Applied Physics Reviews

RESEARCH ARTICLE | APRIL 17 2024

CMOS on-chip thermometry at deep cryogenic temperatures

Grayson M. Noah ; Thomas H. Swift ; Mathieu de Kruijf ; Alberto Gomez-Saiz ;
John J. L. Morton ; M. Fernando Gonzalez-Zalba

Check for updates

Appl. Phys. Rev. 11, 021414 (2024)
https://doi.org/10.1063/5.0190040

View Online
Export Citation

24 April 2024 15:06:33


APL Quantum
First Articles Online
No Article Processing Charges for Submissions
Through December 31, 2024
Read Now
AIP Publishing

AIP Publishing

Applied Physics Reviews ARTICLE pubs.aip.org/aip/are

CMOS on-chip thermometry at deep cryogenic temperatures

Cite as: Appl. Phys. Rev. **11**, 021414 (2024); doi: 10.1063/5.0190040
Submitted: 2 December 2023 · Accepted: 1 April 2024 ·
Published Online: 17 April 2024

View Online Export Citation CrossMark

Grayson M. Noah,[1] Thomas H. Swift,[1,2] Mathieu de Kruijf,[1,2] Alberto Gomez-Saiz,[1] John J. L. Morton,[1,2,a)] and M. Fernando Gonzalez-Zalba[1,a)]

AFFILIATIONS

[1]Quantum Motion, 9 Sterling Way, London N7 9HJ, United Kingdom
[2]London Centre for Nanotechnology, UCL, London WC1H 0AH, United Kingdom

[a)]**Authors to whom correspondence should be addressed:** jjl.morton@ucl.ac.uk and fernando@quantummotion.tech

ABSTRACT

Accurate on-chip temperature sensing is critical for the optimal performance of modern complementary metal-oxide-semiconductor (CMOS) integrated circuits (ICs), to understand and monitor localized heating around the chip during operation. The development of quantum computers has stimulated much interest in ICs operating at deep cryogenic temperatures (typically 0.01–4 K), in which the reduced thermal conductivity of silicon and silicon oxide and the limited cooling power budgets make local on-chip temperature sensing even more important. Here, we report four different methods for on-chip temperature measurements native to CMOS industrial fabrication processes. These include secondary and primary thermometry methods and cover conventional thermometry structures used at room temperature as well as methods exploiting phenomena that emerge at cryogenic temperatures, such as superconductivity and Coulomb blockade. We benchmark the sensitivity of the methods as a function of temperature and use them to measure local excess temperature produced by on-chip heating elements. Our results demonstrate thermometry methods that may be readily integrated in CMOS chips with operation from the millikelvin range to room temperature.

© 2024 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/). https://doi.org/10.1063/5.0190040

24 April 2024 15:06:33

I. INTRODUCTION

Heat is a common by-product of information-processing technologies.[1] Inefficiencies in the information conversion process have led to power dissipation densities approaching those inside a nuclear reactor core, compromising the chip integrity due to the elevated temperatures and as a result steering the development of modern computing technologies.[2,3] Quantum information processing technologies, although theoretically dissipationless, are also subject to these power dissipation challenges, particularly taking into account that useful quantum computers will require large numbers of qubits operating dynamically at microwave frequencies.[4–7] Moreover, the "classical" digital and analog electronics that are used for the control, addressing, and readout of quantum processors dissipate power like any conventional circuit. The majority of leading quantum processor platforms demonstrate optimum behavior at deep cryogenic temperatures (0.01–4 K), including superconducting,[8,9] spin,[10,11] trapped ion,[12,13] and photonic[14] qubits, and hence, the goal of more tightly integrating the control electronics with quantum processors has motivated developing cryogenic integrated circuits (ICs).[15–18]

At these deep cryogenic temperatures, cooling power is orders of magnitude smaller,[19] while the thermal conductivity of silicon and silicon oxide substantially reduce when compared to room temperature.[20] Therefore, cryogenic ICs require power management solutions and on-chip thermometry methods capable of monitoring hotspots across the chip operating at temperatures down to tens of millikelvin (mK).

A number of on-chip thermometry techniques have been adopted across the semiconductor industry. Silicon bandgap temperature sensing utilizes the temperature dependence of the voltage across a P-N junction in a bipolar junction transistor (BJT) at a given applied current—this is typically used in the $T = 30 - 500$ K temperature range; however, with a combination of advanced diode design and low bias current, improved sensitivity for temperatures approaching 1.5 K has been demonstrated.[21–24] Local thermal monitoring can also help determine the optimal chip biasing conditions at different temperatures. Silicided polycrystalline silicon (polysilicon) structures are also used as on-chip temperature sensors based on their temperature coefficient of resistance (TCR), which can be constant down to 50 K, but vanishes around 30 K[25] (depending on the silicide material).

© Author(s) 2024

As silicided polysilicon is also often used in the gate stack of complementary metal-oxide-semiconductor (CMOS) processes, this allows for temperature measurement directly above a field effect transistor (FET) in a technique known as gate resistance thermometry (GRT).[26] This high degree of localization is especially important in ultra-thin-body (UTB) technologies such as silicon-on-insulator (SOI) and FinFETs, where oxide barriers below and/or on the sides of the UTB trap heat within the channel. GRT has been used to show that the temperature of a current-carrying transistor can be more than 10 K higher than that of other transistors on the same chip.[27]

Here, we investigate four solutions for on-chip thermometry for CMOS ICs, which can be operated down to the mK temperature range relevant for many quantum technologies. We explore to what extent existing approaches can be adapted to operate in such conditions, in addition to studying new opportunities for thermometry that emerge at these low temperatures. In particular, we study how diode thermometry (DT) based on silicon P-N junctions and GRT perform even at temperatures below 1 K, well outside of their usual operating ranges, and we identify a method of superconducting phase transition thermometry (SPTT) based on the critical current of the superconducting thin films that are present as part of the CMOS process. Each of the above methods provides a secondary thermometer that requires calibration. We also present a primary method, quantum dot thermometry (QDT), based on tunneling conductance measurements through discrete energy levels of a quantum dot formed using the CMOS process. We first introduce the four different thermometry methods in Sec. II. Next, we present their operating principle and calibration procedure in Sec. III followed by a comparative benchmark in terms of sensitivity in Sec. IV. Finally, we utilize the four methods to measure excess on-chip heating in Sec. V and discuss the results in Sec. VI.

II. ON-CHIP TEMPERATURE-SENSING TECHNIQUES

We first present the devices utilized for on-chip thermometry as well as describe the measurement setup [see Fig. 1(a)]. All devices are contained within a $3 \times 3\,\text{mm}^2$ chip fabricated using a UTB SOI process from an industrial foundry. Most devices are placed in arrays, with each individual device addressable using on-chip multiplexing. We power up the chip using 0.8 and 1.8 V supplies resulting in static power dissipation during operation. We bond the chip to a printed circuit board (PCB) that sits inside a sample puck using silicon-doped aluminum bondwires. The puck is physically connected to the mixing chamber (MXC) of a dilution refrigerator with a base temperature of 20 mK. At the MXC stage, two thermometers are placed approximately 40 cm from the sample: a BlueFors ruthenium oxide and a PT100 platinum. These thermometers measure the MXC temperature (T_{MXC}) and can be used for temperature calibration (see Sec. VII A), but as we discuss later, the accuracy of the MXC thermometers to chip temperature is compromised by the physical separation, highlighting the need for on-chip temperature readings.

A. Diode thermometry (DT)

We start with silicon vertical diode structures, specifically the two varieties PNP (top) and NPN (bottom) shown in Fig. 1(b). Both structures use shallow trench isolation (STI) to define the junction. The PNP diode is forward-biased by applying a positive voltage to a highly doped P-type contact region relative to the voltage of the N-well directly beneath it, whereas the NPN structure is designed to operate with the opposite bias configuration. We use a constant current to forward-bias the diode, and measure the voltage between contacts 1 and 2 as a function of temperature, $V_{\text{Diode}}(T)$. Unless otherwise stated, we use the NPN diode in the rest of the article.

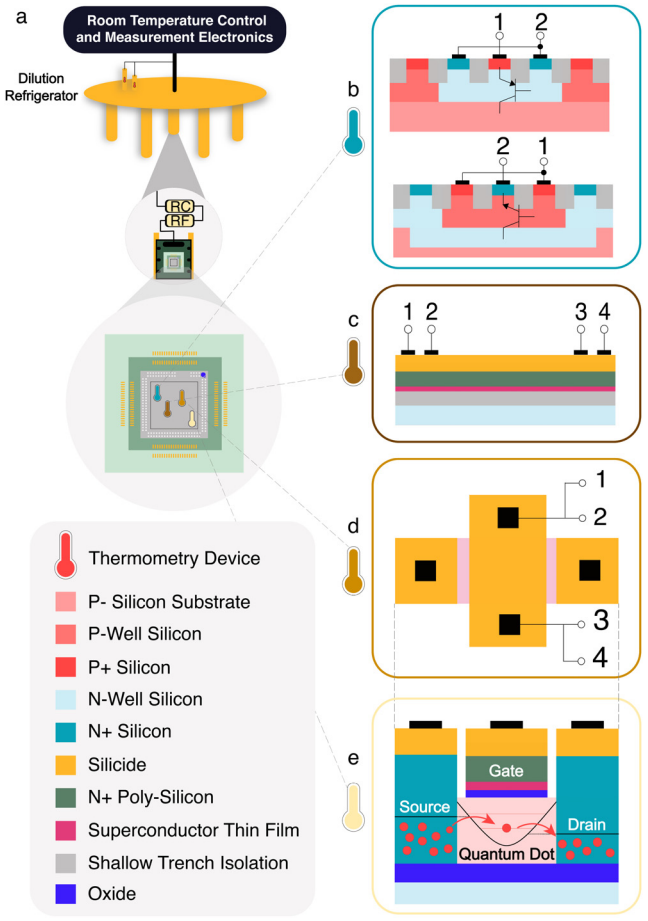

FIG. 1. (a) Diagram of cryogenic measurement setup. DC signals are carried from room-temperature electronics to the mixing chamber using DC PhBr looms thermalized at each stage of a dilution refrigerator and then RF and RC filtered. The $3 \times 3\,\text{mm}^2$ chip is glued to a PCB. The legend in the bottom left gives the materials in (b)–(e). (b) PNP and NPN diode structures. (c) Silicided polysilicon resistor structure with contacts on silicide layer allowing 4-point measurement. (d) Field-effect transistor with gate stack similar to **c**, measurement contacts are separated away from device. (e) Quantum dot transistor with overlay showing the energy level structure that results in sequential single-electron tunneling.

B. Superconducting phase transition thermometry (SPTT)

Next, in Fig. 1(c), we show the layered structure of a polysilicon resistor commonly used in CMOS processes due to the stability of its resistance value at high temperatures. This example consists of a silicided highly doped N-type polysilicon layer on top of a thin adhesion layer of superconducting thin film on SiO_2. As discussed later, the superconducting layer transitions at a critical temperature $T_C \approx 1.2$ K. We probe the superconducting nature of the film using a 4-point measurement, enabling investigation of the dependence on temperature of the critical current, $I_C(T)$.

24 April 2024 15:06:33

© Author(s) 2024

C. Gate resistance thermometry (GRT)

A layered structure similar to the polysilicon resistor also forms part of the gate stack of an FET (length, $L = 150$ nm, and width, $W = 2\,\mu$m) as seen in Fig. 1(d). We extract the temperature-dependent resistance of the gate using a 4-point measurement. For thermometry, we use the normal-state resistance $R_{\text{gate}}(T)$ above T_C and the critical current of the superconducting thin film below. This method provides the most local temperature reading of the corresponding FET.

D. Quantum dot thermometry (QDT)

Finally, we consider the FET in Fig. 1(e). Due to its small dimensions ($L = 40$ and $W = 80$ nm), the transistor can be utilized to trap individual electrons in an electrostatically defined quantum dot (QD) formed in the silicon channel directly under the gate electrode.[28,29] When operated at deep cryogenic temperatures near threshold, the silicon below the gate spacers is highly resistive due to the gradual drop of doping density from the Ohmic to the channel, effectively forming tunnel barriers between the source/drain reservoirs and the QD.[30] The line shape of the drain-source current vs gate-source voltage can be utilized as a local temperature sensor.[31,32]

1. Calibration

We now describe the calibration sequence for the different thermometers. We first calibrate DT to the MXC since the diodes can be measured without powering the auxiliary circuitry, minimizing the on-chip static power dissipation and, hence, allowing DT to closely track the MXC temperature down to 20 mK (see the calibration protocol in Sec. VII A). The remaining sensors are all measured through multiplexers, which require powering the additional circuitry leading to a minimum temperature of $T_{\text{on}} \approx 600$ mK. This figure does not represent a hard physical lower limit but is specific to the auxiliary circuit used here. Therefore, using T_{MXC} to directly calibrate the remaining sensors may lead to inaccurate readings. To increase the accuracy in the calibration, we use the diode to calibrate the remaining secondary thermometers, SPTT and GRT. Finally, QDT can be self-calibrated.

III. OPERATION PRINCIPLE

Next, we describe in detail the operation principle of all four different sensors.

A. DT

Figure 2(a) shows an example of a PNP diode IV curve measured at $T_{\text{MXC}} = 20$ mK, showing a typical exponential dependence of the current (I_{Diode}) with applied voltage across the junction (V_{Diode}). The threshold voltage of the junction is temperature-dependent, which effectively results in a change in V_{Diode} as the temperature is changed. Figure 2(b) shows this temperature dependence for a bias current $I_{\text{Diode}} = 1$ nA. We select this low bias current as it minimizes self-heating while maintaining the precision of the measurement ($\sigma_V/V_{\text{Diode}}$, where σ_V is the standard deviation) below 0.02% from 20 mK to room temperature, see bottom of Fig. 2(b). We note that significant temperature sensitivity is maintained down to 100 mK, outside the specified operating range of any commercial silicon diode thermometers.[22,23] This result may be due to the comparatively lower I_{Diode} used in this experiment or other factors such as doping concentrations and geometry. We calibrate the DT response to a weighted combination of the readings

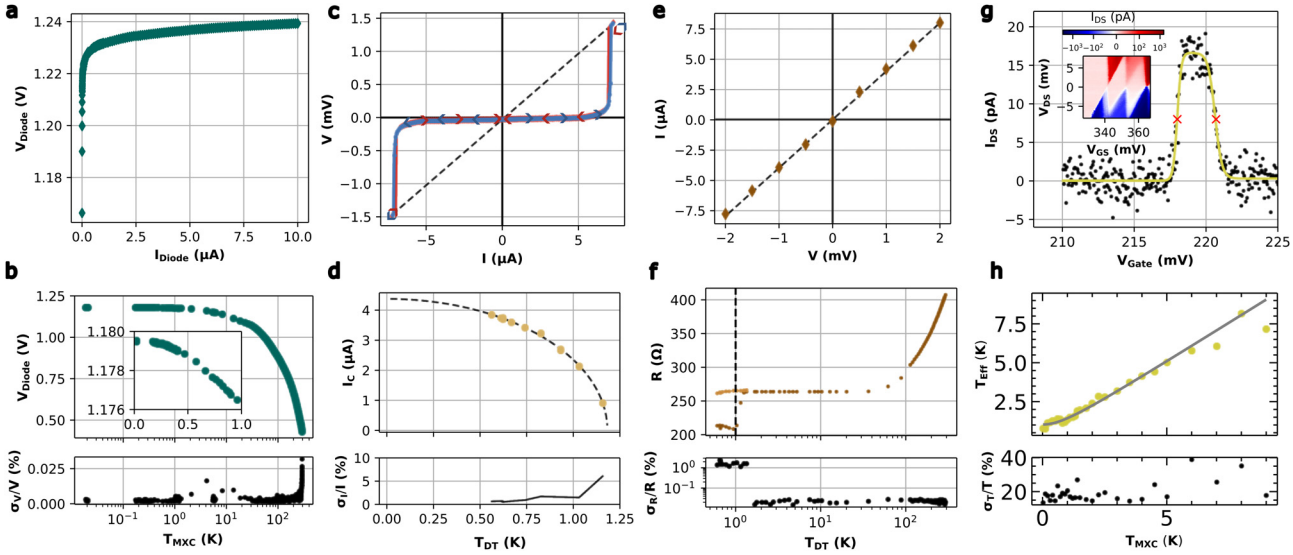

FIG. 2. (a) PNP diode IV sweep at 20 mK. (b) V_{Diode} (top) and associated measurement precision (bottom) at $I_{\text{Diode}} = 1$ nA vs MXC temperature for an NPN diode. (c) IV curve for 4-point measurement of superconducting thin film in the polysilicon resistor. (d) Critical current vs diode temperature fitted with adapted Bardeen formula [Eq. (1)]. (Bottom) Precision of the measurement vs diode temperature. (e) IV curve of quasi-4-point measurement of the gate of an FET with linear regression fit. (f) Resistance (top) and precision (bottom) vs diode temperature for the gate showing a superconducting transition at 1.2 K. (g) IV sweep of single-electron transistor at constant source-drain bias ($V_{\text{ds}} = 2$ mV) showing the fitted top hat line shape (yellow line) taken at a back-gate voltage of 2 V. The red crosses mark the $V_{0,s(d)}$ points that enable extracting the lever arm, $\alpha_g = V_{\text{ds}}/[V_{0d} - V_{0s}]$. (Inset) Coulomb blockade map taken at a back gate voltage of 0 V. (h) (Top) Extracted temperature from the Fermi fits T_{eff} vs T_{MXC} (yellow dots) and fit $T_{\text{eff}} = \sqrt{T_{\text{MXC}}^2 + T_0^2}$ (black line) showing the one-to-one dependence and saturating temperature T_0. (Bottom) Precision of the measurement vs T_{MXC}.

24 April 2024 15:06:33

© Author(s) 2024

from the MXC ruthenium oxide and the platinum thermometer assuming perfect thermalization (see Sec. VII A).

B. SPTT

We probe the superconducting thin films using a 4-point measurement to eliminate the influence of series resistances including both that of the multiplexer and vertical contact resistances through the metal and polysilicon layers. Figure 2(c) shows a typical hysteretic IV curve with a slightly larger upsweep (switching) critical current (blue line) and a slightly smaller downsweep (retrapping) critical current (red line). The switching critical current (I_C) is an intrinsic property of the superconductor, which is stochastic in nature and has a temperature-dependence following the Bardeen formula.[33] The retrapping current arises due to resistive heating,[34] which makes it less sensitive for use in thermometry, as the temperature of the thin film is already elevated above the phonon bath. Figure 2(d) shows the measured I_C vs T_{DT} (varied by changing the MXC temperature), fitted using an adapted form of the Bardeen formula (dashed line), which has been used to accurately fit the dependence of the critical current with temperature in the vicinity of the critical temperature,[35]

$$I_C(T) = I_C(T=0)\left[1-\left(\frac{T}{T_C}\right)^2\right]^i. \tag{1}$$

Here, we use the critical temperature T_C, the critical current at zero temperature $I_C(0)$, and the exponent i as fitting parameters. We find $I_C(0) = 7.95 \pm 0.06$ μA, $T_C(0) = 1.21 \pm 0.01$ K, and $i = 0.44 \pm 0.02$. The form of $I_C(T)$ means it is most sensitive to T as it approaches T_C and by applying an out-of-plane magnetic field, the sensor's point of maximum sensitivity can be tuned (see Sec. VII B). We calibrate I_C to the diode temperature, making the SPTT a secondary thermometer. We also note that we use the SPTT in a complementary way to standard SPTT, where the sharp slope of the $R-T$ curve at the superconducting-normal transition is used for sensitive thermometry over a narrow range of temperatures.[36] Finally, in the bottom panel, we show the precision as a function of T_{DT} which varies from 0.1% to 4% as the temperature approaches T_C.

C. GRT

We measure the resistance across the silicided gate structure of an FET by performing a 4-point IV measurement and using linear regression to fit the data as shown in Fig. 2(e). The resistance decreases with temperature as shown in Fig. 2(f), which is the expected result for a metal.[19] The TCR is constant down to 50 K and then goes to zero at 30 K below which the resistance does not change. However, the layer of superconducting thin film transitions at T_C which is shown by a sharp drop in resistance below 1.2 K (the residual resistance is due to the metal-silicide contact which is part of the 4-wire measurement for the FET gate). The diamond points denote the normal resistance for a bias current above I_C, demonstrating that this is a superconducting transition rather than a normal resistivity change of the material. We calibrate the thermometer to the diode, making GRT a secondary thermometer. We note the precision in the resistance measurement σ_R/R is of the order of 0.01% in the normal state and deteriorates to approximately 1% in the superconducting region.

D. QDT

The technique uses IV measurements through a discrete state in a QD to infer the electronic temperature of the charge reservoirs. In the inset of Fig. 2(g), we first confirm the presence of a QD in the channel of the FET by monitoring the drain-source current, I_{ds}, as a function of the gate-source V_{gs} and drain-source V_{ds} voltages. We observe the characteristic diamond-shaped regions of low conductance, i.e., Coulomb diamonds,[37] indicative of charge quantization. We perform a gate voltage sweep ($V_{ds} = 2$ mV) through the first Coulomb oscillation to reveal a top-hat line shape in the measured current. This line shape is characteristic of electronic transport through a discrete quantum state.[38] The data can be fit to a sequential single-electron tunneling expression,

$$I_{ds} = e\frac{\Gamma_s\Gamma_d}{\Gamma_s+\Gamma_d}, \tag{2}$$

in which e is the electronic charge, $\Gamma_{s(d)} = \Gamma_{0,s(d)} f(\varepsilon_{s(d)})$ is the source (drain) reservoir-to-QD tunnel rate,[32,39] and $f_{s(d)}(\varepsilon)$ is the Fermi distribution evaluated at the source(drain). Here, $\Gamma_{0,s(d)}$ is the maximum source(drain) tunnel rate and $\varepsilon_{s(d)} = -e\alpha_g[V_{gs(d)} - V_{0,s(d)}]$ is the energy detuning between the source(drain) Fermi level and the QD electrochemical level, which align at $V_{gs(d)} = V_{0,s(d)}$. Finally, the lever arm, α_g, the ratio between the gate and total capacitance of the QD, can be extracted from the Coulomb diamond measurement[40] giving this sensor a self-calibrating nature. We obtain $\alpha_g = 0.91 \pm 0.10$ [see Fig. 2(g)]. Since the current through the device is related to temperature by the Fermi function, which is a known physical law, the sensor is a primary thermometer. We utilize the falling edge of the line shape for temperature extraction since it avoids the impact of excited states in the measurement bias window.[41] In Fig. 2(h), we plot the extracted temperature T_{QDT} as a function of the MXC temperature. From 6 K down to approximately 1.5 K, we observe a one-to-one linear dependence between the QDT and the MXC. Below 1.5 K, the QDT reading starts to saturate and the thermometer becomes inaccurate. At $T_{MXC} = 1.12 \pm 0.05$ K, the reading plateaus. The reason for this deviation at the lower-temperature end is associated with an elevated charge noise in the sample that broadens the top-hat line shape beyond the thermal limit. The precision of the measurement is predominantly below 25% for the low temperature range but increases as the temperature exceeds 6 K, i.e., the Fermi level broadening approaches the drain-source excitation ($3.5k_B T_{MXC} \approx eV_{ds}$).

IV. SENSITIVITY BENCHMARK

Next, we proceed to benchmark the different thermometry methods. To compare thermometers, we calculate the bandwidth-normalized sensitivity,

$$\delta T = \frac{1}{\sqrt{BW}}\frac{\sigma_A}{\partial A/\partial T}, \tag{3}$$

which corresponds to the minimum resolvable temperature change in a measurement of bandwidth BW. Here, A and σ_A are the average and the standard deviation of the observable, respectively. Using the results presented in Fig. 2, we extract the sensitivity of each method and present the data in Fig. 3 as a function of temperature. For a discussion of the measurement bandwidth of the different methods, see Sec. VII C.

We first discuss the sensitivity of DT. The low-current-bias diode ($I_{Diode} = 1$ nA, light blue trace) shows a sensitivity O(1 mK/$\sqrt{\text{Hz}}$) from

24 April 2024 15:06:33

© Author(s) 2024

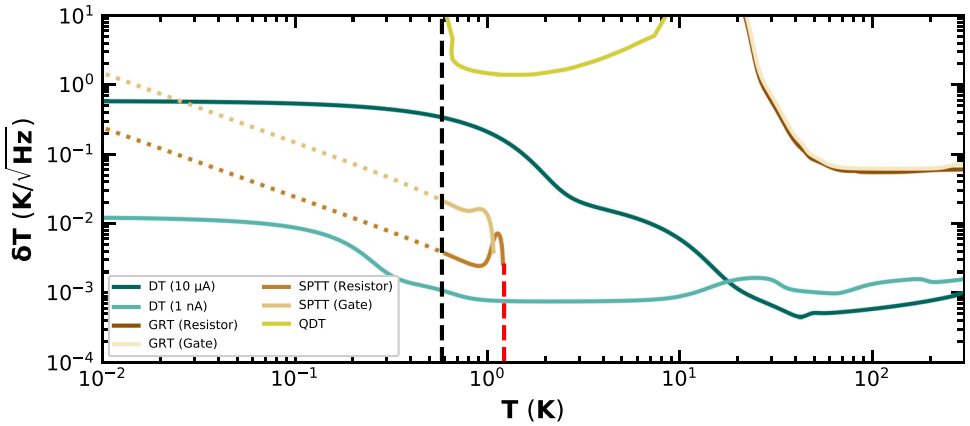

FIG. 3. Plots of the bandwidth-normalized sensitivity vs temperature (MXC for DT and QDT and diode temperature for SPTT and GRT). Diode thermometry for $I_{Diode} = 1$ nA (light blue) and $I_{Diode} = 10\ \mu$A (dark green). Superconducting phase transition thermometry for the polysilicon resistor (SPTT resistor—light brown) and for the gate material (SPTT gate—cream) used below T_C. Gate resistance thermometry for the polysilicon resistor (GRT resistor—dark brown) and for the gate material (GRT gate—light cream) used above T_C. The red dashed line represents the superconducting transition temperature of the resistor and the black dashed line is the minimum operation temperature of the chip, T_{on}.

room temperature down to 300 mK before deteriorating due to self-heating of the sensor. However, a substantial sensitivity remains even down to the lowest measured temperatures. For comparison, we show a high-current-bias result (industry-standard current of $I_{Diode} = 10\ \mu$A, dark green trace) where the sensitivity start to deteriorate below 15 K. For the other measurement techniques, the chip needs to be powered on, setting a lower bound on the temperature that can be measured at $T_{on} \approx 600$ mK (black dashed line).

Next, we discuss the SPTT sensitivity for both the standalone polysilicon resistor (SPTT resistor) and the FET gate material (SPTT gate), below T_C (red dashed line). Just below T_C, the sensitivity presents a minimum due to the sharp derivative of I_C near T_C. However, as the temperature is reduced, the high error in the determination of the critical current close to T_C and the reduced derivative results in a local maximum. As the temperature is reduced further, σ_I is reduced and the sensitivity improves reaching an optimal sensitivity point. Finally, as the temperature lowers well below T_C, the sensitivity deteriorates due to the flattening of the $I_C - T$ dependence. For temperatures below T_{on}, we use the fit to the modified Bardeen formula to extend the expected sensitivity which we indicate by the dashed extensions of the data. The sensitivity of SPTT in the FET gate material is generally worse than that of that of the standalone polysilicon resistor due to its lower resistance (due to the physical dimensions) and inclusion of contact resistances in its 4-wire measurements increasing the error in detection of the switching current transition. We also include the sensitivity of the normal-state resistances of each device, i.e., GRT gate and GRT resistor (regular resistance thermometry in the case of the standalone polysilicon resistor). Above 50 K, the sensitivity is stable at O(0.1 K/$\sqrt{\text{Hz}}$) and deteriorates sharply below due to the saturation of the resistance by scattering crystal lattice imperfections.[19] The sensitivity dependence with temperature is similar for both the polysilicon resistor (dark brown trace) and the FET gate material (cream-colored trace).

Finally, we discuss QDT. We note that since QDT is a primary thermometry technique, the observable is temperature, and hence, Eq. (3) reduces to $\delta T = \sigma_T/\sqrt{\text{BW}}$. The sensitivity is, therefore, determined by the standard deviation for a given measurement bandwidth. At high temperatures, $T \gtrsim 6$ K, the sensor's sensitivity is limited by the standard deviation of the measurement. Particularly, when the Fermi width approaches the applied source-drain bias ($3.5k_BT \approx eV_{ds}$), the current passing through the device decreases, reducing the precision as well as the accuracy of the measurement. At intermediate temperatures ($1.5 - 6$ K), the sensor presents its region of optimal sensitivity O(1 K/$\sqrt{\text{Hz}}$). At even lower temperatures, the sensitivity is limited by a combination of the elevated T_{on} and charge noise in the device. Furthermore, we must note that below 1.5 K, the sensor accuracy deteriorates, as described in Sec. III. The comparatively poor sensitivity of QDT with respect to the other methods is a consequence of the low bandwidth of the measurement (common to primary thermometers that require a functional fit to extract temperature[42–44]) needing to acquire a full IV trace with relatively small currents (tens of picoamperes).

Having discussed the operation principles and benchmarked the sensitivity of the four different methods, we present a summary of the different specifications in Table I. Overall, we find that DT provides the most sensitive and wide-ranging method for on-chip thermometry, particularly when biased using a low current (1 nA). One of the reasons for this superior sensitivity is the high measurement bandwidth as only a single data point is utilized to extract the temperature reading. Bandwidth is reduced for other thermometry methods that require a series of data points to be acquired for valid fit of the data; this is particularly intensive for SPTT and QDT, both of which require a very fine step size in their IV curves to extract a precise temperature reading. With regard to the temperature range of operation, the physical mechanism for DT is robust in the range of temperatures studied. On the other hand, SPTT and QDT are limited in range due to the physical nature of their mechanisms. The normal-superconducting transition of the superconducting thin film, with a critical temperature of $T_C \approx 1.2$ K, sets an upper limit for the SPTT while the Coulomb blockade effect does so for the QDT. In the case of QDT, the elevated charge noise is a limiting factor at the low-temperature end. This limitation is not an intrinsic effect to QDT but rather to the particular technology of the sample. In the case of GRT, in its resistive form, a low-temperature limit exists (~ 30 K) due to impurity scattering.

24 April 2024 15:06:33

© Author(s) 2024

TABLE I. Benchmark table. Comparison among the four different methods in terms of minimum and maximum temperature, sensitivity, precision, self-heating power density, type of thermometer, and integration capability within an integrated circuit (including readout). Precision is defined as σ_A/A. Integration difficulty here represents a relative measure of the level of design challenge and quantity of independent circuit functionalities required (e.g., DAC, ADC, etc.) to be integrated on the same chip. We consider an integrated sensor a circuit without any external analog electronic instruments such that a digital code representing temperature could be returned by the chip. [a]Power up limited. [b]Charge noise limited. [c]Coulomb blockade limited.

Specification	DT	GRT	SPTT	QDT
Minimum temperature (K)	0.13	30	0.6[a]	1.12[b]
Maximum temperature (K)	300+	300+	1.2	6[c]
Sensitivity (K/$\sqrt{\text{Hz}}$)	0.001–0.01	0.1–1	0.05–0.2	1–10
Precision (%)	$(2–7)\times10^{-3}$	$(2–3.5)\times10^{-2}$	0.07–4	10–25
Maximum self-heating planar power density (W/m^2)	4–10	$(3–5)\times10^{3}$	4–14	13–19
Type	Secondary	Secondary	Secondary	Primary
Integration difficulty	Low	Low	Moderate	Very high

Additionally, we compare the maximum self-heating, characterized here by the planar power density at the operation point of maximum dissipation in the temperature range of the sensor (see Sec. VII A). We find that DT, SPTT, and QDT present relatively low self-heating compared to GRT making them more suitable for low temperature thermometry as it was the case in this demonstration. Interestingly, DT performs best showing even lower self-heating than QDT despite the low power dissipation of the latter. In QDT, the power is highly concentrated at the location of the QD whereas the diode has a better surface power distribution. Finally, we note that although QDT presents an inferior sensitivity to DT (and SPTT below T_C), its primary nature allows for self-calibration without the need of a separate thermometer.

V. MEASURING ON-CHIP HEATING

Having calibrated the different thermometers and described their specifications, we now measure locally the effects of on-chip power dissipation. To heat up the chip, rather than using the MXC heaters, we utilize on-chip diodes and FETs located at different distances from the actual thermometers (see Fig. 9 in Sec. VII F).

We first present local heating measurements using DT [see Fig. 4(a)]. We heat the calibrated NPN diode thermometer by driving a current through two other diodes located at distances of 60 and 240 μm. We observe a sublinear increase in the diode temperature with heater power, P, and a more pronounced increase for the 60 μm case, as expected. We observe a similar sublinear temperature increase with power for the SPTT, GRT, and QDT as we shall discuss later [Figs. 4(b) and 4(d)]. Additionally, in Fig. 4(a), we plot the MXC temperature at each power level (black dots) and plot it along with the on-chip temperature demonstrating the disparity between the two readings and underlining the need for on-chip thermometry. In this example, T_{DT} reaches a temperature higher than 1.8 K for the 60 μm case, whereas monitoring only the MXC (which remains below 50 mK throughout) would provide no indication of such on-chip temperature rise.

To obtain a quantitative understanding of the thermal mechanisms playing a role in the steady-state on-chip temperature, we develop a model that takes into account the power dissipation, cooling power and thermal resistances present within the system. These include the quadratic dependence of the cooling power of a dilution refrigerator with temperature[44,45] as well as the dependencies of the thermal conductivity of the metals ($\propto T$) and insulating materials ($\propto T^3$) involved[46] (see Sec. VII G). This leads to a relation between the sensor temperature, T, and the heater power, P, as well as the background power sources, P_s, that elevate the chip temperature above T_{MXC} when no power is applied,

$$T^4 = T^3\alpha\sqrt{P+P_s} + \beta_h P + \beta_s P_s. \quad (4)$$

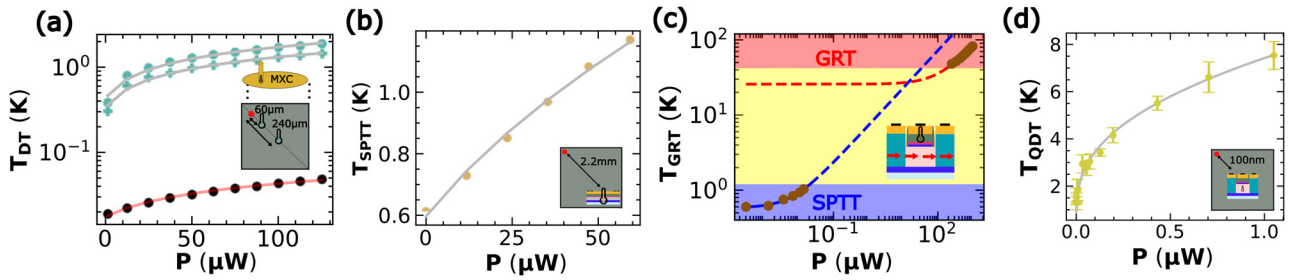

FIG. 4. On-chip temperature vs dissipated power measurements. (a) Measured using DT and heating with NPN diodes at 60 and 240 μm (blue dots). Corresponding MXC temperature (black dots). (b) Measured using SPTT and heating with a diode at 2.2 mm. (c) Self-heating measurements for FET devices using GRT (high-temperature region) and SPTT (low-temperature region). The blue fit extrapolates from the low-temperature (SPTT) points only, whereas the red fit extrapolates from the high-temperature (GRT) points only. As such, the red fit's low-temperature tapering in the yellow region should not be taken to represent a real effect since the sensor cannot measure low enough temperatures to determine its true T_0, but the blue fit's predictions at high power are more reasonable, as we expect the square root trend to continue upwards. (d) Measured using QDT and heating with an FET at 100 nm.

24 April 2024 15:06:33

© Author(s) 2024

The term including the proportionality constant α represents the joint effects of the varying metal thermal conductivity and MXC cooling power with temperature. Overall, the term presents a square root dependence of the sensor temperature with applied power. On the other hand, the terms preceded by the fitting constants β_h and β_s represent, respectively, the effect of the heater power and the background power on the sensor temperature via the insulating materials. In this case, a fourth root dependence is observed.

Coming back to DT, we perform a fit to Eq. (4) considering that the background sources of power are comparatively small ($P_s \rightarrow 0$ W) since the digital and auxiliary circuitry in this experiment are powered off. We find that both the square and fourth root terms are important in performing an accurate fit of the reading, particularly at the lowest temperatures. However, we find that in the case of SPTT, GRT, and QDT, a fit to the simplified version of Eq. (4), without the contribution of the insulating elements,

$$T = \sqrt{\alpha^2 P + T_0^2}, \tag{5}$$

is sufficient to perform accurate fits of the different thermometer readings (see Sec. VII G). Here, we have defined $T_0 = \alpha^2 P_s$ as the base temperature of the sensor. We recall that, in these cases, the support circuitry is powered up, elevating the base temperature of the chip to $T_{on} \approx 600$ mK. Our results suggest that at temperatures close to the MXC base temperature, the insulating materials play an important role in thermalizing the sensor. However, as the temperature increases, the dominant cooling mechanisms are the combined effect of the increased thermal conductivity of the metals, further facilitating thermalization, and the additional MXC cooling power.

We now look more closely at SPTT. A PNP heater diode is approximately 2.2 mm from the polysilicon resistor structure, allowing a similar experiment to be conducted using the critical current of the superconducting structure across different power levels as shown in Fig. 4(b). This method can only be used while the chip is powered up, and hence, the temperature range is limited to values between T_{on} and T_C. A fit to Eq. (5) reveals an $\alpha^{SPTT} = 130 \pm 20\,\mathrm{K\,W^{-1/2}}$ and $T_0^{SPTT} = 597 \pm 13$ mK, the latter in close agreement with T_{on}.

When performing GRT, the FET remains fully operational and, given the close distance from the channel of just a few nanometers, the experiment can effectively be considered a self-heating test. To vary the power dissipated in the channel, we sweep V_{ds} at fixed V_{gs}. For small power dissipation, we use the superconductivity of the gate stack, effectively an SPTT measurement which for the purpose of differentiation we refer to as superconducting GRT (SGRT). We show the resulting points below 1.2 K in Fig. 4(c). From the fit, we find $\alpha^{SGRT} = (7.2 \pm 1.5) \times 10^3\,\mathrm{K\,W^{-1/2}}$, and $T_0^{SGRT} = 583 \pm 12$ mK due to on-chip power dissipation of the auxiliary circuitry, as discussed above. For higher powers, the gate temperature increases to the point where the normal resistance of the silicide layer becomes temperature-sensitive, as shown in the points above 30 K. Here, we find $\alpha^{GRT} = (2.46 \pm 0.02) \times 10^3\,\mathrm{K\,W^{-1/2}}$ and $T_0^{GRT} = 25.7 \pm 1.3$ K. The lower α, compared to that of the superconducting layer, is at least partially a consequence of the larger separation between the silicide layer and the channel. The extracted T_0^{GRT} is within the range in which TCR tends to zero. The technique shows sensitivity to the most extreme levels of on-chip heating allowing self-heating characterization across 6 orders of magnitude in dissipated power with a dead zone in between. More particularly, we see that the FET gate temperature exceeds 1 K with just a few tens of nanowatts and reaches nearly 100 K with 1 mW of power dissipation.

Finally, we discuss QDT. To heat up the QDT, we use another identical FET placed parallel to the QDT at an edge-to-edge distance between channels of 100 nm. We drive a current through the FET by using a constant V_{gs} and variable V_{ds} resulting in a static power dissipation of up to 1 μW. Again, we observe a sublinear increase in temperature as a function of power well-modeled by Eq. (5). We find $\alpha^{QDT} = (8 \pm 2) \times 10^3\,\mathrm{K\,W^{-1/2}}$ indicating a higher degree of sensitivity to power dissipation than SGRT despite the higher locality of the latter. The reason for the larger α could be associated with the larger power density produced in the FET channel heating the QD (recall $W \times L = 80 \times 40$ nm) than in the FET used in the GRT experiments ($W \times L = 2000 \times 150$ nm).

Overall, the measurements presented in this section demonstrate the capability of the four different sensors to measure local excess heating well above the temperatures detected by the thermometers at the MXC. These results highlight the major importance of on-chip thermometry when assessing local temperature hotspots since they provide a much more accurate reading of the true chip thermal environment and shed light on the cooling mechanisms playing a role.

VI. CONCLUSIONS

The performance of solid-state quantum computers and cryoelectronics microcircuits is closely linked to the temperature of their environment. Understanding, hence, what the local temperature is in dynamically operated circuits and architectures is of primary importance. Here, we have introduced and benchmarked four different methods for local on-chip thermometry native to CMOS technology such as they could be seemingly integrated in scaled-up circuits. We find that DT, particularly when biased with a low current (1 nA) is the most sensitive method when compared to GRT, SPTT, and QDT for the whole temperature range studied (20 mK–300 K). Particularly, the low current bias technique enables sensitive thermometry well below 1.5 K, the common lower limit for commercial cryogenic diode thermometers. We envision that the sensitivity of the DT technique could be improved further by operating the sensor in conjunction with fast readout techniques.[47] Such an approach could surpass state-of-the-art thermal sensitivity figures[48,49] while remaining compatible with industry-standard fabrication processes and may enable the study of thermal dynamics at deep cryogenic temperatures in the microsecond timescale.

VII. METHODS

A. Chip temperature calibration and thermometer self-heating

Apart from the diodes, all the structures used in this study required powering up the additional digital and analog circuitry to enable access through the cryogenic multiplexers. The power used by the support circuitry consumes 4.3 μW and raises the on-chip temperature above that of the MXC temperature. To extract the base chip temperature when the support electronics were active, T_{on}, we used DT.

In discussing Fig. 2(b), we described the calibration procedure of the DT to the MXC thermometer which we now use to calibrate T_{on}. Specifically, we use the diode at 1 nA forward bias current and measure the temperature, T_{DT} as a function of T_{MXC} (see Fig. 5).

24 April 2024 15:06:33

© Author(s) 2024

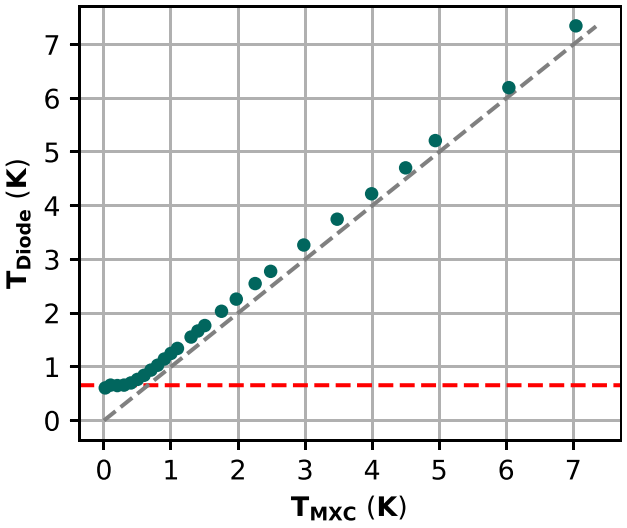

FIG. 5. Plot of measured PNP diode temperature vs MXC chamber temperature with the chip powered on. Gray dashed line shows the path the measurement should track for perfect thermalization and no power dissipation (y = x). The red dashed line is the point at which the diode temperature plateaus ($T_{on} \approx 600$ mK) which represents the minimum operation temperature of the chip when powered.

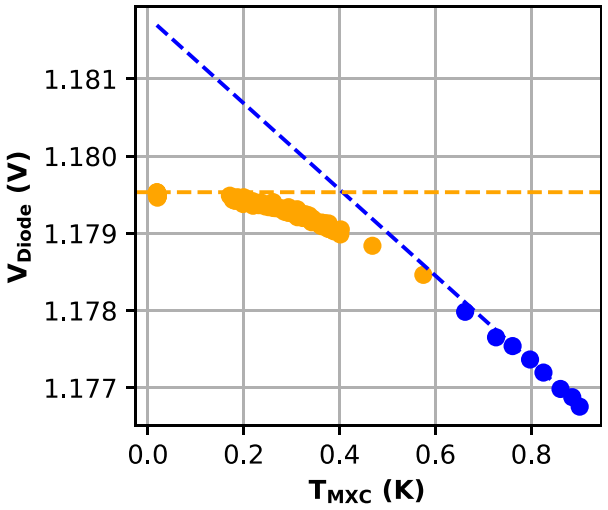

FIG. 6. V_{Diode} at $I_{diode} = 1$ nA vs MXC temperature for an NPN diode with no other on-chip power dissipation. Blue dotted line shows the assumed path the measurement should track for perfect thermalization and no self-heating. The orange dotted line is the level at which the diode voltage plateaus. The intersection of these two dotted lines near 400 mK represents the estimated minimum local temperature achieved at the diode due to self-heating.

We fit $T_{DT} = \sqrt{T_{MXC}^2 + T_{on}^2}$ and extract $T_{on} = 631 \pm 24$ mK. We conclude that the static power dissipation raises the chip temperature above the reading of the MXC thermometers highlighting the necessity of placing thermometers on chip. We make the assumption that the diode is perfectly thermalized to the MXC when the chip is powered off and that all on-chip thermometer structures are at the same temperature when the digital and auxiliary circuitry is powered up and no other heat sources are applied. Each of the on-chip thermometer structures is a similar distance from the digital circuitry at the center of the chip where the majority of the power is dissipated when the chip is powered, but the distribution of devices and auxiliary circuitry across the chip still leads to some level of error. Nevertheless, it provides a much more accurate picture of the chip temperature than direct readings from the MXC thermometer. A more advanced version may benefit from creating an array of thermometers to extract a thermal image of the powered-on chip.

We briefly dwell on the assumption of perfect thermalization of the diode to the MXC when the rest of the chip is powered off. The ~1 nW being dissipated in the diode results in some level of self-heating. The method is still useful, as the goal of a thermometer is often not to actually measure its own temperature but rather to provide an estimate of the temperature of its immediate surroundings based on the effect to its own temperature, for which the assumptions still hold well enough down to ~100 mK given the finite sensitivity of the sensor (i.e., the real assumption is that the region surrounding the diode is well thermalized to the MXC). With this understanding, we estimate the actual diode temperature due to self-heating with the MXC at base temperature. The data from the inset of Fig. 2(b) is reproduced in Fig. 6 for this purpose. Assuming the voltage deviation from the linear behavior at the low-temperature end of this plot is due to diode self-heating, a linear regression can be applied to the data in the ~0.75–1+ K region (in which the diode appears to be thermalized and follow the theoretical behavior) to describe the diode voltage vs diode temperature. The temperature at which the observed diode voltage at MXC base temperature intersects the linear regression is the minimum diode temperature due to self-heating; in this case, the estimated value is ~400 mK.

Without excellent models of cryogenic material properties and very detailed knowledge of the device structure (often not possible with commercial foundry processes), one cannot exactly quantify the self-heating in the material in which power is dissipated for most of the other discussed techniques. We suggest a figure of merit for comparing relative self-heating among the techniques: the planar power density (see Table I). Even for SPTT where no power dissipation takes place in the sensitive material, there is still a small amount of heating in the normal materials in the structure as noted in the table. Additionally, heating within contacts and routing very near the devices contributes to the actual device temperature.

B. Magnetic field dependence of SPTT

As discussed in Sec. III, the critical current of a superconductor is a temperature-dependent quantity that can be used in thermometry, as given by Eq. (1). However, in Fig. 3, we saw that the sensitivity of this technique is best for $T \lesssim T_C$. To increase the range of high sensitivity, the dependence of the critical current with magnetic field can be exploited. More particularly, both T_C and I_C are magnetic-field-dependent properties and are given by Eqs. (6) and (7).

$$T_C(B_\perp) = T_C(B_\perp = 0)\sqrt{1 - \frac{B_\perp}{B_C^\perp(T = 0)}}, \tag{6}$$

$$I_C(B_\perp) = \frac{I_C(B_\perp = 0)}{1 + \frac{B_\perp}{B_0}}. \tag{7}$$

24 April 2024 15:06:33

© Author(s) 2024

These equations can be combined to fit critical current data across a wide range of temperatures and magnetic fields. The equations use as fitting parameters: the critical out-of-plane magnetic field at zero temperature $B_C^{\perp}(T=0)$; the critical current and temperature at zero magnetic field, $I_C(B_\perp = 0)$ and $T_C(B_\perp = 0)$; and B_0 which is a macroscopic materials parameter based on a Kim-Type fit.[50] Based on the equations, we see that an increase in the out-of-plane magnetic field leads to a decrease in the critical current and the critical temperature. This decrease in critical temperature shifts the region of optimal sensitivity of the technique to lower temperatures, showing how the optimal sensitivity of the SPTT can be tuned using external magnetic fields.

C. Measurement bandwidth

In Sec. IV, we use the bandwidth-normalized sensitivity to benchmark the different thermometers. This ensures that the link between standard deviation of the measurement and integration time is taken into account. Here, we describe the details of the measurement bandwidth of each method used to quantify the sensitivity.

For DT, we use an integration time of 20 ms per point resulting in a measurement bandwidth of 50 Hz. For SPTT and GRT (and resistance measurements in general), we use 20 ms integration time for each point and take a 5 ms settling time between points. The bandwidth for these techniques is, therefore, $BW = \frac{1}{N*25\,\mathrm{ms}}$, where N is the number of points (9 for resistance measurements and variable for critical current measurements depending on step size and critical current magnitude). The determination of the resistance measurements from a set of points rather than single point measurements ensures that temperature variations in the chip leakage currents do not affect the resistance reading. Finally, for the Coulomb blockade measurements the bandwidth is 0.1 Hz, as each peak trace acquisition takes 10 s before it is then fit to extract the temperature.

D. Precision-limiting factors

The variability-based precision figure plotted in Fig. 3 indicates the amount of information conveyed by a single iteration of a measurement procedure. A smaller precision means that a single measurement conveys more information (lower uncertainty of the center value of the distribution that would be obtained from multiple measurements). This is different from measurement accuracy for which an exact knowledge of the true value to be measured is needed.

We explore the different contributions to precision for the different sensors. **For DT**, the suspected primary contributors are current source inaccuracy (0.035% + 600 pA), current source noise (~5 pA p-p 0.1–10 Hz), voltage measurement inaccuracy (0.012% + 300 μV), and additional noise from stray EM signals picked up by cabling or otherwise coupled to the system. Similar sources of error apply to the other measurement techniques, but additional factors dominate in those cases. **For GRT**, a source-measure unit (SMU) was used in 4-wire voltage-source current-measure mode. However, the inline resistances on the source and measurement lines are beyond the rated maximum values of the instrument for 4-wire mode, meaning instrument datasheet accuracy values are no longer valid and are likely significantly degraded, specifically regarding the applied voltage at the actual device terminals. This is suspected to be the leading factor limiting precision for GRT.

For SPTT, the observed variation of the switching current as shown across ten iterations of measurements in Fig. 7 is of significantly larger magnitude than expected instrument inaccuracies, implying the inherent stochastic nature of the device switching current is a leading source of inaccuracy. Additionally, since SPTT relies on detecting a hysteretic transition, it is more sensitive to noise-induced current spikes which may also play a significant role. **As for QDT**, the primary precision-limiting factor is charge noise. It is the same physical mechanism that limits the low end of the temperature range to ~1 K. In this context, this manifests itself in fluctuations from trace to trace which cause the fitting function to give a spread of values. In our measurements, we take 20 traces and use the standard deviation of these values as the precision in our measurements.

E. QDT calibration

To calibrate the QDT sensor, we need to accurately estimate the gate lever arm α_g. This factor translates the applied gate voltage to an energy scale of the QD. We obtain α_g from the width of the Coulomb oscillation in gate voltage (ΔV_g) at a fixed V_{ds} of 2 mV. The width in gate voltage is accurately fit using the Fermi-based current formula in Eq. (2) leading to $\alpha_g = V_{ds}/\Delta V_g = 0.91 \pm 0.15$.

In the top panel of Fig. 8, we plot the uncalibrated (black dots) and calibrated measurements (mustard dots). The ratio between the points is given by the constant α_g. Furthermore, in the bottom panel, we calculate the fractional accuracy of the calibrated data defined as $(T_{cal} - T_{MXC})/T_{MXC}$. At the lowest temperature, the sensor reading deviates from the MXC reading due to charge noise. Above ~7 K, the accuracy starts to increase again as the level-splitting in the QD becomes smaller than the thermal energy.

F. Chip floorplan and multiplexer implementation

On the left-hand side of Fig. 9, we show an optical image of the chip used for these experiment. The chip is bonded to an FR4 PCB using Al:Si (1%) bondwires. To the right, we show a diagram indicating the approximate location of the different thermometers and heaters.

The devices used for GRT, SPTT, and QDT are accessed via on-chip multiplexers as shown in the schematic in Fig. 10(a). To access a single device under test (DUT), the registers containing the row_en and col_en information are written such that each has only a single bit

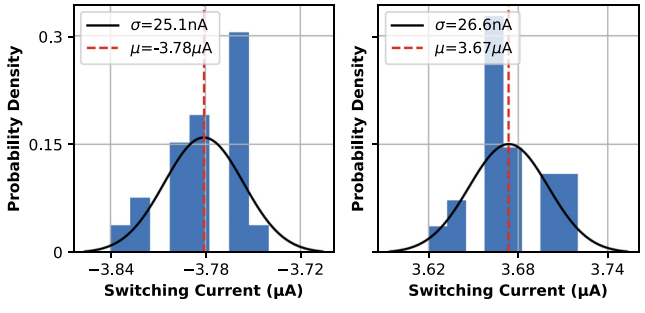

FIG. 7. Probability density of the switching current for both the positive and negative current flow directions. We fit a Gaussian distribution (black line) of mean μ (vertical red dashed line) and standard deviation σ to quantify the spread of values.

24 April 2024 15:06:33

© Author(s) 2024

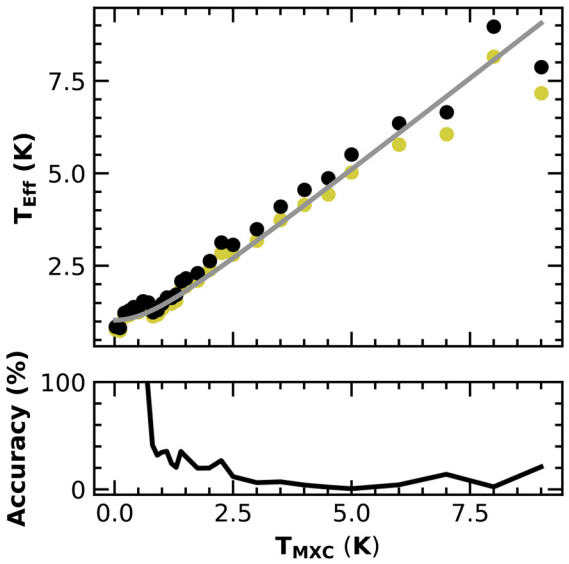

FIG. 8. QDT calibration. Top panel: Comparison of the uncalibrated (black dots) and calibrated effective temperature (mustard dots) as a function of the MXC temperature. In gray, we include a fit to the calibrated data using the expression $T_{\text{eff}} = \sqrt{T_{\text{MXC}}^2 + T_0^2}$. Bottom panel: fractional accuracy of the calibrated data with respect to the MXC temperature.

high, enabling access to the device at the corresponding x-y position in the layout in Fig. 10(b).

G. Thermal model

To fit the data in Fig. 4, we developed a steady-state (no transient effects) thermal model including the temperature dependence of thermal conductivity of the various materials and the cooling power at the mixing chamber of the dilution refrigerator. The thermal model can be illustrated in the same way as an electrical circuit, in which power is represented by current and temperature is represented by voltage [see Fig. 11(a)]. Working from the mixing chamber to the chip, the first component to model is the mixing chamber temperature. The mixing chamber cooling power is known to be proportional to the square of

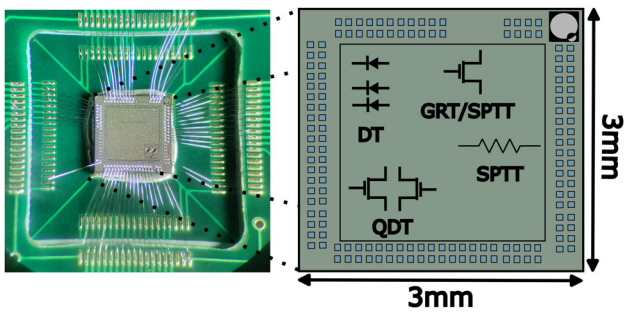

FIG. 9. The silicon chip. (Left) Optical image of the Bloomsbury chip, glued and wire-bonded to a PCB. (Right) Diagram of the Bloomsbury chip showing approximate locations of devices used for thermometry, the pad ring and the dimensions.

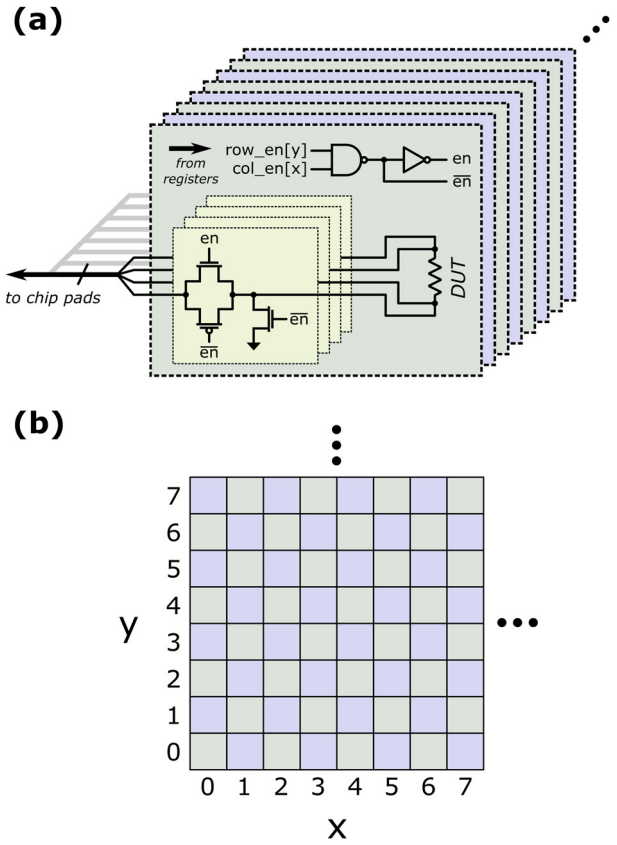

FIG. 10. (a) High-level schematics of on-chip multiplexer structure for a resistor DUT such as one shown in Fig. 1(c). FET and SET DUTs use separate similar multiplexer structures with a different number of DUT terminals as required for that DUT type. (b) Sample layout of devices to be accessed using a multiplexer.

its temperature, T_{m}.[51] We can consider a power-balanced system in which the MXC will reach a stable temperature when its cooling power matches that of the combined background sources of power dissipation P_{s} and any intentional on-chip heater power P_{h} (assuming no other heat sources on the MXC). Thus, the temperature at the mixing chamber can be thought of as a power-controlled temperature source with the following functional dependence:

$$T_{\text{m}} = k\sqrt{P_{\text{h}} + P_{\text{s}}}, \tag{8}$$

where k is a proportionality constant.

Next, we model the thermal conductivity, or equivalently thermal resistance, of the materials in the system. At low temperatures, the thermal conductivity of metals, such as Cu, Al, and Au, scale linearly with temperature, while the thermal conductivity of insulators, such as Si, SiO_2, and Si_3N_4 in the chip, scale with the cube of temperature.[46] The functional dependence of the thermal conductivity in metals is attributed to electronic thermal transport and in insulators to phononic thermal transport through the material lattice. The MXC plate and the puck in which the PCB is placed are metallic (Au-plated OFC), meaning the thermal path from the MXC to the PCB can be modeled as one lumped thermal resistance R_{m} that is inversely

24 April 2024 15:06:33

© Author(s) 2024

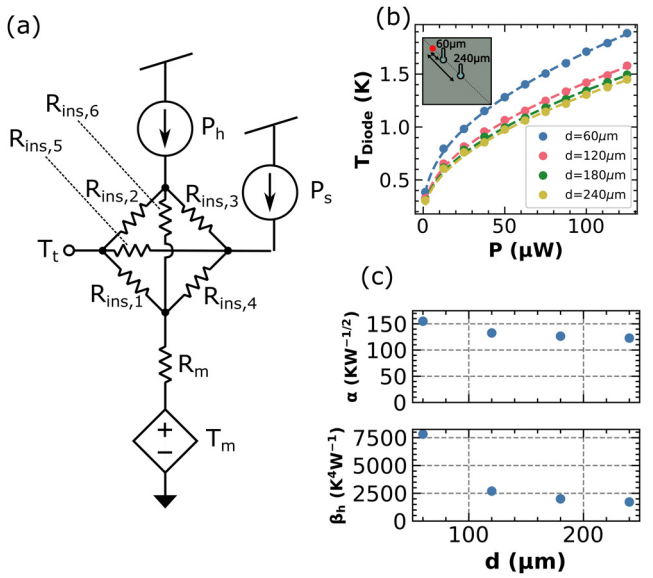

FIG. 11. Thermal model. (a) Schematic of the thermal circuit used to model the system. (b) Diode temperature vs heater power for different separations, d. Dashed lines are fits to Eq. (12). (Inset) Schematic indicating the approximate location of the closest and furthest heating diodes. (c) Extracted parameters α and β_h as a function of separation.

proportional to temperature. Furthermore, the electrical connections to the PCB are all metallic (PhBr). The net effective temperature value used for this proportionality is assumed to track linearly with the mixing chamber temperature (not necessarily equal to T_m, but some constant times T_m), meaning that the metal resistance, R_m, can be expressed as

$$R_m = a_m/T_m, \tag{9}$$

where a_m is a proportionality constant.

Several metal contact points occur along the path represented by R_m, namely, mixing chamber plate to puck head, puck head to puck rails, and puck rails to PCB. The thermal resistance of the contact points is complex to model, as their thermal conductivities likely scale with temperature raised to the power of some unknown value in the range 0.75–2.5.[52] For simplicity, we ignore the thermal contact resistance in this model and note that all contact points are Au-plated and were fastened as tightly and with as much surface area as possible to minimize the effects of thermal contact resistance in measurements. As we see in Fig. 4(a), this approximation does not hinder the quality of the fit.

After the Au-plated contact between puck rail to PCB, the PCB and chip are both made of a combination of insulator and metal materials. However, we assume that the thermal transport in both cases is to be dominated by the phononic mechanism in the insulator portions which make up the bulk of the material. Four distinct nodes must be modeled here representing four different physical locations in the system: the point at which the puck rail contacts the PCB, the location of the on-chip thermometer, the location of the on-chip heater, and the effective location of the on-chip background supply power dissipation. A complete mesh of thermal resistances between each of these nodes and each of the other three nodes is constructed, resulting in six thermal resistances $R_{ins,1}$ through $R_{ins,6}$. The temperature at the on-chip thermometer node is denoted as T_t. All six of these thermal resistances are modeled as proportional to the inverse of T_t^3 such that

$$R_{ins,n} = a_{ins,n}/(T_t^3), \tag{10}$$

where n is a number between 1 and 6 and $a_{ins,n}$ are constants. Again, this does not mean the effective temperature of each of the thermal resistances is exactly T_t, just that the effective temperature scales with T_t in a constant relationship. This assumption and the similar one for mixing chamber metal effective temperature are likely to be the most important sources of error of the model. One way to remove these assumptions would be to have a full 3D physical model of the system with temperature, heat flow, and thermal properties evaluated at many points along a spatial mesh. However, such an approach does not yield a useful analytical form of relationships like the model presented here.

The model is now sufficiently well-defined to determine T_t as a function of P_h and P_s. Tracing along the path from the mixing chamber to T_t via $R_{ins,1}$ yields the following equation:

$$T_t = T_m + R_m(P_h + P_s) + R_{ins,1}(r_h P_h + r_s P_s), \tag{11}$$

where r_h is a value between 0 and 1 representing the portion of P_h which flows through $R_{ins,1}$ and r_s is a similar value representing the portion of P_s which flows through $R_{ins,1}$. The values of r_h and r_s are constant and can be expressed in terms of $a_{ins,1}$ through $a_{ins,6}$. The key point is that the six resistances $R_{ins,1}$ through $R_{ins,6}$ form a constant power divider network through which P_h and P_s flow. Equation (11) can then be rewritten as

$$T_t^4 = T_t^3 \alpha \sqrt{P_h + P_s} + \beta_h P_h + \beta_s P_s, \tag{12}$$

where $\alpha = k + a_m/k$, $\beta_h = r_h a_{ins,1}$, and $\beta_s = r_s a_{ins,1}$. For a given P_h and P_s, the positive real-valued solution for T_t gives the modeled on-chip temperature.

For each point in the measured datasets shown in Fig. 4, T_t, P_h, and P_s are known. Thus, each dataset can be considered as a series of linear equations in α, β_h, and β_s. Since each dataset has more than three points, a Moore–Penrose inverse can be used to determine the least-squares solution for α, β_h, and β_s. Each of these three values has a physical meaning useful for analysis. The parameter α represents the joint effects of metal thermal conductance in the system and cooling power at the mixing chamber changing with temperature. The parameters β_h and β_s represent the effect of thermal conductance of the insulating materials in the chip and PCB changing with temperature as they relate to the heater power and the background supply power, respectively. It is also worth noting that for more than two on-chip heat sources, this model can be generalized to include any number of heat sources as follows:

$$T_t^4 = T_t^3 \alpha \sqrt{\sum P_x} + \sum \beta_x P_x. \tag{13}$$

Additionally, in Fig. 11(b), we plot the diode temperature vs heater power for different thermometer-heater separations. We fit the data to Eq. (12), considering $P_s \to 0$ W, to extract the thermal coefficients α and β_h and plot them in Fig. 11(c). We see a decay of the values as d is increased consistent with the reduced heat transfer efficiency at larger distances.

24 April 2024 15:06:33

© Author(s) 2024

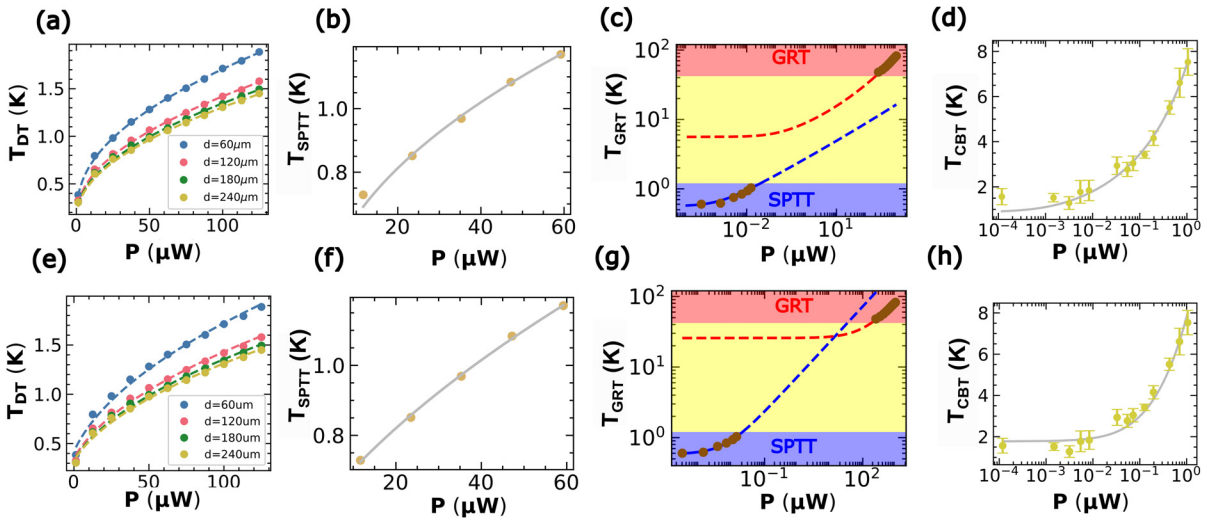

FIG. 12. Fit quality comparison of the heating data. (a)–(d) DT, SPTT, GRT, and QDT data, respectively, fitted using Eq. (4) and (e)–(h) using square root fits, Eq. (5). For the data in (c) and (g), the blue fit extrapolates from the low-temperature (SPTT) points only, whereas the red fit extrapolates from the high-temperature (GRT) points only. As such, the red fit's low-temperature tapering in the yellow region should not be taken to represent a real effect.

Finally, in Fig. 12, we show a comparison of quality of the fits to the heating data using the advanced model, including the effect of the insulating materials [Eq. (4)], and the simplified model [Eq. (5)]; see panels (a)–(h), respectively. In Table II, we compare the quality of the fits through χ^2. We observe that the diode data benefits from including the contribution of the insulating materials, whereas for the rest of the thermometers (SPTT, GRT, and QDT) can be well fit by the square root dependence.

VIII. EXPERIMENTAL SETUP

Figure 13 gives a more detailed diagram of the measurement setup used for the experiments described in Sec. II. Digital control equipment, power supplies, and low-noise SMUs are all connected to a breakout box that is connected to the top of a *Bluefors* XLD dilution refrigerator through shielded cables. These signals are then carried by PhBr twisted pairs to the MXC of the dilution refrigerator. At each temperature stage, the twisted pairs are thermalized and furthermore

TABLE II. χ^2 of the fits in Fig. 12 for the different sensor-heater arrangements. Middle column, fit to Eq. (4). Right column, fit to Eq. (5).

Thermometer	χ^2 full model	χ^2 sqrt model
Diode 60 μm	0.002	0.025
Diode 120 μm	0.009	0.018
Diode 180 μm	0.010	0.017
Diode 240 μm	0.014	0.016
SPTT	0.002	0.000
SGRT	0.013	0.001
QDT	0.749	0.622

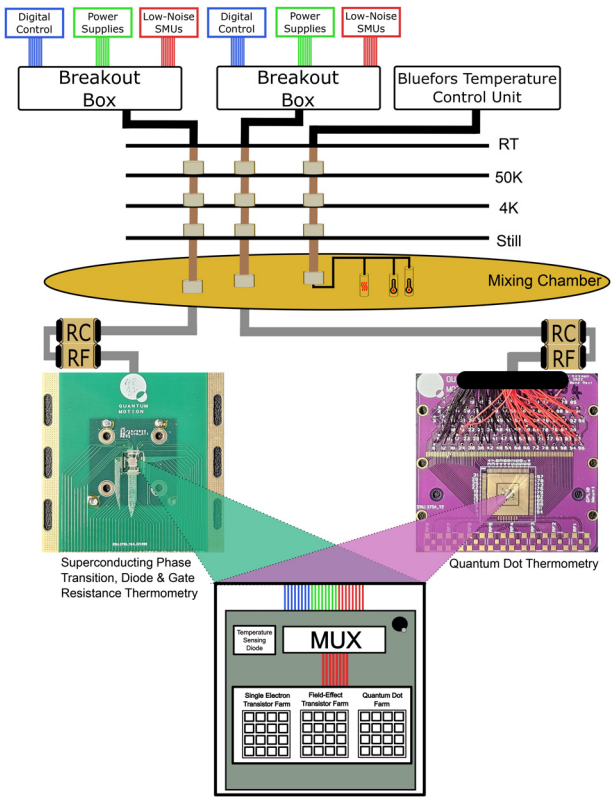

FIG. 13. Diagram showing the measurement setup used for the experiments described in Sec. II. DC wiring in the fridge is made of phosphor bronze and thermalized at each temperature stage.

24 April 2024 15:06:33

© Author(s) 2024

at the MXC stage the signals pass through a set of *QDevil* RC and RF filters before reaching the PCB and sample. We performed conventional thermometry and heating using a *Bluefors* temperature control unit combined with Bluefors ruthenium oxide and platinum thermometers and a heater, respectively, all attached to the MXC plate. The DC lines used for conventional thermometry are the same as those described above for the experimental setups. We glue the sample to a PCB (FR4, 0.8 mm thick, Au-plated Cu) using silver paste. We use two different PCB designs with nominally identical properties (see Fig. 13). The device is bonded to the PCB using 25 μm diameter Al:Si bondwire and the PCB is screwed onto a Au-plated copper bracket inside a sample puck that attaches to the bottom of the MXC plate.

ACKNOWLEDGMENTS

All authors acknowledge Jonathan Warren and James Kirkman of Quantum Motion for their technical support during this work. T.H.S. acknowledges the Engineering and Physical Sciences Research Council (EPSRC) through the Centre for Doctoral Training in Delivering Quantum Technologies (Grant No. EP/S021582/1). M.F.G.Z. acknowledges a UKRI Future Leaders Fellowship (Grant No. MR/V023284/1).

AUTHOR DECLARATIONS

Conflict of Interest

Yes, the authors declare a relevant patent application: European Patent Application No. 23157647.1.

Author Contributions

Grayson M. Noah, Thomas H. Swift, and Mathieu de Kruijf contributed equally to this work.

Grayson M. Noah: Formal analysis (equal); Investigation (equal); Validation (equal); Writing – original draft (equal); Writing – review & editing (equal). **Thomas Hugh Swift:** Formal analysis (equal); Investigation (equal); Writing – original draft (equal); Writing – review & editing (equal). **Mathieu De Kruijf:** Formal analysis (equal); Investigation (equal); Writing – original draft (equal); Writing – review & editing (equal). **Alberto Gomez-Saiz:** Methodology (equal); Writing – review & editing (equal). **John J. L. Morton:** Conceptualization (equal); Writing – review & editing (equal). **Miguel Fernando Gonzalez-Zalba:** Conceptualization (equal); Formal analysis (equal); Investigation (equal); Supervision (equal); Writing – review & editing (equal).

DATA AVAILABILITY

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

REFERENCES

[1]R. Landauer, "Irreversibility and heat generation in the computing process," IBM J. Res. Dev. **5**, 183 (1961).
[2]E. Pop, "Energy dissipation and transport in nanoscale devices," Nano Res. **3**, 147 (2010).
[3]I. Ferain, C. A. Colinge, and J.-P. Colinge, "Multigate transistors as the future of classical metal–oxide–semiconductor field-effect transistors," Nature **479**, 310 (2011).
[4]A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, "Surface codes: Towards practical large-scale quantum computation," Phys. Rev. A **86**, 032324 (2012).
[5]J. O'Gorman and E. T. Campbell, "Quantum computation with realistic magic-state factories," Phys. Rev. A **95**, 032338 (2017).
[6]D. J. Reilly, "Challenges in scaling-up the control interface of a quantum computer," in *2019 IEEE International Electron Devices Meeting (IEDM)* (IEEE, 2019), pp. 31.7.1–31.7.6.
[7]S. Bravyi, O. Dial, J. M. Gambetta, D. Gil, and Z. Nazario, "The future of quantum computing with superconducting qubits," J. Appl. Phys. **132**, 160902 (2022).
[8]J. C. Bardin, E. Jeffrey, E. Lucero, T. Huang, S. Das, D. T. Sank, O. Naaman, A. E. Megrant, R. Barends, T. White, M. Giustina, K. J. Satzinger, K. Arya, P. Roushan, B. Chiaro, J. Kelly, Z. Chen, B. Burkett, Y. Chen, A. Dunsworth, A. Fowler, B. Foxen, C. Gidney, R. Graff, P. Klimov, J. Mutus, M. J. McEwen, M. Neeley, C. J. Neill, C. Quintana, A. Vainsencher, H. Neven, and J. Martinis, "Design and characterization of a 28-nm bulk-CMOS cryogenic quantum controller dissipating less than 2 mW at 3 K," IEEE J. Solid-State Circuits **54**, 3043 (2019).
[9]R. Acharya, I. Aleiner, R. Allen, T. I. Andersen, M. Ansmann, F. Arute, K. Arya, A. Asfaw, J. Atalaya, R. Babbush, D. Bacon, J. C. Bardin, J. Basso, A. Bengtsson, S. Boixo, G. Bortoli, A. Bourassa, J. Bovaird, L. Brill, M. Broughton, B. B. Buckley, D. A. Buell, T. Burger, B. Burkett, N. Bushnell, Y. Chen, Z. Chen, B. Chiaro, J. Cogan, R. Collins, P. Conner, W. Courtney, A. L. Crook, B. Curtin, D. M. Debroy, A. Del Toro Barba, S. Demura, A. Dunsworth, D. Eppens, C. Erickson, L. Faoro, E. Farhi, R. Fatemi, L. Flores Burgos, E. Forati, A. G. Fowler, B. Foxen, W. Giang, C. Gidney, D. Gilboa, M. Giustina, A. Grajales Dau, J. A. Gross, S. Habegger, M. C. Hamilton, M. P. Harrigan, S. D. Harrington, O. Higgott, J. Hilton, M. Hoffmann, S. Hong, T. Huang, A. Huff, W. J. Huggins, L. B. Ioffe, S. V. Isakov, J. Iveland, E. Jeffrey, Z. Jiang, C. Jones, P. Juhas, D. Kafri, K. Kechedzhi, J. Kelly, T. Khattar, M. Khezri, M. Kieferová, S. Kim, A. Kitaev, P. V. Klimov, A. R. Klots, A. N. Korotkov, F. Kostritsa, J. M. Kreikebaum, D. Landhuis, P. Laptev, K.-M. Lau, L. Laws, J. Lee, K. Lee, B. J. Lester, A. Lill, W. Liu, A. Locharla, E. Lucero, F. D. Malone, J. Marshall, O. Martin, J. R. McClean, T. McCourt, M. McEwen, A. Megrant, B. Meurer Costa, X. Mi, K. C. Miao, M. Mohseni, S. Montazeri, A. Morvan, E. Mount, W. Mruczkiewicz, O. Naaman, M. Neeley, C. Neill, A. Nersisyan, H. Neven, M. Newman, J. H. Ng, A. Nguyen, M. Nguyen, M. Y. Niu, T. E. O'Brien, A. Opremcak, J. Platt, A. Petukhov, R. Potter, L. P. Pryadko, C. Quintana, P. Roushan, N. C. Rubin, N. Saei, D. Sank, K. Sankaragomathi, K. J. Satzinger, H. F. Schurkus, C. Schuster, M. J. Shearn, A. Shorter, V. Shvarts, J. Skruzny, V. Smelyanskiy, W. C. Smith, G. Sterling, D. Strain, M. Szalay, A. Torres, G. Vidal, B. Villalonga, C. Vollgraff Heidweiller, T. White, C. Xing, Z. J. Yao, P. Yeh, J. Yoo, G. Young, A. Zalcman, Y. Zhang, N. Zhu, and G. Q. AI, "Suppressing quantum errors by scaling a surface code logical qubit," Nature **614**, 676 (2023).
[10]S. Schaal, A. Rossi, V. N. Ciriano-Tejel, T.-Y. Yang, S. Barraud, J. J. L. Morton, and M. F. Gonzalez-Zalba, "A CMOS dynamic random access architecture for radio-frequency readout of quantum devices," Nat. Electron. **2**, 236 (2019).
[11]S. G. J. Philips, M. T. Madzik, S. V. Amitonov, S. L. de Snoo, M. Russ, N. Kalhor, C. Volk, W. I. L. Lawrie, D. Brousse, L. Tryputen, B. P. Wuetz, A. Sammak, M. Veldhorst, G. Scappucci, and L. M. K. Vandersypen, "Universal control of a six-qubit quantum processor in silicon," Nature **609**, 919 (2022).
[12]G. Pagano, P. W. Hess, H. B. Kaplan, W. L. Tan, P. Richerme, P. Becker, A. Kyprianidis, J. Zhang, E. Birckelbaw, M. R. Hernandez, Y. Wu, and C. Monroe, "Cryogenic trapped-ion system for large scale quantum simulation," Quantum Sci. Technol. **4**, 014004 (2018).
[13]M. F. Brandl, M. W. van Mourik, L. Postler, A. Nolf, K. Lakhmanskiy, R. R. Paiva, S. Möller, N. Daniilidis, H. Häffner, V. Kaushal, T. Ruster, C. Warschburger, H. Kaufmann, U. G. Poschinger, F. Schmidt-Kaler, P. Schindler, T. Monz, and R. Blatt, "Cryogenic setup for trapped ion quantum computing," Rev. Sci. Instrum. **87**, 113103 (2016).
[14]M. Dong, G. Clark, A. J. Leenheer, M. Zimmermann, D. Dominguez, A. J. Menssen, D. Heim, G. Gilbert, D. Englund, and M. Eichenfield, "High-speed programmable photonic circuits in a cryogenically compatible, visible–near-infrared 200 mm CMOS architecture," Nat. Photonics **16**, 59 (2022).

24 April 2024 15:06:33

© Author(s) 2024

[15]J. M. Hornibrook, J. I. Colless, I. D. Conway Lamb, S. J. Pauka, H. Lu, A. C. Gossard, J. D. Watson, G. C. Gardner, S. Fallahi, M. J. Manfra, and D. J. Reilly, "Cryogenic control architecture for large-scale quantum computing," Phys. Rev. Appl. **3**, 024010 (2015).
[16]E. Charbon, F. Sebastiano, A. Vladimirescu, H. Homulle, S. Visser, L. Song, and R. M. Incandela, "Cryo-CMOS for quantum computing," in *2016 IEEE International Electron Devices Meeting (IEDM)* (IEEE, 2016), pp. 13.5.1–13.5.4.
[17]B. Patra, R. M. Incandela, J. P. G. van Dijk, H. A. R. Homulle, L. Song, M. Shahmohammadi, R. B. Staszewski, A. Vladimirescu, M. Babaie, F. Sebastiano, and E. Charbon, "Cryo-CMOS circuits and systems for quantum computing applications," IEEE J. Solid-State Circuits **53**, 309 (2018).
[18]M. F. Gonzalez-Zalba, S. de Franceschi, E. Charbon, T. Meunier, M. Vinet, and A. S. Dzurak, "Scaling silicon-based quantum computing using CMOS technology," Nat. Electron. **4**, 872 (2021).
[19]F. Pobell, *Matter and Methods at Low Temperatures* (Springer, Berlin, Heidelberg, 1996).
[20]G. A. Slack, "Thermal conductivity of pure and impure silicon, silicon carbide, and diamond," J. Appl. Phys. **35**, 3460 (1964).
[21]S. S. Courts, P. R. Swinehart, and C. J. Yeager, "A new cryogenic diode thermometer," AIP Conf. Proc. **613**, 1620 (2002).
[22]Y. M. Shwarts, M. M. Shwarts, and S. V. Sapon, "A new generation of cryogenic silicon diode temperature sensors," in *2008 International Conference on Advanced Semiconductor Devices and Microsystems* (IEEE, 2008), p. 239.
[23]S. S. Courts, "A standardized diode cryogenic temperature sensor for aerospace applications," in *2015 Space Cryogenics Workshop*, *Phoenix, AZ, June 24–26, 2015*, [Cryogenics **74** 172 (2016)].
[24]X. Xue, B. Patra, J. P. G. van Dijk, N. Samkharadze, S. Subramanian, A. Corna, B. P. Wuetz, C. Jeon, F. Sheikh, E. Juarez-Hernandez, B. P. Esparza, H. Rampurawala, B. Carlton, S. Ravikumar, C. Nieva, S. Kim, H.-J. Lee, A. Sammak, G. Scappucci, M. Veldhorst, F. Sebastiano, M. Babaie, S. Pellerano, E. Charbon, and L. M. K. Vandersypen, "CMOS-based cryogenic control of silicon quantum circuits," Nature **593**, 205 (2021).
[25]P. A. T Hart, M. Babaie, A. Vladimirescu, and F. Sebastiano, "Characterization and modeling of self-heating in nanometer bulk-CMOS at cryogenic temperatures," IEEE J. Electron Devices Soc. **9**, 891 (2021).
[26]G. Pavlidis, S. Pavlidis, E. R. Heller, E. A. Moore, R. Vetury, and S. Graham, "Characterization of AlGaN/GaN HEMTs using gate resistance thermometry," IEEE Trans. Electron Devices **64**, 78 (2017).
[27]K. Triantopoulos, M. Cassé, S. Barraud, S. Haendler, E. Vincent, M. Vinet, F. Gaillard, and G. Ghibaudo, "Self-heating effect in FDSOI transistors down to cryogenic operation at 4.2 k," IEEE Trans. Electron Devices **66**, 3498 (2019).
[28]M. Hofheinz, X. Jehl, M. Sanquer, G. Molas, M. Vinet, and S. Deleonibus, "Simple and controlled single electron transistor based on doping modulation in silicon nanowires," Appl. Phys. Lett. **89**, 143504 (2006).
[29]T.-Y. Yang, A. Ruffino, J. Michniewicz, Y. Peng, E. Charbon, and M. F. Gonzalez-Zalba, "Quantum transport in 40-nm MOSFETs at deep-cryogenic temperatures," IEEE Electron Device Lett. **41**, 981 (2020).
[30]B. Voisin, V.-H. Nguyen, J. Renard, X. Jehl, S. Barraud, F. Triozon, M. Vinet, I. Duchemin, Y.-M. Niquet, S. de Franceschi, and M. Sanquer, "Few-electron edge-state quantum dots in a silicon nanowire field-effect transistor," Nano Lett. **14**, 2094 (2014).
[31]F. A. Zwanenburg, A. S. Dzurak, A. Morello, M. Y. Simmons, L. C. L. Hollenberg, G. Klimeck, S. Rogge, S. N. Coppersmith, and M. A. Eriksson, "Silicon quantum electronics," Rev. Mod. Phys. **85**, 961 (2013).
[32]D. Maradan, L. Casparis, T.-M. Liu, D. E. F. Biesinger, C. P. Scheller, D. M. Zumbühl, J. D. Zimmerman, and A. C. Gossard, "GaAs quantum dot thermometry using direct transport and charge sensing," J. Low Temp. Phys. **175**, 784 (2014).
[33]J. Bardeen, "Critical fields and currents in superconductors," Rev. Mod. Phys. **34**, 667 (1962).
[34]M. Tinkham, J. U. Free, C. N. Lau, and N. Markovic, "Hysteretic $i-v$ curves of superconducting nanowires," Phys. Rev. B **68**, 134515 (2003).
[35]A. V. Kuznetsov, I. I. Sannikov, and A. A. Ivanov, "Temperature dependence of critical current in $YBa_2Cu_3O_7$-δ films," J. Phys.: Conf. Ser. **941**, 012071 (2017).
[36]B. J. Luo, P. Huff, J. Schmaler, D. Hauff, F. Pröbst, and J. Wang, "Superconducting phase transition thermometers for the CRESST-experiment," AIP Conf. Proc. **1573**, 932 (2014).
[37]L. P. Kouwenhoven *et al.*, "Few-electron quantum dots," Rep. Prog. Phys. **64**, 701 (2001).
[38]I. Ahmed, J. A. Haigh, S. Schaal, S. Barraud, Y. Zhu, C-m Lee, M. Amado, J. W. A. Robinson, A. Rossi, J. J. L. Morton, and M. F. Gonzalez-Zalba, "Radio-frequency capacitive gate-based sensing," Phys. Rev. Appl. **10**, 014018 (2018).
[39]H. V. Houten, C. W. J. Beenakker, and A. A. M. Staring, "Coulomb-blockade oscillations in semiconductor nanostructures," in *Single Charge Tunneling*, NATO ASI Series (Springer, 1992), pp. 167–216.
[40]T. Ihn, *Semiconductor Nanostructures: Quantum States and Electronic Transport* (Oxford University Press Inc., 2010).
[41]I. Ahmed, A. Chatterjee, S. Barraud, J. J. L. Morton, J. A. Haigh, and M. F. Gonzalez-Zalba, "Primary thermometry of a single reservoir using cyclic electron tunneling to a quantum dot," Commun. Phys. **1**, 66 (2018).
[42]J. P. Pekola, K. P. Hirvi, J. P. Kauppinen, and M. A. Paalanen, "Thermometry by arrays of tunnel junctions," Phys. Rev. Lett. **73**, 2903 (1994).
[43]L. Spietz, K. W. Lehnert, I. Siddiqi, and R. J. Schoelkopf, "Primary electronic thermometry using the shot noise of a tunnel junction," Science **300**, 1929 (2003).
[44]Z. Iftikhar, A. Anthore, S. Jezouin, F. D. Parmentier, Y. Jin, A. Cavanna, A. Ouerghi, U. Gennser, and F. Pierre, "Primary thermometry triad at 6 mk in mesoscopic circuits," Nat. Commun. **7**, 12908 (2016).
[45]R. Richardson, *Experimental Techniques in Condensed Matter Physics at Low Temperatures* (CRC Press, 1988).
[46]P. Duthil, "Material properties at low temperature," (CAS - CERN Accelerator School, 2014), pp. 77–95.
[47]F. Vigneau, F. Fedele, A. Chatterjee, D. Reilly, F. Kuemmeth, M. F. Gonzalez-Zalba, E. Laird, and N. Ares, "Probing quantum devices with radio-frequency reflectometry," Appl. Phys. Rev. **10**, 021305 (2023).
[48]J. Chawner, S. Barraud, M. Gonzalez-Zalba, S. Holt, E. Laird, Y. A. Pashkin, and J. Prance, "Nongalvanic calibration and operation of a quantum dot thermometer," Phys. Rev. Appl. **15**, 034044 (2021).
[49]F. Blanchet, Y.-C. Chang, B. Karimi, J. T. Peltonen, and J. P. Pekola, "Radio-frequency coulomb-blockade thermometry," Phys. Rev. Appl. **17**, L011003 (2022).
[50]Y. B. Kim, C. F. Hempstead, and A. R. Strnad, "Critical persistent currents in hard superconductors," Phys. Rev. Lett. **9**, 306 (1962).
[51]Poole. *Electron Spin Resonance: A Comprehensive Treatise on Experimental Techniques* (John Wiley & Sons, 1983).
[52]L. Salerno and P. Kittel, "Thermal contact conductance," NASA Technical Memorandum No. X-215 (National Aeronautics and Space Administration, AMES Research Center, 1997).

24 April 2024 15:06:33

© Author(s) 2024