Supplementary Information

Experimental Calibrations

The classical limit $|S| \leq 2$ for the CHSH experiment is derived with very minimal assumptions. These include the reproducibility of the measurement axes a, a', b, and b', the space-like separation and thus independent measurement of the particles (basis for locality loophole), and the completeness of the ensemble measurement (basis for detection loophole). But the derivation is, for example, not based on any assumptions about the actual state of the particle pair before separation, the choice of measurement axes, or even the coherence of the states or fidelities of the measurement as long as all introduced errors act on the individual qubits and do not introduce correlations. Thus, it is possible to calibrate almost all parameters describing the experiment with a global optimization process that maximizes the Bell signal $|S|$.

In our experiment, these parameters include all numbers describing the sequence shown in Fig. 1e, including the phase, frequency, and shape of the initial π-pulse, the shape of the pulses that sweep the qubits into resonance with the resonator, and even the shape of the measurement pulses, while ensuring that the two qubits are kept off-resonance from the resonator to avoid further coupling. The optimal values for most of these parameters depend on sample properties such as the coupling strengths between the qubits and the resonator, and are thus not predictable in a useful way. However, the optimal rotation angles for the measurement, i.e. the measurement axes, are predicted, although not uniquely, by quantum mechanics and can thus be used to verify the optimization process.

Quantum mechanics predicts a maximal violation, for example, using measurement axes in the Y/Z-plane that form angles with the z-axis of $a = -135°$, $a' = 135°$, $b = 0°$, and $b' = 90°$. Our optimization resulted in angles of $a = -149°$, $a' = 156°$, $b = 1°$, and $b' = 92°$ that lie in planes deviating from the Y/Z-plane by less than 15°. Given the other non-idealities of the experiment and the fact that, around the maximum, the obtained S-Value depends only to second order on these angles, this good match with theory makes us confident that the optimization found a sensible solution. This confidence is supported further by the fact that several different optimization schemes yield parameters that are consistently close to these.

Measurement Crosstalk

As measurement crosstalk poses the greatest challenge to our experiment, we devised a sensitive test to quantify this error mechanism. This test consists of keeping one qubit in the $|0\rangle$ state while driving a Rabi oscillation on the other qubit. If the qubits are kept off resonance from the resonator and each other during this experiment, the qubit in the $|0\rangle$ state should ideally remain unaffected by the state of the other qubit. However, measurement crosstalk does cause a small oscillation on the measured state populations of the inactive qubit at the same frequency as the Rabi oscillation on the other qubit. Thus, a comparison of the Fourier amplitudes of the observed oscillations in the state populations of the two qubits yields a direct number for the strength of the measurement crosstalk. Fig. 1 shows the data resulting from the experiment and yields a value for the measurement crosstalk $p_c^a = 0.59\%$ from qubit A to B and $p_c^b = 0.31\%$ from qubit B to A.

This crosstalk leads to a correction in the limits on the Bell signal dictated by a local hidden variable theory [1]:

$$-2 + 4\,\min\{p_c^a,\, p_c^b\} \leq S \leq 2 + 2\left|p_c^a - p_c^b\right| \tag{1}$$

Using the values for the measurement crosstalk in our sample, we find the new classical limit to be:

$$-1.9876 \leq S \leq 2.0056 \tag{2}$$

This correction is small enough to not challenge our claim of a violation.

Statistical Analysis

For the measured Bell signal to carry statistical meaning, it needs to be supplemented with an estimate of its standard error. As S is determined by sampling the multinomial distributions that describe the qubits' state, the standard error on S is dominated by statistical sampling noise for small sample sizes. As the sample size increases, though, the error on S shows more and more influence from experimental drifts and $1/f$ noise. The estimation of the standard error for large sample sizes therefore requires a noise and drift model that accounts for these experimental systematic errors.

To circumvent this, we divided the entire dataset into sections, each of which is small enough to be dominated by statistical sampling noise. For this, we analyzed the internal variance in our dataset, as shown in Fig. 2 to determine the maximum acceptable section size. We found that for sections up to 1.55 million samples, or about 20 minutes worth of data taking, the variance is sampling-noise-limited, allowing us to employ standard statistical analysis techniques to estimate the standard error on S for each section. We therefore divided our dataset of 34.1 million samples into 22 sections that produce violations with values of S ranging from 2.0666 to 2.0806 and standard errors around 0.0014, corresponding to violations by about 50 standard deviations. These standard errors can be used in one-sided hypothesis tests to estimate the certainty with which each respective section indicates a non-classical Bell signal. If the 22 sections are combined to yield an overall certainty, a corresponding standard error can be inferred, with which we arrive at our final violation claim of 244 standard deviations.

Quantum Simulation and Sample Performance Parameters

To further verify the experiment, we employed quantum simulations to predict the Bell signal. For the purposes of the simulation, the resonator is treated as a third qubit, which is acceptable in the special case of this experiment since the entire quantum circuit never contains more than one photon while the qubits are coupling to the resonator. The state of the system is then expressed by an 8×8 density matrix in the basis of the system's eight states $|000\rangle$, $|001\rangle$, $|010\rangle$, $\ldots$, $|111\rangle$. Rotation operations on the qubits are simulated via the matrix exponentials of the appropriate Pauli matrices, e.g. a 90° x-rotation on qubit A would be simulated via:

$$\rho_{out} = e^{i\,\pi\,\sigma_x \otimes \mathbf{I} \otimes \mathbf{I}/4}\,\rho_{in} \tag{3}$$

Coupling operations are simulated using matrix exponentials of the coupling matrix

$$\mathbf{C} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{4}$$

For example, a swap operation between the resonator and qubit B is simulated via:

$$\rho_{out} = e^{i\,\pi\,\mathbf{I} \otimes \mathbf{C}/2}\,\rho_{in} \tag{5}$$

Single qubit decoherence and dephasing are added by applying the operations in small steps interleaved with the Kraus operators [CITE] that relax or dephase the state. Measurement errors are included by modelling them with a classical probability to misidentify the individual qubits' states.

Using just the single qubit and resonator performance characteristics T_1, T_2, F_0, and F_1 as shown in Table I and assuming that the coupling operation is ideal, we were able to explain our data with very high fidelity. From this we conclude that efforts to improve our architecture need to be focused primarily on single qubit performance, while a scaling to a larger collection of qubits should not introduce any new error mechanisms (beyond measurement crosstalk) and should thus be relatively straight forward to achieve.

It is important to note that the quantum simulations did not contain any fit parameters, and were instead based solely on the actual sequence parameters and the numbers in Table I, which in turn were measured directly using standard decay and Ramsey techniques. The measurement fidelities, specifically F_1, were somewhat non-trivial to measure well without assumptions about other experimental fidelities. We devised an experiment based on multiple pulse-amplitude-driven Rabi oscillations as shown in Fig. 3. With this method, we found the highest measurement fidelities ever reported in phase qubits as shown in Table I- well above 90% and within a few percent of the theoretically expected maximum of 96.6%.

Experimental Data and Measurement Correction

Since the reduced measurement visibilities classically affect the two qubits independently and do not introduce correlations into the measurement, it is theoretically legitimate to correct our data for these to estimate the Bell signal that we would have obtained with perfect fidelities. Table II shows the raw state probabilities observed in our experiment on which the violation claim in this paper is based. Table III shows the corrected state probabilities and the resulting estimated Bell signal for ideal measurement. The observed number matches the simulated value of $S = 2.337$ very well.

We provide this corrected value of S not to claim a larger Bell violation, but instead as a benchmark of the fidelity of the quantum operations we performed on the qubit pair. The separation between quantum operations and qubit readout is useful, in our opinion, as the number of quantum operations required to implement any significant quantum calculation will probably outscale the number of qubit readouts required by a large factor.

TABLE I: Performance parameters for qubits . T_1 and T_2 are the qubit energy and phase relaxation times, T_φ the pure phase decoherence time, F_0 and F_1 the measurement fidelities for the $|0\rangle$ and $|1\rangle$ state measurements,

Parameter	Value
Qubit A:	
T_1	$296\,\mathrm{ns}$
T_2	$135\,\mathrm{ns}$
T_φ	$175\,\mathrm{ns}$
F_0	$97.04\,\%$
F_1	$96.32\,\%$
Qubit B:	
T_1	$392\,\mathrm{ns}$
T_2	$146\,\mathrm{ns}$
T_φ	$179\,\mathrm{ns}$
F_0	$96.18\,\%$
F_1	$98.42\,\%$
Resonator:	
T_1	$2,552\,\mathrm{ns}$
T_2	$\sim 5,200\,\mathrm{ns}$
T_φ	$\sim \infty$
Coupling:	
Qubit $A \leftrightarrow$ resonator	$36.2\,\mathrm{MHz}$
Qubit $B \leftrightarrow$ resonator	$26.1\,\mathrm{MHz}$
Measurement Crosstalk:	
Qubit $A \rightarrow$ qubit B	$0.31\,\%$
Qubit $B \rightarrow$ qubit A	$0.59\,\%$

TABLE II: Bell violation results

Parameter	ab	$a'b$	ab'	$a'b'$
P_{00}	0.4162	0.3978	0.1046	0.3612
P_{01}	0.1575	0.1759	0.3700	0.1136
P_{10}	0.0852	0.0731	0.3904	0.1185
P_{11}	0.3412	0.3531	0.1350	0.4066
E	0.5147	0.5019	-0.5208	0.5358
S	2.0732			

TABLE III: Bell violation results, corrected

Parameter	ab	$a'b$	ab'	$a'b'$
P_{00}	0.4406	0.4213	0.0900	0.3813
P_{01}	0.1343	0.1539	0.3790	0.0880
P_{10}	0.0726	0.0599	0.4166	0.1092
P_{11}	0.3525	0.3649	0.1145	0.4215
E	0.5862	0.5724	-0.5911	0.6055
S	2.3552			

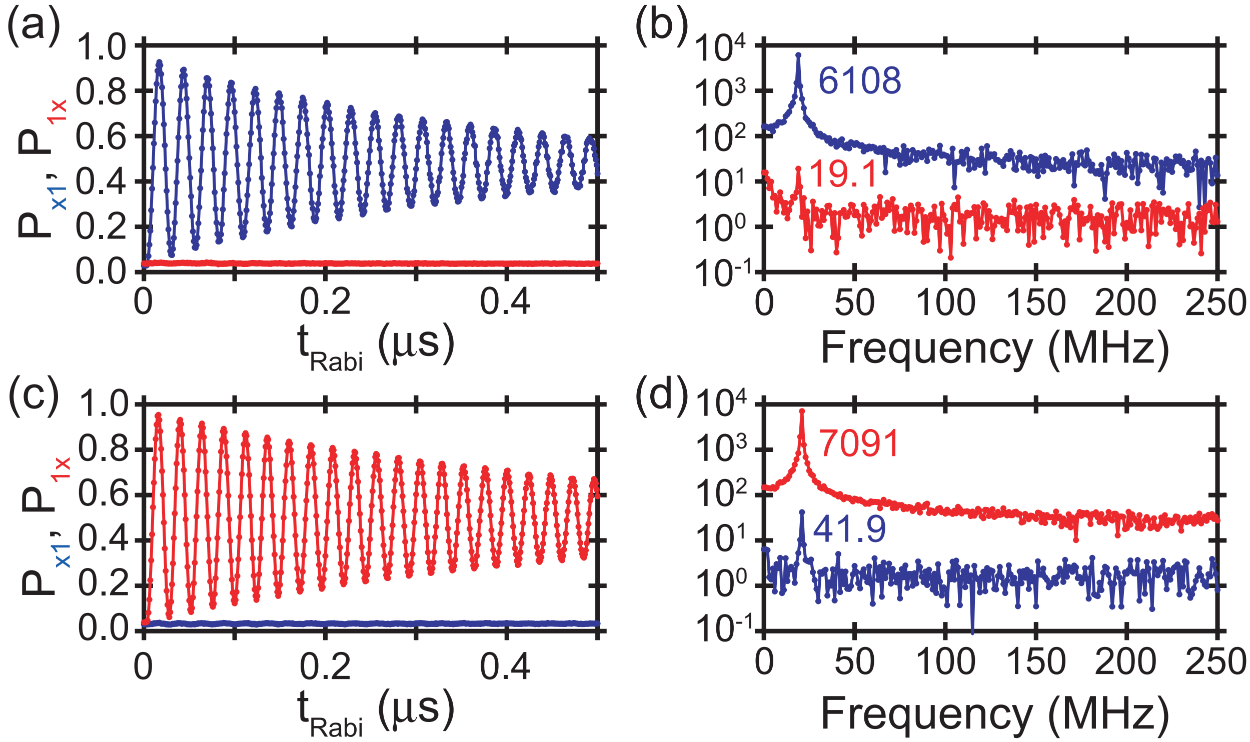

FIG. 1: Quantifying measurement crosstalk: Measurement crosstalk can be quantified by driving a Rabi oscillation on one qubit and observing the other qubit's response. Fourier transforming the data allows the isolation of the relevant features. (a) Rabi oscillation for qubit A (blue). The measured state of the qubit B (red) only shows a very weak dependence on whether the qubit A is in the $|1\rangle$ or $|0\rangle$ state. Here, x represents a sum over the probabilities for 0 and 1. (b) Fourier transform of (a). The ratio of the responses of the two qubits at the same frequency as the Rabi oscillation on A gives a number for the measurement crosstalk, here $19.1/6108 = 0.31\,\%$. (c) Rabi oscillation for qubit B (red); qubit A in blue; x represents a sum over the probabilities for 0 and 1. (d) Fourier transform of (c): Data shows $41.9/7091 = 0.59\,\%$ for measurement crosstalk.

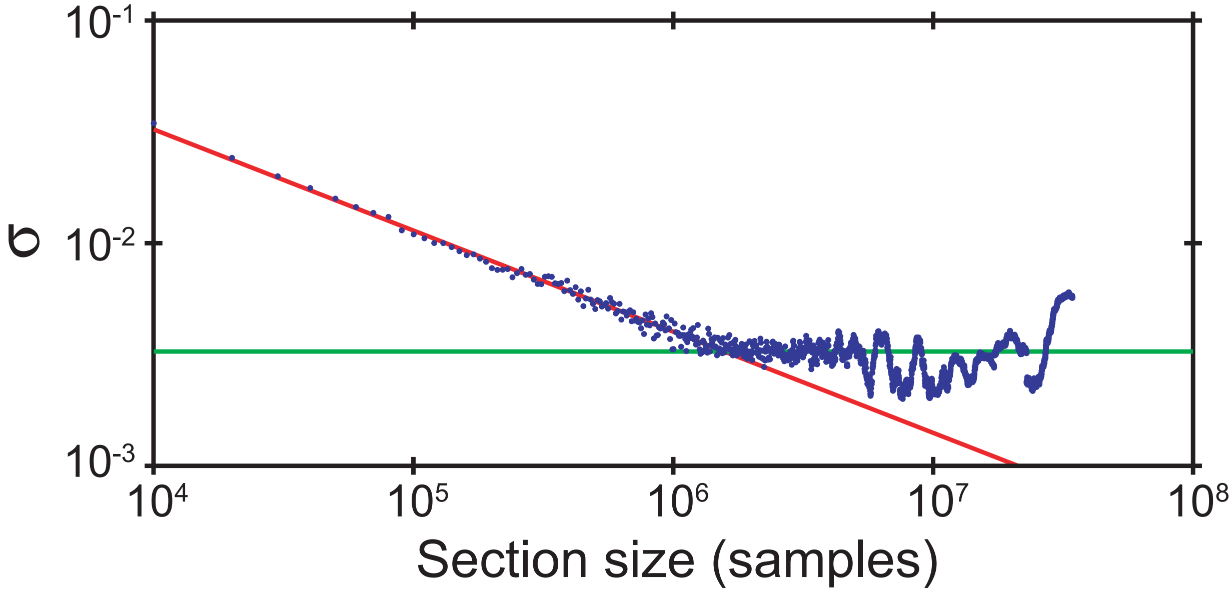

FIG. 2: Standard error analysis: As the sample size increases, the standard error of the estimated mean changes from being dominated by statistical sampling noise (red line) to being dominated by $1/f$ drift in the experiments (green line). The point where the two lines cross gives the maximum sample size that can be statistically analyzed in a meaningful way, without modeling drifts and $1/f$ noise in the experiment.

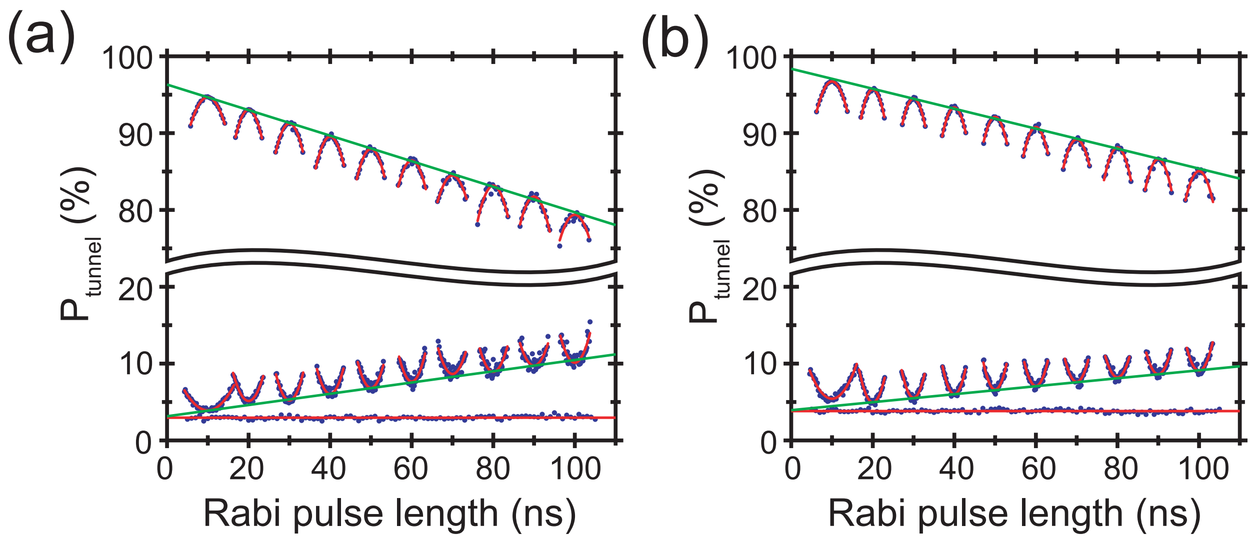

FIG. 3: Visibility analysis (composite of several data sets). Blue dots represent data, red lines are fits through the data, and green lines are fits through the extrema of the red lines. The upper parabolas correspond to Rabi oscillations driven with pulses at fixed length and increasing amplitude around the point where they yield a π-pulse. The bottom parabolas are Rabis driven around pulse amplitudes that yield a 2π pulse. The horizontal dataset at the bottom corresponds to no drive on the qubit. The green fits through the parabolas' extrema (optimal π or 2π pulses) give the measurement visibility when extrapolated to $t = 0$, i.e. to an optimal, instantaneous pulse. The horizontal line checks the method by providing a direct measurement of the $|0\rangle$ state visibility. Since the measurements agree to high precision, the method can be trusted to extract a $|1\rangle$ state fidelity, for which no direct measurement is available. Results are for (a) Qubit A: $F_{0,\mathrm{Rabi}} = 96.86\,\%$, $F_{0,\mathrm{direct}} = 97.04\,\%$, $F_1 = 96.32\,\%$, and (b) Qubit B: $F_{0,\mathrm{Rabi}} = 96.06\,\%$, $F_{0,\mathrm{direct}} = 96.18\,\%$, $F_1 = 98.42\,\%$.

References

[1] Kofman, A. G. & Korotkov, A. N. Analysis of Bell inequality violation in superconducting phase qubits. *Physical Review B (Condensed Matter and Materials Physics)* **77**, 104502 (2008).