Direct Wigner tomography of a superconducting anharmonic oscillator

Yoni Shalibo, Roy Resh, Ofer Fogel, David Shwa, Radoslaw Bialczak, John M. Martinis and Nadav Katz

Supplementary Information

Materials and methods. The Josephson phase circuit [1] used in the experiment has the following design parameters: critical current $I_0 \approx 1.5\,\mu$A, capacitance C≈1.3 pF and inductance L≈940 pH. The qubit has a tunable frequency f_{01} in the 6-9 GHz range. During the experiment the device is thermally anchored to the mixing chamber of a dilution refrigerator at 30 mK, where thermal excitations of the qubit are negligible.

We use a custom built arbitrary waveform generator (AWG) having a fast (1 ns time resolution), 14-bit digital-to-analog converter to shape the microwave pulses which control the quantum state of the circuit. We control the phase and amplitude of the drive by modulating a high-frequency oscillator of frequency f_{LO} with the AWG, using an IQ-mixer. The modulation signals are fed into the I and Q ports of the IQ-mixer to give an amplitude $\sqrt{I(t)^2+Q(t)^2}$ and a relative phase $\phi(t) = \arctan\left(Q(t)/I(t)\right)$ at its output.

Harmonic response. As indicated in the manuscript, the bandwidth of the tomography pulse must be large relative to the anharmonicity, for a nearly harmonic response. This is easily understood by looking at the position of the transitions in the spectral domain, relative to the local oscillator frequency (see Fig. S1). As seen in the figure, the effective drive amplitude of the lowest 6 transitions is the same, to within a 10 % variation. This is consistent with our data and simulations (see Sec. II), showing small systematic errors within the corresponding subspace. In principle, more transitions can fit within this region by positioning the local oscillator lower in frequency.

Parity measurement. In the manuscript we made the assumption that the parity operator can be calculated within the harmonic approximation, that is: $\langle \Pi \rangle \approx W(\alpha) = (2/\pi)\sum_n (-1)^n P_n$. By direct calculation of the parity from the eigenstates of our system we find that the bottom 5 states have a well defined parity $(-1)^n$ to within 1 %. The parity of higher states decreases to ~ 0.6 at $n = 10$. We separately checked using a full simulation, taking into account all the matrix elements, that our parity approximation yields the same results.

Populations measurement. The measured escape probabilities due to a set of experiments with different measurement pulses are converted into level occupation probabilities using the measured escape curves of the lowest levels in the well [2]. In this experiment, we used a novel measurement pulse shape, to reduce distortions of the escape curves of higher levels. This measurement pulse is shown in Fig. S2. Typically, one must use a pulse with a slowly decaying end, in order to reduce the effect of population re-trapping in the original well after the pulse ends. The effect of re-trapping becomes more prominent as the anharmonicity is reduced due to the corresponding deepening of the potential well, and therefore requires increasing both τ_2 and τ_3 to eliminate re-trapping. At the working bias point used in our experiment, we use $\tau_2 = 25$ ns and $\tau_3 = 15$ ns. These parameters are non-negligible relative to the decay time and cause artificial increase of the extracted population at lower levels, which in turn distorts the Wigner image. To avoid this effect we add a small step in amplitude at the beginning of the pulse ($\tau_1 = 1$ ns) to preselect the population we want to escape, and then reduce the amplitude by an amount δA. Even the small difference δA is

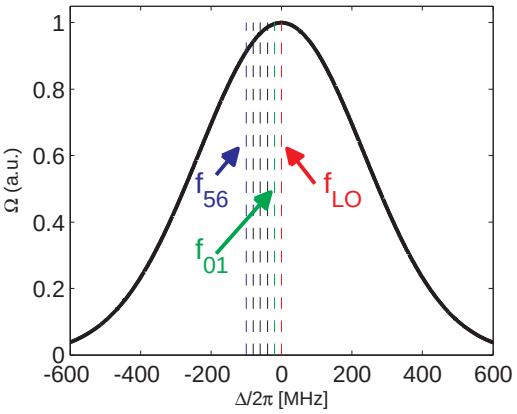

Figure S1: Tomography pulse envelope in the frequency domain. Dashed lines are the lowest five transitions $f_{n,n+1}$ and the local oscillator frequency position in the experiment. The pulse envelope (solid line) is the normalized fourier transform of a 1.6 ns FWHM gaussian function.

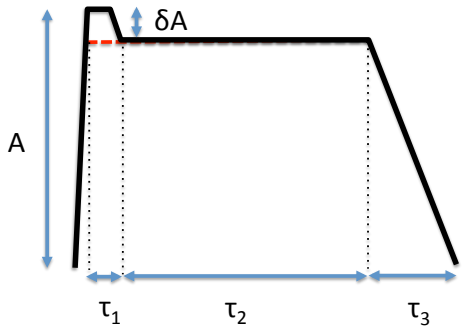

Figure S2: Measurement pulse used in the experiment. All the pulse parameters except for the total amplitude A are fixed.

sufficient to protect undesired population at lower levels from tunneling out during the long waiting time in which the escaped population decays in the external well.

Density matrix fit. In the Wigner tomography experiments we extract the density matrix from the populations of the displaced states [3]. We use 200 homogeneously distributed random displacements within a $|\alpha| < 2$ circle to fit the density matrix, while restricting the density matrix to a 6×6 subspace. It should be noted that both the measured Wigner distribution and the extracted density matrix represent the state after a rotation that occurs during the tomography pulse. To get more accurate phases, one can apply an inverse propagator on the density matrix $U = \exp(-iH_0\Delta t/\hbar)$, where H_0 is the drive-free Hamiltonian and Δt is the effective pulse length for the rotation.

I. STATE PREPARATION BY GENETIC OPTIMIZATION

Finite anharmonicity makes it possible to prepare states, composed of arbitrary superpositions of eigenstates within our system. However, fourier broadening of the drive, together with power broadening of the transition energies causes nontrivial excitation at relatively small anharmonicities. Adding the short decay and coherence time, it becomes challenging to prepare a desired, yet pure state. We solve this difficulty, by optimizing the state produced with a feedback from the experiment. Our target state in the optimization algorithm is a superposition of eigenstates of our system: $|\psi_l\rangle = \left(|0\rangle + e^{i\phi}|l\rangle\right)/\sqrt{2}$. Using the measured probabilities we are able to optimize such a superposition up to an unknown phase ϕ that causes only a free rotation of the measured Wigner function.

Our optimization algorithm is based on guided evolution. It is constructed in the following steps:

(a) We define a set of N_G pulse sequences (genomes). Each genome is a sequence of N_t complex numbers that represent the amplitude and phase of the drive's pulse envelope at each time step, with 1 ns time resolution. In addition, each genome is associated with a population overlap, defined as: $\chi = \sqrt{\overrightarrow{P_{ideal}}} \cdot \sqrt{\overrightarrow{P_{meas}}}$, where $\overrightarrow{P_{ideal}}$ and $\overrightarrow{P_{meas}}$ are the desired and measured population vectors . The genomes are initialized with random complex numbers, each having a maximal amplitude Ω_{max}, and $\chi = 0$.

(b) At each iteration of the algorithm (generation), we assume that the current set of genomes is sorted in order of decreasing χ. The upper N_a genomes are kept, and the next N_a genomes are replaced by a combination of the upper $N_a + 1$ genomes: $\overrightarrow{G_{N_a+k}} = \overrightarrow{P}(G_k, G_{k+1}) + \overrightarrow{\epsilon}$, where the function $\overrightarrow{P}(x, y)$ randomly selects an amplitude between the sets x and y at each time step and $\overrightarrow{\epsilon}$ is a set of random complex numbers with a small amplitude ϵ_{max} (noise). The bottom $N_G - 2N_a$ genomes are replaced by random complex numbers with an amplitude Ω_{max}.

(c) The bottom $N_G - N_a$ genomes of the next generation are then applied and the resulting probabilities $\overrightarrow{P^k_{meas}}$ for each genome are measured. After calculating χ of each, we sort our new set G in order of decreasing χ, and mark genome sequences that have (potentially) higher χ than the current optimal χ. For each genome having a potentially higher χ we repeat the measurement N_{rep} times, recalculate χ and reposition them in the set G. Step (b) is run again.

Fidelity measure. Our population overlap χ measures the distance between two states. There are many possibilities of defining the fidelity, and in general each gives a different bias for the algorithm towards finding a specific family of states. In our definition for example, χ is highly sensitive to the overall population that exists outside the subspace that is populated in the target state. However it is much less sensitive to the distribution among levels that are populated in the target state. Our definition is particularly useful for optimizing our target states $|\psi_l\rangle$ for a Wigner measurement, because we wish to eliminate the population of states other than $|0\rangle$ and $|l\rangle$.

There are two important factors that limit the performance of any optimization algorithm in experimental systems: shot noise, and drifts in the physical parameters of the system.

Shot noise. Any finite number of repetitions results in some uncertainty in the measured probabilities. To determine n occupation probabilities in the final state we measure the tunneling probabilities after n measurement pulses having different amplitudes. In a typical optimization algorithm, we repeat the measurement of a single tunneling probability 900 times. Statistical analysis gives a typical uncertainty (standard deviation) of $\sim$2 % in χ for sufficiently high values of χ ($\chi \gtrsim 80\,\%$) and therefore one cannot distinguish between two genomes with a χ difference which is smaller than $\sim$3 %. We therefore expect the algorithm to run significantly more slowly at χ's approaching unity, due to false increase of χ. To increase the efficiency of the algorithm without increasing the number of repetitions of each measurement, we repeat the measurement only for genomes with potentially increased χ, as described in step (c).

Drifts. The response of the system to the application of a given genome can change in time due to drifts in its physical parameters. We correct for drifts in the energies by performing a spectroscopic measurement of the qubit transition frequency every 10 minutes and adjusting the flux bias to restore the qubit frequency to its original value. In addition, we correct for drifts in the offset voltages applied to the IQ mixer that eliminate leakage of

the microwave drive while it is turned off.

II. SYSTEMATIC ERRORS

To quantify the errors in the measured Wigner function and density matrix caused by finite anharmonicity and decoherence, we perform numerical simulations . In our simulations, we propagate an anharmonic system, initialized with a pure state $\rho = |\psi\rangle\langle\psi|$, with resonant pulses. Each pulse is assumed to be resonant with the first transition and has a gaussian envelope with the same width as in the experiment. We use the rotating wave approximation for constructing the Hamiltonian, and neglect corrections to the drive coupling beyond the harmonic approximation, namely $\langle n|\,\delta\,|m\rangle = 0$ for $n \neq m \pm 1$ and $\langle n|\,\delta\,|m\rangle \propto \sqrt{n}, \sqrt{n+1}$ for $m = n \pm 1$. This results in the following time dependent unitary propagator: $\mathrm{U}(t) = \exp\left(i\frac{\delta t}{2}\left[\left(\Omega(t)a^\dagger + \Omega(t)^* a\right) + \beta a^\dagger a(a^\dagger a - 1)\right]\right)$, where δt is the time step in the simulation, $\Omega(t)$ is the time dependent drive envelope amplitude and $\beta = 2\pi\left(f_{01} - f_{12}\right)$ is the anharmonicity. Decoherence and energy relaxation are taken into account using quantum operations, assuming only two phenomenological parameters: the energy relaxation time of the qubit T_1 and its pure dephasing time T_2.

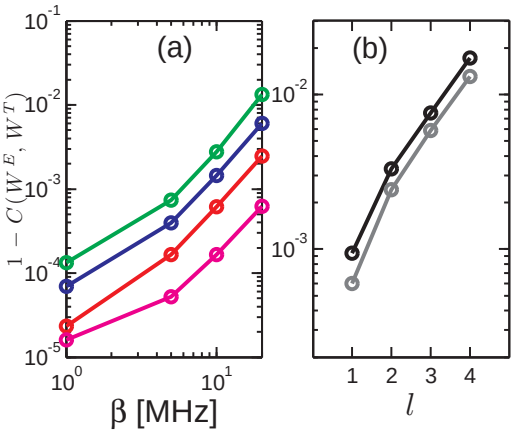

Figure S3: Systematic errors in the Wigner distribution, for initial states $|\psi_l\rangle = (|0\rangle + |l\rangle)/\sqrt{2}$, where $l = 1, 2, 3, 4$. The cross-correlation deviation as a function of the anharmonicity (a) and initial state. In (a), decoherence is not included, and the magenta, red, blue, green lines correspond to the initial state $l = 1, 2, 3, 4$ respectively. In (b) decoherence is included and the anharmonicity is fixed to $\beta/2\pi = 20\,\mathrm{MHz}$. The black and grey lines correspond to decoherence parameters $T_1 = 150\,\mathrm{ns}$, $T_2 = 200\,\mathrm{ns}$ and $T_1 = 600\,\mathrm{ns}$, $T_2 = 600\,\mathrm{ns}$ respectively.

We expect to have negligible systematic errors due to finite anharmonicity when the second term in the exponent of the propagator becomes negligible. By separating the terms in the exponent to first order, and assuming a constant drive amplitude Ω of total duration T, we can approximate the propagator to to $U \approx D(\alpha)\Gamma(T)\exp(-i\frac{T\beta}{4}[(\alpha a^\dagger - \alpha^* a), a^\dagger a(a^\dagger a - 1)])$ where $D(\alpha)$ is the displacement operator, $\alpha \equiv i\Omega T$, and $\Gamma(T) = \exp(i\frac{T\beta}{2}a^\dagger a(a^\dagger a - 1))$ is an overall dispersion operator. The second term can be considered negligible in the limit, $|\alpha| T\beta m^2/4 \ll 1$, where m is the maximal occupied state. For our experimental parameters ($\beta/2\pi = 20\,\mathrm{MHz}$, $T = 1.6\,\mathrm{ns}$), the second term can be neglected only for $|\alpha| \ll 1$, however, the error in the Wigner distribution, obtained from the state populations after the pulse are negligible even for $\alpha \approx 2$, as we show in our analysis. The reason for this is the insensitivity of the parity value to excited states larger than $n \approx 4$. Despite that fact that these states are populated with occupation probabilities that differ significantly from the harmonic case, their contribution cancels out in the calculation of the parity.

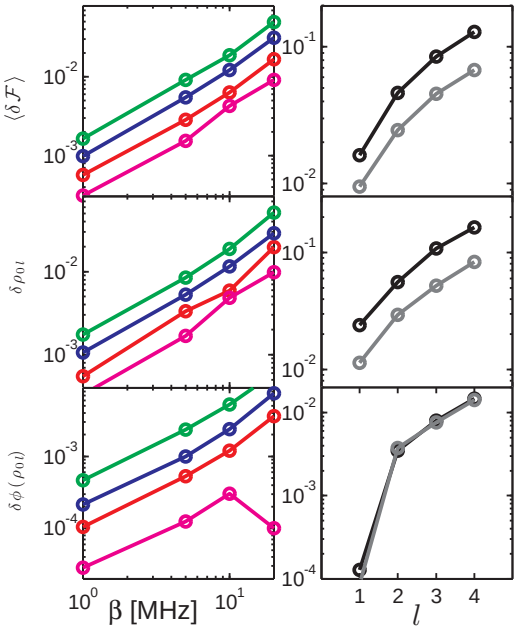

Figure S4: Systematic errors in the extracted density matrix, for initial states $|\psi_l\rangle = (|0\rangle + |l\rangle)/\sqrt{2}$, where $l = 1, 2, 3, 4$. The left column of plots shows the fidelity error, and errors in the off-diagonal matrix element ρ_{0l} as a function of anharmonicity, in the absence of decoherence. The magenta, red, blue and green lines correspond to the initial state $l = 1, 2, 3, 4$ respectively. The same error measures are plotted in the right column, as a function of initial state l, with decoherence included, at $\beta/2\pi = 20\,\mathrm{MHz}$. The black and grey lines correspond to decoherence parameters $T_1 = 150\,\mathrm{ns}$, $T_2 = 200\,\mathrm{ns}$ and $T_1 = 600\,\mathrm{ns}$, $T_2 = 600\,\mathrm{ns}$ respectively.

We plot the errors in the Wigner distribution and the extracted density matrices, as a function of the anharmonicity β, initial state and decoherence parameters. The density matrix is extracted, as in the experimental data, using only the populations of the lowest 6 levels.

To quantify the error in the Wigner distribution, we calculate the cross-correlation (at zero offsets) between the ideal Wigner distribution of the initial state, and the one obtained from the expectation value of the parity operator after a set of displacements. The cross correlation is defined as $C(f(x,y),g(x,y)) = \sum_{x',y'} \frac{\left(f(x',y')-\langle f\rangle\right)\left(g(x',y')-\langle g\rangle\right)}{\sqrt{\sum_{x'',y''}(f(x'',y'')-\langle f\rangle)^2 \sum_{x,y}(g(x'',y'')-\langle g\rangle)^2}}$, where $\langle f\rangle, \langle g\rangle$ are the average values of f, g. The results are plotted in Fig. S3. As seen in the figure, the error increases sharply with both anharmonicity and maximal populated level l. To quantify the error in the extracted density matrices, we use the standard fidelity definition: $\langle \delta\mathcal{F}\rangle = 1 - Tr\sqrt{\sqrt{\rho^F}\rho^I\sqrt{\rho^F}}$, where ρ^I is the initial density matrix and ρ^F is the density matrix obtained from a fit to the populations of the displaced states. We choose the initial density matrix ρ^I to be $|\psi_l\rangle\langle\psi_l|$, where $|\psi_l\rangle = (|0\rangle + |l\rangle)/\sqrt{2}$. In addition, we calculate the error in the non-zero off-diagonal elements of the density matrix using the following definitions: $\delta\rho_{0l} = \left||\rho^F_{0l}| - |\rho^I_{0l}|\right|$ is the error in the amplitude of the matrix element, and $\delta\phi(\rho_{0l}) = \left|\phi\left(\rho^F_{0l}\right) - \phi\left(\rho^I_{0l}\right)\right|/2\pi$ is the normalized error in the phase of the matrix element.

As seen in the Fig. S4, all the error measures are negligible (smaller than ~ 0.05) when decoherence is neglected. However, when included, both the fidelity and the amplitude of the off-diagonal elements have non-negligible errors. For currently available samples having $T_1 > 600\,\text{ns}$, and correspondingly longer T_2, the errors due to decoherence can be substantially reduced; In this case, the errors become smaller than 0.1 in all the measures. As expected, the phase error is quite insensitive to decoherence.

III. SHOT NOISE IN WIGNER TOMOGRAPHY VS. STANDARD STATE TOMOGRAPHY

As pointed out in the manuscript, phase space is a convenient basis to experimentally acquire global information about the state (e.g. phase distribution, average energy etc.) *fast*, at the expense of accurate knowledge of the state in the eigenstate basis. Therefore, when extracting the density matrix in the latter basis from a Wigner measurement, one requires an excess number of measurements compared to standard state tomography (SST) to achieve similar uncertainties. It turns out that this is only true for a certain class of states. For states that are dispersed in phase space, such as states composed of a coherent superposition of a small number of eigenstates, SST requires significantly less tomography pulses than Wigner tomography (WT) to achieve a comparable error in the density matrix but the *same* number of pulses for states that are localized in phase space. The overhead in the number of measurements is an important parameter from an experimental standpoint, and therefore we perform numerical simulations to calculate it. In the following we describe our simulation methods for both cases.

WT. We start with a pure initial state $|\psi\rangle$. We use the same method described in Sec. II to calculate the density matrix after a coherent displacement. We keep the diagonal elements of the final density matrix, for a set of N_W random displacements α that are uniformly distributed in the complex plane. For each experiment (a particular displacement) we calculate an ensemble of M possible outcomes for the measurements of the diagonal elements, assuming r repetitions in the experiment, and a binomial distribution for the escape probabilities, from which the diagonal elements $P(n)$ are calculated. For each outcome, we extract the density matrix and calculate its fidelity $\mathcal{F} = Tr\sqrt{\sqrt{\rho^{tom}}\rho^{ideal}\sqrt{\rho^{tom}}}$. We then calculate the average fidelity $\langle\mathcal{F}\rangle$ of the ensemble to find the experimental error $\langle\delta\mathcal{F}\rangle = 1 - \langle\mathcal{F}\rangle$ due to shot noise.

SST. We start with a pure initial state $|\psi\rangle$. We construct a set of $N_{SST} = d^2$ orthogonal, unitary operations U_j to span a d-level subspace. The set of operations is chosen to be the generators of $SU(n)$ for convenience. From the diagonal elements of the rotated density matrices, we are able to extract the expectation values of each generator $\langle U_j\rangle = Tr(\rho U_j)$, and therefore reconstruct the original density matrix: $\rho = \sum_j \langle U_j\rangle \mathbf{U}_j$ [4]. As before, for each diagonal we calculate an ensemble of M possible measurement outcomes due to shot noise. From each of the measurements in the ensemble, we extract the expectation values of the operators U_j and calculate the corresponding density matrix. The average fidelity error of the ensemble of density matrices is then evaluated.

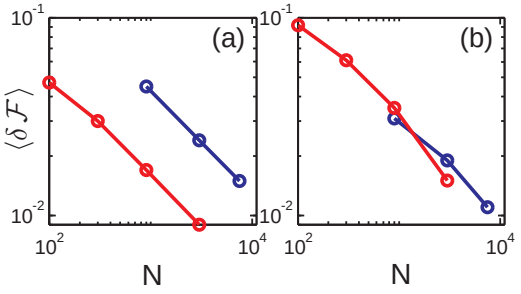

Figure S5: Error in the fidelity due to shot noise in standard state tomography and WT. $\langle\delta\mathcal{F}\rangle$ vs. N in WT (blue) and SST (red) of the initial states $|\psi\rangle = (|0\rangle + |4\rangle)/\sqrt{2}$ (a) and $|\psi\rangle = \left(1/\sqrt{5}\right)\sum_{k=0}^{4}|k\rangle$ (b). Both follow the $1/\sqrt{N}$ trend.

Figure S5 shows the results of both simulations. We plot $\langle\delta\mathcal{F}\rangle$ as a function of $N = R$, where R is the number of repetitions of the experiment per tomography pulse. In the WT simulation we define N relative to the SST case:

$N = R\left(\frac{N_W}{N_{SST}}\right)$, where N_W is the number of tomography pulses in the Wigner simulation, and N_{SST} is the number of tomography pulses in SST. While N_{SST} is fixed, we vary N in the WT simulation by using many displacement pulses while fixing R, and in the SST simulation we vary N by changing the statistics R. In all Wigner simulations we fix R to be 900 (as in our experiment) and in the SST simulations we vary R from 100 to 3000.

We see that for an initial state, composed of a coherent superposition of only 2 states (Fig. S5a), SST outperforms WT by a factor of 8, in terms of the fidelity $\mathcal{F}$. In contrast, for states composed of a coherent superposition of 5 eigenstates, the amount of information that is extracted per pulse in SST and WT is similar. This is because our chosen state is partially localized in phase space, and therefore requires less displacement pulses to extract its properties.

[1] J. M. Martinis, S. Nam, J. Aumentado, C. Urbina, Phys. Rev. Lett., **89**, 117901 (2002).
[2] Y. Shalibo et al. Phys. Rev. Lett., **108**, 037701 (2012).
[3] M. Hofheinz *et al.*, Nature, **459**, 546 (2009).
[4] R. T. Thew, K. Nemoto , A. G. White, W. J. Munro, Phys. Rev. A, **66**, 012303 (2002).