Title

Temperature-Insensitive CMOS Quantum Controller With >99% Fidelity for Universal Single-Qubit Gates

Permalink

Author

Wang, Yen-Hsiang

Publication Date

2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Temperature-Insensitive CMOS Quantum Controller

With >99% Fidelity for Universal Single-Qubit Gates

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Electrical Engineering

by

Yen-Hsiang Wang

2023

ABSTRACT OF THE DISSERTATION

Temperature-Insensitive CMOS Quantum Controller

With >99% Fidelity for Universal Single-Qubit Gates

by

Yen-Hsiang Wang

Doctor of Philosophy in Electrical Engineering

University of California, Los Angeles, 2023

Professor Mau-Chung Frank Chang, Chair

Fault-tolerant quantum computing is built upon high-fidelity universal quantum gates enabled by scalable quantum processors. However, the scalability of microwave-based quantum processors is hindered by large layout pitch size of 1mm due to high-Q resonant circuits and for crosstalk reduction. We therefore choose to work on the more scalable dc-pulse-based quantum processor utilizing the position-based double-quantum-dot charge qubit with a 200nm pitch.

To alleviate prior difficulty encountered in transmitting fast-edged waveforms generated by arbitrary waveform generators via meter-long cables, we propose a scalable architecture to partition the quantum controller and house only its low-power output stage with the qubit chip within low-parasitic packaging. To enable high-fidelity universal single-qubit gates, we derive requirements for the precision of control waveforms, propose temperature-insensitive design methodologies without device models at cryogenic temperatures, and develop an inductor-less,

precise, and fully integrated dc-pulse-based quantum controller achieving 4.3ps maximum error in pulse width and highly linear pulse height control. A closed-loop, automated, multi-variable calibration using an on-chip 2ps/step pulse-swallowing cyclic TDC is devised to ensure temperature-insensitive performances.

Our quantum controller prototype is fabricated in 28nm bulk CMOS and verified through measurements down to 77K. Comparing the performance with state-of-art dc-pulse-based quantum controllers, our work has reduced the shortest duration of a pulse by over 3X, improved the linear resolution of width tuning by over 40X, reduced the width uncertainty by three orders of magnitude, realized an on-chip high-precision pulse width calibration, and largely improved the precision of height tuning to enable high-fidelity universal single-qubit gates.

The dissertation of Yen-Hsiang Wang is approved.

Kang Lung Wang

Yuanxun Wang

Wentai Liu

Mau-Chung Frank Chang, Committee Chair

University of California, Los Angeles

2023

To my family

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

VITA

2008   B.S., Electrical Engineering, National Taiwan University, Taiwan

2011   M.S., Electrical Engineering, University of California, Los Angeles, CA

2013-2016   Technologies Incubation Scholarship, Ministry of Education, Taiwan

2014-2016   High-Speed I/O Design Intern, Bell Labs, New Providence, NJ

2016-present   Mixed-signal IC Design Engineer, Broadcom Inc., Irvine, CA

SELECTED PUBLICATIONS

Wang, Y. H., Chen, J. W., … & Chang, M. C. F. A scalable CMOS quantum state controller with >99% fidelity for universal single-qubit gates. Submitted to journal for publication.

Chiang, H. L., …, Wang, Y. H., ... & Radu, I. P. (2023, June). How Fault-Tolerant Quantum Computing Benefits from Cryo-CMOS Technology. In 2023 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits) (pp. 1-2). IEEE.

Liu, C. C., Wang, Y. H., … & Chang, M. C. F. (2016, June). A 2.2 GHz SRAM with high temperature variation immunity for deep learning application under 28nm. In Proceedings of the 53rd Annual Design Automation Conference (pp. 1-6).

Chen, Z. Z., Wang, Y. H., ... & Chang, M. C. F. (2015, February). 14.9 sub-sampling all-digital fractional-N frequency synthesizer with$-$111dBc/Hz in-band phase noise and an FOM of$-$242dB. In 2015 IEEE International Solid-State Circuits Conference-(ISSCC) Digest of Technical Papers (pp. 1-3). IEEE.

Tang, A., ..., Wang, Y. H., ... & Chang, M. C. F. (2012). A D-Band CMOS Transmitter With IF-Envelope Feed-Forward Pre-Distortion and Injection-Locked Frequency-Tripling Synthesizer. IEEE transactions on microwave theory and techniques, 60(12), 4129-4137.

Tang, A., ..., Wang, Y. H., ... & Chang, M. C. F. (2012, June). A CMOS 135–150 GHz 0.4 dBm EIRP transmitter with 5.1 dB P1dB extension using IF envelope feed-forward gain compensation. In 2012 IEEE/MTT-S International Microwave Symposium Digest (pp. 1-3). IEEE.

Tang, A., Virbila, G., Wang, Y. H., ... & Chang, M. C. F. (2012, June). A 200 GHz 16-pixel focal plane array imager using CMOS super regenerative receivers with quench synchronization. In 2012 IEEE/MTT-S International Microwave Symposium Digest (pp. 1-3). IEEE.

Tang, A., …, Wang, Y. H., ... & Chang, M. C. F. (2012, February). A low-overhead self-healing embedded system for ensuring high yield and long-term sustainability of 60GHz 4Gb/s radio-on-a-chip. In 2012 IEEE International Solid-State Circuits Conference (pp. 316-318). IEEE.

Tang, A., …, Wang, Y. H., ... & Chang, M. C. F. (2012, February). A 144GHz 0.76 cm-resolution sub-carrier SAR phase radar for 3D imaging in 65nm CMOS. In 2012 IEEE International Solid-State Circuits Conference (pp. 264-266). IEEE.

Tang, A., …, Wang, Y. H., ... & Chang, M. C. (2012). A Formula Not Shown-Band CMOS Transmitter With IF-Envelope Feed-Forward Pre-Distortion and Injection-Locked Frequency-Tripling Synthesizer. IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES MTT, 60(12), 4129-4137.

Ku, I. N., …, Wang, Y. H., & Chang, M. C. F. (2011, September). A 40-mW 7-bit 2.2-GS/s time-interleaved subranging ADC for low-power gigabit wireless communications in 65-nm CMOS. In 2011 IEEE Custom Integrated Circuits Conference (CICC) (pp. 1-4). IEEE.

Chapter 1　Introduction to quantum computing

1.1. Significance of quantum computing

Quantum computing can accelerate revolutionary breakthrough by solving complex problems intractable to classical counterparts. Key fundamental properties that distinguish quantum computers include the superposition of basis states and the entanglement between quantum bits (qubits). Unlike a classical bit which is either 0 or 1 at a given time, the state of a qubit, $|\psi\rangle$, is a superposition of the two computational basis states, e.g., $|0\rangle$ and $|1\rangle$. Mathematically it can be expressed as

$$|\psi\rangle = c_0|0\rangle + c_1|1\rangle$$

where c_0 and c_1 represent complex probability amplitudes, and $c_0{}^2 + c_1{}^2 = 1$. Entanglement between qubits further expands the information density of a quantum computer. A two-qubit system stores the superposition of 2^2 states at any given time, and can be written as

$$|\psi\rangle = c_{00}|00\rangle + c_{01}|01\rangle + c_{10}|10\rangle + c_{11}|11\rangle$$

where $c_{00}{}^2 + c_{01}{}^2 + c_{10}{}^2 + c_{11}{}^2 = 1$. In contrast, a two-bit classical computer still stores only one state at a time. Therefore, as the number of bits increases, the information density of a quantum computer is exponentially higher than that of a classical counterpart.

In an N-qubit system, the entanglement also causes the probability amplitude of all 2^N states to change when the state of a single qubit changes. On the other hand, flipping a bit in an N-bit classical system only affects its current state. Above properties enable quantum computers to solve complex optimization problems with large data sets in a parallel fashion that is exponentially faster than classical computers. An example of computational advantage of quantum computers is

shown in Table 1-1. Applications of quantum computing include cryptography, searching, molecular formations, material science, artificial intelligence, etc.

Table 1-1: Analysis of computational advantage of quantum computers (IBM Corporation 2018).

Type of scaling	Time to solve problem				
Classical algorithm with exponential runtime	10 secs	2 mins	330 years	3300 years	Age of the universe
Quantum algorithm with polynomial runtime	1 min	2 mins	10 mins	11 mins	~24 mins

1.2. Cornerstones for quantum computing

To harness the power of quantum computing, in a fault-tolerant quantum computer at least 1,000 almost-error-free logical qubits need to be integrated. With an error-tolerant quantum error correction (QEC) algorithm such as the surface code, each logical qubit will consist of at least 1,000 physical qubits with less than 1% error rate (Fowler, et al. 2012). Also, it must be capable of performing universal quantum gate operations. These requirements impose scalability, universality, and fidelity constraints on the quantum processor, the combination of physical qubits and their quantum controller.

1.2.1. Scalability

To integrate over one million physical qubits (1,000 logical qubits times 1,000 physical qubits per logical qubit) in a fault-tolerant quantum computer, scalability of the quantum processor is crucial. Being able to leverage mature semiconductor fabrication processes and as a result easier

to be integrated with quantum controllers enhances the scalability of the qubit. Thus, in this work we focus on highly scalable, all-electrically-controlled, CMOS-compatible qubits.

Popular types of qubits in this category can be divided into microwave-controlled and dc-pulse-controlled. Microwave-controlled qubits include superconducting qubits and quantum-dot (QD) qubits utilizing the spin of charges. Their minimum layout pitch is limited to ~1mm due to the required high-Q LC resonant circuits and to alleviate crosstalk. On the contrary, the layout pitch of dc-pulse-controlled qubits such as the position-based double-quantum-dot (DQD) charge qubit is primarily limited by the resolution of lithography during fabrication, and it is on the order of 200nm. Examples are shown in Figure 1-1.

<div align="center">(a)   (b)</div>
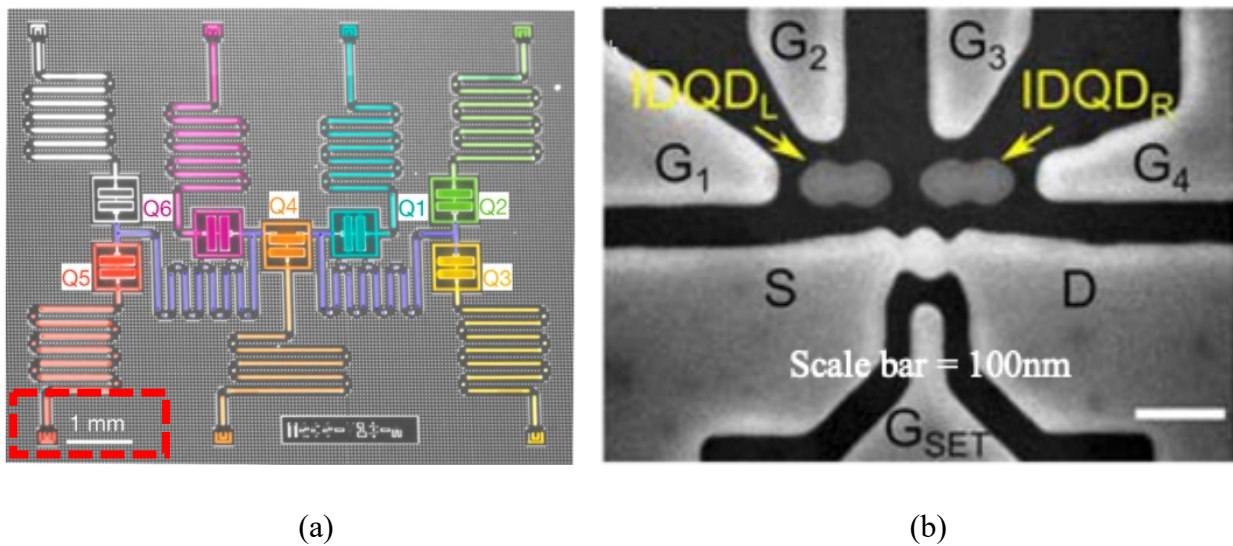
Figure 1-1: (a) 1mm-pitch superconducting qubits with resonant circuits (Kandala, et al. 2017). (b) 200nm-pitch position-based DQD charge qubits (Yang, et al. 2016).

1.2.2. Universality

The state of a qubit can be visualized as a vector on the Bloch sphere as shown in Figure 1-2. Its mathematical expression in a spherical coordinate is

$$|\psi\rangle = cos(\theta/2)|0\rangle + e^{j\phi}sin(\theta/2)|1\rangle$$

where θ represents the polar angle from the Z axis and ϕ is the azimuthal angle from the X axis. When the quantum system is at rest, the state vector of the qubit precesses around the axis of eigenstates, and the angular frequency ω of this Larmor precession is related to the difference in energy of the eigenstates ΔE through

$$\omega = \frac{\Delta E}{\hbar}$$

Universality mandates arbitrary manipulation of all 2^N states in an N-qubit system, which can be broken down into a set of basic quantum gate operations called a universal quantum gate set. The constituents include single- and two-qubit gates. One example includes arbitrary X and Z rotation (Rx and Rz) for single-qubit gates and controlled X rotation for the two-qubit gate as illustrated in Figure 1-3. The combination of arbitrary X and Z rotation can rotate the state vector to any point on the Bloch sphere.
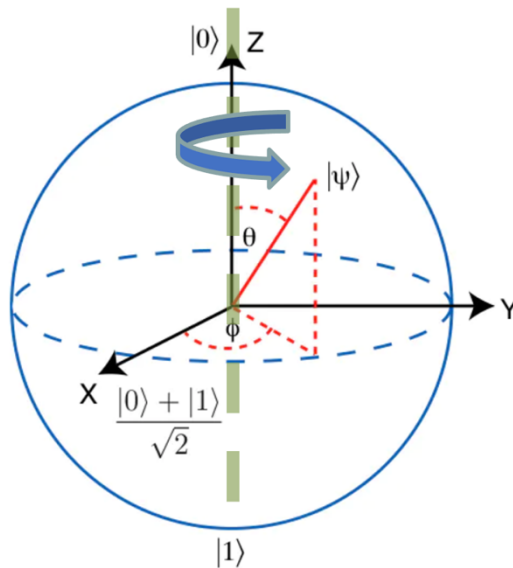
Figure 1-2: Representation of a qubit's state on the Bloch sphere.

Single-qubit gate	
X rotation	Z rotation
Control the probability of $\lvert 0 \rangle$ and $\lvert 1 \rangle$	Control the phase difference between $\lvert 0 \rangle$ and $\lvert 1 \rangle$

Rx

Rz
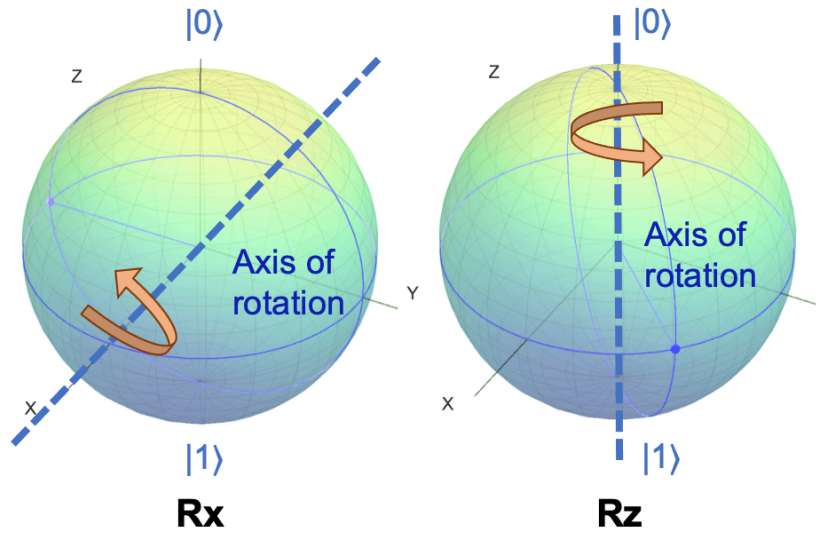
Two-qubit gate
Controlled X rotation
Perform X rotation on target qubit depending on the state of control qubit

Control qubit

Target qubit
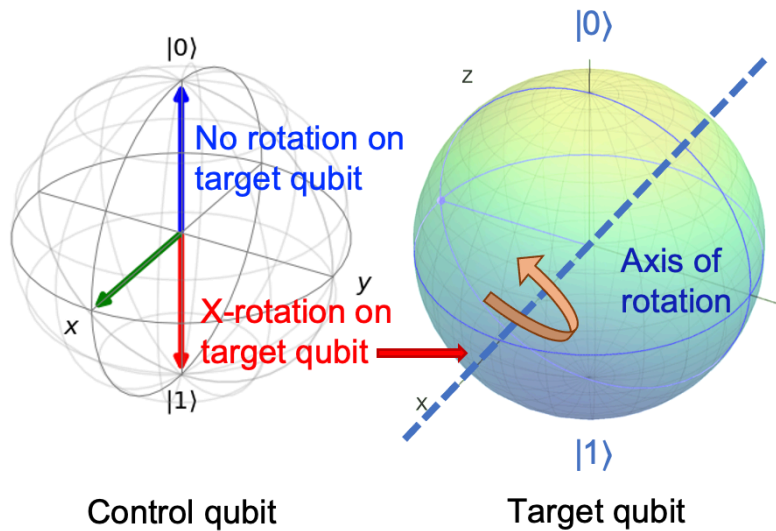
Figure 1-3: An example of universal quantum gate set.

1.2.3. Fidelity

Nonideal control waveforms and noise lead to a finite error rate when operating physical qubits. Even with a small error rate of 0.1% per operation, the accumulated error rate after repeating the operation 1,000 times will reach 63%, which is intolerable for practical purposes. Thus, algorithms for QEC are necessary.

One of the most popular QEC algorithms for its high tolerable error rate of physical qubits is the surface code, which gets its name from the required two-dimensional placement of qubits. A sample placement of qubits compatible with the surface code from Google is shown in Figure 1-4. The error rate of the constructed logical qubit after error correction, P_L, can be estimated from

$$P_L \approx 0.03 \times \left(\frac{P_{ph}}{P_{th}}\right)^{\sqrt{m/3}}$$

where P_{th} is the error threshold for each physical qubit set by the chosen QEC algorithm, P_{ph} is the actual error rate when operating a physical qubit, and m is the number of physical qubits required to achieve a certain P_L (Fowler, et al. 2012). Given the error rate of a logical qubit, as the error rate of constituting physical qubits increases towards the error threshold, more physical qubits will be needed to construct a logical qubit. For almost-error-free logical qubits utilizing surface-code QEC, the error threshold is 1%, and thus the error rate of a physical qubit needs to be below this threshold. Equivalently, the fidelity of operating a physical qubit must exceed 99%.
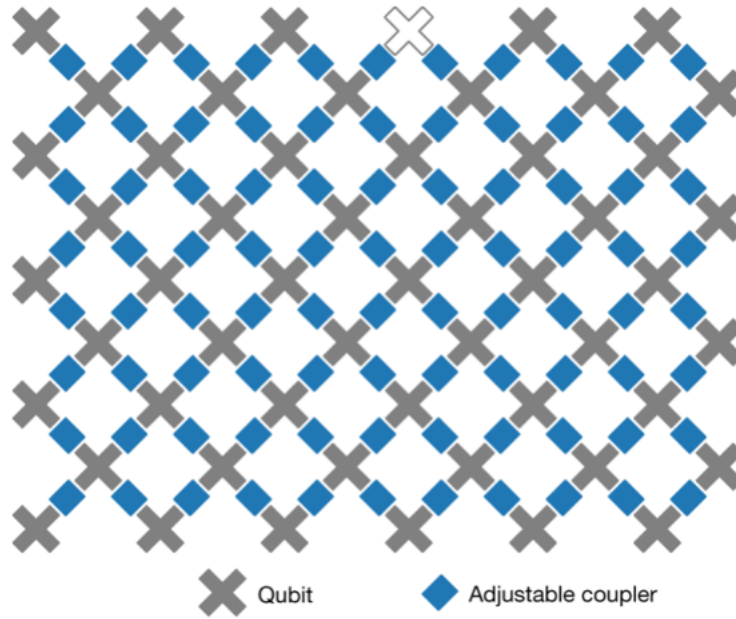
Figure 1-4: Surface-code-compatible qubit placement in the 54-qubit "Sycamore" quantum

processor from Google (Arute, et al. 2019).

1.3. Cooling power constraints

To preserve the delicate quantum properties against thermal noise, the typical operating temperature of our qubit ranges from tens to hundreds of millikelvin. Maintaining at the desired temperature relies on sufficient cooling power of the refrigerator to overcome the heat generated within the cryogenic chamber. The available cooling power reduces exponentially as the target temperature drops, and this imposes a stringent requirement on the power consumption of the quantum processor. I/O cabling will also increase the thermal load to the chamber and reduce the available power budget. Therefore, in practical quantum systems, the quantum controller is placed as close to the qubit as possible with a minimum I/O count. The cooling power of various chambers in a commercial dilution refrigerator is shown in Figure 1-5. At 4K the cooling power is around

1W. It drops to 10mW at 1K, and only 10μW at 20mK. Further enhancement of the cooling power

of refrigerators is another important area of research for scaled-up quantum computing.
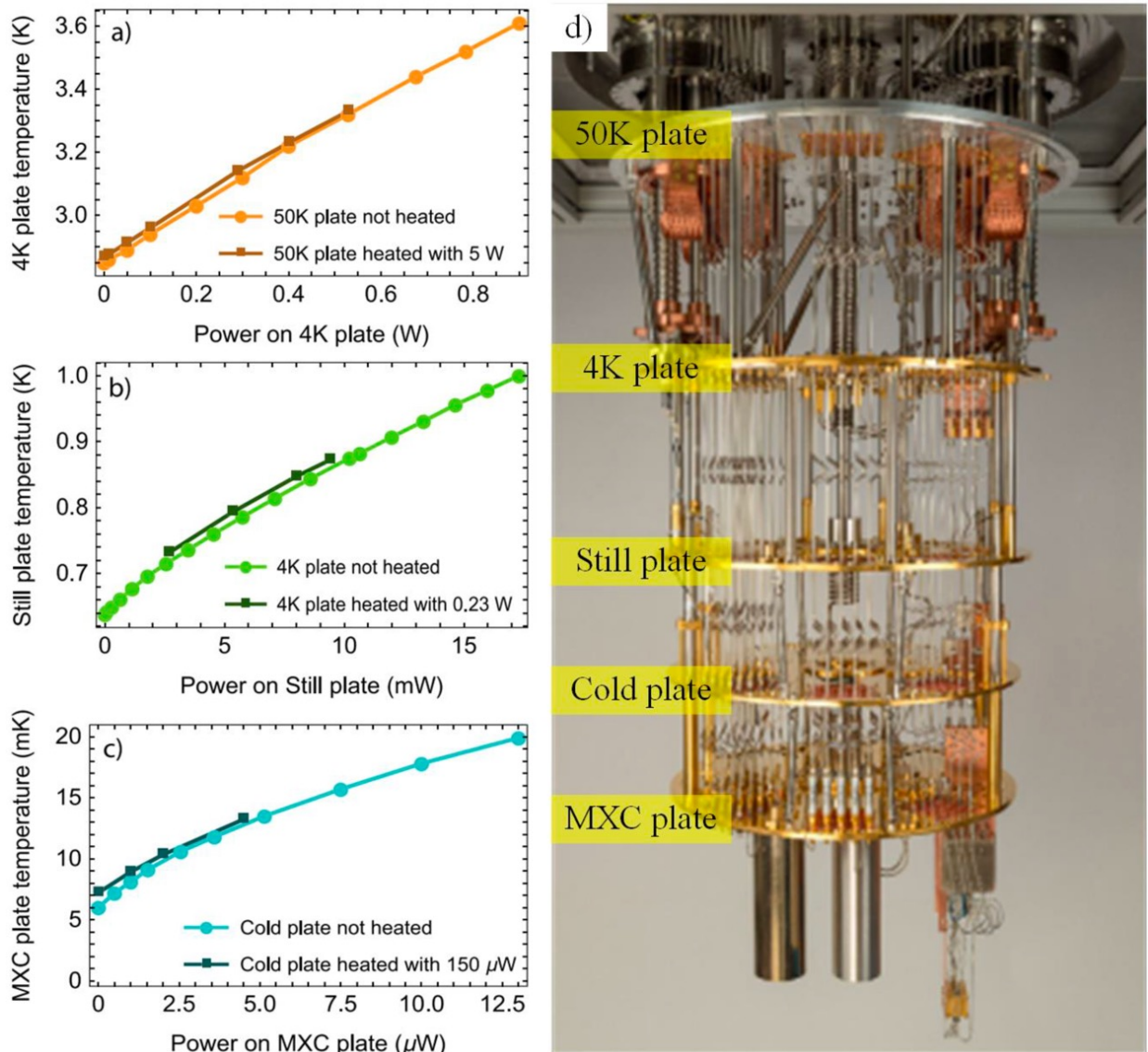
Figure 1-5: Cooling power of a commercial dilution refrigerator in various chambers (Bonen

2020).

Chapter 2   Scalable quantum processors

2.1.   Microwave-based versus dc-pulse-based quantum processors

- **Microwave-based quantum processors**

    The current mainstream microwave-based quantum processor is based on the superconductor qubit. In 2019, Google demonstrated quantum supremacy with a 54-qubit chip, which can encode an amount of information surpassing any classical computer. In 2022, IBM announced a quantum computer with 433 qubits and plans to achieve quantum advantage by 2030. However, as discussed in section 1.2.1, the 1mm pitch will hinder scaling up. This is limited by the dimension of the high-Q LC resonant circuits which is proportional to the carrier's wavelength. The typical carrier frequency of 4-8GHz sets the practical layout pitch. Surface-code-compatible qubit placement shown in Figure 2-1 further increases the equivalent pitch between qubits. On the quantum controller side, for spectral purity, similar high-Q resonant circuitry mandates large on-chip inductors and consumes high power. A 2-channel non-multiplexed quantum controller from IBM in 14nm FinFET is shown in Figure 2-1. Each channel occupied 0.8mm by 1.95mm.

- **DC-pulse-based quantum processors**

    To address the scalability issue, we choose to work on the highly scalable, dc-pulse-controlled position-based DQD charge qubit. The hard-wall-confined DQD shown in Figure 2-2 is of particular interest due to its long decoherence time on the order of tens of microseconds (Yang, et al. 2016) (Gorman, Hasko and Williams 2005). This allows many high-fidelity quantum gate operations as a typical 2π X rotation only needs ~200ps. Aside from the ~200nm layout pitch, its quantum state can be fully controlled by dc pulses, and its capacitive terminals enable the adoption of low-power capacitive-digital-to-analog-converter-based (CDAC-based) output stage for the

controller. Since the controller no longer needs on-chip inductors, its area will be drastically smaller than microwaved-based counterpart. The qubit's required voltage range is also compatible with advanced CMOS technologies. As shown in the Figure 2-2, the needed range of dc bias for state initialization is less than 1.8V, which is a common supply voltage for thick-oxide I/O FETs. Since an adiabatic process is desired during initialization to avoid unwanted excitation, thick-oxide FETs are the perfect candidate for these dc sampling switches. On the other hand, less-than-0.9V pulse height is needed for diabatic pulsing during quantum gate operations. Since this is a typical supply voltage for thin-oxide core FETs, they can be used to generate these fast pulses. Due to these promising properties, our CMOS quantum controller is designed for this type of qubit.
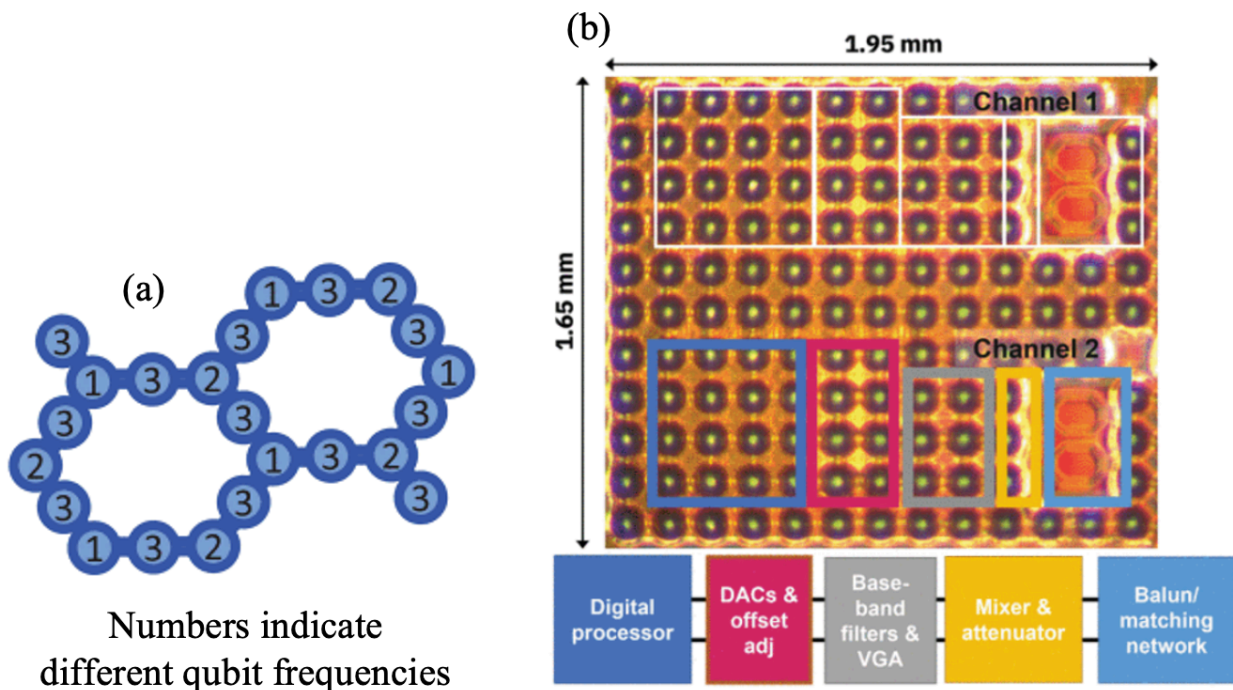
Figure 2-1: (a) Placement of superconducting qubits compatible with heavy hexagonal surface code (Chakraborty, et al. 2022). (b) A 2-channel non-multiplexed quantum controller in 14nm FinFET for superconducting qubits (Chakraborty, et al. 2022).
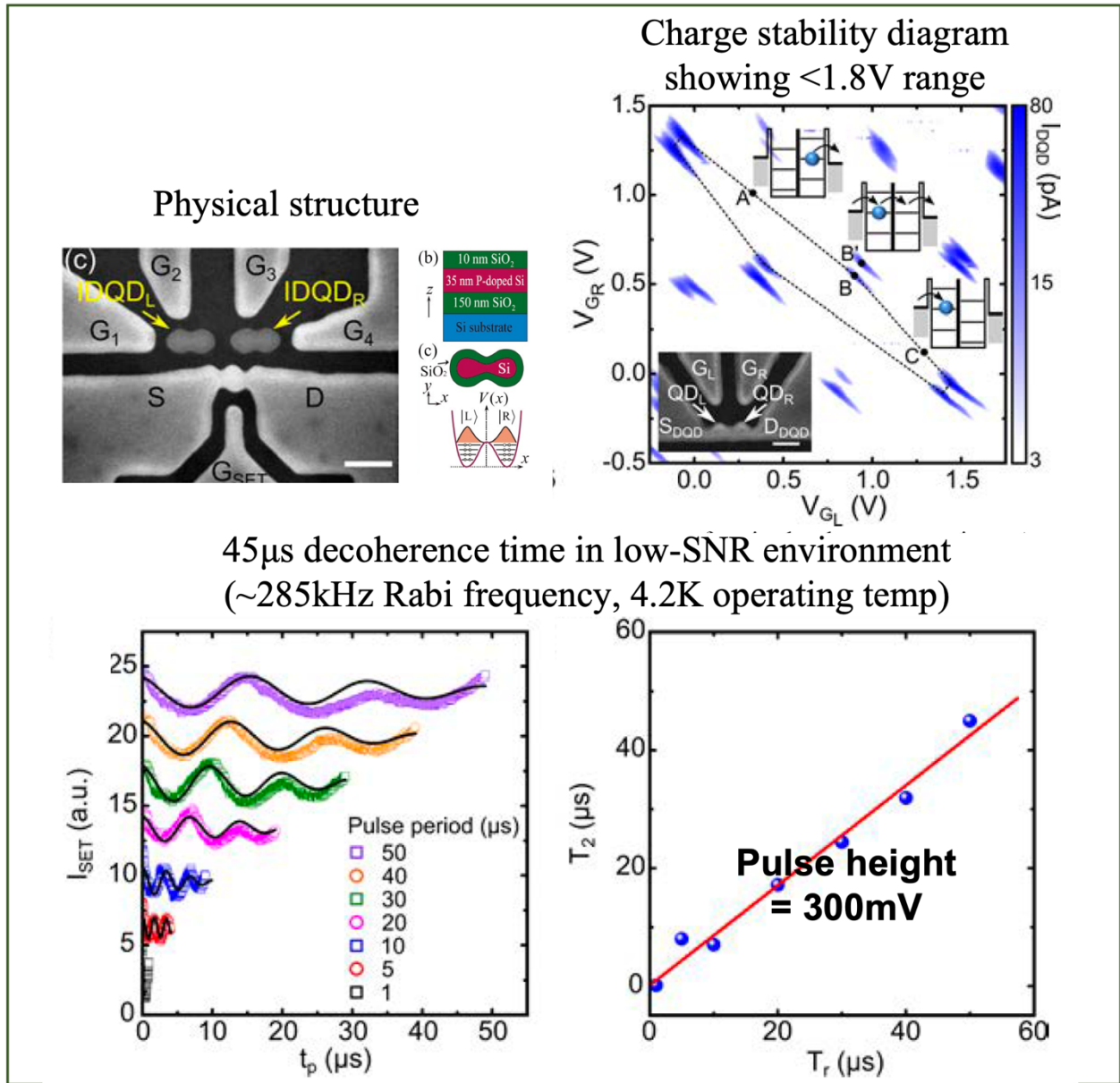
Figure 2-2: Hard-wall-confined position-based DQD charge qubit (Yang, et al. 2016) (Gorman, Hasko and Williams 2005).

2.2. Position-based DQD charge qubit

2.2.1. Universal single-qubit quantum gates

- **Principle of operation**

The position-based DQD charge qubit consists of a left and a right QD separated by a tunable tunneling barrier. As the size of the QD reduces thanks to advanced fabrication technologies, the required charging energy to add extra charges to the QD increases due to the stronger Coulomb repulsion (Rabouw and Donega 2017). This results in clearly discretized charge state of the QD. The position-based DQD charge qubit typically utilizes the location of the charge to encode its quantum state. Mathematically, the basis state $|L\rangle$ denotes that the charge is on the left dot, and $|R\rangle$ on the right dot. The energy diagram shown in Figure 2-4 changes with the interdot coupling related to the potential of the barrier. The general Hamiltonian is written as

$$H = \begin{pmatrix} \varepsilon/2 & \Delta/2 \\ \Delta/2 & -\varepsilon/2 \end{pmatrix}$$

where ε is the detuning energy between the two dots and Δ is the interdot coupling energy at the anti-crossing in the energy diagram. The average energy of the two QDs only contributes to the global phase of the system, and thus it is irrelevant to the quantum gate operations. The Hamiltonian determines the eigenstates and their corresponding eigenenergy as expressed below.

$$E_{\pm} = \frac{1}{2}\sqrt{\Delta^2 + \varepsilon^2}$$

$$\begin{cases} |\psi_+\rangle = \cos(\theta_0/2)\,|L\rangle + \sin(\theta_0/2)\,|R\rangle \\ |\psi_-\rangle = \sin(\theta_0/2)\,|L\rangle - \cos(\theta_0/2)\,|R\rangle \end{cases}$$

$$\theta_0 = \tan^{-1}\frac{\Delta}{\varepsilon}$$

As described in section 1.2.2, universal single-qubit gates require arbitrary rotation around any two orthogonal axes on the Bloch sphere. Typically for our qubit, the two axes chosen are the

12

X and Z axis since the eigenstates align with these axes as explained in following paragraphs and the qubit state naturally precesses around them.

When the barrier potential is high and thus makes the interdot coupling negligible, or $\Delta \ll \varepsilon$, resulting eigenstates are $|L\rangle$ and $|R\rangle$. The corresponding eigenenergy for $|L\rangle$ is $\varepsilon/2$, and for $|R\rangle$ is $-\varepsilon/2$. Solving the time-dependent Schrödinger's equation yields a precession frequency of ε/h. Since $|L\rangle$ and $|R\rangle$ are aligned with the Z axis on the Bloch sphere as shown in Figure 2-4, setting a high barrier potential and introducing detuning between the dots result in a Z rotation.

On the other hand, when the barrier potential is lowered and thus the detuning is minimized, or $\Delta \gg \varepsilon$, resulting eigenstates become $|0\rangle$ and $|1\rangle$. The corresponding eigenenergy for $|0\rangle$ is $-\Delta/2$, and for $|1\rangle$ is $\Delta/2$. Again, solving the time-dependent Schrödinger's equation yields a precession frequency of Δ/h. Since $|0\rangle$ and $|1\rangle$ are aligned with the X axis on the Bloch sphere, lowering the barrier potential and nulling the detuning result in an X rotation.
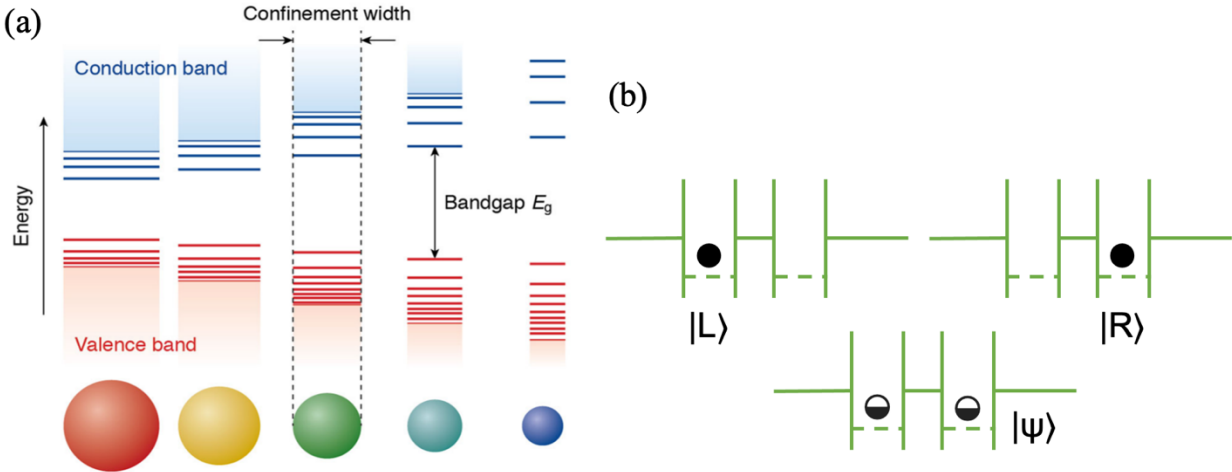
Figure 2-3: (a) Discretized energy states of QD (Rabouw and Donega 2017). (b) Illustration of quantum states of a position-based DQD charge qubit.
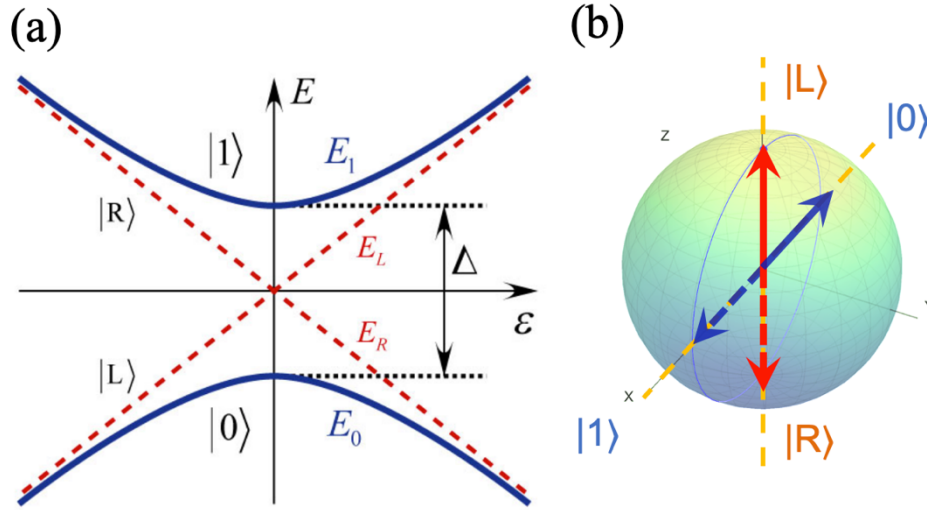
Figure 2-4: (a) Energy diagram with (blue) and without (red) interdot coupling. (b) Eigenstates on the Bloch sphere with (blue) and without (red) interdot coupling.

- **Actual implementation of X rotation (Rabi oscillation)**

To initialize the qubit's state to $|L\rangle$, we minimize the interdot coupling and set the potential of the left and right gate based on the charge stability diagram of the DQD such that there is one charge on the left dot and none on the right. After initialization, the detuning energy is reduced to zero by equalizing the potential of both gates to minimized unintended Z rotation in subsequent phases. To initiate the X rotation, as described earlier in this section, we lower the barrier potential by sending a fast pulse to the barrier terminal. As derived in section 2.2.2, for high fidelity it is important to minimize the edge transition time relative to the Rabi oscillation period. Through this diabatic process, the initial state, $|L\rangle$, starts to rotate around the X axis and oscillate between $|L\rangle$ and $|R\rangle$. The exponential relationship between barrier voltage and interdot coupling energy, and thus the Rabi oscillation frequency, can be explored through a highly programmable quantum controller with precise tuning in pulse width and height. To determine the state of the qubit after the X rotation, we raise the barrier potential to enter the hold phase and perform read-out through

14

neighboring charge sensors. This series of actions need to be repeated periodically for a projective read-out to determine the probability amplitude of each basis state.

Critical requirements on control waveforms include the following.

- Low-noise and high-resolution pulse height tuning: On barrier, this enables accurate control of the Rabi frequency. On gates, this minimizes unintended Z rotation.

- High-resolution, high-linearity and low uncertainty pulse width tuning for barrier pulses: This enables accurate control of degree of rotation.

- Minimized edge transition time during pulsing: This minimizes unintended rotation during edge transitions due to the intermediate time-varying Hamiltonian.
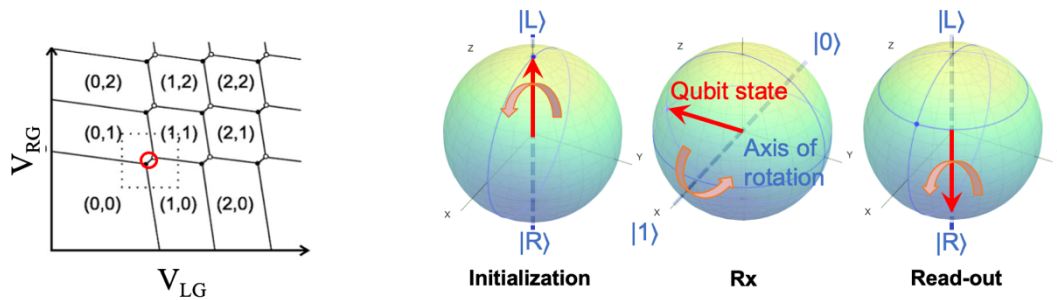
(a) Operating point of Rx on charge stability diagram   (b) State evolution on the Bloch sphere
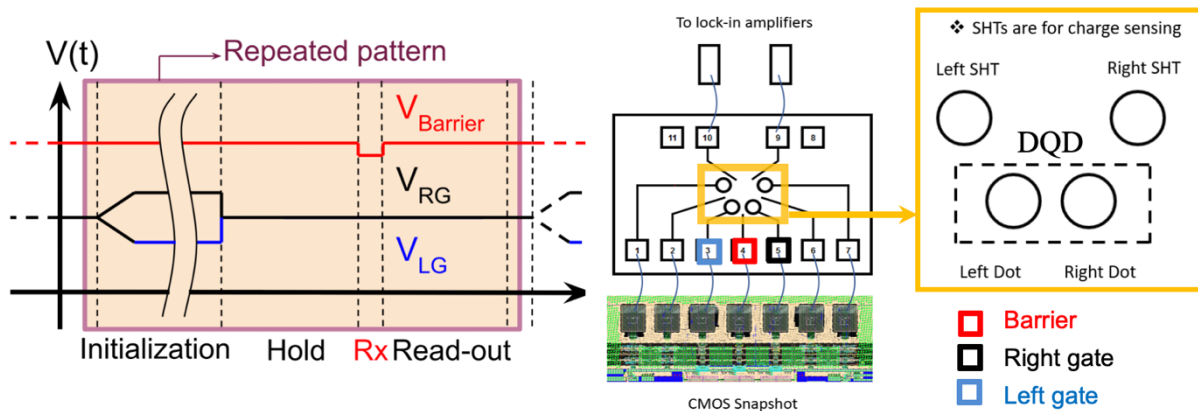
(c) Timing diagram and corresponding terminals in intended CMOS-qubit integration

Figure 2-5: Principle and implementation of the X rotation, or Rabi oscillation.

- **Actual implementation of Z rotation (Ramsey fringes)**

For Z rotations, since the qubit state is initialized in the +Z direction (|L⟩) on the Bloch sphere, we need to perform a 90° X rotation prior to and after the intended Z rotation to observe its effect. After the hold phase following the first 90° X rotation, we increase the detuning by pulsing both gates in opposite direction to facilitate a Z rotation. Between the hold phase and the Z-rotation phase, since the energy eigenstates remain the same, the edge rate does not need to be fast, and only the total area under the pulse determines the degree of Z rotation.

Critical requirements on control waveforms include the following.

- High-resolution, high-linearity, and low-uncertainty pulse height and width tuning: This enables accurate control of the frequency and degree of rotation.

- High-fidelity 90° X rotation: This is needed for precise characterization of the Z rotation.

(a) Operating point of Rz on charge stability diagram   (b) State evolution on the Bloch sphere

(c) Timing diagram and corresponding terminals in intended CMOS-qubit integration
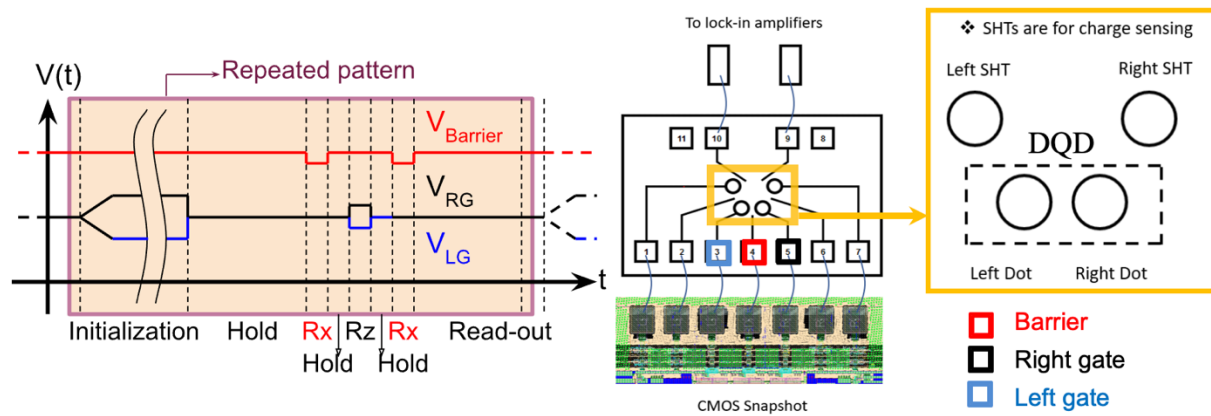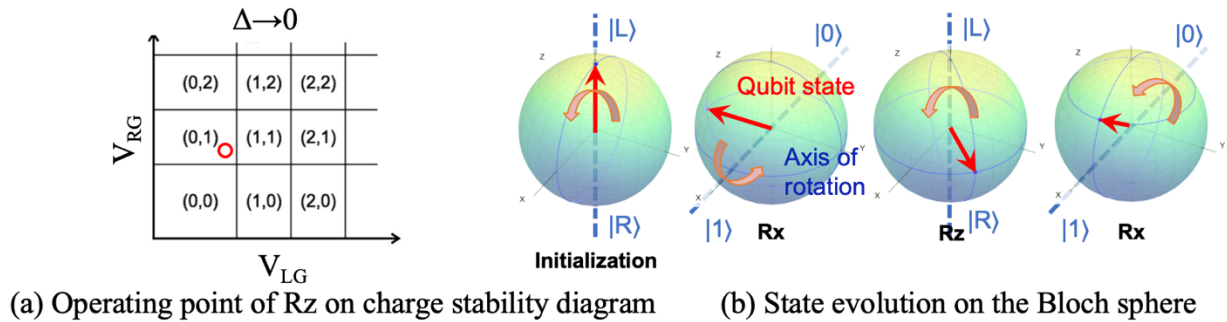
Figure 2-6: Principle and implementation of the Z rotation, or Ramsey fringes.

2.2.2. Requirements on control waveforms for high fidelity

- **Maximum error in pulse width**

Nonideal pulse width results in error in the degree of rotation. For high fidelity, the achieved maximum error in pulse width should restrict this error to less than 1%. Assuming the qubit's state is initialized in $|L\rangle$, or

$$|\psi_0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

at t=0 with nulled detuning, a pulse is applied to the barrier to facilitate the X rotation. The Hamiltonian with the interdot coupling can be written as

$$H = \begin{pmatrix} 0 & \Delta/2 \\ \Delta/2 & 0 \end{pmatrix}$$

By solving the time-dependent Schrödinger's equation, the time evolution of the qubit's state is

$$|\psi\rangle = U(t)|\psi_0\rangle = \begin{pmatrix} \cos\dfrac{\omega t}{2} & -i\sin\dfrac{\omega t}{2} \\ -i\sin\dfrac{\omega t}{2} & \cos\dfrac{\omega t}{2} \end{pmatrix} \times \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos\dfrac{\omega t}{2} \\ e^{-i\pi/2}\sin\dfrac{\omega t}{2} \end{pmatrix}$$

The fidelity after the X rotation (Jozsa 1994) then is

$$\mathcal{F}(\rho,\sigma) = \left(tr\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}} \right)^2 \approx |\langle\psi_\rho|\psi_\sigma\rangle|^2 = |\langle\psi_0|U(t_{final})U(t_{target})|\psi_0\rangle|^2$$

$$= \frac{1}{2} + \frac{1}{2}\cos(\omega t_{err})$$

The approximation holds due to the assumption of a decoherence time much longer than the time for the rotation. Finally, the error in pulse width

$$t_{err} = \frac{\cos^{-1}(2\mathcal{F}-1)}{\omega}$$

For $\mathcal{F} > 99\%$, the maximum error in pulse width can be calculated as a function of the Rabi frequency.

- **Maximum edge transition time**

During the edge transitions of the barrier pulse in an X rotation, non-zero edge transition time contributes to unintended rotation due to the intermediate time-varying Hamiltonian. After initializing the qubit's state to $|L\rangle$ and minimizing the detuning, edge transition occurs from time t_0 to t_1. The intermediate time-varying Hamiltonian and the time evolution matrix are

$$H = \begin{pmatrix} \varepsilon/2 & \Delta(t)/2 \\ \Delta(t)/2 & -\varepsilon/2 \end{pmatrix}$$

$$\hat{U}(t_1, t_0) \approx 1 + \frac{\tau\bar{H}}{i\hbar}$$

where $\bar{H}$ is the average Hamiltonian, $\overline{\Delta(t)} \approx (0 + \hbar\omega)/2$, and τ is the edge transition time. The expression of $\hat{U}(t_1, t_0)$ is simplified via the perturbation approximation to the first order. This simplification is valid due to the short transition time, $\frac{\tau\bar{H}}{i\hbar} \ll 1$. The impact of finite detuning shows up in higher-order terms and is eliminated during the simplification due to its trivial quantity. After the edge transition at t_1, fidelity $\mathcal{F} = 1 - \zeta$, where ζ is the adiabaticity defined in Adiabatic Theorem.

$$\zeta = 1 - \mathcal{F} = 1 - |\langle\psi_{t0}|\psi_{t1}\rangle|^2 \approx \frac{\tau^2}{\hbar^2}\left(\langle L|\bar{H}^2|L\rangle - \langle L|\bar{H}|L\rangle^2\right) = \left(\frac{\tau\omega}{4}\right)^2$$

Therefore, the edge transition time

$$\tau = \frac{4\sqrt{1 - \mathcal{F}}}{\omega}$$

For $\mathcal{F} > 99\%$, the maximum edge transition time can also be calculated based on the Rabi frequency. For advanced CMOS technologies, the practical minimum edge transition time is on the order of 10ps.

- **Trade-off between waveform requirements and operating temperature of the qubit**

Based on aforementioned requirements, we plot the maximum error in pulse width and maximum edge transition time for high fidelity versus the qubit's operating temperature in Figure 2-7 in various signal-to-noise ratio (SNR) scenarios. The SNR of our qubit is set by the ratio between the interdot coupling energy Δ and the thermal noise kT. At a given temperature, larger interdot coupling energy increases the SNR, but it also leads to a higher Rabi frequency, which makes the waveform requirements harder to meet and increases the power and area of the quantum controller. Finally, considering the shortest practical transition time in advanced CMOS technologies, we choose to operate our qubit at 50mK with a 5GHz Rabi frequency and a reasonable SNR of 5. The corresponding maximum error in pulse width is 6.4ps, and the maximum edge transition time is 12.8ps.
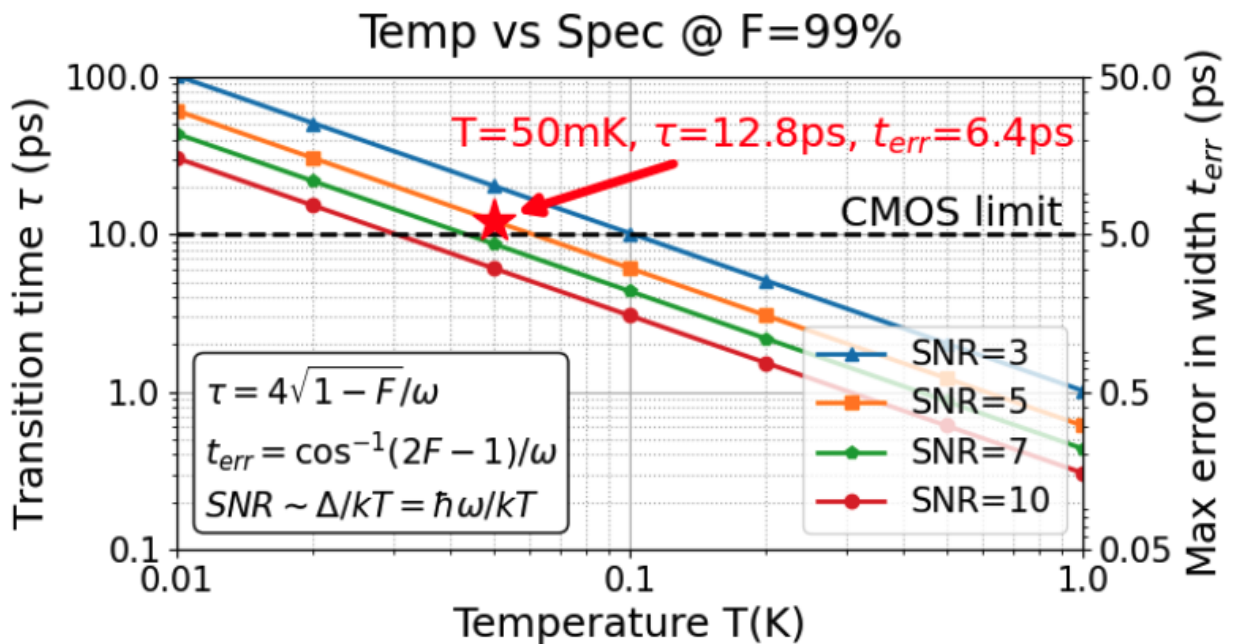
Figure 2-7: Precision requirements on control waveforms versus qubit operating temperature.

2.2.3. Fidelity bottleneck in physics experiments

Stringent requirements derived in the previous section is only realizable with a minimized interconnect between the quantum controller and the qubit. Meter-long cabling from the arbitrary waveform generator (AWG) to the qubit in physics experiments largely limited the fidelity of single-qubit quantum gates due to nonideal waveforms (Cao, et al. 2013). To alleviate the transmission line effect arising from the long cabling, some form of low-pass filtering is needed to limit the voltage overshoot and the prolonged voltage ringing immediately after transition edges. The filtered narrow-band pulse therefore has long edge transition times. Combining with the pulse-height-dependent overshoot and insufficient precision in width control of the AWG, the fidelity then is limited. Our work addresses these issues by proposing to integrate the output stage of the quantum controller with the qubit chip through a low-parasitic packaging. Together with high-precision tuning in pulse width and height, it enables wide-band pulsing with sufficiently short edge transition time and precise waveforms to achieve high-fidelity arbitrary X and Z rotations.

2.3. Proposed architecture for a scalable quantum processor

To achieve scalable quantum computing with high fidelity, we propose the architecture for the quantum processor illustrated in Figure 2-8. Aside from meeting stringent requirements on control waveforms, we envision that the quantum controller to be divided into two parts to be compatible with the available cooling power from the dilution refrigerator. Only the necessary low-power addressing and height control circuitry should be integrated with the qubit chip through a low-parasitic packaging inside the millikelvin chamber. The rest of the circuitry including the width control and the high-speed logic will be housed in a separate chamber at a higher temperature,

e.g., 4K, for its higher cooling power. The number of interconnects between chambers should be minimized to limit the thermal load.
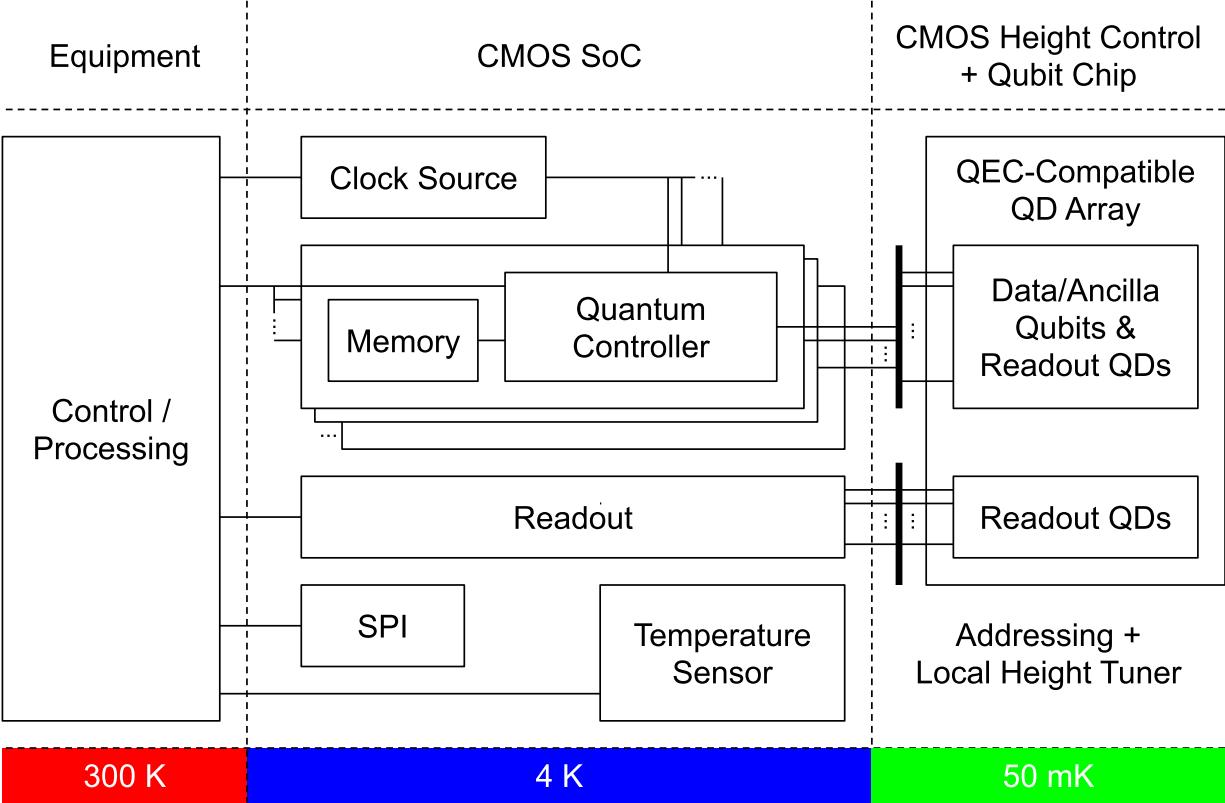
Figure 2-8: Proposed architecture of a scalable quantum processor.

Chapter 3 Temperature-insensitive quantum controller for high-fidelity universal single-qubit gates

3.1. Temperature-insensitive design methodology

A major challenge in designing complex, high-performance circuits for cryogenic operations is the lack of device models at cryogenic temperatures for simulations. To guarantee the circuit's performance for high-fidelity quantum gates, we propose temperature-insensitive design methodologies based on changes of device characteristics from room temperature to cryogenic temperatures reported in existing literature.

- **Digital-centric architecture**

A comparison of transistor characteristics at 300K and 4K is summarized in Table 3-1. Since our controller is designed with 28nm bulk CMOS technology, we focus on the measurement data from the 40nm CMOS samples. For analog circuits to main similar performances at cryogenic temperatures, the significantly increased threshold voltage (V_{th}) and variation in charge mobility mandate wide-range programmability. Worsened matching of FETs (Van Dijk, et al. 2020) and increased 1/f noise (Chang, Abidi and Viswanathan 1994) at cryogenic temperatures result in significant area penalty. Thankfully, effects such as carrier freeze-out and kink effect are not found to occur in nanometer-scale technologies (Van Dijk, et al. 2020).

Digital circuits with fast edges, on the other hand, are less sensitive to changes in V_{th} and charge mobility due to rail-to-rail signaling. Shorter gate delay leads to faster logic, and steeper subthreshold slope lowers the leakage. A key design parameter for synthesized digital circuits is the hold time margin. Later in this section we propose a hold time margin at 300K to ensure robust functionality at cryogenic temperatures while maintaining efficiency in area and power.

From previous analysis, we propose a digital-centric implementation and strive to maximize its performance. For analog amplifiers, we incorporate sufficient programmability to cover the extreme variations in the I-V characteristics of devices.

Table 3-1: Comparison of transistor characteristics at 300K and 4K (Van Dijk, et al. 2020).

Technology		0.16 μm		40 nm	
Temperature		4 K	300 K	4 K	300 K
Device W/L	[$\mu m/\mu m$]	2.32 / 0.16		1.2 / 0.04	
V_T	[V]	0.55	0.40	0.50	0.38
SS	[mV/dec]	22.8	87.0	27.7	88.2
n	[-]	28.7	1.5	34.9	1.5
I_{on}	[A]	$2 \cdot 10^{-3}$	$1.5 \cdot 10^{-3}$	$6 \cdot 10^{-4}$	$5.3 \cdot 10^{-4}$
I_{off}	[A]	$< 3 \cdot 10^{-11}$	$< 1.6 \cdot 10^{-10}$	$< 1.5 \cdot 10^{-12}$	$< 1.4 \cdot 10^{-10}$
I_{on}/I_{off}	[A/A]	$> 6.7 \cdot 10^7$	$> 9.4 \cdot 10^6$	$> 4.0 \cdot 10^8$	$> 3.8 \cdot 10^6$
Gate delay	[ps]	30.60	38.30	-	-
λ	[V^{-1}]	3.3	0.6	4.0	1.3
		Weak Inversion			
g_m/I_D	[V^{-1}]	70	27	92	27
Intrinsic gain = $g_m/(\lambda I_D)$	[V/V]	21.2	45.0	23.0	20.8
		Strong Inversion (at $V_{ov} = 0.2\ V$)			
g_m/I_D	[V^{-1}]	6	9	9	10
Intrinsic gain = $g_m/(\lambda I_D)$	[V/V]	1.8	15.0	2.2	7.7
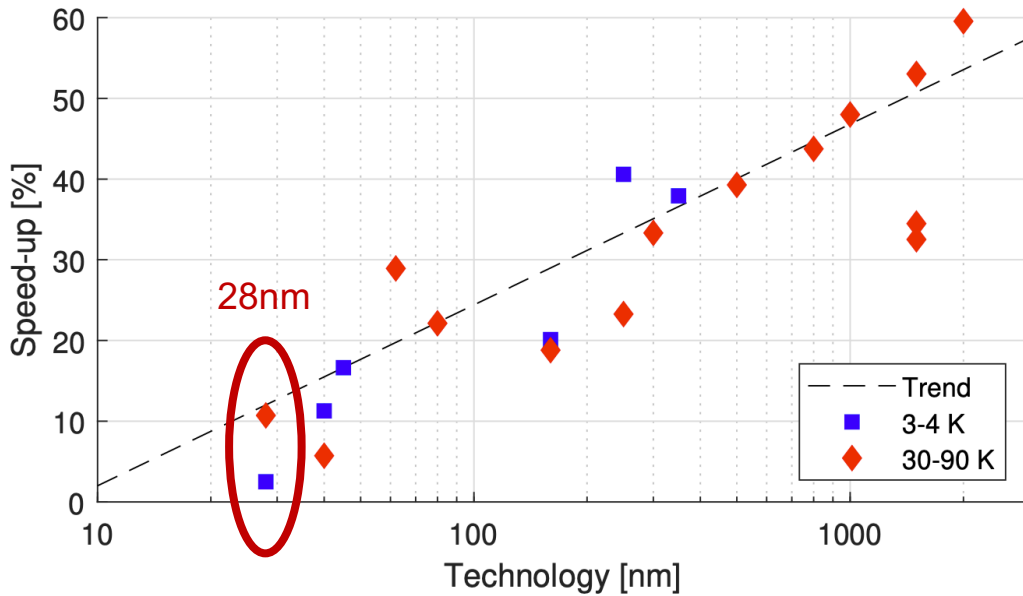
Figure 3-1: Speed-up of digital gates at cryogenic temperatures (Van Dijk, et al. 2020).

- **Capacitor-matching-based linearity-sensitive circuits**

As explained in section 2.2, precise control over the pulse shape is the key to achieve high fidelity. Thus, highly linear converters are indispensable. To fulfill required speed and resolution, current digital-to-analog converter (DAC) and capacitive DAC (CDAC) are two viable options.

Current DACs utilize matching of FETs for linearity. From ('t Hart, et al. 2018) it was reported that the matching property (σ/μ) of the drain current of long-channel FETs, suitable for implementing current sources, can deteriorate up to 20% from 300K to 4.2K, and this is mainly due to increased mismatches in V_{th}. Due to the worsened matching, these FETs need to be sized 20% larger just to meet the same linearity at 4.2K.

CDACs, on the other hand, utilize matching of capacitors for linearity. To our best knowledge, there has not been any literature systematically characterizing matching of capacitors at cryogenic temperatures. However, since the matching solely depends on the ratio of geometry, it is not expected to degrade as temperatures drops (Van Dijk, et al. 2020). In this work, we report the indirectly measured matching property of metal-oxide-metal (MOM) capacitors at 300K and 77K. Another advantage of choosing CDAC-based width and height control is its drastically lower power consumption. All sub-circuits can be designed in the digital fashion and, excluding leakage, consume only dynamic power during switching activities. For the height control, the purely capacitive qubit terminal is also a perfect fit for the CDAC's charge-redistribution operating principle. Therefore, we propose to utilize capacitor matching for linearity-sensitive circuits.
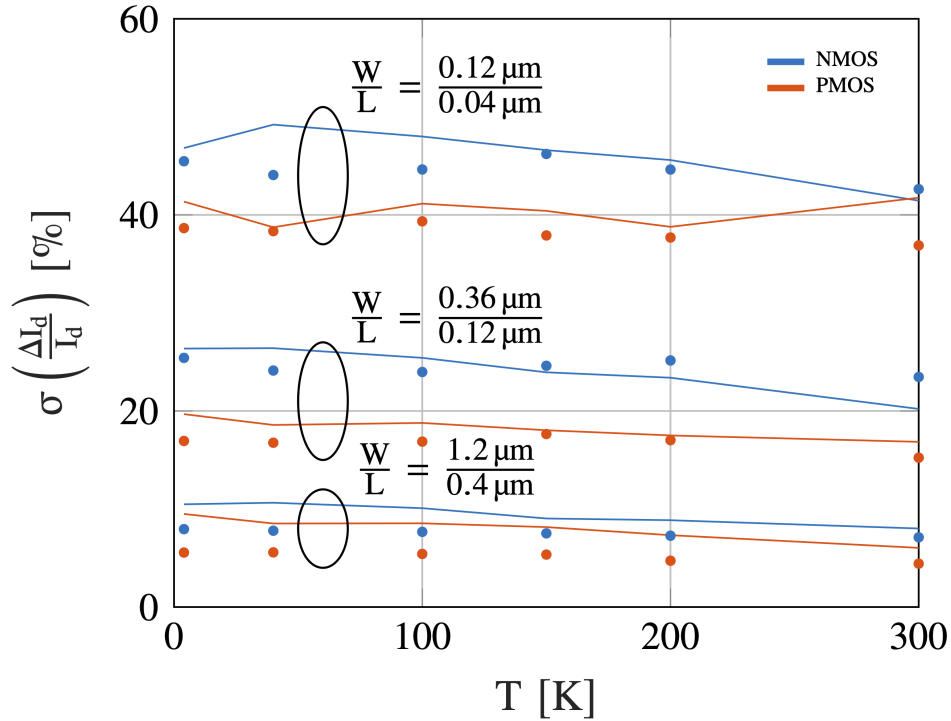
Figure 3-2: Worsened matching in drain current as temperature drops (biased at a fixed g_m/I_D of $10V^{-1}$ in saturation) ('t Hart, et al. 2018).

- **50ps hold time margin**

During the design of synthesized sequential logic digital circuits, closing timing with proper setup and hold time margin is arguably the most important step. However, without modeling at cryogenic temperatures, it is impossible to predict the margin at those temperatures accurately. Insufficient setup or hold time leads to loss of circuit functionality. Setup time violation can be resolved during the chip testing by slowing down the clock rate and increasing the voltage supply to speed up the circuits. Hold time violation, however, cannot be fixed on silicon, and requires extra iterations of costly tape-out to correct. As illustrated in Figure 3-3, for the register REG2 to properly read the output of a combinational logic, the output must hold steady for at least T_h after the arrival of clock. Proper hold time margin is reserved to guarantee positive hold slack in all cases. Hold time violation occurs when the hold slack becomes negative. Since the relative

timing between the combinational logic delay and T_h has no dependency on the clock rate and the supply voltage, hold time violations cannot be resolved during testing.

To ensure robust functionality at cryogenic temperatures, we must reserve sufficient hold time margin in the design phase. Inserting buffers in the combinational logic can increase the margin. However, this comes with a cost of larger chip area and higher power consumption. Thus, the margin must be decided methodically.
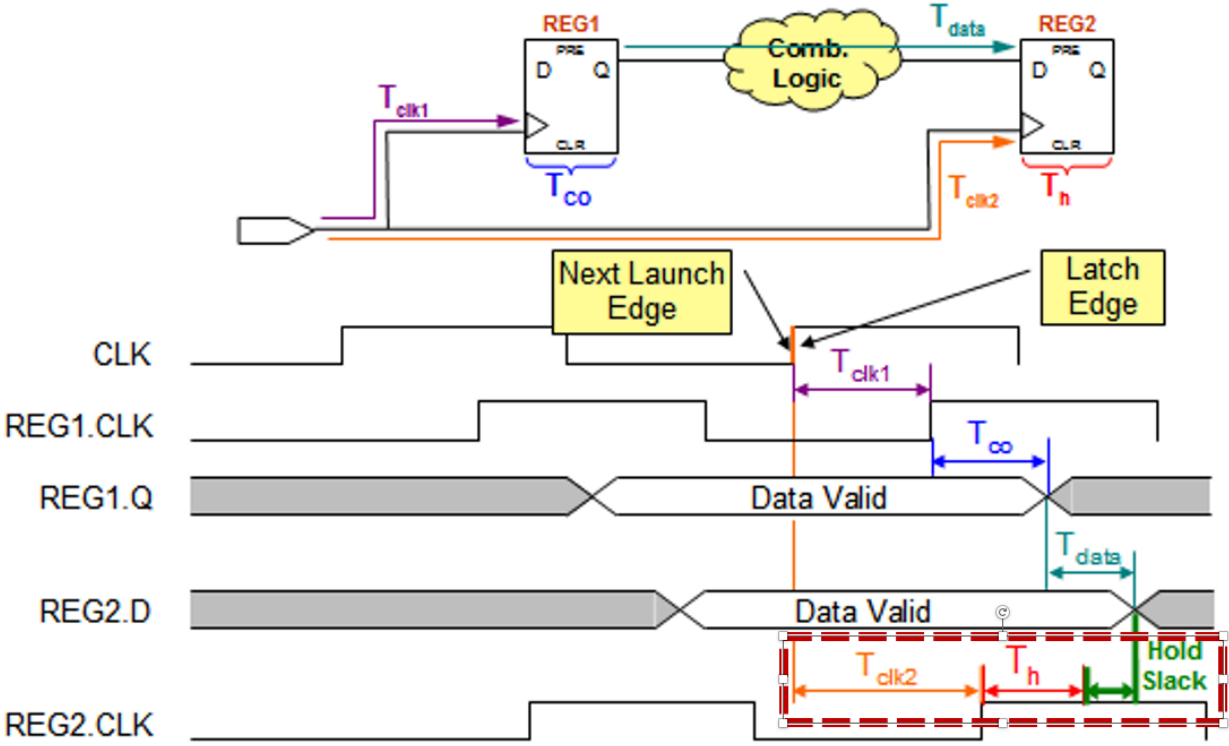
Figure 3-3: Illustration of hold time requirement (Zhang 2016).

As temperature drops, the change in CMOS gate delay is dictated by competing factors such as increased V_{th} and variation in charge mobility. Depending on their changes relative to the voltage supply (V_{DD}), different trends of gate delay have been observed. When the increase in V_{th} dominates, FETs turn on slower and the gate delay increases. From the delay per stage of a 101-stage ring oscillator in 28nm bulk CMOS (Ismail-zade, et al. 2020) shown in Figure 3-4, the increase in gate delay below 300K can be as high as 70%. On the other extreme, when the increase in charge mobility dominates, the "on" current increases. Assuming that the changes in parasitic capacitance of FETs is trivial, the input of the succeeding gate gets charged faster, which leads to reduced gate delay. From the measured "on" current versus temperature reported in (Beckers, et al. 2017) shown in Figure 3-5, the reduction in gate delay from 300K to 4.2K can be as high as 30%.
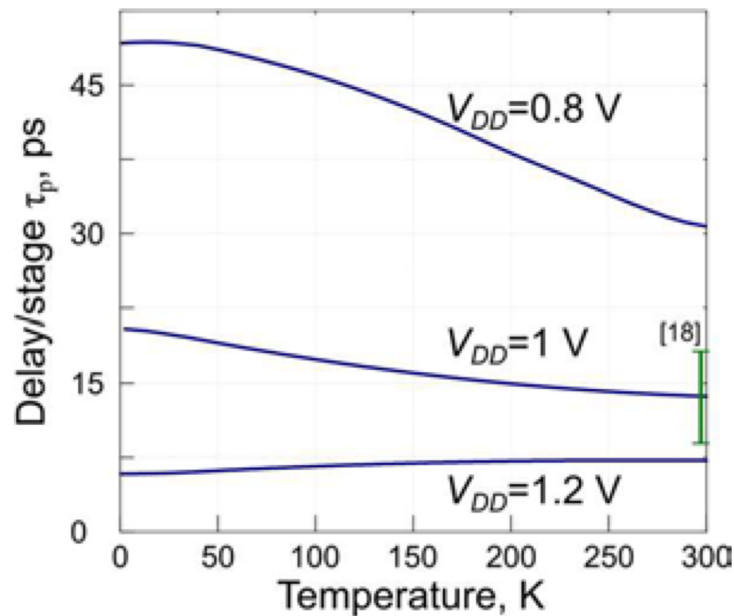
Figure 3-4: Delay per stage of a 101-stage ring oscillator in 28nm bulk CMOS versus temperature (Ismail-zade, et al. 2020).
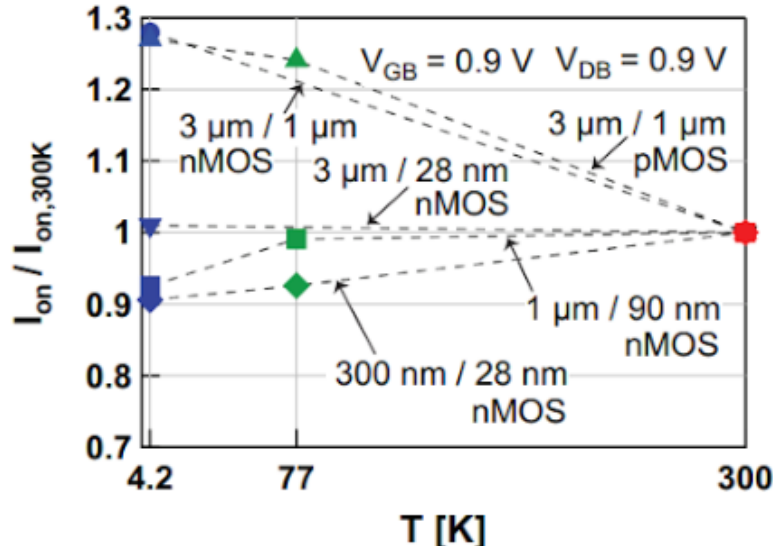
Figure 3-5: "On" current versus temperature in 28nm bulk CMOS (Beckers, et al. 2017).

With the knowledge of extreme variations in the gate delay, we calculate the required hold time margin at 300K to ensure robust functionality at cryogenic temperatures. At room temperature, T_h in our 28nm standard cell is 5ps, and a typical hold time margin is 10ps. The worst scenario for the hold time constraint occurs when the T_h increases to the upper bound and $T_{CQ} + T_{data}$ reduces to the lower bound. Here we assume that with the synthesized clock tree, clock arrives at each register at the same time and is insensitive to temperature changes. To have the same ratio of T_h and hold slack at cryogenic temperatures, the required hold time margin can be derived below.

$$(5p + Hold\ slack|_{300K}) \times 70\% = (5p + 10p) \times 170\%$$

$$\Rightarrow Hold\ slack|_{300K} = 31ps$$

To reserve more margin for millikelvin operations, we specify the hold time margin to be 50ps for all digital blocks, which should guarantee their functionalities at cryogenic temperatures.

- **Summary of temperature-insensitive design methodologies**

  o Adopt digital-centric architecture with fast edge transitions.

  o Incorporate sufficient programmability for analog circuitry.

  o Utilize capacitor matching for linearity-sensitive circuits.

  o Design with 50ps hold time margin at 300K for synthesized sequential digital circuits.

3.2. Quantum controller design with temperature-insensitive performance

3.2.1. Limitation of state-of-art dc-pulse-based controllers

To enable high-fidelity universal single-qubit gates for our qubit, a maximum error in pulse width of 6.4ps is required. The major contributors to this error include the finite tuning resolution, maximum absolute value of differential nonlinearity (max |DNL|), and uncertainty in pulse width due to random noise. In state-of-art dc-pulse-based quantum controllers, the tuning resolution of pulse width in (Esmailiyan, et al. 2020) (Ekanayake, et al. 2008) was limited to one period of their input clock since digital counters triggered by the input clock were used to program the pulse width. This is too coarse since even with an input clock as fast as 6GHz (Esmailiyan, et al. 2020), the resolution was merely 166ps per step, which is 26 times larger than the required maximum error in pulse width. In (Pauka, et al. 2021), an on-chip free-running ring oscillator generated the master clock, and the linear width tuning was done through varying the number of stages in the oscillator. The tuning resolution could be as fine as two times the delay of an inverter, which is on the order of 20ps, but this delay could vary significantly since there was no PVT-compensation or calibration implemented. The shortest pulse width was also limited to one clock period due to the digital-counter tuning mechanism. The 166ps-wide pulse reported in (Esmailiyan, et al. 2020) is already

~300° of a 200ps Rabi period. Max |DNL| was not reported in any of the state-of-art controllers. Uncertainty in pulse width was only reported in (Ekanayake, et al. 2008), but it was on the order of hundreds of picoseconds, which is at least 15 times over our targeted maximum error in pulse width. Also, among all these works there was no calibration to ensure accuracy of the pulse width.

Precise pulse height tuning can enable accurate control of the rotation frequency, minimize unintended rotations, and reduce crosstalk between qubit terminals. In state-of-art controllers, only (Esmailiyan, et al. 2020) implemented on-chip control of pulse height. However, its linearity was limited to a max |DNL| of 10.4LSB with 8-bit programmability.

The concept of integrating a CDAC-based output stage with the qubit chip in the millikelvin chamber was demonstrated in (Pauka, et al. 2021). However, its maximum repetition rate of successive quantum gates was limited to around 10MHz to keep its power consumption lower than the cooling power of the 100mK chamber. This bottleneck can be broken by moving most of the controller to another chamber at a higher temperature while leaving only the output stage in the millikelvin chamber. (Esmailiyan, et al. 2020) integrated the controller with the qubit at 3.4K. Although switching activities in its CDAC were limited to only the enabled units, continuous clocking of those units by the master clock at 2GHz still incurred noticeable overhead in power consumption. This can be improved by replacing the continuous clocking scheme with a pulse-activated design.

Lastly, on-chip generation of all control waveforms needed for universal single-qubit gates is preferrable as it reduces the number of high-speed I/O needed for the cryogenic chamber. I/O interconnects with higher bandwidth result in larger thermal load, and thus further consume the already limited cooling power budget. Among the state-of-art controllers, only (Pauka, et al. 2021) was capable of generating arbitrary waveforms through on-chip memories.

3.2.2. Proposed architecture for the quantum controller

To support high-fidelity universal single-qubit gates for our qubit, we have implemented a temperature-insensitive quantum controller in 28nm bulk CMOS as shown in Figure 3-6. Figure 3-7 details the chip layout and the intended CMOS-qubit integration. This prototype is designed not only for high-fidelity operations, but also to explore the optimal operating point and to study the decoherence and relaxation behavior of the qubit. To do the latter, significant amount of extra circuitry is added to the design. Once the qubit is fully characterized, the extra circuitry can be removed without sacrificing the precision, and the controller can be further partitioned into two parts as described in section 2.3.

Illustrated in Figure 3-6 is the architecture and timing diagram of the proposed quantum controller. A 50Ω-terminated, highly programmable clock amplifier at the front end amplifies the external 1GHz sine wave to a rail-to-rail clock signal, and it is the only analog amplifier in the entire chip. This clock is then utilized by a 12-bit, 3.9ps/step width controller to generate pulses needed for quantum gates. It also drives the dual-edge-triggered core logic for all dynamic controls, pulse gating for power efficiency, and to generate all required waveforms for universal single-qubit gates. The linearity of the width tuning is calibrated by a 2ps/step pulse-swallowing cyclic TDC to ensure temperature-insensitive performance. After the width controller, the 9-bit, nominally 1mV/step, CDAC-based height controller fine tunes the pulse height and delivers final control waveforms to the left gate, barrier, and right gate of the DQD. DC biasing circuitry for the source and drain of the DQD is not shown. For synchronized lock-in read-out from charge sensors next to the DQD, our controller can also embed the lock-in pattern in all output waveforms.
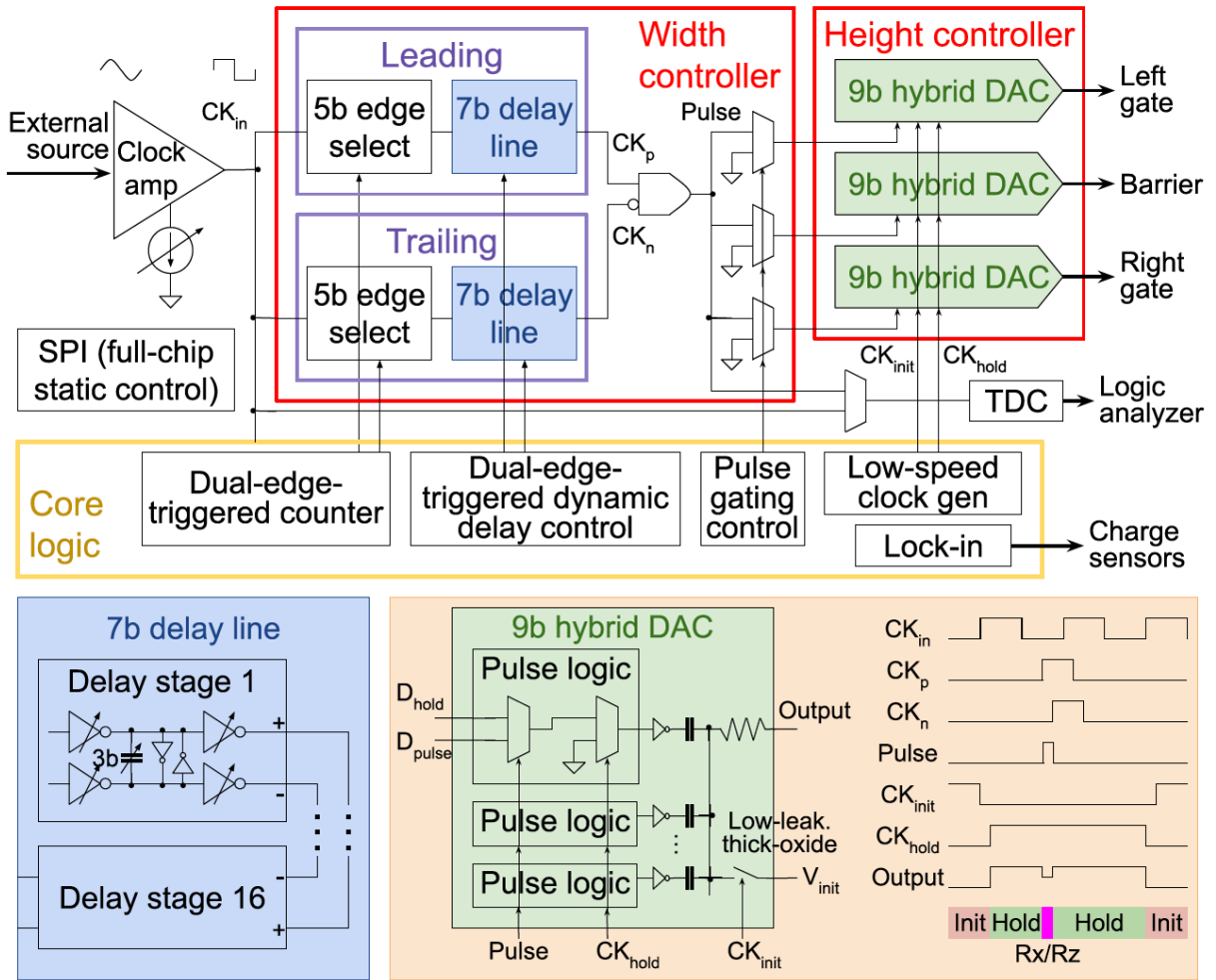
Figure 3-6: Architecture and timing diagram of proposed quantum controller.
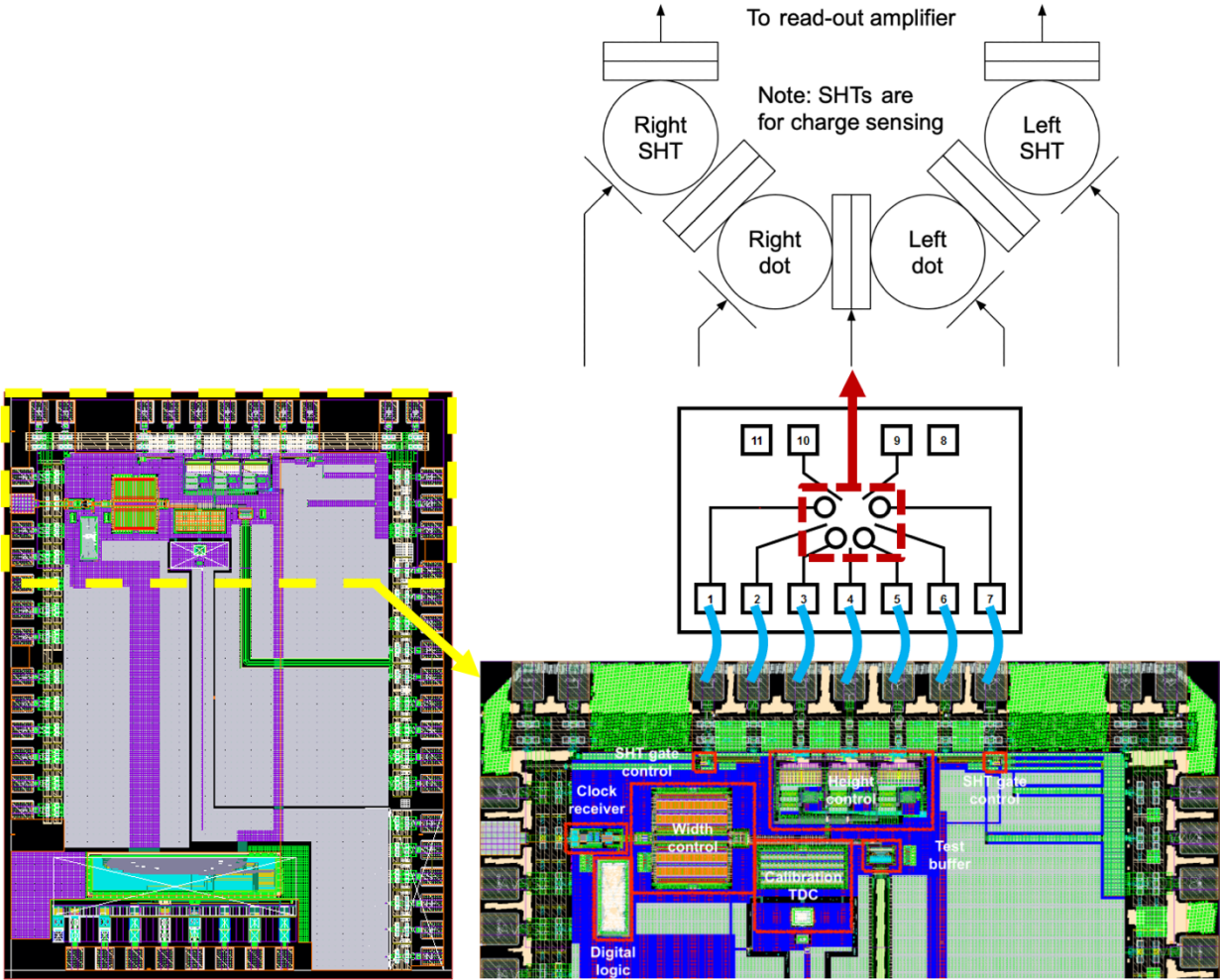
Figure 3-7: Chip layout and intended integration with the qubit.

3.2.3. Temperature-insensitive width controller

- **Tuning resolution**

A two-step architecture is employed to implement the 12-bit width tuning. The pulse width is defined by a leading and a trailing 1GHz clock with different delays as shown in Figure 3-8. The 5-bit coarse tuning covers a 16ns range with a 500ps step. Then the 7-bit fine tuning further dissects the 500ps range with a 3.9ps step.

The coarse tuning is achieved by selecting two edges of the 1GHz input clock. Extensive simulations have been done to close the timing including the synthesized core logic and the edge

selection circuitry on the signal path. The duty cycle of the input clock can be distorted due to nonideal operating point of the clock amplifier and unbalanced rising and falling edge transition on the signal path. This introduces errors to the nominal step size of 500ps. The error is calibrated by the cyclic TDC and thus the tuning resolution is temperature-insensitive. During the calibration, the biasing voltage of the NFET current sources, V_{bn}, in the clock amplifier is adjusted with an 8-bit voltage DAC (VDAC) as shown in Figure 3-9. All VDACs implemented in our controller has an output range from ground to its supply voltage to ensure coverage of extreme variations in transistor characteristics.

The fine tuning is achieved through varying the capacitive load in the 3-bit inverter-C delay stages in the 7-bit delay line. The step size of 3.9ps is also made temperature-insensitive thanks to the calibration by adjusting its supply and the bulk voltage of the NFETs in the delay stages, V_{bb}, through the 8-bit VDAC as shown in Figure 3-10.
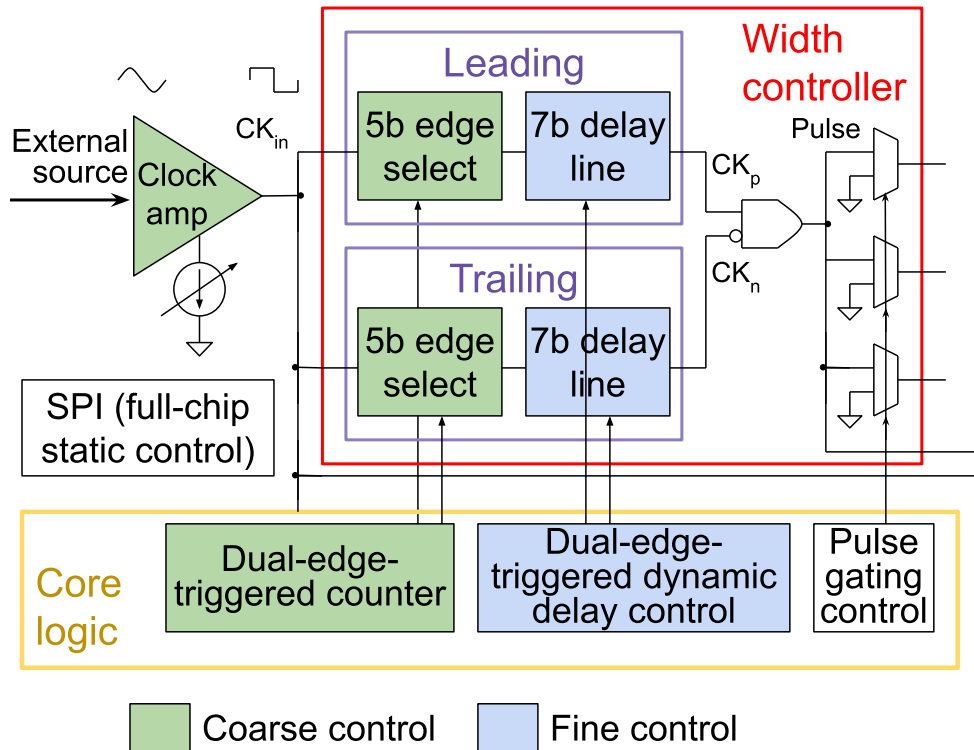
Figure 3-8: Two-step architecture for width control.

34
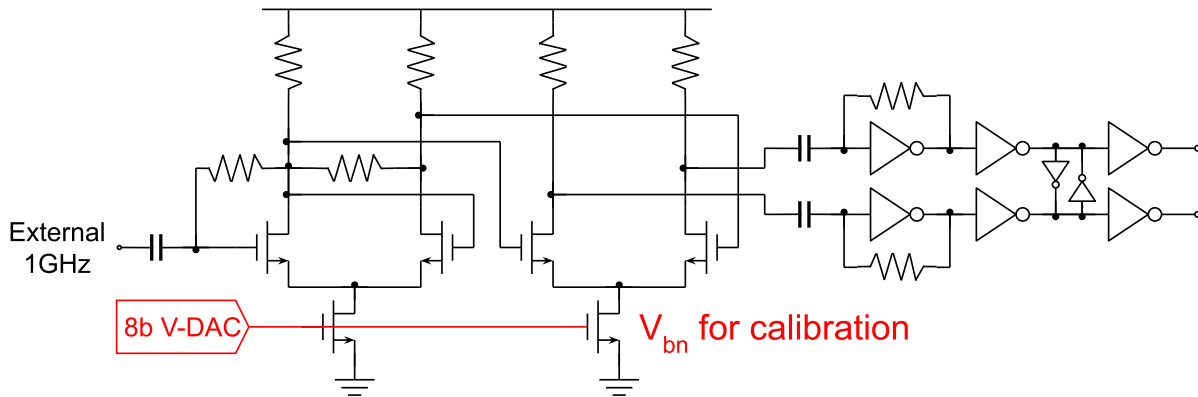
Figure 3-9: Schematic of the programmable clock amplifier.
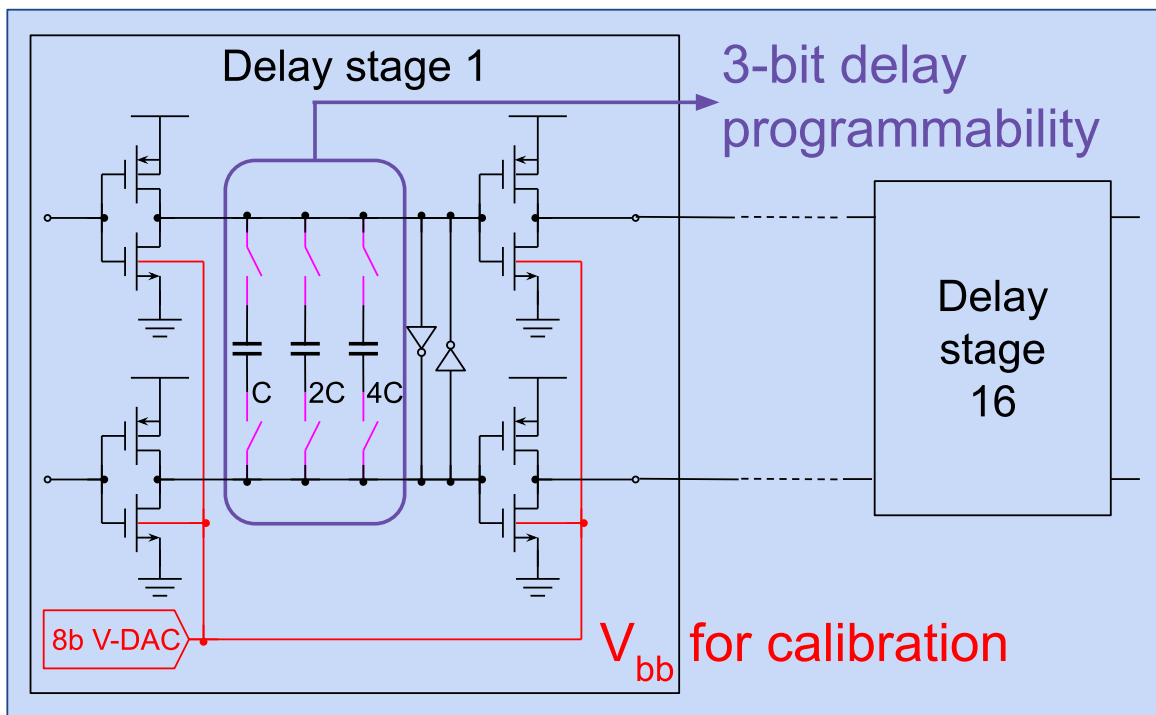
Figure 3-10: Schematic of inverter-C delay stage in the 7-bit delay line.

- **Linearity**

To minimize max |DNL|, we first analyze the major contributors to the nonlinearity in width tuning. In Figure 3-11, the deviation of the full scale of fine tuning from 500ps introduces gain error in its transfer curve. The nonideal duty cycle of the input clock further adds to the nonlinearity when the coarse tuning advances by one step, which is from code 127 to 128 and from code 255 to 256. Thanks to the well-matched inverter-C delay stages, the max |DNL| occurs either from code 127 to 128 or from code 255 to 256. Lastly, supply ripples due to transient current at transition edges also temporarily change the local supply voltage and introduce nonlinearity.

The max |DNL| is calibrated by the cyclic TDC through adjusting V_{bn} and V_{bb} as shown in Figure 3-9 and Figure 3-10, and thus made temperature-insensitive. For a linear fine tuning, we utilize the matching of MOM capacitors and adopt a segmented capacitor array (4-bit thermometer-coded MSBs + 3-bit binary-coded LSBs). The FETs in the inverter-C delay stage are sized for good matching such that the worst integral nonlinearity (INL) of the 7-bit delay line is around 0.45LSB (see Figure 3-12). To reduce supply ripple, we design a pseudo-differential, cross-coupled signal path with abundant bypass capacitors under the supply as shown in Figure 3-10. Illustrated in Figure 3-13 is the final compact, well-matched layout of the 7-bit delay line.

➢ Nonideal slope:

range of fine tuning $\neq$ 500ps

➢ Large |DNL| near code 128 and 256:

imperfect duty cycle of input 1GHz
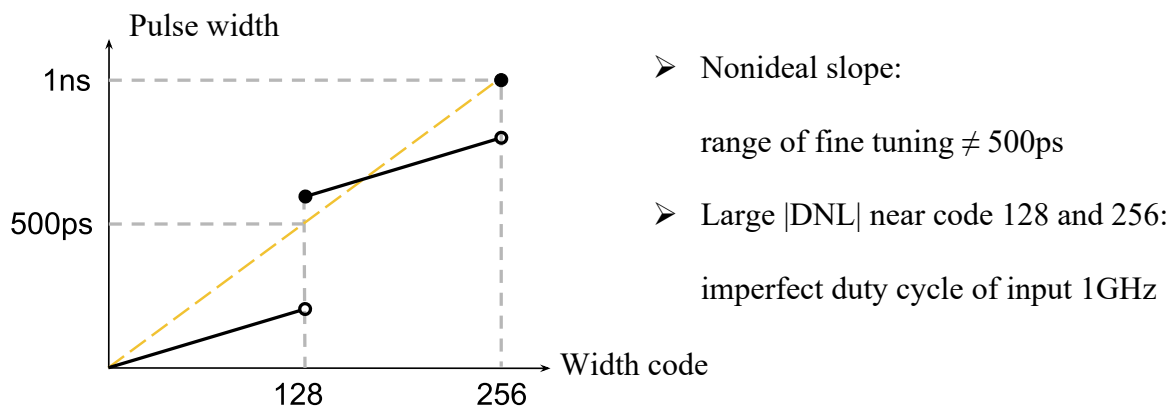
Figure 3-11: Major contributors to the nonlinearity of width control.
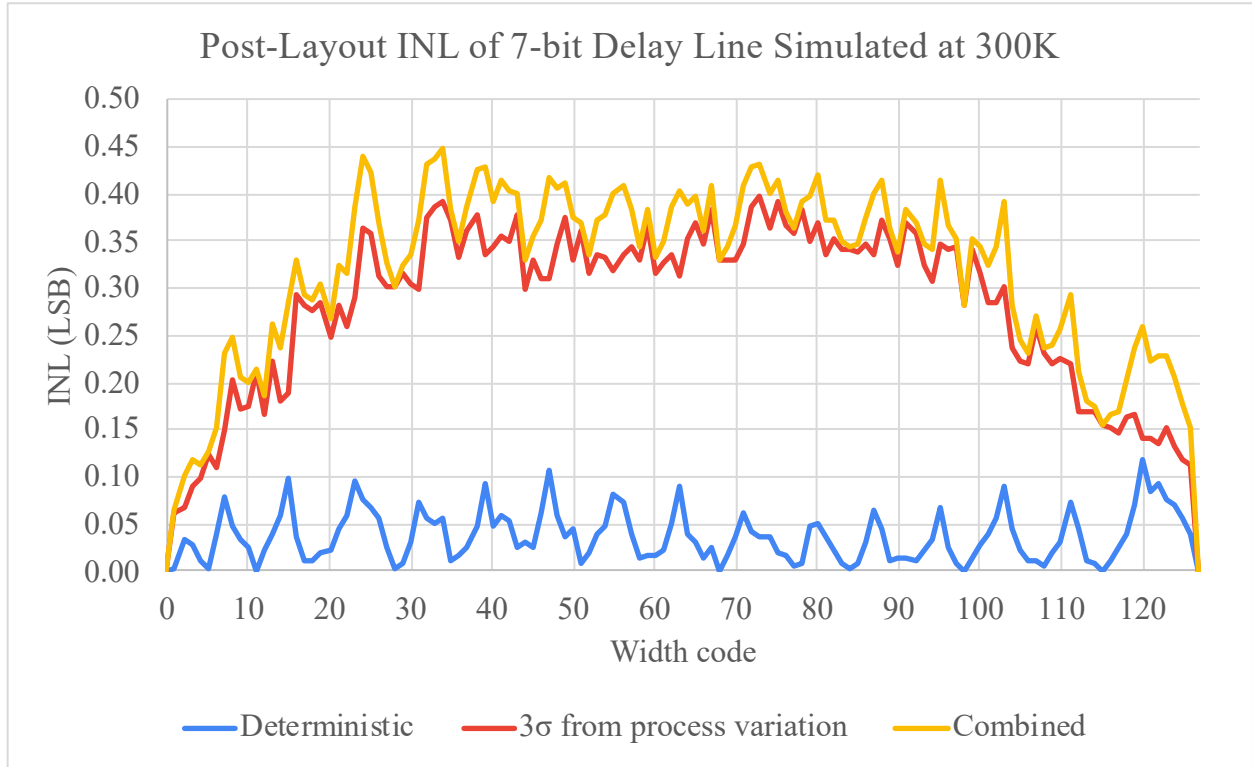
Figure 3-12: Post-layout simulated INL profile of the 7-bit delay line.
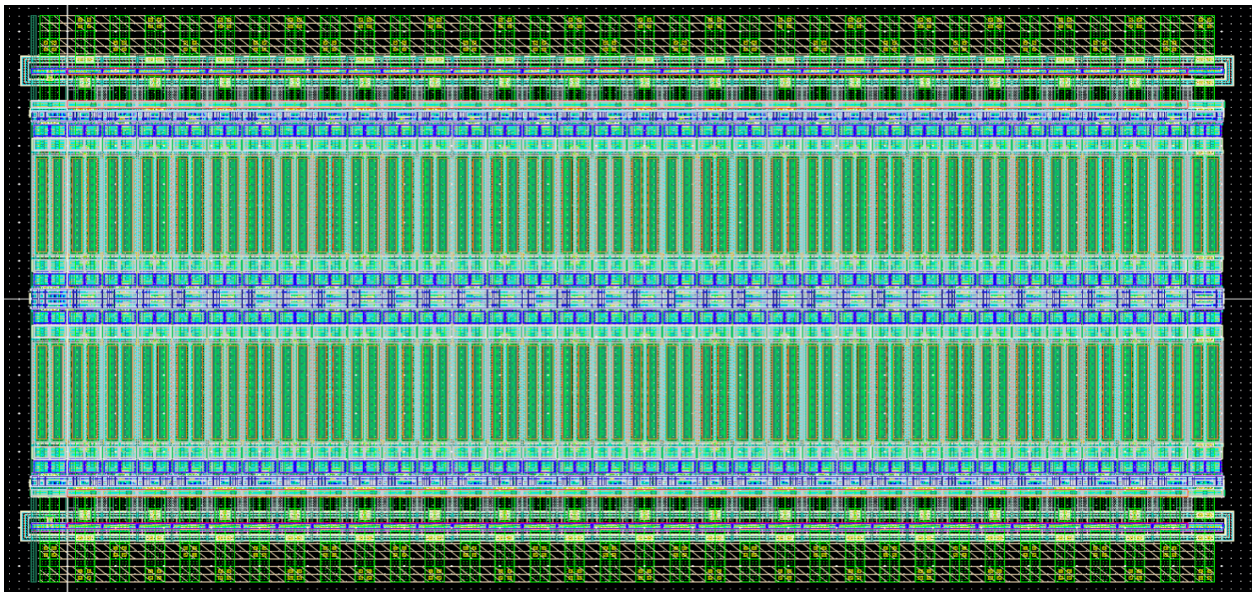
Figure 3-13: Layout of the 7-bit delay line for good linearity.

- **Uncertainty in pulse width**

We aim to minimize the uncertainty in pulse width for pulses shorter than 500ps. Since the expected Rabi period is 200ps, pulses up to 500ps long are sufficient to facilitate arbitrary X rotations. Pulses longer than 500ps are mainly used for decoherence and relaxation studies, and thus their accuracy requirement is not as stringent.

The rising and falling edge of pulses up to 500ps are generated from the same edge of the input clock. Therefore, noise from circuitry before the width controller is largely cancelled. Edge transitions throughout the entire signal path are designed to be fast to minimize the translation from noise to uncertainty in width. In the 7-bit delay line, the wider range covered in each delay stage will reduce the number of delay stages but result in slower edge transitions as the same inverter has to drive larger capacitive load, and this increases the width uncertainty. Therefore, through simulations we strike a balance between the number of bits per delay stage and the worst-case uncertainty in pulse width. At cryogenic temperatures, since the calibrated delay stages have a similar edge rate as that at 300K, lower thermal noise leads to smaller uncertainty. Signal path in the subsequent height control also benefits from lower temperature due to similar reasons.

- **Calibration with pulse-swallowing cyclic TDC**

A 2ps/step pulse-swallowing cyclic TDC is developed to maximize the precision of width tuning as there is no oscilloscope with such high resolution in time. The TDC consists of a loop of skewed inverters, whose PFET and NFET have asymmetric driving strengths. Due to this asymmetry, the rising and falling edge at the input of the inverter experience different amount of delay, and this difference further depends linearly on the output load. Therefore, by sizing one of the inverters larger than the others, the pulse will be shrunk. Figure 3-14 illustrates the operating

principle explained above and the governing equation for pulse shrinkage. Finally, the pulse width can be inferred simply by counting the number of times the pulse is shrunk before vanishing.

To realize 2ps shrinkage per stage with small uncertainty, we select a β of 3 under a 0.9V supply. A total of 64 shrinkage stages plus one pulse entry stage comprise the loop and exhibit a loop delay longer than 1ns. Since the loop delay must be longer than the maximum pulse width for proper functionality, the designed delay enables calibration of the TDC against the accurate 1ns period of the external clock. To enhance the linearity of the TDC, similar to the signal path, it is designed with a pseudo-differential, cross-coupled architecture to reduce supply ripple. The schematic and layout are shown in Figure 3-15 and Figure 3-16, respectively.

From the analysis in (Chen, Lin and Hwang 2014), we expect the resolution of the cyclic TDC to change as temperature drops. From testing, the shrinkage per stage reduces by 5% from 300K to 77K, which can be easily compensated by slightly lowering the supply voltage. Tuning the supply is a more desirable solution than adding circuits for temperature compensation, since the latter requires a much higher supply and thus significantly increases the power consumption.

During testing, the step size of the TDC will be first calibrated against the accurate 1ns period of the 1GHz input clock. Afterwards, the calibrated TDC will be used to optimize the precision of the width tuning.

$$\Delta W = T_{\text{in}} - T_{\text{out}} = \left(\beta - \frac{1}{\beta}\right) C_i \left(\frac{1}{k_{P_i}} - \frac{1}{k_{N_i}}\right) \times \phi$$

Loop delay > max pulse width for proper functionality
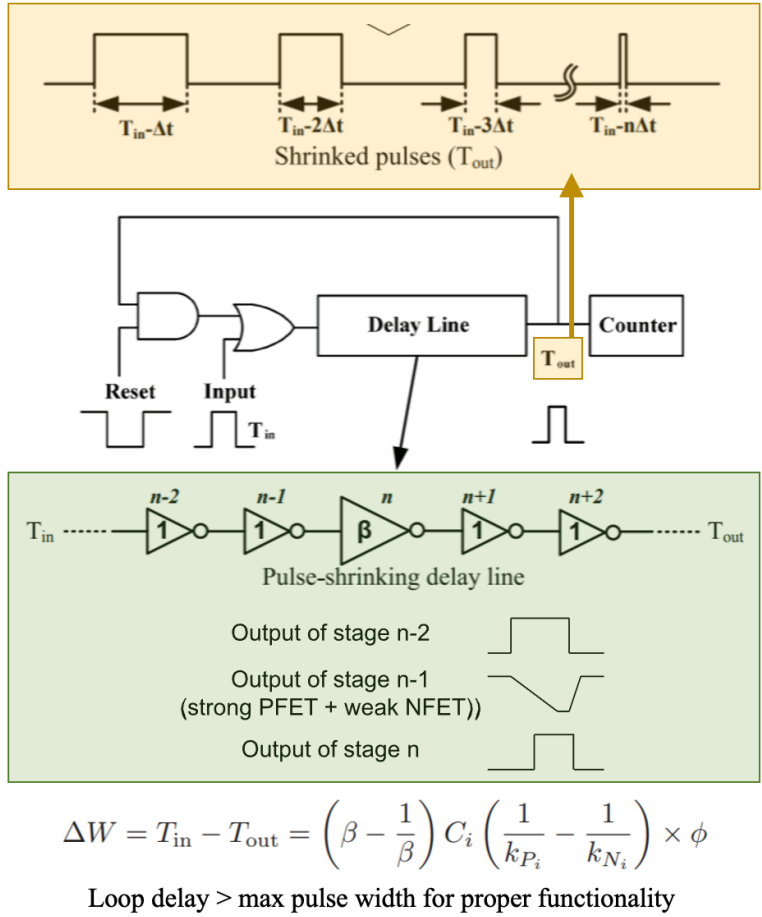
Figure 3-14: Principle of pulse-swallowing cyclic TDC (Chen, Lin and Hwang 2014).
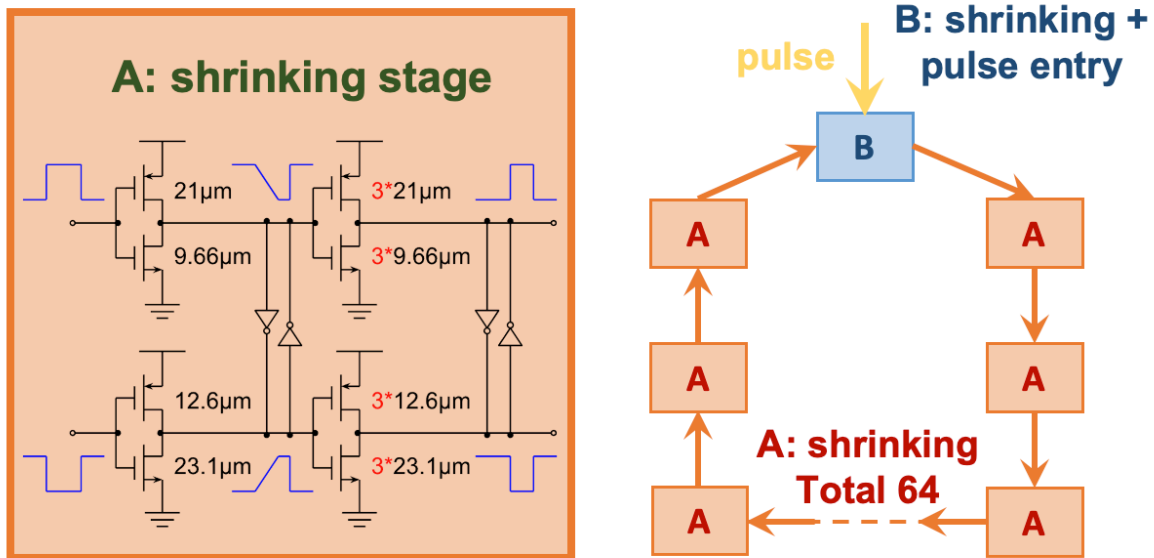
Figure 3-15: Schematic of our pulse-swallowing cyclic TDC.
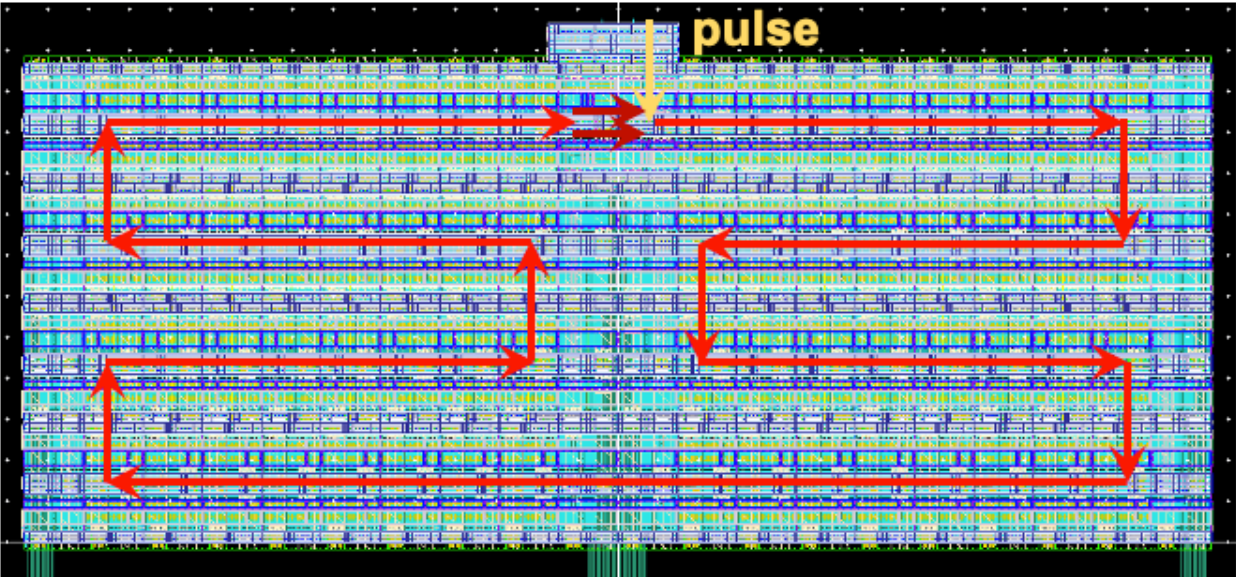
Figure 3-16: Cyclic and matched layout of the TDC for good linearity.

➢Cal ref:
External 1-GHz clock
➢Cal result:
2ps shrinkage/stage

➢Cal ref:
2ps-shrinkage/stage
TDC
➢Cal result:
Minimized max |DNL| of
width control
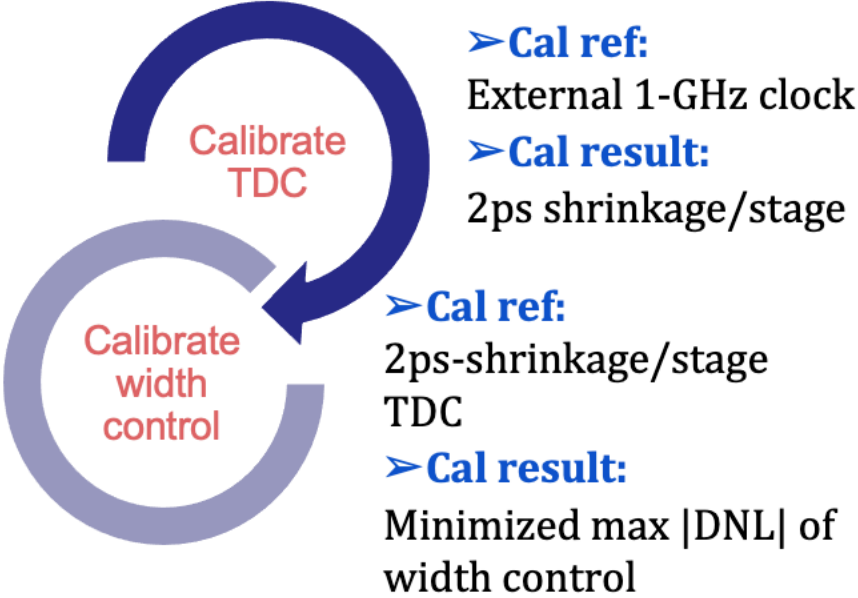
Figure 3-17: Proposed calibration flow for optimal precision in width control.

3.2.4. Temperature-insensitive height controller

The height controller is the output stage of our quantum controller, which directly interfaces with the qubit and charge sensors to perform initialization, manipulation, and read-out. Low-power CDAC-based design is adopted to take advantage of the purely capacitive terminals on the qubit chip. Each phase of operation has its own requirement on the control waveform. Therefore, aside from the precise CDAC, additional features supporting all phases of qubit operations are also incorporated to support high-fidelity quantum gates.

- **Features for high-fidelity operations**

During initialization, at the CMOS-qubit interface we need wide-range, low-speed sampling of dc voltages to properly bias the qubit at the desired charge regime and to avoid unintended excitation. These dc sampling switches are implemented with thick-oxide I/O FETs which also exhibit low leakage when they are turned off in other phases of operation. During manipulation and read-out, the fast pulses are generated by switching the CDAC to satisfy the transition time requirement as derived in section 2.2.2. A series pulse-shaping polysilicon-based resistor is added to the output to minimize the voltage overshoot, and its negligible leakage to the substrate is verified through simulations. In these phases we keep the impedance of the CMOS-qubit inface high to avoid significant voltage drooping which undermines the biasing of qubit and in turn shortens its decoherence time. The ESD diodes are sized just for sufficient protection. Per-terminal hybrid DAC is implemented for independent pulsing to address potential crosstalk between qubit terminals. Lock-in pattern in the output waveforms can be enabled for lock-in read-out in low-SNR settings. All features mentioned above are shown in Figure 3-6 and Figure 3-18.

- **Linearity and resolution**

Requirements on the resolution, full scale, and linearity of the CDAC are dictated by the actual physical properties of the DQD. In this work we define the specifications according to the hard-wall-confined DQD (Yang, et al. 2016) (Gorman, Hasko and Williams 2005) mentioned in section 2.1. To cover a full scale of 500mV with output inverter drives powered by a 0.9V reference voltage, we co-design the 1pF total switchable capacitors in the CDAC with attenuation capacitors, extracted parasitic capacitance at the output, and the anticipated loading from the qubit. To determine the segmentation scheme of the capacitor bank and the sizing of the custom-made unit MOM capacitors for good linearity, we simulate the worst-case nonlinearity through 10,000-point Monte Carlo simulations in MATLAB. Figure 3-19 shows the DNL and INL profile with the final segmentation scheme and assuming conservative capacitor matching property. To reduce the layout size of the 9-bit CDAC, the capacitor with weighting of 1LSB is made of a 2fF unit capacitor while higher-weighting capacitors consist of 4fF unit capacitors. Figure 3-20 shows the mixed-unit-capacitor implementation. The smaller capacitor banks also shorten the routing to qubit terminals and further reduce parasitic capacitance. The estimated matching of the 2fF unit capacitor is 0.7% (Tseng, et al. 2016) (Tsai, et al. 2015) (Saberi, et al. 2011), and the simulated max |DNL| is 0.2LSB.

To stabilize the reference voltage during full-scale pulsing, sufficient bypass capacitors are added to reduce the voltage drop at the pulse edges. The simulated integrated voltage noise at the output is negligible in comparison with 1LSB, and the thermal floor will be even lower at millikelvin temperature. Finally, the pulse-gating mechanism in the width controller in concert with the pulse-activated logic in the CDAC minimize switching activities and lower the power consumption for millikelvin operations.
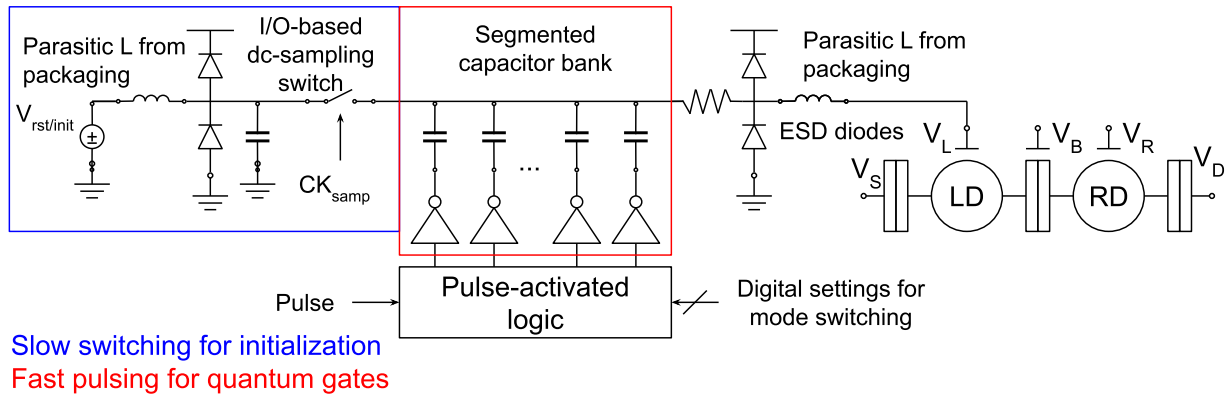
Figure 3-18: Conceptual diagram of the height controller.
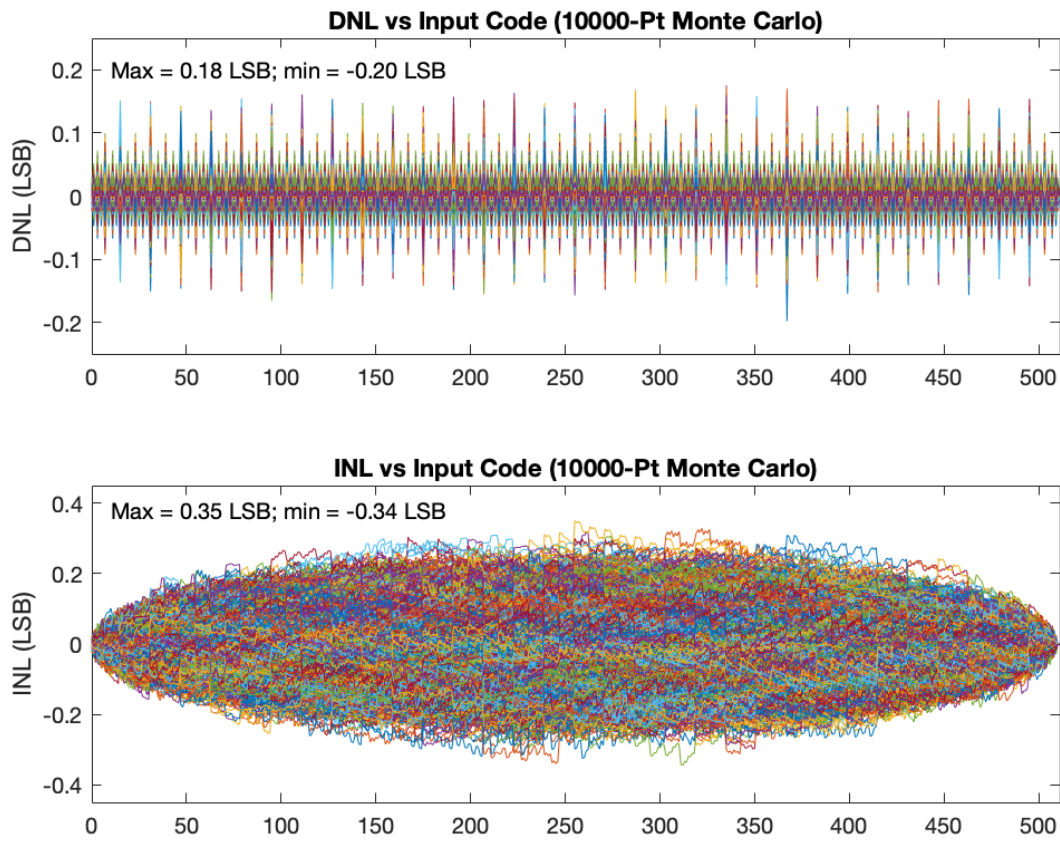
Figure 3-19: Linearity of the segmented CDAC (5 thermometer-coded MSBs + 4 binary-coded

LSBs) with 0.7% σ_C/μ_C of the 2fF unit capacitor.
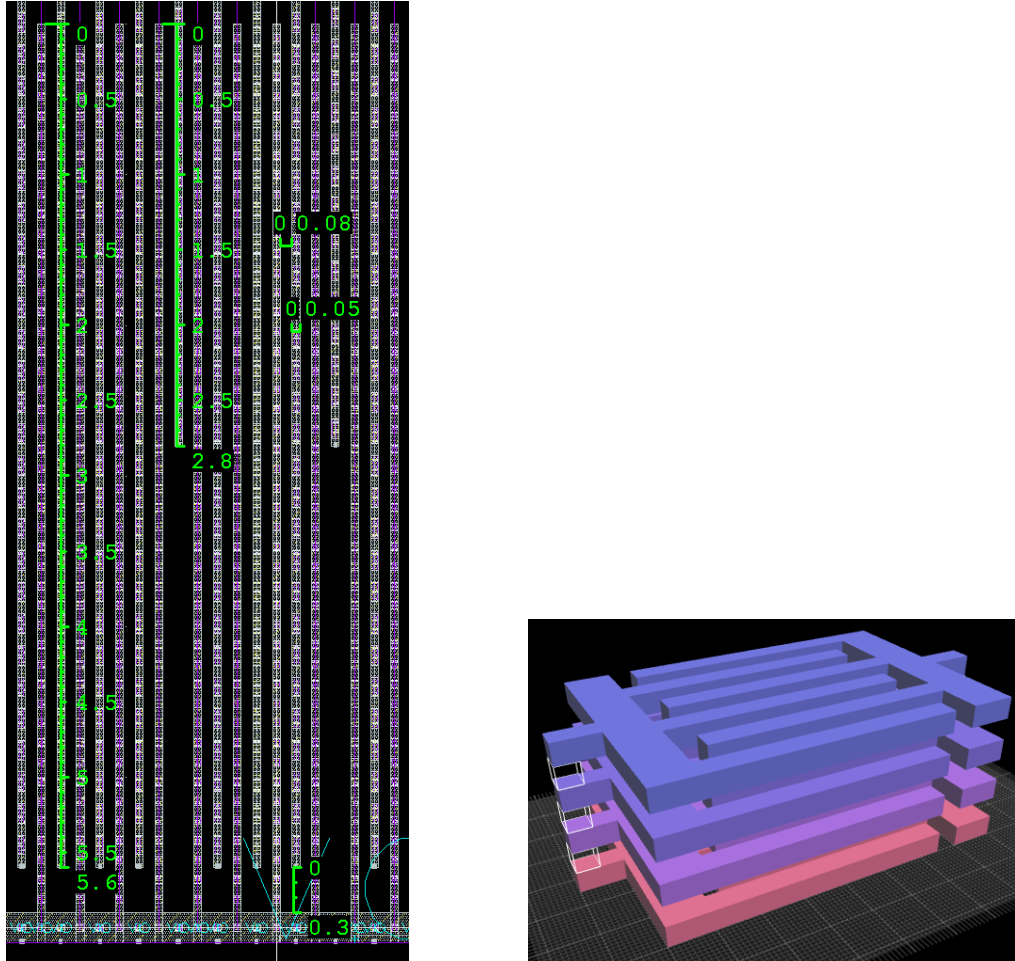
Figure 3-20: Layout and conceptual diagram (Matthias 2022) of custom-made MOM capacitor

using metal 2 to metal 5.

3.2.5. Core logic and SPI

The synthesized core logic and serial peripheral interface (SPI) are necessary for all dynamic controls and static programmability for the controller. The SPI is low-speed and relatively straightforward, while the core logic controls the width and height controller at 1GHz speed. To avoid glitches at its outputs which introduce unexpected distortion to the control waveform, the core logic is implemented as a Moore machine, which is a finite-state machine whose output only depends on its internal state. Also, the outputs are intentionally buffered such that the timing

analysis results are independent of the external load. The core logic defines synchronously the control waveforms for all terminals of the DQD and charge-sensors, lock-in read-out, triggering clock for the TDC's digital output, and digital settings for the whole controller.

To achieve low worst-case uncertainty in pulse width, we limit the full scale of the fine tuning to 500ps, which mandates a step size of 500ps in the coarse tuning. Since the coarse tuning is done through selecting edges of the 1GHz clock by the digital counter in the core logic, the counter needs to be dual-edge-triggered. As dual-edge-triggered flip-flops are not available in our 28nm standard-cell library, the counter design is separated into two parts, i.e., the rising- and falling-edge-triggered. To avoid glitches on the output edge selection signal sent to the signal path, we combine the output of both parts with an XOR gate, which is the simplest circuitry possible.

Chapter 4 Liquid nitrogen dip testing

4.1. Cryogenic testing setup

4.1.1. Mixed-signal automated testing

As depicted in Figure 4-1, we develop a highly automated mixed-signal testing environment to calibrate and optimize performance at 300K and 77K. Numerous dc supplies (Keysight E3631A) are used for biasing and a signal generator (Keysight E4433B) generates the input clock. The chip's functions are controlled through a microcontroller (Arduino Nano) that converts the commands from the PC to the on-chip SPI. Output bits of the on-chip TDC are collected by a logic analyzer (Digilent Analog Discovery 2), and the output control waveforms are displayed through a calibrated oscilloscope (Keysight 54845A/54810A). An additional semiconductor parameter analyzer (Keysight 4155C) is used to calibrate the nonlinearity of the on-board op-amp-based buffer and the oscilloscope for linearity measurement of the pulse height.

- **Cross-platform integration for automated measurements**

To achieve closed-loop calibration for the quantum controller, we first build customized software and firmware for fully automated measurements and data processing. This requires a vertical integration across platforms including the PC, microcontroller, logic analyzer, lab instruments, and our quantum controller.

On the microcontroller, the firmware is written to facilitate the communication between the PC and the quantum controller. It converts messages between the USB format and USART format in real time. In addition, the signals transmitted between the microcontroller and the quantum controller are level-shifted from 5V to 1.8V on a breadboard for reliability of the chip. Through this, the software on the PC directly controls and monitors the behavior of the quantum

47

controller and automates the testing sequence. The logic analyzer receives the 50Mbps serialized output from the on-chip TDC. We do not use the microcontroller for this purpose since its clock rate is only 16MHz. The analyzer is configured to operate at its maximum speed of 100MHz to log the data for further processing. To control and collect data from lab instruments automatically, we use Keysight GPIB-to-USB adapters for the link. It helps convert the signals from USB to the standard GPIB protocol (IEEE 488.2) for instruments. This way we can directly control the instruments through instrument-specific SCPI commands.

Aided with the microcontroller, logic analyzer, and GPIB-to-USB adapters, the software on the PC acts as a control center directly interacting with our quantum controller and instruments. In addition, to verify the communication link during the dip testing at 77K, the software checks whether the return data format from the quantum controller is as expected or not after every command. Upon detection of errors, the testing sequence is interrupted, and the corresponding error messages are displayed for debugging.

Measurements of the pulse width are fully automated. We program the software to sweep the pulse width from the narrowest measurable to the widest acceptable for the TDC, which covers at least 1ns range. This is sufficient to characterize the width controller because, from the signal path point of view, the next 1ns range utilizes the same circuitry, and thus similar nonlinearity is expected. From the distribution of the TDC's output, the pulse width is obtained by averaging, and the DNL and INL are calculated automatically. The automation is a key enabler for multi-variable performance calibration. Since the variables are interdependent as discovered in section 4.2.1, simultaneous optimization of all variables is needed to converge to the global optimum. This optimization in the large search space can only be done with automated testing.

Pulse height measurements are also fully automated. The software is programmed to sweep the pulse height, adjust the semiconductor parameter analyzer's dc output, and collect the measured data from the oscilloscope. To reduce deterministic errors, the timing of measurements is automatically adjusted according to the pulse edges to ensure sufficient settling. The dc output of the semiconductor parameter analyzer is also automatically adjusted for each height setting to minimize the impact of nonlinearity. To alleviate the effect of noise, the software turns on the averaging function of the oscilloscope after changing the height setting. Before advancing to the next height setting, the software turns off the averaging function to clear the memory from the last setting. In this way, we can collect measurement data with the highest fidelity.
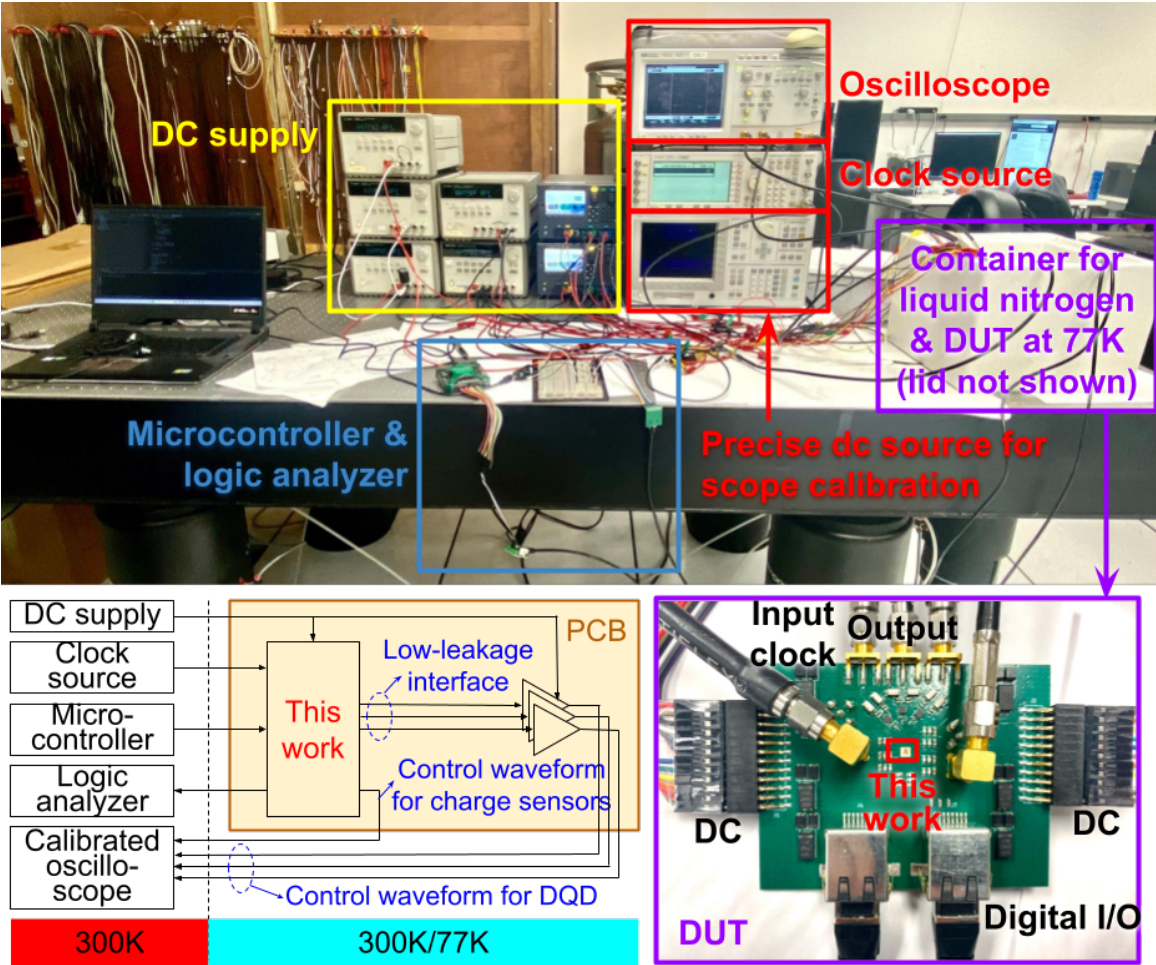
Figure 4-1: Mixed-signal automated setup for liquid nitrogen dip testing.

4.1.2. Testing with thermal shock

To perform dip-testing at 77K, a major challenge is to minimize mismatches in the coefficient of thermal expansion (CTE) to alleviate mechanical stress at the interface between materials due to the thermal shock. In fact, as shown in Figure 4-2, our first chip after dipping is completely broken due to the mechanical stress from the encapsulating epoxy. This chip seizes to function entirely even after thawing.

The unbearable stress comes from significant mismatches in the CTE. The silicon of our CMOS chip has a CTE of 2.6-3.3ppm/K while that of the epoxy is around 55-80ppm/K. Also, the epoxy is hardened after curing. Thus, after dipping the epoxy shrinks much faster which cracks the chip. To address this issue, no encapsulating epoxy is applied to following chips. Chemically, due to the oxidation process at the surface of the chip, it is isolated from the liquid nitrogen. But mechanically it is not protected, and so extra precaution is taken during assembly and testing.

Mismatches in CTE also increase the risk of detached bonding wires from the on-chip aluminum pads. Without the encapsulating epoxy, the bonding wires can move after thermal cycles and short to adjacent wires. To alleviate risks, we opt for aluminum wires to match the material of on-chip pads. One bonding wire per pad instead of multiple in parallel also prevents the short.

Finally, CTE mismatches between the gold-plated pad on the PCB, the adhesive for attaching the CMOS chip to the PCB, and the silicon substrate of the chip contribute to significant stress on the chip as well. This stress does not visibly break the chip, but 10-times-higher leakage current is observed, and the chips frequently lose their function. High-CTE GE Varnish, Super Glue, and silver epoxy are later replaced with a special low-CTE adhesive from MasterBond to enable reliable dip-testing at 77K. Table 4-1 summarizes CTE of the adhesives.
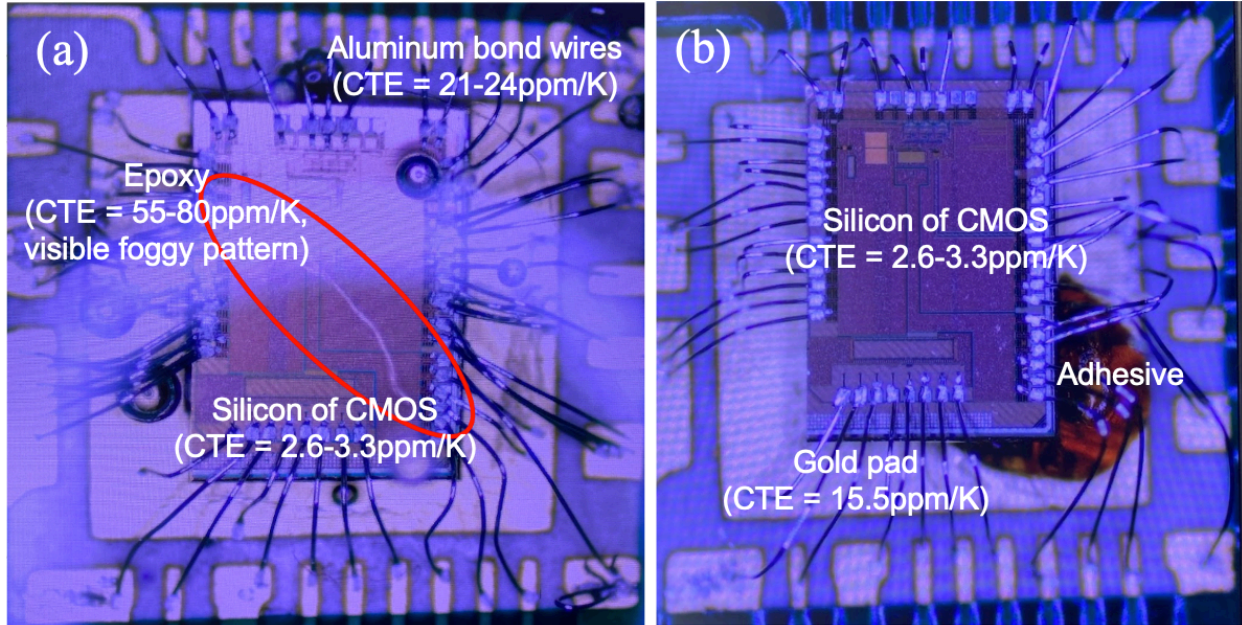
Figure 4-2: Chip-board assembly: (a) Cracked chip due to large CTE mismatches between CMOS chip and epoxy. (b) Final working design with low CTE mismatches.

Table 4-1: CTE of adhesives.

Adhesive	CTE (ppm/K)
MasterBond EP21TCHT-1	18-21 (Masterbond n.d.)
GE Varnish	~50 (Morelock, Suchomel and Wilkinson 2013)
Loctite Super Glue	80 (Loctite 1996)
MG 8331D (silver epoxy)	234 (MG Chemicals 2021)

4.1.3. Suitable on-board components and PCB material

- **On-board RC**

On-board RC components with low temperature dependency are crucial for proper chip operations at cryogenic temperatures. For resistors, the metal film type has the lowest temperature coefficient down to 83K, especially for small-value resistors (R < 100Ω). For larger-value resistors, the thick film type has the lowest temperature coefficient (Patterson, Hammoud and Gerber 2001).

51

For capacitors, capacitance of NP0 ceramic capacitors stays almost unchanged down to 4K. But since commercial ceramic capacitors are only available up to 470nF, for larger-value capacitors, tantalum-polymer capacitors work best (Teyssandier and Prêle 2010). Detailed comparisons are shown in Figure 4-3 and Figure 4-4.

- **PCB material**

Low-moisture-absorbing PCB materials such as RO4000 is quite reliable for cryogenic testing. On the other hand, the much more economic FR-4 is also found to work down to 4K (Homulle 2019). To maximize the compatibility of our PCB with popular cryogenic testing setup such as the dewar, we lay out our PCB as compact as possible. For high-speed 50Ω transmission lines, to preserve signal integrity and to minimize the area, we choose RO3010 for the dielectric layer underneath which has a high dielectric constant of 10.2 and a low loss tangent of 0.002. FR-4 is chosen for other dielectric layers for its much lower cost.
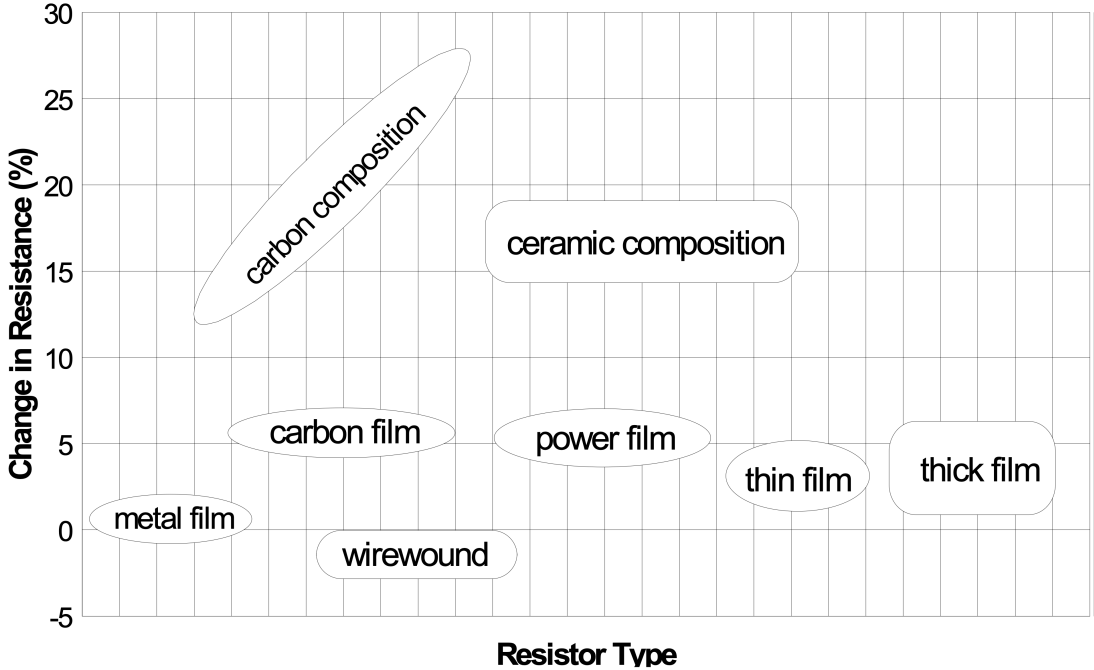
Figure 4-3: Change in resistance of various resistor types from 298K to 83K (Patterson, Hammoud and Gerber 2001).
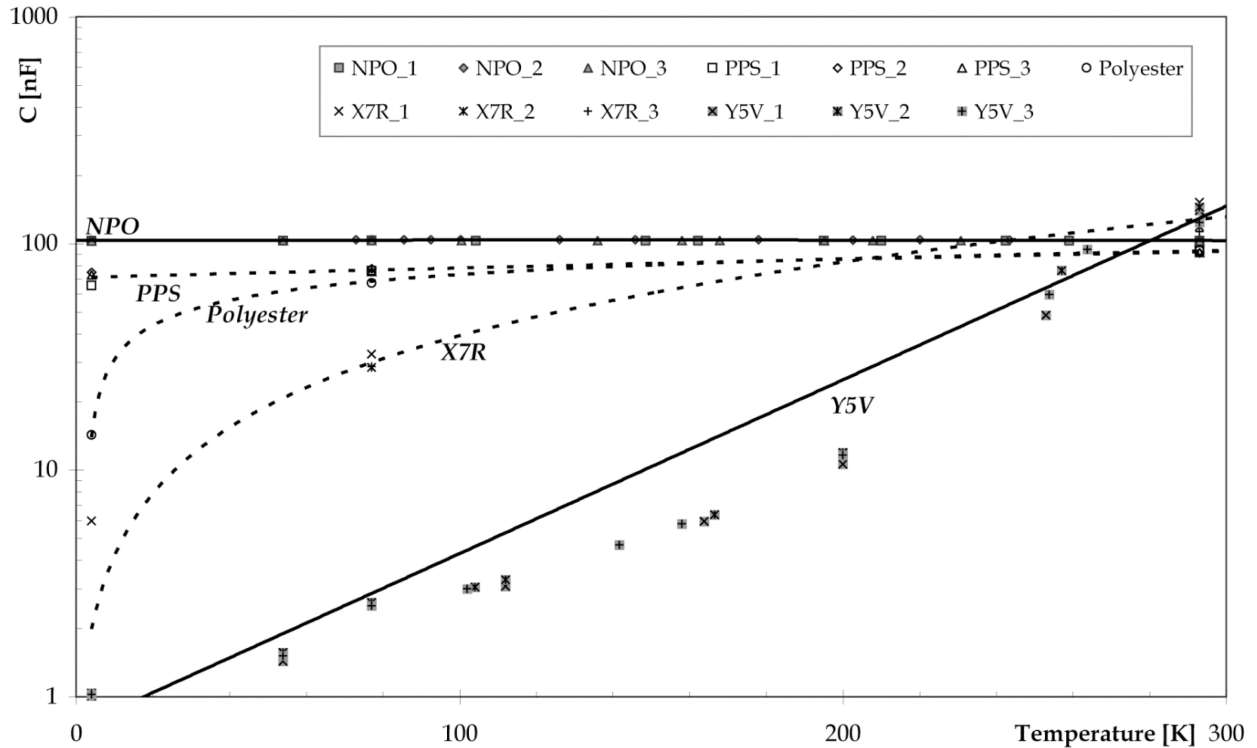
Figure 4-4: Capacitance of various capacitor types versus temperature from 300K to 4K

(Teyssandier and Prêle 2010).

4.1.4. On-board output buffer

Since our height controller is designed to drive the purely capacitive terminals of a qubit, to drive the 50Ω scope for pulse height measurements and the display of control waveforms, we need a capacitive-input on-board buffer with an output impedance matched to 50Ω. Leakage at its input must be sufficiently low to avoid large voltage drooping during CDAC switching. Our first attempt is a common-source amplifier built with a microwave pseudomorphic high-electron-mobility-transistor (pHEMT). However, to resolve an oscillation issue preventing the measurement, in our second attempt we replace it with a lower-speed, high-linearity operational-amplifier-based feedback amplifier.

- **pHEMT-based common-source amplifier**

Our first version of the on-board output buffer is a common-source amplifier built with a GaAs pHEMT (CE3521M4 from California Eastern Laboratories). Based on its datasheet, when biased at the operating point as highlighted in Figure 4-5, it has a g_m of ~50mS and a noise figure of 0.7dB at room temperature. At cryogenic temperatures, it performs better by the nature of pHEMT. At its input gate terminal, the parasitic capacitance is around 0.24pF and the leakage current is 0.4μA at -3V V_{GS}. Assuming the leakage is resistive due to the imperfect isolation between its gate and ground, the equivalent resistance is 7.5MΩ. Parasitic capacitance due to on-board components and traces is around 2pF. Combining all above yields an RC time constant of 17μs. For the widest 16ns pulse, the maximum voltage drooping during CDAC switching then is 0.1%, which is acceptable for the measurement.

The gate of the pHEMT is ac-coupled to the output of the controller, and dc-biased through a 100MΩ resistor. This resistor is 13 times larger than the equivalent leakage resistance calculated above and barely affects the time constant. The 1pF ac coupling capacitor is chosen to preserve the signal amplitude while maximizing the bandwidth of the link thanks to its high self-resonance frequency of 18GHz. The 68Ω load resistor in parallel with the 200Ω output resistance of the pHEMT provide a matched 50Ω and a voltage gain of 2.5V/V.

Although we can reproduce the above operating point in testing, due to impedance mismatches at the gate of the pHEMT and its high gain at microwave frequencies, we observe oscillation at multiple frequencies above 1GHz as shown in Figure 4-6. The oscillation severely distorts the output waveform and prohibits an accurate pulse height measurement.
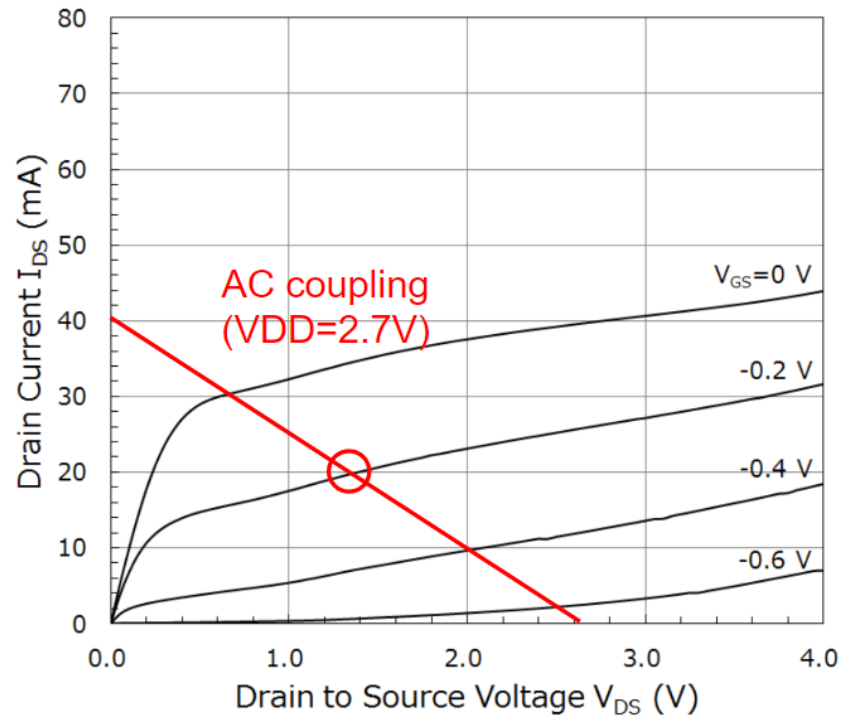
Figure 4-5: Desired operating point of pHEMT and the load line of the 68Ω load.
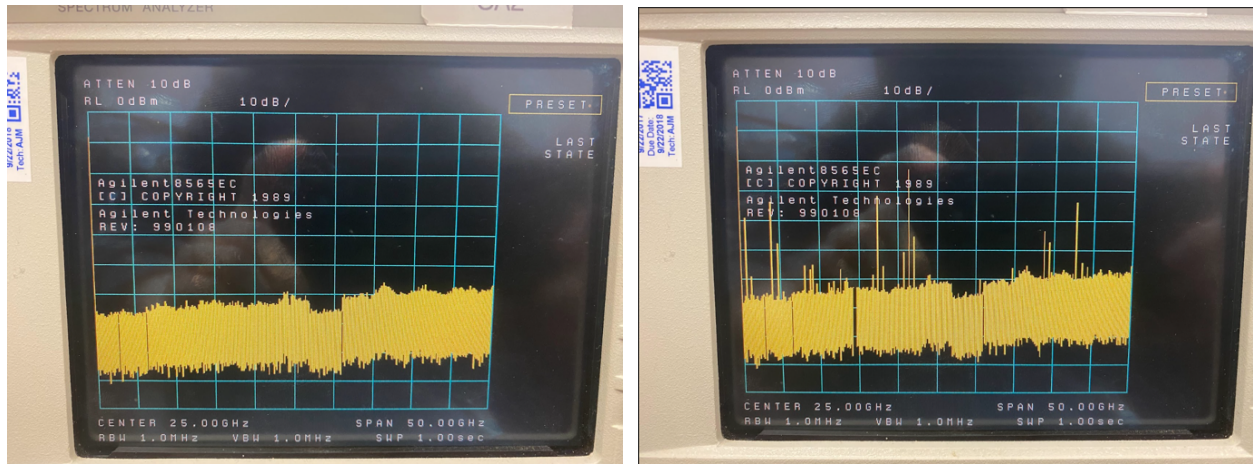
Figure 4-6: Spectrum at the output of the buffer before (left) and after (right) powering up.

- **Operational-amplifier-based feedback amplifier**

To resolve the oscillation, we switch to using a lower-bandwidth feedback amplifier based on an operational amplifier, or op amp (OPA851 from Texas Instrument). This op amp has a gain-bandwidth product of 900MHz, which eliminates the possibility of instability at above 1GHz. A series 10Ω is added at its input to kill the potential standing wave. Due to the feedback architecture, parasitic capacitance at its input and output are minimized to avoid possible sources of instability.

A standalone testing board for this op amp as shown in Figure 4-7 is designed to verify its stability and functionality at 300K and 77K. Figure 4-8 shows the schematic of the signal path including the modeling of our quantum controller's output stage, the bond wire, and the on-board output buffer. The three signal paths on board mimic the actual design on our final testing board with everything integrated.

Through testing all paths are verified to be stable across temperatures. When applying square waves to the inputs, clear output waveforms are observed on an oscilloscope (Figure 4-9). We also perform a sanity check on the linearity of all signal paths. As shown in Figure 4-10, the right channel mimicking the signal path for the right gate has slightly larger |INL|. This may be due to the weaker connection to its power and ground on the board limited by the available space. The measured max |INL| of the left and middle channel are about 0.5mV within a 900mV range. The absolute value of the nonlinearity is not critical as we de-embed it during pulse height measurements by creating a look-up table.
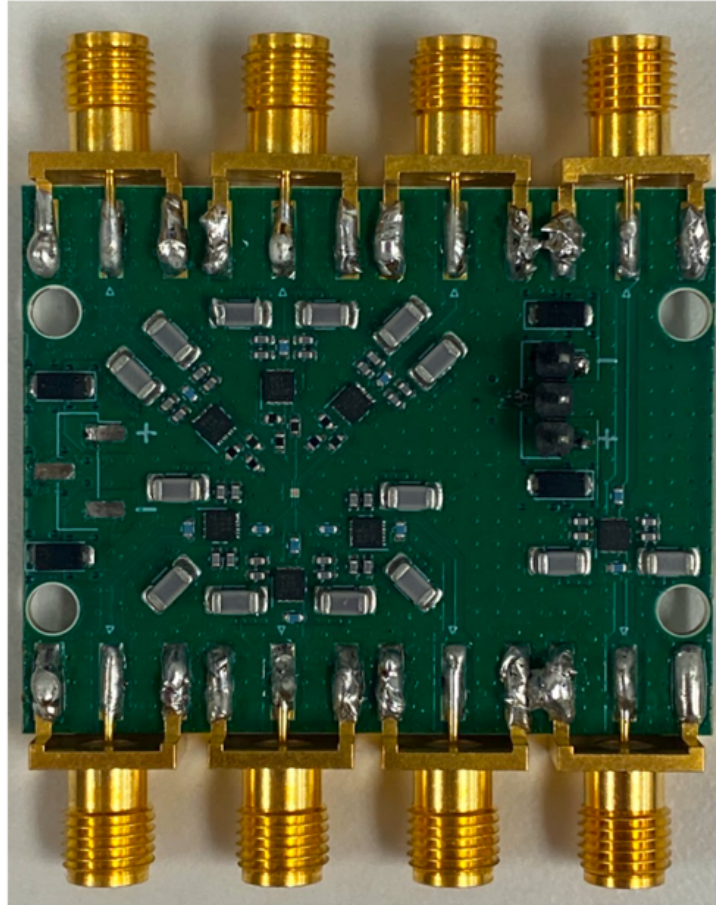
Figure 4-7: Op-amp-only testing board.
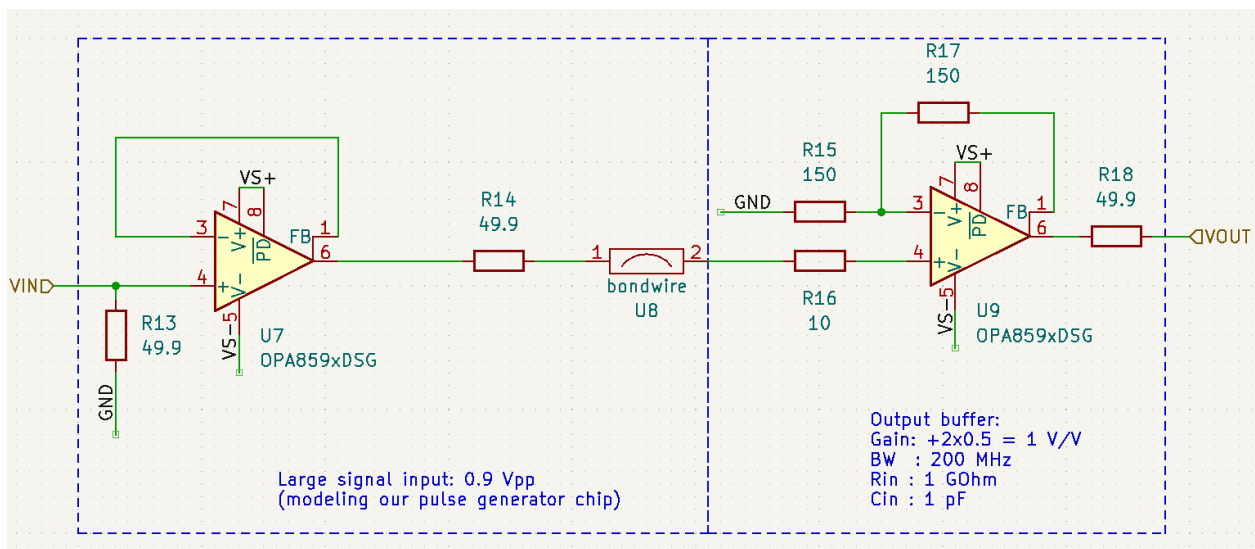
Figure 4-8: Schematic of a signal path on the op-amp board mimicking the final testing board.
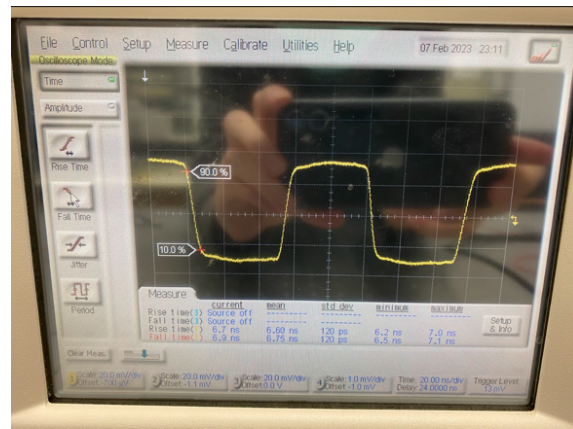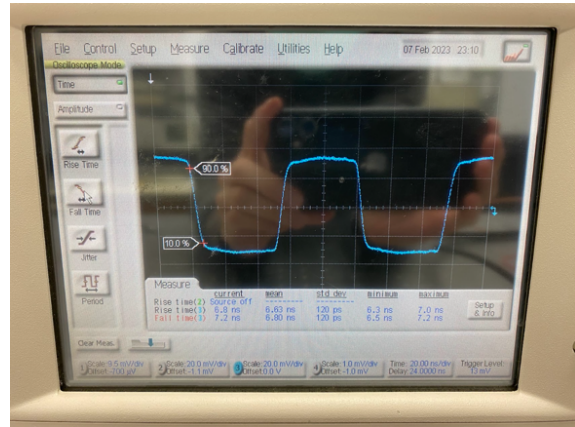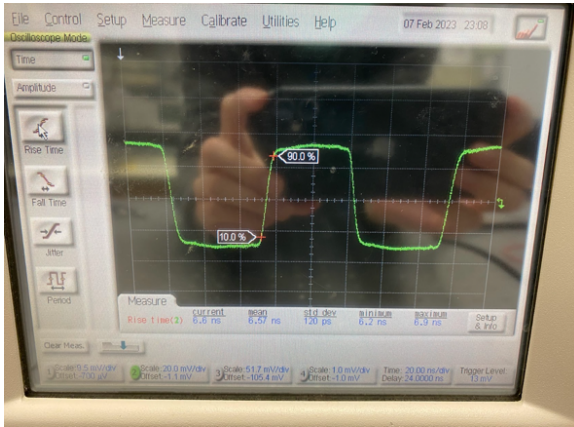
Figure 4-9: Clear output waveforms from all three paths measured at 77K.
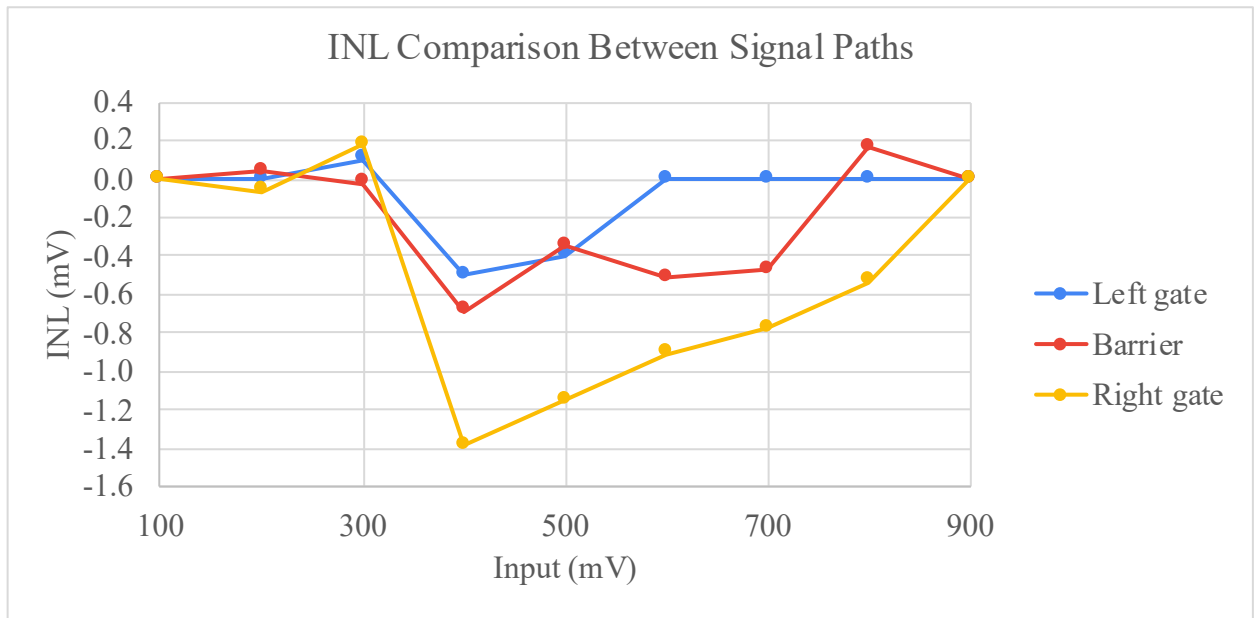
Figure 4-10: Sanity check on linearity of all three paths.

4.1.5. Final testing board

Shown in Figure 4-11 is the final testing board integrating our quantum controller with the op-amp-based output buffers. The three signal paths at the controller's output are carefully laid out to match the verified design on the op-amp-only testing board. All connectors are chosen for solid connections at 77K even after several thermal cycles.
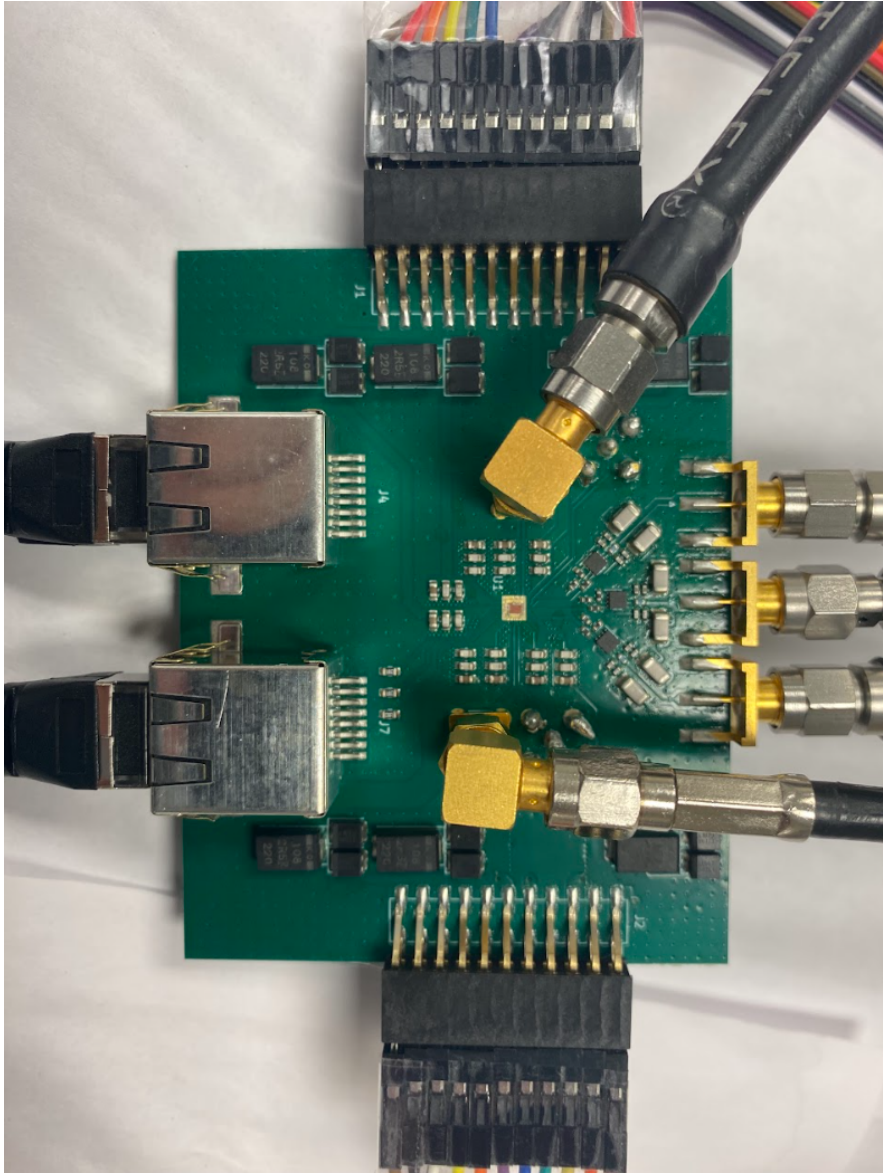
Figure 4-11: Final testing board with op-amp-based output buffer.

4.2. Temperature-insensitive performance for high-fidelity universal single-qubit gates

4.2.1. Successive approximating algorithm for pulse width calibration

The width control requires accurate calibration for the best performance to satisfy the stringent requirements for high-fidelity quantum gates. From the measured DNL shown in Figure 4-14, before calibration there are large spikes from code 127 to 128 and from code 255 to 256. The two major sources of these spikes are the gain mismatch between coarse and fine tuning and the nonideal duty cycle of the input clock as discussed in section 3.2.3.

The best operating point to linearize the transfer curve can be found by adjusting the biasing of the clock amplifier and the bulk voltage of the NFETs in the delay stages. Unfortunately, as evident in Figure 4-12, these two variables are interdependent when it comes to minimizing the two major sources of nonlinearity. Therefore, the calibration needs to optimize these variables simultaneously to converge to the global optimum. This interdependence enlarges the search space substantially, which makes the exhaustive search method unfeasible.

To efficiently search for the best operating condition, we transform this searching problem into an optimization problem aiming to minimize the objective below.

$$|DNL|^2|_{code\ 127\ to\ 128} + |DNL|^2|_{code\ 255\ to\ 256}$$

However, as this objective is a discrete function and is not totally convex, existing general-purpose minimization algorithms may not converge to the global optimum. To solve this, we develop a customized calibration algorithm.

Although the objective is non-convex, there is still a general trend that the value of the objective decreases when approaching the optimum. Based on this, we design a successive

approximating algorithm that gradually converges to the global optimum with one bit per variable per iteration and one-bit redundancy between consecutive iterations. Initially, the search space is restricted to the range with proper chip functionality. The algorithm then samples 3-by-3 points uniformly over the entire search space. In each iteration, it selects a variable to halve the search space while keeping one-bit redundancy to correct for errors due to measurement uncertainty and the nonconvexity. Successively, the algorithm converges to the optimal point for the best linearity in width tuning. From the measured DNL in Figure 4-14, after calibration the |DNL| from code 127 to 128 and from code 255 to 256 are drastically reduced and no longer dictate the max |DNL|.
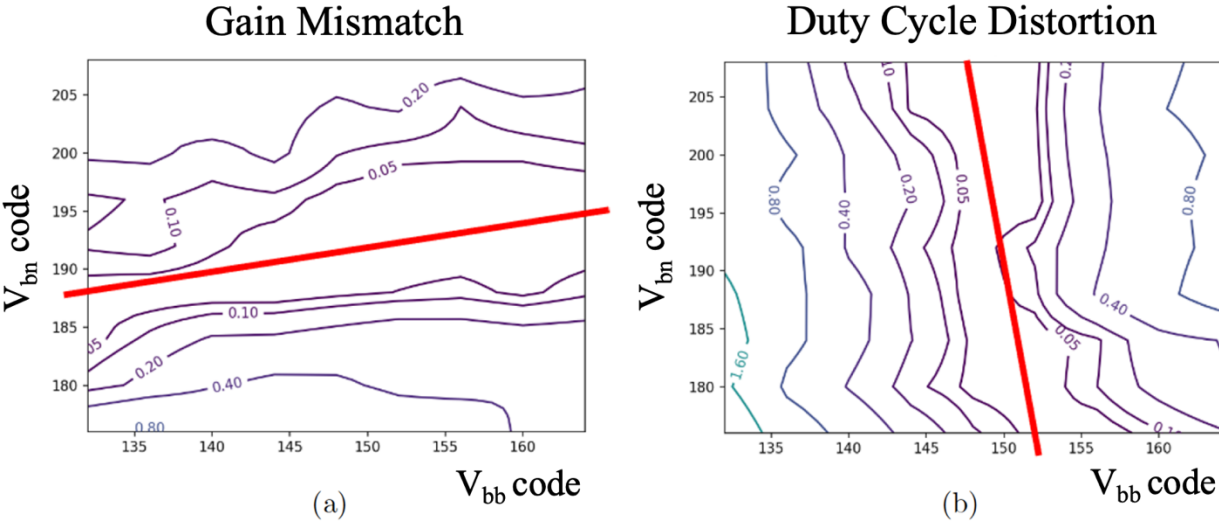
Figure 4-12: Dependency of (a) gain mismatch and (b) duty cycle distortion on the two calibration variables on X and Y axis. Red lines indicate the best combination.
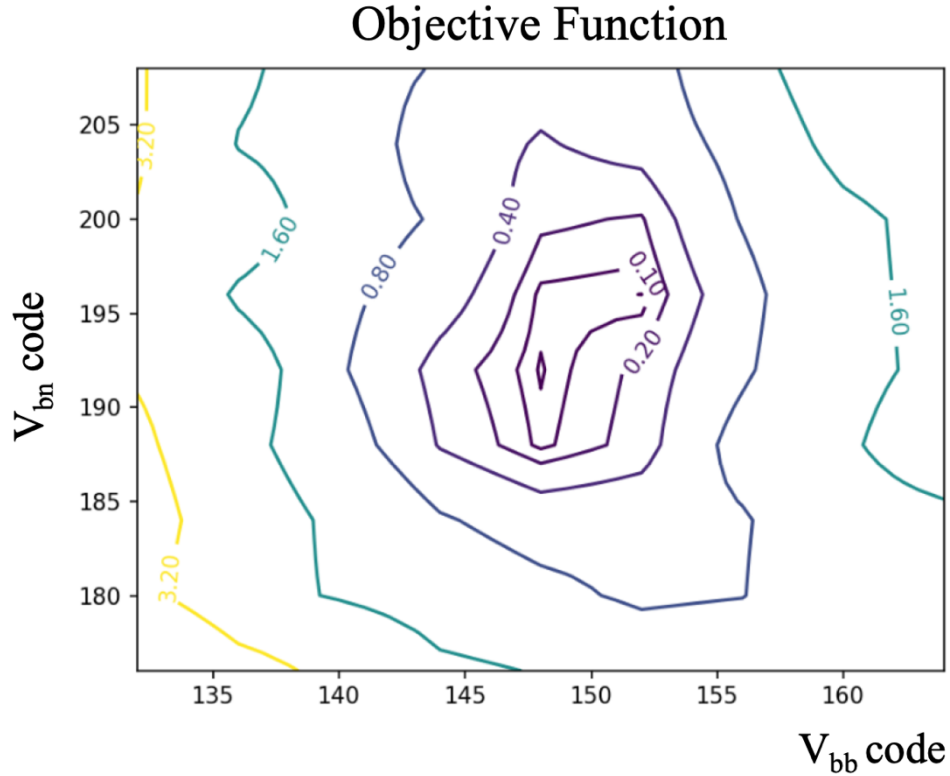
Figure 4-13: Plot of the objective function near the global optimum. Rugged contours indicate the non-convexity of this objective.

4.2.2. Temperature-insensitive maximum error in pulse width

- **Linearity and tuning resolution**

The measured nonlinearity of width control including DNL and INL are shown in Figure 4-14 and Figure 4-15. Firstly, the efficacy of calibration is evident by comparing the DNL profile at both 300K and 77K and the INL profile at 300K. We suspect that the slightly increased max |INL| at 77K is due to larger supply ripple caused by the higher "on" current. To assess the error in pulse width contributed by the nonlinearity, max |DNL| is the meaningful measure since it directly impacts the actual step size and determines how close we can get to the ideal width for 100% fidelity. The INL profile can used as a lookup table when plotting the Rabi oscillation and Ramsey fringes over time. The measured calibrated 1LSB max |DNL| and 3.9ps effective tuning

resolution are temperature-insensitive thanks to the calibration. To define end points of the transfer curve to calculate the nonlinearity, we pick code 12 and 268. Code 12 corresponds to a 47ps pulse width, which is the shortest pulse width reliably measured by the TDC. Code 268, on the other hand, is chosen such that the whole range covers 1ns. The reasoning for the 1ns range is described in section 4.1.1.
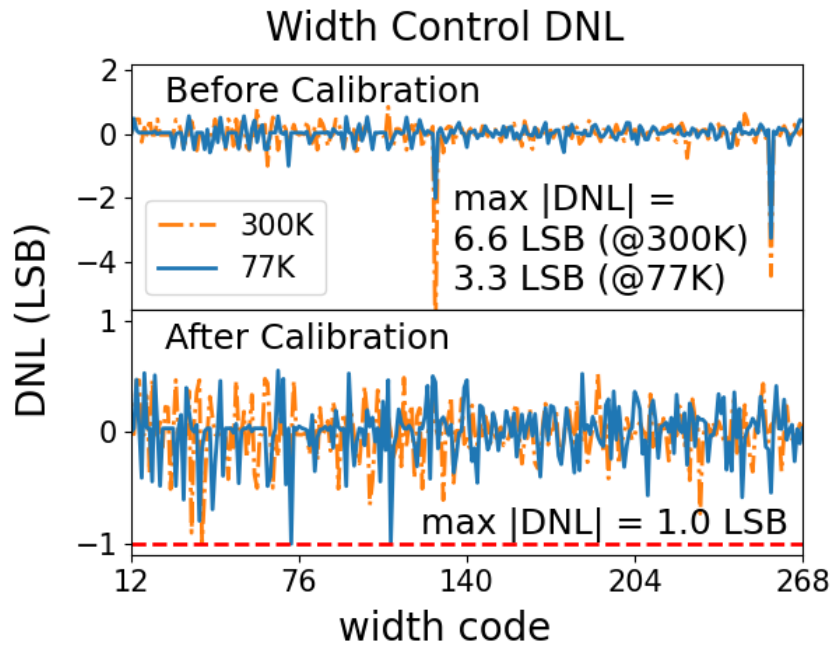
Figure 4-14: DNL profile of width control at 300K and 77K.
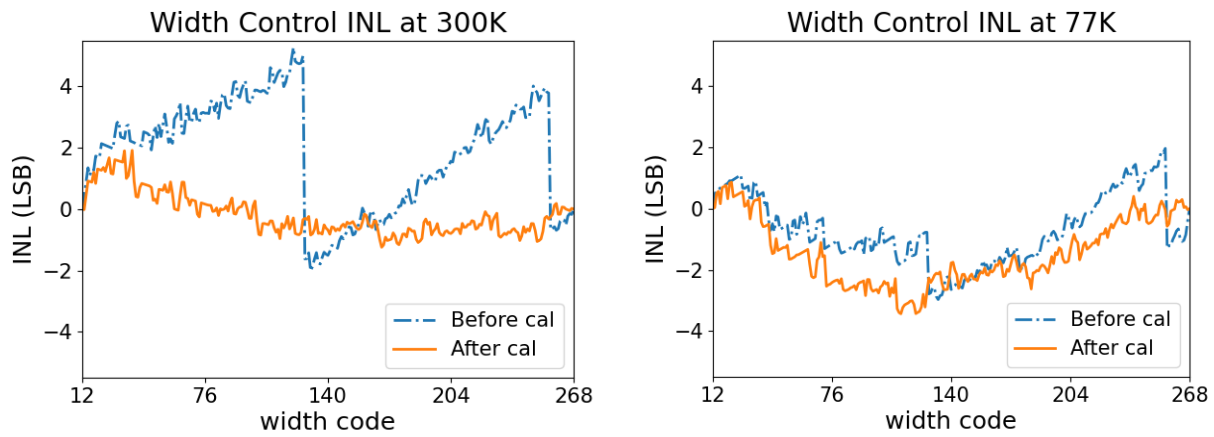
Figure 4-15: INL profile of width control at 300K and 77K.

- **Uncertainty in pulse width**

From simulation results we expect the uncertainty in pulse width to be below 1ps, which is too small for the 2ps/step cyclic TDC to directly measure. Raising the supply voltage increases the TDC resolution, but it quickly saturates to about 1.2ps at 1.4V supply as shown in Figure 4-16. Further raising the supply can result in reliability concerns for the devices. Fortunately, after calibration, adjusting the bulk voltage of the NFETs in the delay stages can fine-tune the pulse width by 0.19ps per LSB of the 8-bit VDAC at 300K and 0.14ps at 77K. This is observed through small changes in the distribution of the TDC's output. When the distribution of pulse width is centered at the middle of two consecutive decision thresholds of the TDC, a slight change in pulse width barely affects the distribution of the TDC's output. On the other hand, when the distribution of pulse width is centered at a decision threshold, a slight change in pulse width shifts the distribution of the TDC's significantly. Figure 4-16 shows above effects measured at 300K.

By approximating a linear dependency of the pulse width on the bulk voltage in a small region and observing changes in the distribution of TDC's output as the pulse width crosses its decision thresholds, the uncertainty in pulse width can be derived with fine resolution. Shown in Figure 4-17, the extracted 1σ uncertainty in pulse width in the worst-case scenario using the full scale of fine tuning, and thus the slowest edge rate, is 0.21ps at 300K and 0.12ps at 77K. The associated 1σ error from measurements is 0.03ps at both 300K and 77K. Thanks to the lower thermal noise, the uncertainty reduces by 43% from 300K to 77K which further improves the control fidelity.
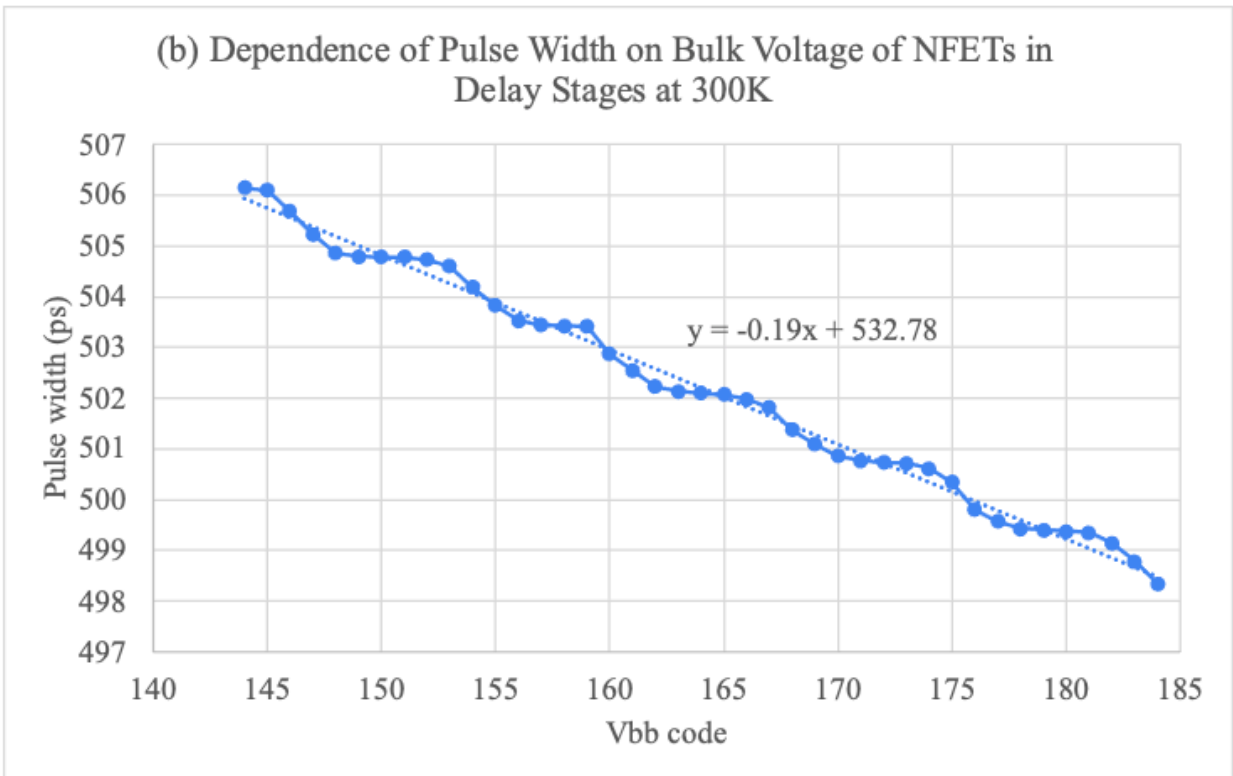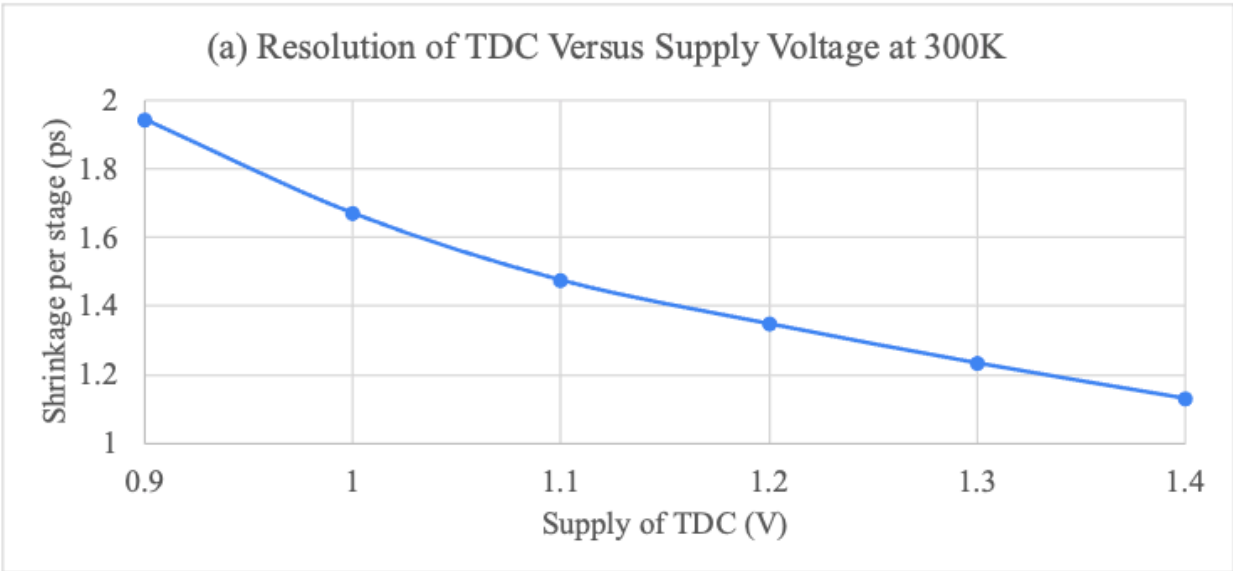
Figure 4-16: (a) Resolution of TDC versus its supply at 300K. (b) Pulse width versus bulk

voltage of NFETs in delay stages at 300K.
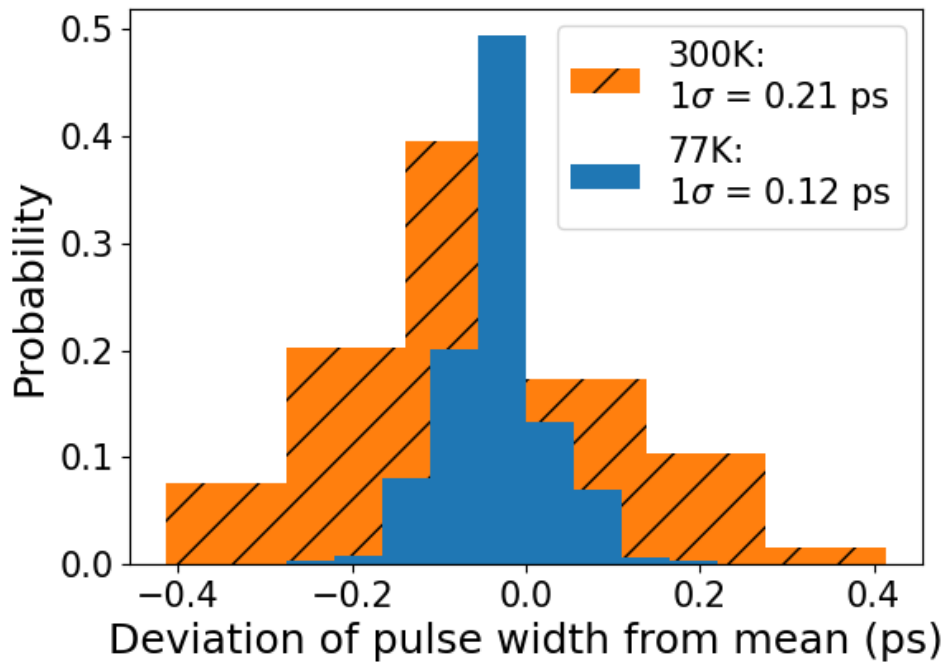
Figure 4-17: Worst-case uncertainty in pulse width at 300K and 77K after linearity calibration.

- **Maximum error in pulse width**

Combining measured temperature-insensitive performance of major contributors to the maximum error in pulse width, our quantum controller achieves similar high control fidelity across temperatures, i.e., 99.50% at 300K and 99.54% at 77K.

Table 4-2: Temperature-insensitive performance in width precision for high fidelity.

Measurement Temperature	300K	77K
Maximum Error in Width (t_{err})	4.5ps	4.3ps
Tuning Resolution (t_{LSB})	3.9ps	3.9ps
Maximum |DNL|	3.9ps	3.9ps
Width Jitter (σ_{width})	0.2ps	0.1ps
Control Fidelity	99.50%	99.54%

$$max(t_{err}) = \frac{t_{LSB} + max(|DNL|)}{2} + 3\sigma_{width}$$

4.2.3. Matching of MOM capacitors

To predict changes in linearity of data converters at cryogenic temperatures, measurement data for matching property of basic matching elements across temperatures are crucial. Unlike the matching of FETs, which has been extensively characterized in existing literature, to the author's best knowledge, there has not been any research systematically comparing measured matching of MOM capacitors at room temperature and cryogenic temperatures. To characterize this with our highly linear CDAC, the main obstacle is the poor SNR ($\ll 1$) caused by the mere 0.7% estimated mismatch between unit capacitors (σ_C/μ_C) and the uncertainty in measurement.

To improve the SNR, we first remove the deterministic nonlinearity of the on-board output buffer and the oscilloscope from measurement results. A highly linear semiconductor parameter analyzer, with a maximum error of ~20μV in its output, is used as the reference to create a look-up table for the nonlinearity of the buffer and the oscilloscope. Shown in Figure 4-18 is the combined INL profile. To ensure settling accuracy, we intentionally slow down the input clock, maximize the pulse width, and measure the pulse height 95μs after its transition edge. During this 95μs, since the output of the CDAC is a high-impedance node, the leakage causes significant voltage drooping, which is nonlinear with respect to the pulse height setting. This nonlinearity is minimized by equalizing the starting point of voltage drooping across all height settings.

Secondly, we alleviate the uncertainty in measurement due to random noise. During testing, a very slow variation, on the order of minutes, in the oscilloscope's voltage reading is observed if instruments are not fully warmed up. This low-frequency variation is practically eliminated by warming up the instruments for at least a day. For high-frequency noise, averaging the measured data by ~1 million times reduces it to an acceptable level within reasonable time. The final 1σ of measurement uncertainty is 30μV.

Lastly, we increase the signal strength by switching 64LSB at a time instead of 1LSB. Statistically, the standard deviation of DNL when we switch N unit capacitors at a time is $\sqrt{N}\,\sigma_C/\mu_C$, where σ_C/μ_C is the matching property of the unit capacitor. Therefore, by measuring the DNL with 64LSB switched at a time increase the signal power by 64 times. However, due to the fixed 9-bit full scale, this leads to only 7 DNL data from one board. To increase the amount of data for a more accurate Gaussian distribution fitting, we measure across multiple boards.

The indirectly measured matching property of our 2fF unit MOM capacitor at 300K is 0.72% and is 0.82% at 77K, which are similar to our estimate, i.e., 0.7%. These results and their error bar calculated from error propagation are shown in Figure 4-19. Unfortunately, due to board-related issues after dipping into the liquid nitrogen, the amount of valid data at 77K is less than that at 300K. This results in less accuracy when fitting a Gaussian distribution, and thus the calculated matching property. We still believe that the matching property should remain similar across temperatures since it solely depends on the ratio of geometry of the capacitor.

During testing we also try to characterize the max |DNL| of the height controller. However, even with effective noise reduction techniques mentioned above, the signal strength when switching 1LSB at a time is simply too weak compared to the measurement uncertainty. With a 0.7% σ_C/μ_C matching of the 2fF unit capacitor, through 10,000-point Monte Carlo simulations the expected max |DNL| is 0.2LSB as shown in Figure 3-19. As explained in section 3.2.4, the required precision in pulse height tuning can only be defined after the physical properties of the DQD are known. Nevertheless, comparing with the state-of-the-art (Esmailiyan, et al. 2020), our work has improved the resolution by 1 bit and significantly reduced the max |DNL| from the previously reported 10.4LSB.
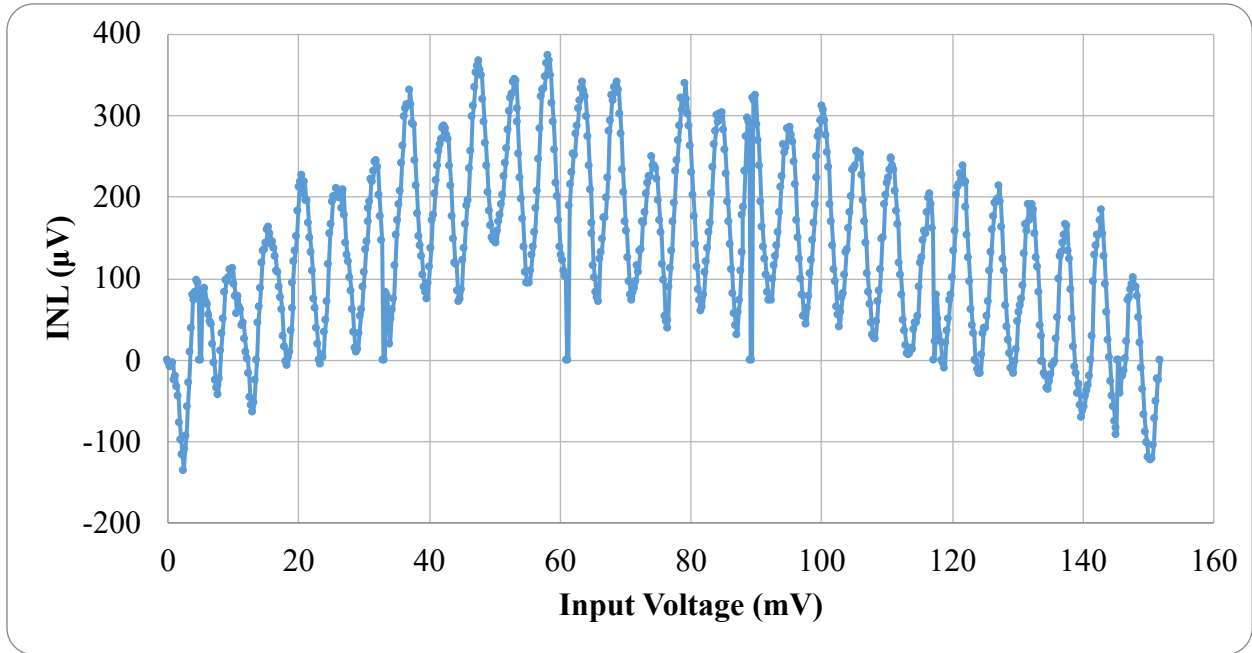
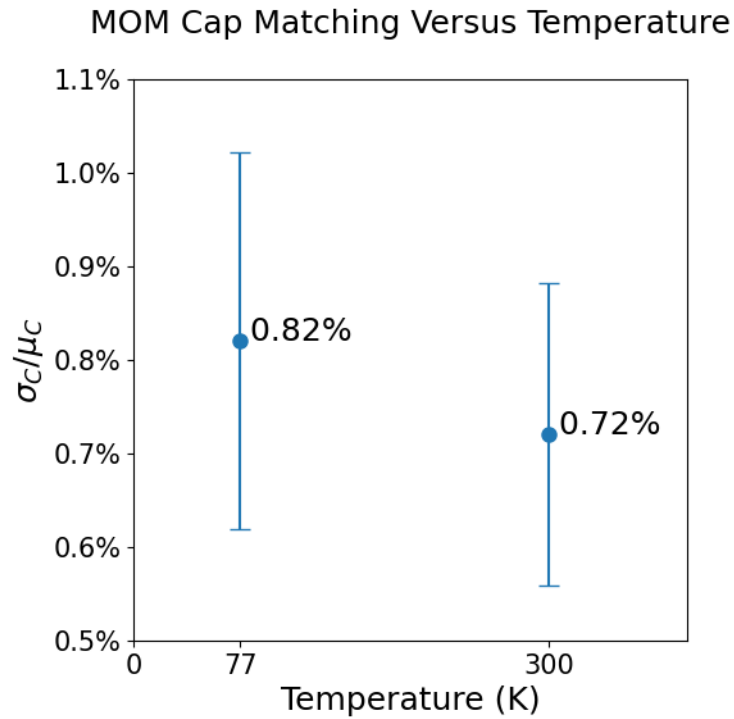Figure 4-18: INL of the output buffer plus the oscilloscope.

Figure 4-19: Matching property of the 2fF unit MOM capacitor versus temperature.

4.2.4. Waveforms for universal single-qubit gates and power consumption

After resolving the oscillation at the output with the lower-bandwidth op-amp-based buffer, we are able to demonstrate control waveforms for universal single-qubit gates and the Hahn echo experiment with the lock-in capability as shown in Figure 4-20. The actual lock-in period in typical quantum gate experiments will be much longer and is covered by our controller. For the Hahn echo experiment, the width of the second pulse on the barrier can be independently programmed. Our controller can deliver these control waveforms at speed, but due to the limited bandwidth of the output buffer, a lower-speed version is demonstrated. The at-speed capability is proven through a measured minimum pulse width of 47ps and post-layout simulations.

Due to limited cooling power of the refrigerator, power consumption of the controller is critical. As explained in section 2.3, we envision the height control to be integrated with the qubit through a low-parasitic packaging in the millikelvin chamber, and the rest of the circuitry to be housed in a higher-temperature chamber. Figure 4-20 details the breakdown of power consumption when the controller is programmed to perform the Rabi oscillation experiment at a 100MHz repetition rate. The 0.25mW power consumption of the height controller is dominated by the leakage through the ultra-low-V_{th} output inverter driver in the CDAC, which is 0.21mW from measurements in idle mode. This can be greatly reduced by replacing devices with the low-V_{th} version. Through simulations we verify that the low-V_{th} drivers can still deliver adequately fast transition edges to satisfy the maximum transition time required for high fidelity. Also, as described in section 3.2.2, this prototype includes significant extra programmability and circuitry for complete qubit characterization including its decoherence and relaxation behavior. Once the qubit is fully characterized, extra parts irrelevant to high-fidelity quantum gates can be removed without sacrificing precision to further reduce the power consumption.
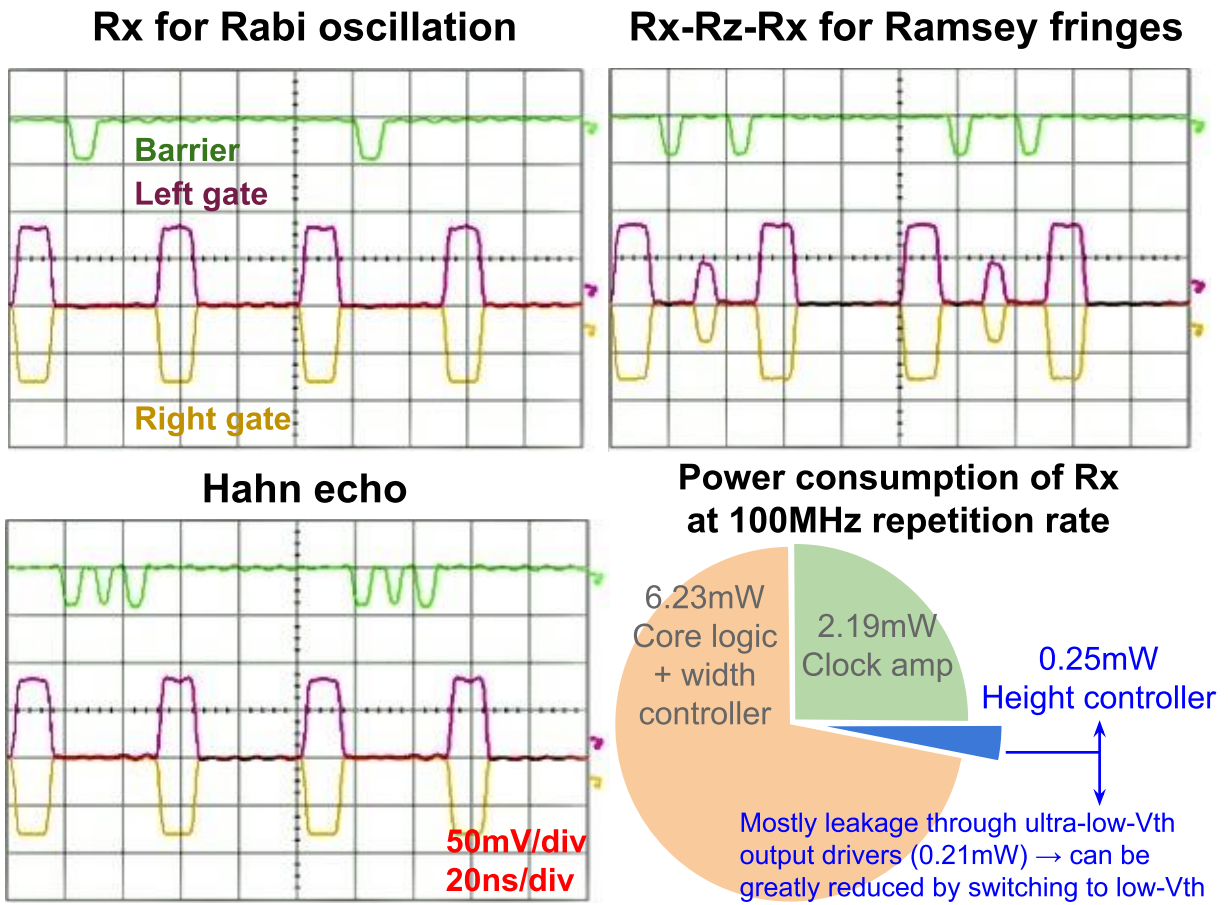
Figure 4-20: Measured waveforms for universal single-qubit gates with lock-in capability and at-speed power breakdown at 77K.
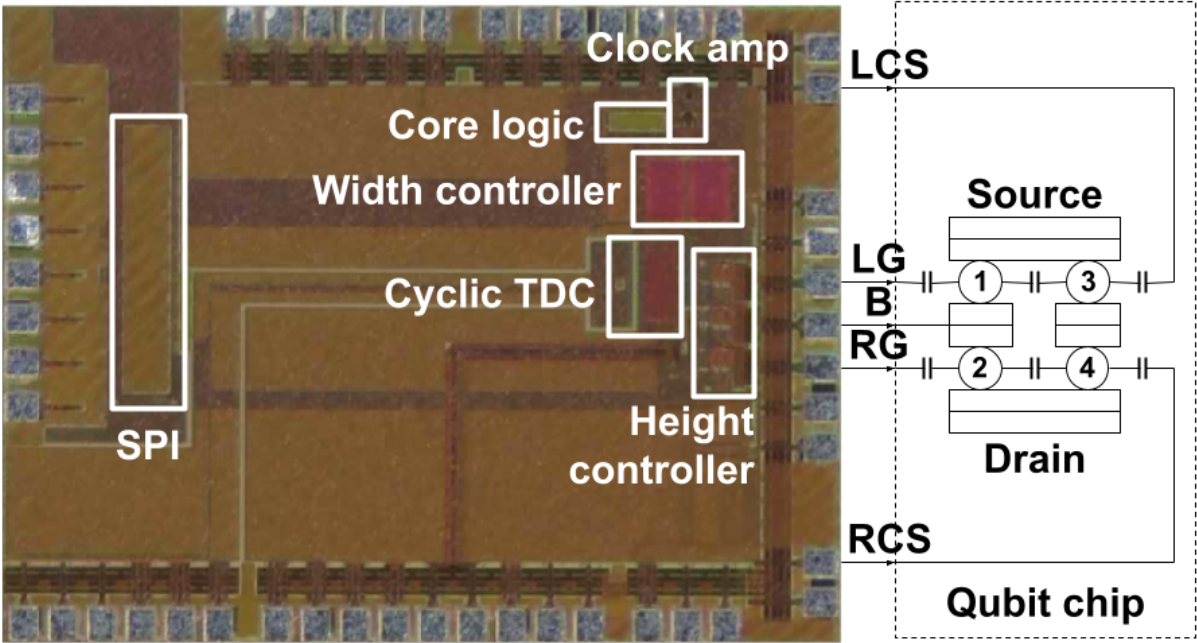
Chapter 5 Conclusion and Future Work

Scalability, universality, and fidelity are the three cornerstones for quantum computing. Among the scalable all-electrically-controlled, CMOS-compatible qubits, the dc-pulse-controlled position-based DQD charge qubit shows great potential for integration of millions of qubits thanks to its mere 200nm pitch. Due to the wide-band nature of control waveforms, the quantum gate fidelity in prior physics experiments was limited by the slow edge rate, which was a direct result of the necessary low-pass filtering to reduce high-frequency reflections in the meter-long cabling. On the controller side, state-of-art dc-pulse-based quantum controllers lacked the required tuning precision in the pulse width and height to enable high-fidelity quantum gates.

In this dissertation we address these issues by proposing an architecture for the scalable dc-pulse-based quantum processor, deriving requirements on control waveforms for high-fidelity quantum gates, proposing temperature-insensitive design methodologies to overcome the lack of device models for simulations, and successfully demonstrating an instrument-like quantum controller prototype in 28nm bulk CMOS down to 77K for complete qubit characterization and high-fidelity universal single-qubit gates. A micrograph of the fabricated chip is shown in Figure 5-1. To meet the stringent 6.4ps maximum error in pulse width for high fidelity, we develop a closed-loop automated multi-variable calibration utilizing an on-chip 2ps/step cyclic TDC.

Comparing achieved precision with state-of-art dc-pulse-based controllers as summarized in Table 5-1, our design has reduced the shortest duration of a pulse by over 3X, boosted the linear width control resolution by over 40X, reduced the width uncertainty by three orders of magnitude, realized an on-chip sub-period width calibration, and largely improved the precision of height tuning to facilitate high-fidelity universal single-qubit gates.

To further improve our controller for multi-qubit, high-fidelity universal quantum gate operations, future research includes full qubit characterization through an integrated quantum processor, simplification of the controller design based on measured qubit properties, expansion of control waveforms for two-qubit quantum gates, and development of time-multiplexed addressing circuits.

Legend for qubit chip

DQD (1: left dot; 2: right dot)	Charge sensors
• LG: gate of left dot	(3: sensor for left dot; 4: sensor for right dot)
• B: barrier	• LCS: gate of charge sensor for left dot
• RG: gate of right dot	• RCS: gate of charge sensor for right dot

Figure 5-1: Chip micrograph and intended integration with the qubit chip.

Table 5-1: Comparison with state-of-art inductor-less dc-pulse-based quantum controllers.

		This Work	Pauka, Nat. Electron. '21	Esmailiyan, ESSCIRC '20	Ekanayake, NANO '08
CMOS Technology		28nm bulk	28nm FDSOI	22nm FDSOI	500nm SOS-CMOS
Measurement Temperature		77K	100mK	3.4K	4.2K
On-Chip Supported Waveforms		Arbitrary Rx and Rz, Hahn echo	Arbitrary	Relied on external pattern generator	
Width	Tuning Mechanism	5b dig. counter + 7b delay line	Free-running ring oscillator, digital counter	External 6GHz pattern gen., no on-chip tuning	Free-running ring oscillator, digital counter
	Shortest Duration	47ps	Not explicitly mentioned, highest frequency measured was only 10MHz	166ps	8.2ns
	Max Error	4.3ps		N/A	N/A
	Step Size†	3.9ps		166ps	2.7ns
	Max |DNL|†	3.9ps		N/A	N/A
	Jitter (1σ)†	0.1ps		N/A	Hundreds of ps
	Calibration Capability	Yes, through cyclic TDC		N/A	N/A
Height	Number of Bits	9	1	8	1
	Max |DNL|	0.2LSB††	N/A	10.4LSB	N/A

† Major contributors to max error

†† Simulated through 10,000-point Monte Carlo with 0.7% mismatches in 2fF unit capacitor

Chapter 6 Bibliography

Arute, Frank, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, Rami Barends, Rupak Biswas, et al. 2019. "Quantum supremacy using a programmable superconducting processor." *Nature* 574 (7779): 505-510.

Beckers, Arnout, Farzan Jazaeri, Andrea Ruffino, Claudio Bruschini, Andrea Baschirotto, and Christian Enz. 2017. "Cryogenic characterization of 28 nm bulk CMOS technology for quantum computing." *2017 47th European Solid-State Device Research Conference (ESSDERC).* IEEE. pp. 62-65.

Bonen, Shai. 2020. "Monolithically Integrated FDSOI CMOS Electron- and Hole-quantum Dot Qubits and Readout Electronics for Large-scale Quantum Computing Processors." *University of Toronto (Canada).*

Cao, Gang, Hai-Ou Li, Tao Tu, Li Wang, Cheng Zhou, Ming Xiao, Guang-Can Guo, Hong-Wen Jiang, and Guo-Ping Guo. 2013. "Ultrafast universal quantum control of a quantum-dot charge qubit using Landau–Zener–Stückelberg interference." *Nature Communications* 4 (1): 1401.

Chakraborty, Sudipto , David J. Frank, Kevin Tien, Pat Rosno, Mark Yeck, Joseph A. Glick, Raphael Robertazzi, et al. 2022. "A Cryo-CMOS Low-Power Semi-Autonomous Transmon Qubit State Controller in 14-nm FinFET Technology." *IEEE Journal of Solid-State Circuits* 57 (11): 3258-3273.

Chang, Jimmin, A.A. Abidi, and C.R. Viswanathan. 1994. "Flicker noise in CMOS transistors from subthreshold to strong inversion at various temperatures." *IEEE Transactions on Electron Devices* 41 (11): 1965-1971.

Chen, Chun-Chi, Shih-Hao Lin, and Chorng-Sii Hwang. 2014. "An area-efficient CMOS time-to-digital converter based on a pulse-shrinking scheme." *IEEE Transactions on Circuits and Systems II: Express Briefs* 61 (3): 163-167.

Ekanayake, S. Ramesh, Torsten Lehmann, Andrew S. Dzurak, and Robert G. Clark. 2008. "Qubit control-pulse generator circuits for operation at cryogenic temperatures." *2008 8th IEEE Conference on Nanotechnology.* IEEE. pp. 472-475.

Esmailiyan, Ali, Hongying Wang, Mike Asker, Eugene Koskin, Dirk Leipold, Imran Bashir, Kai Xu, Anna Koziol, Elena Blokhina, and R. Bogdan Staszewski. 2020. "A fully integrated DAC for CMOS position-based charge qubits with single-electron detector loopback testing." *IEEE Solid-State Circuits Letters* 3: 354-357.

Fowler, Austin G., Matteo Mariantoni, John M. Martinis, and Andrew N. Cleland. 2012. "Surface codes: Towards practical large-scale quantum computation." *Physical Review A* 86 (3): 032324.

Gorman, J., D. G. Hasko, and D. A. Williams. 2005. "Charge-qubit operation of an isolated double quantum dot." *Physical review letters* 95 (9): 090502.

Homulle, Harald. 2019. *Cryogenic electronics for the read-out of quantum processors.* https://doi.org/10.4233/uuid:e833f394-c8b1-46e2-86b8-da0c71559538.

IBM Corporation. 2018. *Coming soon to your business – Quantum computing.* https://www.ibm.com/thought-leadership/institute-business-value/report/quantumstrategy.

Ismail-zade, Mamed R., Konstantin O. Petrosyants, Lev M. Sambursky, Xu Zhang, Bo Li, Jiajun Luo, and Zhengsheng Han. 2020. "SPICE Modeling of Small-Size Bulk, SOI and SOS MOSFETs at Deep-Cryogenic Temperatures." *2020 26th International Workshop on Thermal Investigations of ICs and Systems (THERMINIC).* IEEE. pp. 97-103.

Jozsa, Richard. 1994. "Fidelity for mixed quantum states." *Journal of modern optics* 41 (12): 2315-2323.

Kandala, Abhinav, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M. Chow, and Jay M. Gambetta . 2017. "Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets." *Nature* 549 (7671): 242-246.

Loctite. 1996. "Technical Data Sheet Product 401." *https://docs.rs-online.com/7331/0900766b801df7d6.pdf.*

Masterbond. n.d. *EP21TCHT-1 Product Information.* https://www.masterbond.com/tds/ep21tcht-1.

Matthias. 2022. *MOM capacitor extraction.* https://www.klayout.de/forum/discussion/2045/mom-capacitor-extraction.

MG Chemicals . 2021. "8331D." https://mgchemicals.com/downloads/tds/tds-8331D-2parts.pdf.

Morelock, C. R., M. R. Suchomel, and A. P. Wilkinson. 2013. "A cautionary tale on the use of GE-7031 varnish: low-temperature thermal expansion studies of ScF3." *Journal of Applied Crystallography* 46 (3): 823-825.

Patterson, Richard, Ahmad Hammoud, and Scott Gerber. 2001. *Performance of various types of resistors at low temperatures.* GESS Rep, Cleveland, OH, USA: NASA Glenn Res. Center, NAS3-00142.

Pauka, S. J., K. Das, R. Kalra, A. Moini, Y. Yang, M. Trainer, A. Bousquet, et al. 2021. "A cryogenic CMOS chip for generating control signals for multiple qubits." *Nature Electronics* 4 (1): 64-70.

Rabouw, Freddy T. , and Celso de Mello Donega. 2017. "Excited-state dynamics in colloidal semiconductor nanocrystals." In *Photoactive Semiconductor Nanocrystal Quantum Dots: Fundamentals and Applications*, 1-30.

Saberi, Mehdi, Reza Lotfi, Khalil Mafinezhad, and Wouter A. Serdijn. 2011. "Analysis of power consumption and linearity in capacitive digital-to-analog converters used in successive approximation ADCs." *IEEE Transactions on Circuits and Systems I: Regular Papers* 58 (8): 1736-1748.

't Hart, P. A., J. P. G. Van Dijk, M. Babaie, E. Charbon, A. Vladimircscu, and F. Sebastiano. 2018. "Characterization and model validation of mismatch in nanometer CMOS at cryogenic temperatures." *2018 48th European Solid-State Device Research Conference (ESSDERC)*. IEEE. pp. 246-249.

Teyssandier, F., and D. Prêle. 2010. "Commercially available capacitors at cryogenic temperatures." *Ninth International Workshop on Low Temperature Electronics-WOLTE9*.

Tsai, Jen-Huan, Hui-Huan Wang, Yang-Chi Yen, Chang-Ming Lai, Yen-Ju Chen, Po-Chuin Huang, Ping-Hsuan Hsieh, Hsin Chen, and Chao-Cheng Lee. 2015. "A 0.003 mm ^{2} 10 b 240 MS/s 0.7 mW SAR ADC in 28 nm CMOS With Digital Error Correction and Correlated-Reversed Switching." *IEEE Journal of Solid-State Circuits* 50 (6): 1382-1398.

Tseng, Wei-Hsin, Wei-Liang Lee, Chang-Yang Huang, and Pao-Cheng Chiu. 2016. "A 12-bit 104 MS/s SAR ADC in 28 nm CMOS for digitally-assisted wireless transmitters." *IEEE Journal of Solid-State Circuits* 51 (10): 2222-2231.

Van Dijk, Jeroen , Pascal't Hart, Gerd Kiene, Ramon Overwater, Pinakin Padalia, Job Van Staveren, Masoud Babaie, Andrei Vladimirescu, Edoardo Charbon, and Fabio Sebastiano.

2020. "Cryo-CMOS for analog/mixed-signal circuits and systems." *2020 IEEE Custom Integrated Circuits Conference (CICC).* IEEE. pp. 1-8.

Yang, Tsung-Yeh , Aleksey Andreev, Yu Yamaoka, Thierry Ferrus, Shunri Oda, Tetsuo Kodera, and David A. Williams. 2016. "Quantum information processing in a silicon-based system." *IEEE International Electron Devices Meeting (IEDM)* pp. 34-2.

Zhang, Xuan Silvia. 2016. "ESE 461: Design Automation for Integrated Circuit Systems." https://classes.engineering.wustl.edu/ese461/Lecture/week7b.pdf.