TU Delft

Delft University of Technology

Heterogeneous integration of spin–photon interfaces with a CMOS platform

Li, Linsen; Santis, Lorenzo De; Harris, Isaac B.W.; Chen, Kevin C.; Gao, Yihuai; Trusheim, Matthew; Errando-Herranz, Carlos; Du, Jiahui; Ibrahim, Mohamed I.; Englund, Dirk

DOI
10.1038/s41586-024-07371-7

Publication date
2024

Document Version
Final published version

Published in
Nature

Citation (APA)
Li, L., Santis, L. D., Harris, I. B. W., Chen, K. C., Gao, Y., Trusheim, M., Errando-Herranz, C., Du, J., Ibrahim, M. I., Englund, D., & More Authors (2024). Heterogeneous integration of spin–photon interfaces with a CMOS platform. *Nature*, *630*(8015), 70-76. https://doi.org/10.1038/s41586-024-07371-7

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Green Open Access added to TU Delft Institutional Repository

'You share, we take care!' - Taverne project

https://www.openaccess.nl/en/you-share-we-take-care

Otherwise as indicated in the copyright section: the publisher is the copyright holder of this work and the author uses the Dutch legislation to make this work public.

Article

Heterogeneous integration of spin–photon interfaces with a CMOS platform

https://doi.org/10.1038/s41586-024-07371-7

Received: 14 August 2023

Accepted: 2 April 2024

Published online: 29 May 2024

Check for updates

Linsen Li[1,2]✉, Lorenzo De Santis[1,3], Isaac B. W. Harris[1,2], Kevin C. Chen[1,2], Yihuai Gao[1], Ian Christen[1,2], Hyeongrak Choi[1,2], Matthew Trusheim[1,2,4], Yixuan Song[1], Carlos Errando-Herranz[1,5], Jiahui Du[1], Yong Hu[1,2], Genevieve Clark[1,6], Mohamed I. Ibrahim[7], Gerald Gilbert[8], Ruonan Han[1,2] & Dirk Englund[1,2]✉

Colour centres in diamond have emerged as a leading solid-state platform for advancing quantum technologies, satisfying the DiVincenzo criteria[1] and recently achieving quantum advantage in secret key distribution[2]. Blueprint studies[3–5] indicate that general-purpose quantum computing using local quantum communication networks will require millions of physical qubits to encode thousands of logical qubits, presenting an open scalability challenge. Here we introduce a modular quantum system-on-chip (QSoC) architecture that integrates thousands of individually addressable tin-vacancy spin qubits in two-dimensional arrays of quantum microchiplets into an application-specific integrated circuit designed for cryogenic control. We demonstrate crucial fabrication steps and architectural subcomponents, including QSoC transfer by means of a 'lock-and-release' method for large-scale heterogeneous integration, high-throughput spin-qubit calibration and spectral tuning, and efficient spin state preparation and measurement. This QSoC architecture supports full connectivity for quantum memory arrays by spectral tuning across spin–photon frequency channels. Design studies building on these measurements indicate further scaling potential by means of increased qubit density, larger QSoC active regions and optical networking across QSoC modules.

Modularity plays a critical role in computing architectures, allowing the segregation and combination of diverse system components. This principle has been applied to quantum information processing, resulting in quantum networks consisting of several processing units interconnected through coherent channels. Such networks have been proposed for trapped ion[6], neutral atom[7] and spin-based systems[3–5], with the aim of achieving scalable distributed quantum processing. However, scalability of the qubit layer is essential for building large scale quantum systems. In this study, we introduce a hardware modular architecture, quantum system-on-chip (QSoC), that leverages the fabrication of the qubit layer with modern mass microfabrication processes for scalability. The central qubit platform in our proposed architecture uses electron-nuclear spin systems of diamond colour centres that can be generated with foundry-process ion implantation for large scale fabrication unlike systems based on atoms or ions. Diamond colour centres have emerged as promising solid-state qubits, demonstrating deterministic remote entanglement[8], minute-long coherence times with more than ten auxiliary qubits[9] and large-scale heterogeneous integration into photonic integrated circuits using diamond quantum microchiplets (QMCs)[10]. The implanted diamond colour centre provides benefits in scalability, but will suffer from naturally inhomogeneous spectral broadening[10]. To address that, we integrate the implanted diamond qubit layer on a commercially processed complementary metal–oxide–semiconductor (CMOS) backplane, allowing for local spectral tuning. The system's codesign with CMOS electronics allows the compact two-dimensional array of qubit arrangements and substantially minimizes the size of the system's control elements. CMOS electronics have been used in gate-defined quantum dot control[11], low-power cryogenic microwave control[12] and integration with nitrogen vacancy centres for sensing applications[13,14].

The QSoC demonstrated here contains 64 QMCs with a codesigned CMOS application-specific integrated circuit (ASIC). QSoC not only facilitates qubit scaling but also enables qubit inhomogeneous compensation for full connectivity, which benefits cluster state computational power since this depends on the largest achievable connected qubit graph[3]. The connected qubit graph can be mapped to a regular qubit square lattice layout in the quantum circuit representation for error correction[3]. In this paper, we show the architectural design concept first, followed by the QSoC fabrication, compensation, demonstrating that our QSoC architecture provides the functionality required for creation of interconnected qubit graph states. Here we report an unprecedented scale comprising more than 10,000 individual resolved diamond spin–photon interfaces for the fully connected qubit graphs with QSoC tuneability in the architecture. We also demonstrate that we can maintain the quantum emitter's spin and optical properties

[1]Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, USA. [2]Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA. [3]QuTech, Delft University of Technology, Delft, Netherlands. [4]DEVCOM, Army Research Laboratory, Adelphi, MD, USA. [5]Institute of Physics, University of Münster, Münster, Germany. [6]The MITRE Corporation, Bedford, MA, USA. [7]School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, USA. [8]The MITRE Corporation, Princeton, NJ, USA. ✉e-mail: linsenli@mit.edu; englund@mit.edu

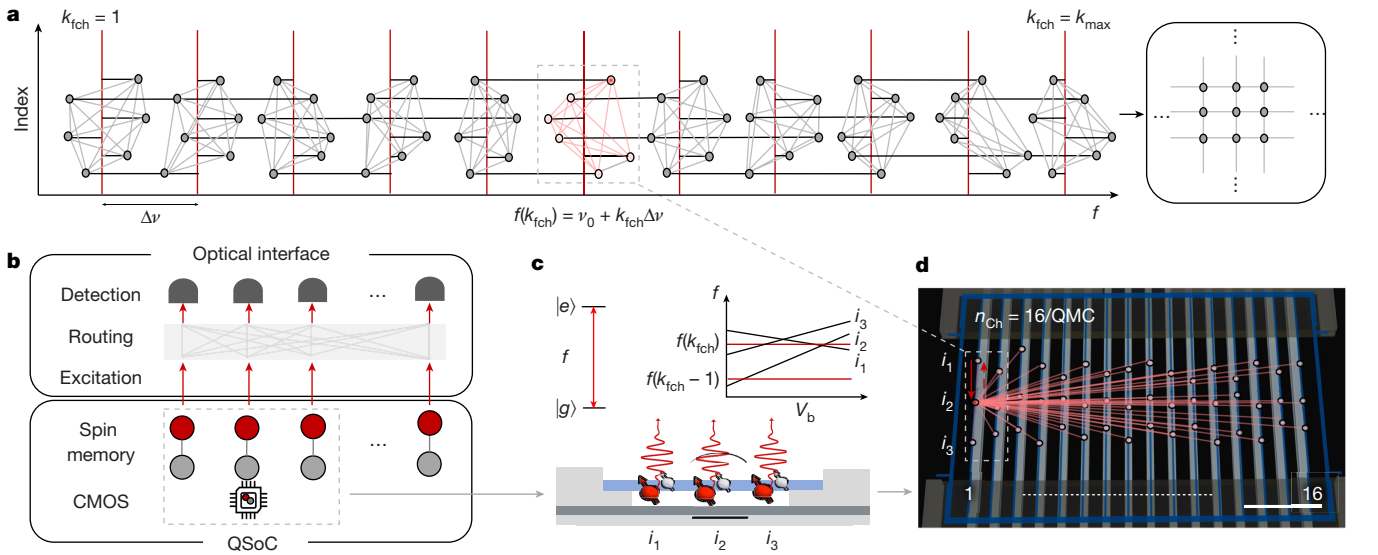


Fig. 1 | Comprehensive architectural design. a, Illustration of the architecture goal for building a connected qubit graph with qubits resonant with a set of frequencies (red lines, $f(k_{fch}) = \upsilon_0 + k_{fch}\Delta\upsilon$). The grey dot represents the artificial atom qubit that, index marked by location, can tune to a specific frequency channel k_{fch} with voltage tuning. The horizontal black line represents the spectral tuneability of this qubit. Several grey dots in one horizontal line mean that it is the same spin qubit at different tuning voltages. The capability of several qubits to tune to identical frequencies, allowing entanglement, is symbolized by a connecting grey line. **b**, A comprehensive architecture diagram, illustrating the optical interface (including optical excitation, routing and detection) and the QSoC (including the CMOS ASIC chip and spin memories). The red dots indicate the electronic spin (quantum emitter), and the grey dots represent the nuclear spin (quantum memory). **c**, A cross-section diagram of three quantum emitters (i_1, i_2 and i_3) located in a waveguide. The optical transition from ground to excited state has a transition frequency f, which is tuneable with the system tuning voltage bias V_b. The f versus V_b of different emitters is shown here, demonstrating the diverse behaviours. **d**, An expanded view illustrating quantum channels, numbered from 1 to 16, within a QMC. The illustration indicates that an emitter can interact with all other emitters of the same resonant frequency through optical routing and detection. The grey box region provides a practical example of the diagram in **a**, but with more resonant quantum emitters a certain quantum emitter can interact across the entire chip. Scale bar, 10 μm.

with efficient spin state preparation and measurement on such a new platform and analyse the QSoC tuneability requirement as we further scale up the system size.

Architecture goal for building a fully connected qubit graph

The largest connected qubit graph determines the quantum computational power of the cluster state[3] so we aim to build a fully connected graph to maximize our use of qubit resources. Figure 1a illustrates the goal of creating an interconnected qubit graph, in which each qubit resonates within a specific set of frequency channels, which number k_{fch} labelled from 1 to k_{max} for herald entanglement. This is due to the wide spectral range of the zero phonon line (ZPL) in the implanted diamond colour centres. We show an example of using the uniform frequency channel distribution with spacing $\Delta\upsilon$, around a reference frequency υ_0 here. During the precharacterization phase, we get a lookup table for each frequency. This table indexes various quantum emitters, detailing their positions and the corresponding voltages needed to resonate within a particular frequency channel.

In each frequency channel depicted, a few indexed quantum emitters are shown as examples. A red line indicates a resonant frequency channel we aim to match by tuning the quantum emitter ZPL with the CMOS backplane. Each grey dot in the figure represents a qubit, and the horizontal black line shows its potential tuning range by the CMOS backplane. A qubit that can be tuned to match a specific frequency channel $f(k_{fch})$ is indexed in the lookup table for that channel. We consider two quantum emitters interference-able[15] when they align with the same frequency channel, and their emitted ZPL photons, sharing identical polarization after filtering, fall within twice the transform-limited linewidth. This is visually represented by a pale grey line connecting the dots.

Consequently, all the grey dots form a large, interconnected graph, allowing for communication between every qubit. This graph is dynamically reconfigurable to suit various task requirements, including potential cluster connections. For example, the qubit graph is able to reprogrammed into a surface code grid structure to support error correction algorithms[3]. Moreover, by selectively using high-quality quantum emitters with narrow linewidths and bright spin–photon interfaces, this architecture is made robust against fabrication variations in the hardware.

Comprehensive system architecture

Figure 1b presents a comprehensive architecture diagram that includes the optical interface and QSoC. The optical interface encompasses optical excitation, routing and detection. The QSoC consists of a CMOS ASIC cooled to 4 K in a cryostat, which is heterogeneously integrated with a 64-diamond QMC array. The spin qubit chosen for this demonstration in the QMC is the negatively charged tin vacancy (SnV^-), offering high quantum efficiency[16] and spin performance that is compatible with cryogenic temperatures above 1 K (ref. 17).

QSoC module detail

Figure 1c illustrates the basic function of the QSoC module. The ASIC provides a voltage bias V_b to tune the ZPL transition frequency f of the quantum emitters. These can be tuned to a predefined set of frequency channels labelled as $f(k_{fch})$. An example cross-section in the QSoC showcases the tuning response behaviour of different quantum emitters (i_1, i_2 and i_3) with varying f as a function of V_b. Some quantum emitters, such as i_2, can couple to a resonant dielectric antenna to enhance free-space coupling (Methods).

Figure 1d shows a 3D representation of the CMOS circuit layout cointegrated with diamond QMCs, each featuring $n_{Ch} = 16$ channels.

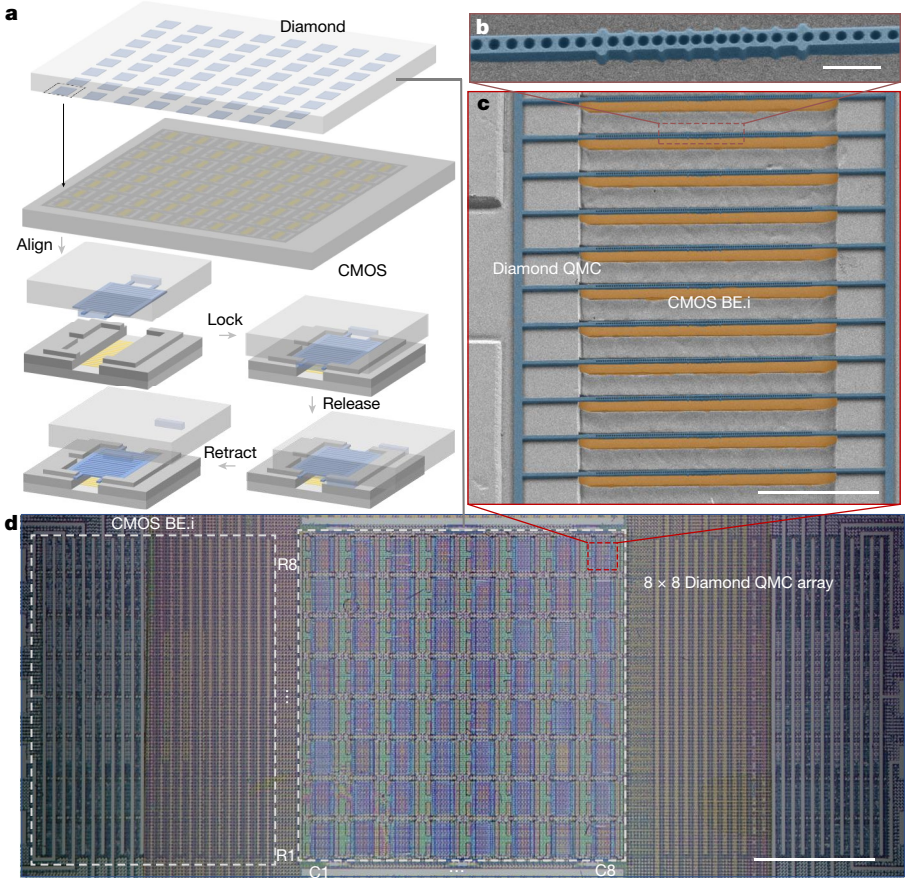


Fig. 2 | QSoC fabrication. a, Schematic representation of the lock-and-release heterogeneous integration process used to transfer the QMC arrays from the diamond parent chip to the postprocessed CMOS chip. This process involves alignment, locking, releasing and retracting. **b**, An SEM image of a quantum channel (light blue) within the QMC, using a dielectric antenna design optimized for photon collection in free space. **c**, An enlarged SEM image of **d** highlighting the central region of a single QMC. The orange colour indicates the visible CMOS individual backplane electrodes (BE.i). **d**, An optical microscope image of 1,024 diamond-resonant dielectric antennas integrated on the CMOS control chip, providing a broader view of the QMC array with rows 1–8 and columns 1–8. Scale bars, 1 µm (**b**), 10 µm (**c**), 200 µm (**d**).

Each quantum channel is integrated with the diamond-resonant dielectric antenna that designs an efficient optical interface based on a one-dimensional photonic crystal cavity design. The antenna incorporates a cointegrated vertically radiating grating coupler, resulting in a greater efficiency of 96% free-space collection efficiency within a numerical aperture (NA) of 0.9 in simulation (Methods). A metal layer on the CMOS backplane chip facilitates electronic signal routing from the external electronic source to each QMC. Future iterations of the CMOS chip can incorporate built-in digital logic and an analogue pulse sequence for routing of quantum control signals with external sources[18].

Fabrication with lock-and-release integration

We introduced SnV^- centres through ion implantation and high temperature annealing in diamond, followed by QMC nanofabrication to define the transferable diamond nanostructure on the bulk parent diamond surface (Methods). The lock-and-release heterogeneous integration technology, a crucial step in QSoC fabrication, is illustrated in Fig. 2a (the detail of the cross-section in Extended Data Fig. 3b). This procedure enables the parallel transfer of a quantum memory matrix consisting of eight columns (C1–C8) and eight rows (R1–R8) to the central region (500 × 500 µm) of the CMOS chip socket, which includes $n_{sys} = 1{,}024$ of quantum channels in total (Methods). Here, we flip the diamond parent chip and align it with a locking structure that has been postfabricated on the Taiwan Semiconductor Manufacturing Company (TSMC) 180 nm CMOS chip, followed by fine adjustment of the QMCs with two probes (Methods). After alignment, we move the parent bulk diamond vertically to lock the QMC and then horizontally, as depicted in Fig. 2a, to break the bridges between the QMC and the bulk diamond for release, followed by retraction of the bulk diamond. Figure 2b shows a scanning electron microscopy (SEM) image of a single central quantum channel. Figure 2c presents an SEM image of a single QMC region of the chip, with the orange region indicating the individual CMOS backplane electrode region beneath the QMCs. Figure 2d shows an optical microscope image of the 1,024 quantum channels integrated into the CMOS control chip. For each quantum channel, we expect to have around three resonant quantum emitters on average at a certain optical frequency (refer to the discussion later). The number of quantum channels in this design can be readily scaled by using a larger CMOS chip corresponding to the diamond parent chip size.

System parameters

In the following sections, we discuss the measurement of essential parameters of the QSoC, which include (1) system size n_{sys}, determined by the target number of quantum channels achievable through lock-and-release heterogeneous integration; (2) emitter number per quantum channel $n_{emitter}$, representing the maximum proportion of the quantum emitter in each quantum channel that can be tuned to the same ZPL frequency (Methods); (3) spin-qubit state preparation and measurement error e_{spam} and (4) the spin–photon interface efficiency, characterized by the ZPL photon detection probability P_{det}

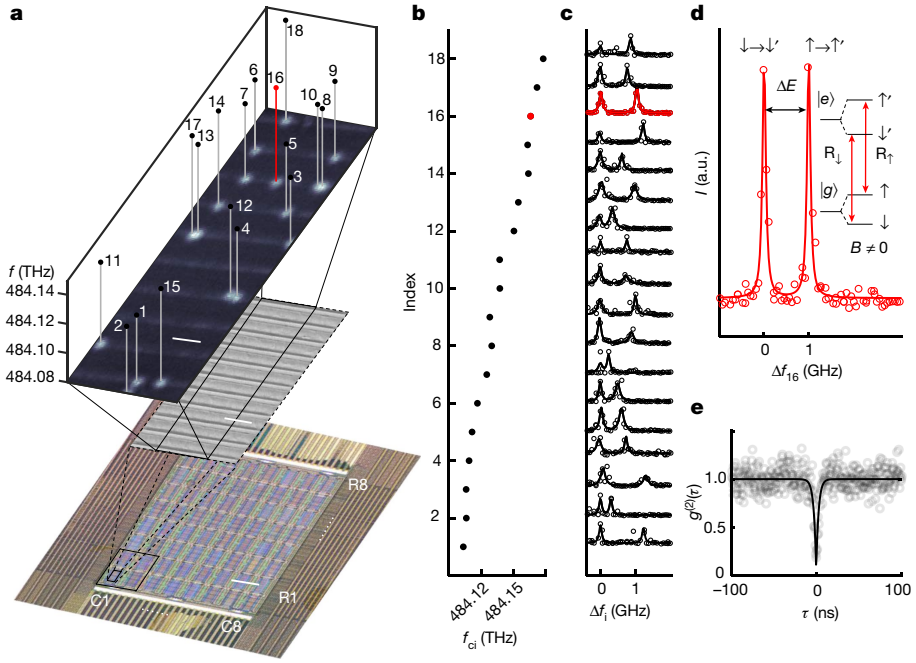


Fig. 3 | QSoC characterization. a, An overlay of the optical microscope image (scale bar, 100 µm), SEM image (scale bar, 5 µm) and superposition EMCCD image of the emitters' bright frames (scale bar, 3 µm). These images show an optically bright SnV^- under a resonant laser excitation frequency ranging from 484.123 to 484.153 THz, with corresponding index mappings indicated. The location of this example on the CMOS chip is also marked. **b**, The ZPL frequency f_{ci} of the emitter with index i in **a**. The indices are sorted according to the peak ZPL frequency from low to high, in which f_{ci} consistently represents the lowest frequency value of the double peaks in the PLE spectrum. **c**, The PLE spectrum of the marked emitters in **a** with ZPL frequency shifted by f_{ci}. **d**, An example PLE spectra of SnV^- for emitter $i = 16$ in **c**. Two spin-state transitions correspond to two peaks in the PLE with splitting ΔE. The inset illustrates the energy diagram of the SnV^- with and without the magnetic field. **e**, Autocorrelation measurements of the single SnV^- in **d**.

after spin-state initialization. We also consider the potential nanocavity enhancement (Purcell factor F_p) of a resonant dielectric antenna, which can boost the P_{det}.

Large-scale SnV^- characterization

We perform high-throughput characterization of our SnV^- qubits using optical excitation with wide-field illumination and read out from an electron-multiplying charge-coupled device (EMCCD)[19] (Methods). We show an example of this measurement on a QMC in Fig. 3. Figure 3a presents the spatial locations of SnV^- in the central QMC region. Figure 3b shows the ZPL frequency of the colour centres f_{ci}, and their normalized photoluminescence excitation (PLE) spectra are reported in Fig. 3c. All measurements and discussions regarding the transition between the ground state and excited state of SnV^- throughout the paper refer to the ZPL C line. More statistical results of the emitters are shown in Extended Data Fig. 6. Figure 3d shows an example PLE of a SnV^- with an external magnetic field $B = 0.13$ T along the diamond axis [001], showing the two spin-conserving transitions used for spin initialization and read out. The spin state can be controlled by an external microwave signal[17,20] or a modulated laser[21]. The ZPL frequency of SnV^- can be tuned to overcome the inhomogeneous distribution by means of strain[10,22]. To confirm the presence of a single quantum emitter, we performed a second-order autocorrelation ($g^{(2)}$) measurement using resonant laser excitation and PSB collection. The collected light is split into two avalanche photodiode (APD) collection paths. The $g^{(2)}$ measurement of SnV^- in Fig. 3d is illustrated in Fig. 3e. Here, $g^{(2)}(0)$ is 0.07 without background correction, indicating the presence of a single emitter as $g^{(2)}(0) < 0.5$.

SnV^- spectral tuning

Individual SnV^- can typically be spectrally resolved due to the inherent inhomogeneous distribution of their optical transitions[23]. The SnV^- optical frequency can be tuned by means of a capacitive actuator controlled by voltage[10,24] (detailed in Extended Data Fig. 7). On the basis of the statistical result, we estimate that the average tuning range is around 2 GHz within the applicable voltage range. Here, we chose a set of uniform 11 frequency channels with spacing $\Delta\nu = 2$ GHz to use inhomogeneously distributed quantum emitters within the mode-hop free laser tuning range (Methods). The QSoC can tune a quantum emitter's ZPL within the laser tuning range into one of the resonant frequency channels set on average. Figure 4a shows a wide-field image of the eight quantum channels of a QMC, focused on $k_{fch} = 1$ frequency channels. This is superimposed with the anticipated edges of the diamond nanobeam. The image highlights bright emitters, each encircled in a colour that corresponds to the optimal tuning voltage to align it with the resonant wavelength. As detailed in the inset, the process involves sweeping the voltage from 0 to 40 V on the CMOS backplane to identify the voltage level that maximizes the emitter's intensity. Within the field of view (FOV), the emitter data are represented as a point cloud shown in the inset. Here, each black dot marks the location of a quantum emitter, whereas the connecting grey lines indicate that these emitters can be tuned to resonate with the same frequency channel. Figure 4b illustrates the tuning of an emitter from frequency channel $k_{fch} = 7$ to 8 at varying voltages. The spectral tuning range from the minimum tuning voltage to the maximum tuning voltage, defined as $\Delta\nu_m$, is shown in the figure. Extended Data Fig. 7 includes a simulation of the spectral tuning effect from voltage-induced strain. The histograms underneath represent the number of emitters found in each frequency channel with light grey for a single FOV and dark grey for all 1,024 quantum channels measured in the entire CMOS chip. The number of bright spots in a single FOV at a specific frequency corresponds to the potential direct connections per qubit enabled by the all-to-all optical connectivity of the QSoC architecture (Methods). From an estimate of approximately 2,400 resonant emitters in the central region of the CMOS chip socket, we

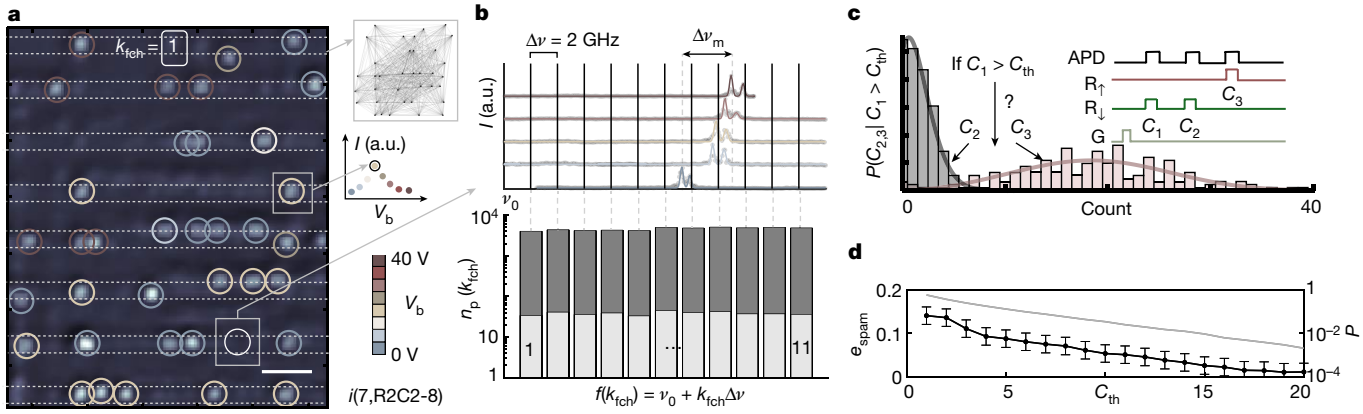


Fig. 4 | Quantum emitter spectral tuning and spin state preparation and measurement. a, Wide-field images of the QMC region at $k_{fch} = 1$. Each emitter is labelled with a circle whose colour indicates the voltage tuning V_b at which it is brightest (scale bar, 3 µm). **b**, An example of voltage-induced tuning for an emitter crossing various frequency channels. The statistics of the emitter bright spots number are shown underneath for 11 frequency channels connected by the lines, representing both shallow grey (in one FOV) and deep grey (among all the targeted 1,024 quantum channels). **c**, The spin state preparation and measurement pulse sequence, involving four programmable channels: green repump laser (G), resonant laser 1 ($R_↓$) resonating with the lower frequency transition ($|↓\rangle$ to $|↓\rangle'$), resonant laser 2 ($R_↑$) resonating with the other higher frequency transition ($|↑\rangle$ to $|↑\rangle'$) and APD. We have three collection time bins (each of duration T_M). The APD time bin 1 serves as the spin state preparation signal for postselection. We show histogram measurement counts, postselected with state preparation threshold counts $C_{th} = 18$ (APD time bin 1 read out) with $T_M = 50$ µs. Following postselection, APD time bin 2 measures the dark-state count, whereas APD time bin 3 measures the bright state counts. **d**, The relationship between state preparation and measurement error e_{spam} (black line with left axis) and successful postselection probabilities, P (grey line with right axis) with C_{th}.

calculate $n_{emitter}$ at a specific laser frequency with a maximum of 40 V CMOS backplane tuning.

Spin state preparation and measurement

Figure 4c demonstrates spin state preparation and measurement of the SnV^- presented in Fig. 3d. We collect the phonon sideband (PSB) emissions confocally using an APD. The inset shows the pulse time sequence used for state preparation and measurement (Methods). Initially, a green laser resets the SnV^- to the negative charge state[25]. Then, we herald the correct spin state of SnV^- using a laser pulse resonant with the $R_↓$ transition. We set a signal threshold count C_{th}; when the APD time bin 1 counts exceeds C_{th}, we consider the SnV^- spin state to be successfully prepared. We illustrate the count histogram of C_2 and C_3 (inset of Fig. 4c) with postselection on successful state preparation events. On the basis of the histogram result, we can calculate the spin state preparation and measurement error e_{spam} (Methods). Figure 4d shows the relationship between e_{spam} and C_{th} on the black line, with the probabilities of successful events for different C_{th} values by the grey line on the logarithmic scale. By selecting $C_{th} = 18$ as an example, we can reduce e_{spam} to 3% after postselection. Although the probability of successful initialization is about 3% in this case, initialization can be attempted several times until it is successful. The statistical result without postselection is shown in Extended Data Fig. 8.

Efficient spin–photon interface

A dielectric antenna structure optimizes the photon emission of the quantum emitter for efficient free-space collection. On the basis of the average PSB photon read-out counts from the bright state of the SnV^- in Fig. 4b, we estimate that the lower bound of P_{det} is above 2.4×10^{-3} (Methods), surpassing the previously reported P_{det} of 4×10^{-4} (refs. 8,15,26). This higher collection efficiency indicates the potential for improved entanglement rates in future experiments. The incorporation of a nanocavity Purcell effect by the resonant dielectric antenna can further enhance the ZPL photon emission rate. With an extract $F_p = 2.9$ for a typical SnV^-, the potential improvement in the ZPL photon collection rate is significant through such an efficient spin–photon interface (Methods). Enhanced fabrication and alignment of the location and orientation of the emitter can further improve the F_p, as simulations have shown a quality factor ten times higher compared to current devices[27,28].

Vision for QSoC to enable large-scale fully connected qubit graph

Figure 5a presents our vision for the construction of a connected qubit graph using experimental single FOV data at 11 frequency channels (k_{fch} from 1 to 11). In Fig. 5a, an expanded version of Fig. 4a is presented, encompassing all 11 frequency channels. Similar to the previous figure, bright emitters are encircled, with the circle's colour denoting the tuning voltage best suited for aligning the emitter with its resonant wavelength. To determine the positions of quantum emitters, we assess whether the centres of two quantum emitters, within their respective frequency channels, are within one pixel of each other in the wide-field characterization. Furthermore, if their difference in peak brightness is less than 50%, we group them as the same quantum emitters at varying backplane voltages. Emitters meeting these criteria are tuneable across several frequency channels, and their connections are marked in the graph. In a FOV, the emitters have all-to-all connectivity within frequency channels with optical routing[29–31], and some of the emitters can be tuned to connect the neighbour frequency channels. Figure 5b presents the quantum circuit of equivalence with the cluster in Fig. 5a, assuming each frequency channel k_{fch}, it has $m_{k_{fch}}$ quantum emitters. A desired quantum algorithm can be compiled for the connectivity of the system. The red line labelled here corresponds to the red line in Fig. 5a here. Figure 5c indicates the ratio of the fully connected qubit graph to the total number of qubit nodes (p_c) as a function of the average frequency tuneability ratio $(\overline{\Delta\nu_m}/\nu_{inh})$ across the entire inhomogeneous range (ν_{inh}) under different sizes of the qubit system (n_{qubit}) (Methods). This plot shows that for larger system scales, the tuneability requirement for achieving a fully connected qubit graph is reduced, allowing the system to operate with lower tuning voltages for better energy efficiency. A fully connected qubit graph allows for the implementation of any error correction code with finite degree and weight. In particular, our optical reconfigurability allows for non-planar operation, opening the possibility of using recently

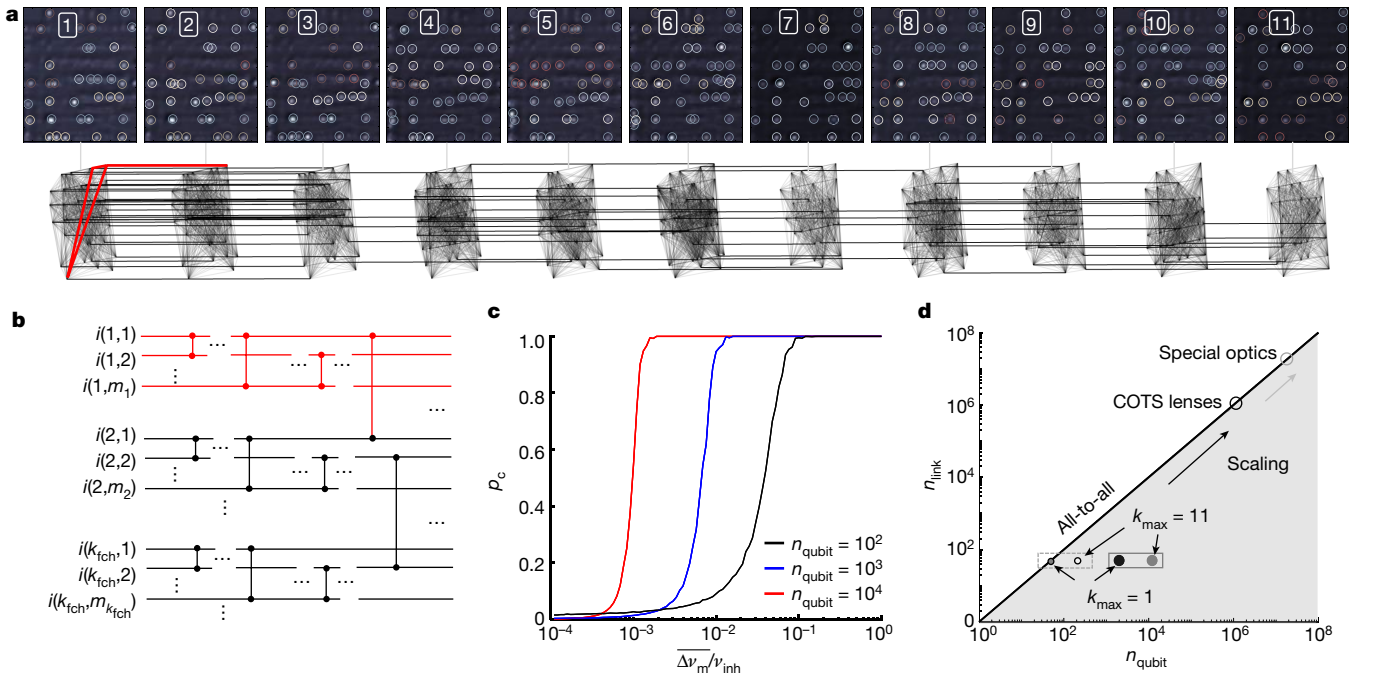


Fig. 5 | QSoC enables large-scale fully connected qubit graph with further scaling. a, Wide-field images of the emitter spot region across 11 frequency channels (k_{fch} from 1 to 11) as well as the illustration of connected cluster using the wide-field data from a single FOV. **b**, The quantum circuit representations of the connected qubit graph. **c**, The ratio of the fully connected qubit graph p_c against $\overline{\Delta\nu_m}/\nu_{inh}$ under various n_{qubit}. **d**, The scaling potential of the QSoC platform considering n_{qubit} and n_{link}. The shading indicates the potential scaling of the corresponding system with the all-to-all connectivity line.

developed codes such as hyperproduct code[32] or other quantum low-density parity-checking codes[33] with a higher encoding rate. For practical applications, we consider the surface code[3,34,35] (shown in Fig. 1a) for its high error threshold and performances in decoding and magic state distillation[36], as well as hashing-bound saturation in handling biased errors[37].

Further scaling

The QSoC module presented here highlighted the advantages of scalability in terms of qubit numbers and connectivity. Connectivity refers to the number of distinct qubits that a single qubit could be coupled to within an entanglement trial, using optical heralding. Figure 5d illustrates the scaling potential of the QSoC platform, considering the number of qubits (n_{qubit}) and the number of direct connections per qubit (n_{link}). The black line, $n_{link} = n_{qubit}$, corresponds to all-to-all connectivity, and the shaded region below represents the physically accessible region on the hardware. The light grey box, including the smaller scattered points, represents the data from a single FOV, whereas the dark grey box, including the larger scattered points, represents the data from the entire sample area. Two grey boxes include the results of a single frequency channel ($k_{max} = 1$) and 11 frequency channels ($k_{max} = 11$).

To achieve further scaling, a larger FOV objective can be used (for example, a commercial ×10 Olympus plan objective with 0.3 NA and a 2.65 mm diameter FOV). Coupled with a denser QMC design (2.52 µm spacing for diffraction-limited spots at SnV^- ZPL with the given objective NA), this imaging system would enable up to 8.7×10^5 directly resolvable spots using commercial-off-the-shelf lenses, as shown in Fig. 5. When a custom-designed lens is used, the number of diffraction-limited spot sites can be expanded to more than ten million, with each spot capable of hosting hundreds of quantum emitters. This level of qubit density of the QSoC is close to the transistor density of the most advanced semiconductor processes (3×10^8 transistors per mm^2 in TSMC N3 process[38]). With a broader inhomogeneous distribution range, our qubit density will not be limited by the transistor density, although the transistor density would determine the number of individual voltages the system can apply simultaneously.

Conclusion

In summary, we have demonstrated the fabrication and characterization of the core QSoC module through a lock-and-release transfer method for high-yield, large-scale integration of artificial atom arrays with CMOS chips, along with a high-throughput characterization approach. The summarized parameters of the characterized QSoC module are presented in Table 1.

The QSoC module shows significant advantages in terms of the parameters n_{sys} and $n_{emitter}$ parameters. Other parameters, such as e_{spam}, P_{det} and F_p can be further improved through refined material processes and control sequences. The architecture can be expanded by incorporating a superconducting nanowire for efficient single-photon detection[39], large-scale CMOS chip control for low-latency solid-state spin control[13,40], reconfigurable qubit connectivity[41] and heralded spin entanglement[42].

This architecture applies to other promising solid-state quantum memory platforms. Particularly suitable for heterogeneous integration would be the thin film Si waveguide and a cavity containing isolated colour centres[43,44], focused-ion-beam-fabricated yttrium orthovanadate crystal comprising rare-earth ion[45], thin film SiC membrane with

Table 1 | QSoC parameters

Parameter	Reported work	This work
n_{sys}	128 (ref. 10)	1,024
$n_{emitter}$	1 (ref. 10)	2.3
e_{spam} (SnV^-)	25% (ref. 25)	3%
P_{det} (F_p)	4×10^{-4} (ref. 8) (1)	2.4×10^{-3} (2.9)

4H–C colour centres[46,47], semiconductor quantum dots in a film[48] and other emerging material platforms[49,50].

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-024-07371-7.

1. DiVincenzo, D. P. The physical implementation of quantum computation. *Fortschr. Phys.* **48**, 771–783 (2000).
2. Bhaskar, M. K. et al. Experimental demonstration of memory-enhanced quantum communication. *Nature* **580**, 60–64 (2020).
3. Choi, H., Pant, M., Guha, S. & Englund, D. Percolation-based architecture for cluster state creation using photon-mediated entanglement between atomic memories. *npj Quantum Inf.* **5**, 104 (2019).
4. Nickerson, N. H., Fitzsimons, J. F. & Benjamin, S. C. Freely scalable quantum technologies using cells of 5-to-50 qubits with very lossy and noisy photonic links. *Phys. Rev. X* **4**, 041041 (2014).
5. Nemoto, K. et al. Photonic architecture for scalable quantum information processing in diamond. *Phys. Rev. X* **4**, 031022 (2014).
6. Monroe, C. et al. Large-scale modular quantum-computer architecture with atomic memory and photonic interconnects. *Phys. Rev. A* **89**, 022317 (2014).
7. Welte, S., Hacker, B., Daiss, S., Ritter, S. & Rempe, G. Photon-mediated quantum gate between two neutral atoms in an optical cavity. *Phys. Rev. X* **8**, 011018 (2018).
8. Humphreys, P. C. et al. Deterministic delivery of remote entanglement on a quantum network. *Nature* **558**, 268–273 (2018).
9. Bradley, C. E. et al. A ten-qubit solid-state spin register with quantum memory up to one minute. *Phys. Rev. X* **9**, 031045 (2019).
10. Wan, N. H. et al. Large-scale integration of artificial atoms in hybrid photonic circuits. *Nature* **583**, 226–231 (2020).
11. Xue, X. et al. CMOS-based cryogenic control of silicon quantum circuits. *Nature* **593**, 205–210 (2021).
12. Bardin, J. C. et al. Design and characterization of a 28-nm bulk-CMOS cryogenic quantum controller dissipating less than 2 MW at 3K. *IEEE J. Solid-State Circuits* **54**, 3043–3060 (2019).
13. Kim, D. et al. A CMOS-integrated quantum sensor based on nitrogen–vacancy centres. *Nat. Electron.* **2**, 284–289 (2019).
14. Ibrahim, M. I., Foy, C., Englund, D. R. & Han, R. High-scalability CMOS quantum magnetometer with spin-state excitation and detection of diamond color centers. *IEEE J. Solid-State Circuits* **56**, 1001–1014 (2021).
15. Bernien, H. et al. Heralded entanglement between solid-state qubits separated by three metres. *Nature* **497**, 86–90 (2013).
16. Iwasaki, T. et al. Tin-vacancy quantum emitters in diamond. *Phys. Rev. Lett.* **119**, 253601 (2017).
17. Trusheim, M. E. et al. Transform-limited photons from a coherent tin-vacancy spin in diamond. *Phys. Rev. Lett.* **124**, 023602 (2020).
18. Pauka, S. et al. A cryogenic CMOS chip for generating control signals for multiple qubits. *Nat. Electron.* **4**, 64–70 (2021).
19. Sutula, M. et al. Large-scale optical characterization of solid-state quantum emitters. *Nat. Mater.* **22**, 1338–1344 (2023).
20. Golter, D. A. et al. Selective and scalable control of spin quantum memories in a photonic circuit. *Nano Lett.* **23**, 7852–7858 (2023).
21. Debroux, R. et al. Quantum control of the tin-vacancy spin qubit in diamond. *Phys. Rev. X* **11**, 041041 (2021).
22. Aghaeimeibodi, S., Riedel, D., Rugar, A. E., Dory, C. & Vučković, J. Electrical tuning of tin-vacancy centers in diamond. *Phys. Rev. Appl.* **15**, 064010 (2021).
23. Bersin, E. et al. Individual control and readout of qubits in a sub-diffraction volume. *npj Quantum Inf.* **5**, 38 (2019).
24. Meesala, S. et al. Strain engineering of the silicon-vacancy center in diamond. *Phys. Rev. B* **97**, 205444 (2018).
25. Görlitz, J. et al. Coherence of a charge stabilised tin-vacancy spin in diamond. *npj Quantum Inf.* **8**, 45 (2022).
26. Pompili, M. et al. Realization of a multinode quantum network of remote solid-state qubits. *Science* **372**, 259–264 (2021).
27. Rugar, A. E. et al. Quantum photonic interface for tin-vacancy centers in diamond. *Phys. Rev. X* **11**, 031021 (2021).
28. Kuruma, K. et al. Coupling of a single tin-vacancy center to a photonic crystal cavity in diamond. *Appl. Phys. Lett.* **118**, 230601 (2021).
29. Christen, I. et al. An integrated photonic engine for programmable atomic control. Preprint at https://arxiv.org/abs/2208.06732 (2022).
30. Kim, J. et al. 1100 x 1100 port mems-based optical crossconnect with 4-db maximum loss. *IEEE Photonics Technol. Lett.* **15**, 1537–1539 (2003).
31. Palm, K. J. et al. Modular chip-integrated photonic control of artificial atoms in diamond waveguides. *Optica* **10**, 634–641 (2023).
32. Tillich, J.-P. & Zémor, G. Quantum ldpc codes with positive rate and minimum distance proportional to the square root of the blocklength. *IEEE Trans. Inform. Theory* **60**, 1193–1202 (2013).
33. Breuckmann, N. P. & Eberhardt, J. N. Quantum low-density parity-check codes. *PRX Quantum* **2**, 040101 (2021).
34. Kitaev, A. Y. in *Quantum Communication, Computing, and Measurement* (eds Hirota, O. et al.) 181–188 (Springer, 1997).
35. Fowler, A. G., Whiteside, A. C. & Hollenberg, L. C. Towards practical classical processing for the surface code. *Phys. Rev. Lett.* **108**, 180501 (2012).
36. Litinski, D. Magic state distillation: not as costly as you think. *Quantum* **3**, 205 (2019).
37. Bonilla Ataides, J. P., Tuckett, D. K., Bartlett, S. D., Flammia, S. T. & Brown, B. J. The XZZX surface code. *Nat. Commun.* **12**, 2172 (2021).
38. Wu, S.-Y. et al. A 3 nm CMOS FinFlex™ platform technology with enhanced power efficiency and performance for mobile SoC and high performance computing applications. *International Electron Devices Meeting*, 27.5.1–27.5.4 (2022).
39. Zhu, D. et al. Superconducting nanowire single-photon detector on aluminum nitride. In *Conference on Lasers and Electro-Optics* FTu4C.1 (Optical Society of America, 2016).
40. Xue, X. et al. Quantum logic with spin qubits crossing the surface code threshold. *Nature* **601**, 343–347 (2022).
41. Gyger, S. et al. Reconfigurable photonics with on-chip single-photon detectors. *Nat. Commun.* **12**, 1408 (2021).
42. Lomonte, E. et al. Single-photon detection and cryogenic reconfigurability in lithium niobate nanophotonic circuits. *Nat. Commun.* **12**, 6847 (2021).
43. Prabhu, M. et al. Individually addressable and spectrally programmable artificial atoms in silicon photonics. *Nat. Commun.* **14**, 2380 (2023).
44. Higginbottom, D. B. et al. Optical observation of single spins in silicon. *Nature* **607**, 266–270 (2022).
45. Ruskuc, A., Wu, C.-J., Rochman, J., Choi, J. & Faraon, A. Nuclear spin-wave quantum register for a solid-state qubit. *Nature* **602**, 408–413 (2022).
46. Crook, A. L. et al. Purcell enhancement of a single silicon carbide color center with coherent spin control. *Nano Lett.* **20**, 3427–3434 (2020).
47. Lukin, D. M. et al. 4h-silicon-carbide-on-insulator for integrated quantum and nonlinear photonics. *Nat. Photon.* **14**, 330–334 (2020).
48. Barik, S. et al. A topological quantum optics interface. *Science* **359**, 666–668 (2018).
49. Kim, J.-H., Aghaeimeibodi, S., Carolan, J., Englund, D. & Waks, E. Hybrid integration methods for on-chip quantum photonics. *Optica* **7**, 291–308 (2020).
50. Elshaari, A. W., Pernice, W., Srinivasan, K., Benson, O. & Zwiller, V. Hybrid integrated quantum photonic circuits. *Nat. Photon.* **14**, 285–298 (2020).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2024

Methods

Diamond colour centre creation and fabrication

We etched the first 10 μm of the diamond (surface orientation [001]) using plasma etching in Ar or Cl_2 followed by the O_2 etching to relieve the strained surface[10]. Then we performed a plane ion implantation of ^{120}Sn in Innovion at an effective areal dose of 5×10^{11} ions per cm^2 at 350 keV. The implantation target depth is 86 nm, with a longitudinal straggle of 11.7 nm and a lateral straggle of 10.6 nm, estimated from the simulations of stopping and range of ions in matter simulations[51]. After implantation, the diamond was annealed at 1,200 °C in an ultrahigh vacuum furnace 10^{-7} mbar, followed by a boiling mixture of 1:1:1 sulfuric acid, nitric acid and perchloric acid cleaning process. The implantation was at a high dose to guarantee more than one quantum emitter in each quantum channel on average. We estimated a SnV^- creation efficiency of roughly 12 ± 5% in this process, on the basis of the spectrum characterization of the diamond pillar array following diamond fabrication. This implantation dose density yields an average of 3 SnV^- per resonant dielectric antenna within the mode volume. After the generation of SnV^- in the diamond, we deposited 180 nm of silicon nitride (Si_3N_4) followed by electron beam lithography and reactive ion etching (RIE) to create a hard mask for the QMC design. We etched Si_3N_4 mask into the diamond layer by O_2 RIE, followed by an 18 nm of conformal alumina atomic layer deposition. After a breakthrough of alumina with CF_4 RIE etching, we used oxygen plasma to isotropically undercut the diamond QMC. The Si_3N_4 and alumina were removed in hydrofluoric acid[52].

QMC spin–photon interface design

A perturbative cavity design was implemented in each quantum channel as a dielectric antenna optimized for free-space collection. These diamond cavities were connected in groups of 16 by means of mechanical trusses to parallelize the subsequent integration process. The perturbative cavity's collection efficiency η was simulated and optimized with finite-difference time-domain simulation in Lumerical[53]. The final cavity design features a diamond waveguide with a width of 280 nm and thickness of 200 nm, periodically patterned with holes of 64 nm radius and a 202 nm spacing (a). The defect supporting the cavity mode is introduced by the Gaussian modulation $x(m) = d_0 + a \times \left[1 - \exp\left(-\frac{m^2}{2\sigma^2}\right)\right]$, where $x(m)$ denotes the distance of the mth holes on one side from the centre of the cavity, d_0 is the distance from the centre of the first hole to the centre of the cavity and σ is a dimensionless number that controls the distribution of the location of the holes. The optimized parameters are $d_0 = 113$ nm and $\sigma = 4.6$. Half-round perturbation bumps are introduced along the cavity region as a grating to improve the free-space collection. The design of the dielectric resonant antenna achieved both a quality factor of 2,000 and a collection efficiency of $\eta = 96\%$ with an NA of 0.9, assuming a perfect metal reflector with a distance of $\Delta z = 250$ nm below the diamond cavity.

CMOS postprocessing

We postprocessed the CMOS chip with several photolithography steps as shown in Extended Data Fig. 1. We started from a bare die of the CMOS chip, and did the first photolithography to define the chiplet locking structure. We removed the fluorescent surface layer by CF_4 RIE and performed photolithography followed by RIE etching of the spacer material to form the QMC locking structure. We applied a combination of iron chloride, ammonium hydroxide and hydrogen peroxide to wet etch the top metal routing layer, barrier layer and via, respectively, forming the QMC locking structure made of oxide. The key feature for the QMC locking structure was it was slightly larger than the QMC's edge, so it served as the nanostructure that restricted the QMC movement during the lock and release. We then performed a second photolithography and RIE etching, forming the QMC platform region. We did the third lithography for the subsequent etching to expose the wire-bonding region. After QMC transfer, we performed extra photolithography followed by e-beam evaporation of a 50 nm Au–5 nm Ti metal layer[10]. This metal layer provided a better bonding interface on the CMOS copper and allowed the capacitive strain tuning of the quantum emitters[54]. Extended Data Fig. 2 shows the surface of the postprocessed CMOS chip, featuring the QMC locking structure, which facilitates the large-scale transfer process. The magnetic field distribution on QSoC under global microwave control is also shown, in which the peak magnetic field will appear right on top of the CMOS backplane lines.

Heterogeneous integration setup and process

Extended Data Fig. 3 shows large-scale heterogeneous integration setup and the changes in the cross-section of the large-scale heterogeneous integration process (visit ref. 55 for a 3D and cross-sectional GIF animation). As depicted in Fig. 2b, we aligned the parent diamond towards the CMOS chip using motorized control and moved down the diamond, allowing the QMC array to lock to the postprocessed CMOS surface microstructure. We then horizontally moved the parent diamond using the piezo control to release the QMC array. The fabricated QMC locking structure restricted relative movement between the QMC and the parent diamond, breaking the weak hinge (about 150 nm wide) between the QMC and the parent diamond in parallel. This method accelerated the heterogeneous integration process by avoiding breaking and picking an individual chiplet. The QMC transfer yield reached 100% after the fine-tuning process adjusted with probes for better alignment with the bottom metal. Before fine-tuning, the yield for successfully separating the bridges between the QMCs and the bulk diamond was about 70% per release attempt. However, we could perform several releases until all bridges were effectively separated. Roughly 50% of the QMCs required posttransfer alignment adjustments using probes for such a socket design, a process that was significantly faster than the individual transfer steps. Here, we created an etched locking socket that was 3 μm larger than the diamond QMC in both length and width. A larger QMC design or a finer resolution of socket fabrication reduced the requirement for fine-tuning.

Experiment setup

Our experimental setup is shown in Extended Data Fig. 9. Our cryogenic system used a closed-cycle helium cryostat with a base temperature of 4 K (Montana Instruments), along with a custom cryogenic microscope objective (×100, a NA of 0.9). Within the cryogenic system, we used three-axis nano position steppers (Attocube ANP-x,z-50) to move the sample, which was glued to a custom cryogenic printed circuit board. A programmable voltage source (Keithley 2400) was used to apply the voltage to the printed circuit board. The collected signal was directed to fibres[56], single-photon counting modules in free space or the duo camera system, composed of a polarized beam splitter connected to both a scientific CMOS camera (Thorlabs, catalogue no. CS235U) and an EMCCD (Hamamatsu ImagEM X2) with a long-pass filter. The tuneable resonant laser (MSquared SolsTis with an external mixing module) was modulated with an acoustic-optic modulator (AOM) for pulsing and an electrical-optic modulator (EOM) for frequency sideband tuning. A 515 nm cobalt laser was used to off-resonantly repump the SnV^-. We used two different signal generators (Rohde & Schwarz, catalogue nos. SMIQ03 and SMIQ06) with a fast radio frequency switch (Mini-Circuits, catalogue no. ZASWA-2-50DR+) to control the modulation signal fed to the EOM. The signal from the signal generator was amplified using a high-power amplifier (Mini-Circuits, catalogue no. ZHL-16W-43-S+) operating in the frequency range 1.8 to 4 GHz, which was within the EOM PLE range shown in Fig. 4b. Controlling the radio frequency switch allows us to switch between the $R_\downarrow$ and $R_\uparrow$ channels, with individual radio frequency amplitudes controlled by two different signal generators. The AOM simultaneously turned off the $R_\downarrow$ and $R_\uparrow$ channels. In each sample measurement, we had three APD read-out time bins. APD bin 1

was positioned when the laser was in $R_\downarrow$ channel while during APD bin 2 the laser was also the same. The delay between APD bins 1 and 2 was 50 µs. APD bin 3 had another delay time 50 µs of from APD bin 2, during which the laser was also in the $R_\uparrow$ channel. The lifetime measurements were performed using a pulse pattern generator (Anritsu, catalogue no. MP1763B), in which the APD signal (SPCM-AQRH-16) and the trigger signal of the pulse pattern generator were connected to the PicoHarp 300 for data analysis. Confocal collection fluorescence in free space was filtered using an long-pass filter (Thorlabs, catalogue no. FEL0650). In the lifetime measurement, the laser was modulated by an EOM (EOSpace) with a digital Mach–Zehnder modulator bias controller (iXblue). We biased the Mach–Zehnder modulator to allow a minimum light output and connected the radio frequency control port of the amplitude modulator to the pulse pattern generator. The pulse pattern generator generated a 2 V pulse that turned on the laser within 500 ps providing pulse laser excitation. The pulse trigger signal was analysed in conjunction with the APD signal received in the PicoHarp 300 for signal processing.

Gas tuning and Purcell enhancement

The inhomogeneous cavity fabrication variation was overcome by using a gas-tuning method[27,28]. We used a supercontinuum laser (SuperK) to monitor the QMC cavity spectrum during the gas-tuning process as shown in Extended Data Fig. 4. Gas deposition was a global redshift process for cavity resonant, whereas pulsed high-power laser-assisted desorption was an individual opposite control process for blue-shift. The gas-tuned cavity remained stable in a cryogenic environment without significant temperature fluctuations[2]. As depicted in Extended Data Fig. 5, we simulated Purcell factor and the far-field distribution of the electromagnetic field for an emitter optically coupled to the resonant dielectric antenna mode. The experimental quality factor of $Q = 220$ remained unchanged before and after gas deposition[57]. We observed a tuning range $\Delta_{gas} > 10$ nm, and this tuning method could accommodate up to 36 nm without reducing the quality factor[58,59]. We compared the impact of spectrally tuning the nanophotonic resonance on the lifetime of the emitter to calculate the Purcell factor $F_p = \left(\frac{\tau_{bulk}}{\tau_{on}} - \frac{\tau_{bulk}}{\tau_{off}}\right)/\xi_{ZPL} = 2.9$. The ξ_{ZPL} was about 0.36, representing the fraction of the total emission in the strongest of the two ZPL transitions visible at the temperature of 4 K for SnV^- (ref. 28).

Strain tuning

Extended Data Fig. 7 presents the result of the simulation of stationary electromechanics conducted using COMSOL Multiphysics. The simulation structure comprised a single diamond waveguide with holes and bumps, as described in the paper, an electrode on the top right providing voltage and the oxide platform region on the CMOS chip to avoid severe deformation and grounding at the base of the substrates. We applied several constraints and boundary conditions to the stationary solver. In the solid mechanics' portion, the diamond body was in the free deformation state, whereas the surfaces on both ends, which connected to the CMOS chip in reality, are fixed. We fixed the intersecting surface between the oxide and the ground. For the electrostatics portion, we applied charge conservation to the entire object and added terminal and ground boundary conditions to the electrode and the ground, respectively. The simulation results indicated the feasibility of strain tuning through the varying voltage applied to the diamond.

State preparation and measurement error

The preparation and measurement error e_{spam} was calculated on the basis of the histogram result $p_{dark}(n)$ and $p_{bright}(n)$, in which we choose an optimum measurement threshold n_m to minimize $e_{spam} = \min(0.5(\Sigma_{n=n_m}^{\max(n)} p_{dark}(n) + \Sigma_{n=0}^{n_m-1} p_{bright}(n)))$. The e_{spam} accounted for errors due to the bright state reading as dark or the dark-state reading as bright during preparation and measurement. The average value of $p_{dark}(n)$ was 1.6 whereas the average value of $p_{bright}(n)$ was 18. We fitted the histogram with the Poisson distribution as shown in Fig. 4b.

Extended Data Fig. 8a shows the pulse control sequence, whereas Extended Data Fig. 8b–d presents the histogram of the three APD time bin read-out results in Fig. 4a, along with a maximum likelihood estimation of mixture Poisson distribution fitting (black curve), $P(n) = (1-P_0)\frac{\lambda_1^n \exp(-\lambda_1)}{n!} + P_0\frac{\lambda_2^n \exp(-\lambda_2)}{n!}$. Here, P_0 is the fitted probability between two distributions, whereas λ_1 and λ_2 represent the average count numbers of the bright- and dark-state events, respectively. The fitted individual Poisson distributions for the bright and dark states are plotted separately in red and blue. We estimated the read-out threshold count n_m to minimize the read-out error by satisfying $(1-P_0)\lambda_1^{n_m}\exp(-\lambda_1) = P_0\lambda_2^{n_m}\exp(-\lambda_2)$. Bin 1 and bin 3 are fitted to the mixture Poisson distribution, whereas bin 2 is fitted to a single Poisson distribution. The fitted n_m is 3.5 here. On the basis of the relationship between the APD count and n_m, the correlation was divided into four quadrants, labelled with the colours grey (not counting), red (success), blue (error) and magenta (error). We repeated the sequence 10,000 times, and obtained 172 error events and 1,508 success events in the red region, indicating a preparation and measurement error of 10.2% without postselection of the first APD bin count.

Coherent photon detection efficiency P_{det} calculation

To estimate the detection efficiency, we considered the SnV^- in bulk. The quantum efficiency of SnV^- emission is 80%, and four ZPL lines represent 57% of the energy of the emitted photon. The spin-conserved C line is the bright line representing 80% of the emission energy among all ZPL lines[27]. The PSB of the emission is 43%, and the PSB after applying a 637 nm long-pass filter is 35%. Consequently, the intensity ratio between the ZPL C line emission and PSB after the 637 nm long-pass filter is roughly 1.3. In our measurement in Fig. 4c, we observed an average of 18 counts in PSB after 633 nm long-pass filter within 50 µs read-out time. The SnV^-'s lifetime is about 5 ns so its maximum emitted $ph_{total} = 10{,}000$ photons in our read-out time. We detect ZPL C line photons $ph_{ZPL} = 24$ with our 65% quantum efficiency APD. Therefore, we estimated our $P_{det} = ph_{ZPL}/ph_{total} > 2.4 \times 10^{-3}$. With a higher efficiency detector, such as superconducting nanowire single-photon detectors[39], we can get a $P_{det} > 3.5 \times 10^{-3}$.

Emitter number per quantum channel calculation

The $n_{emitter}$ is the average number of resonant quantum emitters per frequency channel within a quantum channel. We summed up all spatially resolvable spots resonating with each channel while tuning the emitter frequency to estimate the total number of quantum emitters across the 11 frequency channels, in which we identified $n_{spot} = 52{,}322$ spots[55]. Taking into account that each SnV^- is counted twice due to its two spin transition tuning across neighbour frequency channels on average, the total number of distinct spots $n'_{spot} \approx n_{spot}/2 = 26{,}161$. Dividing this by the 11 frequency channels (k_{max}), we got an average of $n'_{spot}/k_{max} = 2{,}378$ qubits per frequency channel. Furthermore, dividing this by n_{sys} gave an average of $n_{emitter} = n'_{spot}/(k_{max}n_{sys}) = 2.3$ resonant quantum emitters per quantum channel per nanobeam. To estimate the size of the connected qubit graph, we used the formula: $n_{qubit} = n_{emitter}n_{sys}p_c k_{max}$. Here, p_c represents the proportion of the fully connected graph relative to the total number of nodes, a relationship graphically illustrated in Fig. 5c.

Fully connected qubit number ratio versus system tuneability

To estimate the fully connected qubit number ratio p_c in Fig. 5, we performed the Monte Carlo simulation with a uniform distributed n_{qubit} random emitter zero bias frequency in the inhomogeneous range v_{inh} from 0 to 100%. We assumed a zero-mean Gaussian distribution of maximum tuneability for each individual emitter, and fitted the standard deviation to match the experimental sampled statistics. With a dynamic programming calculation counting the number of the

quantum emitter that can be connected within all the possible emitters' tuning ranges, we can estimate the number of connected node n_{con} among n_{qubit} samples we have. This value is an average of ten times in the Monte Carlo simulation. The tuneability ratio was sweep for getting the Fig. 5b curve of p_c against $\overline{\Delta\nu_m}/\nu_{inh}$ under various n_{qubit}. Although we used 11 frequency channels to cover a 20 GHz emitter spectral range, ν_{inh} can reach a 150 GHz for SnV^-, which makes our $\overline{\Delta\nu_m}/\nu_{inh}$ reach 1.3%. When expanding the spectral range of the emitter, we expect to see much more numbers for n_{qubit} in the same FOV. Various methods can be applied for exciting and routing photons to enable optical interactions between different quantum emitters[60]. An example of collecting two spin-entangled photons into two sub-FOVs is shown in Extended Data Fig. 10.

Architecture modelling

We provide a hardware architecture model, which is accessible at ref. 55. Included in this repository is a Python notebook covering various topics: (1) key parameters for general-purpose quantum computing; (2) architectural entanglement sequence for estimating entanglement rates; (3) spectral diffusion's impact on entanglement rate; (4) effect of spectral diffusion and Purcell enhancement on visibility; (5) strain tuning models for ZPL shift and spin splitting; (6) thermal budget for the architecture's operation and (7) emitter statistical data. This modelling provides a relationship of the figure of merit among various architectural parameter settings.

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request. The data that support the findings of this study are also openly available at https://doi.org/10.6084/m9.figshare.25374583.v1 (ref. 61).

Code availability

Our code and model checkpoint are available at https://doi.org/10.6084/m9.figshare.25374610.v1 (ref. 55).

51. Ziegler, J. F., Ziegler, M. D. & Biersack, J. P. SRIM–the stopping and range of ions in matter (2010). *Nucl. Instrum. Methods Phys. Res. B: Beam Interact. Mater. At.* **268**, 1818–1823 (2010).
52. Mouradian, S., Wan, N. H., Schröder, T. & Englund, D. Rectangular photonic crystal nanobeam cavities in bulk diamond. *Appl. Phys. Lett.* **111**, 021103 (2017).
53. Li, L., Choi, H., Heuck, M. & Englund, D. Field-based design of a resonant dielectric antenna for coherent spin-photon interfaces. *Optics Expr.* **29**, 16469–16476 (2021).
54. Clark, G. et al. Nanoelectromechanical control of spin–photon interfaces in a hybrid quantum system on chip. *Nano Lett.* **24**, 1316-1323 (2024).
55. Li, L. Architecture modeling for 'Heterogeneous integration of spin-photon interfaces with a scalable CMOS platform'. QSoC architecture modeling. figshare https://doi.org/10.6084/m9.figshare.25374610.v1 (2024).
56. Starling, D. J. et al. Fully packaged multichannel cryogenic quantum memory module. *Phys. Rev. Appl.* **19**, 064028 (2023).
57. Evans, R. E. et al. Photon-mediated interactions between quantum emitters in a diamond nanocavity. *Science* **362**, 662–665 (2018).
58. Li, L. et al. Coherent spin control of a nanocavity-enhanced qubit in diamond. *Nat. Commun.* **6**, 6173 (2015).
59. Chen, K. C. et al. A scalable cavity-based spin–photon interface in a photonic integrated circuit. *Optica Quantum* **2**, 124–132 (2024).
60. Martínez, J. A. et al. Photonic indistinguishability of the tin-vacancy center in nanostructured diamond. *Phys. Rev. Lett.* **129**, 173603 (2022).
61. Li, L. Data in 'Heterogeneous integration of spin-photon interfaces with a scalable CMOS platform'. figshare https://doi.org/10.6084/m9.figshare.25374583.v1 (2024).

Acknowledgements This work was supported by the MITRE Corporation Quantum Moonshot Program, the National Science Foundation (NSF) STC Center for Integrated Quantum Materials (grant no. DMR-1231319), the ARO MURI W911NF2110325 and the NSF Engineering Research Center for Quantum Networks (grant no. EEC-1941583). L.L. acknowledges funding from NSF QISE-NET Award (grant no. DMR-1747426). L.D. and C.E.-H. acknowledge funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement nos. 840393 and 896401. I.H. acknowledges the funding support by the NSF RAISE-TAQS (grant no. 1839159). K.C.C. acknowledges funding support by the NSF Graduate Fellowship. M.E.T acknowledges support from the ARL ENIAC Distinguished Postdoctoral Fellowship. H.C. acknowledges the Claude E. Shannon Fellowship. M. I. I. and R. H. acknowledge funding support by the NSF RAISE-TAQS (grant no. 1839159) and MIT MTL Center of Integrated Circuits and Systems. We thank A. Menssen for the discussion of the SLM application. We thank R. Chen for help with the foundry tape out and E. Bersin for useful comments on the manuscript. L.L. thanks S. Wang for the kind support.

Author contributions L.L. conducted device design and fabrication, large-scale heterogeneous integration and measurements of the whole experiment. L.D.S. contributed to the cryogenic measurement setup and the experiment automation code. I.H. performed the tape out of the CMOS chip and contributed to the room-temperature measurement setup. K.C.C. contributed to the diamond fabrication and cryogenic measurement. Y.S., I.C., M.T., Y.H. and Y.G. contributed to the setup of the cryogenic experiment. H.C. contributed to the theoretical analysis of the architecture, C.E.-H. contributed to the transfer print stage for the transfer process, J.D. contributed to the transfer and the scientific plotting. Y.G. performed the COMSOL simulation for the strain tuning. G.C. contributed to characterizing the linewidth of the colour centre to verify the colour centre annealing properties. M.I.I. and R.H. provided an experience for the CMOS tape out. D.E. conceived the QSoC architecture and the frequency-shifted qubit graph algorithms and supervised the whole project. G.G. contributed to quantum algorithms. D.E. and L.L. designed the experiment. L.L. wrote the manuscript with input from all authors.

Competing interests The authors declare no competing interests.

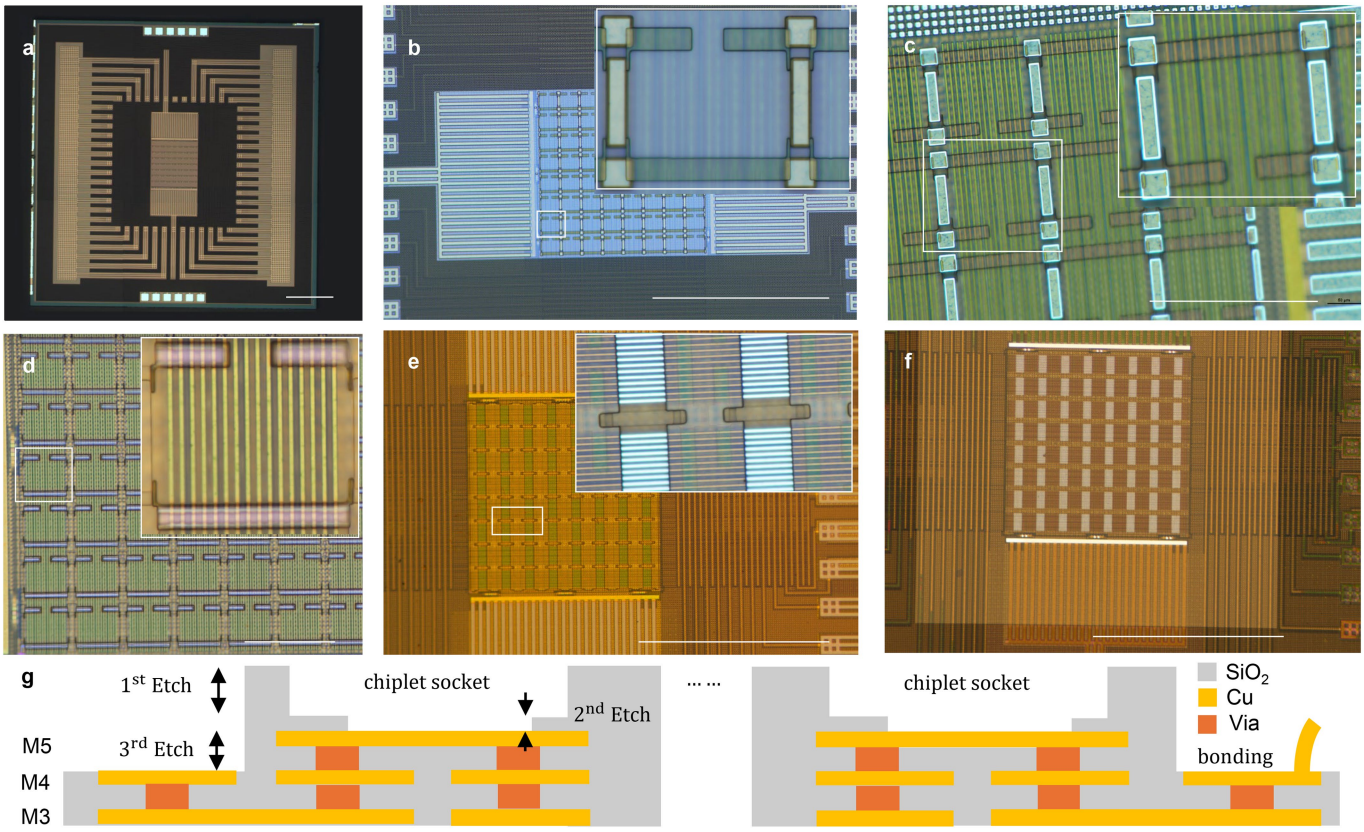
Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-024-07371-7.

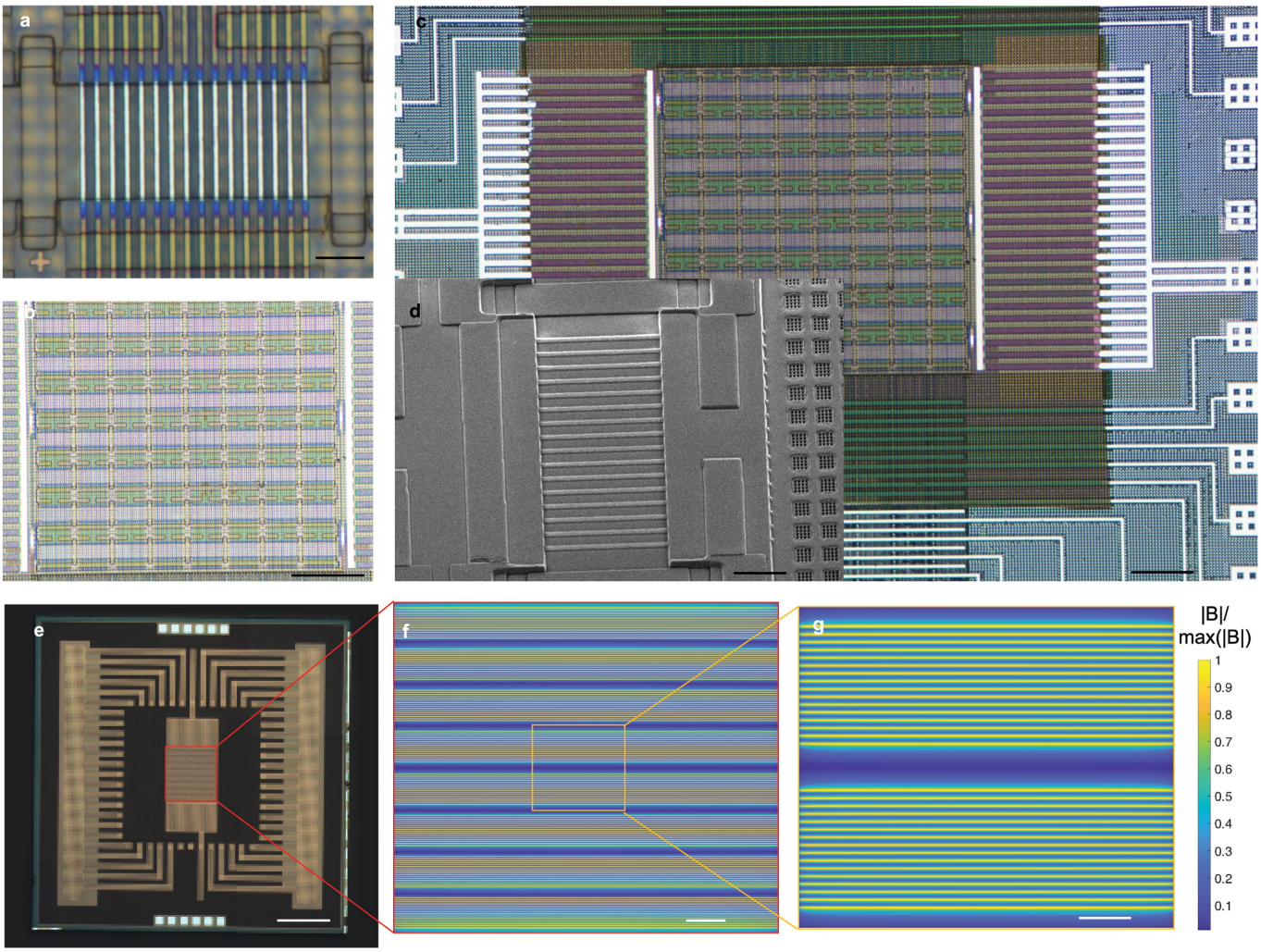
Correspondence and requests for materials should be addressed to Linsen Li or Dirk Englund.

Peer review information *Nature* thanks the anonymous reviewers for their contribution to the peer review of this work.

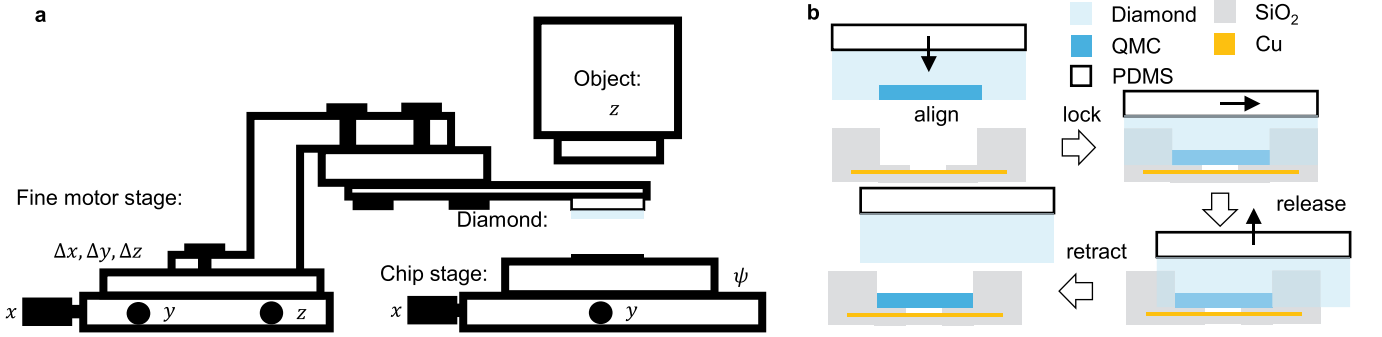
Reprints and permissions information is available at http://www.nature.com/reprints.



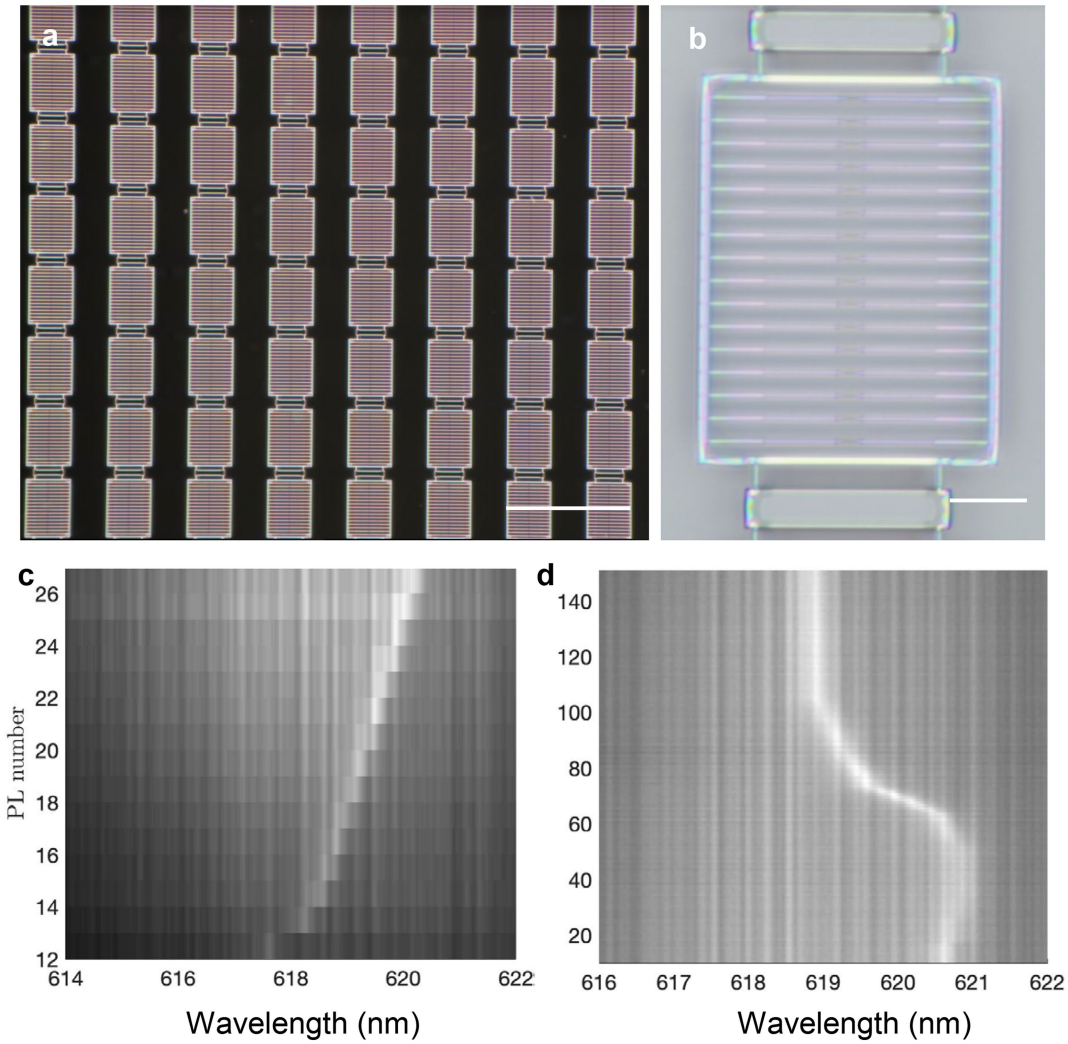
Extended Data Fig. 1 | CMOS post-processing procedure. a, Optical microscope image of a 3 mm × 3 mm bare die CMOS chip from the foundry, with alignment markers for the following process (scale bar 500 µm). **b**, Optical microscope image after the first photolithography (scale bar 500 µm), created the initial photoresist layer defining the locking structure of the chiplet. **c**, Optical microscope image after the first dry etching (scale bar 100 µm). **d**, Optical microscope image after wet etching (scale bar 100 µm), resulting in the QMC oxide locking structure. **e**, Optical microscope image after second photolithography and dry etching (scale bar 500 µm), formed the QMC platform region for QMC placement. **f**, Optical microscope image after third photolithography (scale bar 500 µm), prepared the chip for subsequent etching to expose the internal routing metal layer for the wire bonding region. **g**, Cross-section of the post-processed CMOS chip, with M3-M5 corresponding to metal layers 3 to 5 in the foundry metal definition; this CMOS process has a total of 6 metal layers.



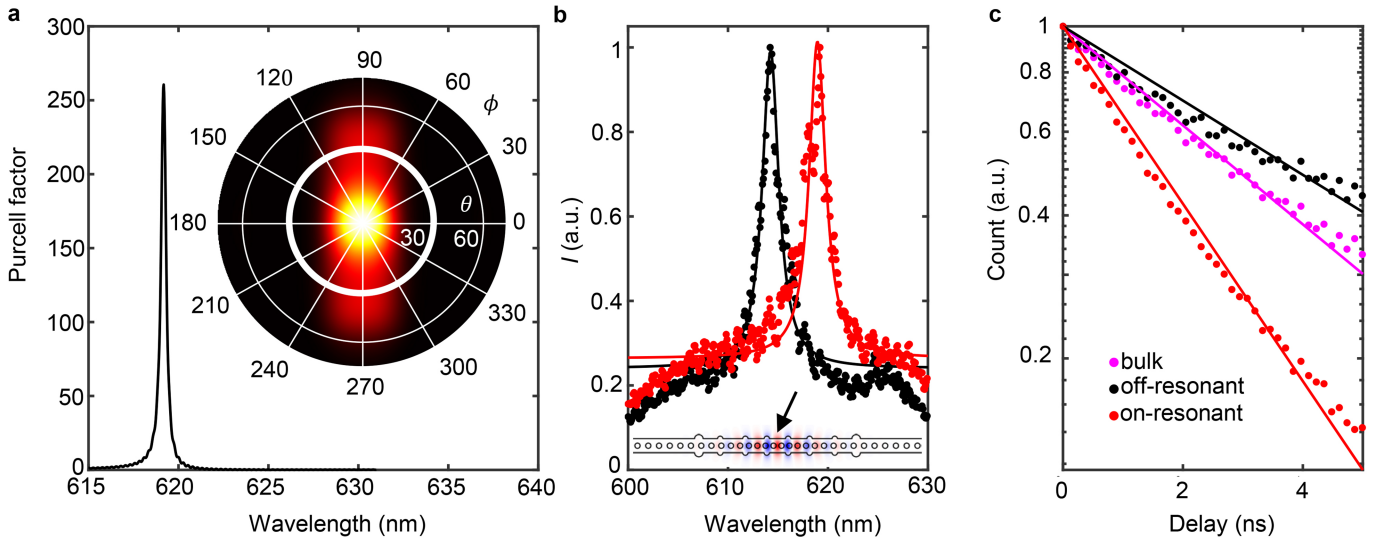
Extended Data Fig. 2 | CMOS post-processing result. a, Optical microscope image of a single post-processed CMOS socket unit (scale bar 10 µm). **b**, Optical microscope image of an 8 × 8 array of post-processed CMOS sockets (scale bar 100 µm). **c**, Optical microscope image of the central region of the post-processed CMOS chip (scale bar 100 µm). **d**, SEM image of the post-processed CMOS socket unit (scale bar 100 µm). **e**, Optical microscope image of a 3 mm × 3 mm bare die CMOS chip from the foundry (scale bar 500 µm). Central red region is the target diamond integration region. **f**, The magnetic field distribution in the QSoC central region (red rectangle region in **e**, where the diamond QMCs locate in (scale bar 50 µm). **g**, The magnetic field distribution of the pink rectangle in **f**, which shows the microwave field of the CMOS backplane (scale bar 20 µm).



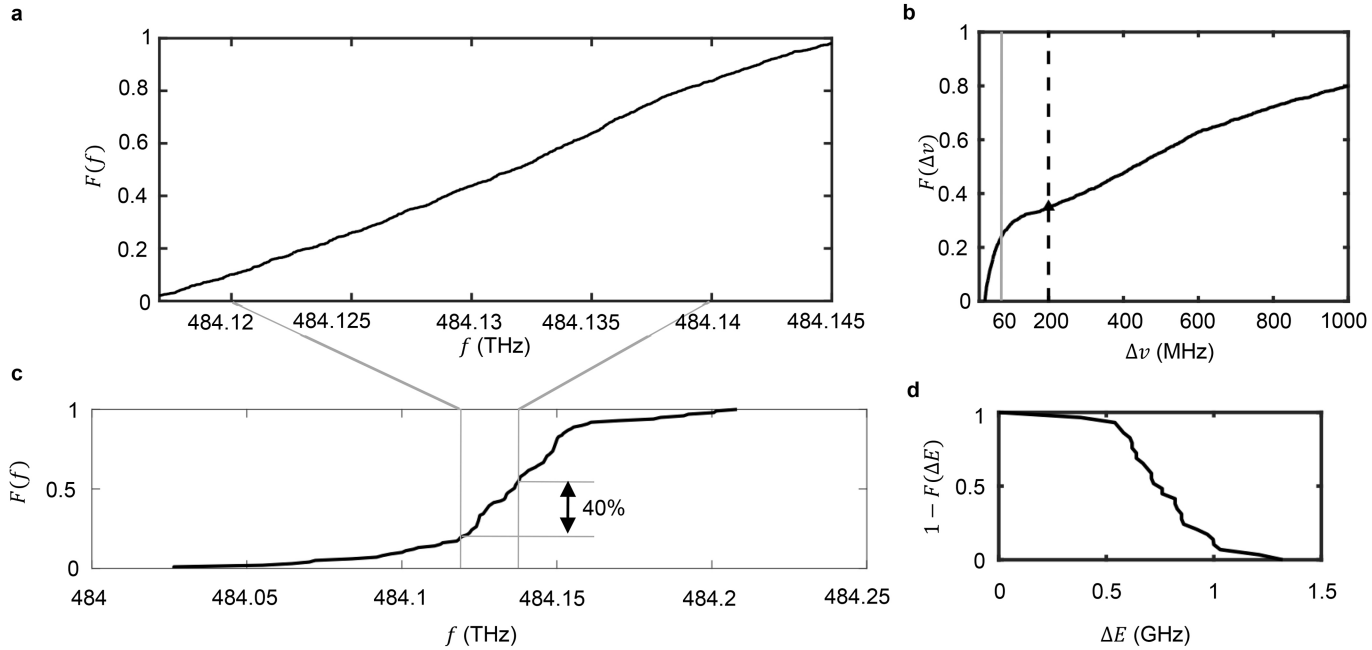
Extended Data Fig. 3 | Detailed of lock-and-release transfer operation for QSoC fabrication. a, Setup for the large-scale heterogeneous integration process between the fabricated parent diamond substrate and the target CMOS chip. The parent diamond was placed on a cut polydimethylsiloxane (PDMS) gel attached to a glass slide. The CMOS chip was mounted on a motorized stage with controllable horizontal (x) and lateral movement (y) as well as rotation (ψ). The diamond-attached glass was screwed to a 3-axis motorized stage with 3-axis piezo control along with the pitch and roll angle control knobs to align the surface of the diamond parallel to the CMOS chip surface. The chiplet alignment can be viewed through a microscope with variable magnification and a long working distance objective. **b**, Cross-section of the unit QMC and its corresponding CMOS locking structure during the transfer process.



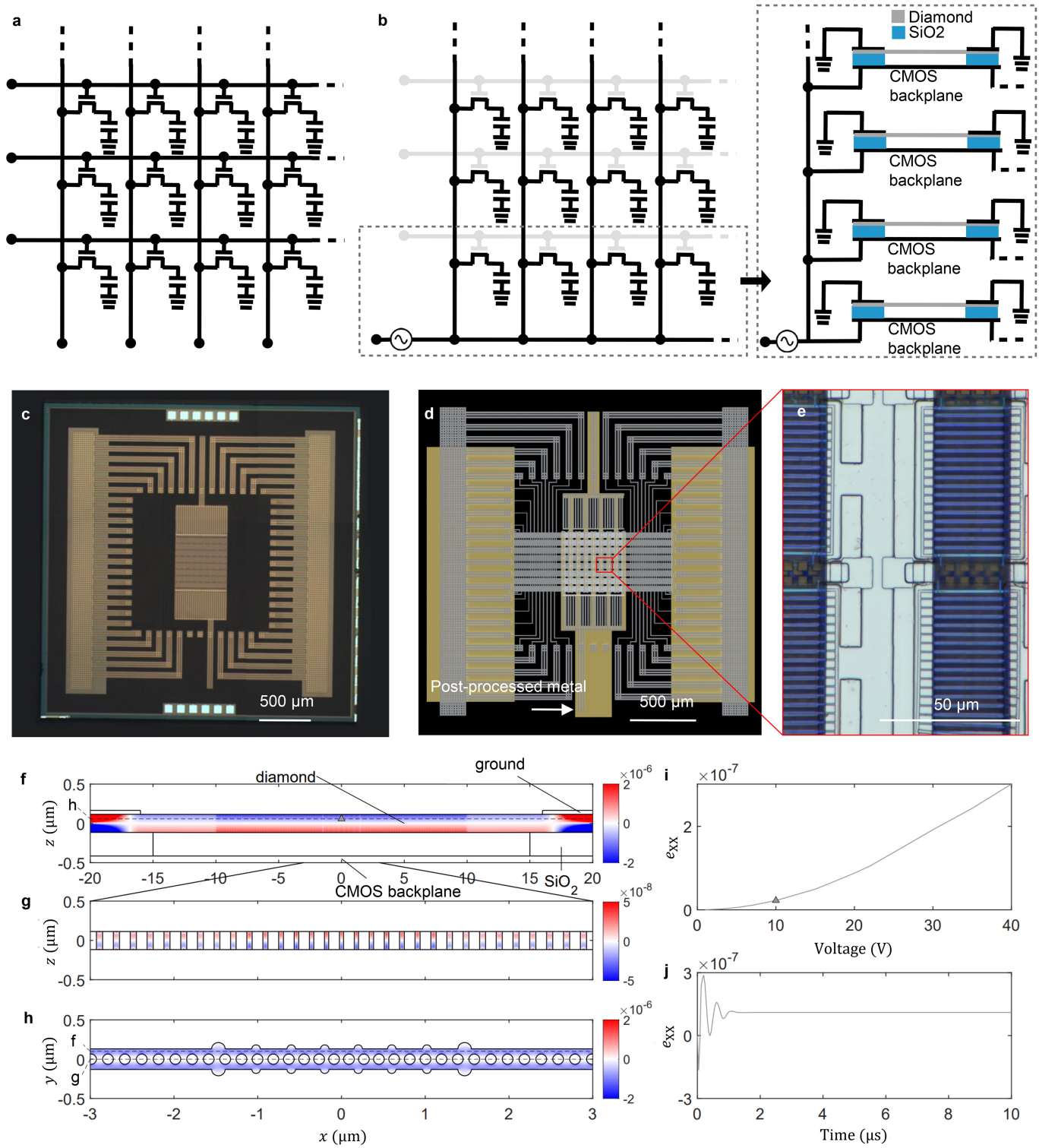
Extended Data Fig. 4 | Diamond QMC array fabrication and gas tuning process for dielectric antenna resonant wavelength adjusting. a, Dark-field optical microscope image of the QMC array on the parent diamond (scale bar 100 µm). **b**, Optical microscope image of a single QMC with 16 quantum channels. (scale bar 10 µm). **c**, Redshift of the cavity resonance due to gas deposition on the cavity, A larger PL number indicates a later measurement time during the gas deposition process. **d**, Blue-shift of the cavity resonance resulting from gas desorption using a pulsed high-power laser.



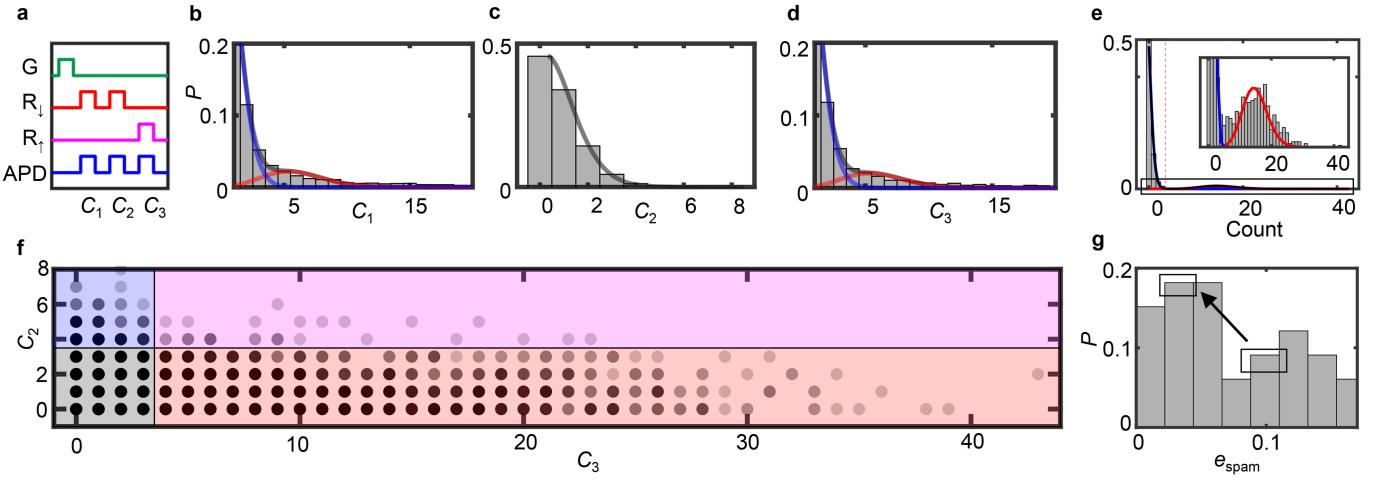
Extended Data Fig. 5 | The Purcell enhancement of the resonant dielectric antenna. a, Simulated Purcell factor and far-field distribution of the quantum emitter in the center of the optimized resonant dielectric antenna. **b**, Resonant dielectric antenna spectrum (interference between laser reflection and resonant dielectric antenna reflection) displaying off-resonant (black) and on-resonant (red) peaks with the SnV^- ZPL wavelength. The inset shows the simulated electric field distribution (depicting the real part of the electric field, with red being positive and blue being negative) overlaid with the edge of the resonant dielectric antenna design, assuming that the dipole is at the center of the resonant dielectric antenna with an in-plane orientation perpendicular to the nanobeam. Four pairs of half-round bumps are placed on each side of the cavity. The centers of the four bumps are 200 nm, 600 nm, 1,070 nm, and 1,470 nm from the cavity center, and their radii are 50 nm, 50 nm, 50 nm, and 80 nm, respectively. **c**, Representative lifetime measurements when the resonant dielectric antenna is on/off-resonant with the SnV^- ZPL after/before gas tuning are shown in red/black, respectively. The lifetime of another SnV^- in the bulk diamond is shown in magenta. The fitted results for these lifetimes are the bulk material $\tau_{bulk} = 4.12$ ns, the emitter in the off-resonance dielectric antenna $\tau_{off} = 5.56$ ns, and $\tau_{on} = 2.32$ ns after tuning the dielectric antenna on-resonance with its ZPL.



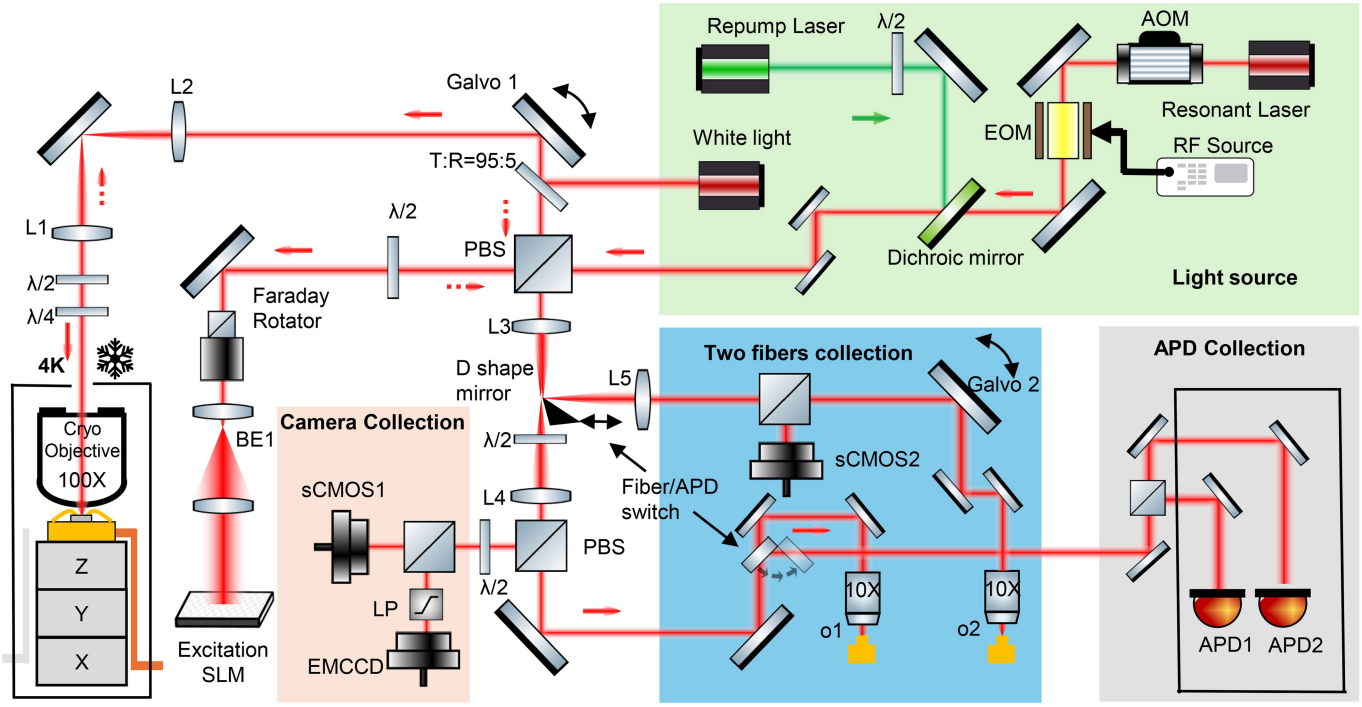
Extended Data Fig. 6 | Diamond quantum emitter widefield PLE cumulative distribution function statistics. a, The cumulative distribution function (cdf) of emitter central frequency f (wide field PLE between 484.115 THz to 484.145 THz), suggesting a uniform distribution of SnV^- centers within this frequency spectrum. **b**, The cdf of the emitter linewidth smaller than $\Delta\nu$, suggesting 20% of emitters have linewidths (inclusive of spectral diffusion) that are within twice the transform-limited linewidth (30 MHz). In the context of optical addressability for the two spin states, color centers with a frequency variation below 200 MHz are considered suitable, accounting for 35% of the sample. **c**, The cdf of the SnV^- ZPL inhomogenous frequency in a FOV, aligning with the near-linear ZPL distribution in a narrow range. **d**, The cdf of the energy splitting larger than ΔE under a 0.13 T magnetic field, revealing 80% of ΔE splitting exceed 0.6 GHz.



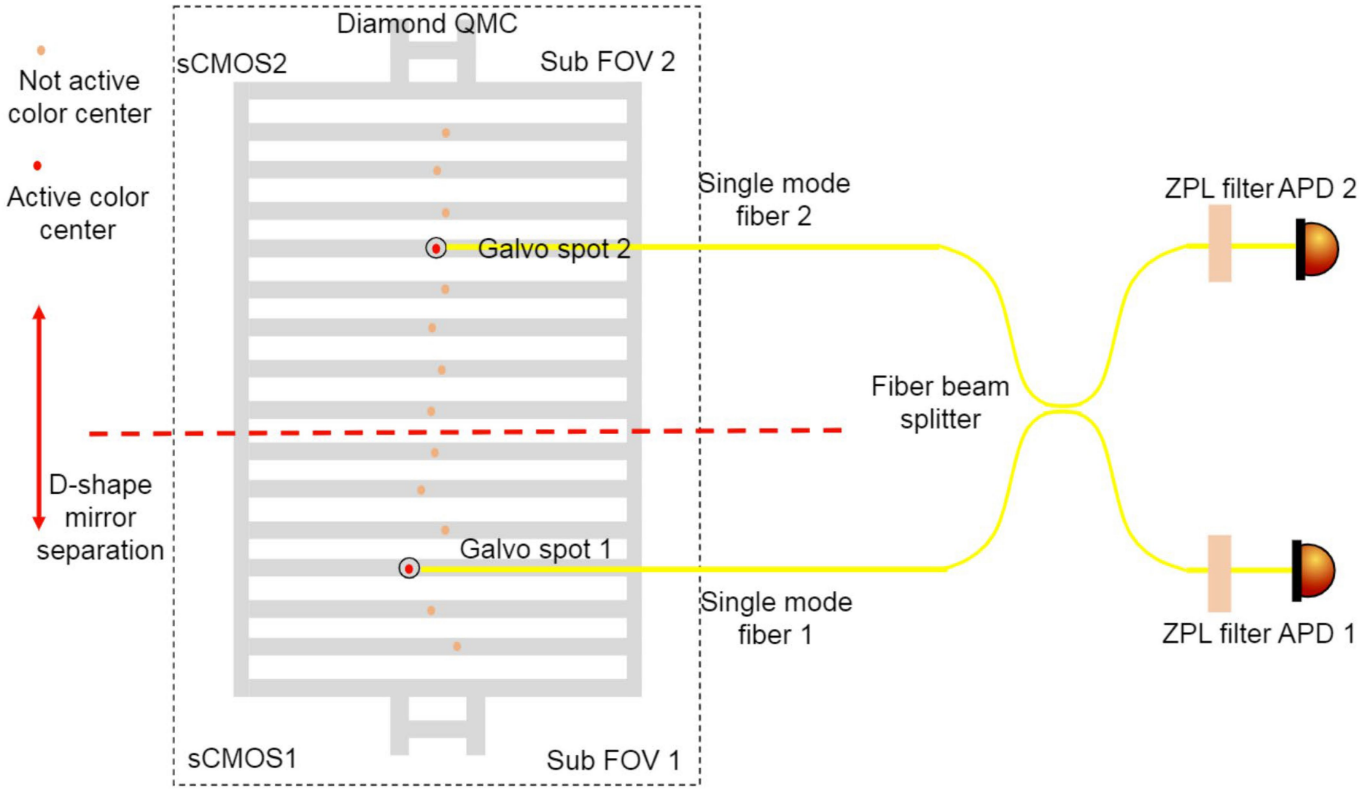
Extended Data Fig. 7 | Strain tuning mechanism and simulation in COMSOL Multiphysics. a, The crossbar structure CMOS circuit diagram, capable of applying individual voltage control to the capacitance of each pixel region. **b**, The circuit in the manuscript, multi-columns are combined to facilitate global bias control. A cross-section image is shown as the sample of the circuit region on the right. **c**, The initial CMOS chip from vendor. **d**, The post-processed CMOS chip design with metalization (labeled in the yellow color, the initial CMOS metal routing is in white color. The chip post-processing metal location highlights the top electrode appearance. Wire bonding is applied to these post-processed metal areas, connecting them to an external Keithley voltage source. **e**, A zoom-in top-view region of the post-processed device layer. **f**, Device strain tensor XX component e_{xx} distribution across the cross-section of the x–z plane at $y = 0.1$ μm in COMSOL simulation. A bias of 10 V is applied between the CMOS backplane and the top ground electrode. **g**, Zoomed-in e_{xx} strain distribution in the x–z plane, with the cavity region on the x-axis ranging from −3 μm to 3 μm and $y = 0$ μm. **h**, e_{xx} strain distribution in the x–y plane with the cavity region at $z = 57.5$ nm, illustrated by the dashed line in **f**. The two dashed lines in **h** shows the x–z plane position in **f** and **g** respectively. **i**, e_{xx} strain relation with bias voltage for the location marked with a gray triangle in **f**. **j**, The transient e_{xx} response when the bias voltage is 10 V at the location marked with a gray triangle in **f**.



Extended Data Fig. 8 | The spin state measurement result without post-selection. a, Pulse control sequence. **b**–**d**, Histogram plots of the spin readout results for the three APD time bins in Fig. 4b, presented without post-selection. **e**, Readout histogram without post-selection, featured a zoomed-in histogram in the inset. **f**, Correlation between the second and third APD time bins, with the correlation plot divided into four quarters, each represented by a different color: gray region ($C_2 < n_m$, $C_3 < n_m$), red region ($C_2 < n_m$, $C_3 > n_m$), blue region ($C_2 > n_m$, $C_3 < n_m$), and magenta region ($C_2 > n_m$, $C_3 > n_m$). n_m is fit from **b** here, contrasting with n_m, which is derived from the post-selected data. In the gray area, we consider the spin to be not initialized to the correct charge state, since APD bins 2 and 3 are both considered dark in the readout result. In the red area, we consider the successful spin preparation and measurement as expected. In the blue and magenta regions, the readout is considered an error for state preparation and measurement. **g**, The histogram of the e_{spam} statistics without post-selection. The red rectangle represents the change of e_{spam} after post-selection corresponding to the data in the manuscript Fig. 4b.



Extended Data Fig. 9 | The optical measurement setup for the hardware architecture demonstration. The setup includes the cryogenic system, laser source, 4f confocal system, SLM excitation, collection box including the duo camera image system (EMCCD and scientific CMOS camera), and collection APD detectors. The resonant laser here is modulated by an AOM and EOM, and mixed with the green repump laser via a dichroic mirror. For SLM excitation, the laser's polarization is optimized with a half-waveplate (HWP) and quarter-waveplate (QWP) before reaching the PBS. A Faraday rotator alters the laser polarization, allowing the SLM-modulated laser signal to be reflected into the cryostat by the PBS. A beam expander (BE) enlarges the beam to cover most of the SLM pixels. The excitation laser beams enter the cryostat through a 4f system (Lenses L1 and L2) after passing through a galvo, which maps the objective back aperture and galvo plane. An HWP and QWP are positioned just before the cryostat window to adjust the setup for cross-polarization or co-polarization collection. The optical setup is designed to confocally collect photons from the quantum emitter, with the signal passing through a PBS. Situated between two lenses (L3 and L4) in the image plane, a movable D-shaped mirror reflects a portion of light in the image plane (depending on the mirror's location) to another optical path, which is then followed by the camera, galvo, and single-mode fiber collection. A scientific CMOS camera monitors the reflected sub-FOV. By adjusting the HWP, the collection signal can be directed either to the camera path or the confocal collection path, which can enter either the free-space APD or fiber-coupled APD for collection via a programmable flippable mirror.



Extended Data Fig. 10 | Approach of the all-to-all routing of photons from quantum emitters. The D-shaped mirror divides the FOV into two sub-FOVs, each monitored with a scientific CMOS camera. The galvo positions are calibrated with potential color center candidates. Photons from each sub-FOV can be collected into single-mode fibres using corresponding galvos. The collected photons then pass through a fiber beam splitter, followed by a ZPL filter and APD for interference measurement.