3D integrated superconducting qubits

Citation for the original published paper (version of record):
Rosenberg, D., Kim, D., Das, R. et al (2017)
3D integrated superconducting qubits
npj Quantum Information, 3
http://dx.doi.org/10.1038/s41534-017-0044-0

N.B. When citing this work, cite the original published paper.
3D integrated superconducting qubits

D. Rosenberg\textsuperscript{1,} D. Kim\textsuperscript{1,} R. Das\textsuperscript{1,} D. Yost\textsuperscript{1,} S. Gustavsson\textsuperscript{2,} D. Hover\textsuperscript{1,} P. Krantz\textsuperscript{3,} A. Melville\textsuperscript{1,} L. Racz\textsuperscript{1,} G. O. Samach\textsuperscript{1,} S. J. Weber\textsuperscript{1,} F. Yan\textsuperscript{1,} J. L. Yoder\textsuperscript{1,} A. J. Kerman\textsuperscript{1} and W. D. Oliver\textsuperscript{1,2,3}

As the field of quantum computing advances from the few-qubit stage to larger-scale processors, qubit addressability and extensibility will necessitate the use of 3D integration and packaging. While 3D integration is well-developed for commercial electronics, relatively little work has been performed to determine its compatibility with high-coherence solid-state qubits. Of particular concern, qubit coherence times can be suppressed by the requisite processing steps and close proximity of another chip. In this work, we use a flip-chip process to bond a chip with superconducting flux qubits to another chip containing structures for qubit readout and control. We demonstrate that high qubit coherence ($T_1, T_{\text{echo}} > 20 \mu s$) is maintained in a flip-chip geometry in the presence of galvanic, capacitive, and inductive coupling between the chips.

**INTRODUCTION**

Superconducting qubits are a prime candidate for constructing large-scale quantum processors due to their lithographic scalability, compatibility with microwave control, gate speed, and relatively long coherence times in planar geometries.\textsuperscript{1,2} Recent increases in coherence times\textsuperscript{3–9} and the development of fast, high-fidelity single-qubit gates\textsuperscript{6–8} and two-qubit gates\textsuperscript{7,9} have yielded control fidelities that exceed the most lenient thresholds required for fault tolerant quantum error correction via the surface code,\textsuperscript{10} a code of particular interest because it requires only nearest-neighbor interactions between qubits. With this motivation, recent experiments have prototyped basic error-detection codes, Bell-state memories, and multi-qubit entangled states using four,\textsuperscript{11} five,\textsuperscript{12} nine,\textsuperscript{13} and ten qubits\textsuperscript{14} in a planar geometry. While these experiments are important demonstrations of the underlying qubit technology, the devices were all controlled and read out using interconnects that laterally addressed the qubits from the perimeter of the same chip. Extending this approach to larger numbers of qubits is impractical due to the interconnect crowding that will occur when addressing qubits within a large two-dimensional array. Moving into the third dimension eases such geometrical constraints, enabling efficient interconnect routing to large 2D arrays, allowing for more compact qubit-qubit coupling geometries, and affording significantly increased connectivity beyond nearest-neighbor interactions that is advantageous for many error correcting codes\textsuperscript{10,15,16} and of importance to quantum annealing and quantum simulation.

One method for accessing the third dimension is to use monolithic fabrication techniques to create a planarized multilayer structure. This method has been used in the D-Wave quantum annealing processors containing more than 2000 qubits (URL https://www.dwavesys.com/). However, with current fabrication techniques the price of monolithic fabrication is a severe penalty on qubit coherence, as evidenced by the low coherence time of the qubits in the D-Wave processor compared with state-of-art in single-layer aluminum devices.\textsuperscript{4,5,17} A previous experiment used a flip-chip architecture with large sapphire spheres setting the spacing between two chips,\textsuperscript{18} but the assembly method used was not scalable and lacked galvanic connection between the chips. More recent efforts have focused on scalable vertical interconnects,\textsuperscript{19–21} but these approaches have not yet demonstrated compatibility with high-coherence superconducting qubits.

Here we describe an approach that leverages heterogeneous 3D integration to create an architecture that enables use of the third dimension without sacrificing qubit performance, and we present proof-of-principle experimental data indicating the feasibility of this approach. Figure 1 shows a schematic of our envisioned structure. The design consists of three chips, attached using superconducting bump bonds, with each chip performing a different function. The top chip contains the superconducting qubits that are the basic logic elements of the quantum processor. The middle interposer chip has patterned surfaces on both sides of the chip, with metalized through-silicon vias (TSVs) providing connectivity between the two surfaces. The bottom chip uses a multilayer planarized process for efficient wire routing\textsuperscript{22} and active Josephson junctions for signal amplification.\textsuperscript{23} In this design, elements on the top surface of the interposer chip, close to the qubits, are galvanically, inductively, or capacitively coupled to the qubits for bias, control, and readout, and these elements connect to the signal readout and interconnect chip through the TSVs and the indium bumps. This design has two significant advantages. First, the fabrication processes for each chip can be performed separately and independently. This is particularly an advantage for fabrication of the qubits, which are notoriously sensitive to materials and processes that can cause decoherence. Second, the thick interposer chip provides a large mode volume for the qubit electromagnetic fields as well as isolation between the qubit and interconnect/readout chip, ensuring that the qubit performance is not degraded by the added system complexity.

A first step towards assessing the practicality of the 3D structure illustrated in Fig. 1 is to determine its impact on qubit...
performance. The presence of an additional surface proximate to the qubit may introduce new sources of noise, reducing qubit coherence times. In addition, 3D integration generally requires additional processing steps, such as depositing additional metal layers and bonding the chips, that may affect qubit performance. To quantify the effect of 3D integration on the qubit, we performed experiments using an intermediate architecture where a qubit chip is bonded to a single chip—an interposer without TSVs—using indium bumps. This allows us to determine the impact of 3D integration and to demonstrate basic desirable functionalities enabled by 3D integration, such as off-chip control and readout of the qubit.

RESULTS

For the experiments described here, we fabricated capacitively shunted aluminum flux qubits using a process that has been described elsewhere. As shown in Fig. 2a, each chip contains six qubits, each of which is inductively coupled to a bias line for applying magnetic flux to shift the qubit energy levels and capacitively coupled to a quarter-wave resonator for control and dispersive readout. The qubits have relatively large loop areas, a design choice related to their application to quantum annealing, and generally have $T_1, T_2 \approx 20 \mu$s, somewhat lower than obtained for gate-based smaller-loop designs with $T_1, T_2 \approx 50 \mu$s. The resonators, which are spectrally spaced by approximately 50 MHz, are all inductively coupled to a transmission line for multiplexed readout and control. Our bump bonding approach included the addition of a patterned under bump metallization (UBM) layer, a metal stack comprising Ti/Pt/Au, to our standard qubit fabrication process, in order to make contact to the aluminum and to provide a diffusion barrier to avoid the formation of intermetallic compounds.

Qubits were designed to be tested either on stand-alone single chips, as shown in Fig. 2a, or in a flip-chip configuration with separate qubit and control/readout chips, as shown in Fig. 2b. For the flip-chip configuration, we bonded the qubit chip to a silicon interposer chip that contained structures (e.g., capacitors, inductors, transmission lines, etc.) patterned from evaporated aluminum. As with the qubit wafers, a patterned under bump metallization layer was used for making electrical contact to the aluminum.

The design shown in Fig. 2b is notable, both because the qubit chip is bonded to another chip, and because all the structures used to control and read out the qubits are on the other chip. With the exception of the underbump metallization structures required for bump bonding and jumpers used to connect disparate sections of ground plane and improve the qubit microwave environment, the only structures on the qubit chip are the qubits themselves. As shown in the color-coded schematic at the bottom of Fig. 2, the flux bias line and readout resonator elements have been relocated to the control/readout chip and are inductively coupled.
and capacitively coupled to the qubit across the gap separating the two chips. To the extent possible, we ensured that the flip-chip qubits were nominally the same as the planar qubits, which served as our experimental controls. This required design modifications in order to account for the vertical spacing (2–10 μm) between the chips and the change in capacitance due to an increased effective dielectric constant resulting from the presence of the extra silicon chip. We took these effects into account and designed the chips so that the shunt capacitance, the capacitive coupling between the qubit and the resonator, and the mutual inductance between the flux bias line and the qubit loop were nominally the same for the qubit designs shown in Fig. 2a, b. Each of the six qubits on the chip in Fig. 2b is designed to be bonded to another chip at one of four possible standoff distances of 2, 5, 10, and 20 μm. Simulations indicate that our design is fairly robust to deviations in the qubit-interposer spacing; for the 10-μm target design, a deviation of 1 μm results in a change in shunt and coupling capacitance of around 2 and 5%, respectively. In practice, we control the chip-to-chip spacing across the 5 mm × 5 mm chips to better than 1 μm.

To determine the impact of bump bonding on qubit coherence, we first tested the capacitively shunted flux qubits in the standard single-chip configuration (Fig. 2a). Based on noise spectroscopy measurements across a range of qubit designs fabricated using the same process, it is expected that both charge and flux noise play a role in limiting the T₁ of these devices. As a result, our measurements are sensitive to increases in both charge and flux noise. The T₁ and T₂,echo times for these qubits were measured to be 10–20 μs, in reasonable agreement with measured flux and charge spectral noise densities. This range represents measurements of nominally identical qubits from different fabrication runs and at different times, and we attribute variations in the measurements to changes in qubit noise sensitivity due to small variations in qubit parameters from chip to chip, the presence of weakly coupled electromagnetic modes that may have small variations from package to package, and temporal fluctuations of quasiparticles.

Figure 3 shows repeated measurements of T₁ and T₂,echo on degeneracy for the flip-chip qubit illustrated in Fig. 2b, where the qubit is biased, controlled, and read out using structures on a separate chip at a standoff distance of 10 μm. We perform repeated measurements and time average over these measurements to account for any long-time temporal variation of the coherence times. We find that the relaxation and echo times of 20.9 and 24.6 μs, respectively, are within the same range as those measured on qubits in the standard configuration (Fig. 2a), indicating that 3D integration did not adversely affect the qubit. The slight increase in T₁ and T₂,echo compared to our control could be due to reduced participation of the surfaces in the qubits electric field, but the increase is within the range of variations we generally observe in similar qubits.

Although the demonstration of off-chip readout and control was enabled by capacitive and inductive coupling alone, direct galvanic connection between bump-bonded chips is required for the full architecture shown in Fig. 1. Using structures such as those shown in Fig. 4, we have measured the inter-chip resistance at low frequencies using chains of bumps and at microwave frequencies using resonators with bump interconnects. For the low-frequency measurements, we performed low-temperature four-wire measurements of the bump chains using both a commercial multimeter and a lock-in amplifier at frequencies ranging from 2 to 200 Hz. We observe changes in resistance at 3 and at 1 K, which we attribute to the indium and the aluminum going through their respective superconducting transitions. Using a chain of 2,704 indium bumps, we measured a DC resistance of 240 nΩ per bump at temperatures well below 1 K, consistent with estimates of the normal state resistance of the under bump metallization layer. To reduce this resistance, we note that the UBM may be replaced by a superconducting material such as TiN.

In addition to DC signals, bumps will likely be used to pass microwave signals, e.g., for qubit control and readout. Therefore, it is important to quantify microwave loss due to the indium bumps, for example, from electrically-active two-level systems (TLS) on the bump surfaces that participate in the qubit electromagnetic mode volume. To measure the bump resistance at microwave frequencies, we designed quarter-wave transmission line resonators with bump interconnects, where the resistance of the bump is manifest as a reduction in the quality factor of the resonator. The resonators, with resonant frequencies ranging from 4.5 to 5.5 GHz, were distributed between two chips, and a single bump with a diameter of 15 μm before compression (30 μm after compression) provided an electrical connection between the sections of the resonator. As shown in Fig. 4b, we designed
The greatest current reduction, since the interconnect is at the position with interconnects at the shorted end of the resonator to have a primary loss mechanism is resistive loss, we expect the resonators with bump interconnects near the middle of the resonator to exhibit a greater interaction with a bath of TLS, we expect the resonators with bump interconnects at the voltage node of the resonator. Depending on the dominant loss mechanism, we expect different results for the position dependence of the loss. If the loss mechanism is primarily through interaction with a bath of TLS, we expect the resonators with interconnects near the middle of the resonator to exhibit a greater reduction in their quality factor, \( Q \). If, on the other hand, the primary loss mechanism is resistive loss, we expect the resonators with interconnects at the shorted end of the resonator to have a greater \( Q \) reduction, since the interconnect is at the position with the greatest current flow.

\[
Q_{\text{bump}} = \left( \frac{1}{Q_i} - \frac{1}{Q_{\text{control}}} \right)^{-1}
\]

The top panel of Fig. 5 shows the reduction of quality factor for seven resonators across two chips. Four of the resonators had interconnects at the voltage node, and three had interconnects near the middle of the resonator. For each chip, the intrinsic quality factor due to other effects (e.g., material losses and non-bump geometry-related effects), \( Q_{\text{control}} \), was measured using a resonator with the same geometry but no bump interconnects. We then measured the internal quality factor of the resonators with bumps, \( Q_i \), and subtracted \( Q_{\text{control}} \) in parallel to obtain the \( Q \) reduction due to the bump \( Q_{\text{bump}} = \left( \frac{1}{Q_i} - \frac{1}{Q_{\text{control}}} \right)^{-1} \). The resonators with the bumps at the voltage nodes clearly show a more pronounced reduction in \( Q \), consistent with the bumps exhibiting a series resistance at microwave frequencies. We extracted this resistance from our data by comparing the reduction in quality factor to simulations of an ideal coplanar waveguide resonator with varying resistance at the appropriate locations along the resonator. The extracted resistances are in the range of 0.1 to 0.5 \( \Omega \) (bottom panel of Fig. 5). The slight power dependence seen in the resonators with bumps in the middle of the resonator is plausibly consistent with the behavior expected from TLS, which should saturate at high powers. If TLS were contributing to the \( Q \) reduction, however, we would expect that at high photon numbers, where the TLS are saturated, the effective bump resistance would be equal to that obtained from the data with bumps at the voltage node. The inconsistency may be related to small systematic differences between the resonators, but remains undetermined.

There are several factors which could contribute to the difference of three orders of magnitude between the DC and microwave resistance. First, the indium could intrinsically be lossier at microwave frequencies compared to DC frequencies. Second, the thin normal underbump layer could result in different current flows at DC and microwave frequencies. Finally, small differences in the design of the DC and microwave structures, such as the spacing of adjacent bumps, could contribute to the resistance difference. Overall, the data indicate that it should be possible to incorporate the bumps in transmission lines with microwave power levels suitable for qubit manipulation and measurement, but, in their present form, not in high-\( Q \) resonators or for transferring quantum information. A next step that may alleviate this issue is the use of superconducting underbump metals such as TiN.

DISCUSSION

Our demonstration of capacitive, inductive, and low-resistance galvanic coupling between two chips is a promising first step...
toward building larger-scale devices for quantum information processing. We have shown that it is possible to control and read out a qubit using off-chip elements while maintaining high qubit coherence. Although we have performed these initial demonstrations with flux qubits, designed for use in quantum annealers, these results are generally applicable to chip-based superconducting and semiconducting qubit modalities used for all forms of quantum information processing, including computation, annealing, and emulation. Additionally, for transmons and other qubits limited by surface dielectric losses, the enhanced capacitance provided by the flip-chip architecture enables the construction of smaller qubits with lower electric field surface participation. We believe these proof-of-principle experiments are the first step towards an architecture that will enable large-scale quantum processing with high-coherence qubits.

METHODS

The capacitively-shunted flux qubits were fabricated on 2" silicon wafers using the process described in ref. 5. The three junctions forming the qubit loop had a ratio between the small and large junction areas of $J_c = 0.43$, with a large junction area of 0.065 $\mu$m$^2$ and critical current density $J_c = 3.3 \times 10^5$ A/cm$^2$. The qubit loop was shunted by a 48 fF planar capacitor, resulting in a qubit frequency of 4.8 GHz at degeneracy.

In the flip-chip configuration, the qubit chips were bonded to a control/ readout chip that had structures patterned from a 250-nm layer of evaporated aluminum, with an underbump metallization layer of Ti/Pt/Au (with layer thickness 20, 50, and 100 nm, respectively) for making contact to the aluminum. Thick (8–30 μm) pillars of indium with diameters of 15 and 30 μm were evaporated on top of the underbump metal (squares in Fig. 2) and patterned using a lift-off step, and a commercial thermocompression bonder was used in force-feedback or distance-feedback mode to bond the chips together at a temperature of 105 °C. Three-dimensional images of the two bonded chips using a white-light interferometer and a confocal microscope indicated that the tilt angle between the two chips was less than 0.4 mRad, and infrared images showed an in-plane alignment error of less than 1 μm.

Data availability

The data that support the findings of this study may be made available from the corresponding author upon reasonable request and with the permission of the US Government sponsors who funded the work.

ACKNOWLEDGEMENTS

We gratefully acknowledge J. Birenbaum, G. Calusine and W. Woods for useful discussions and M. Augeri, P. Baldo, G. Fitch, J. Liddell, K. Magoon, X. Miloski, P. Murphy, B. Osadchy, A. Sevi, R. Slattery, C. Thounmaraj, D. Volfson and T. Weir for valuable technical assistance. This research was funded by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA) and by the Assistant Secretary of Defense for Research & Engineering under Air Force Contract No. FA8721-05-C-0002. The views and conclusions contained in this report are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of ODNI, IARPA, or the US Government.

AUTHOR CONTRIBUTIONS

D.R. performed the measurements. D.R. and W.D.O. wrote the manuscript. D.K., R.D., D.Y., A.M., L.R. and J.Y. fabricated and planned and performed 3D integration of the devices. S.G., D.H., G.O.S., S.J.W., and F.Y. contributed to the qubit design, measurement infrastructure, and data analysis. P.K. rendered the 3D illustrations. A. J.K. and W.D.O. proposed the experiment. All authors discussed the results and commented on the manuscript.

ADDITIONAL INFORMATION

Supplementary Information accompanies the paper on the npj Quantum Information website [https://doi.org/10.1038/s41534-017-0044-0].

Competing interests: The authors declare that they have no competing financial interests.

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

REFERENCES

1. Oliver, W. D. & Welander, P. B. Materials in superconducting quantum bits. MRS Bull. 38, 816825 (2013).
2. Devoret, M. H. & Schoelkopf, R. J. Superconducting circuits for quantum information: an outlook. Science 339, 1169–1174 (2013).
3. Barends, R. et al. Coherent josephson qubit suitable for scalable quantum integrated circuits. Phys. Rev. Lett. 111, 080502 (2013).
4. Rigetti, C. et al. Superconducting qubit in a waveguide cavity with a coherence time approaching 0.1 ms. Phys. Rev. B 86, 100506 (2012).
5. Yan, F. et al. The flux qubit revisited to enhance coherence and reproducibility. Nat. Commun. 7, 12964 (2016).
6. Sheldon, S. et al. Characterizing errors on qubit operations via iterative randomized benchmarking. Phys. Rev. A 93, 012301 (2016).
7. Barends, R. et al. Superconducting quantum circuits at the surface code threshold for fault tolerance. Nature 508, 500–503 (2014).
8. Rol, M. et al. Restless tuneup of high-fidelity qubit gates. Phys. Rev. A 7, D41001 (2017).
9. Sheldon, S., Magesan, E., Chow, J. M. & Gambetta, J. M. Procedure for systematically tuning up cross-talk in the cross-resonance gate. Phys. Rev. A 93, 060302 (2016).
10. Fowler, A. G., Mariotoni, M., Martinis, J. M. & Cleland, A. N. Surface codes: Towards practical large-scale quantum computation. Phys. Rev. A 86, 032324 (2012).
11. Corcoles, A. et al. Demonstration of a quantum error detection code using a square lattice of four superconducting qubits. Nat. Commun. 6, 6979 (2015).
12. Ristè, D. et al. Detecting bit-flip errors in a logical qubit using stabilizer measurements. Nat. Commun. 6, 6983 (2015).
13. Kelly, J. et al. State preservation by repetitive error detection in a superconducting quantum circuit. Nature 519, 66–69 (2015).
14. Song, C. et al. 10-qubit entanglement and parallel logic operations with a superconducting circuit. arXiv:1703.10302 (2017).
15. Bombin, H. & Martin-Delgado, M. A. Topological quantum distillation. Phys. Rev. Lett. 97, 180501 (2006).
16. Kovalev, A. A. & Pryadko, L. P. Fault tolerance of quantum low-density parity check codes with sublinear distance scaling. Phys. Rev. A 87, 020304 (2013).
17. Weber, S. J. et al. Coherent coupled qubits for quantum annealing Phys. Rev. A 88, 014004 (2017).
18. Li, D., da Silva, F. C. S., Braje, D. A., Simmonds, R. W. & Pappas, D. P. Remote sensing and control of phase qubits. Appl. Phys. Lett. 97, 102507 (2010).
19. Béjanin, J. et al. Three-dimensional wiring for extensible quantum computing: The quantum socket. Phys. Rev. Appl. 6, 044010 (2016).
20. Versluis, R. et al. Scalable quantum circuit and control for a superconducting surface code. arXiv:1612.08208 (2016).
21. Liu, Q. et al. Extensible 3d architecture for superconducting quantum computing. Appl. Phys. Lett. 110, 232602 (2017).
22. Tolpygo, S. K. et al. Fabrication process and properties of fully-planarized deep-submicron Nb/AI – AlOx/Nb Josephson junctions for VLSI circuits. IEEE Trans. Appl. Supercond. 25, 1–12 (2015).
23. Macklin, C. et al. A near-quantum-limited josephson traveling-wave parametric amplifier. Science 350, 307–310 (2015).
24. Gustavsson, S. et al. Suppressing relaxation in superconducting qubits by quadratic...