

Design and Characterization of a 28-nm Bulk-CMOS Cryogenic Quantum Controller Dissipating Less than 2 mW at 4 K

Joseph C. Bardin, *Senior Member, IEEE*, Evan Jeffrey, Erik Lucero, Trent Huang, Sayan Das, *Student Member, IEEE*, Ofer Naaman, Rami Barends, Ted White, Marissa Giustina, Daniel Sank, Pedram Roushan, Kunal Arya, Benjamin Chiaro, Julian Kelly, Jimmy Chen, Brian Burkett, Yu Chen, Andrew Dunsworth, Austin Fowler, Brooks Foxen, Craig Gidney, Rob Graff, Paul Klimov, Josh Mutus, Matthew McEwen, Anthony Megrant, Matthew Neeley, Charles Neill, Chris Quintana, Amit Vainsencher, Hartmut Neven, and John Martinis

Abstract—Implementation of an error corrected quantum computer is believed to require a quantum processor with on the order of a million or more physical qubits and, in order to run such a processor, a quantum control system of similar scale will be required. Such a controller will need to be integrated within the cryogenic system and in close proximity with the quantum processor in order to make such a system practical. Here, we present a prototype cryogenic CMOS quantum controller designed in a 28-nm bulk CMOS process and optimized to implement a 4-bit XY gate instruction set for transmon qubits. After introducing the transmon qubit, including a discussion of how it is controlled, design considerations are discussed, with an emphasis on error rates and scalability. The circuit design is then discussed. Cryogenic performance of the underlying technology is presented and the results of several quantum control experiments carried out using the integrated controller are described. The paper ends with a comparison to the state of the art. It has been shown that the quantum control IC achieves comparable performance with a conventional rack mount control system while dissipating less than 2 mW of total AC and DC power and requiring a digital data stream of less than 500 Mb/s.

Index Terms—ADD THEM

I. INTRODUCTION

QUANTUM computers have the potential to solve currently intractable problems in cryptography [?], machine learning [?], computational chemistry [?], and, among others, database searching [?]. Over the past decade, significant progress has been made towards the implementation of such a computer and the field is now at the point where small quantum processors with on the order of 50-to-100 quantum bits, or qubits, are being demonstrated [?], [?], [?], [?].

However, qubits are error-prone devices and, even when cooled to 10 mK., todays best devices still suffer from errors rates on the order of 0.1% per computational step. Performing error-free calculations using these noisy qubits will require the implementation of quantum error correction (QEC) codes and it is currently believed that approximately a million qubits will be required to implement an error protected quantum computer [1]. Scaling contemporary technology to these levels comes with many challenges.

While intense research in the field has pushed the performance of quantum processors beyond that required to achieve

fault tolerance, significant work is still required if a system with on the order of a million qubits is to be realized. In order to operate a quantum processor at these levels, a high performance quantum control and measurement system is required [2]. Todays, these systems are realized using racks of room temperature electronics, connected to the quantum processor through meters of lossy coaxial cable, as shown conceptually in Fig. 1(a) [3]. If quantum computing systems are to be scaled to the million qubit level, these systems must be monolithically integrated placed in close proximity to the quantum controller.

Multiple approaches towards the scaling of the quantum control and measurement system such that it can be integrated within the cryogenic system have been proposed. When operating at deep cryogenic temperatures, the use of mixed signal circuits based upon Josephson junctions becomes feasible. As such, authors have proposed digital control [4] and readout [5], with a large digital processor integrated at 4 K and drive/sensing circuitry co-located on the 10 mK stage of the system [6]. However, even for small-scale circuits, a significant impact to performance has been observed due to heating from the superconductive circuits heatsunk to the 10 mK stage of the system [4] and ongoing work seeks to overcome this limitation.

Other authors have proposed reducing the number of high speed interconnects from room temperature down to the quantum processor by introducing a semiconductor switch matrix at the 10 mK stage in order to multiplex RF control signals [7]. However, this approach also has the problem of introducing dissipation at the base temperature.

Finally, significant work is currently underway towards implementation of the scheme shown conceptually in Fig. 1(b) [8], [9], [?]. In this approach, the quantum-to-classical interface is integrated in one or more (Bi)CMOS ICs and located on the 4-K stage of the system. The use of commercial (Bi)CMOS technologies provides a significant advantage due to the high manufacturing yields and mature design infrastructure. Moreover, 4 K is a particularly convenient temperature, since it is feasible to cool devices dissipating watts of power. Even so, this is a major challenge, as the cryo ICs must be high performance and very low power (eventually < 1 mW/qubit). Preliminary work in this area has focused on

circuit blocks [10], [11], [12] and transistor modeling [10], [13], [14], both of which are steps towards the implementation of integrated cryogenic quantum control and measurement systems.

In this paper, we present the design and implementation of one part of a cryogenic quantum control and measurement system: a CMOS IC that has been designed for operation at 3-K and optimized to perform XY gate operations on transmon qubits. The paper is organized as follows.

- 1) A brief introduction to qubits and quantum control is presented. The fundamentals of transmon qubits are described with an emphasis on microwave control of these devices.
- 2) Design considerations are described. Requirements associated with quantum error correction and scalability are emphasized.
- 3) The design of a cryogenic CMOS quantum control circuit is presented along with a discussion of particular challenges associated with design for operation at deep cryogenic temperatures.
- 4) Experimental results are presented. The temperature dependence the underlying transistor technology is evaluated. A series of quantum control experiments are carried out to validate the efficacy of the implemented quantum controller.
- 5) The results are compared to the state of the art and future steps are discussed.

II. QUANTUM COMPUTING AND QUANTUM CONTROL

The fundamental building block in a quantum computer is the quantum bit, or qubit. A qubit is a two-level quantum mechanical object whose instantaneous state can be described as a superposition of its two basis states,

$$|\psi\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + \exp\{j\phi\}\sin\left(\frac{\theta}{2}\right)|1\rangle. \quad (1)$$

The state of a qubit has a unique interpretation as a point on the surface of a unit sphere, which is referred to as the Bloch sphere. As shown in Fig. 2, the north and south poles of the Bloch sphere correspond to the $|0\rangle$ and $|1\rangle$ states respectively, whereas all other points on the surface of the Bloch sphere correspond to unique superposition states.

This geometric picture provides insight into the control and measurement of single qubits. The act of measuring the qubit will cause its state to collapse to either its $|0\rangle$ or $|1\rangle$ basis state and the probability of these two outcomes depends solely on the elevation angle, θ : $P\{|0\rangle\} = \cos^2(\theta/2)$ and $P\{|1\rangle\} = \sin^2(\theta/2)$. Single qubit gate operations are represented by 2×2 unitary matrices and can be interpreted as deterministic rotations of the Bloch vector on the surface of the Bloch sphere. Thus, while ϕ has no impact on measurement results, it does impact the effect of gate operations on the qubit.

To explain how single qubit gate operations are enabled by the physics of quantum mechanics, we begin by pointing out

that a qubit is a quantum mechanical object and, as such, obeys the time dependent Schrödinger equation [15],

$$\frac{\partial}{\partial t}|\psi(t)\rangle = -\frac{j}{\hbar}\hat{H}|\psi(t)\rangle, \quad (2)$$

where \hbar is the reduced Planck's constant and \hat{H} is the Hamiltonian (total energy) operator, which for a single qubit is a 2×2 matrix. As such, the time evolution of the qubit state is completely determined by its Hamiltonian. Thus, deterministic control of the state of a quantum system requires the ability to systematically enable and disable components of the system's Hamiltonian. In Section II-B, we will explain how this is carried out for transmon qubits. However, first we describe these devices.

A. The Flux-Tunable Transmon Qubit

A schematic diagram of a flux-tunable transmon qubit, complete with XY-drive, Z-drive, and readout ports is shown in Fig. 3. The intrinsic qubit consists of capacitor C_Q in parallel with a Josephson junction loop. A Josephson junction is a superconducting tunnel junction and can be thought of as a nonlinear inductance of the form $L_J = L_{J0}/\cos(2\pi\Phi/\Phi_0)$, where $\Phi = \int v_{JJ}dt$ is the flux across the junction, $\Phi_0 = h/2q$ is the flux quanta, h is Planck's constant, q is the charge of an electron, $L_{J0} = \Phi_0/2\pi I_C$, and I_C is the critical current of the junction and is proportional to junction area [16]. A Josephson junction loop, or SQUID, can be thought of as a composite Josephson junction, with flux-tunable critical current, $I_{C,SQUID} = 2I_{C0}|\cos(\pi\Phi_E/\Phi_0)|$, where I_{C0} is the critical current of each of the (identical) junctions and Φ_E is the external flux used to tune the effective critical current [17]. As such, the flux-tunable transmon qubit can be thought of as a frequency tunable non-linear LC resonator.

If cooled to temperatures such that thermal population of the non-linear resonator is suppressed, a transmon behaves quantum mechanically. For typical qubit frequencies in the vicinity of 5 GHz, the effective photon temperature ($T_{\text{Photon}} = hf/k$) is approximately 250 mK. As such, these devices are typically cooled to the 10 mK range to ensure that thermal excitations are sufficiently suppressed¹.

The eigenenergies of a transmon can be found from the Hamiltonian of the isolated qubit [18]:

$$E_n \approx \hbar\omega_0\left(n + \frac{1}{2}\right) - \frac{E_C}{12}(6n^2 + 6n + 3). \quad (3)$$

Here, the index n is an integer greater than or equal to zero, $\omega_0 = 1/\sqrt{L_{J0}C_Q}$, and $E_C = q^2/2C_Q$ is the energy required to add one electron to the charge on the capacitor. The corresponding energy diagram appears in Fig. 4. In contrast to the ideal qubit described above, the transmon has an infinite number of energy levels. As such, care must be taken to ensure that only the lowest two energy levels are occupied. Since it is only possible to populate higher levels if the $|2\rangle$ state is

¹This may make it seem advantageous to operate at much higher frequencies in hope of reducing the challenges associated with cryogenic cooling. However, there are other disadvantages to operating at higher frequency including decreased coherence time and more challenging specifications for the control electronics.

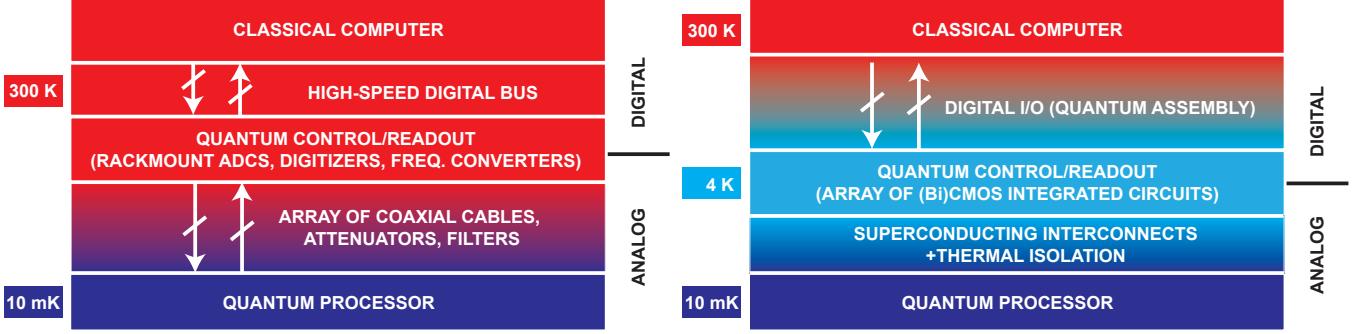


Fig. 1. High level picture of current and proposed approaches to interfacing a quantum processor to a classical computer.

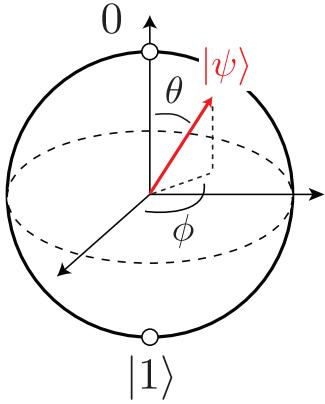


Fig. 2. Bloch sphere representation of a qubit.

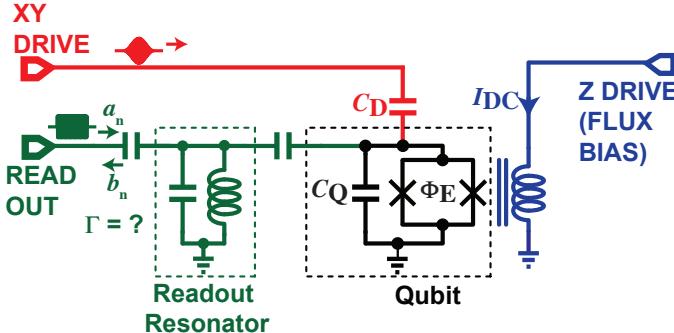


Fig. 3. Schematic diagram of transmon qubit including XY drive, Z drive, and readout ports. The “X” symbols represent Josephson junctions.

already populated, we will constrain our control signals such that the $|2\rangle$ state is not populated and neglect higher levels in the following discussion.

The XY drive port shown in Fig. 3 can be used to excite state transitions. The excitation frequencies corresponding to the first two transitions can be determined directly from Equation (3):

$$\omega_{01} = \frac{E_1 - E_0}{\hbar} = \omega_0 - E_C/\hbar \quad (4)$$

and

$$\omega_{12} = \frac{E_2 - E_1}{\hbar} = \omega_0 - 2E_C/\hbar \quad (5)$$

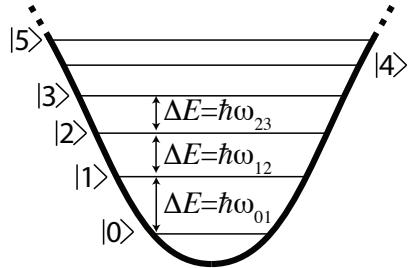


Fig. 4. Energy diagram for transmon qubit. The non-linearity of the Josephson junction creates anharmonicity.

From this, we can determine the anharmonicity, or difference between the ω_{12} and ω_{01} transitions:

$$\eta = \omega_{12} - \omega_{01} = -E_C/\hbar = -\frac{q^2}{2\hbar C_Q}. \quad (6)$$

This metric is significant since it bounds the spectral width of microwave excitation pulses, and hence sets a minimum duration for XY gate operations. Ideally, the anharmonicity (or qubit nonlinearity) would be as large as possible in order to enable fast gates. However, there is a trade-off as larger capacitances help to suppress decoherence due to $1/f$ charge noise [18]. Typical values for η are in the 150–350 MHz range, with the highest coherence time devices having values towards the lower end of this range.

A projective measurement of the transmon state can be accomplished through a reflection measurement of an ancillary linear resonator that is capacitively coupled to the qubit (see Fig. 3). Assuming the readout resonator frequency is below that of the qubit, the coupling between the resonator and the qubit will cause a dispersive frequency shift up or down in frequency if the qubit collapses to the $|0\rangle$ or $|1\rangle$ states, respectively. If the readout resonator is interrogated at its bare frequency—that is, its resonant frequency if the coupling to the qubit were removed—the dispersive frequency shift will produce a state-dependent phase shift on the interrogation signal, which can be detected and used to determine the state of the qubit.

B. XY Control of a Transmon

Assuming a voltage drive referenced to the input of C_D and of the form

$$v_D(t) = a(t) \sin(\omega_D t - \phi_D), \quad (7)$$

the Hamiltonian of the driven circuit in the rotating frame of the drive signal² can be written [19]

$$\hat{H}_D \approx g \frac{a(t)}{2} (\cos(\phi) \hat{\sigma}_X + \sin(\phi) \hat{\sigma}_Y) - \frac{\hbar \Delta \omega}{2} \hat{\sigma}_Z, \quad (8)$$

where $g = (C_D/C_Q) \sqrt{\hbar/2Z_Q}$ is the drive coupling strength, $Z_Q = \sqrt{L_{J0}/C_Q}$ is the qubit impedance, $\Delta\omega = \omega_D - \omega_0$ is the offset in drive frequency from the resonant frequency of the qubit, and

$$\hat{\sigma}_X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \hat{\sigma}_Y = \begin{bmatrix} 0 & -j \\ j & 0 \end{bmatrix}, \quad \hat{\sigma}_Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad (9)$$

are the Pauli spin matrices.

It can be shown that this Hamiltonian produces a time evolution corresponding to a rotation of the qubit state about a vector

$$\begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} = \frac{1}{\sqrt{g^2 a^2(t) + \hbar^2 \Delta \omega^2}} \begin{bmatrix} g a(t) \cos(\phi_D) \\ g a(t) \sin(\phi_D) \\ -\hbar \Delta \omega \end{bmatrix}, \quad (10)$$

and at a frequency **Check**

$$\omega_r = \frac{\sqrt{g^2 a^2(t) + \hbar^2 \Delta \omega^2}}{\hbar}. \quad (11)$$

Referring to equations (10) and (11), we can identify three distinct control knobs, $a(t)$, ϕ_D , and $\Delta\omega$. Moreover, it is apparent that $\Delta\omega$ affects the Z component of the rotation axis, whereas the other two degrees of freedom affect its XY projection.

Since it is desirable to only induce rotations about an axis in the XY plane of the Bloch sphere (Z rotations can be carried out in software [20]), the qubit should nominally be driven on resonance³ ($\Delta\omega = 0$). In this case, the axis of rotation is defined by the carrier phase, $\phi_R = \phi_D$, and the angle of rotation is proportional to the integrated envelope amplitude

$$\theta_R = \frac{g}{\hbar} \int_{t_0}^t a(t') dt', \quad (12)$$

as defined in Fig. ??.

A block diagram of a standard control and readout system appears in Fig. 5 with the XY control portion highlighted. The XY signals are generated using a pair of 1 Gs/s 14-bit DACs whose outputs drive the LO ports of an IQ mixer. Single sideband mixing is typically employed, with LO and RF signals in the 4–8 GHz frequency range. The upconverted output is heavily attenuated at both the 3-K and 10-mK stages

²This is similar to a baseband analysis. Here, we have rotated the coordinate system at the carrier frequency to remove dynamics and have assumed that higher frequency components will average out through the integration that occurs when the Schrödinger equation is solved.

³It should be noted that the more complete description including $\Delta\omega$ above allows describing the behavior that occurs if the drive signal is slightly detuned from the qubit frequency.

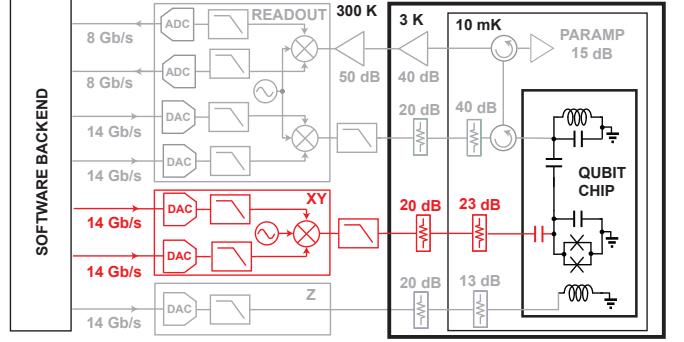


Fig. 5. Block diagram showing standard control and readout hardware, with XY control hardware highlighted.

of the cryogenic system. In the remainder of the paper, we consider simplifying and integrating this system in a manner such that it can be placed on the 3-K stage of the cryostat.

III. XY CONTROLLER REQUIREMENTS

Techniques to accurately control and measure the state of a quantum processor with on the order of a million qubits will be required to enable fault tolerant quantum computing [?]. Here we summarize some of the key design specifications of the XY controller.

A. Physical Temperature

We first consider the physical temperature at which to thermalize the quantum controller, as this is critical in driving other specifications. On the one hand, it is desirable to place the controller as close as possible to the quantum processor so as to minimize interconnects. In fact, if there were no other considerations, it would be appropriate to co-integrate the classical control with the quantum chip and researchers are currently exploring both 3D packaging approaches [21] and integrated control and readout architectures [4], [5], [6] as steps to make this a reality.

However, as discussed above, superconductive quantum processors must be thermalized to temperatures in the vicinity of 10 mK. Unfortunately, the efficiency of cryocoolers is fundamentally limited by the laws of thermodynamics to $\eta \leq T_{\text{BASE}} / (T_A - T_{\text{BASE}})$, where T_{BASE} and T_A are the base and room temperatures, respectively [22]. So, the efficiency in removing heat at 10 mK cannot be more than about 0.003%. In reality, the efficiency of the state-of-the-art dilution refrigerators employed for cooling to 10 mK is orders of magnitude lower than the theoretical Carnot efficiency. In fact, typical systems can handle less than 15 μW at 10 mK despite drawing 5-10 kW of wall power [?]. As such, thermalization of CMOS integrated control electronics to the same stage as the quantum IC is not viewed as a viable approach.

At the other extreme, we could consider placing the quantum controller at room temperature, as done today. The advantage here is that power consumption of the electronics is not critical and existing electronics could be miniaturized and scaled. However, the challenge of interconnects between the

room temperature electronics and the 10 mK stage, which must be able to support microwave transmission with high isolation while preventing as low of a thermal load as possible, is not believed to be scalable to the levels required to control a quantum processor with a million or more qubits.

A more tractable solution is to thermalize the control electronics in the range of 3-to-4 K and use superconductive transmission lines to couple the control signals down to the 10 mK stage [7], [9]. By using superconducting interconnects, it is possible to achieve excellent electrical and thermal performance in a small cross sectional area [23]. Below, we will assume that the quantum control electronics are thermalized in the range of 3-to-4 K.

B. Power Consumption

While much higher than at 10 mK, the power handling capability of the 4 K stage of a typical closed-cycle refrigeration system is still limited. For near-term systems, which will employ standard pulse-tube cryocoolers, the total dissipation at 4 K will be limited to just 2 W at 4 K [24]. As this power budget must include any thermal load associated with cabling, the actual power available for electronic consumption is lower. As such, a reasonable near term power budget for the control and readout electronics is 1 W.

However, for a scaled system, one could envision building a much larger refrigerator employing more advanced cooling techniques. For instance, helium liquifiers can be used to increase the capacity to around 1 kW [25], or about 1 mW/qubit at the million qubit level. As this power consumption must be shared between any incoming cabling (from 300 K), Z-control electronics, and the readout electronics (which can be multiplexed by 5-10x), a reasonable long term limit for the dissipation of the control electronics is 250 μ W per qubit. For this proof of concept work, which is being carried out without cryogenic simulation models, we relax this specification by a factor of 10x to 2.5 mW, with the expectation that the power consumption can be further optimized in a future design.

C. Frequency Range

Typical transmon qubits have transition frequencies in the 4–8 GHz frequency range. As such, we seek to cover this full band.

D. Gate Fidelity and Error Rates

The remainder of the specifications are all performance related and, as such, a metric to benchmark performance is required. Fidelity and error rates provide such metrics. Gate fidelity is a measure of how close the effect of an applied gate is to that of the desired unitary and can be limited by a number of factors including control errors, noise, decoherence, etc. Formally, we can define the gate fidelity of a single qubit gate, averaged over all pure input states as [26]

$$\mathcal{F}_G \equiv \frac{\text{Tr} \left\{ U_{\text{Ideal}}^\dagger U U^\dagger U_{\text{Ideal}} \right\} + \left| \text{Tr} \left\{ U_{\text{Ideal}}^\dagger U \right\} \right|^2}{6}. \quad (13)$$

The first term, $\text{Tr} \left\{ U_{\text{Ideal}}^\dagger U U^\dagger U_{\text{Ideal}} \right\}$ evaluates to two in the case that the realized operator is unitary, but is smaller in the case where the matrix is not unitary due to matrix truncation (e.g., if there is leakage to the $|2\rangle$ state). In a computation, the average error per step is simply $1 - \mathcal{F}_G$.

The degree of redundancy required for fault tolerant quantum computing is strongly dependent up on error rates [1]. In order to reduce this to practical levels, the error rates should be reduced as far as possible. Here, we target an overall error rate of 0.01%, which is consistent with state of the art performance [?]. To achieve this performance, we will limit individual error contributors to an order of magnitude lower error rates.

E. Noise on the Drive Line

Noise at the qubit transition frequency ($w_{01} = 2\pi f_{01}$) causes decoherence. An available noise power spectral density of $S_a(f_{01}) = kT_e$ at the drive port at frequency f_{01} leads to transitions between the $|0\rangle$ and $|1\rangle$ states at a rate [27]

$$R_{\uparrow\downarrow} = 2\pi \frac{T_e(f_{01})}{T_{\text{Photon}}} \Delta f, \quad (14)$$

where $\Delta f = f_0/Q_D$, $Q_D = (C_D/C_Q)^2 (Z_Q/Z_0)$ is the quality factor of the qubit due to the drive circuit, $T_{\text{Photon}} = \hbar\omega_{01}/k$ is the effective temperature of a photon at the qubit frequency, and Z_0 is impedance of the drive port. The noise spectral density $S_a = kT_e$ must be kept low enough such that $R_{\uparrow\downarrow}$ is less than other decoherence channels in the system, such as losses from materials imperfections. This usually means ensuring that the noise temperature seen looking back into the drive line is kept well below the effective temperature of a photon at the drive frequency. As such, it is important both to include attenuation on the drive line at 10 mK and also to minimize noise generated by the XY controller.

F. Spectral Content

Also related to gate fidelity is the spectral content of the qubit drive signal. Due to the finite coherence time of a qubit (on the order of 10-to-100 μ s for transmon qubits) fast gates are desired. However, due to the finite separation between the ω_{01} and ω_{12} transition frequencies, it is important to ensure that the drive signal does not introduce leakage to higher order levels. In optimizing the control waveforms, one has to carefully balance the trade-off between speed and leakage. Typical pulse envelopes include Gaussian and raised cosine shapes and advanced techniques such as derivative removal by adiabatic gate (DRAG) are used regularly to further reduce the pulse duration beyond what can be achieved using envelope shaping alone [19], [28]. In this work, we employ simple raised cosine pulse shaping.

G. Amplitude Control

The drive amplitude sets the range of rotations that can be carried out for a given pulse duration. As discussed in Section II-B, a drive pulse on resonance with the qubit produces a rotation of proportional to the integrated envelope

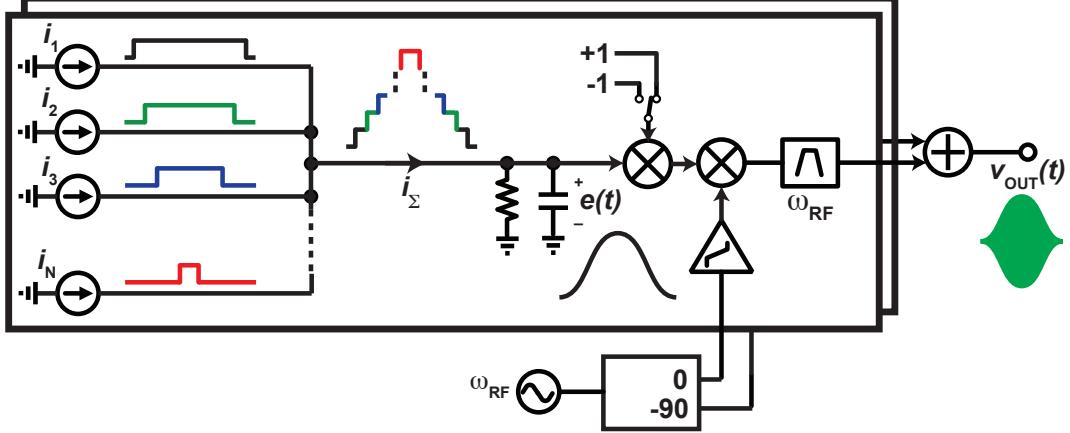


Fig. 6. Block diagram of the proposed pulse generation approach.

amplitude (see Equation (12)). The largest rotation required in a practical quantum algorithm is a π -pulse. For a raised cosine envelope, the peak amplitude required to achieve this rotation is

$$A_{PK,\pi} = 2\pi \frac{\hbar}{g\tau_g}. \quad (15)$$

For typical values of g , the required peak amplitude for a 15 ns pulse is on the order of $100 \mu\text{V}$ (peak available power of -76 dBm). In general, we want to be able to drive with pulses several times larger to allow both shorter π -pulses and also to carry out rotations well beyond 180° . As such, it is desirable that the XY controller be able to drive pulses on the order of 1 mV at the reference plane of the qubit drive port. The signal at 4 K must be significantly larger to account for the attenuation on the drive line that is required at 10 mK to reduce the thermal noise floor.

Amplitude resolution is also important, since any error in pulse amplitude will translate to an over or under rotation and contribute to gate error rates. An error in a rotation of $\Delta\theta_R$ produces an average error rate of

$$\epsilon_{\Delta\theta_R} \approx \frac{\Delta\theta_R^2}{6}. \quad (16)$$

To keep error rates due to envelope energy errors to below 10^{-5} , it is necessary that XY rotation angles are controlled to better than 0.5° . This means that the integrated π -pulse amplitude must have an accuracy better than 0.15%. A resolution of 11 bits in the integrated envelope amplitude is required to provide the required dynamic range.

H. Carrier Phase Control

A carrier phase error of $\Delta\phi_D$ on an XY gate will lead to a rotation around the wrong axis, resulting in an average gate error rate of

$$\epsilon_{\Delta\phi_D} = \frac{2}{3} \Delta\phi_D^2 \sin^2\left(\frac{\theta_R}{2}\right) \left(1 - \frac{\Delta\phi_D^2}{4} \sin^2\left(\frac{\theta_R}{2}\right)\right). \quad (17)$$

For small phase errors, it can be shown that the worst case average error rate will occur for π -pulses. To ensure the contribution of phase to the error rate is below $1e^{-5}$ for all XY

rotations, it is necessary that the carrier phase be controlled to better than 0.25° .

I. ON/OFF Ratio

Carrier leakage causes initialization errors and reduces gate fidelity. This effect is most significant for direct conversion pulse modulation architectures for which the leakage tone is on-resonance with the qubit. To understand the effect on initialization, we consider the case in which the qubit is reset to the ground state at $t = 0$ and idles for a time period of ΔT prior to being intentionally driven. Such a delay is typical to permit settling of the qubit after initialization. Under this scenario, a leakage tone with amplitude A_L produces a persistent XY rotation, resulting in a time dependent $|1\rangle$ state population due to leakage of $P_L\{|1\rangle\} = \sin^2(gA_L\Delta t/2\hbar)$, which for small enough $A_L\Delta t$ products simplifies to $P_L\{|1\rangle\} \approx (gA_L\Delta t/2\hbar)^2$. So, defining the specification for ON/OFF ratio in terms of the peak pulse amplitude required for a π -pulse ($A_{PK,\pi}$) and assuming raised cosine pulse shaping, we find

$$R_{\text{ON/OFF}} \equiv \frac{A_{PK,\pi}}{A_L} \approx \pi \frac{\Delta t}{\tau_g} \frac{1}{\sqrt{P_L\{|1\rangle\}}}, \quad (18)$$

where τ_g is the gate time and $P_L\{|1\rangle\}$ is the initialization error due to leakage. As a numerical example, if $T_g = 15 \text{ ns}$, achieving an initialization error of 0.01% with a settling time of 500 ns requires an ON/OFF ratio of 80 dB.

Another important metric is the $|1\rangle$ state population which is driven by carrier leakage over the course of a coherence time, an upper limit for which can be determined by substituting T_1 for T_g in equation (18). A reasonable limit to this value is 5% for a typical T_1 value of $20 \mu\text{s}$. Reaching this value requires a slightly higher ON/OFF ratio of 85 dB. While this is a stringent specification, it should be feasible to meet this using modern cancellation techniques due to the narrowband nature of the carrier leakage tone [?].

J. Summary

The performance specifications of the XY controller are summarized in Table I.

TABLE I
KEY SPECIFICATIONS FOR CRYOGENIC XY CONTROLLER

	Phys. Temp.	RF Frequency	Envelope	Env. Amp	Pulse Duration	Amp. err.	Phase err	Inst. Set	AC+DC Power
Prototype Goal	3-to-4.2 K	4-to-8 GHz	Symmetric	>10 mV	10-to-30 ns	< 0.15%	< 0.5°	4-bit	<2.50 mW
Long-term Goal	3 K	TBD	DRAG	> 1 mV	<10 ns	< 0.15%	< 0.5°	TBD	<250 μW

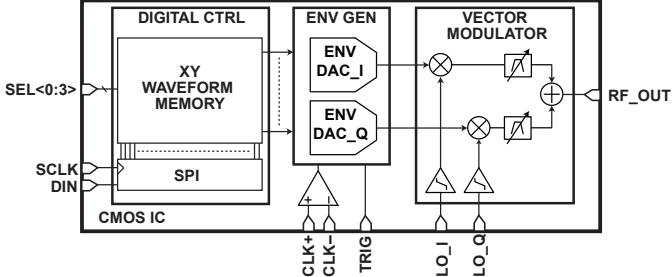


Fig. 7. Chip block diagram

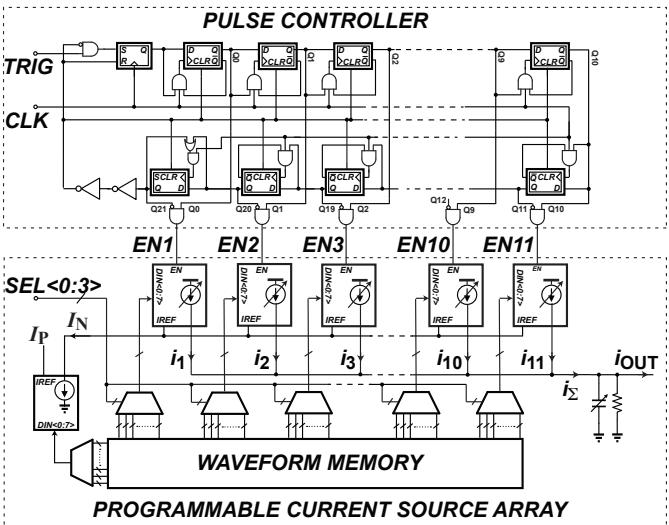


Fig. 8. ED

IV. CIRCUIT DESIGN

A conceptual block diagram explaining the waveform generation approach appears in Fig. 6. A vector modulation approach has been chosen to enable simultaneous control of the envelope amplitude and carrier phase. Symmetric envelopes are generated using an array of current sources which are sequentially enabled and disabled with the appropriate timing to generate a symmetric staircase of current, as shown in Fig. 6. The current waveform is then lowpass-filtered to create a smooth envelope. By controlling the individual weights of each of the current-mode sub-DACs, a wide range of symmetric envelopes can be realized. After low-pass filtering, the envelope currents pass through a polarity switch, are up-converted to the carrier frequency, and the two quadratures are then combined to create the output pulse.

This architecture has been chosen as a trade-off between power consumption, performance, and robustness to the inher-

ent uncertainty associated with designing for operation at 3-K, where foundry design models are currently unavailable. Several research groups are currently investigating the cryogenic performance of nanometer CMOS technology (e.g., (e.g., [13], [14]) and it is known that threshold voltages, transconductances, and subthreshold slope all increase with cryogenic cooling. Unfortunately, so does mismatch, particularly in the sub-threshold regime [29]. A critical feature of the envelope generation approach is guaranteed monotonicity, even if the sub-DACs display non-linearity or non-monotonicity.

A block diagram of the implemented IC appears in Fig. 7. The circuit consists of a pair of baseband envelope DACs, which drive a vector modulator. The clock port of the DACs as well as the LO ports of the vector modulator are buffered to minimize the external RF power required to drive the chip. A small digital memory is incorporated to store configuration parameters and implement a rudimentary 4-bit XY instruction set. Finally, a trigger line is included to initialize a pulse.

A schematic diagram of one of the envelope DACs appears in Fig. 8. The circuit contains a bank of eleven 8-bit current-mode sub-DACs, whose outputs are combined in the current domain and lowpass filtered before being fed to the modulator stage. Each DAC is constructed using a binary-weighted current mirror circuit, with the smallest unit cell being a minimum feature-size transistor. While linearity and monotonicity could be improved by using larger transistors and thermometer coding, respectively, these properties were sacrificed to reduce dynamic power consumption. The bank of DACs share a common reference current, which has 8-bits of programmability. The configuration bits for each of the sub-DACs as well as the master reference current for the sub-DACs (I_N) are configured by a sixteen waveform configuration memory, with the desired waveform selected by the 4-bit instruction interface ($SEL < 0 : 3 >$).

A current waveform of the form shown in Fig. 6 is generated by enabling each of the 11 current sources in the appropriate sequence. The enable signals for this operation are generated by the pulse controller shown in Fig. 8. The pulse controller is a shift-register based design in which the enable signal propagates through the shift register after a trigger pulse has been detected. In order to minimize the power consumption of this block, the clock signal is gated such that each flip-flop is only clocked if its output is to transition during a given clock cycle. The controller was designed to accept clock signals in excess of 2 GHz and produces pulses that are 21 clock cycles in duration.

A schematic diagram of the passive-mixer-based vector modulator appears in Fig. 9. A passive mixer has been selected to avoid introducing noise of a non-thermal nature during the mixing process and the circuit has been designed as a trade-off

between power consumption and bandwidth. Each LO signal is transformer coupled to a fully differential gain block which helps to improve common-mode rejection. The differential output of this cell is further amplified by a standard cell based pseudo differential amplifier chain which is sized to drive the mixer. The LO paths were optimized to be driven with LO signals below $50\ \mu\text{W}$.

The IF currents from each envelope DAC are routed through one arm of a polarity switch—required to achieve a full 360° of carrier phase control—through the mixer, and back through the other arm of the polarity switch back to ground. The output of the mixer is transformer coupled to the output, where the two quadratures are combined in the voltage domain. Finally, all three transformers feature bondpad side capacitive tuning networks which can be used to optimize coupling over the full 4–8 GHz frequency range.

V. RESULTS

The pulse modulator was implemented in a 28-nm CMOS technology and a die photo appears in Fig. 10. The chip was packaged in a module for testing. As shown in Fig. ??, the die was mounted within a cutout in a printed circuit board so as to minimize bondwire lengths. A photograph of the module appears in Fig. ??.

A. Technology Characterization

Prior to characterization of the IC, test structures from each transistor family from the 28-nm CMOS process were characterized on-wafer at both 300 K and 7 K.

B. Room Temperature Characterization

Prior to cryogenic measurements, the chip was first evaluated at room temperature to ensure that it was operating as expected.

C. Quantum Control Experiments

The low-temperature operation of the chip was evaluated through a series of quantum control experiments. To facilitate these measurements, the packaged IC was mounted on the 3-K stage of a commercial dilution refrigerator and interfaced to a qubit mounted on the 10 mK stage of the system. The qubit employed for this experiment was transmon that is part of a 5-qubit processor and is characterized by an anharmonicity parameter, $\eta/2\pi \approx 330\ \text{MHz}$. For all measurements, the other four qubits were tuned to be as far away in frequency as possible so as to make their presence inconsequential.

A detailed block diagram of the test setup appears in Fig. 12. The module containing the CMOS IC was mounted on the 3-K stage of the dilution refrigerator, along with 90- and 180-degree hybrids, which were used to interface to the LO and clock ports of the chip, respectively. The LO and clock signals were generated at room temperature using commercial synthesizers.

The LO signal was split at room temperature to enable feedforward cancellation of LO leakage to the RF output port of the cryo-CMOS IC module. Vector modulation of the

leakage cancellation tone was achieved using a digital step attenuator followed by a digital phase shifter. The amplitude and phase shifted cancellation signal was then combined RF output using a 20-dB coupler, mounted on the 3-K stage of the cryostat. In concert with a power combiner added after the phase shifter, this coupler also provided a mechanism to inject a standard XY control signal, thereby enabling baseline measurements.

A second 20-dB coupler was employed at the 3-K stage to couple out a small fraction of the RF output back to room temperature. This weak monitor signal was further amplified by approximately 50 dB (not shown) such that it could be viewed on an oscilloscope for debugging purposes.

Finally, the RF output was attenuated by 40 dB and connected to the XY port of the qubit. The Z and readout terminals of the qubit were connected to standard Z and readout electronics chains to complete the analog portion of the measurement setup.

The SPI port of the chip was interfaced to through a microprocessor, allowing for simple integration into a software stack. Finally, the trigger and word select lines of the CMOS IC were interfaced to using a bank of AWGs, configured to operate with binary outputs. All instruments were under the control of a standard qubit software stack, with the necessary drivers added to incorporate the CMOS IC within a standard experimental flow.

Once the setup was complete, the system was cooled to base temperature and a series of experiments were employed to evaluate the performance of the IC and compare it to that of conventional room temperature control electronics. These measurements are detailed below. The CMOS IC was powered down for all baseline measurements.

1) LO Leakage Cancellation: Prior to performing quantum control experiments, we first nulled the LO leakage using the cancellation path shown in Fig. 12. The experiment we performed to minimize LO leakage is explained in Fig. ???. The qubit was first initialized to the ground state and then allowed to idle for a period, τ , before its state was measured⁴. The magnitude and phase of the cancellation tone were varied to minimize the measured $|1\rangle$ state population. The resolution employed for amplitude and phase control were 0.25 dB and 1° , respectively, and we found that the nominal settings depended strongly on the impedances presented by the LO and RF ports of the IC (which could be varied through the transformer tuning capacitors). However, we also determined that these values were repeatable and stable on timescales of several days.

Example results in which the cancellation is turned OFF and ON are shown in Fig. 13. With the cancellation disabled, Rabi oscillations were observed with a period of approximately 300 ns, indicating the presence of significant LO leakage. This is not surprising since the RF and LO bonds were

⁴It may seem like a more direct approach one could take to null the LO feedthrough is to look at it on a spectrum analyzer, taking advantage of the monitor port that we have built into our test setup. However, we found that this was actually an ineffective method that did not produce nominal settings. We believe this is either due to the finite directivity of the directional coupler or secondary paths through which the LO signal leaked to the spectrum analyzer.

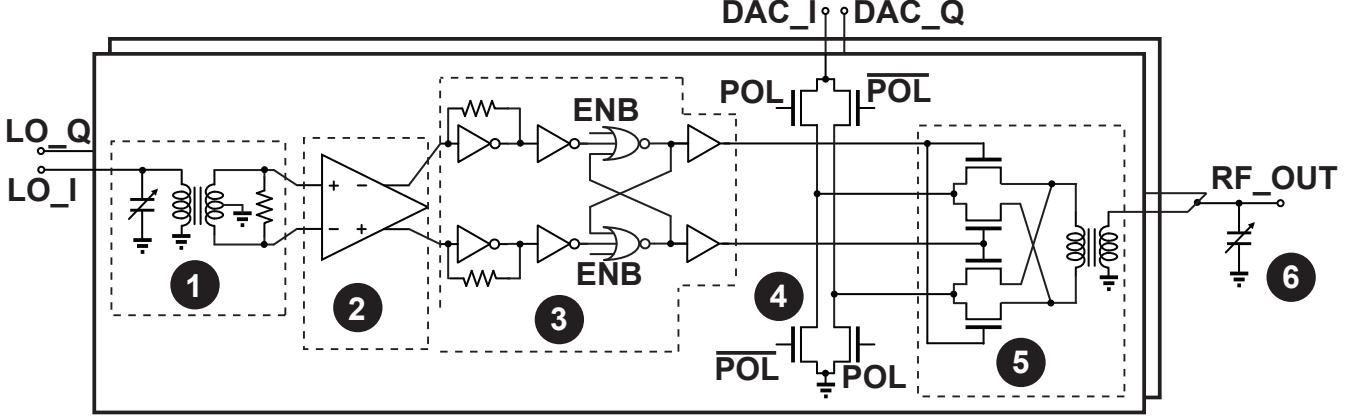


Fig. 9. Schematic diagram of the vector modulator.

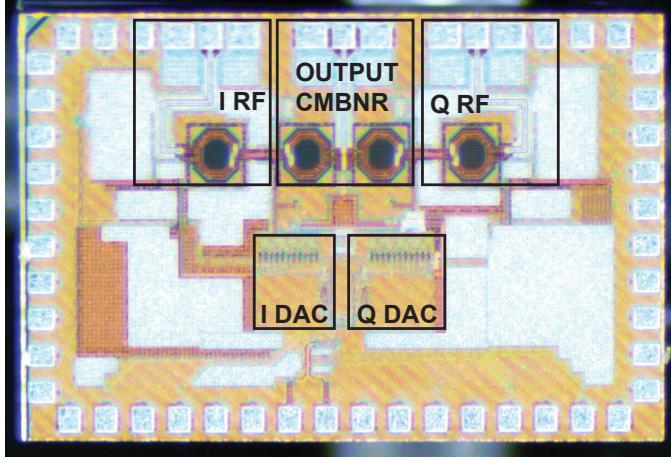


Fig. 10. Die micrograph. The chip measures 1.1 mm by 1.6 mm.

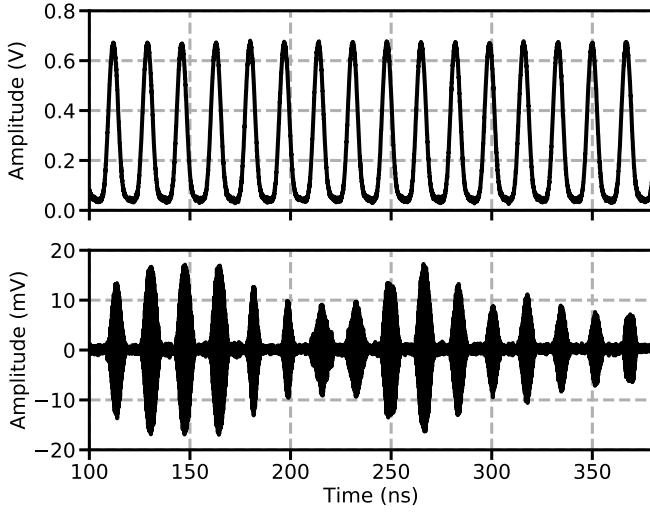


Fig. 11. Example time domain trigger (top) and pulse (bottom) waveforms, measured at 300 K. The chip was initialized with sixteen different waveforms and stepped through the select lines. For these measurementen, the chip was driven with a 2 GHz sample clock. A cable loss of approximately xx dB has not been de-embedded from the measurement results.

in close proximity (see Fig. ??). On the other hand, with the cancellation enabled, we found that the LO leakage was reduced to the level that a $|1\rangle$ population of less than 4% was observed after $20\ \mu\text{s}$. This level of unintended drive is acceptable, as it corresponds to an error of less than 0.01% on the timescale of a gate operation.

2) *Amplitude Control:* Next, we preformed a Rabi oscillation experiment using the procedure delineated in Fig. ???. The qubit was, initialized to the $|0\rangle$ state, excited with a raised cosine pulse of 22-ns in duration, and then the qubit state was read out. The experiment was repeated over a wide range of

3) *Coherent Control and Fast Switching:* A three pulse experiment was carried out to evaluate the phase control and fast switching aspects of the IC. The experimental protocol is described in Fig. ??.

4) *Qubit Relaxation Time (T_1):* The relaxation time, T_1 , for the qubit was measured using the CMOS IC as well as the baseline system and the results were compared to ensure that the CMOS IC was not introducing additional noise that could drive ω_{01} , there by reducing T_1 . For this experiment, the qubit was reset to the $|0\rangle$ state, excited to the $|1\rangle$ state with a π -pulse, and a state measurement was made after a delay, τ . xx statistics were gathered for each value of τ .

Results are plotted in Fig. 16. As expected, we observed that the qubit relaxed exponentially both when controlled by the CMOS IC and by the standard XY control electronics. The measured values of T_1 were found to be $18.3\ \mu\text{s}$ and $17.8\ \mu\text{s}$ when using the standard control electronics and cryo-CMOS IC, respectively. It is believed that these numbers are within the measurement/fitting error, and thus that there was no evidence of any impact on qubit relaxation due to broadband noise.

5) $|2\rangle$ state population: The envelope quality (spectral width) was also evaluated. In this experiment, the qubit was first reset and then excited with a sequence of 200 π -pulses. Since any energy at the ω_{12} transition frequency will drive this undesired transition, the application 200 π -pulses enables will result in substantial $|2\rangle$ state population if the ω_{12} transition is even weakly excited. Finally, the state was measured using a readout scheme that permitted distinguishing the $|0\rangle$, $|1\rangle$ and $|2\rangle$ states. The experiment was repeated for pulse durations

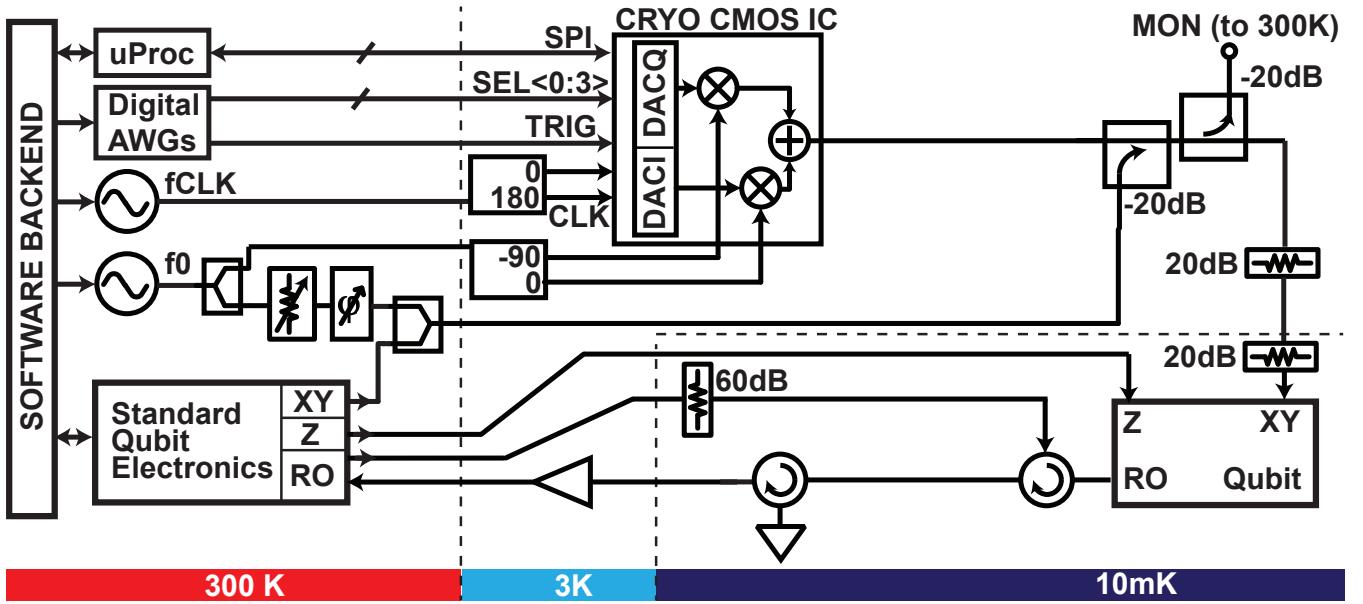


Fig. 12. Simplified block diagram of the test setup used for quantum control experiments.

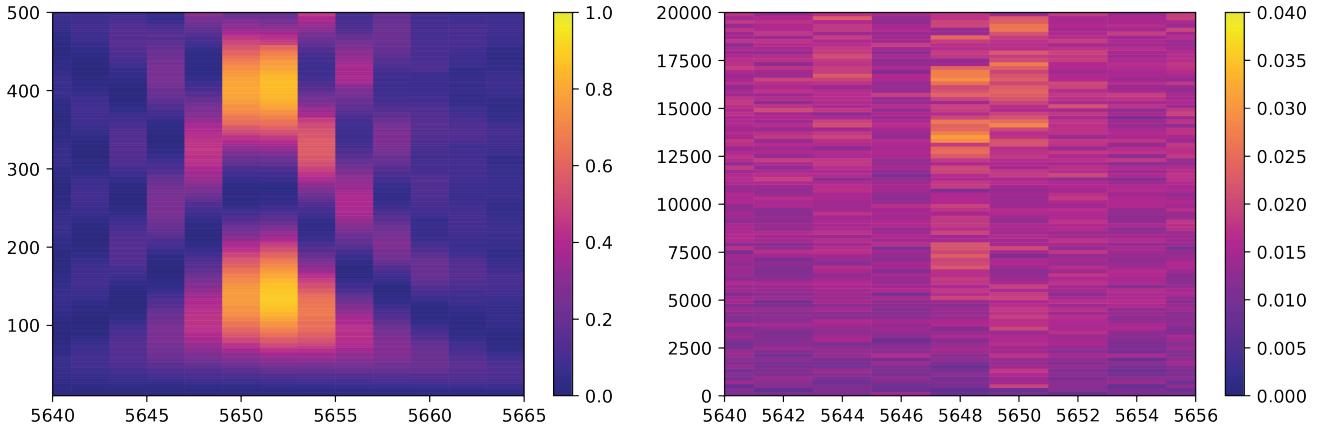


Fig. 13. Effect of local oscillator feedthrough.

ranging from xx-yy ns, corresponding to a clock frequencies in the range of xx-yy GHz and xx statistics were gathered at each point. A nominally raised cosine envelope shape was employed and the amplitude for a nominal π -pulse was determined independently at each clocking frequency.

The $|2\rangle$ -state measurement results appear in Fig. 18 along with the measurement noise floor, limited by the separately measured $|2\rangle$ state readout fidelity of $\text{xx}\%$. The $|2\rangle$ -state population was found to be negligible for pulse durations of approximately xx ns or longer. However, for faster pulses, excitation of the ω_{12} transition was observed, with a maximum $|2\rangle$ state population of just over 2.5%, or 0.0125% per pulse, for a pulse duration of xx ns. The behavior is consistent with expectation, given the relatively large anharmonicity of the qubit, and would be worse for a qubit with smaller anharmonicity.

D. Comparison with state of the art

The integrated circuit is compared with a state-of-the-art control system in Table II. The measured performance is comparable with the standard electronics in most regards. However, the size, digital data-rate, and power consumption required to operate the chip are all at least an order of magnitude lower than the standard readout electronics. Of particular importance is the total power consumption, which is below 2 mW. The number cited is equal to the total AC and DC power required to put out a continuous stream of π -pulses when clocked at 1 GHz and is viewed to be a conservative number as the chip is expected to operate at a lower activity factor in practice. Nonetheless, there is still significant room for improvement in performance that can be achieved with careful design once cryogenic simulation models become available.

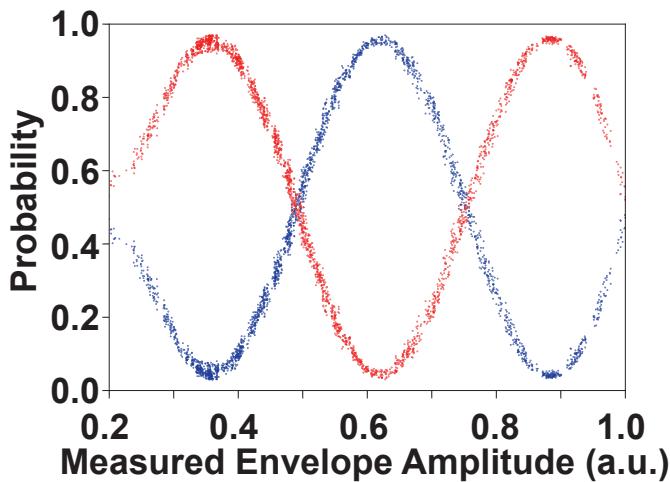


Fig. 14. Chip block diagram

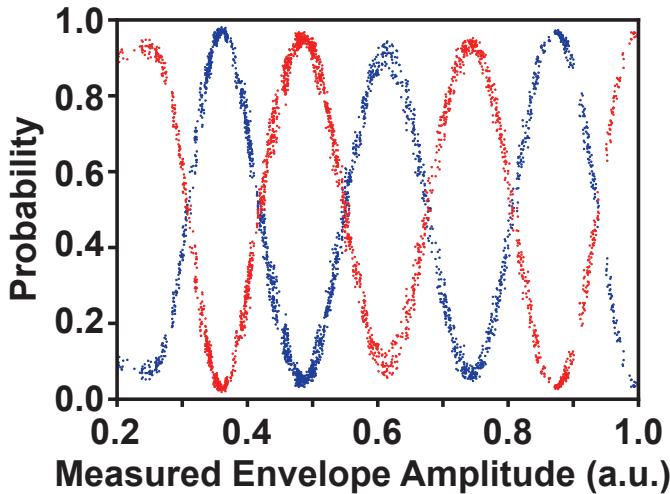


Fig. 15. Chip block diagram

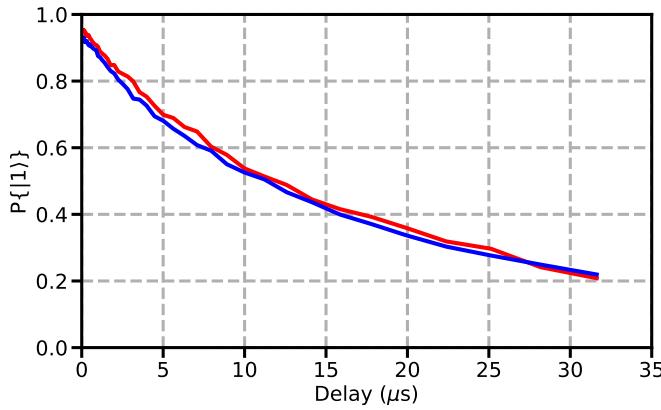


Fig. 16. Chip block diagram

VI. CONCLUSION

REFERENCES

- [1] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, "Surface codes: Towards practical large-scale quantum computation," *Physical*

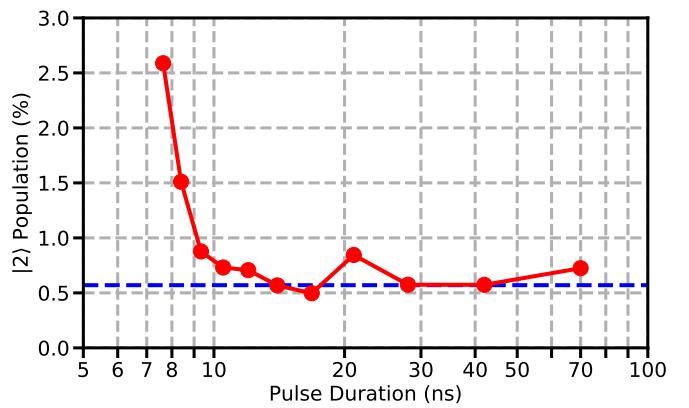


Fig. 17. Add limit

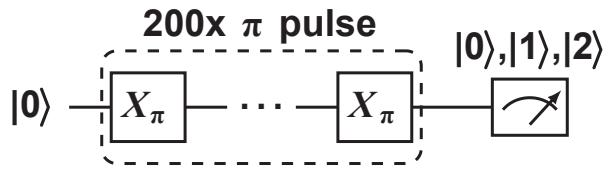


Fig. 18. 2state

TABLE II
COMPARISON OF CMOS IC TO STATE-OF-THE-ART QUANTUM CONTROL SYSTEM

	Conventional Control	Cryo-CMOS IC
Form Factor	Rack Mount	Integrated Circuit
Physical Temp	300 K	3 K
Update Rate	1 Gs/s	1 Gs/s
Instruction Set	N/A	4-bit
Digital Data Rate	28 Gb/s	<0.5 Gb/s
Measured T_1	18.3 μ s	17.8 μ s
$ 2\rangle$ population	Negligible	Negligible
π -Rabi $P\{ 1\rangle\}$	\approx 95 %	\approx 95 %
3 Gate RMS Error	2.5%	11.7%
Total AC+DC Power	> 1 W	< 2 mW

Review A, vol. 86, no. 3, p. 032324, 2012.

- [2] J. P. van Dijk, E. Kawakami, R. N. Schouten, M. Veldhorst, L. M. Vandersypen, M. Babaie, E. Charbon, and F. Sebastian, "The impact of classical control electronics on qubit fidelity," *arXiv preprint arXiv:1803.06176*, 2018.
- [3] 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*, vol. 508, no. 7497, p. 500, 2014.
- [4] E. Leonard Jr, M. A. Beck, J. Nelson, B. G. Christensen, T. Thorbeck, C. Howington, A. Opremcak, I. V. Pechenezhskiy, K. Dodge, N. P. Dupuis *et al.*, "Digital coherent control of a superconducting qubit," *Physical Review Applied*, vol. 11, no. 1, p. 014009, 2019.
- [5] A. Opremcak, I. Pechenezhskiy, C. Howington, B. Christensen, M. Beck, E. Leonard, J. Suttle, C. Wilen, K. Nesterov, G. Ribeill *et al.*, "Measurement of a superconducting qubit with a microwave photon counter," *Science*, vol. 361, no. 6408, pp. 1239–1242, 2018.
- [6] R. McDermott, M. Vavilov, B. Plourde, F. Wilhelm, P. Liebermann, O. Mukhanov, and T. Ohki, "Quantum-classical interface based on single flux quantum digital logic," *Quantum science and technology*, vol. 3, no. 2, p. 024004, 2018.
- [7] J. Hornibrook, J. Colless, I. C. Lamb, S. Pauka, H. Lu, A. Gossard, J. Watson, G. Gardner, S. Fallahi, M. Manfra *et al.*, "Cryogenic control architecture for large-scale quantum computing," *Physical Review Applied*, vol. 3, no. 2, p. 024010, 2015.

- [8] M. Mehrpoo, B. Patra, J. Gong, J. van Dijk, H. Homulle, G. Kiene, A. Vladimirescu, F. Sebastian, E. Charbon, M. Babaie *et al.*, "Benefits and challenges of designing cryogenic cmos rf circuits for quantum computers," in *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2019, pp. 1–5.
- [9] E. Charbon, F. Sebastian, M. Babaie, A. Vladimirescu, M. Shahmohammadi, R. B. Staszewski, H. A. Homulle, B. Patra, J. P. Van Dijk, R. M. Incandela *et al.*, "15.5 cryo-cmos circuits and systems for scalable quantum computing," in *2017 IEEE International Solid-State Circuits Conference (ISSCC)*. Ieee, 2017, pp. 264–265.
- [10] B. Patra, R. M. Incandela, J. P. Van Dijk, H. A. Homulle, L. Song, M. Shahmohammadi, R. B. Staszewski, A. Vladimirescu, M. Babaie, F. Sebastian *et al.*, "Cryo-cmos circuits and systems for quantum computing applications," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 1, pp. 309–321, 2018.
- [11] H. Homulle, F. Sebastian, and E. Charbon, "Deep-cryogenic voltage references in 40-nm cmos," *IEEE Solid-State Circuits Letters*, vol. 1, no. 5, pp. 110–113, 2018.
- [12] S. Bonen, U. Alakusu, Y. Duan, M. Gong, M. Dadash, L. Lucci, D. Daughton, G. Adam, S. Iordănescu, M. Păsteanu *et al.*, "Cryogenic characterization of 22-nm fdsoi cmos technology for quantum computing ics," *IEEE Electron Device Letters*, vol. 40, no. 1, pp. 127–130, 2019.
- [13] R. M. Incandela, L. Song, H. Homulle, E. Charbon, A. Vladimirescu, and F. Sebastian, "Characterization and compact modeling of nanometer cmos transistors at deep-cryogenic temperatures," *IEEE Journal of the Electron Devices Society*, vol. 6, pp. 996–1006, 2018.
- [14] A. Beckers, F. Jazaeri, and C. Enz, "Characterization and modeling of 28-nm bulk cmos technology down to 4.2 k," *IEEE Journal of the Electron Devices Society*, vol. 6, pp. 1007–1018, 2018.
- [15] J. J. Sakurai and E. D. Commins, "Modern quantum mechanics, revised edition," 1995.
- [16] A. M. Kadin, *Introduction to superconducting circuits*. Wiley New York, 1999.
- [17] M. Tinkham, *Introduction to superconductivity*. Courier Corporation, 2004.
- [18] J. Koch, M. Y. Terri, J. Gambetta, A. A. Houck, D. 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*, vol. 76, no. 4, p. 042319, 2007.
- [19] F. Motzoi, J. M. Gambetta, P. Rebentrost, and F. K. Wilhelm, "Simple pulses for elimination of leakage in weakly nonlinear qubits," *Physical review letters*, vol. 103, no. 11, p. 110501, 2009.
- [20] D. C. McKay, C. J. Wood, S. Sheldon, J. M. Chow, and J. M. Gambetta, "Efficient z gates for quantum computing," *Physical Review A*, vol. 96, no. 2, p. 022330, 2017.
- [21] D. Rosenberg, D. Kim, R. Das, D. Yost, S. Gustavsson, D. Hover, P. Krantz, A. Melville, L. Racz, G. Samach *et al.*, "3d integrated superconducting qubits," *npj quantum information*, vol. 3, no. 1, p. 42, 2017.
- [22] C. Kittel, *Thermal Physics*. Wiley, 1969. [Online]. Available: <https://books.google.com/books?id=iCA-AAAAIAAJ>
- [23] D. B. Tuckerman, M. C. Hamilton, D. J. Reilly, R. Bai, G. A. Hernandez, J. M. Hornibrook, J. A. Sellers, and C. D. Ellis, "Flexible superconducting nb transmission lines on thin film polyimide for quantum computing applications," *Superconductor Science and Technology*, vol. 29, no. 8, p. 084007, 2016.
- [24] PT420-CPA1114-Capacity-Curve, Cryomech, 1 2018.
- [25] D. S. Holmes, A. L. Ripple, and M. A. Manheimer, "Energy-efficient superconducting computing—power budgets and requirements," *IEEE Transactions on Applied Superconductivity*, vol. 23, no. 3, pp. 1701 610–1701 610, 2013.
- [26] L. H. Pedersen, N. M. Møller, and K. Mølmer, "Fidelity of quantum operations," *Physics Letters A*, vol. 367, no. 1-2, pp. 47–51, 2007.
- [27] 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*, vol. 82, no. 2, p. 1155, 2010.
- [28] J. M. Gambetta, F. Motzoi, S. Merkel, and F. K. Wilhelm, "Analytic control methods for high-fidelity unitary operations in a weakly nonlinear oscillator," *Physical Review A*, vol. 83, no. 1, p. 012308, 2011.
- [29] J. van Dijk, M. Babaie, E. Charbon, A. Vladimirescu, F. Sebastian *et al.*, "Characterization and model validation of mismatch in nanometer cmos at cryogenic temperatures," in *2018 48th European Solid-State Device Research Conference (ESSDERC)*. IEEE, 2018, pp. 246–249.
- [30] M. Myslinski, W. Wiatr, and D. Schreurs, "A three-step procedure utilizing only two test structures for deembedding transistor from on-wafer s-parameter measurements," in *15th International Conference on Microwaves, Radar and Wireless Communications (IEEE Cat. No. 04EX824)*, vol. 2. IEEE, 2004, pp. 674–677.