

# Fast superconducting qubit control with sub-harmonic drives

Mingkang Xia,<sup>1</sup> Chao Zhou,<sup>1</sup> Chenxu Liu,<sup>2,3,\*</sup> Param Patel,<sup>1</sup> Xi Cao,<sup>1</sup> Pinlei Lu,<sup>1</sup> Boris Mesits,<sup>1</sup> Maria Mucci,<sup>1</sup> David Gorski,<sup>1</sup> David Pekker,<sup>1</sup> and Michael Hatridge<sup>1</sup>

<sup>1</sup>*Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, PA 15213, USA*

<sup>2</sup>*Department of Physics, Virginia Tech, Blacksburg, Virginia 24061, USA*

<sup>3</sup>*Virginia Tech Center for Quantum Information Science and Engineering, Blacksburg, VA 24061, USA*

(Dated: June 21, 2023)

Increasing the fidelity of single-qubit gates requires a combination of faster pulses and increased qubit coherence. However, with resonant qubit drive via a capacitively coupled port, these two objectives are mutually contradictory, as higher qubit quality factor requires a weaker coupling, necessitating longer pulses for the same applied power. Increasing drive power, on the other hand, can heat the qubit's environment and degrade coherence. In this work, by using the inherent non-linearity of the transmon qubit, we circumvent this issue by introducing a new parametric driving scheme to perform single-qubit control. Specifically, we achieve rapid gate speed by pumping the transmon's native Kerr term at approximately one third of the qubit's resonant frequency. Given that transmons typically operate within a fairly narrow range of anharmonicity, this technique is applicable to all transmons. In both theory and experiment, we show that the Rabi rate of the process is proportional to applied drive amplitude cubed, allowing for rapid gate speed with only modest increases in applied power. In addition, we demonstrate that filtering can be used to protect the qubit's coherence while performing rapid gates, and present theoretical calculations indicating that decay due to multi-photon losses, even in very strongly coupled drive lines, will not limit qubit lifetime. We demonstrate  $\pi/2$  pulses as short as tens of nanoseconds with fidelity as high as 99.7%, limited by the modest coherence of our transmon. We also present calculations indicating that this technique could reduce cryostat heating for fast gates, a vital requirement for large-scale quantum computers.

## I. INTRODUCTION

High fidelity single qubit control is one of the fundamental requirements for gate-based quantum computing. While many factors can limit quantum gate fidelity, such as cross-talk [1, 2] and leakage to non-computational states [3], the most fundamental are gate speed and qubit coherence [4–8], with recent improvements driven more by increased coherence than enhanced speed [9, 10]. However, for superconducting qubits, single qubit gates typically use resonant driving of a qubit transition, in which the requirements for fast gates and coherent qubits are often contradictory.

For instance, to protect the qubit coherence time, weakly coupled drive ports and heavily attenuated lines are typically used to suppress qubit photon leakage, thermal noise [10, 11], and out of band radiation [12]. As qubit coherence increases, there is a trend towards more thorough and careful filtering. The increased losses of filtering and weaker qubit-drive line couplings together result in longer gate times at a given drive strength. However, achieving high gate fidelity requires fast gate speeds and so we attempt to compensate by increasing drive strength, which in turn increases heating of the cryostat. The heating of filter elements and drive lines [13] can degrade the qubit by either creating excess thermal population at the qubit transition or higher frequencies

(for instance at the superconducting gap) and/or heating the entire cryostat by exceeding the cooling capacity of the dilution refrigerator [14]. Moreover, the thermal time constants (in the millisecond range) are much slower than the pulses (roughly 10s of nanoseconds range) or qubit coherence times ( $\sim 100\mu s$  to 1 ms), and can produce effects which accumulate and persist across many experiments [13]. This situation is exacerbated by the necessity of scaling to larger quantum machines and thus increased qubit count and associated heating [7].

To combat these limitations, better heat handling of each component, especially the attenuators on the control lines, has proven useful [11, 13]. Another solution is to break the symmetry between input control drives and out-going qubit photon decay in the power domain using a non-linear filter [15]. In this paper, we propose to break the link between qubit coherence and driving by separating the two processes in frequency space, proposing and demonstrating a new single qubit control scheme based on parametric driving. Far off-resonant controls are widely used in parametric amplification [16, 17] and parametric qubit-cavity [18], multi-cavity [19–21] and multi-qubit gates [22, 23]. In each of these scenarios, far off-resonant terms in the Hamiltonian are utilized, consuming pump photons to produce effective lower-order interaction Hamiltonians which need not satisfy energy conservation. The benefits of parametric driving include strong interactions and high on/off ratios; the use of reflective filters to protect mode frequencies while allowing strong driving is also well established [21, 24, 25].

In this work, we introduce a single-qubit paramet-

\* Recently moved to: Pacific Northwest National Laboratory, Richland, WA 99354, USA

ric driving scheme, which we term sub-harmonic driving. In sub-harmonic driving we operate a transmon, a commonly used fixed-frequency qubit, by parametrically driving it at an integer fraction of its  $|g\rangle \leftrightarrow |e\rangle$  transition frequency. The dominant nonlinear term in a transmon, the self-Kerr, is generated from the fourth order term of the Josephson junction's sinusoidal potential, with magnitude typically ranging from 150-250 MHz [26]. However, the potential energy also contains fast-rotating terms that are normally suppressed in the rotating wave approximation. We pick the 4<sup>th</sup> order term  $\hat{q}^\dagger \hat{q} \hat{q} \hat{q} + h.c.$ , where  $\hat{q}^\dagger$  is the annihilation operator of qubit, and parametrically drive it near one third the qubit transition frequency. Three drive photons are consumed to create one qubit photon, generating an effective Hamiltonian  $\varepsilon^3 a^\dagger + h.c.$ , where  $\varepsilon$  is the drive treated as a stiff pump. We emphasize that the term which powers this interaction is both ubiquitous and of very similar strength in every transmon qubit. This scheme can also be adapted to other qubits and systems by choosing an appropriate source of nonlinearity, for instance 3<sup>rd</sup> or 5<sup>th</sup> order terms in qubits with asymmetric Josephson elements [27], and thus can be applied widely in superconducting circuits.

This control scheme has two main advantages. The first advantage is that the drive frequency is separated from the qubit transition frequency. It allows us to suppress a primary relaxation channel in the system without affecting the ability to control the qubit by engineering the impedance at two widely separated frequencies. To this end, we place a reflective low-pass filter (LPF) with very low absorption and good isolation in the stop band at the qubit drive port, as shown in Figure 1(a). This suppresses qubit photon leakage to the environment through resonant decay while allowing low-frequency drive photons to pass freely. Therefore, the drive port can be more strongly coupled to the qubit for fast control without increasing the qubit's direct relaxation rate. The second advantage of sub-harmonic control is that the Rabi rate is proportional to the drive amplitude cubed because it is a three-photon driving process, which allows us to rapidly improve qubit gate speed with only a moderate increase of drive power.

In this work, we experimentally demonstrate the concept of single qubit, sub-harmonic control and achieve gates as fast as 35 ns with gate fidelities up to 99.7% on a typical, unoptimized transmon qubit. We also present calculations which address two key questions about practical implementation of sub-harmonic gates. First, we present a theoretical study of the effects of the low frequency lossy environment on the qubit's coherence, finding that this coupling offers a negligible decay channel even for very strong couplings. Second, based on both theoretical calculations and measured parameters of our system, we find that the combination of low loss at base with realistic filter losses should allow high fidelity qubit gates with lower heating than conventional direct qubit drives.

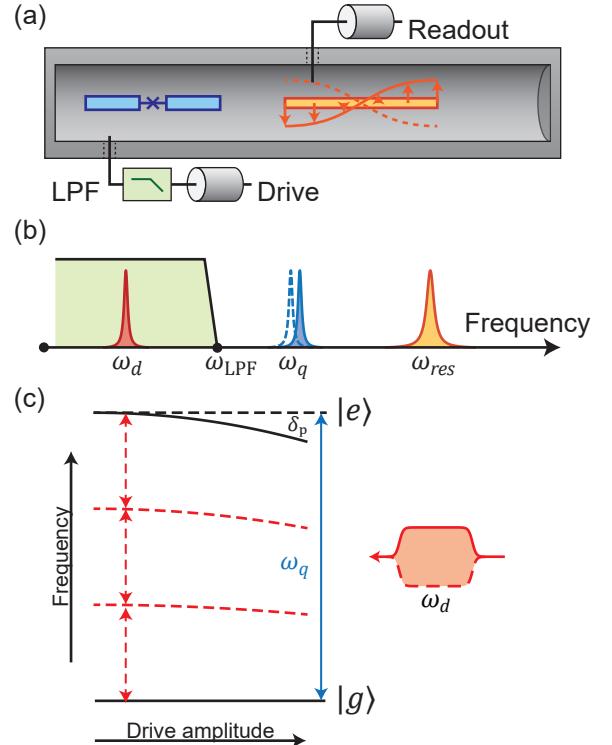


Figure 1. Sub-harmonic driving schematic. (a) A transmon qubit (at left in blue) and a  $\lambda/2$  resonator for readout (at right in orange) are supported on a sapphire chip inside an aluminum tube. A low-pass filter (in green) is placed at the qubit drive port to suppress qubit photon leakage to the environment while allowing fast single qubit control at low frequencies. (b) The frequency distribution of the sub-harmonic drive  $\omega_d$  (red), qubit frequency  $\omega_q$  (blue), and the readout resonator frequency  $\omega_{res}$  (orange). Only the sub-harmonic drive is within the pass-band of the LPF (green). The dashed, shifted peak of  $\omega_q$  represents the AC-Stark effect during the sub-harmonic drive. (c) The energy level diagram of sub-harmonic driving. The AC-Stark effect creates a detuning  $\delta_p$  between the qubit's un-driven and driven frames. Flat-top pulses with the drive frequency  $\omega_d = (\omega_q - \delta_p)/3$  (shown at right) are used to control the qubit state.

## II. SUB-HARMONIC DRIVING THEORY

To consider the simplest case for single qubit control, a transmon qubit driven by a far off resonant single microwave tone, the Hamiltonian of this system can be written as:

$$\hat{H}/\hbar = (\omega_q - \alpha)\hat{q}^\dagger \hat{q} + \frac{\alpha}{12}(\hat{q}^\dagger + \hat{q})^4 + \epsilon(t)(\hat{q}^\dagger + \hat{q}), \quad (1)$$

where  $\hat{q}$  ( $\hat{q}^\dagger$ ) is the photon annihilation (creation) operator,  $\omega_q$  the transmon  $|g\rangle \leftrightarrow |e\rangle$  frequency, and  $\alpha$  the (negative) anharmonicity parameter. The first two terms are the Hamiltonian of the transmon up to 4<sup>th</sup> order, which is a good approximation for analyzing the properties of sub-harmonic driving. Without loss of generality, the

drive strength  $\epsilon(t)$  can be written as:

$$\epsilon(t) = \begin{cases} \varepsilon(t)e^{-i\omega_d t} + \varepsilon^*(t)e^{i\omega_d t} & 0 < t < t_{\text{gate}} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

After transitioning to a displaced frame (see Supplement Sec. I A), the Hamiltonian can be written as

$$\hat{H}^D/\hbar = (\omega_q - \alpha)\hat{q}^\dagger \hat{q} + \frac{\alpha}{12}(\hat{q} + \eta e^{-i\omega_d t} + h.c.)^4, \quad (3)$$

where  $\eta = \frac{2\varepsilon(t)\omega'_q}{\omega_q^2 - \omega'_q{}^2}$  represents the drive strength. Expanding the Hamiltonian  $\hat{H}^D$ , we find numerous 4<sup>th</sup> order terms corresponding to different parametric processes of the transmon, which can be individually activated by driving at the correct frequencies. For example, driving the term  $4\eta e^{i\omega_d t}\hat{q}\hat{q}\hat{q} + h.c.$  at  $3(\omega_q + \alpha)$  activates the three photon transition  $|g\rangle \leftrightarrow |h\rangle$  (where  $|h\rangle$  is the third excited state), while the two-photon transition  $|g\rangle \leftrightarrow |f\rangle$ , which is commonly seen in the transmon spectroscopy (usually termed  $gf/2$ ), can be activated by driving the term  $6(\eta e^{i\omega_d t})^2\hat{q}\hat{q} + h.c.$  near  $(2\omega_q + \alpha)/2$ . For sub-harmonic driving, the terms we are interested in are  $4(\eta e^{i\omega_d t})^3\hat{q} + h.c.$  When moved to the rotating frame at  $3\omega_d \approx \omega_q$ , we acquire the desired single-qubit Rabi drive, as well as a term proportional to  $(\eta\eta^*)\hat{q}^\dagger q$ , which represents the AC-Stark effect during sub-harmonic drive.

In the end, the Hamiltonian has the form:

$$\hat{H}_{\text{sub}}^R/\hbar = (2\alpha|\eta|^2 - 3\delta)\hat{q}^\dagger \hat{q} + \frac{1}{2}\alpha\hat{q}^\dagger \hat{q}^\dagger \hat{q}\hat{q} + \frac{1}{3}\alpha(\eta^3\hat{q}^\dagger + \eta^{*3}\hat{q}), \quad (4)$$

with  $\delta = \omega_d - \frac{1}{3}\omega_q$ . Equation (4) shows two important properties of sub-harmonic driving: the Rabi rate of the process is proportional to  $|\eta|^3$ , and the AC-Stark shift during the sub-harmonic drive is proportional to  $|\eta|^2$ . The AC-Stark effect also adds a drive dependent phase on the qubit which needs to be considered and calibrated in all experiments, as the qubit frequency changes in response to the amplitude of the drive.

One potential concern about sub-harmonic driving is qubit photon decay through the drive line at low frequencies via multi-photon processes, which could limit coherence when we couple the transmon and drive line very strongly. We perform a detailed calculation in the Supplement Sec. I G and summarize key results here. For a transmon qubit coupled to a low-frequency lossy environment, the system-bath coupling strength is

$$\lambda(\nu) = \Theta(\nu) \frac{C_c}{\sqrt{C_r c}} \sqrt{\frac{\omega_q \nu}{2\pi v}}, \quad (5)$$

where  $C_c$  is the coupling capacitance between the qubit and the transmission line,  $C_r$  is the capacitance of the qubit,  $c$  is the characteristic capacitance of the transmission line, and  $v$  is the speed of light in the transmission line [28]. In our case, we put a cut-off (filter) function

$\Theta(\nu)$  to suppress the high-frequency system-bath coupling. Specifically, the filter function is modelled as

$$\Theta(\nu) = \begin{cases} 1 & \nu \leq \omega_q/3 + \vartheta \\ 0 & \nu > \omega_q/3 + \vartheta, \end{cases} \quad (6)$$

where  $\omega_q/3 + \vartheta$  is the bandwidth of the filter pass-band. The decay rate through a three photon decay process can then be written as

$$\Gamma_3 = \frac{243}{32\pi^2} \frac{\gamma_1^3 |\alpha|^2}{\omega_0^4} \left( \frac{\vartheta}{\omega_0} \right)^2 \quad (7)$$

where  $\gamma_3 = 2\pi\lambda^2$ . For a transmon system with typical properties, because  $\alpha$  and  $\gamma_3$  are both much smaller than  $\omega_q$ , the sub-harmonic decay rate  $\Gamma_3$  is orders of magnitude smaller than the decay rate through resonant decay via internal losses or residual coupling to drive ports. Therefore,  $\Gamma_3$  can be safely neglected even for very strong couplings.

### III. EXPERIMENTAL RESULTS

The sub-harmonic driving scheme was tested on a single qubit-resonator system, which had a transmon qubit and a  $\lambda/2$  stripline resonator [29] with parameters commonly used in the field. Specifically, a transmon qubit ( $\omega_{\text{ge}}/2\pi = 3.96$  GHz,  $\alpha/2\pi = -208$  MHz,  $T_1/T_{2R}/T_{2E} = 42/11/23$   $\mu$ s) and a  $\omega_{\text{res}}/2\pi = 6.73$  GHz readout resonator were housed inside a 3 mm diameter aluminum tube. The resonator's coupling rate is  $\kappa_{\text{ext}}/2\pi = 5.42$  MHz and the cross-Kerr between the qubit and resonator is  $\chi/2\pi = 0.80$  MHz. This system was chosen to represent a general transmon-cavity system without any special engineering or specific design requirements, demonstrating that the sub-harmonic driving scheme is widely applicable to transmon-based quantum processors.

To better understand the properties of sub-harmonic driving, we first performed Rabi experiments, driving the qubit between its ground and excited states. Here, we first fixed the drive strength and swept the drive frequency and drive pulse duration. As shown in Figure 2(a), when using flat-top pulses, the sub-harmonic Rabi oscillation had a pattern similar to the one observed when using a resonant drive. By fitting the data, we extracted both the AC-Stark shift induced by the off resonant drive on the qubit and the Rabi rate at the selected drive strength.

The experiment was then repeated with different room temperature amplitudes ranging from 0.05 V to 0.38 V. The voltage here was the output amplitude of the arbitrary waveform generator (AWG) (see Figure S7 for microwave configuration). Assuming the total insertion loss on the input line is frequency independent within the range of frequencies applied, this voltage is linearly proportional to the drive strength  $\eta$  on the qubit. As shown in Figure 2(b), with a drive amplitude of 0.05 V,

the Rabi rate was only 40 kHz. When the drive amplitude was increased by 7.6 times to 0.38 V, the Rabi rate reached 12.1 MHz, which is 300 times faster. The Rabi rate  $\Omega(V_d)$  and the AC-Stark shift  $\Delta f(V_d)$  as a function of drive strength  $V_d$  are fitted simultaneously with a single free parameter  $k$ , which relates the room temperature and cryogenic drive voltages:

$$\begin{cases} \Delta f(V_d) = 2\alpha(kV_d)^2 \\ \Omega(V_d) = \frac{2}{3}\alpha(kV_d)^3, \end{cases} \quad (8)$$

where the transmon self-Kerr  $\alpha$  is a known parameter measured using qubit spectroscopy. As shown in Figure 2(b), the data fits to  $k = -1.197 \text{ V}^{-1}$ , and corresponds well with the Hamiltonian in Eq. (4).

These Rabi experiments demonstrate that the qubit state can be controlled with sub-harmonic driving. To use this scheme in real quantum circuits, we also need to develop a procedure for calibrating a high-fidelity single-qubit gate with well-defined parameters. For on-resonance driving, both deterministic tune-up [30] and randomized benchmarking procedures have been well developed [5]. However, these procedures cannot be directly applied to sub-harmonic gate tune-up without first addressing the issue of drive induced phase correction. As shown in Eq. (4) and Figure 2, because of the AC-Stark shift, the qubit frame rotates at a different speed relative to generator/lab frame during driving, and effectively adds a phase shift to the qubit. When applying a sequence of multiple sub-harmonic gates, the relative phase of the qubit and microwave drive depends on the history of previously applied pulses and delays. To perform gates over the intended direction, we must compensate for these accumulated phase shifts.

In experiment, smoothed flat-top pulses are chosen as the profile of the gate because they are both time-optimal and can simplify gate tune-up procedure for  $X_\pi$  and  $X_{\pi/2}$  gates. Gates with other pulse profiles can be used in principle, but the tune-up procedure may be more complicated. A flat-top pulse can be separated into three parts: ramps up and down, the fixed amplitude top, and a waiting time until the next pulse is applied. Therefore, the drive induced phase shift can be separated into three parts respectively:  $\varphi_{\text{ramp}}$ ,  $\varphi_{\text{flat}}$  and  $\varphi_{\text{gap}}$ . In experiment, the generator frequency is set to the one third of the AC-Stark shifted qubit frequency so that it is resonant with the qubit during the flat-top portion of the pulses, leaving the phase  $\varphi_{\text{flat}}$  equal to zero. For the remaining phases we have:

$$\varphi_{\text{gate}} = \varphi_{\text{ramp}} + (\omega_q - 3\omega_d)t_{\text{gap}}. \quad (9)$$

For flat-top pulses with different durations, only the flat-top part of the pulse is changed and the ramp-up/down profile of the pulse is kept the same. As a result, calibrating  $\varphi_{\text{ramp}}$  of a pulse with one duration is sufficient for phase correction of any other pulse with same amplitude.

To perform full single qubit control, we need to tune up the  $\pi$  and  $\pi/2$  pulses over the  $X$  and  $Y$  axes. The first

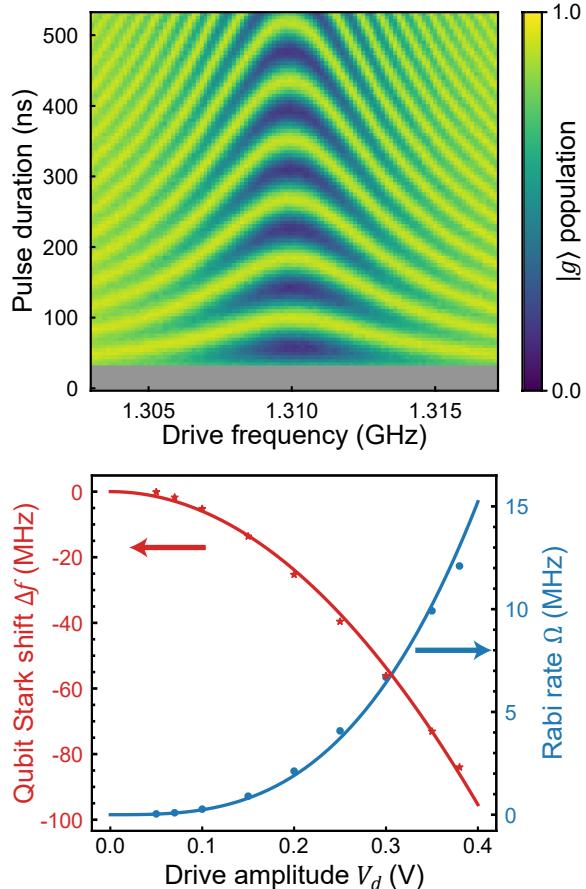


Figure 2. Rabi rate and AC-Stark effect vs. sub-harmonic driving strength. (a) A Rabi experiment for sub-harmonic driving of the qubit as a function of pulse duration and drive frequency at fixed drive amplitude of 0.38 V. The first 30ns of data was not measured because of equipment limitations. At this amplitude, the Rabi rate is 12.1 MHz and the drive-induced qubit AC-Stark shift is -81 MHz. (b) Rabi rate and qubit AC-Stark shift vs. drive amplitude. As expected from theory, the Rabi rate (blue circles) is proportional to drive amplitude cubed, while the AC-Stark shift (red stars) is negative and proportional to drive amplitude squared. A single, simultaneous fit describes both sets of data with only one free parameter, a scaling factor  $k$  converting AWG amplitude to effective driving strength.

step is choosing an optimum drive amplitude and corresponding frequency by Rabi experiment. This also gives a rough estimate of pulse duration. Because driving frequency and phase corrections are amplitude dependent, the drive amplitude is fixed in the later steps to simplify the tune-up procedure. The details of a gate sequence based calibration procedure that calibrates each parameter is explained in Supplement Sec. II C.

After tuning up all parameters, we can successfully implement  $X_\pi$ ,  $X_{\pi/2}$ ,  $Y_\pi$  and  $Y_{\pi/2}$  gates. Randomized benchmarking and interleaved randomized benchmarking are used to calibrate the average fidelity of all Clifford gates

and individual fidelity of specific gates, including  $X_\pi$ ,  $X_{\frac{\pi}{2}}$ ,  $Y_\pi$  and  $Y_{\frac{\pi}{2}}$  gates. In Figure 3, the blue data is the result of randomized benchmarking with sub-harmonic gates. The average Clifford gate fidelity (found by fitting the data to the blue fit curve using an exponential) is 98.6%. The gate fidelity of single Clifford gates is characterized using interleaved randomized benchmarking. The fidelity of the interleaved gate can be calculated using [31],

$$1 - F = \frac{1 - p_{\text{gate}}/p_{\text{ref}}}{2}. \quad (10)$$

We measured gate fidelity of  $X$ ,  $\sqrt{X}$ ,  $Y$  and  $\sqrt{Y}$  to be 99.4%, 99.7%, 99.5%, 99.7% respectively.

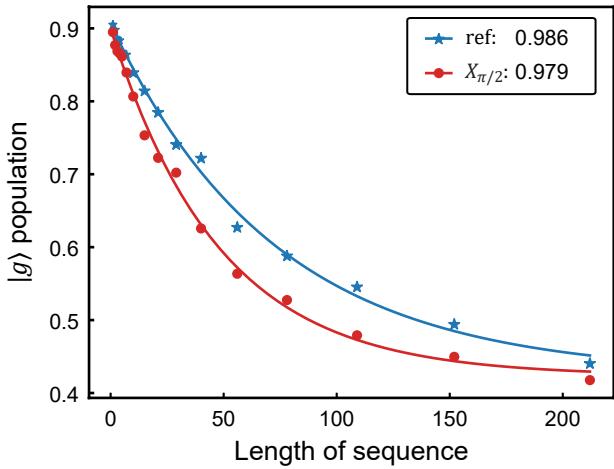


Figure 3. Gate fidelity is calibrated with interleaved randomized benchmarking. The blue trace shows the average fidelity of Clifford gates using sub-harmonic driving. The red trace shows the average fidelity of a particular interleaved Clifford gate ( $X_{\frac{\pi}{2}}$  here). We extract a fidelity of 99.7% for  $X_{\frac{\pi}{2}}$ .

The main limitation here is still the qubit coherence and our ability to generate chains of short pulses (30 ns is our minimum pulse block in our AWG). We note that by using an AWG with smaller minimum pulse durations, the gate speed could be tripled at the cost of dissipating approximately twice the power, while on-resonance driving will require 9 times more power. For faster and higher fidelity gates, the problem of leakage to higher excited states can be suppressed with a modified version of standard methods such as DRAG [3] as we detail in the Supplement Sec. I D.

Since the Rabi rate of the sub-harmonic gate is proportional to drive strength cubed, it also makes the gates more sensitive to the stability of the control electronics. In particular, we found that consistent, high-fidelity sub-harmonic gates required temperature stabilizing our drive circuit, as 1°C change in temperature caused about 1% change in output amplitude, corresponding to a  $\sim 3\%$  change in the gate's effective amplitude. Therefore, we designed and built a PID controlled temperature stabilizing box, which greatly improved the room-temperature

equipment stability (see Supplement Sec. II B). We note that sub-harmonic driving will always be a factor of three more sensitive to such drifts, but this factor does not scale with coherence time, fidelity, etc.

#### IV. POTENTIAL FOR REDUCED HEATING AND FAST SUB-HARMONIC GATES

We have shown experimentally that a sub-harmonic drive can perform fast and high-fidelity single qubit control, and found theoretically that strong couplings to low-frequencies are not associated with enhanced qubit losses. Of far greater concern is the behavior of the low-pass filter which protects the qubit. The filter's figures of merit are its losses in transmission at the drive frequency (which lead to cryostat heating as discussed below) and its absorptive losses in reflection and insertion loss at the qubit's transition frequency (which protect the qubit from the environment and Purcell decay into the filter). In room temperature measurements of the filter used in this experiment (Mini-circuits ZLSS-A2R8G-S+), the measured impedance is  $27.1 - 253.7i\Omega$  at the qubit frequency. Assuming the qubit lifetime is only limited by decaying through the drive port, with the low-pass filter used in the experiment, the resonant decay rate into the filter's losses is improved from  $\kappa_{\text{res}} = 4.7\text{ kHz}$  with the qubit port directly connected to a transmission line to  $\kappa_{\text{LPF}} = 159\text{ Hz}$  using the filter. The corresponding sub-harmonic decay rate of  $\kappa_{\text{sub}} = 8.8 \times 10^{-13}\text{ Hz}$  calculated using Eq. 7 is drastically lower and thus can be neglected. While these calculations show that such a low-pass filter can protect qubit lifetime effectively, suppressing the decay rate by about 30 times, we would clearly benefit from an even lower loss filter, as the qubit-drive port coupling strength could be further increased to improve gate speed and reduce required driving power.

Sub-harmonic driving also offers new possibilities to address cryostat and component heating. It is a common practice to use attenuation of around 60 dB to reduce room temperature thermal noise to a very small residual photon occupancy in the drive lines at the readout resonator and qubit frequencies. In some experiments higher attenuation, as much as 80 dB, is used to further reduce noise and provide better protection to the qubits [32, 33].

Figure 4 shows an estimate of heat dissipated at the base stage of a dilution refrigerator in various driving configurations. To give a fair comparison, the port coupling strengths are chosen to limit qubit lifetime to  $T_1 = 1\text{ ms}$  using finite element circuit modeling and the measured absorption of our current filter. Using resonant driving and configuration 1 shown in Figure 4(a) (a high-attenuation conventional drive line), approximately -30 dBm of power is dissipated at base to achieve a  $\pi$ -pulse time of 10 ns. The current limited cooling capacity of our dilution refrigerator ( $\sim 10\mu\text{W}$  at 20 mK) only allows tens of single qubit drives in parallel before

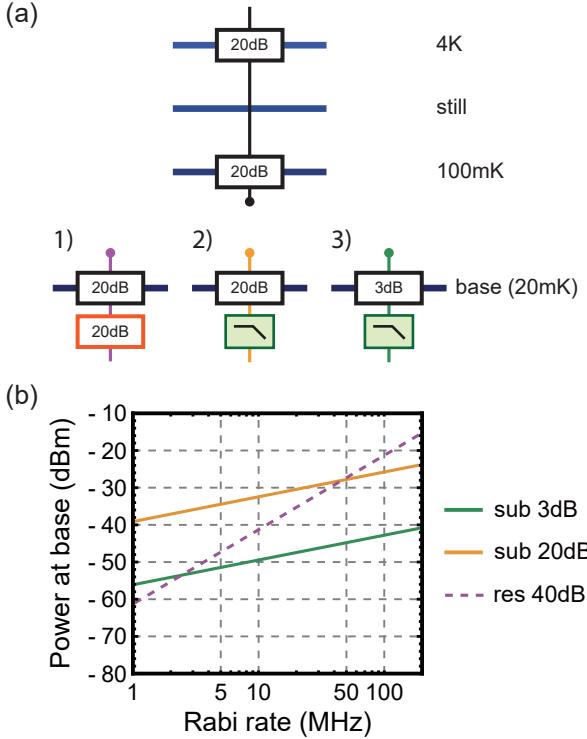


Figure 4. Resonant drive vs. sub-harmonic drive. (a) Different configurations of input line for regular drive and sub-harmonic drive. More attenuation is required to better protect the qubit from thermal noise in resonant drive method, while attenuators can be removed and replaced with reflective filters in sub-harmonic drive method. A commercial LPF is used on the sub-harmonic drive input line. (b) A comparison of dissipated heat vs. Rabi rate  $\Omega$  between different input line configurations that have a  $T_1$  limit of around 1 ms. The regular drive is shown in dashed line, and sub-harmonic drive is shown in solid line.

base temperature rises significantly, which can be an important limit for near-term quantum machines. Moreover, the attenuator itself as a dissipative element can be heated and generate thermal photons in turn; commercial attenuators have been shown to cool down much more slowly than the typical experiment repetition rates, creating still further complications [13]. It would be advantageous to simply remove these elements.

As shown in the Fig. 4, we compare conventional drive lines with 20 dB base attenuation + LPF (configuration

2, close to our experimental configuration) and LPF only at base (configuration 3) for sub-harmonic gates. The third configuration removes attenuators from the base plate completely and alleviates the heat dissipation at base stage. The remaining 3 dB attenuation represents our estimate of realistic losses on components other than RF attenuators, such as Eccosorb filters and coaxial cables. Reducing the losses at base allows us to take fuller advantage of sub-harmonic driving; at a Rabi rate of 50 MHz, heat dissipation for configuration 3 is improved by around 20 dB relative to the conventional drive configuration, allowing 100 times as many qubits to be controlled for the same drive power and realistic assumptions about the drive line configuration.

## V. CONCLUSION

In conclusion, we have shown that sub-harmonic driving is a new single qubit control scheme with fast gate speed and high fidelity. By breaking the symmetry between driving and decaying of a qubit in the frequency domain, we protect the qubit from resonant relaxation and control the qubit without any need for resonant access. Also, the cubic dependence of gate speed on drive strength could greatly reduce the heat dissipated during qubit driving for appropriately designed reflective drive lines. We achieved a gate fidelity of 99.7% as verified by interleaved randomized benchmarking, limited by the qubit coherence and the minimum pulse duration of our control hardware. With longer qubit coherence times and faster room-temperature control system, it's possible to further improve the gate fidelity to 99.99% or even higher, making it a viable method for single qubit gates in future large-scale quantum computers.

## VI. ACKNOWLEDGEMENTS

This research, including funding for M.X., C.Z. and C.L. was supported by the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, the Co-design Center for Quantum Advantage (C2QA) under Contract No. DE-SC0012704. Support was also provided by the Air Force Office of Strategic Research under award FA9550-15-1-0015, and the National Science Foundation PIRE HYBRID program under contract 1743717.

- 
- [1] S. Sheldon, E. Magesan, J. M. Chow, and J. M. Gambetta, Procedure for systematically tuning up crosstalk in the cross-resonance gate, *Physical Review A* **93**, 060302 (2016).
  - [2] P. Mundada, G. Zhang, T. Hazard, and A. Houck, Suppression of qubit crosstalk in a tunable coupling superconducting circuit, *Physical Review Applied* **12**, 054023 (2019).
  - [3] F. Motzoi, J. M. Gambetta, P. Rebentrost, and F. K. Wilhelm, Simple pulses for elimination of leakage in weakly nonlinear qubits, *Physical review letters* **103**, 110501 (2009).
  - [4] R. Barends, J. Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, T. C. White, J. Mutus, A. G. Fowler,

- B. Campbell, *et al.*, Superconducting quantum circuits at the surface code threshold for fault tolerance, *Nature* **508**, 500 (2014).
- [5] J. Kelly, R. Barends, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, I.-C. Hoi, E. Jeffrey, *et al.*, Optimal quantum control using randomized benchmarking, *Physical review letters* **112**, 240504 (2014).
- [6] S. Sheldon, L. S. Bishop, E. Magesan, S. Filipp, J. M. Chow, and J. M. Gambetta, Characterizing errors on qubit operations via iterative randomized benchmarking, *Physical Review A* **93**, 012301 (2016).
- [7] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, Characterizing quantum supremacy in near-term devices, *Nature Physics* **14**, 595 (2018).
- [8] M. Kjaergaard, M. E. Schwartz, J. Braumüller, P. Krantz, J. I.-J. Wang, S. Gustavsson, and W. D. Oliver, Superconducting qubits: Current state of play, *Annual Review of Condensed Matter Physics* **11**, 369 (2020).
- [9] A. P. Place, L. V. Rodgers, P. Mundada, B. M. Smitham, M. Fitzpatrick, Z. Leng, A. Premkumar, J. Bryon, A. Vrajitoarea, S. Sussman, *et al.*, New material platform for superconducting transmon qubits with coherence times exceeding 0.3 milliseconds, *Nature communications* **12**, 1 (2021).
- [10] C. Wang, X. Li, H. Xu, Z. Li, J. Wang, Z. Yang, Z. Mi, X. Liang, T. Su, C. Yang, *et al.*, Transmon qubit with relaxation time exceeding 0.5 milliseconds, *arXiv preprint arXiv:2105.09890* (2021).
- [11] Z. Wang, S. Shankar, Z. Minev, P. Campagne-Ibarcq, A. Narla, and M. H. Devoret, Cavity attenuators for superconducting qubits, *Physical Review Applied* **11**, 014031 (2019).
- [12] K. Serniak, S. Diamond, M. Hays, V. Fatemi, S. Shankar, L. Frunzio, R. Schoelkopf, and M. Devoret, Direct dispersive monitoring of charge parity in offset-charge-sensitive transmons, *Physical Review Applied* **12**, 014052 (2019).
- [13] J.-H. Yeh, J. LeFebvre, S. Premaratne, F. Wellstood, and B. Palmer, Microwave attenuators for use with quantum devices below 100 mK, *Journal of Applied Physics* **121**, 224501 (2017).
- [14] S. Krinner, S. Storz, P. Kurpiers, P. Magnard, J. Heinssoo, R. Keller, J. Luetolf, C. Eichler, and A. Wallraff, Engineering cryogenic setups for 100-qubit scale superconducting circuit systems, *EPJ Quantum Technology* **6**, 2 (2019).
- [15] S. Kono, K. Koshino, D. Lachance-Quirion, A. F. Van Loo, Y. Tabuchi, A. Noguchi, and Y. Nakamura, Breaking the trade-off between fast control and long lifetime of a superconducting qubit, *Nature communications* **11**, 1 (2020).
- [16] A. A. Clerk, M. H. Devoret, S. M. Girvin, F. Marquardt, and R. J. Schoelkopf, Introduction to quantum noise, measurement, and amplification, *Reviews of Modern Physics* **82**, 1155 (2010).
- [17] A. Roy and M. Devoret, Introduction to parametric amplification of quantum signals with Josephson circuits, *Comptes Rendus Physique* **17**, 740 (2016).
- [18] A. Narla, S. Shankar, M. Hatridge, Z. Leghtas, K. M. Sliwa, E. Zalys-Geller, S. O. Mundhada, W. Pfaff, L. Frunzio, R. J. Schoelkopf, *et al.*, Robust concurrent remote entanglement between two superconducting qubits, *Physical Review X* **6**, 031036 (2016).
- [19] A. J. Sirois, M. Castellanos-Beltran, M. DeFeo, L. Ranzeni, F. Lecocq, R. Simmonds, J. Teufel, and J. Aumentado, Coherent-state storage and retrieval between superconducting cavities using parametric frequency conversion, *Applied Physics Letters* **106**, 172603 (2015).
- [20] C. J. Axline, L. D. Burkhardt, W. Pfaff, M. Zhang, K. Chou, P. Campagne-Ibarcq, P. Reinhold, L. Frunzio, S. Girvin, L. Jiang, *et al.*, On-demand quantum state transfer and entanglement between remote microwave cavity memories, *Nature Physics* **14**, 705 (2018).
- [21] C. Zhou, P. Lu, M. Praquin, T.-C. Chien, R. Kaufman, X. Cao, M. Xia, R. S. Mong, W. Pfaff, D. Pekker, *et al.*, Realizing all-to-all couplings among detachable quantum modules using a microwave quantum state router, *npj Quantum Information* **9**, 54 (2023).
- [22] A. Niskanen, K. Harrabi, F. Yoshihara, Y. Nakamura, S. Lloyd, and J. S. Tsai, Quantum coherent tunable coupling of superconducting qubits, *Science* **316**, 723 (2007).
- [23] D. C. McKay, S. Filipp, A. Mezzacapo, E. Magesan, J. M. Chow, and J. M. Gambetta, Universal gate for fixed-frequency qubits via a tunable bus, *Physical Review Applied* **6**, 064007 (2016).
- [24] C. Axline, M. Reagor, R. Heeres, P. Reinhold, C. Wang, K. Shain, W. Pfaff, Y. Chu, L. Frunzio, and R. J. Schoelkopf, An architecture for integrating planar and 3D cqed devices, *Applied Physics Letters* **109**, 042601 (2016).
- [25] R. Rehammar and S. Gasparinetti, Low-pass filter with ultra-wide stopband for quantum computing applications, *ArXiv abs/2205.03941* (2022).
- [26] J. Koch, M. Y. Terri, J. Gambetta, A. A. Houck, D. I. Schuster, J. Majer, A. Blais, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, Charge-insensitive qubit design derived from the cooper pair box, *Physical Review A* **76**, 042319 (2007).
- [27] A. Noguchi, A. Osada, S. Masuda, S. Kono, K. Heya, S. P. Wolski, H. Takahashi, T. Sugiyama, D. Lachance-Quirion, and Y. Nakamura, Fast parametric two-qubit gates with suppressed residual interaction using the second-order nonlinearity of a cubic transmon, *Physical Review A* **102**, 062408 (2020).
- [28] A. Blais, A. L. Grimsmo, S. M. Girvin, and A. Wallraff, Circuit quantum electrodynamics, *Rev. Mod. Phys.* **93**, 025005 (2021).
- [29] C. Wang, Y. Y. Gao, P. Reinhold, R. W. Heeres, N. Ofek, K. Chou, C. Axline, M. Reagor, J. Blumoff, K. Sliwa, *et al.*, A schrödinger cat living in two boxes, *Science* **352**, 1087 (2016).
- [30] M. D. Reed, *Entanglement and Quantum Error Correction with Superconducting Qubits*, Ph.D. thesis, Yale University (2013).
- [31] E. Magesan, J. M. Gambetta, B. R. Johnson, C. A. Ryan, J. M. Chow, S. T. Merkel, M. P. Da Silva, G. A. Keefe, M. B. Rothwell, T. A. Ohki, *et al.*, Efficient measurement of quantum gate error by interleaved randomized benchmarking, *Physical review letters* **109**, 080505 (2012).
- [32] C. Wang, X. Li, H. Xu, Z. Li, J. Wang, Z. Yang, Z. Mi, X. Liang, T. Su, C. Yang, *et al.*, Towards practical quantum computers: Transmon qubit with a lifetime approaching 0.5 milliseconds, *npj Quantum Information* **8**, 3 (2022).
- [33] S. Chakram, A. E. Oriani, R. K. Naik, A. V. Dixit, K. He, A. Agrawal, H. Kwon, and D. I. Schuster, Seamless high-q

microwave cavities for multimode circuit quantum electrodynamics, Physical review letters **127**, 107701 (2021).

# Fast superconducting qubit control with sub-harmonic drives: Supplementary information

Mingkang Xia,<sup>1</sup> Chao Zhou,<sup>1</sup> Chenxu Liu,<sup>2,3,\*</sup> Param Patel,<sup>1</sup> Xi Cao,<sup>1</sup> Pinlei Lu,<sup>1</sup> Boris Mesits,<sup>1</sup> Maria Mucci,<sup>1</sup> David Gorski,<sup>1</sup> David Pekker,<sup>1</sup> and Michael Hatridge<sup>1</sup>

<sup>1</sup>*Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, PA 15213, USA*

<sup>2</sup>*Department of Physics, Virginia Tech, Blacksburg, Virginia 24061, USA*

<sup>3</sup>*Virginia Tech Center for Quantum Information Science and Engineering, Blacksburg, VA 24061, USA*

(Dated: June 21, 2023)

## I. THEORY

### A. Transmon under sub-harmonic drive

The un-driven transmon Hamiltonian can be written as

$$\begin{aligned}\hat{H}_0 &= \sqrt{8E_C E_J} \hat{q}^\dagger \hat{q} - \frac{E_C}{12} (\hat{q}^\dagger + \hat{q})^4 \\ &= \hbar(\omega_q - \alpha) \hat{q}^\dagger \hat{q} + \hbar \frac{\alpha}{12} (\hat{q}^\dagger + \hat{q})^4,\end{aligned}\tag{S1}$$

where  $E_J$  and  $E_c$  are the Josephson energy and charging energy of the transmon qubit, respectively,  $\alpha = -E_c$  is the anharmonicity of the qubit,  $\hbar\omega_q = \sqrt{8E_c E_J} - E_c$  is the qubit angular frequency, and  $\hat{q}$  is the photon annihilation operators of the qubit. In the expansion of  $(\hat{q}^\dagger + \hat{q})^4$ , most 4<sup>th</sup> order terms are fast oscillating in the absence of drives and can be neglected under the rotating wave approximation (RWA). However, these fast-oscillating terms are important for parametric driving, and so we cannot immediately apply the rotating wave approximation. The Hamiltonian of a transmon under a single microwave drive can be written as

$$\begin{aligned}\hat{H} &= \hat{H}_0 + \hat{H}_{\text{drive}} \\ &= \hbar(\omega_q - \alpha) \hat{q}^\dagger \hat{q} + \hbar \frac{\alpha}{12} (\hat{q}^\dagger + \hat{q})^4 + \hbar\epsilon(t)(\hat{q}^\dagger + \hat{q}),\end{aligned}\tag{S2}$$

where

$$\epsilon(t) = \begin{cases} \varepsilon(t)e^{-i\omega_d t} + \varepsilon^*(t)e^{i\omega_d t} & 0 < t < t_{\text{gate}} \\ 0 & \text{otherwise} \end{cases}.\tag{S3}$$

If the drive frequency is nearly on-resonant with the qubit's transitions, this equation will yield the typical directly driven transmon dynamics. On the other hand, for a far-detuned drive, it is useful to continue by applying a displacement transformation  $\hat{D}(t) = e^{z\hat{q}^\dagger - z^*\hat{q}}$  which cancels the driving term, moving its effects into the qubit operator. If we define  $z$  as

$$z = -\frac{\varepsilon(t)}{\omega_d - \omega'_q} e^{-i\omega_d t} + \frac{\varepsilon^*(t)}{\omega_d + \omega'_q} e^{i\omega_d t} \quad \text{and} \quad \omega'_q = \omega - \alpha,\tag{S4}$$

the driving term is canceled up to a scalar  $C$ . After the displacement transformation, the Hamiltonian can be written as

$$\begin{aligned}\hat{H}^D/\hbar &= \hat{D}\hat{H}\hat{D}^\dagger/\hbar + i\dot{\hat{D}}\hat{D}^\dagger \\ &= \omega'_q \hat{q}^\dagger \hat{q} + \frac{\alpha}{12} (\hat{q}^\dagger + \hat{q} - z^* - z)^4 + C,\end{aligned}\tag{S5}$$

---

\* Recently moved to: Pacific Northwest National Laboratory, Richland, WA 99354, USA

which can be further simplified to:

$$\hat{H}^D/\hbar = \omega_q' \hat{q}^\dagger \hat{q} + \frac{\alpha}{12} (\hat{q}^\dagger + \hat{q} + \eta e^{-i\omega_d t} + \eta^* e^{i\omega_d t})^4, \text{ with } \eta = \frac{2\omega_q' \varepsilon(t)}{\omega_d^2 - \omega_q'^2} \quad (\text{S6})$$

The fourth order component of Eq. S6 conceals many potential dynamical behaviors of the qubit that can be activated by applying the drive at specific frequencies  $\omega_d$  [1]. Generally, each term in the expansion of this fourth-order component involves 1-3 qubit operators  $\hat{q}$  or  $\hat{q}^\dagger$ , and, correspondingly 3-1 ‘pump’ waves  $\eta$  or  $\eta^*$ . For our purposes, we are interested in the terms that activate the single-photon excitation ( $|g\rangle \leftrightarrow |e\rangle$  transition) on the qubit, which therefore should contain one qubit operator and three pump waves. Considering the drive we applied is at around one-third of the qubit frequency, i.e.  $\omega_d = \omega_q/3 + \delta$ , we move to the rotating frame that rotates at  $3\omega_d$  and observe the effect of this pump:

$$\hat{H}^R/\hbar = \hat{R} \hat{H}^D \hat{R}^\dagger / \hbar + i \dot{\hat{R}} \hat{R}^\dagger, \text{ with } \hat{R} = e^{3i\omega_d t \hat{q}^\dagger \hat{q}}. \quad (\text{S7})$$

After this transformation, the first term in  $\hat{H}^D$  yields:

$$\hat{R}(\omega_q' \hat{q}^\dagger \hat{q}) \hat{R}^\dagger / \hbar + i \dot{\hat{R}} \hat{R}^\dagger = (-\alpha - 3\delta) \hat{q}^\dagger \hat{q}. \quad (\text{S8})$$

The second term in  $\hat{H}^D$  can be expanded and each individual term has the general form of  $O(\eta^m \eta^{*n}) q^{\dagger i} q^j e^{-i(m-n)\omega_d t}$ , with  $i, j, m, n \geq 0$  and  $i + j + m + n = 4$ . In our chosen frame, the phase of each term is  $(m - n + 3i - 3j)\omega_d t$ . The terms that satisfy  $m - n + 3i - 3j = 0$  survive under RWA and have a significant effect on qubit evolution. In summary, the Hamiltonian has the form:

$$\begin{aligned} \hat{H}^R/\hbar &= (\hat{H}_{\text{Stark}} + \hat{H}_{\text{Kerr}} + \hat{H}_d^{\text{eff}})/\hbar \\ &= (2\alpha|\eta|^2 - 3\delta) q^\dagger q + \frac{\alpha}{2} q^\dagger q^\dagger q q + \frac{\alpha}{3} (\eta^3 q^\dagger + \eta^{*3} q). \end{aligned} \quad (\text{S9})$$

For each driving amplitude  $\varepsilon(t)$ , we can choose a drive detuning  $\delta$ , that makes  $2\alpha|\eta|^2 - 3\delta = 0$ , such that the first term in Eq. S9 goes to 0, and the Hamiltonian becomes:

$$\hat{H}^R/\hbar = \frac{\alpha}{2} q^\dagger q^\dagger q q + \frac{\alpha}{3} (\eta^3 q^\dagger + \eta^{*3} q), \quad (\text{S10})$$

which has exactly the same form as a transmon under  $|g\rangle \leftrightarrow |e\rangle$  on-resonance drive with strength  $\frac{\alpha}{3}\eta^3$ , in the frame rotating at  $\omega_q$  [2].

## B. Flat-top pulse

The most commonly used pulse shape for transmon control is the Gaussian pulse. The main benefit of a Gaussian pulse is that its profile in the frequency domain is also a Gaussian function, and thus has a relatively narrow frequency distribution. It can avoid exciting other unwanted energy transitions in the qubit. In practice, the Gaussian pulse is usually truncated by a few  $\sigma$  to make a (formally) infinitely long pulse possible to use in experiments. The definition can be written as:

$$A_G(t) = \begin{cases} A_0 e^{-\frac{1}{2}(\frac{t-t_0}{\sigma})^2} & -n\sigma < t - t_0 < n\sigma \\ 0 & \text{otherwise} \end{cases}. \quad (\text{S11})$$

Another frequently used pulse shape for transmon control is the flat-top pulse, or the smoothed square pulse. We used flat-top pulses in the sub-harmonic driving experiment for the qubit control. Flat-top pulses have the advantage of a constant amplitude for the duration of the pulse. To narrow the bandwidth in frequency, the edges of the pulses are often smoothed. Flat-top pulse envelopes can be slightly different because of the definition of the ramp-up/down, for example it can be defined using be not limited to piece-wise definition, hyper-Gaussian definition, half-cosine definition and so on. We define the flat-top pulses with tanh functions. The flat-top pulse starting at  $t = 0$  and length of  $t_0$  is defined as:

$$A_F(t) = \begin{cases} \frac{1}{2} A_0 (\tanh(kt - t_1) - \tanh(k(t - t_0) + t_1)) & 0 < t < t_0 \\ 0 & \text{otherwise} \end{cases} \quad (\text{S12})$$

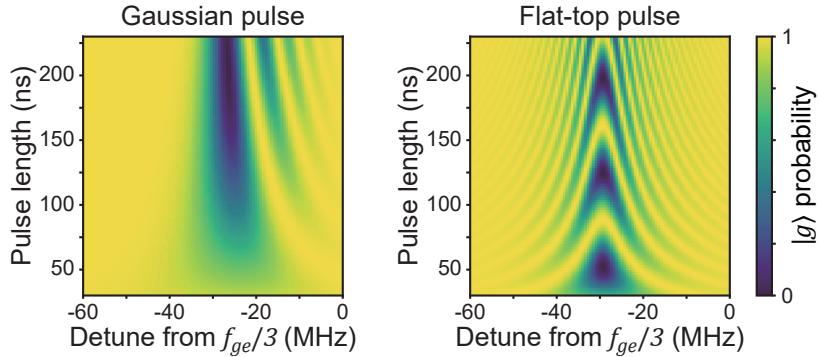


Figure S1. A simulation of Rabi experiment with the Gaussian pulse and the flat-top pulse

where  $k$  defines the ramping speed and  $t_1/k$  defines the mid-points of the pulse ramp. When a different pulse length is required, only  $t_0$  is changed in Eq. S12 in our experiment.

Figure S1 shows a numerical simulation of sub-harmonic Rabi process with fixed driving amplitude  $A_0$  and varying length using Gaussian pulses and flat-top pulses. As a result of the AC-Stark shift, the Rabi experiment of Gaussian pulses looks distorted, because amplitude is changing throughout the pulse. It is complicated to select pulse lengths for  $\pi$  and  $\pi/2$  rotations at a given drive frequency. The result of the flat-top pulse, on the other hand, remains a pattern similar to Rabi experiment result with resonant driving. For Rabi process with fixed pulse length and varying amplitude and frequency, both types of pulse profiles give distorted results. Therefore, it's simpler to tune up a single qubit gate with flat-top pulses with fixed amplitude and varying duration. Choosing flat top pulses has the advantage of facilitating the calibration of the added-phase resulting from the off-resonance drive, which is particularly crucial for achieving phase-coherent qubit control, and will be discussed in the next section.

### C. Phase correction of sub-harmonic gate

Because of the Stark effect during the sub-harmonic drive, performing a pulse will effectively add a phase to the qubit. To compensate for the added phase, phase correction needs to be applied to all of the following pulses to rotate the qubit in the desired direction. It can be considered as adding a virtual-Z gate[3] between two sub-harmonic gates. Such phase correction is also required in other kinds of superconducting qubit control, such as the parametric two-qubit gate and the CR gate.

The phase difference between the drive frame and the qubit frame can be written as

$$\varphi(t_n) = \sum_i \varphi_i = \sum_i \int_{t_i}^{t_{i+1}} \omega_q(t) - 3\omega_d dt \quad (\text{S13})$$

where  $t_i$  is the starting time of pulse  $i$  in the whole sequence. Each time interval of a pulse can be separated into two parts as shown in Figure S2. From  $t_i$  to  $t_i + t_{gate}$  is the time when the pulse is applied, and from  $t_i + t_{gate}$  to  $t_{i+1}$  is the time gap between two pulses. During the time gap between two pulses, the detuning and driving frequency and qubit frequency remains constant. Therefore,  $\varphi_i$  can be written as

$$\varphi_i = \varphi_{\text{gate}_i} + (\omega_q - 3\omega_d)t_{\text{gap}}. \quad (\text{S14})$$

For flat-top pulses,  $\varphi_i$  can be further simplified. The generator frequency is set to be the same as the qubit frequency during the flat-top part of the pulse. Therefore, during the flat-top part of the drive, as shown in green color in Figure S3, the integration value is zero. Therefore, we have  $\varphi_{\text{gate}_i} = \varphi_{\text{ramp}_i}$ . To vary the length of a pulse, only the length of the flat-top part of the pulse is changed, while the ramp-up and ramp-down profiles of the pulse are the same. As a result, the  $\varphi_{\text{gate}_i}$  is a constant for all pulses and only two parameters,  $\varphi_{\text{ramp}}$  and  $\delta\omega = \omega_q - 3\omega_d$ , need to be calibrated.

Another important thing is performing a gate in the y-axis direction. After correcting the phase caused by the Stark shift, the remaining phase that needs to be considered is the phase of the gate itself. For example, a  $Y$  gate has a phase of  $\pi/2$ , and a  $-X$  gate has a phase of  $\pi$ . Because the gate is a three-photon transition, only  $1/3$  of the gate's phase needs to be added to the drive. It might create difficulties in defining pulses. One trick that can be done is

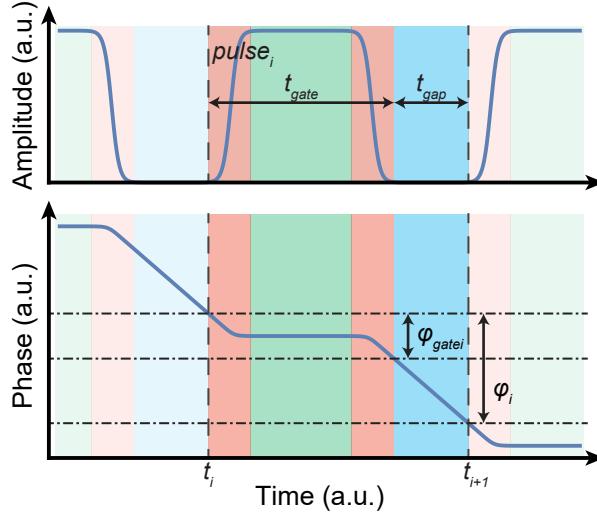


Figure S2. A pulse sequence of flat-top pulses. For phase correction, the pulse is separated into three parts, the ramp up/down, flat and gap parts, represented by red, green, and blue colors in the figure.

flipping the direction of the y-axis of the drive frame and adding the gate phase directly to the drive. However, such a trick only works for the three-photon sub-harmonic gates in  $X$  and  $Y$  direction because

$$\begin{aligned} \varphi \bmod 2\pi &= -3\varphi \bmod 2\pi \\ \Rightarrow \quad \varphi &= \frac{n\pi}{2}, \quad n \in \mathbb{N}. \end{aligned} \quad (\text{S15})$$

#### D. Suppressing leakage to high excitation state

For faster sub-harmonic pulses that can cause leakage to higher excited states, DRAG correction can be applied [4]. Only considering the first three lowest levels of the transmon qubit, the Hamiltonian in Eq. S9 can then be truncated to:

$$\hat{H}^R/\hbar = (2\alpha|\eta|^2 - 3\delta)|e\rangle\langle e| + (\alpha + 4\alpha|\eta|^2 - 6\delta)|f\rangle\langle f| + \frac{\alpha}{3}[\eta^{*3}(|g\rangle\langle e| + \sqrt{2}|e\rangle\langle f|) + h.c.]. \quad (\text{S16})$$

Defining

$$\sigma_{j,k}^x = |j\rangle\langle k| + |j\rangle\langle k|, \quad \sigma_{j,k}^y = -i|j\rangle\langle k| + i|k\rangle\langle j| \quad (\text{S17})$$

$$\zeta^x = \text{Re}(\eta^3), \quad \zeta^y = \text{Im}(\eta^3), \quad (\text{S18})$$

the Hamiltonian can be simplified to

$$\hat{H}/\hbar = (2\alpha|\eta|^2 - 3\delta)|e\rangle\langle e| + (\alpha + 4\alpha|\eta|^2 - 6\delta)|f\rangle\langle f| + \frac{\alpha}{3}[\zeta^x(\sigma_{g,e}^x + \sqrt{2}\sigma_{e,f}^x) + \zeta^y(\sigma_{g,e}^y + \sqrt{2}\sigma_{e,f}^y)]. \quad (\text{S19})$$

To implement DRAG, the adiabatic transformation

$$V(t) = \exp[-i\frac{\zeta^x}{3}(\sigma_{g,e}^y + \sqrt{2}\sigma_{e,f}^y)] \quad (\text{S20})$$

is performed on the Hamiltonian in Eq. S16. Assuming  $\eta, \zeta^x, \zeta^y \ll 1$  and  $\delta \ll \alpha$ , we have

$$\begin{aligned} \hat{H}^V \approx & (2\alpha|\eta|^2 - 3\delta + \frac{2}{9}\alpha\zeta^{x2})|e\rangle\langle e| + (\alpha + 4\alpha|\eta|^2 - 6\delta + \frac{4}{9}\alpha\zeta^{x2})|f\rangle\langle f| \\ & + \frac{1}{3}\alpha\zeta^x\sigma_{g,e}^x - \frac{\sqrt{2}}{3}(2\alpha|\eta|^2 - 3\delta)\zeta^x\sigma_{e,f}^x + \frac{\sqrt{2}}{18}\zeta^{x2}\sigma_{g,f}^x + \frac{1}{3}(\alpha\zeta^y + \dot{\zeta}^x)(\sigma_{g,e}^y + \sqrt{2}\sigma_{e,f}^y). \end{aligned} \quad (\text{S21})$$

It shows that to cancel transition to  $|f\rangle$  and ac-Stark shift to the 1<sup>st</sup> order during the drive,

$$\begin{aligned}\zeta^y &= \frac{-\dot{\zeta}^x}{\alpha}, \\ \delta &= \frac{2}{3}\alpha|\eta|^2 - \frac{2}{27}\alpha\zeta^{x2}.\end{aligned}\quad (\text{S22})$$

Using the definition in Eq. S18, we can find the correction pulse amplitude by solving the differential equation:

$$\begin{aligned}\text{Im}[\eta(t)^3] + \frac{1}{\alpha} \frac{\partial \text{Re}[\eta(t)^3]}{\partial t} &= 0 \\ \Rightarrow \dot{\varepsilon}_x(\varepsilon_x^2 - \varepsilon_y^2) - 6\varepsilon_x\varepsilon_y\dot{\varepsilon}_y + \alpha(3\varepsilon_x^2\varepsilon_y - \varepsilon_y^3) &= 0.\end{aligned}\quad (\text{S23})$$

A pulse with 0 phase only has  $\varepsilon_x$  component before applying DRAG correction. Therefore, we can find the correction amplitude by solving for  $\varepsilon_y$  in Eq. S23. An example of a 50 ns pulse with first-order DRAG correction is given in Figure S3. Pulses with phase  $\varphi$  can be derived from the  $x$  and  $y$  components of zero phase pulse,  $\varepsilon_x$  and  $\varepsilon_y$ ,

$$\begin{aligned}\varepsilon'_x &= \varepsilon_x \cos(\varphi/3) - \varepsilon_y \sin(\varphi/3) \\ \varepsilon'_y &= \varepsilon_x \sin(\varphi/3) + \varepsilon_y \cos(\varphi/3).\end{aligned}\quad (\text{S24})$$

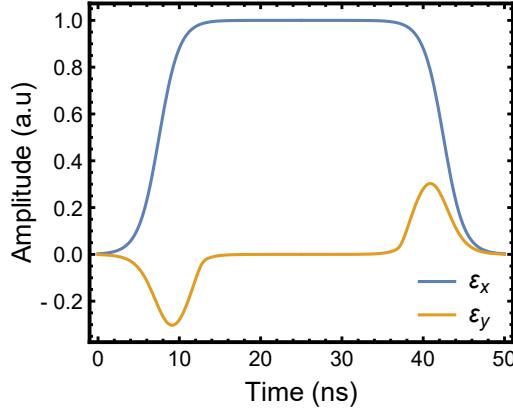


Figure S3. DRAG correction of a 50 ns flat-top pulse. The pulse with zero phase and without DRAG correction is defined to only have  $\varepsilon_x$  component, shown as the blue trace. The orange trace,  $\varepsilon_y$  component, is the DRAG correction of the pulse.

### E. Modelling drive strength

As shown in Figure 1, we consider a transmon qubit placed inside an Aluminum tube and only couples to a driving port. It can be modeled by the lumped-element circuit, shown in Figure S4. The two antenna pads A and B of the transmon are each capacitively coupled to the ground, the wall of the package, and the center pin of the driving port. When the coupling between transmon pads and the external environment is weak, the Hamiltonian of interaction between transmon and external drive can be written as  $\hat{H}_{int} = 2e\hat{n}\beta V_d$ , where  $\hat{n}$  is the number operator of cooper pairs and  $\beta$  is a scaling factor between the voltage at the drive port and the voltage across the transmon's Josephson junction. The factor  $\beta$  is defined as follow[5],

$$C_a = C_{a0} + C_{a1} \quad (\text{S25a})$$

$$C_b = C_{b0} + C_{b1} \quad (\text{S25b})$$

$$C_s = \frac{C_a C_b}{C_a + C_b} + C \quad (\text{S25c})$$

$$\beta = \frac{C_a C_{b1} - C_b C_{a1}}{C_s(C_a + C_b)} \quad (\text{S25d})$$

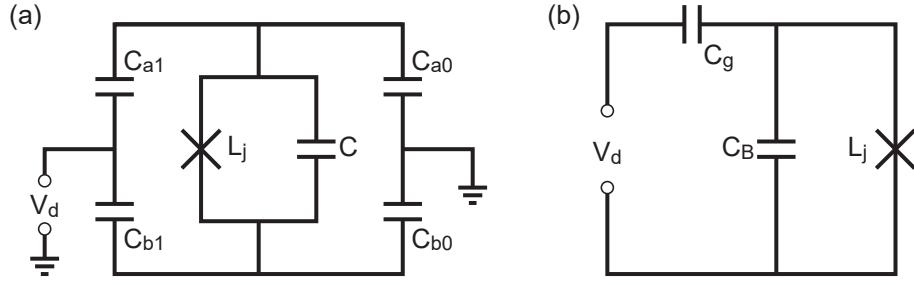


Figure S4. (a) The lumped-element circuit diagram of transmon with single drive port. (b) the reduced equivalent circuit.

Knowing the power sent by room temperature electronics and the total attenuation on the input line, the drive port voltage  $V_d$  can be calculated easily.

The Hamiltonian of a qubit under a drive with frequency of  $\omega_d$  and amplitude  $V$  can be written as,

$$\begin{aligned}\hat{H} &= \hat{H}_0 + \hat{H}_{int} \\ &= \frac{1}{2}\hbar\omega_q\sigma_z + 2e\hat{n}\beta V\cos(\omega_dt),\end{aligned}\quad (\text{S26})$$

where  $\hat{n} = n_0\Sigma\sqrt{i+1}(|i\rangle\langle i+1| + |i+1\rangle\langle i|)$ ,  $n_0 = \sqrt[4]{E_j/32E_c}$ . [6] Since we are only considering the first two energy levels of the transmon,  $\hat{n}$  can be simplified to  $\hat{n} = n_0\hat{\sigma}_x$ . So  $\hat{H} = \hbar\omega_q\hat{\sigma}_z/2 + 2en_0\beta V\cos(\omega_dt)\hat{\sigma}_x$

In the interaction picture we have

$$2en_0\beta V \begin{pmatrix} 0 & e^{-i\omega_q t}\cos(\omega_dt) \\ e^{i\omega_q t}\cos(\omega_dt) & 0 \end{pmatrix} \begin{pmatrix} c_1(t) \\ c_0(t) \end{pmatrix} = i\hbar \begin{pmatrix} \dot{c}_1(t) \\ \dot{c}_0(t) \end{pmatrix}, \quad (\text{S27})$$

Assuming  $\omega_d = \omega_q$  and taking the rotating wave approximation, we have

$$en_0\beta V \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c_1(t) \\ c_0(t) \end{pmatrix} = i\hbar \begin{pmatrix} \dot{c}_1(t) \\ \dot{c}_0(t) \end{pmatrix}, \quad (\text{S28})$$

It can be solved that

$$\begin{cases} c_1(t) = a_1e^{-igt} + a_2e^{igt} \\ c_0(t) = b_1e^{-igt} + b_2e^{igt} \end{cases}, \quad (\text{S29})$$

where  $g = en_0\beta V(t)/\hbar$  and coefficients  $a_1, a_2, b_1, b_2$  depend on the initial state  $|\varphi\rangle = c_1(0)|1\rangle + c_0(0)|0\rangle$ . Assuming the initial state is  $|\varphi\rangle = |0\rangle$ , we have  $c_1(t) = \sin(gt)e^{i\theta_1}$ ,  $c_0(t) = \cos(gt)e^{i\theta_0}$ . The expectation value of  $z$ -axis measurement is  $|c_1|^2 - |c_0|^2 = -\cos(2gt)$ . So the Rabi rate of the process is

$$\Omega = 2en_0\beta V/\hbar. \quad (\text{S30})$$

The Rabi rate of sub-harmonic process can be calculated together with Eq. 4,

$$\Omega_{sub} = \frac{1}{3}\alpha\left(\frac{\omega'_q\Omega}{\omega_d^2 - \omega'_q^2}\right)^3. \quad (\text{S31})$$

#### F. Sub-harmonic driving vs. resonant driving

The cooling power of dilution refrigerator at base temperature ( $\sim 20\text{ mK}$ ) is limited, usually around  $10s\mu\text{W}$  because of the low heat conductivity between liquid helium and metal surface at cryogenic temperature. Therefore, heat dissipation at base stage is an important factor to consider when choosing a controlling method. Meanwhile, the stages with higher temperature have much higher cooling power, which are less likely to be the limiting factors of driving power. Here, we compared the heat dissipated at base stage between resonant drive and sub-harmonic drive.

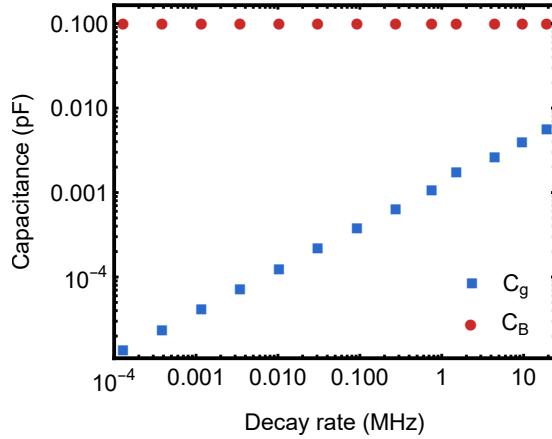


Figure S5. Capacitance of reduced circuit shown in Figure S4(b) that is calculated from simulation using Ansys Maxwell.

Considering the base stage as a one port device, the dissipated heat inside the device is determined by the input driving power  $P_B$  and the total attenuation on base stage drive lines  $T_B$ . Ignoring the heat absorbed by the transmon system and the package, the heat dissipated at the base stage approximately equals to  $P_B(1 - T_B^2)$ . To compare the performance of the two driving scheme, we want to know that, under same gate speed and  $T_1$  limitation, which method dissipates more heat. Therefore, we need to know the relation between driving power  $P_B$  and Rabi rate  $\Omega$  under giving driving port's coupling strength  $\kappa_{\text{ext}}$ .

As discussed in Sec. IE, the Rabi rate of the qubit under certain driving power can be calculated using a lumped circuit model. To extract parameters of the lumped circuit model, we used the FEM software Ansys Maxwell to calculate the capacitance matrix of all conductors, including two antenna pads of the transmon, the tube wall and the probe pin of the drive port. The readout resonator is neglected in the simulation because it has little effects on qubit driving. Figure S5 shows the capacitance in the equivalent lumped-element circuit under varying port's coupling strength.

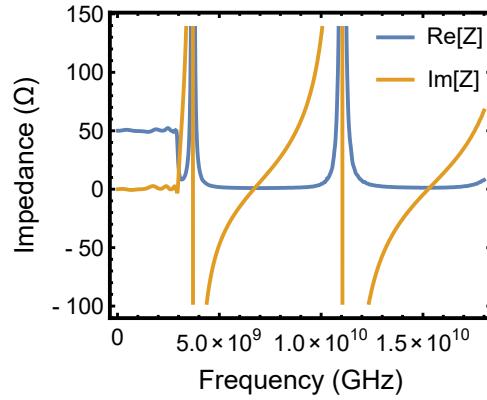


Figure S6. Impedance vs. frequency of low-pass filter used in experiment

The qubit mode properties, such as frequency and lifetime, are acquired using Ansys HFSS. To model the qubit decaying through the port, a lumped RLC port is assigned to the driving port. For resonant driving, the port is assigned a  $50\Omega$  impedance for perfect impedance match. For sub-harmonic driving, to evaluate the effect of the low-pass filter, the impedance of the port is calculated from the S-parameter measured by VNA at room temperature. The commercial low-pass filter Mini-Circuits ZLSS-A2R8G-S+ is measured and used in this experiment. To compare the performance of two driving methods, the coupling port strength is adjusted so that the qubit mode has the same Q factor of  $2.5 \times 10^7$ , which corresponds to  $T_1 \approx 1$  ms.

### G. Qubit relaxation through multi-photon emission process

In this sub-section, we calculate the decay rate of a transmon qubit via sub-harmonic relaxation. We specifically focus on the relaxation from the first excited state  $|1\rangle$  to the ground state  $|0\rangle$  via the emission of three photons, each carrying away approximately  $\hbar\omega_q/3$  of energy.

The Hamiltonian that describes the system (i.e. the transmon qubit) and the photon bath (i.e. the transmission line) is,

$$\begin{aligned} H &= H_{\text{sys}} + H_{\text{bath}} + H_{\text{s-b}} \\ H_{\text{sys}} &= \omega_q q^\dagger q + H_{\text{nl}} \end{aligned} \quad (\text{S32})$$

$$H_{\text{bath}} = \int d\nu \nu b_\nu^\dagger b_\nu \quad (\text{S33})$$

$$H_{\text{s-b}} = \int d\nu \lambda(\nu) (b_\nu^\dagger q + b_\nu q^\dagger) \quad (\text{S34})$$

where we set  $\hbar = 1$ ,  $b_\nu^\dagger$  creates a bath photon with angular frequency  $\nu$  and  $\lambda(\nu)$  describes the system-bath coupling and we define its form below. Terms contributing to  $H_{\text{nl}}$ , which describe the nonlinearity of the transmon, can be found in Eq. (S1). We remark that the terms  $q^4$  and  $(q^\dagger)^4$  do not participate in the lowest order dynamics, and hence we ignore these two terms for simplicity. We specifically focus on the nonlinear Hamiltonian

$$H_{\text{nl}} = g_2(q^\dagger q^\dagger + qq) + g_4(q^\dagger q^\dagger q^\dagger q + q^\dagger qqq) + g'_4 q^\dagger q^\dagger qq, \quad (\text{S35})$$

where we define  $g_4 = \alpha/3$ ,  $g_2 = \alpha/2$  and  $g'_4 = \alpha/2$ , and  $\alpha$  is the anharmonicity of the qubit. The system-bath coupling strength is

$$\lambda(\nu) = \Theta(\nu) \frac{C_c}{\sqrt{C_r c}} \sqrt{\frac{\omega_0 \nu}{2\pi v_{\text{tl}}}}, \quad (\text{S36})$$

where  $C_c$  is the coupling capacitance between the qubit and the transmission line,  $C_r$  is the capacitance of the qubit,  $c$  is the characteristic capacitance of the transmission line,  $v_{\text{tl}}$  is the speed of the transmission line,  $\omega_0$  is the frequency of the qubit [2]. In our case, we put a cut-off (filter) function  $\Theta(\nu)$  to suppress the high-frequency system-bath coupling. Specifically, with the Kerr nonlinearity, we assume

$$\Theta(\nu) = \begin{cases} 1 & |\nu| \leq \omega_0/3 + \vartheta \\ 0 & |\nu| > \omega_0/3 + \vartheta \end{cases} \quad (\text{S37})$$

Due to the energy conservation, we consider a three-photon relaxation process, i.e., a single transmon excitation decays to three bath photons. The lowest-order term contributing to the decay rate of the qubit comes from fourth-order Fermi's golden rule perturbation theory [7]. Specifically we are looking for transitions from the initial state  $|1; \text{vec}\rangle$  to the final state  $|0; 1_{\nu_1}, 1_{\nu_2}, 1_{\nu_3}\rangle$ , where the first half of the ket notation indicates the state of the qubit and the second part the state of the bath. We ignore the transitions that end up with fewer than three photons, as these transitions are suppressed by the system-bath low-pass filter. We remark that transitions that end up with three photons in fewer than three modes are possible and do contribute to the decay rate. However, the rate for these transitions scales as  $(D(\nu)^{-1}/\omega_0)$  (to two distinct modes) and  $(D(\nu)^{-1}/\omega_0)^2$  (all photons go into the same mode), where  $D(\nu)$  is the density of states in the transmission line. The density of states of a length  $l$  transmission line with dispersion relation  $\nu = v_{\text{tl}} k$  is given by

$$D(\nu) = \frac{\nu_{\text{tl}}}{2\pi l}. \quad (\text{S38})$$

Specifically, a 1 m long transmission line with  $v_{\text{tl}} \sim c/1.5$ , the density of states  $D(\nu) \sim 0.032 \text{ GHz}^{-1}$ , which gives  $D(\nu)^{-1}/\omega_0 \sim 0.01$ . This ratio is small compared to other frequency ratios (e.g.  $|\alpha|/\omega_0 \sim 0.1$  order). Therefore, we ignore these transitions [8].

The transition rate can then be calculated from Fermi's golden rule as

$$\Gamma = 2\pi \frac{1}{6} \int_{\omega_0/3-\vartheta}^{\omega_0/3+\vartheta} d\nu_1 d\nu_2 d\nu_3 \delta(\omega_0 - \nu_1 - \nu_2 - \nu_3) \left| \sum_{m,n,p} \frac{\langle 0; 1_{\nu_1}, 1_{\nu_2}, 1_{\nu_3} | V | m \rangle \langle m | V | n \rangle \langle n | V | p \rangle \langle p | V | 1; \text{vac} \rangle}{(\omega_0 - \varepsilon_p + i\eta)(\omega_0 - \varepsilon_n + i\eta)(\omega_0 - \varepsilon_m + i\eta)} \right|^2, \quad (\text{S39})$$

where  $V = H_{\text{nl}} + H_{\text{s-b}}$ , the summation is over all possible system-bath states,  $\varepsilon_{m,n,p}$  are the energies of the system-bath state  $|m\rangle$ ,  $|n\rangle$  and  $|p\rangle$ , the factor of 1/6 is to remove the over-counting induced by the frequency integrals, and we take the limit  $\eta \rightarrow 0_+$ . We then count the virtual paths that are involved in the fourth-order perturbation. Note that there are two distinct sets of possible paths. One set of the paths consists of states

$$|p\rangle = |3; \text{vac}\rangle, \quad |n\rangle = |2; 1_{\nu_1}\rangle, \quad |m\rangle = |1; 1_{\nu_1}, 1_{\nu_2}\rangle, \quad (\text{S40})$$

and all possible permutations of  $\nu_1$ ,  $\nu_2$ , and  $\nu_3$ . The other set of paths includes the path

$$|p\rangle = |0; 1_{\nu_1}\rangle, \quad |n\rangle = |2; 1_{\nu_1}\rangle, \quad |m\rangle = |1; 1_{\nu_1}, 1_{\nu_2}\rangle, \quad (\text{S41})$$

and all the other paths formed by permuting  $\nu_1$ ,  $\nu_2$ , and  $\nu_3$ . Putting these considerations together, we obtain the three-photon relaxation rate

$$\Gamma = \frac{2\pi}{6} \int d\nu_1 d\nu_2 d\nu_3 \delta(\omega_0 - \nu_1 - \nu_2 - \nu_3) \times \left| \sum_{\{\nu_1, \nu_2, \nu_3\}} \frac{6\lambda(\nu_1)\lambda(\nu_2)\lambda(\nu_3)(g_2 + g_4)}{(-\nu_1 - \nu_2 + i\eta)(-\omega_0 - \nu_1 + i\eta)(-2\omega_0 + i\eta)} + \frac{2\lambda(\nu_1)\lambda(\nu_2)\lambda(\nu_3)g_2}{(-\nu_1 - \nu_2 + i\eta)(-\omega_0 - \nu_1 + i\eta)(\omega_0 - \nu_3 + i\eta)} \right|^2, \quad (\text{S42})$$

where  $\sum_{\{\nu_1, \nu_2, \nu_3\}}$  indicates a summation over all terms that are generated by permuting  $\nu_1$ ,  $\nu_2$ , and  $\nu_3$ . We then make the further assumption that  $\vartheta/\omega_0 \sim |\alpha|/\omega_0 \ll 1$ . We find that the relaxation rate, to lowest order in  $\vartheta/\omega_0$ , is

$$\Gamma = \frac{243}{32\pi^2} \frac{\gamma_1^3 |\alpha|^2}{\omega_0^4} \left( \frac{\vartheta}{\omega_0} \right)^2, \quad (\text{S43})$$

where we consider that in the coupling bandwidth, the system-bath coupling is slowly varying in the band with  $\lambda(\nu) \sim \lambda(\omega_0/3)$ , and we define  $\gamma_1 = 2\pi|\lambda(\omega_0/3)|^2$ . We further note that when the transmon is coupled to a thermal bath, which has a mean thermal excitation number

$$\text{Tr}_{\text{bath}}(b_{\nu_1} b_{\nu_1}^\dagger \rho_{\text{b}}) = (\bar{n}_{\nu_1} + 1), \quad (\text{S44})$$

the relaxation rate will depend on the thermal bath temperature. Because the three-photon relaxation process requires coupling the transmon to three bath photons, the thermal relaxation rate is

$$\Gamma_{\text{th, relax}} = (\bar{n}_{\omega_0/3} + 1)^3 \Gamma. \quad (\text{S45})$$

This can be seen from the fact that when calculating the jump term in the master equation, we will have terms in the form of  $\text{Tr}\{b_{\nu_1}^\dagger b_{\nu_2}^\dagger b_{\nu_3}^\dagger \rho_{\text{b}} b_{\nu_3} b_{\nu_2} b_{\nu_1}\}$ . Meanwhile, the bath will thermally excite the transmon with a rate

$$\Gamma_{\text{th, excite}} = \bar{n}_{\omega_0/3}^3 \Gamma. \quad (\text{S46})$$

## II. EXPERIMENT DETAIL

### A. Experiment setup

The metal block housing the qubit and cavity's tube is made of Al-6061 alloy with a 3 mm diameter hollow tube and three coupling ports. One of the ports is coupled to the transmon for single qubit control. The other two ports are coupled to the resonator for the dispersive readout. In this experiment, only one of the readout ports is connected with the output coaxial line and the other one is removed. The transmon qubit and the  $\lambda/2$  stripline resonator are fabricated on a 421 nm thick sapphire substrate with around 200 nm Ta film except for the Josephson junction of the transmon, which is a Al/AlO<sub>x</sub>/Al junction.

The experiment setup is shown in the Figure S7. We used a Mini-circuits low-pass filter ZLSS-A2R8G-S+ for protecting the qubit and performing sub-harmonic drives. The sample is mounted at the base plate of the dilution refrigerator, which is operated at around 20 mK. The qubit driving pulses and resonator readout pulses are generated by upconverting the IF signals from arbitrary waveform generators (Keysight M3202A (1 GSa/s) and M3201A (500 MSa/s)) with LO signals from RF generators (SignalCore SC5511A and SC5506A). For the qubit drive, before the signal is sent to the sample, a small fraction of it is split using a directional coupler, then downconverted by the RF signal and sent to the digitizer for monitoring pulse stability.

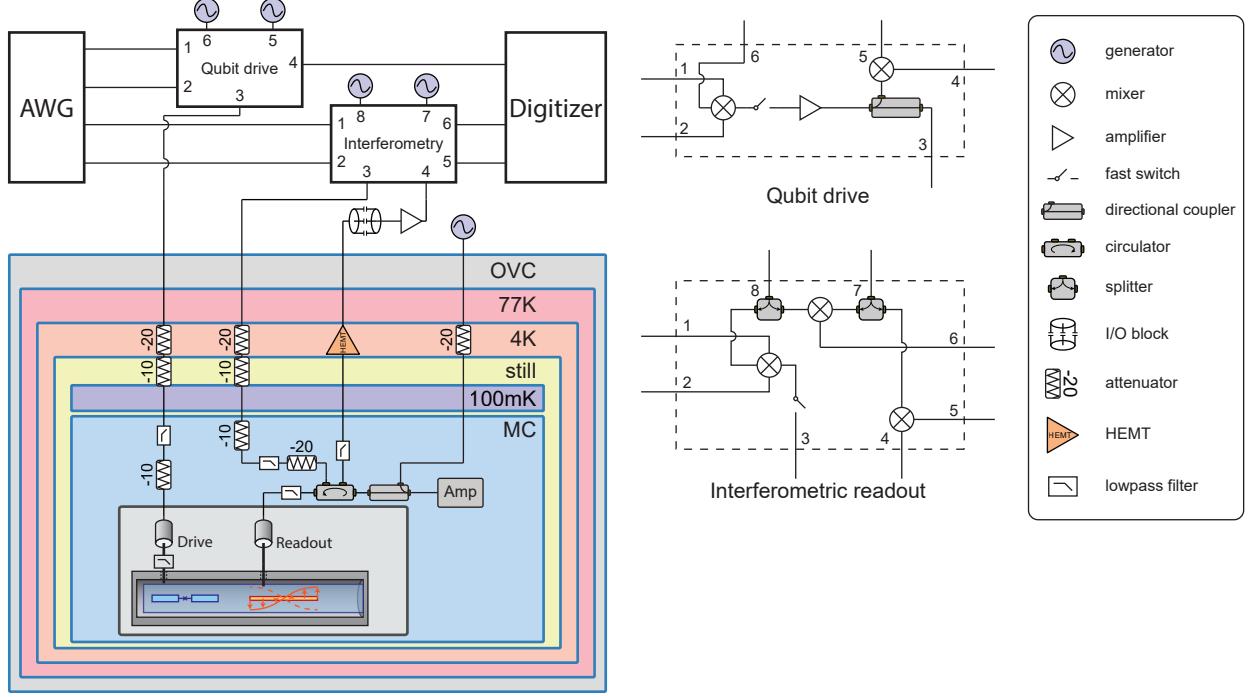


Figure S7. Fridge diagram of the sub-harmonic qubit drive experiment.

## B. Temperature stabilization

The performance of room temperature electronics depends on their operating status and environment. For example, the gain of a power amplifier changes as environment temperature changes. In monitoring our pulse properties, we found the operating temperature, which is regularly perturbed by air-conditioning cycles in the room air, has the most significant effect on our system performance. As shown in Figure S8(a), a clear dependence on devices' temperature can be observed. While a few percents of relative variation in power and phase can often be ignored in simple one or two pulse protocols, such as  $T_1$  or  $T_2$  measurements, it is a limiting factor for high-fidelity quantum control, and it is important for building a stable quantum computer without frequent pulse calibration. Also, this is especially important for sub-harmonic driving, since it is three times more sensitive to the fluctuation of the pulses' magnitude and phase.

Because of the periodic temperature change in the lab, we find about a  $1.5^\circ\text{C}$  change of the devices' temperature, measured from the built-in sensor of the RF generator. In Figure S8 (b), we see a temperature change with a period of approximately 4 hours, which is due to the air conditioner setting in the lab. Such fluctuation results in a 2% pulse magnitude variation. The stability of the electronics improved significantly after we reduce the heat transfer between the devices and the environment as well as use a PID controlled cooler to actively control devices' temperature. As shown in the Figure S8(c), all of the room temperature components, except for the AWGs and ADCs, are mounted on a 1/4 inch thick Al breadboard to spread the heat, and placed inside a Styrofoam wrapped box, which reduced the heat transfer rate between inside and outside. A commercial liquid to-air CPU cooler is mounted on the other side of the Al board, which is designed as the primary heat transfer channel. The heat transfer rate is controlled by the speed of the fans which flow over the cooler's radiator (placed in the room air). The fan speed is determined by a Arduino-based PID controller that records temperature from a digital thermometer mounted to the Al breadboard. The temperature sensor and the adjustable fan speed form a closed feedback loop that fix the box temperature to a desired set point.

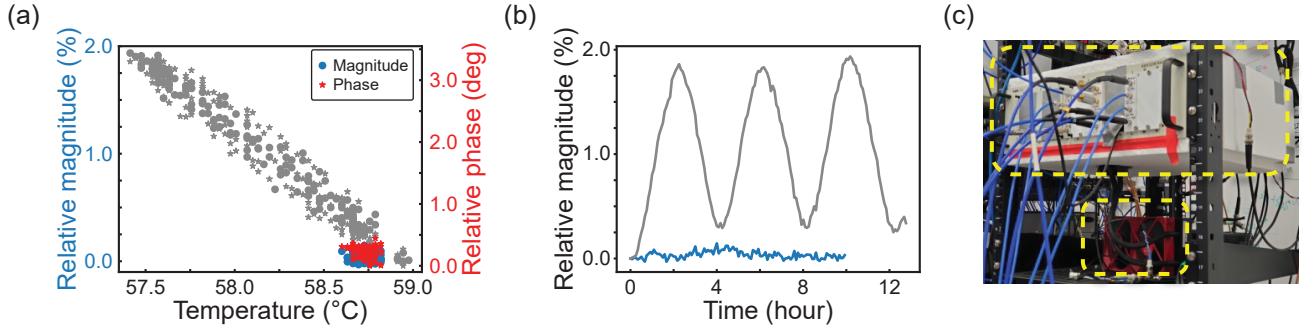


Figure S8. Temperature stability is especially important for sub-harmonic driving, as Rabi rate is cubic proportional to driving amplitude. (a) The amplitude stability vs temperature before (grey circles and stars) and after stabilization (blue circles and red stars). (b) The amplitude stability over time without stabilization (grey trace) and with stabilization (blue trace). The great improvement in pulse stability makes high fidelity control possible. (c) The experiment setup and a PID-controlled commercial liquid cooler are highlighted with yellow dashed boxes. The box are wrapped with Styrofoam except for front and back faces, which are left for connecting coaxial cables. The liquid cooling head are mounted on the Al plate from the bottom of the box.

### C. Pulse tune-up

To perform single qubit control, we need to be able to perform  $\pi$  pulses and  $\pi/2$  pulses in  $X$  and  $Y$  direction. For each gate, as shown in Eq. S12, there are four free parameters that controls the pulse shape: amplitude  $A_0$ , pulse length  $t_0$ , ramping speed  $k$ , cutoff factor  $t_1/k$ . Besides pulse shape, the frequency and overall microwave phase of the pulse also needs to be tuned up.

The pulse tune-up procedure starts from the Rabi experiment. The purpose of this step is to find a desired drive strength and the corresponding drive frequency. Figure 2 in the main paper is a summary of many Rabi experiments performed with different driving amplitudes. The full experimental results are shown in Figure S9. Each plot corresponds to a data point in Figure 2(b). It shows that the Rabi rate increases quickly as drive amplitude increases and drive frequency shifts get larger, corresponding well with theoretical calculations.

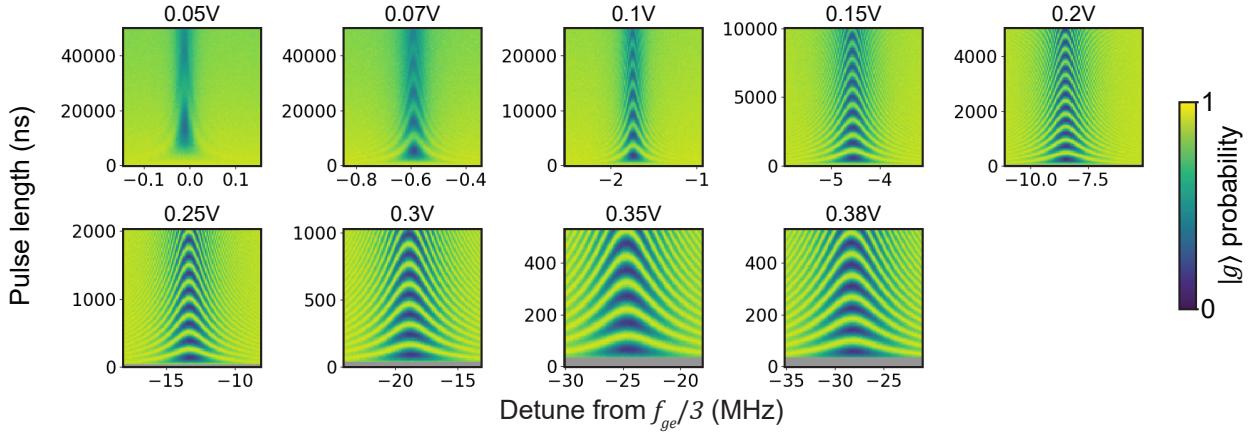


Figure S9. The sub-harmonic Rabi process under different driving amplitudes, ranging from 0.05 V to 0.38 V. The cubic dependence of Rabi rate on drive voltage can be observed (note the time axes vary by approximately 100x across the plots). The grey area in the plots means no data was measured at that region due to hardware limitations.

In the rest of the pulse tune-up procedure, the driving amplitude is fixed so that the drive frequency can be fixed. Also,  $k$  and  $t_1$  are fixed so that the pulse edges are the same shape for all pulses, which simplify the phase correction. There are only two remaining free parameters that need to be tuned up, gate length and gate phase. Gate length is different for  $\pi$  pulse and  $\pi/2$  pulse while the gate phase is the same.

The tune-up procedure is shown in the Figure S10. The length of a  $\pi$  pulse is chosen as the first step of calibration. This is because, ideally, qubit only goes to  $|g\rangle$  and  $|e\rangle$  state, which phase does not matter. So it can calibrated without phase correction. The  $\pi$ -pulse length is calibrated with the sequence shown in the inset of Figure S10(b): a  $\pi/2$ -pulse

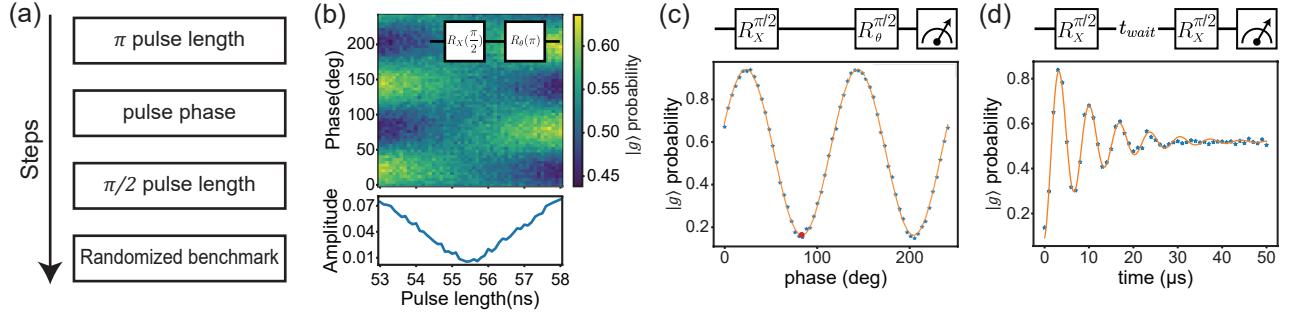


Figure S10. The sub-harmonic gates tune-up procedure. (a). Steps of tuning up sub-harmonic gates. The first three experiments calibrate individual pulse parameters, while randomized benchmarking measures the gate fidelity in the end. (b). The experiment to optimize length of the  $\pi$  pulse. When pulse length is too short or too long, the  $|g\rangle$  probability oscillates depending on the phase of the second pulse. (c) and (d) show the two sequences and experiment results for correcting drive-induced phase. (c) calibrates the time independent phase errors  $\varphi_{ramp}$ , where the red circle shows the optimal point. (d) calibrates time dependent phases errors  $\delta\omega$ .

over  $x$ -axis followed by a  $\pi$ -pulse with additional phase. This pulse sequence allows us to calibrate the length of the  $\pi$  pulse without correcting phase and the amplitude error of the  $\pi/2$  pulse. As shown in Figure S10(b), we swept the length of the first pulse and the phase of the second pulse. The pulse length error will affect the  $|g\rangle$  probability of the final state. As the phase of the second pulse changes, a sinusoidal oscillation of  $|g\rangle$  probability can be observed, and its amplitude is proportional to the pulse length error. A more accurate  $\pi$ -pulse length can be acquired by fitting the minimum of the oscillation amplitude. From this, we find the optimal length for  $\pi$ -pulse is 55.5 ns.

The second step is correcting phase error. As shown in Eq. 9, the phase error induced by the AC-Stark shift consists of two parts, a constant part  $\varphi_{ramp}$  and a time-dependent part  $\delta\omega t$  when pulses are not being applied. They are calibrated with the Ramsey experiment by varying different pulse parameters, as shown in Figure S10 (b) and (c). To find the constant part  $\varphi_{ramp}$ , two  $\pi/2$  pulse are placed with minimum time gap. The phase of the second pulse is swept to find the correct compensating phase. For the time-dependent part, the time gap between two pulses are swept instead. The frequency  $\delta\omega$  is found by fitting the oscillating frequency of the trace. This experiment allows us to calibrate the phase of both the  $\pi$  pulse and  $\pi/2$  pulses for arbitrary times in a sequence.

The third step is to calibrate the length of the  $\pi/2$  pulse. It is calibrated by repeating the same  $\pi/2$  pulse  $n$  times to amplify the rotation errors.

These tune-up procedures may be applied in iteration, each time scanning for finer control of the pulse parameters. More steps of calibration can be applied if DRAG is required. However, with currently achieved gate time and qubit coherence time, leakage to higher excited states is not the limiting factor of gate fidelity. Finally, we use interleaved randomized benchmarking to provide a quantitative measure for the fidelity of each pulse type.

- [1] M. D. Reed, *Entanglement and Quantum Error Correction with Superconducting Qubits*, Ph.D. thesis, Yale University (2013).
- [2] A. Blais, A. L. Grimsmo, S. M. Girvin, and A. Wallraff, Circuit quantum electrodynamics, *Reviews of Modern Physics* **93**, 025005 (2021).
- [3] D. C. McKay, C. J. Wood, S. Sheldon, J. M. Chow, and J. M. Gambetta, Efficient z gates for quantum computing, *Physical Review A* **96**, 022330 (2017).
- [4] F. Motzoi, J. M. Gambetta, P. Rebentrost, and F. K. Wilhelm, Simple pulses for elimination of leakage in weakly nonlinear qubits, *Physical review letters* **103**, 110501 (2009).
- [5] M. Pechal, Microwave photonics in superconducting circuits, PhD Thesis, ETH Zürich (2016).
- [6] J. Koch, M. Y. Terri, J. Gambetta, A. A. Houck, D. I. Schuster, J. Majer, A. Blais, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, Charge-insensitive qubit design derived from the cooper pair box, *Physical Review A* **76**, 042319 (2007).
- [7] J. Sakurai and J. Napolitano, *Modern Quantum Mechanics* (Addison-Wesley, 2011).
- [8] Here we also assume the system-bath coupling bandwidth  $\vartheta/\omega_0$  is around the order 0.1, which is of the same order of  $|\alpha|/\omega_0$ .