

# Theory-independent monitoring of the decoherence of a superconducting qubit with generalized contextuality

Received: 6 December 2024

Accepted: 23 January 2026

Published online: 07 February 2026

 Check for updates

Albert Aloy <sup>1,2</sup> , Matteo Fadel <sup>3</sup> , Thomas D. Galley<sup>1,2</sup> ,  
Caroline L. Jones<sup>1,2</sup>  & Markus P. Müller <sup>1,2,4</sup> 

Characterizing the nonclassicality of quantum systems under minimal assumptions is an important challenge for quantum foundations and technology. Here we introduce a theory-independent method of process tomography and perform it on a superconducting qubit. We demonstrate its decoherence without assuming quantum theory or trusting the devices by modelling the system as a general probabilistic theory. We show that the superconducting system is initially well-described as a quantum bit, but that its realized state space contracts over time, which in quantum terminology indicates its loss of coherence. The system is initially nonclassical in the sense of generalized contextuality: it does not admit of a hidden-variable model where statistically indistinguishable preparations are represented by identical hidden-variable distributions. In finite time, the system becomes non-contextual and hence loses its nonclassicality. Moreover, we demonstrate in a theory-independent way that the system undergoes non-Markovian evolution at late times. Our results extend theory-independent tomography to time-evolving systems, and show how important dynamical physical phenomena can be experimentally monitored without assuming quantum theory.

Demonstrating that some quantum systems have properties that genuinely defy classical explanation is at the forefront of current theoretical and experimental research. The strongest form of nonclassicality, Bell nonlocality<sup>1</sup>, allows us to refute local hidden-variable models for experiments on spatially separated quantum systems. It is often experimentally more tractable to focus on single systems, avoiding the need for spacelike separation, and the Kochen-Specker theorem<sup>2</sup> precludes the existence of noncontextual hidden-variable models for such systems of Hilbert space dimension at least three.

The Kochen-Specker theorem relies, however, on several undesirable assumptions: it assumes the validity of quantum theory, the fact that the performed measurements are noiseless projective measurements, and that the outcomes depend deterministically on

the underlying hidden variables. The latter two assumptions in particular are detrimental to the goal of the practical certification of nonclassicality. Spekkens' notion of generalized noncontextuality overcomes these difficulties<sup>3</sup>. In a nutshell, it stipulates that statistically indistinguishable operations (measurements, preparations or transformations) should have identical representations on the hidden-variable level. Generalized contextuality admits of robust experimental detection<sup>4–6</sup>, and subsumes a variety of natural notions of nonclassicality (such as Wigner negativity<sup>7,8</sup>) into a single simple criterion.

Here, we certify the generalized contextuality of a superconducting qubit experimentally, directly from the measurement statistics and without assuming quantum theory. In contrast to earlier

<sup>1</sup>Institute for Quantum Optics and Quantum Information, Austrian Academy of Sciences, Vienna, Austria. <sup>2</sup>Vienna Center for Quantum Science and Technology (VCQ), Faculty of Physics, University of Vienna, Vienna, Austria. <sup>3</sup>Department of Physics, ETH Zürich, Zürich, Switzerland. <sup>4</sup>Perimeter Institute for Theoretical Physics, Waterloo, ON, Canada. ✉e-mail: [Albert.Aloy@oeaw.ac.at](mailto:Albert.Aloy@oeaw.ac.at); [fadelm@phys.ethz.ch](mailto:fadelm@phys.ethz.ch); [Thomas.Galley@oeaw.ac.at](mailto:Thomas.Galley@oeaw.ac.at); [CarolineLouise.Jones@oeaw.ac.at](mailto:CarolineLouise.Jones@oeaw.ac.at); [Markus.Mueller@oeaw.ac.at](mailto:Markus.Mueller@oeaw.ac.at)

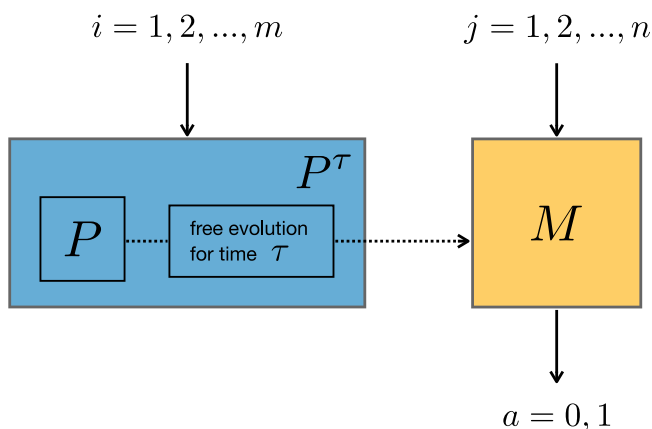
experimental detection of Kochen-Specker contextuality in a superconducting platform<sup>9</sup>, our analysis thus makes significantly weaker assumptions. We analyze how the amount of contextuality changes over time—in particular, the system becomes noncontextual in finite time due to decoherence. In fact, we do more than this: we characterize the system completely, for different evolution times, by generalizing theory-agnostic tomography<sup>10,11</sup>. That is, by performing many preparation and measurement procedures, we determine its space of states and of effects (measurement outcomes) which, when combined linearly, reproduce the experimental statistics. Such pairs are known as generalized probabilistic theories (GPT)<sup>12–15</sup>, with quantum theory and its qubit as a special case. For a quantum bit, the state space is the three-dimensional Bloch ball, but a GPT's state space can be any convex set of any dimension. Without assuming the validity of quantum theory, we determine directly from the data that our superconducting system's normalized state space is likely three-dimensional and in shape close to a ball. Observing the contraction of this “bumpy Bloch ball” yields a theory-independent monitoring of the quantum decoherence process, of its loss of contextuality, and of its non-Markovianity<sup>16</sup> at late times.

Superconducting qubits<sup>17</sup> admit the rapid implementation of a very large number of preparation and measurement procedures, which is necessary for the data-demanding sort of process tomography that we implement here. Our results hence demonstrate the suitability of the superconducting system for this sort of theory-independent analysis.

## Results

### Experimental setup

Our experiments are performed on a transmon superconducting qubit<sup>18</sup> with frequency  $\omega_q = 2\pi \cdot 5.05$  GHz, hosted inside a three-dimensional readout cavity resonator with frequency  $\omega_r = 2\pi \cdot 8.56$  GHz. Strong dispersive coupling between the two allows us to measure the qubit's state by performing a transmission measurement on the resonator to probe its qubit-state-dependent resonance frequency, as standard in circuit-QED<sup>19</sup>. This signal is amplified by both cryogenic and room temperature amplifiers, which give a readout fidelity of 85(1)%. We measure a qubit energy relaxation rate  $T_1 = 21.9 \pm 0.4$   $\mu$ s and Ramsey decoherence rate  $T_2 = 12.7 \pm 0.6$   $\mu$ s, as well as a resonator frequency shift  $\chi = -2\pi \cdot 1.89$  MHz when the qubit is excited. Preparation of the initial state is achieved by applying a 200 ns resonant Rabi pulse with well-defined amplitude and phase to the qubit  $t = 0$  state  $|0\rangle$ . This amplitude and phase are chosen from a list  $\mathcal{V}_m$



**Fig. 1 | Diagram of the prepare-transform-measure setup of the experiment.** We regard the actual state preparation and the subsequent time evolution as a single preparation procedure, which is simply a convenient convention (analogous to the choice of Schrödinger versus Heisenberg picture). As a result, the prepared states will depend on the waiting time  $\tau$ , but the measurements will not.

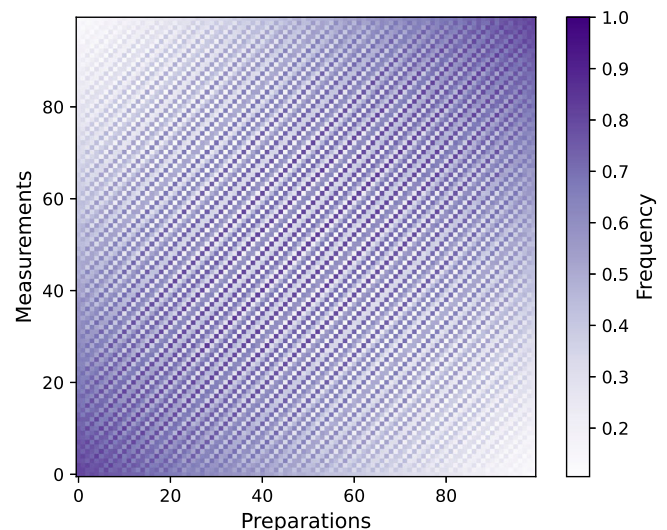
defining the space of  $m$  states. In order to have a uniform distribution of  $m$  points on the surface of the Bloch sphere, we chose a Fibonacci distribution<sup>20</sup>. We then wait for a variable time  $\tau$ , after which a second Rabi pulse is applied in order to specify the measurement basis. Again, amplitude and phase of the pulse are chosen from a list  $\mathcal{V}_n$ , now specifying the space of effects. In our case,  $m = n = 100$ , and the two lists of preparations and measurement directions coincide. Finally, the qubit's state is measured in the computational  $z$ -basis  $\{|0\rangle, |1\rangle\}$  by the standard circuit-QED readout technique in the strong-dispersive-coupling regime, as mentioned above. For each combination of state preparation and measurement, this sequence is repeated 2000 times in order to average over the qubit projection noise. This gives us the frequencies  $F_{ij}$  from which we estimate the probabilities  $p(O|P_i, M_j)$  entering Eq. (3) below.

We describe the experimental setup and the qubit readout characterization procedure in more detail in the Methods' subsection “Experimental setup and qubit readout characterization”.

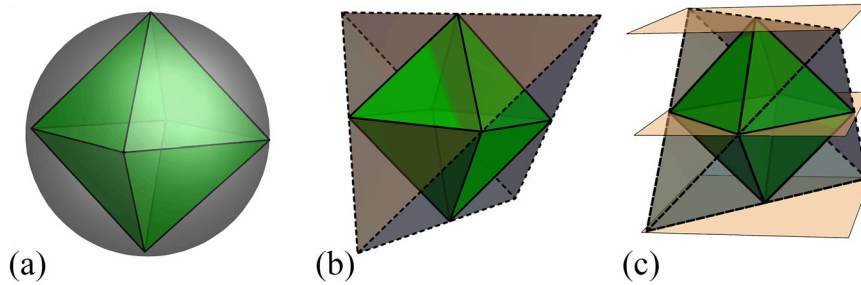
### Theory-independent analysis

To provide a theory-independent analysis, we view the experimental setup as a prepare-and-measure scenario, see Fig. 1. We have  $m$  possible preparation procedures  $P_i$  and  $n$  measurement procedures  $M_j$ , each with outcome  $a \in \{0, 1\}$ . Moreover, we have the option to introduce a waiting time  $\tau$  between preparation and measurement. Without loss of generality, we choose the convention that this evolution for time  $\tau$  is considered a part of the preparation procedure, yielding effective preparation procedures  $P_i^\tau$ . Thus, we probe the system (the superconducting qubit) to estimate the conditional probabilities  $p(a|P_i^\tau, M_j)$ . By normalization, it is sufficient to obtain the statistics for one of the outcomes which we denote by  $a = 0$  – shown in Fig. 2 for  $\tau = 0$ . Statistics are gathered for all the possible  $i, j$  and for  $\tau = \{0, 5, 10, 15, 20, 30, 40, 50\}$   $\mu$ s. For every fixed value of  $\tau$ , this defines a corresponding *operational theory*<sup>3</sup>: a collection of preparation and measurement procedures together with a probability rule that describes the statistical properties of the laboratory system.

Following this operational approach, generalized noncontextuality<sup>3,21</sup> requires that procedures which are statistically equivalent at the operational level (i.e., producing identical statistics if all other processes are fixed) must also be identically represented in the underlying hidden-variable model. A hidden-variable model (also



**Fig. 2 | Frequency data collected in the experiment.** Experimentally measured frequency of occurrence of the outcome  $a = 0$  for the  $m \times n$  table of preparations and measurements, for  $\tau = 0$ .



**Fig. 3 | Relations between the state spaces of the qubit, stabilizer qubit<sup>31</sup> and classical four-level system.** **a** The state space of the stabilizer qubit (in green) is embedded in the state space of the qubit, given by the gray ball; **b** the state space of the stabilizer qubit (in green) is embeddable within a tetrahedron (in gray), the state space of a classical four level system; **c** the embedding of the stabilizer qubit into the tetrahedron pictured here is such that any valid effect on the stabilizer qubit is

also a valid effect on the tetrahedron. This requirement prevents, for example, the embeddability of the qubit into the tetrahedron (we know that the qubit is non-embeddable because it is contextual). Here the orange planes represent the effect corresponding to the +1 outcome of the Pauli Z measurement: the upper plane intersects all states giving probability 1 for the outcome +1, the central plane those giving probability  $\frac{1}{2}$  and the lower plane those giving probability 0.

known as an ontological model) of an operational theory is defined on a measurable space  $\Lambda$ . To each preparation  $P$  in the theory, there corresponds a probability distribution  $\mu_P(\lambda)$  on  $\Lambda$ , and to every outcome  $a$  of every measurement  $M$ , there corresponds a response function  $\chi_{a|M}(\lambda)$ , such that  $\sum_a \chi_{a|M}(\lambda) = 1$ . The assignments are such that the probabilities are reproduced by the laws of classical probability theory,

$$p(a|P, M) = \int_{\Lambda} \chi_{a|M}(\lambda) \mu_P(\lambda) d\lambda. \quad (1)$$

A hidden-variable model is preparation-noncontextual<sup>3,21</sup> if any two operationally equivalent preparations  $P$  and  $P'$  have  $\mu_P(\lambda) = \mu_{P'}(\lambda)$ . Two preparations  $P$  and  $P'$  are operationally equivalent if

$$p(a|P, M) = p(a|P', M) \text{ for all } a, M. \quad (2)$$

Similarly, the model is measurement-noncontextual if operationally equivalent measurement outcomes (i.e., ones that have identical probabilities for all preparation procedures) are represented by the same response function.

For our scenario, we may not only consider the actually implemented procedures  $M_i$  and  $P_j$ , but also statistical mixtures of those, such as the preparation procedure  $P_{\text{mix}}$  which results in implementing either preparation  $P_1$  or  $P_2$  with probability  $\lambda$  or  $1 - \lambda$ , respectively. We can then represent operationally equivalent preparation procedures  $P$  and  $P'$  by some state  $\mathbf{s}_P = \mathbf{s}_{P'}$ , an element of some vector space over the real numbers. Statistical mixtures are then represented by convex combinations, such that, for example,  $\mathbf{s}_{P_{\text{mix}}} = \lambda \mathbf{s}_{P_1} + (1 - \lambda) \mathbf{s}_{P_2}$ . Similarly, we can represent operationally equivalent measurement-outcome pairs  $(a, M)$  and  $(a', M')$  as so-called effects  $\mathbf{e}_{a,M} = \mathbf{e}_{a',M'}$ , elements of the dual space, such that  $p(a|P, M) = \langle \mathbf{e}_{a,M}, \mathbf{s}_P \rangle$ . All possible states of a system form a convex set  $S$ , and all possible effects yield another convex set  $\mathcal{E}$ . This defines a generalized probabilistic theory (GPT)<sup>14,15</sup>, sometimes also called a “GPT system”. Please see the “Methods” section for more details, and for how a quantum bit is described in this formalism. In particular, its space of pure states is the famous Bloch sphere, and the full space  $S$  of all (pure and mixed) states is the Bloch ball.

Quantum and classical systems are examples of GPT systems, but the framework contains many more exotic possibilities, such as theories with superstrong nonlocality<sup>12</sup> and higher-order interference<sup>22,23</sup>. Originally, interest in GPTs originated in the research program to reconstruct the abstract formalism of quantum theory from simple physical principles<sup>24–27</sup>. Here, however, we use the GPT formalism as a tool for a theory-independent description of our experimental statistics. In particular, knowing the GPT associated to an operational theory allows us to determine whether it admits of a preparation- and measurement-noncontextual hidden-variable model, using the

criterion of *simplex embeddability* demonstrated in ref. 28. If it does admit of such a model, we say that the physical system is non-contextual. This can be checked with a linear program<sup>29,30</sup>.

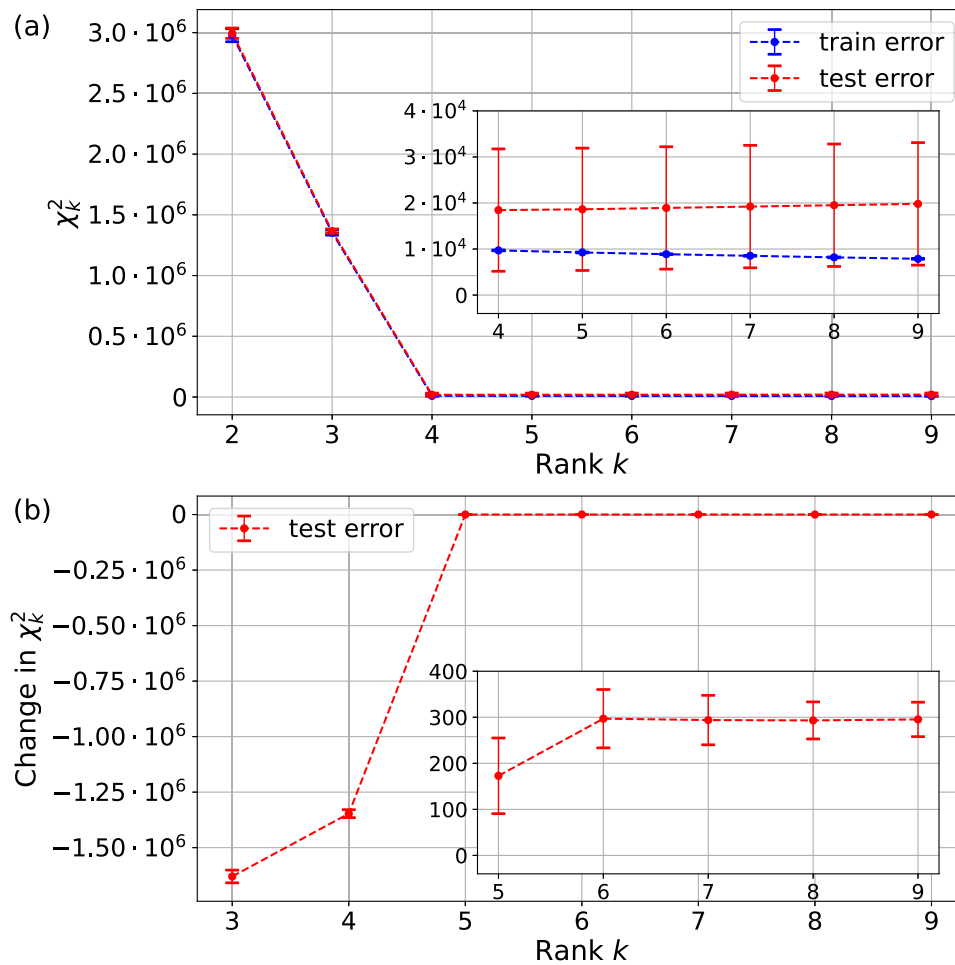
For example, it can be checked that the qubit GPT system is contextual, i.e. cannot be embedded into any classical probability simplex, in contrast to the stabilizer qubit, a GPT system defined by restricting the qubit to convex combinations of its stabilizer states and to stabilizer measurements<sup>31</sup>, as illustrated in Fig. 3. Hence, the stabilizer qubit is noncontextual (when restricted to preparations and measurements) or “classical” in this sense. This is related to the Gottesman-Knill theorem, establishing the efficient classical simulatability of stabilizer quantum computation<sup>32</sup>. More generally, generalized contextuality is a resource for a number of information processing tasks<sup>33–38</sup>.

The obstacle to applying this directly to our experiment is that we do not know the probabilities  $p(a|P, M)$ , but only experimentally determined approximations, namely the frequencies of occurrence of the outcomes (depicted in Fig. 2). Hence, we cannot directly write down the GPT that describes our superconducting system, but have to estimate the GPT from the experimental data. We do so by modifying and generalizing the method of theory-agnostic tomography<sup>10,11</sup>, which leads to the following multi-step procedure.

Let us first discuss the procedure for a fixed value of waiting time  $\tau$ ; say,  $\tau = 0$ . We start by collecting the experimental statistics for a large set of possible preparations  $P_i$  and two-outcome measurements  $M_j$ , with  $i \in \{1, \dots, m\}$  and  $j \in \{1, \dots, n\}$ . The goal is to estimate the conditional probabilities  $p(0|P_i, M_j)$  for obtaining outcome  $a = 0$ , given preparation  $P_i$  and measurement  $M_j$ . We organize the collection of statistics in the form of an  $m \times n$  matrix  $D$  as:

$$D = \begin{pmatrix} p(0|P_1, M_1) & p(0|P_1, M_2) & \dots & p(0|P_1, M_n) \\ p(0|P_2, M_1) & p(0|P_2, M_2) & \dots & p(0|P_2, M_n) \\ \vdots & \dots & \ddots & \vdots \\ p(0|P_m, M_1) & p(0|P_m, M_2) & \dots & p(0|P_m, M_n) \end{pmatrix}. \quad (3)$$

In an actual experiment, the conditional probabilities  $p(0|P_i, M_j)$  are estimated by running the experiment  $N$  times. Thus, when talking about experimental data, instead of the matrix  $D$  we will refer to a matrix  $F$  containing the observed frequencies  $F_{ij} = f(0|P_i, M_j)$ . i.e.,  $F$  is a frequency table. Given that there are  $N$  runs for each pair of preparation and measurement  $(P_i, M_j)$ , we have  $F_{ij} = \frac{N_{ij}}{N}$ , where  $N_{ij}$  is the number of outcomes 0 observed for the pair  $(P_i, M_j)$ . In the case where a large number  $M$  of frequency tables  $\{F^q\}_{q=1}^M$  are obtained, one can directly compute the variance of a frequency table  $F^q$  as  $(\Delta F^q_{ij})^2 = (F^q_{ij} - \bar{F}_{ij})^2$ , where  $\bar{F}_{ij} = \frac{1}{M} \sum_{q=1}^M F^q_{ij}$ . In the case where one does not have a large



**Fig. 4 | Errors of the optimal GPT fits for different ranks.** **a** Test and train errors for the optimal GPT fits for ranks  $k \in \{2, \dots, 9\}$ . Inset: zoomed in for ranks  $k \geq 4$ . The training error is evaluated on 10 frequency tables  $\{F^\alpha\}_{\alpha=1}^{10}$ , whilst the test error is evaluated on the 90 pairs of frequency tables  $\{F^\alpha, F^\beta\}$  with  $\alpha \neq \beta$ . The error bars for the training error and test error are given by the statistical uncertainty over the 10

tables and 90 pairs, respectively. **b** For each pair  $\{F^\alpha, F^\beta\}$  with  $\alpha, \beta \in \{1, \dots, 10\}$ ,  $\alpha \neq \beta$  with  $F^\alpha$  serving as the training data and  $F^\beta$  as the testing data we plot  $\chi_k^2(F^\beta, D_k^\alpha) - \chi_{k-1}^2(F^\beta, D_{k-1}^\alpha)$  which is the change in the test error between the best fit rank  $k$  and best fit rank  $k - 1$  models. For  $k \leq 4$  it is strictly negative, showing that the test error decreases, while for  $k > 4$  it is strictly positive, showing that the test error increases.

number of frequency tables, we can estimate  $(\Delta F_{ij})^2$  by making the following assumptions<sup>10</sup>. We expect that in the limit of  $N \rightarrow \infty$ , these frequencies converge to the conditional probabilities  $p(0|P_i, M_j)$ , under the assumption of ideal repeatability disregarding drift and other imperfections. In this case, the variable  $N_{ij}$  is the number of 0 outcomes of  $N$  independent events in a sequence of experiments. This can be modeled as a binomial distribution with  $N$  events and probability given by the frequency  $F_{ij} = \frac{N_{ij}}{N}$ . The variance in  $N_{ij}$  is therefore  $\frac{N_{ij}(N-N_{ij})}{N}$ , which implies that the variance in the frequency is therefore  $(\Delta F_{ij})^2 = \frac{N_{ij}(N-N_{ij})}{N^3} = \frac{F_{ij}(N-N_{ij})}{N^2} = \frac{F_{ij}(1-F_{ij})}{N}$ .

The matrix  $D$  contains all relevant statistical information about the prepare-and-measure experiment. For every GPT, the probabilities  $p(0|P_i, M_j)$  are given by  $\langle \mathbf{e}_j, \mathbf{s}_i \rangle$ , where the effects  $\mathbf{e}_j$  describe the measurement and its outcome, and the states  $\mathbf{s}_i$  describe the preparation procedures. States are elements of some real vector space of unspecified dimension  $k$  and effects elements of the dual space, with  $\langle \cdot, \cdot \rangle$  denoting the natural pairing, i.e., the application of the covector  $\mathbf{e}_j$  to the vector  $\mathbf{s}_i$ , yielding a real number. The sets of all possible states and effects define the GPT and determine its information-theoretic and physical behavior. Our goal now is to find the GPT state and effect spaces that best fit for  $D$  while minimizing the number of parameters. That is, we want to find a fitting GPT model which assigns a  $k$ -dimensional state vector  $\mathbf{s}_i$  to each preparation  $P_i$  and a  $k$ -dimensional effect

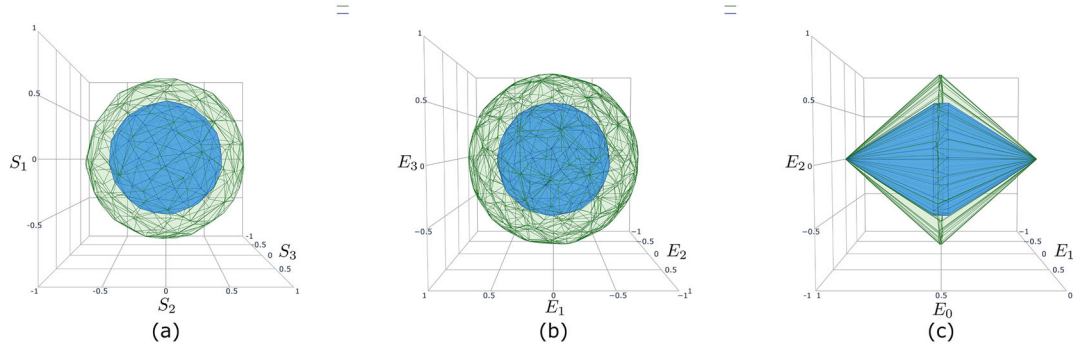
vector  $\mathbf{e}_j$  to each measurement outcome 0 of  $M_j$ . Thus, the  $m \times k$  matrix  $S = (\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_m)^\top$  and the  $k \times n$  matrix  $E = (\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n)$  can be used to factorize the data table  $D = SE$ , resulting in an  $m \times n$  matrix of rank  $k$ , this time with entries  $D_{ij} = \mathbf{s}_i^\top \mathbf{e}_j$  corresponding to the probabilities  $p(0|P_i, M_j)$  predicted by the GPT.

To perform theory-agnostic tomography in our experiment, we first have to determine the appropriate rank  $k$ . To this end, we fit a rank  $k$  matrix  $D = SE$  to the experimentally obtained frequency table  $F$  for various values of  $k$ . The best rank  $k$  matrix  $D$  is determined via minimization of the weighted  $\chi_k^2$  in the following optimization problem<sup>10</sup>:

$$\begin{aligned} \min_{D \in M_{mn}} \quad & \chi_k^2 = \sum_{ij} \frac{(F_{ij} - D_{ij})^2}{\Delta F_{ij}^2} \\ \text{subject to} \quad & \text{rank}(D) = k, 0 \leq D_{ij} \leq 1. \end{aligned} \tag{4}$$

In the Methods' subsection "Solving the weighted low-rank approximation problem of Eq. (4)", we describe how we solve this nonconvex optimization problem numerically, following the method presented in ref. 10.

The experimental data consists of ten frequency tables  $\{F^\alpha\}_{\alpha=1}^{10}$  for the 100 preparations and measurements specified in section "Experimental setup". For every  $k \in \{2, \dots, 9\}$ , the average optimal  $\chi_k^2$  over  $F^\alpha$  is plotted in blue in Fig. 4a. It decreases sharply for  $k < 4$  and decreases slowly for  $k \geq 4$ . The large  $\chi_k^2$  values for  $k < 4$  indicate that the models  $D_k$



**Fig. 5 | Reconstructed state and effect spaces for zero time delay.** Plots showing the realized (blue) and consistent (green) spaces for the rank  $k = 4$  GPT for  $\tau = 0$ . **a** The normalized state spaces are three-dimensional and similar to Bloch balls;

**b** effect space projection onto dimensions 1,2,3; **c** effect space projection onto dimensions 0,1,2 (or similarly for 0,2,3 and 0,1,3) containing the zero and unit effects 0 and **1**.

have too few parameters to properly account for the regularities in the data; they underfit the data. For  $k > 4$ , the  $\chi^2_k$  decrease at a slower rate as the extra parameters in the models fit to the statistical fluctuations in the frequency tables. The inflection point at  $k = 4$  indicates that this is likely the rank of the underlying ‘true’ GPT generating the probability table  $D$  of Eq. (3).

Fitting to statistical fluctuations is a signature of overfitting of a model. Following ref. 11, we can detect overfitting of a model by evaluating the optimal rank  $k$  fit  $D_k$  obtained from a frequency table  $F^a$  on a different frequency table  $F^b$ .  $F^a$  is known as the *training data* (since it is used to train the model  $D_k$ ) and  $F^b$  is known as the *test data* (since it is used to evaluate the model  $D_k$ ). Since the statistical fluctuations of  $F^a$  and  $F^b$  are independent, the more a model  $D_k$  fits the statistical fluctuations in  $F^a$ , the worse it will perform when evaluated on  $F^b$ . Given a training data/test data pair  $(F, F')$  the test error relative to a model  $D$  is

$$\chi_k^{\text{test}} = \sum_{ij} \frac{(F'_{ij} - D_{ij})^2}{(\Delta F'_{ij})^2}. \tag{5}$$

The test error is expected to be high for models which underfit the data (since they have too few parameters to adequately capture the structure of the data), and expected to decrease as  $k$  approaches the true underlying rank  $k_{\text{true}}$ . For  $k > k_{\text{true}}$  the test error is expected to increase. As such, the test error is expected to be minimal for the value of  $k$  which neither underfits nor overfits the data, which is expected to be  $k = k_{\text{true}}$ .

Figure 4a shows the average test error for each  $k \in \{2, \dots, 9\}$  evaluated for every pair  $(F^a, F^b)$  ( $a \neq b$ ), where  $F^a$  is the training data and  $F^b$  is the test data. The large statistical variances in the test error in Fig. 4a do not allow us to infer that the test error reaches its minimum at rank 4. In Fig. 4b, we plot the average difference in test error  $\chi_k^2(F^b, D_k^a) - \chi_{k-1}^2(F^b, D_{k-1}^a)$ . The error bars are sufficiently small to allow us to conclude that the change in test error is negative for  $k \leq 4$  and positive for  $k > 4$ , allowing us to conclude that the minimal test error occurs for  $k = 4$ , and thus that the rank of the underlying true GPT is likely  $k_{\text{true}} = 4$ .

**Dynamical analysis: decoherence**

To determine the dimension of the GPT system, we restrict our attention to the  $\tau = 0$  data. In Fig. 4a, we have computed the optimal models fitting the experimental data  $F$  for different rank  $k$  candidates with  $k \in \{2, \dots, 9\}$ . In particular, following Eq. (4), we minimize  $\chi_k^2$  for each candidate  $k$  obtaining the best-fit GPT models of rank  $k$  to the obtained experimental data.

We have seen that solving Eq. (4) provides a model with the best-fit of rank  $k$  to the probability matrix  $D$  sampling the observed measurement outcomes. Moreover, we have observed that rank  $k = 4$  provides the best estimate of an underlying model describing the

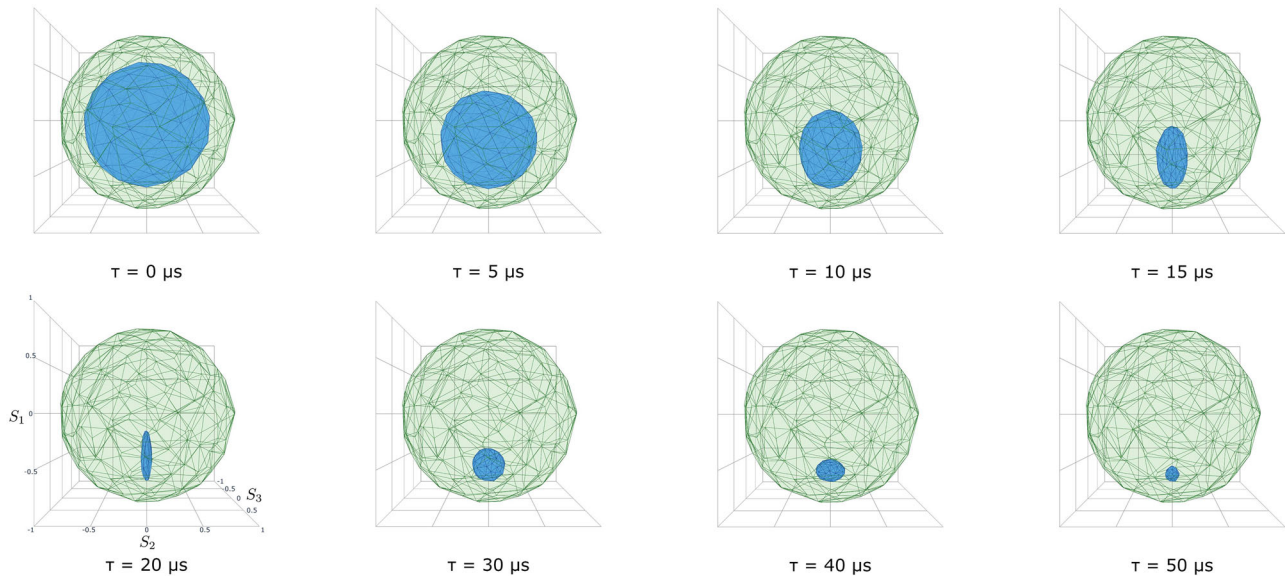
experiment. This could correspond to a qubit description as expected: the dimension of the set of normalized states will then be  $k - 1 = 3$ , which coincides with the dimension of a qubit’s Bloch ball state space.

Let us now characterize the state and effect spaces for the GPT model of rank  $k = 4$ , which generate  $D$ . To obtain  $D$ , we have split the problem into estimating the realized GPT states  $S$  and effects  $\mathcal{E}$ . The decomposition  $D = SE$  is not unique: for every invertible  $k \times k$ -matrix  $L$ , we also have  $D = (SL)(L^{-1}E)$ , and all possible decompositions are of this form, see Methods’ subsection ‘‘Uniqueness of the decomposition  $D = SE$ ’’. Following ref. 10, we choose a parametrization where the first component of a state denotes its normalization, and so the first column of  $S$  consists entirely of ones. Furthermore, the first column of  $E$  is chosen to be the unit (normalization) effect. All further freedom in choosing  $L$  only leads to a different linear reparametrization of the resulting GPT, and this does not change any of its physical properties (see the notion of ‘‘equivalent GPTs’’ in Methods’ subsection ‘‘Generalized probabilistic theories’’).

Figure 5a shows the resulting state space in blue, and Fig. 5b, c shows projections of the effect space in blue for one specific parametrization (see Methods’ subsection ‘‘Generalized probabilistic theories’’). As expected,  $S$  resembles a quantum bit Bloch ball, and  $\mathcal{E}$  resembles the set of Hermitian  $2 \times 2$  operators  $E$  with eigenvalues in  $[0, 1]$  (which is  $(k = 4)$ -dimensional). However, in contrast to a perfect quantum bit,  $S$  and  $\mathcal{E}$  are polytope and not exact duals of each other since only finitely many preparations and measurements have been implemented. That is, given our realized set of effects  $\mathcal{E}$ , we can consider the set  $S_{\text{consistent}}$  of all possible vectors  $s$  which give valid probabilities for all measurements, i.e.,  $0 \leq \langle e, s \rangle \leq 1$  for all effects  $e \in \mathcal{E}$ . Then  $S \subseteq S_{\text{consistent}}$ . Similarly, given  $S$ , we can consider the set  $\mathcal{E}_{\text{consistent}}$  of all possible covectors  $e$  that give valid probabilities on all states, and  $\mathcal{E} \subseteq \mathcal{E}_{\text{consistent}}$ . These sets are shown in green in Fig. 5.

For the sake of the argument, assume for a moment that the no-restriction hypothesis holds<sup>39</sup>, as predicted, e.g., by quantum theory: all possible outcome probability rules are implementable effects, and all possible vectors yielding valid probabilities on the possible measurements are implementable states. In the hypothetical (but unrealistic) case that the measurements have been perfectly noiseless and all possible measurements have actually been implemented, this would imply that  $S_{\text{consistent}}$  describes the set of all possible states of the system, and  $S$  being a strict subset of this means that the preparations have been inherently noisy. Realistically, both preparations and measurements are noisy, but a theory-independent analysis cannot tell us whether the error occurred on the level of the former or the latter. All our analysis can tell us is that the no-restriction hypothesis is violated by the experimentally realized GPT system, and the considerations below allow us to quantify how strong the violation is.

Consider the distinguishability of two states  $s$  and  $s'$ , defined as  $D(s, s') := \max_{e \in \mathcal{E}} |\langle e, s \rangle - \langle e, s' \rangle|$ . For ideal quantum systems, this equals



**Fig. 6 | Evolution of the state space for different delay times.** Plots of the consistent state space  $S_{\text{consistent}}$  (green) for all preparations, and realized state spaces  $S^\tau$  (blue) for varying wait times  $\tau = [0, 5, 10, 15, 20, 30, 40, 50]$   $\mu\text{s}$ .

the trace distance between the two density matrices that represent the states,  $\frac{1}{2} \|\rho_s - \rho_{s'}\|_1$ . For any given state  $s$ , consider the function  $f(s) := \max_{s' \in S} D(s, s')$ , then  $0 \leq f(s) \leq 1$ , and this equals 1 if and only if  $s$  is perfectly distinguishable from some other state. As shown in ref. 40, the no-restriction hypothesis implies that  $f(s) = 1$  for all boundary points  $s$  of the state space, the “perfect distinguishability” axiom of ref. 40. In our case, this function takes values in between  $0.681 \pm 0.003$  and  $0.691 \pm 0.004$ , whose minimum on the boundary points gives us a theory-independent quantity saying “how restricted” the effective GPT system is. Under the additional assumption that quantum theory holds, and that the superconducting system is fundamentally described by a qubit, this tells us that there must have been states  $\rho_s$  prepared (i.e.  $s \in S$ ) that are close to pure, in the sense that their largest eigenvalue is  $\lambda_{\max}(\rho_s) \geq 0.846 \pm 0.002$  (and, assuming approximate rotational symmetry as apparent in Fig. 5, all prepared states for  $\tau = 0$  should have this property approximately); see Methods’ subsection “Lower bound on the purity of the prepared states”. Note that  $\lambda_{\max}(\rho_s) = \langle \psi | \rho_s | \psi \rangle$ , i.e., the fidelity between the prepared state  $\rho_s$  and its eigenstate  $|\psi\rangle$  corresponding to its largest eigenvalue.

Let us now turn to the task of monitoring the time evolution of our superconducting system, via snapshots for different evolution times  $\tau$ . We modify and extend the method of theory-agnostic tomography as follows. For a finite set of  $\{\tau_0, \tau_1, \dots, \tau_T\}$ , we can obtain the frequency tables  $\{F^i\}$ . Since the measurements for each  $F^i$  are treated to be the same by convention (recall Fig. 1), we can write a total frequency table  $F$ , which is a  $m(T+1) \times n$  matrix:

$$F = \begin{pmatrix} F^{\tau_0} \\ F^{\tau_1} \\ \vdots \\ F^{\tau_T} \end{pmatrix} \quad (6)$$

The probability table for a rank  $k$ -GPT in this set of prepare-and-measure experiments with shared measurements is

$$D = \begin{pmatrix} D^{\tau_0} \\ D^{\tau_1} \\ \vdots \\ D^{\tau_T} \end{pmatrix} = \begin{pmatrix} S^{\tau_0} \\ S^{\tau_1} \\ \vdots \\ S^{\tau_T} \end{pmatrix} E = SE, \quad (7)$$

with  $S$  an  $m(T+1) \times k$  matrix and  $E$  a  $k \times n$  matrix.

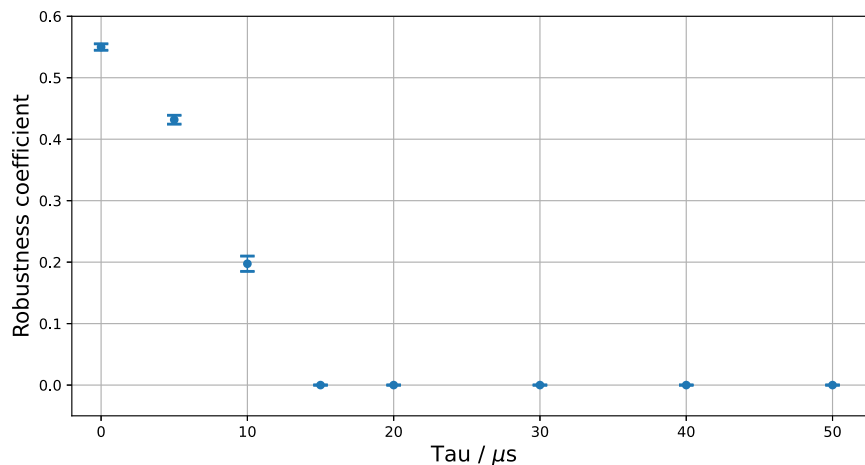
The problem of finding the best rank- $k$  GPT fit for the sequence of prepare-and-measure scenarios  $\{(P^{\tau_0}, \mathcal{M}), (P^{\tau_1}, \mathcal{M}), \dots, (P^{\tau_T}, \mathcal{M})\}$  is thus equivalent to the task of finding the best GPT rank- $k$  GPT fit for  $m(T+1)$  preparations and  $n$  measurements  $\{P', \mathcal{M}\}$ , where  $P' = \bigcup_i P^{\tau_i}$ .

The result is shown in Fig. 6. From our knowledge of quantum physics and the experimental platform, we expect two effects to be in place: decoherence and relaxation to the ground state. Both effects are visible in the theory-independent representation. For every GPT, reversible transformations are represented by linear symmetries of the state space. Hence, under reversible time evolution,  $S$  is preserved. For  $\tau \leq 20 \mu\text{s}$ , we see, however, that this is not the case, and that the time evolution is manifestly irreversible, shrinking  $S$  towards more mixed states. For all times, we see that evolution moves the state space closer to a single distinguished state (potentially a pure state), which describes a relaxation process. Quantum physics intuition tells us that this should correspond to the superconducting qubit’s ground state, but our theory-independent analysis does not allow us to conclude this with certainty.

### Contextuality and its loss under decoherence

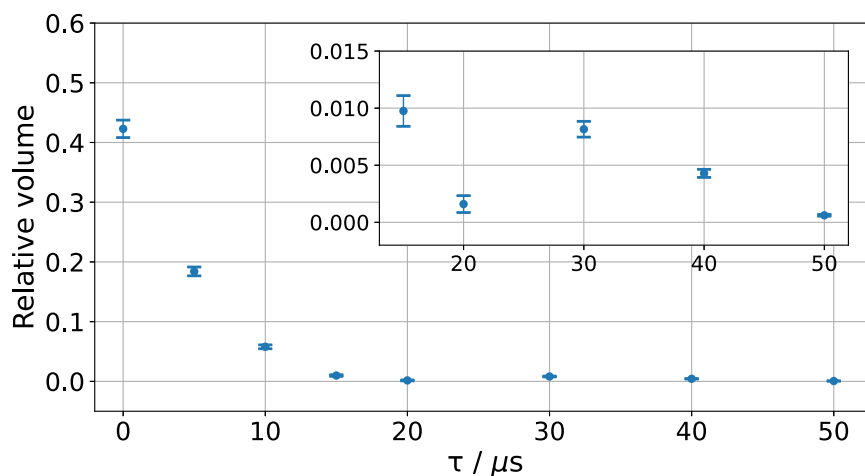
Now that we have a description of the different  $S^\tau$  and  $E^\tau$ , we can check whether the corresponding GPT systems are noncontextual. To do so, we use the linear program of ref. 30 to determine whether the GPT systems are simplex-embeddable. Every system becomes non-contextual if a sufficient amount of noise is added. To quantify this, we use the state  $\mu^\tau$ , which is the average of all extremal states of  $S^\tau$  (think of it as the maximally mixed state in the center of  $S^\tau$ ), and we consider depolarizing noise that replaces every given state by  $\mu^\tau$  with probability  $r$ . The resulting state space  $S_r^\tau := (1-r)S^\tau + r\mu^\tau$  will be noncontextual if  $r$  is large enough. In Fig. 7, we show how large  $r$  has to be chosen such that  $S_r^\tau$  is noncontextual. If the answer is  $r > 0$ , then the system at time  $\tau$  is contextual, and otherwise, if  $r = 0$ , it is noncontextual. The uncertainties in Fig. 7 (resp. Fig. 8) are given by the standard deviations in the robustness (resp. relative volume) computed from 7 frequency tables  $\{F^\alpha\}_{\alpha=1}^7$  of the form given in Eq. (6).

We see that the system is initially contextual, but that it loses its contextuality between the times of  $\tau_2 = 10 \mu\text{s}$  and  $\tau_3 = 15 \mu\text{s}$ . In other words, the decoherence process leads to the system becoming effectively classical and remaining so. Note that the error bars around  $r = 0$  are zero, because small perturbations of noncontextual systems are



**Fig. 7 | Degree of contextuality of the GPT systems for different delay times.** Plot showing the robustness coefficient, i.e., the necessary amount of depolarizing noise  $r$  that has to be added such that a noncontextual ontological model exists for

the GPT system, tested for increasing delay time  $\tau$  (with 7 runs per value of  $\tau$ , with the standard deviation then determining error bars). For robustness  $r = 0$ , the superconducting system is noncontextual, and otherwise contextual.



**Fig. 8 | Signature of non-Markovian time evolution at late times.** Numerically determined relative volumes of the realized to consistent state spaces  $\text{Vol}(\mathcal{S}^\tau)/\text{Vol}(\mathcal{S}_{\text{consistent}})$  for different waiting times  $\tau$ , again with the standard

deviation determining error bars. The volume always decreases, as required by Markovian time evolution, except between  $\tau_4 = 20 \mu\text{s}$  and  $\tau_5 = 30 \mu\text{s}$ .

typically noncontextual too. Indeed, we have obtained  $r = 0$  for  $\tau \geq 15 \mu\text{s}$  in all repetitions.

It is well-known that the quantum bit is noncontextual in the sense of Kochen and Specker<sup>2</sup>, and that it admits of simple hidden-variable models, including one that has already been given by Bell<sup>41</sup>. However, as shown by Spekkens<sup>3</sup>, all such models must be contextual according to the generalized notion introduced in Results' subsection "Theory-independent analysis". As we have demonstrated above, the same is true for the "noisy qubits" that describe the superconducting system for sufficiently small evolution times.

### Non-Markovianity at late times

In Fig. 6, it can be seen that the state space shrinks during all time steps, except between  $\tau_4 = 20 \mu\text{s}$  and  $\tau_5 = 30 \mu\text{s}$ , where it appears to expand. This is a signature of non-Markovianity. Let us define what this means for arbitrary GPTs, including but not restricted to quantum theory. We say that a system  $S$  has Markovian time evolution from time  $\tau_0$  to  $\tau_1 > \tau_0$  if there is a transformation  $T$ , defined on the system  $S$  alone, such that  $\mathbf{s}_S(\tau_1) = T\mathbf{s}_S(\tau_0)$ , with  $\mathbf{s}_S(\tau)$  the state of  $S$  at time  $\tau$  (in QT; this would be a trace-preserving completely positive map). This requirement is

automatically fulfilled if we are able to prepare the system at time  $\tau_0$  in some arbitrary state of  $S$ , uncorrelated with other systems and the environment. Mathematically, the corresponding transformation  $T$  will then be a linear map on the set of unnormalized states (a well-known fact for GPTs). Thus, it must act as an affine-linear map on the set of normalized states  $\mathcal{S}$ . Since it maps states to valid states, it holds that  $T(\mathcal{S}) \subset \mathcal{S}$ , and so it must be volume-non-increasing (see Methods' subsection "Markovian evolutions do not increase the volume" for details).

On the other hand, if there are initial correlations between the system  $S$  and its environment  $E$  at time  $\tau_0$ , then initial state  $\mathbf{s}_{SE}$  and final state  $\mathbf{s}'_{SE}$  are related by some transformation of the total system  $SE$ ,  $\mathbf{s}'_{SE} = T_{SE}\mathbf{s}_{SE}$ . Thus, there will not, in general, be a transformation (or other linear map)  $T$  that relates the possible initial and final marginal states  $\mathbf{s}'_S$  and  $\mathbf{s}_S$ . In the quantum case, this fact has been pointed out by Schmid et al.<sup>42</sup>. Intuitively, this sort of non-Markovianity happens if there is some information backflow from the environment to the system.

We have numerically determined the volumes of the state spaces  $\mathcal{S}^\tau$  (see Fig. 8), and see that the volume increases between times

$\tau_4 = 20 \mu\text{s}$  and  $\tau_5 = 30 \mu\text{s}$ , as is visible in Fig. 6. In particular, for each  $\tau$ , we calculated the volume of the realized state space  $\text{Vol}(S^\tau)$  relative to the volume of the consistent state space  $\text{Vol}(S_{\text{consistent}})$ . The observation that the relative volume increases between  $\tau_4$  and  $\tau_5$  proves that there is non-Markovian evolution during this time interval. This is physically not unexpected, since the transmon is only approximately a qubit, and higher-lying levels cannot be perfectly ignored. For the experiment we consider, the non-Markovianity may thus be due to the residual off-resonant coupling with an additional degree of freedom, such as another qubit hosted in the same readout cavity resonator. On the other hand, successful state preparation at time  $\tau_0 = 0$  implies that time evolution from time  $\tau_0 = 0$  to any other time  $\tau_1 = \tau$  must always be Markovian, and so  $\text{Vol}(S^0) \geq \text{Vol}(S^\tau)$  for all  $\tau \geq 0$ , which is confirmed by our data.

We fit an exponential decay function  $Ae^{-\frac{\tau}{B}}$  to the relative volumes for different  $\tau$  (weighted by the uncertainties in the relative volume), obtaining values  $A = 0.45 \pm 0.01$  and  $B = 0.227 \pm 0.004 \mu\text{s}^{-1}$ . This provides a theory-independent estimate of the relaxation rate  $\frac{1}{B} = 4.409 \mu\text{s}$ .

The two phenomena discussed in this and the previous subsection have very different roles to play: generalized contextuality is a precious resource that is hard to obtain; non-Markovianity, on the other hand, is a frequent statistical phenomenon that is often detrimental rather than desirable. What we claim to be remarkable is not that there is non-Markovianity in the experiment, but that we are able to demonstrate it as a property of the system's time evolution without assuming the validity of quantum theory.

## Discussion

Theory-agnostic tomography for preparations and measurements was introduced in ref. 10, where it was applied to a photonic qubit, and subsequently in ref. 11 to a photonic qutrit. The present work makes use of the methods of refs. 10,11 to construct the state and effect space of the best fit GPT model of the superconducting qubit.

Following ref. 11, we determine the best rank fit as the one which least overfits the data, requiring fewer assumptions than the AIC criterion used in ref. 10. However, unlike ref. 11, the 10 different frequency tables obtained provide us with 90 pairs of training and test data, allowing us to give error bars for the best fit  $\chi^2$  values for each dimension.

In the present work, we extend theory-agnostic tomography of refs. 10,11 to include a form of process tomography. We use this for theory-independent monitoring of decoherence, and to obtain a theory-independent witness of non-Markovianity. We also monitor the time evolution of generalized contextuality of the superconducting qubit, showing that it is contextual for times  $\tau = 0, 5, 10 \mu\text{s}$  and non-contextual for times greater than  $15 \mu\text{s}$ .

An experimental demonstration of Kochen-Specker contextuality in a three-level superconducting system was shown in ref. 9. The notion of generalized contextuality used in the present work is weaker than Kochen-Specker contextuality in that it does not require projective measurements (but rather applies to POVMs more generally) nor does it require outcome-determinism. Outcome-determinism is the requirement that outcomes of measurements depend deterministically on the underlying hidden variable  $\lambda$ . As argued in ref. 3, outcome-determinism is not entailed by non-contextuality, and thus should be viewed as an independent assumption in proofs of contextuality.

In addition to the assumption of outcome-determinism/projective measurements, a number of additional assumptions are needed in the demonstration of contextuality of ref. 9, as specified in ref. 9, Supplementary Material. We briefly comment on these assumptions (1.(a), 1.(b), 2. and 3.) and how they relate to the present work. Assumptions 2. and 3. are concerned with sequential measurements and compatibility of measurements, neither of which are needed in

present work. Assumption 1.(a) concerns the probability of picking a particular context, which is not needed in this work either. Assumption 1.(b), which states that the probability distribution over the ontic states is the same in every run, is shared with the present work.

As discussed in section "Contextuality of GPTs", we make the assumption that the preparations and measurements implemented in the experiment are tomographic relative to one another, which is not needed in ref. 9. The tomography loophole for experimentally demonstrating generalized contextuality can be viewed as analogous to the finite precision measurement loophole<sup>43</sup> for experimental demonstrations of Kochen-Specker contextuality.

Finally, we briefly contrast the theory-independent approach based on GPTs used here to the device-independent (DI) approach<sup>44</sup>, which can also be used to certify nonclassicality using minimal assumptions. In both the DI approach and GPT tomography, the basic objects are frequencies/probabilities over outputs given some inputs, which together with some well-motivated principles form the basis of the analysis. We note that the DI approach is not necessarily theory-independent, for instance, self-testing protocols verify that a certain quantum state (up to a family of local transformations) has been prepared based solely on the Bell inequality violation of the observed statistics<sup>45</sup>. The DI approach typically uses assumptions about causal structure (motivated by special relativity), whereas the GPT tomography approach makes both a causal assumption ( $\lambda$ -mediation<sup>21</sup>) and a tomographic completeness assumption, which are not motivated by appeal to another physical theory.

To summarize, in this work we have monitored the decoherence of a physical system, its evolution of generalized contextuality, and the (non-)Markovianity of its time evolution in a theory-independent way. This means that we have not assumed the validity of quantum theory in any of the data analysis or the conclusions that we have drawn. All that we had to assume was that the experimental setup includes a physical system that is probed in a way such that the implemented preparations and measurements are tomographically complete for each other, which allowed us to use the GPT formalism that does not depend on any specific choice of theory.

Our analysis has shown us that our experimental data is best described by a GPT that has a four-dimensional space of states and effects (consistent with the description as a qubit). We have shown that the state space shrinks over time (decoherence), loses its contextuality (emergence of classicality), and undergoes non-Markovian evolution at late times. Since we do not need to assume quantum theory in the analysis of the experiment, these properties are demonstrated to hold irrespective of the theoretical description of the physical system. That is, even if quantum theory were to be overturned by another theory in the future, our conclusions would still hold, as long as our experiment and its analysis have been correctly performed, and our assumption of tomographic completeness (see Methods' subsection "Contextuality of GPTs") is satisfied. This is to some extent similar to experimental demonstrations of the violation of Bell inequalities<sup>46-48</sup>, which show the failure of local realism not only within quantum theory, but as a property of nature that will survive all future revisions of our theoretical description. While we do not claim the same sort of device independence as for Bell experiments (mainly due to the assumption of tomographic completeness), we believe that our results represent a significant step forward towards the goal of a theory-independent reevaluation of experiments under minimal assumptions, which includes subjecting quantum theory to rigorous scrutiny.

It is instructive to be slightly more specific about what the GPT system really is that we have determined. Our experimental choice of 100 preparation and measurement procedures defines an operational theory, to which we have fitted an effective GPT system with state

space  $\mathcal{S}$  and effect space  $\mathcal{E}$ . As discussed in the Methods' subsection "Determining the set of consistent states", the set of all possible preparations and measurements on the superconducting qubit is larger, with state and effect spaces  $\mathcal{S}_{\text{phys}} \supset \mathcal{S}$  and  $\mathcal{E}_{\text{phys}} \supset \mathcal{E}$ . Due to our assumption of tomographic completeness, however, our effective GPT is, in the terminology of ref. 49, a *fragment* of the physical GPT of the superconducting system. Hence, the proof of generalized contextuality for the former (for times 10  $\mu\text{s}$  or less) applies directly to the latter, and so does the proof of non-Markovianity and the observation that the state space contracts over time.

Our work raises several interesting questions. For example, what can such experiments tell us if we do not make any assumption of tomographic completeness, such as in cases where many-body systems are probed with coarse-grained or collective measurements only<sup>50</sup>? Some progress has recently been made on this question<sup>51</sup>, in particular by showing that a notion of *relative tomographic completeness* is sufficient for demonstrating generalized contextuality, but there are still important open questions on how to apply this to the analysis of concrete experiments. Furthermore, can (standardly performed) measurements of the Wigner function, interpretable as large collections of preparation or measurement procedures, reveal generalized contextuality or other phenomena related to the GPT representation<sup>37</sup>? We believe that further theoretical analyses and experimental exploration may lead to interesting new insights into the foundations and practical certification of nonclassicality, as well as novel precision tests of quantum theory.

## Methods

To complement the methods section, in the repository<sup>52</sup> we provide an open-source Python code and the actual data used in this paper.

### Generalized probabilistic theories

A GPT system (without transformations) is described by a convex set of (normalized) states  $\mathcal{S} \subset V$  ( $V \simeq \mathbb{R}^d$ ) and a convex set of effects  $\mathcal{E} \subset V^*$  such that  $0 \leq \langle \mathbf{e}, \mathbf{s} \rangle \leq 1$  for all states and effects. The set of effects contains the  $\mathbf{0}$  effect:  $\langle \mathbf{0}, \mathbf{s} \rangle = 0$  for all  $\mathbf{s} \in \mathcal{S}$  and the unit effect  $\mathbf{u}$ :  $\langle \mathbf{u}, \mathbf{s} \rangle = 1$  for all  $\mathbf{s} \in \mathcal{S}$ . The number  $\langle \mathbf{e}, \mathbf{s} \rangle$  gives the probability that the measurement outcome represented by  $\mathbf{e}$  occurs when the system is prepared in state  $\mathbf{s}$ . Given a set of effects  $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$  and a set of states  $\{\mathbf{s}_1, \dots, \mathbf{s}_m\}$ , the associated probability table  $D$  is a  $m \times n$  matrix with entries  $D_{ij} = \langle \mathbf{e}_j, \mathbf{s}_i \rangle$ . Two GPT systems  $(\mathcal{S}, \mathcal{E})$  and  $(\mathcal{S}', \mathcal{E}')$  are *equivalent* if exists an invertible linear map  $L: V \rightarrow V'$  such that  $L(\mathcal{S}) = \mathcal{S}'$  and  $(L^{-1})^*(\mathcal{E}') = \mathcal{E}$ , where  $T^*$  denotes to adjoint of a linear map  $T$ . Therefore two equivalent systems yield the same probabilities:  $\langle \mathbf{e}', \mathbf{s}' \rangle = \langle (L^{-1})^*(\mathbf{e}), L(\mathbf{s}) \rangle = \langle \mathbf{e}, L^{-1}L(\mathbf{s}) \rangle = \langle \mathbf{e}, \mathbf{s} \rangle$ .

As a simple example of a GPT system, consider a qubit in quantum theory. The normalized states of the qubit are the  $2 \times 2$  density operators, which form a convex subset of the real vector space of Hermitian operators on  $\mathbb{C}^2$ . The effects of a qubit are the positive semidefinite operators  $E$  such that  $0 \leq E \leq \mathbf{1}$ , where  $\mathbf{0}$  is the zero effect and  $\mathbf{1}$  the unit effect. The probability of effect  $E$  given state  $\rho$  is given by  $\langle E, \rho \rangle = \text{Tr}(\rho E)$ .

The vector representation of a qubit is given by the Bloch representation. The state of a qubit can be written in the form  $\rho = \frac{1}{2} \left( \mathbf{1} + \sum_{i=1}^3 a_i \sigma_i \right)$ , where the  $\sigma_i$  are the Pauli matrices, which together with the identity  $\mathbf{1}$  form a basis for the 4-dimensional real vector space of Hermitian operators on  $\mathbb{C}^2$ . Positive semidefiniteness of  $\rho$  is equivalent to  $\sum_i a_i^2 \leq 1$ . Thus, the states of a qubit can be equivalently expressed in Bloch vector form as  $\mathbf{s}_\rho = \frac{1}{2}(1, a_1, a_2, a_3)^T$  with  $\|\mathbf{a}\| \leq 1$ . The effects can be expressed in the same basis as a vector  $\mathbf{e}_E = (c_0, c_1, c_2, c_3)^T$ , where  $c_i = \frac{1}{2} \text{Tr}(E \sigma_i)$ . Hence, the outcome

probabilities are not  $w_p(0|P_i, M_j) = \text{Tr}(\rho_i E_j) = \mathbf{s}_i^T \mathbf{e}_j$ . The full probability table can now be written as:

$$D = \begin{pmatrix} \mathbf{s}_1^T \\ \mathbf{s}_2^T \\ \vdots \\ \mathbf{s}_m^T \end{pmatrix} (\mathbf{e}_1 \ \mathbf{e}_2 \ \dots \ \mathbf{e}_n) = SE. \tag{8}$$

Thus, the  $m \times n$  probability table  $D$  for the qubit can be factored into a  $m \times 4$  matrix  $S$  and a  $4 \times n$  matrix  $E$ . This implies that  $D$  is a rank-4 matrix.

Another canonical example of a GPT system is the  $d$ -dimensional classical system  $\Delta_d$ . The normalized states of  $\Delta_d$  consist of probability distributions over  $d$  outcomes, which form a convex subset of  $\mathbb{R}^d$ . The effects are the response functions, namely all linear functionals  $e$  which map states to  $[0, 1]$ . In the vector representation, the normalized states are  $d$ -dimensional vectors with entries in  $[0, 1]$  which sum to 1 and the effects are vectors with entries in  $[0, 1]$ . The  $\mathbf{0}$  effect is the vector  $(0, \dots, 0)^T$  and the unit effect is the vector  $(1, \dots, 1)^T$ . The states form a  $d$ -dimensional simplex, whilst the effects form a  $d$ -dimensional hypercube.

### Uniqueness of the decomposition $D = SE$

**Lemma 1.** Suppose that  $D = SE$ , where  $S$  is a real  $m \times k$  matrix,  $E$  is a real  $k \times n$  matrix, and  $k = \text{rank } D$ . If  $D = S'E'$  is another decomposition with these properties, then there is an invertible matrix  $L$  such that  $S' = SL$  and  $E' = L^{-1}E$ .

**Proof..** If  $M$  is some matrix, we denote its Moore-Penrose pseudoinverse<sup>53</sup> by  $M^+$ . Set  $L := E(E')^+$ , then

$$SL = SE(E')^+ = D(E')^+ = S'E'(E')^+. \tag{9}$$

Now, the  $k$  rows of  $E'$  are linearly independent, because otherwise, we would have  $\text{rank } D \leq \text{rank } E' < k$ . This implies  $E'(E')^+ = \mathbf{1}$ , and hence  $SL = S'$ . Similarly, the  $k$  columns of  $E$  are linearly independent, which implies both  $E^+E = \mathbf{1}$  and  $(E(E')^+)^+ = E'E^+$ . Thus

$$L^+E = (E(E')^+)^+E = E'E^+E = E'. \tag{10}$$

It remains to be shown that  $L$  is invertible. Using all of above, this follows from

$$L^+L = (E(E')^+)^+E(E')^+ = E'E^+E(E')^+ = \mathbf{1}, \tag{11}$$

hence  $L^+ = L^{-1}$ .

### Solving the weighted low-rank approximation problem of Eq. (4)

In this section, we outline the methodology introduced in ref. 10 to find a rank- $k$  matrix  $D^{\text{realized}}$ , which serves as the best estimate for the ideal probability table  $D$  leading to the experimentally observed statistics  $F$ . Let us recall that this is achieved by minimizing the weighted  $\chi_k^2$  statistics in the following optimization problem (4):

$$\min_{D \in M_{mn}} \chi_k^2 = \sum_{ij} \frac{(F_{ij} - D_{ij})^2}{\Delta F_{ij}^2} \tag{12}$$

subject to  $\text{rank}(D) = k, 0 \leq D_{ij} \leq 1$ ,

where  $\Delta F_{ij}^2$  is the statistical variance in  $F_{ij}$ . Note that  $D = SE$  contains the inner products  $\mathbf{s}_i^T \mathbf{e}_j$ , which makes Eq. (4) a non-convex optimization

problem. Moreover, this minimization problem is known to not have an analytical solution<sup>54,55</sup>. Following ref. 10, we employ an iterative approach that decomposes the optimization into two convex subproblems. Specifically, we alternate between optimizing the states while fixing the effects and optimizing the effects while fixing the states, with each optimization being a convex quadratic program<sup>10</sup>. This iterative process, known as the see-saw algorithm<sup>56,57</sup>, continues until the cost function, here the  $\chi_k^2$ , converges to a desired numerical precision. However, it is important to note that due to the non-convexity of the general problem, this procedure does not guarantee convergence to a global minimum. We refer the reader to ref. 10 [Appendix C] for an explicit description of how the  $\chi_k^2$  minimization in Eq. (4) can be translated into a series of alternating convex optimizations.

The output of the optimization algorithm is a rank  $k$  matrix  $D$  with entries in  $[0, 1]$ . In order to obtain the corresponding GPT states and effects, we must factor  $D$  into two matrices  $S$  and  $E$ . Note that this factorization is non-unique. As in ref. 10, we first append a column of 1's to  $D$ , which reflects the probability associated with the unit effect (corresponding to the identity operator in quantum theory), assuming no experimental losses. Then, as described in ref. 10 [Appendix C], we perform a QR decomposition of  $D$  to express it as  $D = SE$ . This ensures that the first column of  $S$  consists of 1's, thus effectively encoding the normalization of the GPT states (analogous to the trace degree of freedom for quantum states).

Hence the matrices of states and effects have the following forms:

$$S = \begin{pmatrix} 1 & s_1^1 & \dots & s_1^{k-1} \\ 1 & s_2^1 & \dots & s_2^{k-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & s_m^1 & \dots & s_m^{k-1} \end{pmatrix} = \begin{pmatrix} \mathbf{s}_1^\top \\ \mathbf{s}_2^\top \\ \vdots \\ \mathbf{s}_m^\top \end{pmatrix}, \quad (13)$$

$$E = \begin{pmatrix} 1 & e_1^0 & \dots & e_n^0 \\ 0 & e_1^1 & \dots & e_n^1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & e_1^{k-1} & \dots & e_n^{k-1} \end{pmatrix} = (\mathbf{e}_1 \ \mathbf{e}_2 \ \dots \ \mathbf{e}_n). \quad (14)$$

From the optimization problem and choice of factorization, we obtain the unit effect  $\mathbf{1} = \mathbf{e}_0$  and  $n$  effects  $\mathbf{e}_1, \dots, \mathbf{e}_n$ , where  $\mathbf{s}_i^\top \mathbf{e}_j \in [0, 1]$  for all  $i, j$  by construction. Every measurement  $M_j$  has two outcomes 0 and 1 with  $p(0|P_i, M_j) = \mathbf{s}_i^\top \mathbf{e}_j$ . Since  $P(0|P_i, M_j) + P(1|P_i, M_j) = 1$  it follows that  $p(1|P_i, M_j) = 1 - \mathbf{s}_i^\top \mathbf{e}_j = \mathbf{s}_i^\top (\mathbf{1} - \mathbf{e}_j)$ .

Hence, in order to account for the probabilities  $p(1|P_i, M_j)$ , the effect space should include the complement effect  $(\mathbf{1} - \mathbf{e}_j)$  for every  $\mathbf{e}_j$ . We append the  $(n + 1) \times k$  matrix of complement effects  $E'$  with columns  $(\mathbf{1} - \mathbf{e}_j)$  to the matrix  $E$  to obtain a  $2(n + 1) \times k$  matrix of effects (which we shall label as  $E$  hereon for convenience). Augmenting the set of effects in this manner does not change the set of consistent states or effects.

### Determining the best fit rank $k$

When solving Eq. (4), note that as the chosen rank  $k$  increases, the error of the fit  $\chi_k^2$  naturally decreases. However, note that in general the experimental noise causes  $F$  to be a full rank matrix, even when the underlying  $D$  it approximates is not. Therefore, increasing  $k$  can result in a model that overfits the experimental data, which is to say that it fits the noise in the training set data  $F$ . In practice, this can be determined by separating the data into a training set  $F = F^{\text{train}}$  and a test set  $F' = F^{\text{test}}$ , where the  $\chi_k^2$  is evaluated according to Eq. (5). The  $\chi_k^2$  for the training data determines how much the model underfits the data; the higher the  $\chi_k^2$  value the worse the fit is. To test for overfitting, the model is applied to the test set  $F^{\text{est}}$ . Models that overfit will have a

significantly worse  $\chi_k^2$  since they will fit to noise that was present in  $F^{\text{train}}$  but is not present in  $F^{\text{test}}$ . There is a value  $k^{\text{opt}}$  for which the models will transition from underfitting to overfitting. Then, the matrix  $D_{k^{\text{opt}}}$  is chosen as the best fit.

In summary, to estimate the rank  $k^{\text{opt}}$ , we hypothesize several values  $k$  for  $k^{\text{opt}}$  and test each one. Then, for each hypothesized rank, we compute the optimal estimate of the rank- $k$  matrix  $D$  and evaluate its fit to the data using the  $\chi^2$  statistic. Additionally, to decide which  $k$  provides the best balance between fitting the experimental data and maintaining simplicity in the model, we split the analysis with a training and a test set to ensure that the model is not overfitting the data. As we have seen in subsection “Generalized probabilistic theories”, if the experiment is described by a qubit quantum model (as we expect, since we have prepared and probed a system that is, from quantum physics and the experimental setup, expected to be a qubit), then the rank is expected to be 4. Indeed, following this method, we have shown in Fig. 4a that this is the case. Nonetheless, recall that we are keeping it general to follow a theory-agnostic approach, where we obtain the best GPT fit without assuming the correctness of quantum theory.

### Determining the set of consistent states

The methodology so far estimates one possible set of states and effects compatible with the experimental data, which we denote  $\mathcal{S}$  and  $\mathcal{E}$ . However, as in ref. 10, we can further estimate the set of all logically consistent states and effects for the experiment, named respectively  $\mathcal{S}_{\text{consistent}}$  and  $\mathcal{E}_{\text{consistent}}$ . The former is given by the set of all possible vectors  $\mathbf{s}$ , normalized such that  $(\mathbf{1}, \mathbf{s}) = 1$ , which, when acted on by any covector of the realized effect space, give valid probabilities, i.e. all  $\mathbf{s}$  such that  $0 \leq (\mathbf{e}, \mathbf{s}) \leq 1$  for all  $\mathbf{e} \in \mathcal{E}$ . Conversely, the latter is defined by the set of all covectors  $\mathbf{e}$  that take the vectors of the realized state space to valid probabilities, i.e. all  $\mathbf{e}$  such that  $0 \leq (\mathbf{e}, \mathbf{s}) \leq 1$  for all  $\mathbf{s} \in \mathcal{S}$ . These inequalities tell us that the consistent spaces are the *duals* of the realized spaces, i.e.,  $\mathcal{S}_{\text{consistent}} = \text{dual}(\mathcal{E})$  and  $\mathcal{E}_{\text{consistent}} = \text{dual}(\mathcal{S})$ . If the realized state space  $\mathcal{S}$ /effect space  $\mathcal{E}$  is smaller (in particular, if it is a strict subset of the set of states  $\mathcal{S}_{\text{phys}}$ /effects  $\mathcal{E}_{\text{phys}}$  that could in principle be implemented on the system, which is always the case in experiments on quantum systems with infinitely many pure states), then its corresponding dual is larger – since there is a larger set of covectors/vectors that combine appropriately to give valid probabilities. Accordingly, we know that  $\mathcal{S}_{\text{phys}}$  and  $\mathcal{E}_{\text{phys}}$  must lie somewhere between the realized and consistent state/effect spaces, comprising lower and upper bounds, respectively.

The duals are calculated via the `odd` library in Python<sup>58</sup>, and amounts (in their terminology) to converting from the H-representation of a convex polyhedron  $P$  to its V-representation. For a given matrix of points, say that of  $\mathcal{S}$ , that must be combined with  $\mathcal{E}_{\text{consistent}}$  to give valid probabilities, a convex polyhedron of the form  $P = \{x | Ax \leq b\}$  can be defined, where  $A = [-[\mathcal{S}], [\mathcal{S}]]^\top$  and  $b = [[0]^m, [\mathbf{1}]^m]^\top$ . The H-representation is the matrix  $[b - A]$ , representing the set of linear inequalities characterizing  $P$ . The polyhedron can also be represented by the convex hull of its vertices, i.e.,  $P = \text{conv}(v_1, \dots, v_n)$ , which is called its V-representation. We can use the `get_generators()` method of the `odd` library to obtain the V-representation  $[t - V]$ , where  $t = [[\mathbf{1}]^m]^\top$  and  $V = [\mathcal{E}_{\text{consistent}}]$ .

### Reparametrization of the state space

In order to compare the different  $\mathcal{S}^\tau$  for different decoherence times  $\tau$ , we are interested in finding an appropriate parametrization for these distinct sets. This is achieved by applying a linear transformation that approximates their respective  $\mathcal{S}_{\text{consistent}}$  set to a unit sphere. Note that  $\mathcal{S}_{\text{consistent}}$  is a function of  $\mathcal{E}$ , which does by construction not depend of  $\tau$ . As explained in Subsection “Generalized probabilistic theories”, we can always reparametrize GPT systems linearly without changing any

of their relevant properties. The method we present is split in three steps: (1) we make sure that the set of data points we work with are all extremal points; (2) we center the boundary on the origin, and (3) we pose the search for the suitable linear transformation as an optimization problem. After finding an appropriate centering and transformation for the boundary of the consistent state space  $\mathcal{S}_{\text{consistent}}$ , we apply the same to all realized state spaces  $\mathcal{S}^r$ , so they can be appropriately compared.

Let us start with a cautionary preliminary step by removing all the interior points of a given set of states  $\mathcal{S}$ . This is to guarantee that we are left only with boundary points. We do so by probing whether a given state  $\mathbf{s}_i \in \mathcal{S}$  can be obtained as a convex combination of all the other points  $\mathbf{s}_j \in \mathcal{S} \setminus \{\mathbf{s}_i\}$  for  $i \neq j$ . This can be posed as a feasibility problem using linear programming. If the linear program is feasible, then we know that the state being probed  $\mathbf{s}_i$  is a mixture of the other points; i.e., there exists a convex combination  $\sum_{j \neq i} \lambda_j \mathbf{s}_j = \mathbf{s}_i$  with  $\mathbf{s}_j \in \mathcal{S} \setminus \{\mathbf{s}_i\}$ ,  $0 \leq \lambda_j \leq 1$  and  $\sum_{j \neq i} \lambda_j = 1$ . The feasibility linear program is the following:

$$\begin{aligned} & \text{find } \lambda \\ & \text{subject to } \sum_{j \neq i} \lambda_j \mathbf{s}_j = \mathbf{s}_i, \quad \sum_{j \neq i} \lambda_j = 1, \quad 0 \leq \lambda_j \leq 1 \text{ for all } j \neq i. \end{aligned} \quad (15)$$

Therefore, when the linear program is feasible, we discard the state  $\mathbf{s}_i$  and update the set of states to  $\mathcal{S}' = \mathcal{S} \setminus \{\mathbf{s}_i\}$ . We iterate this procedure for all  $i \in \{1, \dots, n\}$  until we have tested all states  $\mathbf{s}_i$  and we are left only with the ones which cannot be obtained as convex combinations of the rest.

Next, we center the boundary  $\mathcal{S}'$  on the origin by subtracting the columnwise mean; namely  $\mathcal{S}_{\text{centered}} = \mathcal{S}' - \boldsymbol{\mu}$ , where  $\boldsymbol{\mu}$  is a column vector with entries  $\mu_j = \frac{1}{n} \sum_{i=1}^n S_{i,j}$  denoting the mean for column  $j$  of  $\mathcal{S}'$ .

Finally, let us now provide a method to find a linear transformation that maps  $\mathcal{S}_{\text{centered}}$  closer to the unit sphere. Consider a state  $\mathbf{s}_i \in \mathcal{S}_{\text{centered}}$ ; the idea is to find a transformation  $L$  such that  $\|\mathbf{s}'_i\| \approx 1 \forall i$  with  $\mathbf{s}'_i := L\mathbf{s}_i$ . If we focus in the case where the state space has rank  $k = 4$ , then  $L$  will be a  $(k - 1) \times (k - 1) = 3 \times 3$  matrix. By means of the singular value decomposition,  $L = U\Sigma V^T$ , note that we are only interested in finding a diagonal matrix  $\Sigma := \text{diag}(\sigma_1, \sigma_2, \sigma_3)$  with  $\sigma_i \in \mathbb{R}_{\geq 0}$  and a  $3 \times 3$  real orthogonal matrix  $V$ . We can neglect  $U$  since, by definition,  $\|\mathbf{s}'_i\| = \sqrt{\mathbf{s}_i^T V \Sigma^T U^T U \Sigma V^T \mathbf{s}_i} = \sqrt{\mathbf{s}_i^T V \Sigma^2 V^T \mathbf{s}_i}$ . Furthermore, we use Euler angles parametrization for  $V$ , which guarantees the constraint  $VV^T = \mathbf{1}$  while having full coverage of  $\text{SO}(3)$  with 3 parameters (angles)  $\alpha, \beta, \gamma \in \mathbb{R}$ :

$$V := \begin{pmatrix} \cos(\alpha) \cos(\gamma) - \sin(\alpha) \cos(\beta) \sin(\gamma) & -\cos(\alpha) \sin(\gamma) - \sin(\alpha) \cos(\beta) \cos(\gamma) & \sin(\alpha) \sin(\beta) \\ \sin(\alpha) \cos(\gamma) + \cos(\alpha) \cos(\beta) \sin(\gamma) & -\sin(\alpha) \sin(\gamma) + \cos(\alpha) \cos(\beta) \cos(\gamma) & -\cos(\alpha) \sin(\beta) \\ \sin(\beta) \sin(\gamma) & \sin(\beta) \cos(\gamma) & \cos(\beta) \end{pmatrix}. \quad (16)$$

Therefore, to find the suitable  $\Sigma$  and  $V$ , we consider the following constrained minimization problem:

$$\begin{aligned} & \min_{\sigma_1, \sigma_2, \sigma_3, \alpha, \beta, \gamma \in \mathbb{R}} \sum_{i=1}^n (1 - \|\mathbf{s}'_i\|^2)^2 \\ & \text{subject to } \Sigma := \text{diag}(\sigma_1, \sigma_2, \sigma_3), \\ & \quad \sigma_1, \sigma_2, \sigma_3 \geq 0. \end{aligned} \quad (17)$$

Once we find the suitable translation  $\boldsymbol{\mu}$  and transformation  $L = \Sigma V^T$  for a given  $\mathcal{S}_{\text{consistent}}$ , we apply them in turn to all realized state spaces  $L(\mathcal{S}^r - \boldsymbol{\mu})$  so that we can properly compare the sets as shown in Fig. 6.

### Lower bound on the purity of the prepared states

If we assume that quantum theory holds, then there will be a valid qubit density matrix  $\rho_{\mathbf{s}}$  for every state  $\mathbf{s} \in \mathcal{S}$ . Let us assume that  $\lambda_{\max}(\rho_{\mathbf{s}}) \leq c$  for all  $\mathbf{s} \in \mathcal{S}$ , where  $\frac{1}{2} \leq c \leq 1$ . Then the Bloch vectors of all  $\rho_{\mathbf{s}}$  are contained in ball of radius  $2c - 1$  around the origin in the Bloch representation. Since the one-norm distance of quantum states corresponds to the Euclidean distance in the Bloch ball, this means that  $D(\mathbf{s}, \mathbf{s}') \leq \frac{1}{2} \|\rho_{\mathbf{s}} - \rho_{\mathbf{s}'}\|_1 \leq 2c - 1$  for all  $\mathbf{s}, \mathbf{s}' \in \mathcal{S}$ . Hence  $c \geq \frac{1}{2} (1 + \max_{\mathbf{s}, \mathbf{s}' \in \mathcal{S}} D(\mathbf{s}, \mathbf{s}'))$ , and the distinguishability maximum is numerically determined to be at least  $0.691 \pm 0.004$ , thus  $c \geq 0.846 \pm 0.002$ .

### Markovian evolutions do not increase the volume

If  $T$  is any transformation on a GPT system, then it acts linearly on the set of unnormalized states in  $\mathbb{R}^{k \times 15}$ . In particular, in the parametrization where the first component of a state vector is its normalization, we have  $\mathbf{s}' = \begin{pmatrix} 1 \\ \mathbf{s}' \end{pmatrix} = T \begin{pmatrix} 1 \\ \mathbf{s} \end{pmatrix}$ , and so  $\mathbf{s}' = T' \mathbf{s} + t$ , with some vector  $t$  and linear transformation  $T'$  on  $\mathbb{R}^{k-1}$ . (To see this explicitly, write  $T = (T'_{ij})_{i,j=0}^{k-1}$ , then  $\mathbf{s}'_j = T_{j0} + \sum_{n=1}^{k-1} T_{jn} \mathbf{s}_n = : t + T' \mathbf{s}$ .) Clearly, transformations map valid states to valid states, so the image of the full state space must satisfy  $T' \mathcal{S} + t \subseteq \mathcal{S}$ , and so

$$\text{Vol}(\mathcal{S}) \geq \text{Vol}(T' \mathcal{S} + t) = \text{Vol}(T' \mathcal{S}) = |\det T'| \text{Vol}(\mathcal{S}). \quad (18)$$

Thus,  $|\det T'| \leq 1$ , and hence the action of the transformation on the normalized state space cannot increase the volume of any region.

To compute the volumes shown in Fig. 8 for each  $\mathcal{S}^r$ , we have used the *Delaunay Triangulation*<sup>59</sup> on their boundary points. This splits the polytope into non-overlapping simplices (tetrahedras in our case), for which the volume formula can be easily calculated. The total volume of  $\mathcal{S}^r$  is then estimated by summing the volumes of these simplices.

### Decoherence in GPTs

In the main text, we say that the state of the superconducting system decoheres, even though the notion of “decoherence” is usually only used in QT. Here, we point out that (1) there is a rigorous definition of decoherence for GPTs in the literature which applies to the long-time behavior of the superconducting system; and (2) for finite times, the system evolves irreversibly, and this allows us to draw very similar conclusions to the quantum version of decoherence, justifying the name also beyond QT.

Regarding (1), we will use the definition of ref. 60, which is slightly more general than that of ref. 61 or ref. 62: a complete decoherence map  $T$  is a transformation such that  $T^2 = T$ , i.e., applying the transformation twice is the same as applying it once. In QT, an example is given by the map that removes all off-diagonal elements of a density matrix. Figure 6 indicates that the long-time effect of the time evolution is to map every initial state  $\omega$  to the same final state  $\omega_0$  (which we physically expect to be the ground state),  $T(\omega) = \omega_0$ . This is clearly a decoherence map according to the previous definition, since  $T(T(\omega)) = T(\omega_0) = \omega_0$ .

To see (2), note that reversible transformations  $T$  map pure states of a GPT system to pure states, and they preserve the state space  $\mathcal{S}$ , i.e.,  $T(\mathcal{S}) = \mathcal{S}$ . This is clearly not the case in our system’s time evolution: most pure states are mapped to mixed states (in particular, for times

$\tau \leq 15 \mu\text{s}$ ), and the image of  $S$  is strictly smaller than  $S$ . Hence, we have irreversible time evolution on the superconducting system  $S$  that we are probing. However, if we assume that the total evolution of  $S$  and its environment  $E$  is reversible (in QT, this would be equivalent to global unitarity), then it cannot act on  $S$  and on  $E$  independently (otherwise,  $S$  and  $E$  would individually preserve their purity). In other words, it must be the reversible interaction between  $S$  and  $E$  that makes the initially close to pure state of  $S$  mixed. This resembles the phenomenology of decoherence of QT and justifies the use of this term in the more general GPT context.

### Contextuality of GPTs

Having shown the image of the GPT state space according to the time delay  $\tau$ , we want a theory-independent test of their (non-)contextuality. It was shown in ref. 29 that non-contextuality can be established via a linear program which takes as an input a set of states and a set of effects, and determines whether their statistics, via a specified probability rule, can be reproduced by a classical (non-contextual) model. We make use of an open-source version of this program<sup>30</sup>, which, moreover, in the case that the GPT is contextual, computes the amount of noise that must be added such that a noncontextual model can be fitted. More specifically, the linear program of ref. 30 is a test of *simplex embeddability*: the property that a GPT's state space can be embedded into a classical probability simplex, with its effect space as the dual of that simplex. This was shown in refs. 28,49 to be equivalent to the existence of a non-contextual ontological model.

For the nonembeddability of the GPT to meaningfully imply non-classicality, we must assume that there is some system that our experiment probes for which our preparations and measurements are *relatively tomographically complete*<sup>51</sup>. We note that the system being probed will in general be a subsystem of some larger system, for instance in this case it corresponds to the first two energy levels of an anharmonic oscillator. However, crucially, it must be such that the preparations and measurement of that subsystem are tomographically complete relative to each other, or equivalently, that the subsystem being probed is embeddable within the larger system<sup>60</sup>. Without assuming so, it is always possible that the statistics derive from some higher-dimensional, noncontextual system, c.f. the Holevo projection<sup>63</sup>. Tomographic completeness in itself is a theory-independent assumption, but can be strongly motivated here by quantum physics—in particular, by the state space of the qubit being a Bloch ball, for which we have chosen a tomographically highly over-complete distribution of points to sample ( $k = 4$  linearly independent states and effects would have been sufficient for tomographic completeness, but many more of these procedures are needed to obtain an accurate estimate of the shape of the state and effect spaces as depicted in Figs. 5 and 6).

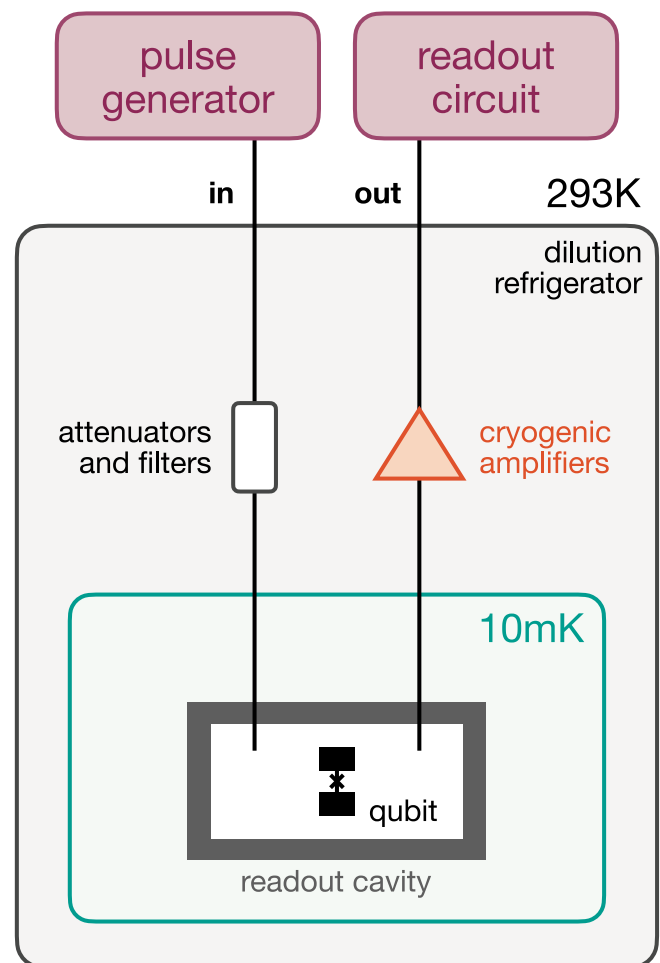
To use the linear program, one inputs the GPT's set of states  $S^T$  and set of effects  $\mathcal{E}$ , which were found via the theory-independent analysis of Results' subsection "Theory-independent analysis", as well as two row vectors specifying the unit effect  $\mathbf{u}$  and the average state  $\mathbf{m}^T$ . The unit effect is given by  $(1, 0, 0, 0)$ , which is always the first entry of the set of effects, whilst the average state is calculated by taking the columnwise mean of the set of states. The program then outputs a coefficient  $r$ , termed the *robustness of nonclassicality*, as well as a set of epistemic states and response functions (which specify a noncontextual ontological model for the scenario). The important information is encoded in  $r$ , which quantifies how much depolarizing noise must be added such that a noncontextual ontological model can be fitted. In particular, the maximally mixed state defines a depolarization map  $D_r^T(\mathbf{s}) = (1 - r)\mathbf{s} + r\mathbf{m}^T$ ; the minimum  $r$  such that a noncontextual ontological model can be fitted to the scenario thus constitutes an operational measure of nonclassicality. More specifically, a robustness of  $r = 0$  indicates that a noncontextual

ontological model can be constructed precisely, without adding noise, whilst  $r > 0$  indicates that some amount of depolarizing noise is necessary for an ontological model—thereby witnessing contextuality.

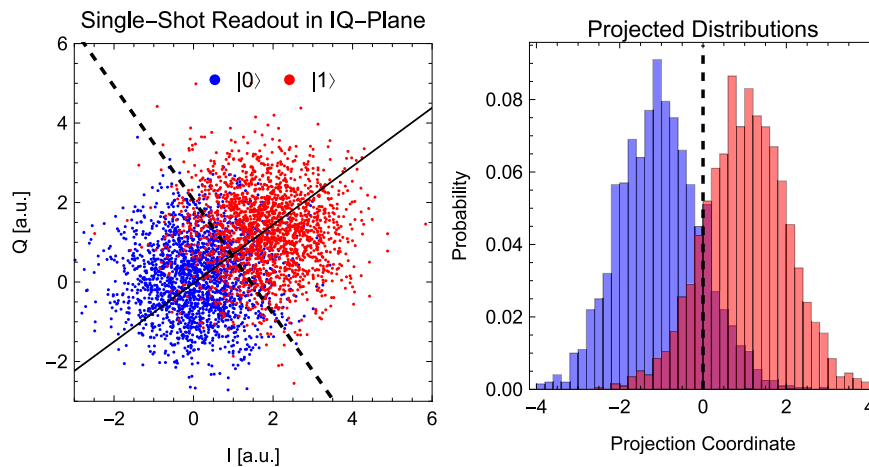
### Experimental setup and qubit readout characterization

The device used for this work is a superconducting transmon qubit hosted inside a three-dimensional readout cavity resonator. For operation, this device is cooled down at  $\sim 10$  mK temperature by a dilution refrigerator and operated by commercial microwave electronics, see Fig. 9 for an illustration. Qubit rotation and readout pulses are generated at room temperature by an arbitrary waveform generator and then up-converted to gigahertz frequencies. After a room-temperature amplification stage, these pulses are sent to the sample inside the refrigerator through a line of cryogenic attenuators and filters protecting the qubit from undesired electromagnetic radiation. The qubit readout pulse is transmitted through the cavity, amplified by both cryogenic and room-temperature components, down-converted to megahertz frequencies, and finally digitalized by a FPGA before being recorded on a PC.

Following the standard practice in superconducting qubit experiments, the state of our qubit is measured in the strong dispersive regime of circuit-QED<sup>19</sup>. To calibrate our readout and characterize its fidelity, we collect 2000 measurement shots for the qubit



**Fig. 9 | Schematic illustration of the experimental setup used in this work.** The superconducting qubit device used in this work is hosted inside a three-dimensional microwave cavity cooled down at cryogenic temperature by a dilution refrigerator. Qubit state preparation and readout are controlled by microwave electronics at room temperature.



**Fig. 10 | Illustration of the qubit readout characterization.** Left: for each qubit initial state, 2000 measurements are collected for identifying the decision boundary (dashed line) in the IQ-plane. Right: projection of the distributions along the direction connecting their means (solid line).

**Table 1 | Assignment probabilities for prepared quantum states**

	Assigned $ 0\rangle$	Assigned $ 1\rangle$
Prepared $ 0\rangle$	0.846	0.154
prepared $ 1\rangle$	0.148	0.852

in its ground state  $|0\rangle$  and an equal number of shots for the qubit prepared in  $|1\rangle$ . Representative results are shown in Fig. 10 (left), which allow us to identify the decision boundary for the  $|0\rangle$  vs.  $|1\rangle$  state. For clarity, we also show a 1D projection of the measurement distributions along the direction connecting their means in Fig. 10 (right).

From the collected statistics, we can identify the conditional probability  $p(a|P)$  of assigning the label  $a \in \{|0\rangle, |1\rangle\}$  when state  $P \in \{|0\rangle, |1\rangle\}$  was prepared. These are given in Table 1.

The average readout fidelity is then defined as  $1 - (p(0|1) + p(1|0))/2$ , which gives 85(1)%. Contributions to the infidelity come from the finite cavity dispersive shift, qubit state decay during the  $\sim 10 \mu\text{s}$  long readout pulse, as well as finite thermal population of  $\sim 10^{-3}$  at equilibrium.

### Data availability

The data that support the findings of this study are publicly available in the Zenodo repository<sup>52</sup>.

### Code availability

The custom code used to produce the results presented in this paper is publicly available in the Zenodo repository<sup>52</sup>.

### References

- Bell, J. S. On the Einstein-Podolsky-Rosen paradox. *Physics* **1**, 195–200 (1964).
- Kochen, S. & Specker, E. P. The problem of hidden variables in quantum mechanics. *J. Math. Mech.* **17**, 59–87 (1967).
- Spekkens, R. W. Contextuality for preparations, transformations, and unsharp measurements. *Phys. Rev. A* **71**, 052108 (2005).
- Mazurek, M. D., Pusey, M. F., Kunjwal, R., Resch, K. J. & Spekkens, R. W. An experimental test of noncontextuality without unphysical idealizations. *Nat. Commun.* **7**, 11780 (2016).
- Kunjwal, R. *Contextuality beyond the Kochen-Specker Theorem*. Ph.D. thesis, Homi Bhabha National Institute (2016).
- Krishna, A., Spekkens, R. W. & Wolfe, E. Deriving robust non-contextuality inequalities from algebraic proofs of the Kochen-Specker theorem: the Peres-Mermin square. *New J. Phys.* **19**, 123031 (2017).
- Spekkens, R. W. Negativity and contextuality are equivalent notions of nonclassicality. *Phys. Rev. Lett.* **101**, 020401 (2008).
- Schmid, D., Selby, J. H., Pusey, M. F. & Spekkens, R. W. A structure theorem for generalized-noncontextual ontological models. *Quantum* **8**, 1283 (2024).
- Jerger, M. et al. Contextuality without nonlocality in a superconducting quantum system. *Nat. Commun.* **7**, 12930 (2016).
- Mazurek, M. D., Pusey, M. F., Resch, K. J. & Spekkens, R. W. Experimentally bounding deviations from quantum theory in the landscape of generalized probabilistic theories. *PRX Quantum* **2**, 020302 (2021).
- Grabowecky, M., Pollack, C., Cameron, A., Spekkens, R. W. & Resch, K. J. Experimentally bounding deviations from quantum theory for a photonic three-level system using theory-agnostic tomography. *Phys. Rev. A* **105**, 032204 (2022).
- Barrett, J. Information processing in generalized probabilistic theories. *Phys. Rev. A* **75**, 032304 (2007).
- Janotta, P. & Hinrichsen, H. Generalized probability theories: what determines the structure of quantum theory? *J. Phys. A Math. Theor.* **47**, 323001, (2014).
- Müller, M. P. Probabilistic theories and reconstructions of quantum theory (Les Houches 2019 lecture notes). *SciPost Phys. Lect. Notes* **28**, 1–41 (2021).
- Plávala, M. General probabilistic theories: an introduction. *Phys. Rep.* **1033**, 1–64 (2023).
- Li, L., Hall, M. J. & Wiseman, H. M. Concepts of quantum non-Markovianity: a hierarchy. *Phys. Rep.* **759**, 1–51 (2018).
- Kjaergaard, M. et al. Superconducting qubits: current state of play. *Annu. Rev. Condens. Matter Phys.* **11**, 369–395 (2020).
- Koch, J. et al. Charge-insensitive qubit design derived from the Cooper pair box. *Phys. Rev. A* **76**, 042319 (2007).
- Blais, A., Grimsmo, A. L., Girvin, S. M. & Wallraff, A. Circuit quantum electrodynamics. *Rev. Mod. Phys.* **93**, 025005 (2021).
- González, Á. Measurement of areas on a sphere using Fibonacci and latitude-longitude lattices. *Math. Geosci.* **42**, 49–64 (2010).
- Catani, L. & Leifer, M. A mathematical framework for operational fine tunings. *Quantum* **7**, 948 (2023).
- Ududec, C., Barnum, H. & Emerson, J. Three slit experiments and the structure of quantum theory. *Found. Phys.* **41**, 396 (2011).

23. Barnum, H., Müller, M. P. & Ududec, C. Higher-order interference and single-system postulates characterizing quantum theory. *New J. Phys.* **16**, 123029 (2014).
24. Hardy, L. Quantum theory from five reasonable axioms. *arXiv* <https://doi.org/10.48550/arXiv.quant-ph/0101012> (2001).
25. Dakić, B. & Brukner, Č. Quantum theory and beyond: is entanglement special? in *Deep Beauty: Understanding the Quantum World through Mathematical Innovation* (ed. Halvorson, H.) (Cambridge University Press, 2011).
26. Masanes, L. & Müller, M. P. A derivation of quantum theory from physical requirements. *New J. Phys.* **13**, 063001 (2011).
27. Chiribella, G., d'Ariano, G. M. & Perinotti, P. Informational derivation of quantum theory. *Phys. Rev. A* **84**, 012311 (2011).
28. Schmid, D., Selby, J. H., Wolfe, E., Kunjwal, R. & Spekkens, R. W. Characterization of noncontextuality in the framework of generalized probabilistic theories. *PRX Quantum*, **2**, 010331 (2021).
29. Gitton, V. & Woods, M. P. Solvable criterion for the contextuality of any prepare-and-measure scenario. *Quantum* **6**, 732 (2022).
30. Selby, J. H., Wolfe, E., Schmid, D., Sainz, A. B. & Rossi, V. P. A linear program for testing nonclassicality and an open-source implementation. *Phys. Rev. Lett.* **132**, 050202 (2024).
31. Lillystone, P., Wallman, J. J. & Emerson, J. Contextuality and the single-qubit stabilizer subtheory. *Phys. Rev. Lett.* **122**, 140405 (2019).
32. Gottesman, D. The heisenberg representation of quantum computers. *arXiv* <https://doi.org/10.48550/arXiv.quant-ph/9807006> (1998).
33. Spekkens, R. W., Buzacott, D. H., Keehn, A. J., Toner, B. & Pryde, G. J. Preparation contextuality powers parity-oblivious multiplexing. *Phys. Rev. Lett.* **102**, 010401 (2009).
34. Hameedi, A., Tavakoli, A., Marques, B. & Bourennane, M. Communication games reveal preparation contextuality. *Phys. Rev. Lett.* **119**, 220402 (2017).
35. Saha, D. & Chaturvedi, A. Preparation contextuality as an essential feature underlying quantum communication advantage. *Phys. Rev. A* **100**, 022108, (2019).
36. Schmid, D. & Spekkens, R. W. Contextual advantage for state discrimination. *Phys. Rev. X* **8**, 011015 (2018).
37. Schmid, D., Du, H., Selby, J. H. & Pusey, M. F. Uniqueness of non-contextual models for stabilizer subtheories. *Phys. Rev. Lett.* **129**, 120403 (2022).
38. Catani, L., Galley, T. D. & Gonda, T. Resource-theoretic hierarchy of contextuality for general probabilistic theories. *arXiv* <https://doi.org/10.48550/arXiv.2406.00717> (2024).
39. Chiribella, D., D'Ariano, G. M. & Perinotti, P. Probabilistic theories with purification. *Phys. Rev. A* **81**, 062348 (2010).
40. Chiribella, D., D'Ariano, G. M. & Perinotti, P. Informational derivation of quantum theory. *Phys. Rev. A* **84**, 012311 (2011).
41. Bell, J. S. On the problem of hidden variables in quantum mechanics. *Rev. Mod. Phys.* **38**, 447–452 (1966).
42. Schmid, D., Ried, K. & Spekkens, R. W. Why initial system-environment correlations do not imply the failure of complete positivity: a causal perspective. *Phys. Rev. A* **100**, 022122 (2019).
43. Meyer, D. A. Finite precision measurement nullifies the Kochen-Specker theorem. *Phys. Rev. Lett.* **83**, 3751 (1999).
44. Pironio, S., Scarani, V. & Vidick, T. Focus on device independent quantum information. *New J. Phys.* **18**, 100202 (2016).
45. Mayers, D. & Yao, A. Quantum cryptography with imperfect apparatus. In *Proc. 39th Annual Symposium on Foundations of Computer Science, FOCS, Palo Alto 503–509* (IEEE, 1998).
46. Hensen, B. et al. Loophole-free Bell inequality violation using electron spins separated by 1.4 kilometres. *Nature* **526**, 682 (2015).
47. Giustina, M. et al. Significant-loophole-free test of Bell's theorem with entangled photons. *Phys. Rev. Lett.* **115**, 250401 (2015).
48. Shalm, L. K. et al. Strong loophole-free test of local realism. *Phys. Rev. Lett.* **115**, 250402 (2015).
49. Selby, J. H. et al. Accessible fragments of generalized probabilistic theories, cone equivalence, and applications to witnessing non-classicality. *Phys. Rev. A* **107**, 062203 (2023).
50. Schmied, R. et al. Bell correlations in a Bose-Einstein condensate. *Science* **352**, 6284 (2016).
51. Schmid, D., Selby, J. H., Rossi, V. P., Baldijão, R. D. & Sainz, A. B. Shadows and subsystems of generalized probabilistic theories: when tomographic incompleteness is not a loophole for contextuality proofs. *arXiv* <https://doi.org/10.48550/arXiv.2409.13024> (2024).
52. Aloy, A., Fadel, M., Galley, T. D., Jones, C. L. & Müller, M. P. Code and data for “Theory-independent monitoring of the decoherence of a superconducting qubit with generalized contextuality”. *Zenodo* <https://doi.org/10.5281/zenodo.17979407> (2024).
53. Barata, J. C. A. & Hussein, M. S. The Moore-Penrose pseudoinverse: a tutorial review of the theory. *Braz. J. Phys.* **42**, 146–165 (2012).
54. Gillis, N. & Glineur, F. Low-rank matrix approximation with weights or missing data is NP-hard. *SIAM J. Matrix Anal. Appl.* **32**, 1149–1165 (2011).
55. Markovsky, I. *Low Rank Approximation: Algorithms, Implementation, Applications* 906 (Springer, 2012).
56. Pál, K. F. & Vértesi, T. Maximal violation of a bipartite three-setting, two-outcome Bell inequality using infinite-dimensional quantum systems. *Phys. Rev. A* **82**, 022116 (2010).
57. Werner, R. F. & Wolf, M. M. Bell inequalities and entanglement. *Quantum Inf. Comput.* **1**, 1–25 (2001).
58. Troffaes, M. C. M. Working with polyhedron representations. <https://pypi.org/project/pycddlib/> (2025).
59. Lee, D. T. & Schachter, B. J. Two algorithms for constructing a Delaunay triangulation. *Int. J. Comput. Inf. Sci.* **9**, 219–242 (1980).
60. Müller, M. P. & Garner, J. P. Testing quantum theory by generalizing noncontextuality. *Phys. Rev. X* **13**, 041001 (2023).
61. Lee, C. M. & Selby, J. H. A no-go theorem for theories that decohere to quantum mechanics. *Proc. R. Soc. A* **474**, 20170732 (2018).
62. Richens, J. G., Selby, J. H. & Al-Safi, S. W. Entanglement is necessary for emergent classicality in all physical theories. *Phys. Rev. Lett.* **119**, 080503 (2017).
63. Holevo, A. *Probabilistic and Statistical Aspects of Quantum Theory* (North Holland, 1982).

## Acknowledgements

T.G. acknowledges helpful discussions with Robert Spekkens and Patrick Daley, and thanks Michael Mazurek for providing him with the data from ref. 10. These contributions helped with writing an initial version of part of the code used in the present paper. This research was funded by the Austrian Science Fund (FWF) 10.55776/PAT2839723; funded by the European Union—NextGenerationEU (AA, TDG, CLJ, MPM). Furthermore, this research was supported in part by Perimeter Institute for Theoretical Physics (MPM). Research at Perimeter Institute is supported by the Government of Canada through the Department of Innovation, Science, and Economic Development, and by the Province of Ontario through the Ministry of Colleges and Universities. M.F. was supported by the Swiss National Science Foundation Ambizione Grant No. 208886, and by The Branco Weiss Fellowship—Society in Science, administered by the ETH Zürich.

## Author contributions

M.F. performed the experiment, A.A., C.J., T.G., and M.M. worked out the theoretical results and wrote the paper, A.A., C.J., and T.G. performed the numerical analysis, and M.M. directed the research.

## 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/s41467-026-69030-x>.

**Correspondence** and requests for materials should be addressed to Albert Aloy, Matteo Fadel, Thomas D. Galley, Caroline L. Jones or Markus P. Müller.

**Peer review information** *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

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

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

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026