$$\text{CZ}.$$  

(9)

(10)

The flux drive introduces phase shifts $\beta_n$ to the Hamiltonian, which depend on both the amplitude and phase of modulation. The Rabi frequency is maximal, and the gate time minimal, where $J_n$ is maximal. For the first harmonic this occurs for $\tilde{\omega}_T/2\omega_p \approx 1.84$, giving $J_1 \approx 0.582$.

The interaction Hamiltonian shown above is derived under a rotating wave approximation, neglecting couplings between highly detuned levels. This approximation is valid for modulation frequencies $2\omega_p$ and couplings $g$ well below the qubit frequencies. We also neglected the modulation of the anharmonicity and have
kept only the leading term in the coupling strengths between the different transitions. Taking these effects into account leads to small adjustments in the effective coupling rate for the range of modulation amplitudes used in this work. A more comprehensive analysis of this interaction is presented in Ref. [30].

A parametric drive that resonantly couples two levels produces swapping in the subspace of those levels, described by the unitary

$$\hat{U} = \begin{pmatrix} \cos(\theta/2) & -ie^{-i\phi}\sin(\theta/2) \\ -ie^{i\phi}\sin(\theta/2) & \cos(\theta/2) \end{pmatrix}, \quad (11)$$

with the population exchange given by $\theta = 2 \int_0^\tau g_{\text{eff}}(t) \, dt$ during the flux pulse, and the phase $\phi$ of the exchange given by $\beta_n$. The $i$SWAP gate, mapping $|10\rangle + \gamma|11\rangle - \delta|01\rangle + \alpha|00\rangle \rightarrow \alpha|00\rangle + \gamma|01\rangle + \delta|10\rangle + \beta_n|11\rangle$, is realized by selecting a modulation amplitude and frequency to satisfy the $i$SWAP condition (Equation 6) and pulsing the modulation for a time $\tau$ that yields $\theta = \pi$. The CZ gates, mapping $|00\rangle + \beta|01\rangle + \gamma|10\rangle + \delta|11\rangle \rightarrow |00\rangle + \beta|01\rangle + \gamma|10\rangle - \delta|11\rangle$, are realized by activating one of the CZ resonances (Equations 7 and 8) for a time that yields $\theta = 2\pi$. Physically, this causes the $|11\rangle$ state to fully leave and return to the computational basis, picking up the $SU(2)$ phase of $-1$ in the process. Corrections of local phases, necessary to realize the correct unitary for each gate, are discussed below.

I. EXPERIMENTAL RESULTS

We implemented this proposal on a two-transmon device fabricated on a high-resistivity (>10 kΩ·cm) silicon substrate. In successive steps, Ti/Pd alignment marks and large Al features (ground planes, signal lines, and compact LC readout resonators) were defined using photolithography, electron-beam deposition, and liftoff. The transmon qubits, including Josephson junctions, were fabricated using the controlled undercut technique [31]. A bilayer resist stack consisting of MMA (8.5) MAA copolymer and PMMA 950 was written with a 100-kV electron-beam lithography tool, and Al films were deposited in a Plassys MEB 550S at opposite substrate angles with an intervening controlled oxidation to form tunnel junctions.

The transmons were capacitively coupled to each other with a strength of $g/2\pi = 6.3$ MHz, and to separate linear readout resonators with strengths of $\approx 70$ MHz. The microwave $XY$ control signals were delivered through the readout resonators, and the $Z$ control of the tunable transmon was delivered on a single-ended, on-chip flux-bias line. All RF signals used for qubit control and readout were transmitted (Tx) and received (Rx) by the Ettsus Research USRP X300 software-defined radio platform, modified with custom gateware. A schematic of the experiment is given in Fig. 6. The tunable transmon frequencies were $\omega_p^{\text{max}}/2\pi = 4582$ MHz, $\eta_p^{\text{max}}/2\pi = 173$ MHz, $\omega_p^{\text{min}}/2\pi = 3285$ MHz, $\eta_p^{\text{min}}/2\pi = 185$ MHz. The fixed transmon frequencies were $\omega_F/2\pi = 3940$ MHz.
and $\eta_F/2\pi = 180$ MHz.

The chip was operated with the tunable transmon parked at $\omega_p^{\text{max}}$, where its relaxation and coherence times were $T_1 = 13 \mu s$ and $T_2^* = 10 \mu s$. The fixed transmon had $T_1 = 34 \mu s$ and $T_2^* = 20 \mu s$. Single-shot readout assignment fidelities of 0.85 (fixed) and 0.92 (tunable) were achieved with dispersive readouts combined with a binary classifier trained on $|0\rangle$ and $|1\rangle$ state preparation for each qubit [32]. This classifier was used to classify the simultaneous readout signals obtained in the two-qubit tomography and randomized benchmarking results described below. Single-qubit randomized benchmarking decay constants of greater than 0.98 were obtained. Application of the flux modulation increased the tunable transmon’s sensitivity to flux noise, such that we measured $T_2^* \approx 4 \mu s$ with the drive applied. In principle the flux modulation may also increase the qubit’s relaxation rate [15, 33], but we found that $T_1$ was consistent with the flux drive on versus off.

As illustrated in Fig. 2, we observed the resonance conditions for the $i$SWAP and CZ interactions as predicted by theory and numerical simulations. Our simulations were performed by evolving the Lindblad master equation with the full Hamiltonian of the circuit-QED system under flux modulation. The master equation included dephasing and dissipation consistent with the measured values of $T_1$ and $T_2^*$ under modulation, reported in Table 1. To obtain this concordance we measured the transfer function of flux pulses to the tunable transmon by measuring the average qubit frequency $\tilde{\omega}_p$ during modulation, as a function of both amplitude and frequency over the nominal 0–500 MHz passband of the signal chain. For data-taking we then applied a frequency-dependent amplitude correction, similar to [26], to flatten the frequency response.

Closer inspection of the resonant features also showed good agreement between theory and observations of a characteristic “chevron” signal (see Fig. 3). This signal was used to confirm the resonant frequency at the chosen value of modulation amplitude. The parametric $i$SWAP and CZ gates were then realized by determining which pulse durations $\tau$ yielded the appropriate values of $\theta$.

Once a fully entangling interaction is specified, realizing the desired unitary for each gate requires a number of corrections to be applied.

1. The mean qubit frequencies are shifted during the modulation and thus acquire local phases with respect to their microwave drives. These local phases were measured with a Ramsey experiment where only the phase of the second $\pi/2$ pulse was varied. The corrections for the phases were then implemented as local updates to the phases of downstream microwave pulses. The mean shifts depend in general on the initial phase of each flux pulse; however, this effect can be suppressed by using risetimes much slower than the modulation frequency. For this reason we used risetimes of 30–40 ns.

2. The frequency of the flux pulse must be chosen according to the resonant condition (Equation 4), which is affected by the shifts in qubit frequencies. However, when the flux pulse is not applied, the dynamical phase between the two levels accumulates at the unshifted detuning of the two levels. Because the flux pulses were generated using a local oscillator tuned to the shifted frequency, the phases of successive pulses required correction for the mean shift of the qubit frequencies.

3. The phase $\beta_p$ is present in the $i$SWAP interaction, where the off-diagonal elements of Equation 11 are non-zero, and is set by the phase $\theta_p$ of flux modulation with respect to the individual qubit frames. This interaction phase can be handled either by updating the phase of the flux drive or pushed to equal and opposite local frame changes applied after the gate.

4. The exchange of states in the $i$SWAP gate entails the exchange of their local phases. This was tracked by swapping the microwave phases after each $i$SWAP gate.

![FIG. 3. A zoomed-in view of resonant features, similar to those shown in Fig. 2, reveals a “chevron” pattern characteristic of Rabi oscillations. For each type of gate, a modulation amplitude was chosen to provide a suitable $g_{\text{eff}}$. With that amplitude held constant, the flux-pulse duration and frequency were varied and values were chosen to give the roughly the desired gate up to local phases. The patterns observed in the experimental data (left) were in good agreement with the predictions of a dynamical simulation (right), as shown here for transitions $|01\rangle \leftrightarrow |10\rangle$ at $\tilde{\Phi} = 0.318\Phi_0$ (bottom) and $|11\rangle \leftrightarrow |20\rangle$ at $\tilde{\Phi} = 0.325\Phi_0$ (top). The only free parameters in the model are the effective relaxation and coherence times in the presence of flux drives.](image-url)
Characteristics of each gate are given in Table I. The average gate fidelity of each operation was estimated using interleaved randomized benchmarking (IRB) [34, 35] as well as maximum-likelihood quantum process tomography (QPT) [36]. In using QPT to estimate the gate fidelities [37, 38] we compensated for the known readout infidelities [32, 39] and applied complete-positivity (CP) constraints. We estimate $F_{\text{SWAP}} = 0.93$ and $F_{\text{CZ}} = 0.92$ for both CZ$_{02}$ and CZ$_{20}$. The constrained reconstructions, along with the ideal target gates, are depicted in Fig. 5. Similar devices have also yielded average fidelities for CZ above 0.9 [32]. The result of the IRB experiment for the iSWAP gate is shown in Fig. 4.

In order to roughly identify the source of the observed infidelities, we estimate the fidelity to the unitary closest to the reconstructed evolution. If the fidelity to the closest unitary is similar to the fidelity to the ideal target gate, we conclude there were no significant coherent errors, and that the fidelity is limited by the decoherence of the evolution. Otherwise, we conclude there are significant coherent errors in the evolutions. There are two natural notions of closest unitary for a reconstructed superoperator $\mathcal{E}$ with corresponding canonical Kraus representation $\mathcal{E}(\rho) = \sum_i K_i \rho K_i^\dagger$ [40] and with Liouville representation singular value decomposition $\mathcal{E} = U S V^\dagger$. One is given by the unitary $U_0$ where $\|K_0\| \geq \|K_{i \neq 0}\|$, and where $K_0 = U_0 P_0$ is a polar decomposition [41], which results in the approximate interferometric bound

$$F \lesssim \frac{\text{tr} \left( (U_0^* \otimes U_0)^\dagger \mathcal{E} + d \right)}{d^2 + d}.$$  

(12)

where $d = 4$ is the dimension of the Hilbert space of interest. The other is given by $\mathcal{U} V^\dagger$, which results in the strict Procrustean unitarity bound

$$\tilde{F} \leq \frac{\text{tr} \ S + d}{d^2 + d}.$$  

(13)

These upper bounds were consistent with our measured fidelity estimates, indicating that infidelities are largely explained by the decoherence during the gate evolution.

\textbf{a. Summary} — We have proposed and demonstrated parametric iSWAP and CZ gates activated by modulating the frequency of a tunable transmon. We have shown close agreement between observed and predicted resonance conditions over a wide range of frequencies and multiple drive amplitudes, and we have described and implemented local phase corrections necessary to produce correct and repeatable unitaries. Gate fidelities, estimated by interleaved randomized benchmarking and quantum process tomography, were estimated to be 0.91-0.94 and appeared to be limited by decoherence during the parametric drive. The underlying interactions used in this technique are highly selective and provide a way to mitigate frequency crowding in a scalable quantum processor architecture.

\textbf{ACKNOWLEDGMENTS}

Part of this work was performed at the Stanford Nano Shared Facilities (SNSF), supported by the National Science Foundation under award ECCS-1542152.
FIG. 5. Two-qubit process tomography parametric CZ (top) and iSWAP (bottom) processes. The corresponding average process fidelities are $F = 0.92$ for both CZ’s, and $F = 0.93$ for the iSWAP.

FIG. 6. Schematic of signal delivery to the two-qubit chip studied in this paper. Qubit drive and readout tones were transmitted, and the readout tone received, by Ettus Research USRP modules running with X300 daughterboards. The combined slow- and fast-flux components were generated by a Yokogawa GS200 DC power supply and another USRP X300, attenuated by 20dB at 4K and delivered through a UHF low-pass filter and a custom, dissipative Eccosorb filter. The on-chip flux-bias line terminated in an inductive short to ground.


