Integrated multiplexed microwave readout of silicon quantum dots in a cryogenic CMOS chip

A. Ruffino, 1,∗ T.-Y. Yang, 2,† J. Michniewicz, 3 Y. Peng, 1 E. Charbon, 1,‡ and M. F. Gonzalez-Zalba 2, 4,⊥

1 Advanced Quantum Architecture Laboratory, École Polytechnique Fédérale de Lausanne, Rue de la Maladière 71b, 2002, Neuchâtel, Switzerland
2 Hitachi Cambridge Laboratory, Hitachi Europe Ltd., J. J. Thomson Avenue, CB3 0HE, Cambridge, United Kingdom
3 Cavendish Laboratory, University of Cambridge, J. J. Thomson Avenue, CB3 0HE, Cambridge, United Kingdom
4 Quantum Motion Technologies, Nexus, Discovery Way, LS2 3AA, Leeds, United Kingdom

Solid-state quantum computers require classical electronics to control and readout individual qubits and to enable fast classical data processing [1–3]. Integrating both subsystems at deep cryogenic temperatures [4], where solid-state quantum processors operate best, may solve some major scaling challenges, such as system size and input/output (I/O) data management [5]. Spin qubits in silicon quantum dots (QDs) could be monolithically integrated with complementary metal-oxide-semiconductor (CMOS) electronics using very-large-scale integration (VLSI) and thus leveraging over wide manufacturing experience in the semiconductor industry [6]. However, experimental demonstrations of integration using industrial CMOS at mK temperatures are still in their infancy.

Here we present a cryogenic integrated circuit (IC) fabricated using industrial CMOS technology that hosts three key ingredients of a silicon-based quantum processor: QD arrays (arranged here in a non-interacting 3×3 configuration), digital electronics to minimize control lines using row-column addressing and analog LC resonators for multiplexed readout, all operating at 50 mK. With the microwave resonators (6–8 GHz range), we show dispersive readout of the charge state of the QDs and perform combined time- and frequency-domain multiplexing, enabling scalable readout while reducing the overall chip footprint.

This modular architecture probes the limits towards the realization of a large-scale silicon quantum computer integrating quantum and classical electronics using industrial CMOS technology.

Quantum computing is poised to be an innovation driver of the decade, given its theoretically-demonstrated capability to solve certain computational problems more efficiently than classical computers [7]. However, constructing the required quantum hardware is one of the greatest technological challenges for the scientific community. Single electron spins isolated in silicon QDs are one of the most promising solid-state systems to achieve that goal: recent demonstrations of long coherence times [8], high-fidelity spin readout [9], and one- and two-qubit gates [10–12], fulfill the basic requirements to build a quantum computer approaching fault-tolerant thresholds [13]. Until now, silicon QDs have been typically fabricated using custom processes [8, 14], but recent results have revealed they can be manufactured at scale using industry-compatible [6] or even industry-standard processes [15, 16]. This allows to leverage the integration capabilities of the semiconductor industry to scale up.

Researchers have already produced blueprints of large-scale quantum computers in silicon [17–19]. The proposals share common concepts: MOS-based QD arrays to host the qubits, digital electronics for I/O data management and analog classical electronics for control and readout. Full system integration will bring the benefits of reduced footprint, ease of signal synchronization, reduced latency and minimized inter-chip wiring. However, the ultimate level of possible integration is still uncertain, given the reduced cooling power availability at mK temperatures for typical classical electronics. Exploring the limits of integration is hence paramount to realize any fully-fledged solid-state quantum processor [20, 21].

For silicon, one approach has been to operate qubits at elevated temperatures (≈1.2 K) [22, 23] to enable larger cooling power budgets for the classical electronics [24], but so far this has come at the cost of reduced fidelity and/or coherence times. Alternatively, ICs that operate at mK temperatures have been produced for control [25], readout [26], signal multiplexing [27–29], or single devices have been used for time-multiplexed readout [30], but these circuits have not been fully integrated with the quantum devices.

Here we present an IC fabricated using industrial 40-nm CMOS technology that enables scalable multiplexed microwave readout of a non-interacting array of silicon QDs, all operating at mK temperatures in the same monolithic chip. The QDs are hosted in the channel of minimum size transistors and placed in a 3×3 array. Individual QDs can be randomly addressed via digital

* Equally-contributing authors
† Equally-credited authors
FIG. 1: Fully-integrated cryogenic CMOS quantum-classical matrix. a, Micrograph of the IC, with corresponding control signals. b, Schematic of the cryogenic CMOS readout matrix consisting of three integrated LC resonators, i.e., Resonator$_1$ (blue), Resonator$_2$ (red), and Resonator$_3$ (green), access transistors $T_{ij}$ and QD transistors $Q_{ij}$, word-lines $V_{WL,j}$, data-lines $V_{DL,i}$, source voltages $V_{Sij}$, bias-tees $R_{Tj}$, and storage capacitors $C_C$. Inset: A $3\times3$ grid representing the $3\times3$ QD transistor array. c, Coulomb diamonds at $V_{WL,3}=1.5$ V with $V_{DL,3}$ ranging from 0.39 V to 0.56 V, as marked by the red line section in d. d, Stability diagram of $V_{WL,3}$ vs $V_{DL,3}$ with $V_{S33}=0.01$ V. e, Frequency spectrum of the multi-resonator for Set-up A at 300 K and Set-up B at 50 mK, showing the 3 resonance frequencies. Set-up A: direct measurement with vector network analyzer (VNA) → cable → chip → cable → VNA. Set-up B: the chip is placed in the dilution fridge, and connected via VNA → attenuator → cable → circulator → chip → circulator → low noise amplifiers → cable → VNA (see Methods). The resonance frequencies are found to be $(f_1, f_2, f_3)=(6.810, 7.374, 7.941)$ GHz for (Resonator$_1$, Resonator$_2$, Resonator$_3$) at 300 K (see Supplementary Information S2). The black and blue arrows indicate the resonance frequencies at 300 K and 50 mK, respectively. Inset: zoom-in for Resonator$_1$ at different $V_{WL,1}$ and $V_{DL,1}$.

transistors using a row-column architecture to minimize the number of inputs. The readout is performed using gate-based microwave reflectometry, a technique readily compatible with industrial CMOS technology [31]. For the first time, we combine time- and frequency-domain multiplexing in one demonstration to minimize circuit footprint, otherwise compromised by using one readout resonator per qubit, while preserving a degree of parallel readout, ideal for quantum error correction.

A FULLY-INTEGRATED DRAM-LIKE READOUT MATRIX ARCHITECTURE

The proposed quantum-classical readout interface has been designed and fabricated in an industrial 40-nm bulk CMOS technology. Fig. 1a and Fig. 1b present the chip micrograph and its schematic, respectively. The modular architecture consists of a $3\times3$ array of 9 identical cells, arranged in a row-column random-access configuration similar to a dynamic random-access memory (DRAM). Each cell contains a silicon QD device, i.e., $Q_{ij}$ for $i,j=1,2,3$, implemented as a MOS transistor. Furthermore, the gate of each quantum device is connected to the source of an access MOS transistor with a channel length $L=40$ nm and a channel width $W=15$ µm, allowing conditional access, and to a storage capacitor (200 fF), to store the voltage at the gate of the quantum device (see Supplementary Information S1). The access transistors in column $j$ are controlled by the same word-line signal $V_{WL,j}$. The cells in row $i$ are connected to both a shared data-line signal $V_{DL,i}$, providing a bias voltage to control the QDs through an on-chip bias-tee, and a LC resonator.
Resonator1, to perform microwave reflectometry readout [32, 33]. When applying data-line voltage $V_{DLi}$ and word-line voltage $V_{WLj}$, a quantum device $Q_{ij}$ is accessed, and its charge state can be read out via Resonator1. We also included independent source voltages $V_{Sij}$ to perform a full electrical characterization of the individual QD devices in the quantum regime.

This architecture reduces the resources needed for control and readout of $N$ quantum devices from $N$ to just $2\sqrt{N}$ lines and $\sqrt{N}$ resonators, respectively, substantially improving upon the circuit complexity of current paradigms.

QUANTUM-CLASSICAL CIRCUITS IN STANDARD CMOS

The QD devices are designed as minimum size $(W/L=120 \text{ nm}/40 \text{ nm})$ nMOS transistors, to minimize channel volume. The devices have a low threshold voltage $Vt$ and they are implemented as single gate planar NPN transistors in a deep n-well, to isolate them from substrate noise. When we cool these devices down to 50 mK and apply a positive $V_{DLj}$ voltage ($> V_{th}$) with high $V_{WLj}$, single electrons are trapped in the small area under the gate at the oxide-silicon interface, i.e. few-electron QDs [34] are formed, which we observe at low $V_{Sij}$, as shown in Fig. 1c.

Here we measured the source-drain current $I_s$ as a function of the $V_{DLj}$ and $V_{Sij}$ voltages for the QD transistor $Q_{33}$ and we observe Coulomb diamonds. From this plot, a charging energy $E_c\approx e^2/\epsilon S\approx 21.42 \text{ meV}$ ($e$ is the electron charge) can be extracted, which enables electron trapping even at 4.2 K as electron QDs [34] are formed, which we observe at low $V_{Sij}$, as shown in Fig. 1c.

In Fig. 1d we demonstrate the functionality of the access transistor-QD cell and highlight the allowed and forbidden regions for readout. When $V_{WL3} - V_{DL3} < 0.277 \text{ V}$, the access transistor channel resistance $R_{acc}$ becomes comparable to the QD transistor gate leakage resistance $R_G$, therefore the effective gate voltage for the QD transistor $V_{DL, eff} = (V_{DL} - R_G)/(R_{acc} + R_G)$ is reduced. So, the QD becomes effectively decoupled from the data-line (see Supplementary Information S1). However, when $V_{WL3} - V_{DL3}$ is far higher than this threshold, i.e., above the dashed black line in Fig. 1d, then $V_{DL, eff} \approx V_{DL}$, and Coulomb blockade oscillations are observed as a function of $V_{DL3}$.

This demonstrates the realization of a quantum-classical integrated circuit in industrial bulk CMOS technology at 50 mK.

FIG. 2: Integrated gate-based reflectometry. a, DC transport measurement for device $Q_{13}$: $I_{S13}$ vs $V_{DL1}$. b, Normalized microwave reflected voltage as a function of $V_{WL3}$ and $V_{DL1}$ at $V_{S13}=2 \text{ mV}$. This diagram is the union of five individual $V_{WL3}$ vs $V_{DL1}$ measurements, resulting in the repetitive blue/pink pattern in the Forbidden region. c, Microwave reflectometry stability diagrams as a function of $V_{S13}$ and $V_{DL1}$ for the Coulomb peaks $\alpha$, $\beta$, $\gamma$, $\delta$, and $\epsilon$ at $V_{WL3}=1.5 \text{ V}$. Inset: energy diagram of Source-QD transition generating the signal $\beta$. The measurements are performed at 50 mK, and the carrier frequency is $f_{\text{sw1}}=6.877 \text{ GHz}$.

RADIO-FREQUENCY CHARACTERIZATION

We designed the resonators $\text{Resonator}_j$ to have different resonant frequencies $f_i$, so as to enable frequency selective readout over the corresponding rows. Fig. 1e shows the frequency spectrum for the integrated LC resonators. At 300 K, the reflection coefficient $S_{11}$ presents 3 minima at the resonant frequencies $(f_1, f_2, f_3)=(6.810, 7.374, 7.941) \text{ GHz}$ (see grey trace). At 50 mK, when all $V_{WLj}$ are set to 0 V, the spectrum shows high mismatch (see blue trace), while, by turning on, for example, the column voltage $V_{WL1}$, all the access transistors on that column turn on, modifying the total impedance of the system and hence the reflected power (see black trace in the inset). The resonators are designed to match the high impedance of the QD device gate to the 50 $\Omega$ microwave input when the access transistors are on, allowing maximum power transfer and enhanced sensitivity [35]. To demonstrate the frequency selectivity of the readout, we then increase the data line voltage $V_{DLj}$ to turn off the access transistor $T_{11}$ (see pink dashed line in the inset). Only when $T_{11}$ is switched off, the reflected power at
$f_1$ returns to its original value. At 50 mK, the probing frequencies are identified as $(f_1, f_2, f_3) = (6.872, 7.420, 7.951)$ GHz.

Frequency-multiplexing interfaces have been previously reported [26], however typically readout frequencies below 1 GHz have been used [31]. More recently, silicon QDs were interfaced with microwave resonators in the 6–8 GHz range to explore coherent spin-photon interactions [36, 37]. Although these resonators enabled fast state readout [38], hybrid manufacturing was necessary. Here, the resonators and the QDs are co-integrated in the same industrial CMOS process.

Operating at higher readout frequency has several advantages. Firstly, it reduces the footprint of the inductors, the largest elements of the architecture. Furthermore, the resonator quality factor, critical for the sensitivity of the technique [39], is higher for smaller inductors used at higher frequencies. Our quality factors are modest ($Q < 100$), compared to superconductor-based resonators, but show the state-of-the-art of what can be achieved with standard CMOS. Finally, the sensitivity of gate-based dispersive readout is higher at higher frequencies [40].

GATE-BASED DISPERSIVE READOUT

We use the resonators as sensors to perform integrated gate-based dispersive readout of the QD charge states. The resonators produce an oscillatory voltage on the gate of the QD which can result in cyclic tunneling of electrons back and forth the electronic reservoirs [41]. This results in an equivalent capacitance that modifies the impedance of the resonator producing a change in the reflected voltage ($V_{\text{ref}}$).

To benchmark the method, we performed DC transport measurements, shown in Fig. 2a, for device $Q_{13}$, and observe Coulomb peaks $\gamma$, $\delta$, and $\varepsilon$. We compare these results with those measured via reflectometry in the same $V_{\text{DLi}}$ region (Fig. 2b), while exploring the dependence of the state of the access transistor. At low $V_{\text{WL3}} - V_{\text{DL1}} < (0.786 - 0.340)$ V, the access transistor is highly resistive and the microwave signal is highly attenuated (Off region). At intermediate $V_{\text{WL3}} - V_{\text{DL1}}$, the access transistor is in the depletion region and the oscillatory voltage at the transistor input produces changes in the capacitance that are picked-up as large changes in the reflected signal (Forbidden region). Finally, at high $V_{\text{WL3}} - V_{\text{DL1}} > (1.278 - 0.340)$ V, the access transistor presents a low resistance state and the microwave signal can travel through (On region), exciting cyclic tunneling in the QD, which manifests as regions of enhanced $V_{\text{ref}}$. Besides the same transitions as in Fig. 2a, two additional peaks, $\alpha$ and $\beta$, are observed. These are results of cyclic tunneling to one of the electron reservoirs only (as opposed to current, that requires sizable tunneling rates to both source and drain). This highlights the efficiency of gate-based sensing in detecting electronic transitions even if the QDs are offset from the center of the channel and present low tunnel rates to one ohmic contact. To provide further evidence of the nature of these transitions, we present individual $V_{\text{ref}}$-time maps in Fig. 2c. The dependence on $V_S$ suggests that $\alpha$ and $\beta$ peaks are originated from cyclic tunneling to the source. A lineshape analysis (see Methods) from the data in Fig. 2c, reveals a signal-to-noise ratio (SNR) of 28.7 in 400 ms of integration time and a tunnel rate to the source of 48.3 GHz. These measurements represent the first demonstration of fully-integrated conditional gate-based readout of quantum dots implemented in an industrial CMOS technology.

FIG. 3: Integrated time-multiplexed readout. a, Individual gate-based reflectometry measurements for quantum devices $Q_{12}$ (blue), $Q_{13}$ (green), and $Q_{11}$ (gold) with reflectometry signals $V_{\text{ref}}$ as a function of $V_{\text{DL1}}$ and $V_{\text{S11}}$. b, Sequences of $V_{\text{WL1}}$ and $V_{\text{DL1}}$ for time-domain multiplexing reflectometry sensing of $Q_{12}$, $Q_{13}$, $Q_{11}$. $V_{\text{WL2}}$ (blue), $V_{\text{WL3}}$ (green), and $V_{\text{WL1}}$ (gold) follow a square wave between 1.5 V and 0.5 V in time. $V_{\text{DL1}}$ follows a ramping wave synchronized to $V_{\text{WL1}}$ in time domain. c, Stability diagrams for $Q_{12}\rightarrow Q_{13}\rightarrow Q_{11}$ in time domain corresponding to sequences of $V_{\text{WL1}}$ and $V_{\text{DL1}}$ in b. Measurements are performed at 50 mK and the carrier frequency is $f_{\text{ave1}}=6.872$ GHz for the measurements in a and c. Data presented in c are processed by using the data processing method in Methods.
TIME-MULTIPLEXED READOUT

We then performed time-multiplexed reflectometry measurements of QD devices in the same row, by addressing them at the frequency of the shared resonator and activating the corresponding columns, one after the other. We chose Resonator$_1$ with a carrier frequency $f_{\text{ref1}} = 6.872 \text{ GHz}$ and used the $1 \times 3$ quantum dot array [$Q_{ij}$].

We first measured the charge stability diagrams for each individual $Q_{ij}$ (Fig. 3a), when the corresponding access transistor is active. We used such data as the control set, we then performed the dynamic characterization in time-domain multiplexing and compared the results. The dynamic voltage sequence consists of a digital high $V_{\text{WLij}}$ voltage ($V_{\text{High}} = 1.5 \text{ V}$) applied to the cell to be read, while the other two cells are set at low $V_{\text{WLij}}$ voltages ($V_{\text{Low}} = 0.5 \text{ V}$). The digital values are selected according to the On-Off regions in Fig. 2b. During that period, we simultaneously applied a voltage ramp to the data-line $V_{\text{DLij}}$ to acquire the data from the $Q_{ij}$ of the corresponding cell. We then sequentially raised the $V_{\text{WL1j}}$ voltage of the next cell while keeping the other two low. The full sequence is illustrated in Fig. 3b, where we first measured $Q_{12}$ followed by $Q_{13}$ and $Q_{11}$. Finally, we repeated the full sequence as we stepped the $V_{\text{Sij}}$ voltages to acquire the charge stability maps in Fig. 3c. The match between the control set and the dynamic measurements indicates the success of the protocol.

We note that time-domain multiplexing does not necessarily require sequential addressing but can be performed in a random-access manner similar to DRAM architectures. These results represent fully-integrated time-multiplexed reflectometry measurements of silicon QDs and demonstrate an important reduction in the analog readout infrastructure of quantum circuits, since multiple devices can be read out by a single resonator.

FREQUENCY-MULTIPLEXED READOUT

Once time-multiplexing (addressing columns) in reflectometry has been demonstrated, using the second degree of freedom in the matrix, we demonstrate frequency-multiplexing (addressing rows). For this experiment, we used Resonator$_2$ at $f_{\text{ref2}} = 7.419 \text{ GHz}$ with Resonator$_1$ at $f_{\text{ref1}} = 6.873 \text{ GHz}$, to perform parallel readout of two independent QDs on different rows (addressable at different frequencies). As shown in Fig. 4a (left panels, up to 8.33 ms), the word-line voltage $V_{\text{WL2}}$ was kept high to activate column 2, while both $V_{\text{DL1}}$ and $V_{\text{DL2}}$ were ramped up to activate row 1 and 2. As a result, a $2 \times 1$ parallel set of Coulomb peak transitions from $Q_{12}$ and $Q_{22}$ was obtained (Fig. 4b, left panels, up to 8.33 ms).

Parallel readout reduces the overall integration time to read large quantum circuits by clustering in the same step the readout of a number of devices. This feature is of particular relevance to quantum error correction sequences, such as the surface code, that requires continuous qubit readout at scale [13]. These results demonstrate fully-integrated frequency-multiplexing gate-based readout of silicon quantum dots on a single chip and in the 6-8 GHz range.

TIME- AND FREQUENCY-MULTIPLEXED READOUT

Finally, we combined time- and frequency domain multiplexing. Resonator$_1$ and Resonator$_2$ were used simultaneously for sensing the sub-arrays [$Q_{12}$, $Q_{13}$] and [$Q_{22}$, $Q_{23}$], respectively. Similar to the time-domain multiplexing readout, a sequence of square waves $V_{\text{WL1}} \rightarrow V_{\text{WL3}} \rightarrow V_{\text{WL2}} \rightarrow V_{\text{WL1}}$ was applied to the word-lines, as shown in Fig. 4a, to selectively read out the...

![Image](image-url)
waves in $V_{DL1}$ and $V_{DL2}$ synchronized to $V_{WL2}$ and $V_{WL3}$, as shown in Fig. 4a, were applied to read out two parallel rows. The result is a $2 \times 2$ matrix of Coulomb maps from the addressed quantum dots (Fig. 4b). The results presented here demonstrate, for the first time, the combination of time- and frequency-multiplexing gate-based readout for silicon QDs, in a fully-integrated platform, with a scalable architecture.

CONCLUSION

We have presented a cryogenic IC fabricated using industrial 40-nm CMOS technology that contains three elements required in a silicon-based quantum computer: QDs, I/O management and multiplexed readout electronics. Our results probe the limits of integration using commercial CMOS technology and pave the way to the realization of an integrated silicon quantum processor. On the way forward, spin control will require attention. Single qubit control could be achieved by embedding the IC in a 3D cavity to perform electron spin resonance conditionally controlled by Stark shifts [42]. Two-qubit gates will require bringing the cells in close proximity to enable spin exchange interaction. Current industrial design rules for this process set the gate pitch limit to 120 nm, therefore some process customization will be required to bring it down to $\sim 70$ nm, where sizeable tunnel coupling occurs [43]. Finally, the SNR could be improved by lowering the tunnel rates closer to the probing frequency, by increasing the quality factor of the inductors using industry-compatible superconductors (TiN) [44] and by using traveling-wave or Josephson parametric amplification [45]. These improvements could reduce the minimum integration time by a factor of 100 or more, to bring it below the coherence time of the qubits [8], a necessity for error correction protocols.

[1] F. Arute et al., “Quantum supremacy using a programmable superconducting processor,” Nature 574, 505–510 (2019).
[2] T. F. Watson et al., “A programmable two-qubit quantum processor in silicon,” Nature 555, 63–67 (2018).
[3] D. J. Reilly, “Engineering the quantum-classical interface of solid-state qubits,” npj Quantum Information 1, 15011 (2015).
[4] E. Charbon et al., “Cryo-CMOS for quantum computing,” in 2016 IEEE International Electron Devices Meeting (IEDM) (2016) pp. 13.5.1–13.5.4.
[5] D. J. Reilly, “Challenges in Scaling-up the Control Interface of a Quantum Computer,” in 2019 IEEE International Electron Devices Meeting (IEDM) (2019) pp. 31.7.1–31.7.6.
[6] R. Maurand et al., “A CMOS silicon spin qubit,” Nature Communications 7, 13575 (2016).
[7] A. Montanaro, “Quantum algorithms: an overview,” npj Quantum Information 2, 15023 (2016).
[8] M. Veldhorst et al., “An addressable quantum dot qubit with fault-tolerant control-fidelity,” Nature Nanotechnology 9, 981–985 (2014).
[9] M. Urdampilleta et al., “Gate-based high fidelity spin readout in a CMOS device,” Nature Nanotechnology 14, 737–741 (2019).
[10] J. Yoneda et al., “A quantum-dot spin qubit with coherence limited by charge noise and fidelity higher than 99.9%,” Nature Nanotechnology 13, 102–106 (2018).
[11] W. Huang et al., “Fidelity benchmarks for two-qubit gates in silicon,” Nature 569, 532–536 (2019).
[12] D. M. Zajac et al., “Resonantly driven CNOT gate for electron spins,” Science 359, 439–442 (2018).
[13] A. G. Fowler et al., “Surface codes: Towards practical large-scale quantum computation,” Physical Review A 86, 032324 (2012).
[14] E. Kawakami et al., “Electrical control of a long-lived spin qubit in a Si/SiGe quantum dot,” Nature Nanotechnology 9, 666–670 (2014).
[15] T.-Y. Yang et al., “Quantum Transport in 40-nm MOSFETs at Deep-Cryogenic Temperatures,” IEEE Electron Device Letters 41, 981–984 (2020).
[16] S. Bonen et al., “Cryogenic Characterization of 22-nm FDSOI CMOS Technology for Quantum Computing ICs,” IEEE Electron Device Letters 40, 127–130 (2019).
[17] M. Veldhorst et al., “Silicon CMOS architecture for a spin-based quantum computer,” Nature Communications 8, 1766 (2017).
[18] L. M. K. Vandersypen et al., “Interfacing spin qubits in quantum dots and donors-hot, dense, and coherent,” npj Quantum Information 3, 34 (2017).
[19] R. Li et al., “A crossbar network for silicon quantum dot qubits,” Science Advances 4, 3960 (2018).
[20] L. Le Guevel et al., “A 110mK 295µW 28nm FDSOI CMOS Quantum Integrated Circuit with a 2.8GHz Excitation and nA Current Sensing of an On-Chip Double Quantum Dot,” in 2020 IEEE International Solid-State Circuits Conference - (ISSCC) (2020) pp. 306–308.
[21] A. Opremcak et al., “Measurement of a superconducting qubit with a microwave photon counter,” Science 361, 1239–1242 (2018).
[22] C. H. Yang et al., “Operation of a silicon quantum processor unit cell above one Kelvin,” Nature 580, 350–354 (2020).
[23] L. Petit et al., “Universal quantum logic in hot silicon qubits,” Nature 580, 355–359 (2020).
[24] X. Xue et al., “CMOS-based cryogenic control of silicon quantum circuits,” (2020), arXiv:2009.14185 [quant-ph].
[25] S. J. Pauka et al., “A Cryogenic Interface for Controlling Many Qubits,” (2019), arXiv:1912.01299 [quant-ph].
[26] J. M. Hornibrook et al., “Frequency multiplexing for readout of spin qubits,” Applied Physics Letters 104, 103108 (2014).
[27] S. J. Pauka et al., “Characterizing Quantum Devices at Scale with Custom Cryo-CMOS,” Phys. Rev. Applied 13, 054072 (2020).
[28] A. Potočnik et al., “Millikelvin temperature cryo-CMOS multiplexer for scalable quantum device characterisation,” (2020), arXiv:2011.11514 [quant-ph].
[29] B. Paquedt Wuetz et al., “Multiplexed quantum transport using commercial off-the-shelf CMOS at sub-Kelvin temperatures,” npj Quantum Information 6, 43 (2020).
[30] S. Schaal et al., “A CMOS dynamic random access architecture for radio-frequency readout of quantum devices,” Nature Electronics 2, 236–242 (2019).
[31] A. West et al., “Gate-based single-shot readout of spins in silicon,” Nature Nanotechnology 14, 437–441 (2019).
[32] J. I. Colless et al., “Dispersive Readout of a Few-Electron Double Quantum Dot with Fast rf Gate Sensors,” Phys. Rev. Lett. 110, 046805 (2013).
[33] A. Wallraff et al., “Strong coupling of a single photon to a superconducting qubit using circuit quantum electrodynamics,” Nature 431, 162–167 (2004).
[34] F. A. Zwanenburg et al., “Silicon quantum electronics,” Rev. Mod. Phys. 85, 961–1019 (2013).
[35] N. Ares et al., “Sensitive Radio-Frequency Measurements of a Quantum Dot by Tuning to Perfect Impedance Matching,” Phys. Rev. Applied 5, 034011 (2016).
[36] N. Samkharadze et al., “Strong spin-photon coupling in silicon,” Science 359, 1123–1127 (2018).
[37] X. Mi et al., “A coherent spin-photon interface in silicon,” Nature 555, 599–603 (2018).
[38] G. Zheng et al., “Rapid gate-based spin read-out in silicon using an on-chip resonator,” Nature Nanotechnology 14, 742–746 (2019).
[39] I. Ahmed et al., “Radio-Frequency Capacitive Gate-Based Sensing,” Phys. Rev. Applied 10, 014018 (2018).
[40] M. F. Gonzalez-Zalba et al., “Probing the limits of gate-based charge sensing,” Nature Communications 6, 6084 (2015).
[41] I. Ahmed et al., “Primary thermometry of a single reservoir using cyclic electron tunneling to a quantum dot,” Communications Physics 1, 66 (2018).
[42] A. Laucht et al., “Electrically controlling single-spin qubits in a continuous microwave field,” Science Advances 1, e1500022 (2015).
[43] F. Ansaloni et al., “Single-electron operations in a foundry-fabricated array of quantum dots,” Nature Communications 11, 6399 (2020).
[44] A. Shearrow et al., “Atomic layer deposition of titanium nitride for quantum circuits,” Applied Physics Letters 113, 212601 (2018).
[45] S. Schaal et al., “Fast Gate-Based Readout of Silicon Quantum Dots Using Josephson Parametric Amplification,” Phys. Rev. Lett. 124, 067701 (2020).
[46] A. Beckers et al., “Cryogenic MOS Transistor Model,” IEEE Transactions on Electron Devices 65, 3617–3625 (2018).
[47] B. Patra et al., “Characterization and Analysis of On-Chip Microwave Passive Components at Cryogenic Temperatures,” IEEE Journal of the Electron Devices Society 8, 448–456 (2020).
METHODS

Chip design and implementation

The chip architecture was designed and simulated using Cadence Virtuoso, and industry-standard design tool for integrated circuit design and implementation. The initial design of quantum dot transistors, access transistors and passive elements, such as inductors, capacitors and resistors in the resonators and bias-tees, was based on existing device models at 300 K provided in the process design kit (PDK) of the 40-nm CMOS foundry. Subsequently, cryogenic models [46, 47] have been used to establish more predictive circuit design at 1-4 K temperatures, based on modified compact models for transistors, adjusted lumped-element equivalent models for resistors, capacitors and inductors, and electromagnetic (EM) simulations with modified cryogenic substrate for custom microwave signal lines and connections. This allowed to provide a closer prediction of the chip performance at 50 mK, given that no established models exist for cryogenic circuit design at 1-4 K, let alone deep-cryogenic temperatures of 50 mK. Final simulations of the chip operation and performance have been performed using Cadence Virtuoso with such custom-modified cryogenic models and Keysight ADS Momentum (an industry-standard 3D planar EM solver) for electromagnetic simulations of the whole chip, including all microwave lines, resonators, input ground-signal-ground (GSG) pads, bond-wires and external printed-circuit board (PCB) substrate, using modified material properties for cryogenic operation. The physical layout implementation was also performed with Cadence Virtuoso, while the physical design verification in terms of design rule check (DRC) and layout versus schematic (LVS) was performed using Mentor Graphics Calibre. The chip was finally fabricated by a standard VLSI manufacturing process in 40-nm technology in a multi-project wafer (MPW) without any custom modification. The values of the circuit components in Fig. 1b are listed in Table I.

Experimental setup

All the measurements reported at 50 mK were performed in an Oxford Instruments Triton 200 dilution fridge. The designed chip was glued to a PCB and wire-bonded to a microwave transmission line on board. The device under test (DUT) was placed on a higher temperature (1-4 K), and a discrete cryogenic circulator (QUINSTAR QCY-G0400801) placed at the mixing chamber plate of the dilution fridge. A vector network analyzer (Rohde & Schwarz ZVA 24) has been used to measure the frequency spectrum. To perform gate-based reflectometry measurements, externally, additional low-noise amplifiers (PASTERNACK PE1524 and/or PE1522) and an IQ mixer (Marki IQ-0618MXP) have been used at room temperature to perform signal demodulation. Microwave sources (Anritsu 3692B and 3694C) were employed to provide the local oscillator signal for down-conversion in a homodyne scheme, and an oscilloscope (Teledyne LeCroy HDO4054A) was used to acquire the IF signals after they were amplified by low noise pre-amplifiers (Stanford Research Systems SR560). For time-multiplexing, arbitrary waveform generators (Keysight 33500B) were used to generate the square wave and ramp signal sequences used for \( V_{WLj} \) and \( V_{DLi} \), respectively. Finally, for (time- and frequency-multiplexed) measurements, two microwave signals have been used at room temperature with a power combiner (PASTERNACK PE2068) to generate multiplexed single-tone microwave probing signals towards the DUT and two IQ mixers have been used to demodulate the two tones separately by using the same microwave sources as local oscillators, respectively. The obtained signals have been acquired by the oscilloscope simultaneously. The complete setup is shown in Fig. 5.

Data processing

The time- (and frequency-)multiplexing readout data presented in this paper are processed data, and this is the data processing applied to it. In order to selectively address different columns in the quantum dot transistor array, square waves are applied to the word-lines when performing time-domain multiplexing readout. However, the rapid switches in the word-line voltages create large backgrounds which mask the reflectometry signals, as shown in Fig. 6a. Hence, we use polynomial fits to the data as backgrounds. After subtracting the backgrounds from the original data, as shown in Fig. 6b, the Coulomb blockade peaks are revealed and consistent with the individual reflectometry measurement for transistors \( Q_{12}, Q_{13}, \) and \( Q_{11} \) in Fig. 3a.

Data analysis

The signal-to-noise ratio (SNR) can be derived from the reflectometry signal \( V_{\text{mw}} \), as shown in Fig. 7a, and SNR=\( A/\sigma=28.7 \), where \( A \) is the Coulomb blockade peak amplitude and \( \sigma \) is the standard deviation of the noise signals. To extract the amplitude, we use a Lorentzian...
fit corresponding to the expected lineshape for lifetime broadened transitions [41]:

$$V_{mw} \propto \frac{h\gamma}{(h\gamma)^2 + (e\alpha(V_{DL1} - V_{DL1,offset}))^2},$$

(1)

where $\gamma$ is the tunnel rate, $\alpha$ the lever arm or gate coupling factor and $V_{DL1,offset}$ the data-line voltage at which the source and gate Fermi levels align. We compare the individual device measurement and the time-domain multiplexing measurement in Fig. 7b and Fig. 7c respectively, and show a benchmark in Table II. Furthermore, we estimate the minimum integration time (defined as the time to reach $\text{SNR}=1$) from the bandwidth of the measurement and the number of averages. The tunnel rate for Source-QD is estimated as well. These results are shown in Table II.

DATA AVAILABILITY

The data that support the plots within this paper and all the findings of this study are available from the corresponding author upon reasonable request.

ACKNOWLEDGMENTS

We are grateful to Simon Schaal for providing useful comments.

The research leading to these results has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 688539 and 951852. M. F. G.-Z. acknowledges support from the Royal Society.

AUTHOR CONTRIBUTIONS

A. R., M. F. G.-Z. and E. C. conceived the architecture and devised the experiments; A. R. and Y. P. designed the chip with inputs from E. C. and M. F. G.-Z.; T.-Y. Y., J. M., M. F. G.-Z. and A. R. performed the experiments and analyzed the results; A. R., T.-Y. Y. and M. F. G.-Z. wrote the manuscript with inputs from all the coauthors, M. F. G.-Z. and E. C. supervised all the experiments.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information is available for this paper.

Correspondence should be addressed to A. R. (andrea.ruffino@epfl.ch).

Request for materials should be addressed to T.-Y. Y. (tyy20@cam.ac.uk).
SUPPLEMENTARY INFORMATION

S1. Retention time study

We characterize the charge retention time for an individual cell at 50 mK. A simple equivalent circuit model is shown in Fig. 8a [30]. To extract the retention time, we apply the sequence (i) charging and (ii) discharging. For (i) charging: the cell is firstly charged by applying $V_{WL}=1.49$ V, much higher than the threshold voltage of the access transistor, while $V_{DL}=0.8$ V, as in Fig. 8b, setting the access transistor well in the ON state. This is then followed by (ii) discharging: $V_{WL}$ is reduced to 0.5 V, where the access transistor is highly resistive. The effective voltage on the QD transistor gate $V_G$ as a function of time can be expressed as:

$$V_G = V_0 \left[1 + \frac{R_{acc}}{R_G} e^{-\frac{t}{\tau}}\right],$$

where $V_0 = \frac{V_{th} R_{acc}}{R_G}$ is the equilibrium voltage at the QD transistor gate at $t \rightarrow \infty$, and $R_G$ and $R_{acc}$ are the gate leakage resistance of the QD transistor and the channel resistance of the access transistor, respectively. $\tau = \frac{C_{cell} R_G}{R_G + R_{acc}}$ is the circuit time constant, i.e., retention time, where $C_{cell}$ is the parallel sum of the QD transistor gate capacitance and storage capacitance $C_C$ in Fig. 1b. By monitoring $I_S$ after $V_{WL}$ is switched from 1.49 V to 0.5 V, Coulomb oscillations are observed as a function of time due to the decay of $V_G$ in time, as shown in Fig. 8b. The observed Coulomb peaks in the time domain, marked as $P_1$, $P_2$, $P_3$, and $P_4$, have their counterparts in the voltage domain, as shown in Fig. 8c. Combining the marked Coulomb peaks in time and voltage domains, we fit the data points to Equation (2) and find the time constant $\tau \approx 207$ ms in this cell, as shown in Fig. 8d. From Fig. 1d in the main text, we deduce that at $V_{WL}=0.5$ V, $R_{acc} > R_G$ and leakage occurs primarily through the gate resistance of the QD device. Considering $C_{cell} \sim C_C=200 \, \text{fF}$, we obtain $R_G \sim 1 \, \text{T}\Omega$. Further improvements in the gate voltage retention time will require increasing $R_G$.

It is also worth mentioning that limited charge injection is visible in Fig. 8b after switching. The design only uses a single nMOS pass transistor, instead of a transmission gate (nMOS and pMOS) for the access transistor, but additional dummy single-finger transistors have been included in the layout next to the access transistor to absorb and minimize the charge injected by switching.

S2. Resonator characterization at room temperature

We characterize the integrated LC resonators by measuring the frequency spectrum, as shown in Fig. 9. We perform the measurement at room temperature by directly connecting a VNA to the microwave port on the PCB (Set-up A). We perform numerical fittings to the data by initially estimating the resonance frequencies ($f_{res}$) where the reflection coefficient has local minima, i.e., around (6.8, 7.4, 7.9) GHz for (Resonator$_1$, Resonator$_2$, Resonator$_3$), respectively. The resonance frequencies and the reflection coefficients at the resonance frequencies ($\Delta S_{11}$) are then extracted from the fitted results (red curves in Fig. 9). Furthermore the quality factors ($Q$) can be derived from $Q = f_{res}/\Delta f$, where $\Delta f$ is the full width at half maximum of the fit. The characteristics of the three integrated resonators are listed in Table III.

It is worth mentioning that the resonators on chip are, instead of more conventional L-shaped LC matching networks, $\pi$ CLC matching networks ($C_S$, $L$, $C_P$), since the former would impose a fixed matching network quality factor determined by the ratio of source (50 $\Omega$) and load impedance (the gate of the QD device), while the latter introduces an additional degree of freedom, thus allowing to independently determine the matching network quality factor. Therefore, all components in the matching network are functional, not due to parasitics. Finally, however, the measured quality factor is in any case determined by the component quality factor, in this case mostly by inductors.
FIG. 5: Experimental measurement setup. Dilution fridge measurement set-up for the time- and frequency-multiplexed gate-based readout experiment. The DUT is placed at 50 mK on the sample stage of the dilution fridge. A low-noise amplifier and attenuators are placed at intermediate cryogenic temperatures (~3 K), and a cryogenic circulator is placed at the mixing chamber plate. At room temperature, two microwave signal sources are power-combined to generate two multiplexed single-tone probing signals $f_{mw1}$ and $f_{mw2}$, and two IQ mixers are used to demodulate the reflected signals $RF_1$ and $RF_2$ at the two frequencies, by using the microwaves sources as respective local oscillators $LO_1$ and $LO_2$. An oscilloscope acquires the I/Q outputs for each of the two tones.

FIG. 6: Data processing. a, A single sweep of reflectometry signal $V_{mw}$ (grey curve) as a function of time at $V_S=0$ V corresponding to Fig. 3c. The data show the reflectometry signals from $Q_{12}$ ($t=0$ to 8.33 ms), $Q_{13}$ ($t=8.33$ to 16.67 ms), and $Q_{11}$ ($t=16.67$ to 25 ms). The light blue, light green, and gold curves are fits to the data and used as backgrounds for data processing. b, Processed reflectometry signals of $Q_{12}$ (blue), $Q_{13}$ (green), and $Q_{11}$ (gold) after subtracting the background fits. The arrows indicate Coulomb blockade peaks.
FIG. 7: Signal-to-noise ratio analysis. a, Trace of reflectometry signal $V_{mw}$ as a function of $V_{DL1}$ from an individual device measurement, corresponding to the black line section in the bottom left inset ($Q_{13}$, left panel in Fig. 2c). The red curve is a Lorentzian fit to the experimental data. b, Trace of $V_{mw}$ as a function of $V_{DL1}$ for $Q_{13}$ (middle panel in Fig. 3a) from an individual device measurement. $A$ is the Coulomb peak amplitude from the fit. c, Trace of $V_{mw}$ as a function of $V_{DL1}$ for $Q_{13}$ in Fig. 3c from the time-domain multiplexing measurement. Top left inset: zoom in of the background noise. The measurement in a was performed in a different cool down thermal cycle from those in b and c, while the measurements in b and c are within the same cool down.

FIG. 8: Retention time experiment. a, Equivalent circuit of a single cell. b, Source current ($I_s$) as a function of time ($t$) with the application of word-line voltage $V_{WL}^{High}=1.49$ V for $t=0$ to 200 ms and $V_{WL}^{Low}=0.5$ V afterwards (green line), and a constant data-line voltage $V_{DL}=0.8$ V (blue dashed line). c, $I_s$ as a function of $V_{DL}$ at $V_{WL}=1.49$ V. d, Locations of Coulomb blockade peaks $P_1$, $P_2$, $P_3$, and $P_4$ in time (in b), as a function of the corresponding Coulomb peaks in voltage (in c). The red dashed line is an exponential decay fit.
FIG. 9: Frequency spectrum analysis. Frequency spectrum of integrated LC resonators at 300 K. The black traces are the experimental data in linear scale (upper panel) and decibel scale (bottom panel), and the red curves are squared Lorentzian fits. $\Delta S_{11}$, $f_{\text{res}}$, and $\Delta f$ are the dip amplitude, center of the dip, and the dip full width at half maximum from the fit, respectively.
TABLE I: Values of the circuit components.

| Row (i) | $C_T$ (pF) | $R_T$ (kΩ) | $C_S$ (pF) | $L$ (nH) | $C_P$ (pF) | $C_C$ (pF) |
|---------|------------|------------|------------|---------|------------|------------|
| 1       | 231        | 92         | 231        | 2.42    | 540        | 200        |
| 2       | 81         | 92         | 243        | 1.95    | 1450       | 200        |
| 3       | 79         | 92         | 210        | 1.84    | 1160       | 200        |

TABLE II: Benchmark of signal-to-noise ratio analysis for device $Q_{13}$. Coulomb peak amplitude ($A$), peak full width at half maximum (FWHM), noise standard deviation ($\sigma$), signal-to-noise ratio (SNR), measurement integration time ($t_{\text{int}}$), estimated minimum integration time ($t_{\text{min}}$), gate lever arm ($\alpha$), Source-QD electron tunnel rate, and coefficient of determination of the fit ($R^2$) for multiple data sets.

| Parameter | Formula | Individual, $Q_{13}$ (Fig. 2c) | Individual, $Q_{13}$ (Fig. 3a) | Time multiplexing, $Q_{13}$ (Fig. 3c) |
|-----------|---------|--------------------------------|--------------------------------|-----------------------------------|
| Peak amplitude ($|A|$, mV) | 783     | 10.2                           | 3.44                            |
| FWHM (mV) | 0.721   | 2.05                           | 2.13                            |
| Noise standard deviation ($\sigma$, mV) | 27.3    | 0.569                          | 0.513                           |
| Signal-to-noise ratio (SNR) | $A/\sigma$ | 28.7                           | 18.0                            | 6.70 |
| Integration time ($t_{\text{int}}$, ms) | 400 sweeps*/1 kHz** | 400                            | 400                            | 400 |
| Minimum integration time ($t_{\text{min}}$, ms) | $t_{\text{int}}/\text{SNR}^2$ | 0.487                          | 1.24                           | 8.90 |
| Lever arm ($\alpha$, eV/V) | $\Delta V_s/\Delta V_G$* | 0.555                          | 0.555                           | 0.555 |
| Tunnel rate (GHz) | $\alpha \times \text{FWHM}/2h$§ | 48.3                           | 137                            | 143 |
| $R^2$ (coefficient of determination) | 0.941   | 0.622                          | 0.486                           |

* Number of sweeps for data integration in oscilloscope
** Low-pass filter frequency
§ Planck constant
¶ $\Delta V_s$ and $\Delta V_G$ are derived from the Coulomb diamond diagram as demonstrated in Fig. 1c.

TABLE III: Characteristics of the integrated LC resonators at 300 K. Resonance frequency ($f_{\text{res}}$), reflection coefficient at the resonance frequency ($\Delta S_{11}$), and quality factor ($Q$).

| Parameter | Resonator$_1$ | Resonator$_2$ | Resonator$_3$ |
|-----------|---------------|---------------|---------------|
| $f_{\text{res}}$ (GHz) | 6.810 | 7.374 | 7.941 |
| $\Delta S_{11}$ (dB) | -6.709 | -10.830 | -11.169 |
| $Q$ factor | 14.478 | 20.111 | 14.850 |