

# Experimentally informed decoding of stabilizer codes based on syndrome correlations

Ants Remm<sup>1,2,\*</sup>, Nathan Lacroix<sup>1,2,‡</sup>, Lukas Bödeker<sup>3,4</sup>, Elie Genois<sup>5</sup>, Christoph Hellings<sup>1,2</sup>, François Swiadek<sup>1,2</sup>, Graham J. Norris<sup>1,2</sup>, Christopher Eichler<sup>1,¶</sup>, Alexandre Blais<sup>5,6</sup>, Markus Müller<sup>3,4</sup>, Sebastian Krinner<sup>1,2,||</sup> and Andreas Wallraff<sup>1,2,7</sup>

<sup>1</sup>Department of Physics, *ETH Zurich*, CH-8093 Zurich, Switzerland

<sup>2</sup>Quantum Center, *ETH Zurich*, CH-8093 Zurich, Switzerland

<sup>3</sup>Institute for Theoretical Nanoelectronics (PGI-2), *Forschungszentrum Jülich*, 52428 Jülich, Germany

<sup>4</sup>Institute for Quantum Information, *RWTH Aachen University*, 52056 Aachen, Germany

<sup>5</sup>Institut Quantique, Département de Physique, *Université de Sherbrooke*, Sherbrooke, Quebec J1K 2R1, Canada

<sup>6</sup>Canadian Institute for Advanced Research, Toronto, Ontario M5G 1M1, Canada

<sup>7</sup>*ETH Zurich—PSI Quantum Computing Hub, Paul Scherrer Institute*, CH-5232 Villigen, Switzerland



(Received 3 March 2025; accepted 27 November 2025; published 16 January 2026)

High-fidelity decoding of quantum error correction codes relies on an accurate experimental model of the physical errors occurring in the device. Because error probabilities can depend on the context of the applied operations, the error model is ideally calibrated using the same circuit as is used for the error correction experiment. Here, we present an experimental approach guided by an analytical formula to characterize the probability of independent errors using correlations in the syndrome data generated by executing the error correction circuit. Using the method on a distance-three surface code, we analyze error channels that flip an arbitrary number of syndrome elements, including Pauli  $\hat{Y}$  errors, hook errors, multiqubit errors, and leakage, in addition to standard Pauli  $\hat{X}$  and  $\hat{Z}$  errors. We use the method to find the optimal weights for a minimum-weight perfect matching decoder without relying on a theoretical error model. Additionally, we investigate whether improved knowledge of the Pauli  $\hat{Y}$  error channel, based on correlating the X- and Z-type error syndromes, can be exploited to enhance matching decoding. Furthermore, we find correlated errors that flip many syndrome elements over up to eight cycles, potentially caused by leakage of the data qubits out of the computational subspace. The presented method provides the tools for accurately calibrating a broad family of decoders, beyond the minimum-weight perfect matching decoder, without relying on prior knowledge of the error model.

DOI: [10.1103/PhysRevResearch.8.013044](https://doi.org/10.1103/PhysRevResearch.8.013044)

## I. INTRODUCTION

Recent performance advances in quantum computing with superconducting qubits [1–6] have enabled experimental demonstrations of complex quantum computations [7–9]. State-of-the-art devices, however, still lack the performance to solve real-world problems [10]. Quantum error correction (QEC) promises to exponentially reduce the effective logical error rate at the cost of a polynomial increase in the number of qubits [11–13], as recently demonstrated experimen-

tally [6,14,15]. A widely pursued route to achieve this error suppression is the use of stabilizer codes [16], which rely on repeated measurements of a set of mutually commuting stabilizer operators. Physical errors are accompanied by changes in the stabilizer values, called the syndrome. The change of logical operator values can be decoded from the syndromes. While a main challenge in implementing quantum error-corrected circuits remains the realization of large-scale devices with low physical error rates, a topic of rising relevance is the fast and accurate decoding of error syndromes extracted from large circuits [17]. Decoders for experimental quantum error correction data have so far been calibrated by numerically optimizing for maximal logical fidelity [18,19], analyzing the correlations between syndrome elements [3,4,9], or numerically optimizing to match the higher-order correlations between syndrome elements [20–22]. The most performant decoders often rely on either an accurate error model of the device or a large dataset of training data in the case of machine-learning-based decoders.

In this work, we focus on developing an accurate error model by analyzing correlations in the experimental syndrome data of the target error correction circuit without relying on a theoretical device error model or on conducting separate calibration experiments. We present an analytical expression to calculate the probability of any independent error event

\*Contact author: ants.remm@gmail.com

†Present address: Atlantic Quantum, Cambridge, Massachusetts 02139, USA.

‡Contact author: nathanlacroix@google.com

§Present address: Google Quantum AI, Goleta, CA, USA.

¶Present address: Department of Physics, Friedrich-Alexander University Erlangen-Nürnberg (FAU), Erlangen, Germany.

||Present address: Zurich Instruments, CH-8005 Zurich, Switzerland.

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

that has a unique signature in the stabilizer measurement outcomes, based on the higher-order correlations in the experimental syndrome data. This equation generalizes the result by Spitz *et al.* [23], where an equation was derived for the probability of errors that flip at most two syndrome elements. A similar result was recently independently derived in Ref. [24]. Similarly to the error analysis method presented in Ref. [21], our method allows for the characterization of the full error model using the same circuit as is used for the QEC experiment. This error model can be used to optimize the parameters of various decoding algorithms, which we illustrate by presenting a correlated minimum-weight perfect matching (MWPM) decoder as an example. Specifically, the probabilities of the different error events inferred with our method translate to the weights of the matching graph. In addition, the full error model can be used for characterizing device performance under the same conditions as for the QEC experiment. For example, crosstalk or control errors can be identified from a discrepancy between Pauli  $\hat{X}$  and  $\hat{Y}$  error probabilities, since simple gate error models that include only dephasing and relaxation processes predict equal physical error probabilities for  $\hat{X}$  and  $\hat{Y}$  [25]. Correlated errors that flip syndrome elements over multiple rounds, on the other hand, can be indicative of qubit leakage [26–28] or error sources such as high-energy impact events [29] that are detrimental to the performance of the logical qubit. We apply the tools we present to experimental data from a distance-three surface code logical state preservation experiment performed on the 17-qubit device first introduced in Ref. [4].

The paper is structured as follows. First, we introduce our implementation of the distance-three surface code and detail how we extract the syndromes for a logical state preservation experiment (Sec. II). We then explain how we have implemented one of the most common decoding algorithms, the minimum-weight perfect matching decoder, in the experimental demonstration of quantum error correction with the surface code [4] (Sec. III). Finally, we present the theoretical tools that allow us to characterize the error model of the device based on the experimental syndrome data (Sec. IV). We find a significant presence of errors that correlate with leakage of the data qubits out of the computational subspace and have signatures that span multiple syndrome elements over many cycles, which we discuss in detail in Sec. V. These results highlight the capability of our method to determine the error probability associated with any observed syndrome signature and demonstrate its utility both in diagnosing device errors and in optimizing decoder parameters.

## II. FROM ERRORS TO SYNDROME

The experimental data we present were taken for a distance  $d = 3$  surface code [4]. We start by giving a brief description of the code and setting the terminology for the rest of the paper. The surface code consists of a  $d \times d$  square lattice of data qubits; see red circles in Fig. 1(a), which encode a protected quantum state. As a stabilizer code [16], it protects states that are simultaneous eigenstates of a set of commuting stabilizer generator operators (called *stabilizers* here), each taking the value of  $\pm 1$ . If we consider a square lattice with the data qubits at the vertices, the stabilizers of the surface

code are the products of Pauli operators of the data qubits on the vertices of each plaquette. The stabilizers  $\hat{S}_{Ai}$  alternate in the lattice between products of  $\hat{Z}$  and  $\hat{X}$  [see green and blue squares in Fig. 1(a)],

$$\hat{S}_{Xi} = \prod_{Dj \in N(Xi)} \hat{X}_{Dj} \quad \text{and} \quad \hat{S}_{Zi} = \prod_{Dj \in N(Zi)} \hat{Z}_{Dj}, \quad (1)$$

where  $N(Ai)$  denotes the set of data qubits on the vertices of the stabilizer plaquette. At the center of each plaquette, an auxiliary qubit  $Ai = Xi$  or  $Zi$  is used to measure the respective stabilizer. Additionally, there are stabilizers located at the boundary of the surface code lattice with a support on only two data qubits. In total, the stabilizers of a distance  $d$  surface code are adding up to  $(d-1)^2 + 2(d-1) = d^2 - 1$  stabilizers. By constraining the protected state to a mutual eigenstate of all the stabilizers, the dimension of the protected Hilbert space is reduced by a factor of 2 for each of the stabilizers, down to  $2^{d^2}/2^{d^2-1} = 2$ , meaning that the protected subspace corresponds to a single *logical qubit*. The Pauli operators of the logical qubit are defined as two anticommuting operators,

$$\hat{Z}_L = \prod_{Dj \in \text{row}} \hat{Z}_{Dj} \quad \text{and} \quad \hat{X}_L = \prod_{Dj \in \text{column}} \hat{X}_{Dj}, \quad (2)$$

which commute with all the stabilizers; see black lines in Fig. 1(a).

The stabilizers are repeatedly measured during the operation of the error correction code, yielding values  $s_m^{Ai}$  with  $m$  the index of the error correction cycle and  $Ai$  the associated auxiliary qubit. In our implementation, we use the circuit shown in Fig. 1(b), which consists of single-qubit Hadamard gates implemented using  $\pi/2$  rotations around  $\hat{Y}$  and virtual  $\hat{Z}$  rotations [30], and four conditional phase flip (CZ) gates [31–33] that map the parity of the data qubits to the auxiliary qubits, which are thereafter read out [34]. To reduce the number of two-qubit gates that are executed in parallel, we apply the gates for Z-type stabilizers while reading out the X-type auxiliary qubits and vice versa [4,35]. Therefore, the stabilizers of different types are read out at full and half-integer values of the cycle index  $m = 1, 1.5, 2, \dots$ , respectively. If the stabilizer value is  $-1$  (indicating an odd parity), the circuit flips the state of the corresponding auxiliary qubit. Since we do not reset the auxiliary qubits between error correction cycles, the measurement outcome for an auxiliary qubit associated with a stabilizer value of  $-1$  will alternate between  $-1$  and  $+1$  in consecutive cycles. Therefore, we infer the stabilizer value  $s_m^{Ai} = M_{m-1}^{Ai} M_m^{Ai}$  from the change in consecutive readout outcomes  $M_m^{Ai} = \pm 1$  rather than by the measurement outcome in a given cycle.

Because the stabilizers involving  $\hat{X}$  and  $\hat{Z}$  operators of each data qubit are repeatedly measured, all physical single-qubit errors are projected onto bit and phase flips [36,37]. The bit and phase flips of the physical qubits will cause some stabilizers to flip and might also flip the logical operator values  $\hat{X}_L$  and  $\hat{Z}_L$ , depending on where the errors occurred. We express the change of the stabilizer values in terms of the *syndrome elements*  $\sigma_m^{Ai} = (1 - s_{m-1}^{Ai} s_m^{Ai})/2$ , which have the value  $\sigma_m^{Ai} = 1$  or 0, if the value of stabilizer  $\hat{S}_{Ai}$  at cycle  $m$  changed or did not change, respectively. The process of deciding, based on

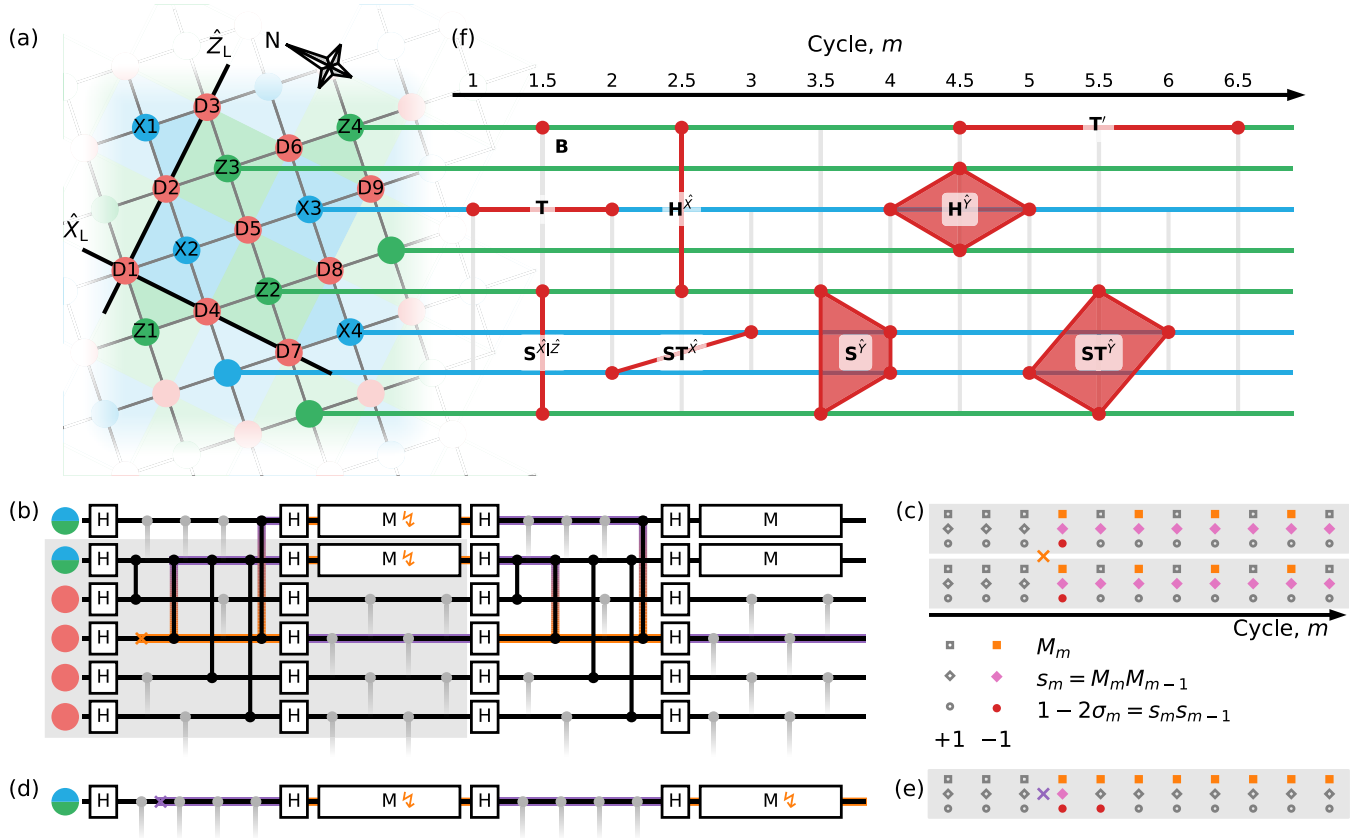


FIG. 1. (a) The surface code grid with data qubits in red and auxiliary qubits for Z- and X-type stabilizer measurement in green and blue, respectively. The physical device used in this work consists of the 17 qubits that are labeled. The support of the logical  $\hat{X}_L$  and  $\hat{Z}_L$  operators is indicated with black lines. A compass indicating the orientation of the lattice (pointing north) is used to define the East and West boundaries (see text). (b) The circuit for two cycles of stabilizer measurements, with data qubits as red circles and auxiliary qubits as split blue-green circles. The circuit section highlighted with gray background implements the measurement of a single stabilizer, the building block of the surface code. The propagation of an  $\hat{X}$  error (orange cross) is indicated with orange (as  $\hat{X}$ ) and purple (as  $\hat{Z}$ ) lines. The readout outcomes that are flipped as a result of the  $\hat{X}$  error are marked with  $\hat{z}$ . The grayed CZ gates involve qubits that are not explicitly shown in this circuit. (c) Conversion of the errors shown in panel (b) into stabilizer values  $s_m$  and syndrome elements  $\sigma_m$ . (d) The propagation of a  $\hat{Z}$  error (purple cross) on an auxiliary qubit during the parity map. (e) Conversion of the error in panel (d) into stabilizer values and syndrome elements. (f) Examples of error signatures due to common single-qubit Pauli errors. Pauli  $\hat{X}$  and  $\hat{Z}$  errors lead to two nonzero syndrome elements, categorized into timelike (T), spacelike ( $S^{\hat{X}|\hat{Z}}$ ), space-time ( $ST^{\hat{X}}$ ), and hook ( $H^{\hat{X}}$ ) error classes. Pauli  $\hat{Y}$  errors lead to error signatures with support on up to four auxiliary qubits and classified into spacelike ( $S^{\hat{Y}}$ ), space-time ( $ST^{\hat{Y}}$ ), and hook ( $H^{\hat{Y}}$ ) error classes. Errors on the boundary of the lattice can have a signature on a single auxiliary qubit only (B) and readout misclassification errors lead to timelike signatures on syndrome elements two cycles apart (T').

the syndrome elements, whether the logical operator values have flipped is called decoding.

To successfully decode errors, we must know which syndrome elements are flipped by each independent error process, i.e., the *signature* of that error, and whether it flips any of the logical qubit Pauli operators. As an example, let us consider a bit flip ( $\hat{X}$ ) error on one of the data qubits, indicated by the orange cross in Fig. 1(b). Note that this error is equivalent to a phase flip ( $\hat{Z}$ ) before the preceding Hadamard gate on that qubit. The error propagates to two neighboring auxiliary qubits as phase flips via the CZ gates, and the phase flips, in turn, change the outcomes of the following auxiliary qubit readouts. As the effect of the error remains on the data qubit, the auxiliary qubits will be flipped in every consecutive cycle, leading to an alternating pattern of readout outcomes; see Fig. 1(c). Converting the flips of readout outcomes  $M_m^{Ai}$  to

flips of stabilizer values  $s_m^{Ai}$  and to flips of syndrome elements  $\sigma_m^{Ai}$ , we find that the signature of such an error includes two syndrome element flips. The events are detected on neighboring auxiliary qubits during the same cycle. We label all errors with such signatures  $S^{\hat{X}|\hat{Z}}$  for errors with space-separated signatures due to  $\hat{X}$  and  $\hat{Z}$  errors on data qubits. Another common error is a phase flip ( $\hat{Z}$ ) error on an auxiliary qubit [see the purple cross in Fig. 1(d)], or equivalently, a bit flip error just before, during, or after readout. In this case, the error flips all following readout outcomes  $M_m^{Ai}$  on that auxiliary qubit  $Ai$ , which then results in two consecutive nonzero syndrome elements; see Fig. 1(e). We label all errors with such signatures as T for errors with time-separated signatures.

A few additional error classes are shown in Fig. 1(f); see Appendix A for examples of error propagation and Table I for a summary of all error classes we consider in this work. If an

TABLE I. Summary of error classes and symbols.

Error class	Symbol
Timelike	<b>T</b>
Spacelike ( $\hat{X}$ , $\hat{Z}$ )	$\mathbf{S}^{\hat{X} \hat{Z}}$
Spacelike ( $\hat{Y}$ )	$\mathbf{S}^{\hat{Y}}$
Space-time ( $\hat{X}$ )	$\mathbf{ST}^{\hat{X}}$
Space-time ( $\hat{Y}$ )	$\mathbf{ST}^{\hat{Y}}$
Hook ( $\hat{X}$ )	$\mathbf{H}^{\hat{X}}$
Hook ( $\hat{Y}$ )	$\mathbf{H}^{\hat{Y}}$
Boundary	<b>B</b>
Readout misclassification	$\mathbf{T}'$

$\hat{X}$  error occurs on a data qubit between the two **CZ** gates of one half cycle, then the syndrome elements are flipped on the neighboring auxiliary qubits in consecutive cycles. We label such errors as  $\mathbf{ST}^{\hat{X}}$  for errors with a space-time separated signature. If an  $\hat{X}$  error occurs on an auxiliary qubit that is used to measure a weight-four Z-type stabilizer at the middle of the parity map, then this error propagates as phase flip errors to two neighboring data qubits that perform a gate with that auxiliary qubit next. Depending on the gate order in the parity map, these phase flip errors will flip two or four neighboring syndrome elements of X type. We label such errors as  $\mathbf{H}^{\hat{X}}$  for  $\hat{X}$ -caused hook errors. Bit flip errors on the auxiliary qubit at other times during the parity map lead to a phase flip error on zero, one, three, or four neighboring data qubits. Because the neighboring qubits are part of a Z-type stabilizer, and the application of a stabilizer has no effect on the quantum state, then the error propagating to three or four neighboring qubits is equivalent to errors on the complementary one or zero neighboring data qubits. Therefore, these errors are indistinguishable from phase flip errors on data qubits, which belong to the  $\mathbf{S}^{\hat{X}|\hat{Z}}$  class. Again, due to the X-Z symmetry of the code, the situation is the same for  $\hat{X}$  errors on X-type auxiliary qubits, in which case Z-type syndrome elements are flipped.

In addition to  $\hat{X}$  and  $\hat{Z}$  errors discussed above,  $\hat{Y}$  errors can also occur at the same circuit locations. These errors have signatures flipping both Z- and X-type syndrome elements. We label the corresponding error classes as  $\mathbf{S}^{\hat{Y}}$ ,  $\mathbf{ST}^{\hat{Y}}$ , and  $\mathbf{H}^{\hat{Y}}$ . Near the boundary of the surface code lattice, errors can have signatures that include only a single syndrome element, which we will label boundary errors **B**. Finally, we label readout misclassification errors, for which the auxiliary qubit state is incorrectly classified, but does not get flipped, as  $\mathbf{T}'$  errors. These errors are equivalent to a correlated bit flip just before and after a perfect measurement, and they flip two syndrome elements on a single auxiliary qubit  $\Delta m = 2$  cycles apart.

We also consider two-qubit Pauli errors, not shown in Fig. 1(f), which might arise due to the two-qubit gates. We distinguish between the  $\mathbf{M}^{\hat{Z}\hat{Z}}$  class, corresponding to a correlated phase flip error  $\hat{Z} \otimes \hat{Z}$  on the two qubits involved in a two-qubit gate, and the  $\mathbf{M}^{\hat{X}\hat{Y}}$  class, corresponding to a correlated bit flip error  $\hat{X} \otimes \hat{X}$ ,  $\hat{X} \otimes \hat{Y}$ ,  $\hat{Y} \otimes \hat{X}$ , or  $\hat{Y} \otimes \hat{Y}$ . Other two-qubit Pauli errors are equivalent to a single-qubit Pauli error on one of the qubits before or after the **CZ** gate. While

we do not expect to observe correlated bit flips, errors of the  $\mathbf{M}^{\hat{Z}\hat{Z}}$  class could occur due to residual interactions between the qubits [38] or a miscalibrated conditional phase of the **CZ** gate.

### III. AUXILIARY QUBIT GRAPH AND DECODING

There are various strategies for decoding the syndrome data in the surface code [18,39–43]. Most accurate results can be achieved by maximum likelihood decoding [44], in which case all possible combinations of physical errors that are consistent with the observed syndrome data are considered. Each set of physical errors requires a corresponding correction of the logical qubit state, and the decoder picks the correction with the largest total likelihood. Since the number of physical error combinations is exponentially large in code size and number of executed cycles, maximum likelihood decoding is prohibitively expensive for all but the very smallest of codes [45]. Approximations of likelihood calculation using tensor networks can reduce complexity [46], but these methods remain too slow to meet the stringent timing requirements for large-scale error-corrected logical algorithms for which decoding must be performed between the application of non-Clifford gates in near-real time [9,21,47].

When looking at one type of auxiliary qubits at a time, the signature of each physical error is given by at most two syndrome element flips. Furthermore, a chain of neighboring errors leads to flipped syndrome elements only at the ends of the chain, since the syndrome element flips in the middle of the chain cancel with each other. The resulting property that syndrome flips come in pairs allows a much more efficient MWPM decoder [44,48] to be used. In MWPM decoding, all nonzero elements of the syndrome are matched in pairs. Once the matching is completed, a unique logical correction operator can be inferred based on whether or not the logical operator is expected to be flipped by the error chains in the matched graph. By assigning a weight to each potential pair of nonzero elements of the syndrome according to the likelihood of its occurrence, the problem of finding the most likely matching can be converted to finding the minimum-weight matching.

Compared to maximum likelihood decoding, this leads to two approximations. Foremost, by decoding the Z- and X-type stabilizers separately, we ignore correlations between syndrome types, which can contain extra information in the presence of errors that lead to flips of both types of stabilizers. Second, we find the most likely matching, but ignore that several (each individually less likely) matchings can potentially lead to the same logical correction. Consequently, the logical correction with the highest probability of success might differ from the correction given by the most likely matching.

Next, we will discuss the detailed procedure of minimum-weight perfect matching. We focus on this decoding approach to provide a pedagogical account of the methodology, as it is a widely used technique in the field and was employed in our previous work [4]. The first step, which can be done offline before running the surface code experiment, is to construct an *auxiliary qubit graph* (also called a matching graph [49], ancilla graph [50], or decoding graph), where each vertex  $k$  corresponds to a syndrome element  $\sigma_m^{A_i}$  of one type (X or Z), and each edge



$q = (k, k')$  corresponds to a statistically independent and therefore uncorrelated error process, which flips the connected syndrome elements  $k$  and  $k'$ ; see Fig. 2(a). In addition, there are two virtual vertices, which are connected to the syndrome elements that can flip without a pairing syndrome element at two opposing boundaries of the surface code lattice. Physically, multiple distinct independent errors can have the same syndrome signature. Since we infer probabilities directly from the syndrome data, these errors are indistinguishable and treated as a single process to construct the auxiliary qubit graph. An error probability  $p_q$  is associated with each edge of the auxiliary qubit graph. The edges and their probabilities amount to the effective error model of the device, which can be constructed either from an independent physical error model or based on the correlations in the syndrome data [23], the latter of which is the main topic of this work; see Sec. IV.

The second step of the MWPM decoding process is to construct the *syndrome graph*, a fully connected graph where vertices correspond to nonzero syndrome elements of the auxiliary qubit graph for a given experimental run; see Fig. 2(b) for an example. Each edge  $q$  in this graph is assigned a weight  $w_q$  defined as the negative logarithm of the total probability that any chain of errors flips only the syndrome elements connected by that edge. Up to first order in the error probabilities,  $w_q$  can be calculated as [50]

$$w_q \approx -\ln \left( \sum_{R \in \mathcal{R}} \prod_{r \in R} p_r \right), \quad (3)$$

where  $\mathcal{R}$  denotes the set of possible paths between the end points of  $q$  in the auxiliary qubit graph that do not go through the boundaries, and  $r$  are the edges in one of those paths  $R$ . A low edge weight indicates a high probability that a chain of errors triggers the corresponding syndrome pair, while a high weight suggests a low probability of such an event. These weights can be precalculated once before the decoding, at the cost of only a polynomial overhead in the code distance, as [50]

$$\mathbf{w} = -\ln((\mathbf{1} - \mathbf{A})^{-1} - \mathbf{1}), \quad (4)$$

where  $\mathbf{w}$  is a matrix of weights between all the potential nodes of the syndrome graph,  $\mathbf{A}$  (*adjacency matrix*) is the matrix of error probabilities in the auxiliary qubit graph,  $\mathbf{1}$  is the identity matrix, and the logarithm is taken elementwise.

In a logical state preservation experiment, where the same stabilizers are measured repeatedly, the system is expected to exhibit time-translational invariance. Consequently, the error probabilities are expected to remain constant from cycle to cycle during the middle of the error correction process; see Appendix B. Thus, one can calculate the error probabilities as a function of only the cycle difference  $\Delta m = m' - m$  instead of cycles  $m$  and  $m'$ ,

$$P_{(\sigma_m^{A_i}, \sigma_{m'}^{A_j})} = P_{A_i, A_j}^{(\Delta m)}. \quad (5)$$

Note that in the first measurement cycles after the logical state preparation and in the last measurement cycles before the final measurement, the assumed time-translational invariance of the protocol does not hold. Therefore, using this approximation of time-invariant error probabilities can lead to slightly suboptimal performance.

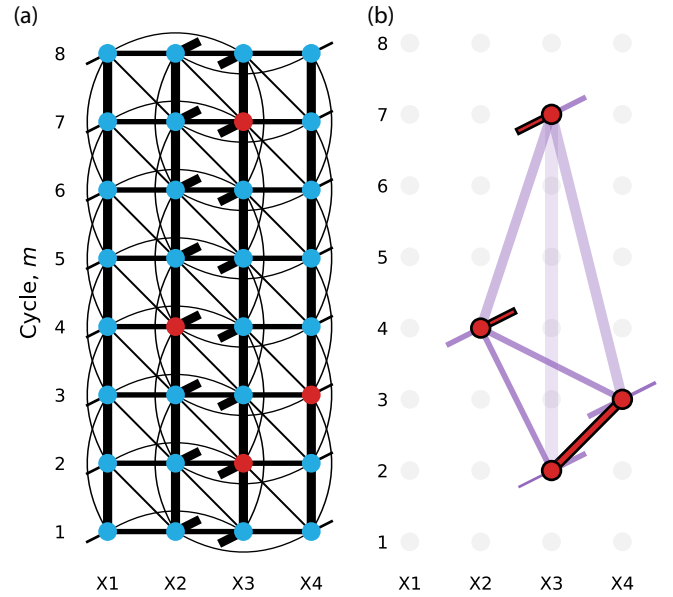


FIG. 2. (a) The auxiliary qubit graph for an example experimental run consisting of eight cycles of syndrome extraction. Vertices correspond to the measured syndrome elements (red if the syndrome element has a value of 1 and blue otherwise), while the edge thickness indicates the per-cycle probability of an independent error that flips the connected syndrome elements. The widths of the single-ended edges pointing to top right and bottom left indicate boundary error probabilities and their direction indicates to which boundary the syndrome element is connected. (b) The syndrome graph. Vertices correspond to the nonzero syndrome elements of the auxiliary qubit graph, and the thickness and transparency of the edges indicate the weight for the MWPM decoding. The minimum-weight matching for this graph is indicated in red with a black border. Note that in the auxiliary qubit graph thick lines indicate probable errors, but in the syndrome graph, thin lines indicate low-weight error chains, i.e., probable error chains. Due to the small distance ( $d = 3$ ) of the code, the four X-type auxiliary qubits can be spacially connected in a one-dimensional chain through data qubits D2, D5, and D8 [see Fig. 1(a)]. This allows the corresponding syndrome graph to be visualized in a planar layout, with the horizontal axis representing the spatial dimension and the vertical axis representing different cycles. For larger code distances, the graph would require a three-dimensional perspective.

For example, for X-type stabilizers in a  $d = 3$  surface code [4], we can calculate the adjacency matrix for a maximum cycle difference of  $\Delta m_{\max} = 2$  as

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}^{(0)} & \mathbf{A}^{(1)} & \mathbf{A}^{(2)} & \mathbf{0} & \mathbf{0} & \mathbf{B} \\ \mathbf{A}^{(-1)} & \mathbf{A}^{(0)} & \mathbf{A}^{(1)} & \mathbf{A}^{(2)} & \mathbf{0} & \mathbf{B} \\ \mathbf{A}^{(-2)} & \mathbf{A}^{(-1)} & \mathbf{A}^{(0)} & \mathbf{A}^{(1)} & \mathbf{A}^{(2)} & \mathbf{B} \\ \mathbf{0} & \mathbf{A}^{(-2)} & \mathbf{A}^{(-1)} & \mathbf{A}^{(0)} & \mathbf{A}^{(1)} & \mathbf{B} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}^{(-2)} & \mathbf{A}^{(-1)} & \mathbf{A}^{(0)} & \mathbf{B} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad (6)$$

with the cycle-shifted adjacency matrices containing the probabilities of errors for which the signature spans  $\Delta m$  cycles

$$(\mathbf{A}^{(\Delta m)})_{i,j} = p_{X_i, X_j}^{(\Delta m)} \quad \text{and} \quad (\mathbf{A}^{(0)})_{i,i} = 0, \quad (7)$$

and the boundary adjacency matrix contains the probabilities of errors with a signature on a single syndrome element

$$\mathbf{B} = \begin{pmatrix} p_{X1, BE} & 0 \\ 0 & p_{X2, BW} \\ p_{X3, BE} & 0 \\ 0 & p_{X4, BW} \end{pmatrix}. \quad (8)$$

Following the distance-three surface code layout and orientation indicated in Fig. 1(a), we define an east and a west boundary (BE and BW). To construct the boundary adjacency matrix, note that the stabilizers X1 and X3 connect only to the east boundary (BE via D3, D6, and D9), and the stabilizers X2 and X4 to the west boundary (BW via D1, D4, and D7). We have set the corresponding unconnected elements of  $\mathbf{B}$  to zero. We then apply Eq. (4) to precalculate the weights for syndrome pairs up to  $\Delta m_{\max}$  cycles apart.

As the final step, we run a minimum-weight perfect matching algorithm on the syndrome graph [49–51]. The space component of each matched edge corresponds to a set of data qubits where an error has occurred. The time component, arising from the space-time and auxiliary qubit errors, has no direct effect on the logical operator value. When decoding the X-type syndrome, each overlap of a data qubit in a matched edge with the logical  $\hat{X}_L$  operator corresponds to a flip of  $\hat{X}_L$  that should be corrected. Correspondingly, flips of the operator  $\hat{Z}_L$  can be inferred from the Z-type syndrome data.

Note that recent implementations of minimal-weight perfect-matching decoders [52,53] achieve faster run times by directly finding the perfect matching on the sparse auxiliary qubit graph. This approach eliminates the need to construct the syndrome graph with all-to-all connected edges and assigned weights [52].

With the goal of further improving the weight-inference-based MWPM decoding, we investigate the use of a correlated MWPM decoder (detailed in Appendix C). This decoder is designed to better correct for  $\hat{Y}$  errors, which flip both X- and Z-type syndrome elements and thereby create correlations between the two syndrome types. For the current distance and error rates of the device, we do not observe significant performance improvements, suggesting that the logical error per cycle is not currently limited by  $\hat{Y}$  errors. Nevertheless, we expect this approach to become increasingly beneficial at larger distances [54,55]; see Appendix C for details.

#### IV. FROM SYNDROME CORRELATIONS TO ERROR PROBABILITIES

Having a good quantitative knowledge of the experimental errors occurring on a device is of high importance for high-fidelity decoding, independently of the exact decoding algorithm used. While many error processes can be characterized using independent measurements, the effective error rates might differ when running the actual error correction experiment, e.g., due to time drift of parameters or unaccounted-for error mechanisms, like crosstalk. Therefore, we ideally want to construct an error model for the decoder based on syndrome data produced by running the same circuit as for the error correction experiment. In this section, we explain, based on experimental data, how this can be done.

Using the device presented in Ref. [4] (see also Appendix D), we prepare the logical state  $|0\rangle_L$ ,  $|1\rangle_L$ ,  $|+\rangle_L$ , or  $|-\rangle_L$ , acquire 16 cycles of stabilizer measurements, and finally read out all qubits. For each state, we perform 500 000 experimental runs, and remove the ones in which any of the physical qubits are measured outside the computational subspace, or the initial state is not the ground state, leaving us with about 54 000 runs per state. In the future, leakage reduction units as well as leakage-aware decoders could be employed to eliminate the need for postselection on no-leakage events [26–28,56,57]. For calculating the syndrome elements for the first and last cycles, we make use of the known initial state of the auxiliary qubits, the initial parity of the data qubits, and the parity of the data qubits from the final readout.

The average per-cycle probability of detecting a nonzero syndrome element on weight-four and weight-two stabilizers is  $\langle \sigma \rangle = 0.165(17)$  and  $0.118(6)$ , respectively, where the uncertainty indicates the standard deviation across different auxiliary qubits; see Appendix B. To visualize the correlations between syndrome elements, characteristic of the error classes discussed in Sec. II, we calculate the covariances  $C_{Ai,Aj}^{(\Delta m)}$  between syndrome elements on auxiliary qubits  $A_i$  and  $A_j$ ,  $\Delta m$  cycles apart, as

$$C_{Ai,Aj}^{(\Delta m)} = \langle \sigma_m^{Ai} \sigma_{m+\Delta m}^{Aj} \rangle - \langle \sigma_m^{Ai} \rangle \langle \sigma_{m+\Delta m}^{Aj} \rangle. \quad (9)$$

The averaging  $\langle \cdot \rangle$  is done first over the experimental runs, and then over the cycle index  $m$ . We omit syndrome elements from the first and the last cycle when calculating the correlations, since these syndrome elements are measured based on the initial or final state of the data qubits, and are not representative of the mean value during the bulk of the experiment.

For  $\Delta m = 0$ , we observe the highest covariance between neighboring auxiliary qubits; see the first off-diagonal elements of the matrix in Fig. 3(a). These correspond to correlated syndrome element flips due to  $\mathbf{S}^{\hat{X}|Z}$  and  $\mathbf{S}^{\hat{Y}}$  errors. The second off diagonal, which has significant but lower covariance, corresponds to correlated syndrome element flips on next-nearest neighbors, caused by hook errors such as  $\mathbf{H}^{\hat{X}}$  and  $\mathbf{H}^{\hat{Y}}$ . We find negligible covariance between syndrome elements that are spatially separated by more than two data qubits (third off diagonal), suggesting that errors remain local, as expected for uncorrelated single-qubit errors. Because we use a pipelined stabilizer measurement circuit in which stabilizers of different types are measured sequentially, the covariance between syndrome elements of different types of auxiliary qubits is detected at a half-cycle separation, i.e.,  $\Delta m = 0.5$ . The nonzero covariances occur between neighboring auxiliary qubits and are caused by  $\mathbf{S}^{\hat{Y}}$  and  $\mathbf{ST}^{\hat{Y}}$  errors. We observe the highest covariance between syndrome elements on a single auxiliary qubit between consecutive cycles of error correction  $\Delta m = 1$ , which corresponds to  $\mathbf{T}$  errors. The high covariance is an indication of the relatively higher probability of  $\mathbf{T}$ -class errors, being caused by auxiliary qubit dephasing and readout errors. The nonzero covariances on the first off diagonal are caused by  $\mathbf{ST}^{\hat{X}}$  and  $\mathbf{ST}^{\hat{Y}}$  errors. There is almost no covariance for  $\Delta m = 1.5$ , and only diagonal elements for  $\Delta m = 2$ . An expected source of errors leading to  $\Delta m = 2$  correlations is readout misclassification errors  $\mathbf{T}'$ , where the auxiliary qubit

state is misclassified without changing the qubit state. However, as we discuss in Sec. V, most of these correlations are due to errors that can flip several syndrome elements over many cycles, possibly related to leakage of the data qubits outside the computational subspace.

For calculating the weights for minimum-weight perfect matching decoding, we have to convert those covariances

into per-cycle error probabilities. An analytical formula for this, assuming that every error flips at most two syndrome elements, is presented by Spitz *et al.* [23]. Here, we have generalized those equations for errors with signatures that flip an arbitrary number of syndrome elements. In the general case, the probability  $p_{i_1, \dots, i_n}$  that an error that flips  $n$  syndrome elements  $\sigma_{i_1}, \dots, \sigma_{i_n}$  occurs can be calculated as

$$p_{i_1, \dots, i_n} = \frac{1}{2} - \frac{1}{2} \frac{\prod_{\{j_1, \dots, j_m\} \subseteq \{i_1, \dots, i_n\}} \langle \tilde{\sigma}_{j_1} \dots \tilde{\sigma}_{j_m} \rangle^{(-1)^{m-1} 2^{-(n-1)}}}{\prod_{\{j_1, \dots, j_m\} \supset \{i_1, \dots, i_n\}} (1 - 2p_{j_1, \dots, j_m})}; \quad (10)$$

see Appendix E for the verification of this formula and how it relates to the formulas introduced in Ref. [23]. Here, we have denoted  $\tilde{\sigma}_i = 1 - 2\sigma_i$  for brevity, and the indices  $i$  and  $j$  each include both the auxiliary qubit index and the cycle number at which the syndrome element is measured. The product in the numerator is taken over all the subsets  $\{j_1, \dots, j_m\}$  of arbitrary length  $m$  of the set of syndrome elements  $\{i_1, \dots, i_n\}$ , including the set itself, whereas the product in the denominator is taken over the supersets of a different arbitrary length  $m$ . We provide the explicit form of the equation when accounting for errors that can trigger up to four syndrome elements simultaneously in Appendix E. Compared to numerical methods for finding the probability of an error that flips a specific set of syndrome elements (see Sec. VI C of Supplemental Material of Ref. [21]), using an explicit formula will likely be faster and will not suffer from convergence problems.

Next, using Eq. (10), we calculate the probabilities of errors that trigger various error signatures in the same dataset as was used to calculate the correlations shown in Fig. 3(a). Because each run of the  $N = 16$ -cycle-long experiment produces  $(d^2 - 1)(2N - 1)/2 = 124$  syndrome elements, it is unfeasible to calculate the error probability for all of the  $2^{124}$  signatures. We therefore only consider error signatures that are caused by single-qubit Pauli errors at any position in the circuit or two-qubit Pauli errors during a two-qubit gate, as presented in Sec. II. To avoid introducing a bias into the extracted error probabilities by omitting the renormalization by  $p_{j_1, \dots, j_m}$  in the denominator in Eq. (10), we need to make sure that we include all processes with highly correlated signatures present in the system; see Appendix F. Failure to do so may result in nonphysical (negative) error probabilities beyond statistical fluctuations that require additional corrections [58]. To address this, we include two additional error signatures in the analysis that account for highly correlated errors, which we label **C**; see Fig. 3(c). First, we consider any subset of flipped syndrome elements on up to nine consecutive cycles on a single auxiliary qubit, and second, we consider any subset of syndrome elements on auxiliary qubits neighboring one data qubit, separated by up to two cycles. The maximum cycle separation for both of these signature classes is limited by the computational power, as the number of signature subsets grows exponentially with the maximum cycle separation. The **C** error class could be associated with undetected leakage of the data qubits to higher transmon states, as we will show

in Sec. V. We average the probabilities of errors with the same signature over the cycles (except the first and the last cycle) to obtain the average probabilities for the 116 error signatures stemming from Pauli errors in the circuit and 4360 error signatures belonging to the **C** class. The total probability of an error per cycle from each of the 12 signature classes is shown in Fig. 3(b).

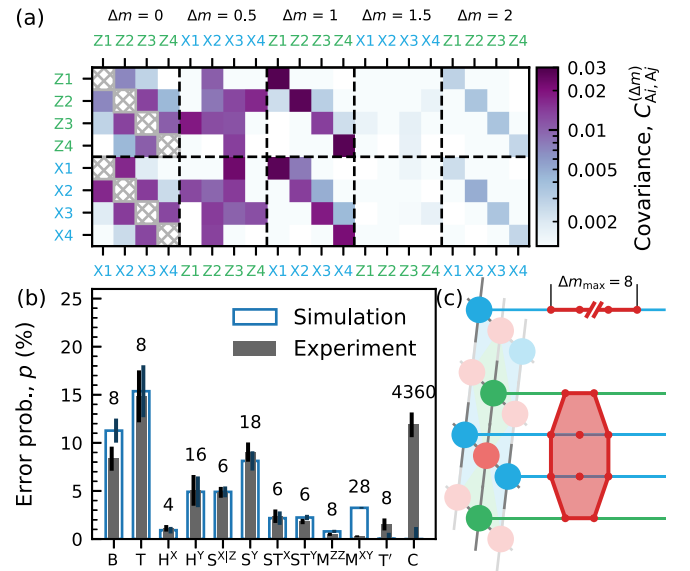


FIG. 3. (a) Average (over 16 cycles) covariance between pairs of syndrome elements as a function of the time separation  $\Delta m$  between them. The syndrome indicated by the column is detected  $\Delta m$  cycles after the syndrome indicated by the row. The covarying qubit is indicated on the top or bottom axis for Z- and X-type auxiliary qubits, respectively. (b) Total error probability  $p$  of the various error classes extracted from the experimental (solid gray bar) and simulated (blue wireframe) syndrome correlations. The number above the bar indicates the number of different error signatures that were considered in that class. The error bars indicate the standard deviation of the total, calculated as the square root of the sum of the squared deviations from the mean within that class. (c) Schematic of the signatures that are considered as the highly correlated **C** error class. That is, any subset of the highlighted syndrome elements on one auxiliary qubit over nine cycles or on multiple auxiliary qubits neighboring one data qubit separated by up to two cycles.

We compare the experimental results to a circuit-level Pauli-error simulation, shown as blue wireframes in Fig. 3(b). The simulation assumes uniform error probabilities across the device, with the probability values based on independent calibration measurements; see Appendix G for details. Despite this simplification, we observe overall good agreement between the sum of error probabilities of each class obtained from simulation and those extracted from the experiment. The most notable difference is in the correlated bit flips due to CZ gates, indicated by the  $M^{\hat{X}\hat{Y}}$  error class. These errors are included in the depolarizing error model of the simulation, but are not as pronounced in the experimental data. In simulations with a depolarizing Pauli noise model, two-qubit gate errors have been found to contribute most significantly to the logical error probability [34]. However, if correlated bit flip errors are rare for the experimental implementation of the CZ gate, as we observe in our data, the effect of two-qubit gate errors on the logical error probability in experimental realizations might be more akin to single-qubit errors. Note that the simulations do not include leakage and readout misclassification errors, which could explain some of the residual differences between simulations and experimental data.

The *in situ* error characterization method presented here is especially useful for stabilizer codes, since it yields the probabilities of errors with a given signature, which is exactly the information needed for decoding, without resorting to simulation of the circuit. Furthermore, the method allows us to analyze spurious correlations between syndrome elements not caused by known Pauli errors, and to identify whether they lead to highly correlated errors, which are known to be particularly harmful in quantum error correction [59,60].

## V. DEVICE DIAGNOSTICS USING ERROR PROBABILITIES

In this section, we study the per-cycle probabilities of specific errors with various signatures. The signature of a circuit-level error in the syndromes depends on the Pauli error itself (for instance, single-qubit  $\hat{X}$ ,  $\hat{Y}$ , or  $\hat{Z}$  error) and on its location in the circuit. In addition, several combinations of Pauli errors and circuit locations can lead to the same error signature. Therefore, we start by categorizing the error signatures according to the number  $\nu$  of possible combinations of Pauli errors and circuit locations that can trigger it. Specifically, we consider 15 two-qubit Pauli errors occurring at any of the 24 two-qubit gates of a single cycle, amounting to a total of  $24 \times 15 = 360$  possible combinations; see Appendix H. We omit signatures from the timelike errors class **T** in this analysis, since readout errors strongly enhance their probability. If uniform depolarizing noise at every two-qubit gate was the only error source, we would expect the error probability to be directly proportional to the number of combinations that can trigger that signature. Overall, we indeed observe a strong correlation between the probability of an error signature and the number of combinations of Pauli errors and circuit locations that can produce it; see Fig. 4(a). This correlation suggests that the variation in the probabilities of different error signatures is largely determined by how frequently errors that can produce each signature arise in the circuit. The correlation is also

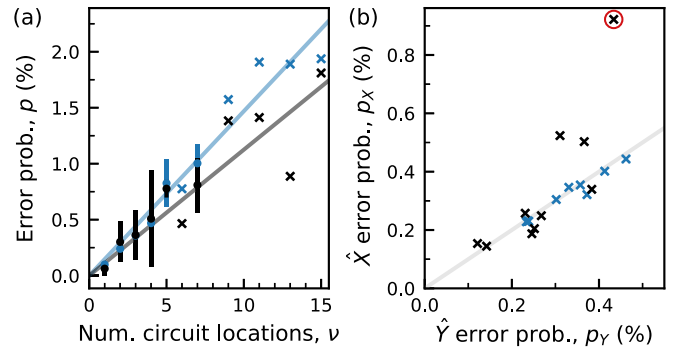


FIG. 4. (a) The probability  $p$  of triggering an error signature as a function of the number of combinations of Pauli errors and circuit locations  $\nu$ , which can cause the signature, extracted from experimental (black) and simulation (blue) data. If there are multiple signatures with the same number of circuit locations  $\nu$ , then their mean probability and standard deviation are shown as dots and error bars, while crosses indicate probabilities of single signatures. The black (blue) line indicates a least-squares fit of the experimental (simulated) data to a linear model. (b) Comparison of error probabilities,  $p_X$  and  $p_Y$ , for signature pairs that can only be caused by an  $\hat{X}$  or a  $\hat{Y}$  error, respectively, at a specific location. Black crosses correspond to experimental data and blue crosses to simulation. Due to the symmetry of the physical error mechanisms with respect to  $\hat{Z}$  rotations, we expect the two probabilities to be identical, which is indicated by the diagonal gray line. The red circle highlights error probabilities of D2 during the X-type parity map, which deviate significantly from the expectation (see text for details).

reproduced by the simulation with independently characterized parameters [blue markers and line in Fig. 4(a)].

Another application of the generalized error probability extraction method is the identification of crosstalk and control errors. For some circuit locations, there are unique signatures for both  $\hat{X}$  and  $\hat{Y}$  errors, belonging to the classes of space-time errors or hook errors due to  $\hat{X}$  or  $\hat{Y}$  type errors, denoted as  $ST^{\hat{X}}$  or  $H^{\hat{X}}$  and  $ST^{\hat{Y}}$  or  $H^{\hat{Y}}$ , respectively. Since the physical energy relaxation and pure dephasing of the qubits is described by an error channel that is symmetric with respect to  $\hat{X}$  and  $\hat{Y}$ , we expect the corresponding extracted error probabilities  $p_X$  and  $p_Y$  to be equal [25]. We find  $p_X \approx p_Y$  for all error pairs in simulation and in experiment, see Fig. 4(b), except for errors on D2 during the X-type parity map (black cross circled in red), for which  $p_X \approx 2p_Y$ . Control errors arising from miscalibrations that cause systematic under- or overrotations could explain this discrepancy because they can increase the likelihood of  $\hat{X}$  errors without increasing the probability of  $\hat{Y}$  errors, causing  $p_X$  to be larger than  $p_Y$ . An alternative source for this undesired rotation could be microwave crosstalk during the X-type parity map affecting D2. However, this explanation appears less likely, as an independent characterization of microwave crosstalk indicates that D2 is not significantly affected by microwave pulses applied to the drive lines of other qubits on the device (see Supplemental Material of Ref. [4]).

Next, we analyze the potential origin of the signature class of errors that cause longer time correlations, introduced as **C**. As the first observation, we investigate the timelike



correlations in the experimental syndrome data, and find that the covariance between syndrome elements on a single auxiliary qubit  $\Delta m$  cycles apart decays exponentially as  $0.89^{\Delta m}$ ; see Fig. 5(a). Ideally, we expect the correlations for  $\Delta m > 2$  to be zero. The slow decay could be an indication of leakage of the data qubits [61] that is not removed by the leakage rejection scheme based on the final readout of the data qubits. This hypothesis is supported by the observation that without including long-timescale error-class signatures in the analysis, the error probability for  $\mathbf{T}'$  signatures is higher at the beginning of the experiment when the leaked data qubits have several cycles to seep [62] back to the computational subspace before the final readout at the end of the experiment; see black filled dots in Fig. 5(b). If we do not reject any runs based on the final data qubit readout, then the error probability does not depend on the cycle number and is comparable to the error probability with leakage rejection close to the first cycle; see open black circles in Fig. 5(b). Alternatively, the slow decay could also be an indication of the probability of bit flip errors varying during the data acquisition period, which would introduce correlations between errors with a large time separation; see Appendix I for an example. The varying error rate could be caused by the asymmetry of the energy relaxation channel and readout errors, implying that the effective bit flip error rate for auxiliary qubits depends on whether they spend more time in the ground or excited state, or by changes in the rate of quasiparticle generation and tunneling, for example, due to impacts with cosmic rays [29]. Because quasiparticle tunneling can also cause leakage [63], then it would be consistent with the suppression of errors when data qubit leakage rejection is used, but asymmetry of the energy relaxation channel alone could not explain our observations.

With the goal of taking into account highly correlated noise processes of the experiment, we include the **C** class signatures in the analysis. Due to the renormalization term in the denominator of Eq. (10), accounting for these additional processes gives us a more accurate value for the probability of  $\mathbf{T}'$  errors. With this analysis, we find that the error probability of readout misclassification errors  $\mathbf{T}'$  only weakly depends on the cycle number and whether data qubit leakage rejection is used or not; see filled and open purple squares in Fig. 5(b). Furthermore, the error probabilities extracted using the improved analysis are consistent with the separately characterized overlap error of auxiliary qubit readout, which is the expected mechanism for  $\mathbf{T}'$  errors; see Fig. 5(c).

## VI. CONCLUSION

An accurate error model of the quantum error correction circuit is a crucial component for any high-fidelity decoder. In this work, we explained in detail how the syndrome is generated under a circuit-level Pauli noise model. We presented a closed-form analytical method for calculating the error probabilities of errors with a given signature from the correlations between syndrome elements of arbitrary weight. We used these error probabilities to calculate the weights of a minimum-weight perfect matching decoder as used in Ref. [4], the nuances of which we also explained. Furthermore, a correlated matching decoder is employed to harness the correlations introduced by  $\hat{Y}$  errors between both

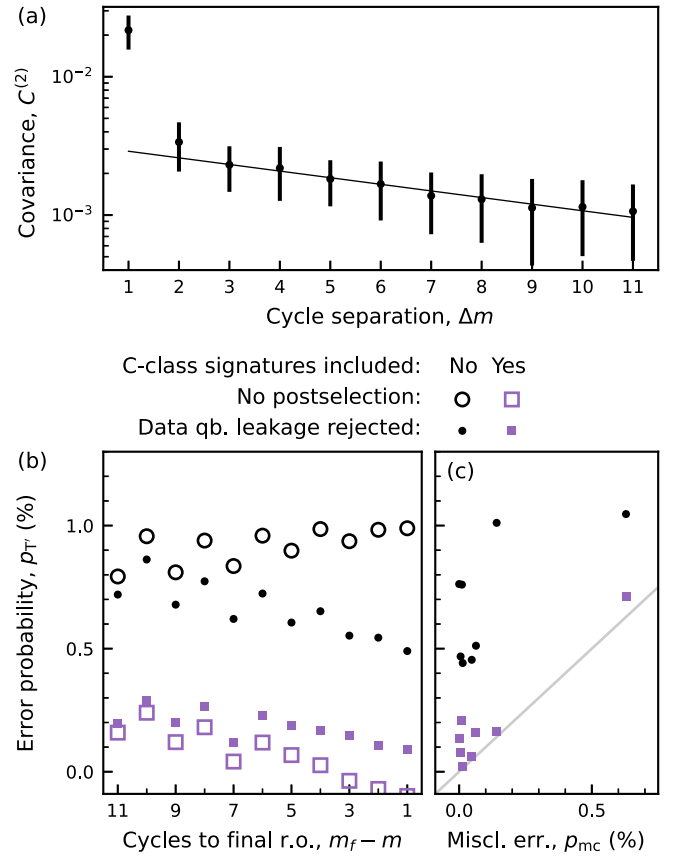


FIG. 5. (a) Covariance between syndrome elements on a single auxiliary qubit  $\Delta m$  cycles apart, averaged over the eight auxiliary qubits. The error bars indicate one standard deviation over the different qubits. A fit to an exponential decay model for  $\Delta m > 2$  is shown as a black line. (b) Extracted probability of an error triggering a syndrome configuration corresponding to measurement misclassification  $\mathbf{T}'$  ( $\Delta m = 2$ ) as a function of the number of cycles until the final data qubit readout. The four data series correspond to using data qubit leakage rejection or not and including leakage-related syndromes in the analysis or not. The data are averaged over the auxiliary qubits. (c) Average (over cycles)  $\mathbf{T}'$  error probability of the eight auxiliary qubits as a function of the independently characterized readout misclassification error  $p_{mc}$ . The gray line indicates the expected identity relation  $p_{\mathbf{T}'} = p_{mc}$ .

syndrome graphs to increase the decoding performance by a modest but stable amount. In addition, the error model can be used to analyze the crosstalk, control, and leakage errors, measured using the same circuit as is used for executing the error correction experiment. We identified control errors on one of the data qubits, leading to an imbalance of  $\hat{X}$  and  $\hat{Y}$  errors. We also identified errors with signatures spanning multiple syndrome extraction cycles, which are consistent with undetected leakage on the data qubits or quasiparticle tunneling.

Although characterizing errors based on syndrome data taken directly from executing the error correction circuit can be a powerful tool, one needs to be aware of its limitations. First, even though the analytical formula allows to easily calculate the probability of errors that trigger an arbitrary number of syndrome elements, it is not feasible to calculate the probabilities for triggering all possible signatures, as the

number of signatures is exponentially large in the number of syndrome elements. Therefore, some errors with nonstandard signatures might be excluded from the model. Second, this method requires many experimental runs of the error correction circuit to accurately estimate the syndrome element correlations used to calculate the error probabilities. Because of the renormalization terms in the denominator of Eq. (10), the more high-weight signatures are included in the model, the more the uncertainties of errors with lower-weight signatures increase. Furthermore, if the error model parameters drift during the time it takes to gather the necessary amount of statistics, then spurious correlations can appear in the data.

The main advantage of the presented method is that the error model is extracted from experimental data obtained by executing the same quantum circuit that we used to run the quantum error correction code. This allows for the identification of error sources, such as crosstalk, that may not be captured in isolated characterization experiments. Furthermore, the fact that the probability of errors with signatures of arbitrary weight can be calculated is beneficial for identifying error models that include  $\hat{Y}$  errors. This can be used for improved decoders that account for correlations between X- and Z-type stabilizers, which we demonstrated for a modified version of the matching decoder [64]. The inference of error processes causing higher-weight signatures can potentially benefit also other decoders, such as belief matching [43] and tensor-network-based [46] decoders. As we have shown, the method can also be used to characterize errors with high-weight signatures, which can cause systematic deviations in extracted single- and two-qubit error probabilities if unaccounted for.

### ACKNOWLEDGMENTS

The authors are grateful for valuable feedback on the manuscript by N. Delfosse, M. S. Peralta, and P. Girardin. We acknowledge financial support by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), through the U.S. Army Research Office Grant No. W911NF-16-1-0071, by the EU Flagship on Quantum Technology H2020-FETFLAG-2018-03 project 820363 OpenSuperQ, by the National Center of Competence in Research “Quantum Science and Technology” (NCCR QSIT), a research instrument of the Swiss National Science Foundation (SNSF, Grant No. 51NF40-185902), by the SNSF R’Equip Grant 206021-170731, by the EU programme H2020-FETOPEN project 828826 Quomorphic, and by ETH Zurich. S.K. acknowledges financial support from Fondation Jean-Jacques et Félicia Lopez-Loreta and the ETH Zurich Foundation. L.B. and M.M. acknowledge support by the Deutsche Forschungsgemeinschaft through Grant No. 449905436. M.M. furthermore acknowledges support by the German Research Foundation (DFG) under Germany’s Excellence Strategy “Cluster of Excellence Matter and Light for Quantum Computing (ML4Q) EXC 2004/1” 390534769, and by the ERC Starting Grant QNets through Grant No. 804247. L.B. gratefully acknowledges the computing time provided to us at the NHR Center NHR4CES at RWTH Aachen University (project number p0020074). This is funded by the BMFTR and the state governments participating based on

the resolutions of the GWK for national high-performance computing at universities. E.G. and A.B. acknowledge support from NSERC, the Canada First Research Excellence Fund, and the Ministère de l’Économie et de l’Innovation du Québec.

The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government.

### DATA AVAILABILITY

The data that support the findings of this article are not publicly available upon publication because it is not technically feasible and/or the cost of preparing, depositing, and hosting the data would be prohibitive within the terms of this research project. The data are available from the authors upon reasonable request.

### APPENDIX A: ERROR SIGNATURES

To see which error signatures are expected due to physical errors, we place single-qubit  $\hat{X}$  and  $\hat{Z}$  errors between each pair of neighboring operations in the syndrome extraction circuit, and propagate the errors through the circuit to see which auxiliary qubit readout outcomes are flipped due to the error. The set of single-qubit Pauli errors shown in Fig. 6 covers all possible error locations if we account for the symmetry of the error extraction circuit with respect to change of the auxiliary qubit types. The propagation and signatures of single-qubit  $\hat{Y}$  errors and errors of class  $\mathbf{M}^{\hat{Z}\hat{Z}}$  and  $\mathbf{M}^{\hat{X}\hat{Y}}$  can be constructed as the sum of the propagations of the constituent  $\hat{X}$  and  $\hat{Z}$  errors.

### APPENDIX B: SYNDROME PROBABILITY VERSUS CYCLE

To motivate the assumption of cycle-independent error rates, we calculate the mean syndrome element values for each stabilizer and each cycle; see Fig. 7. The stabilizers matching the type of the prepared state are measured on whole-integer cycles, while the other stabilizers are read out on half-integer cycles in the pipelined surface code circuit. The syndrome elements at the middle of the experiment are calculated based on auxiliary qubit measurement flips over two error correction cycles and are therefore sensitive to physical errors happening during two error correction cycles. On the other hand, the first and last syndrome elements for the stabilizer type of the prepared logical operator ( $\hat{Z}_L$  or  $\hat{X}_L$ ) make use of data qubit readout results to calculate the value of one of the stabilizers and are sensitive to errors during a single round of error correction. Therefore, the syndrome elements and error probabilities at the time boundaries are expected to have a different mean value. Overall, we find that the mean syndrome element values are increasing very slightly over the course of the 16 experimental cycles.

### APPENDIX C: CORRELATED MWPM DECODING

In this Appendix, we present a correlated MWPM decoder, with the goal of improving the weight-inference-based

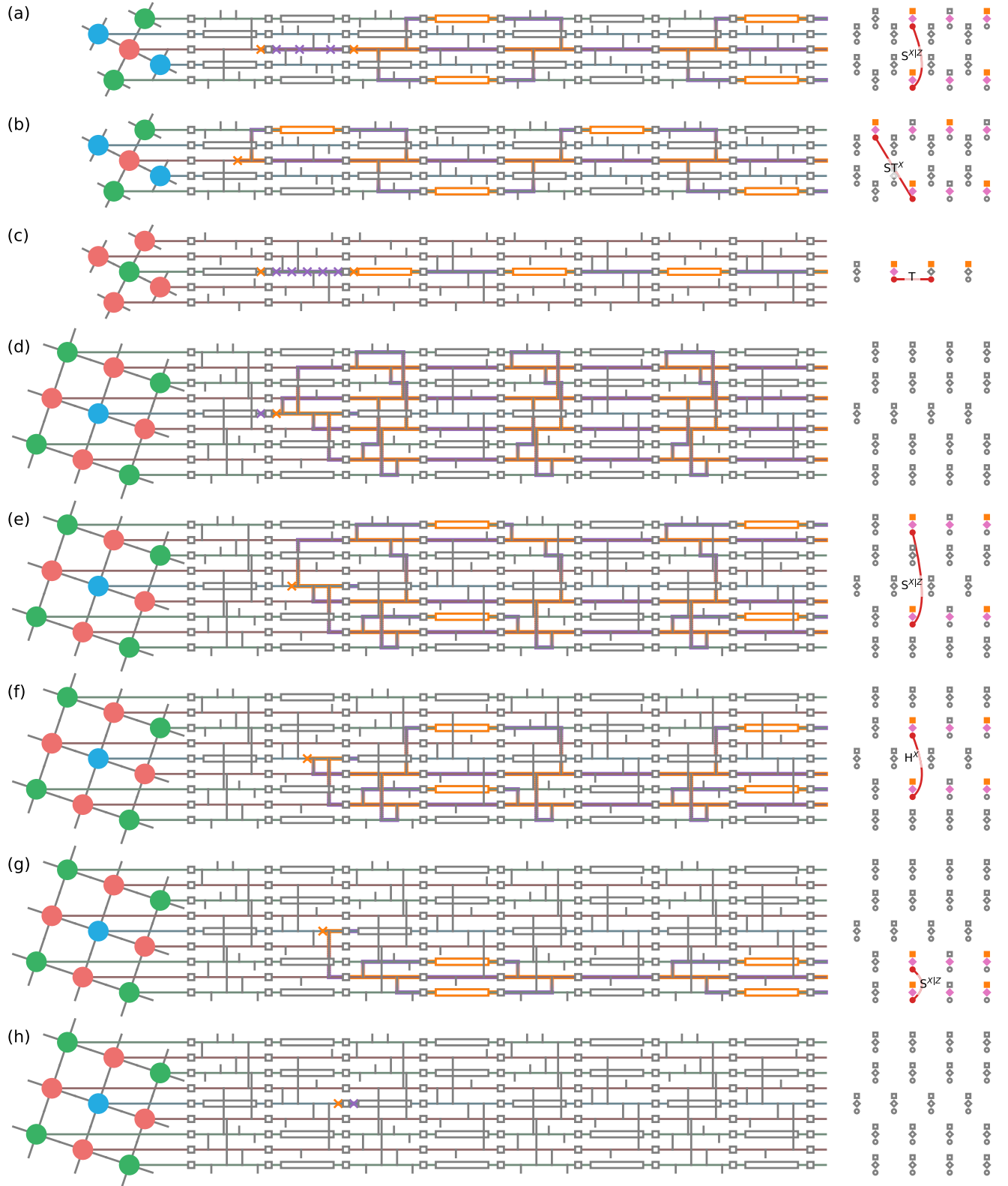


FIG. 6. Examples of syndrome generation due to physical  $\hat{X}$  (orange cross) and  $\hat{Z}$  (purple cross) errors at all possible circuit locations. Multiple crosses in the same panel indicate equivalent errors. The squares indicate Hadamard gates, vertical lines CZ gates, and rectangles readouts. Orange and purple lines indicate the propagation of  $\hat{X}$  and  $\hat{Z}$  flips through the circuit. At the right-hand side of each panel, the readout, stabilizer, and syndrome element flips are indicated in orange, pink, and purple, respectively. Errors on data qubits with signatures of class  $S^{X|Z}$  and  $ST^X$  are shown in panels (a) and (b), respectively. Errors on auxiliary qubits with  $T$ , empty,  $S^{X|Z}$ ,  $H^X$ ,  $S^{X|Z}$ , and empty signature are shown in panels (c)–(h), respectively.

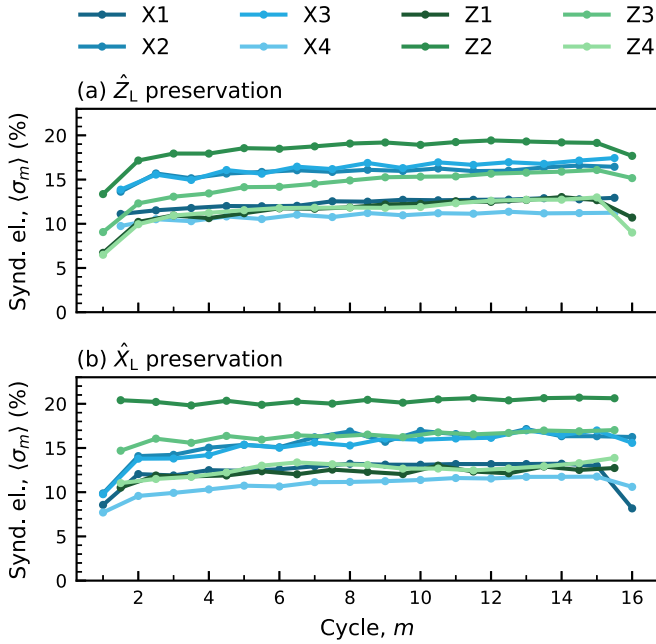


FIG. 7. Mean syndrome element ( $\sigma_m$ ) of each qubit as a function of the measurement cycle number  $m$  when preparing an eigenstate of the (a)  $\hat{Z}_L$  and (b)  $\hat{X}_L$  operators.

MWPM decoding described in the main text. We refer to the latter approach here as “standard” or uncorrelated MWPM. The correlated MWPM decoder is designed to better correct for  $\hat{Y}$  errors, which flip both X- and Z-type syndrome elements and thereby create correlations between the two syndrome types. The idea of harnessing the correlation of both syndrome types and performing a correlated decoding has been explored in complementary works such as Refs. [41–43,54,64–68]. Our strategy for making use of the correlations between syndrome types is to iterate between decoding the two syndrome types, updating the weights based on the results from the other decoding graph.

For syndrome matching, we can only utilize errors with signature weight of at most two. Including higher-order error signatures in the decoding graph would turn it into a hypergraph (where edges can connect more than two vertices), on which finding the minimum matching is no longer efficient. A  $\hat{Y}$  error can have a signature of weight two, three, or four when considering the joint syndrome, but not more than weight two on either syndrome type. Therefore, in the standard MWPM, the high-weight syndromes of  $\hat{Y}$  errors are split and a matching can be performed at the expense of treating single  $\hat{Y}$  errors as two uncorrelated  $\hat{X}$  and  $\hat{Z}$  errors. In the correlated MWPM, we first decode one syndrome graph in the standard way. Using the information about the decoded errors, we update the other auxiliary qubit graph with the conditional error probabilities, given the decoded errors on the first graph.

As an example, consider an error signature flipping three syndrome elements—two Z type and one X type. Let us assume that the X-type syndrome decoding is performed first with nonmodified weights. Then, the deduced correction will serve as a flag for the Z-type syndrome decoding to consider the conditional probability of having a  $\hat{Y}$  error given a  $\hat{Z}$  error

instead of the joint  $\hat{X}$  error probability. For the conditional probability, the leading-order contribution consists of taking the  $\hat{Y}$  error probability divided by the probability of finding the local X-type syndrome that was corrected in the previous Z-type syndrome decoding; see Eq. (C2).

Based on this procedure, we can alternate the decoding of both syndrome types while updating the weights until neither decoding outcome changes any more or until a predetermined maximum number of decoding rounds is performed. In practice, we only find minor changes to the decoded logical fidelity after the first iteration. We therefore restrict ourselves to one iteration of the weight update. In addition, to control the statistical uncertainty of the new weight, we do not fully replace the weight, but take a weighted average between the old and the new value instead. We find the highest logical fidelity for a weight of  $\gamma = 0.09$ , where  $\gamma = 0$  and  $\gamma = 1$  correspond to keeping the old weight and fully replacing the weight, respectively. This implies that for too large values of  $\gamma$ , the modification of the weights to include the possibility of  $\hat{Y}$  errors also breaks the correct decoding for some instances; see Appendix C3 for details. We suspect that statistical uncertainties of the inferred weights are at the origin of this problem.

Overall, we observe a slight improvement in decoding performance when using the correlated MWPM decoding protocol compared to the standard approach, although the difference is not statistically significant (see Fig. 8). We expect this method to deliver more substantial performance gains at larger code distances. In that regime, the decoder’s enhanced ability to accurately account for  $\hat{Y}$  errors allows for more informed decisions among a broader set of correction options, thereby potentially reducing logical error rates. This expectation is supported by simulations [54] and has been experimentally validated in Ref. [55] using data from Ref. [21].

In the Appendixes below, we provide further background on the principles of correlated MWPM using illustrative examples and elaborate on our implementation of the decoder.

### 1. Illustration of the probability update

To give an example of how probabilities can be updated during the correlated matching, we consider a simplified setting where symmetric depolarising errors are placed on data qubits with probabilities  $p^X = p^Z = p^Y = p/3$  before a perfect syndrome readout. Accordingly, all edges in both the X- and the Z-type decoding subgraphs are given a weight of  $w = -\ln 2p/3$ . Now, we assume that a physical  $\hat{Y}$  error occurred and that we perform the matching on the X-type syndrome subgraph first with the initial weights. The decoding yields the correct result of applying  $\hat{Z}$  as a correction. We can now reevaluate the probability of having had an  $\hat{X}$  error for the Z-type syndrome decoding problem. This probability is given by the conditional probability of having  $\hat{X}$  conditioned on having found  $\hat{Z}$ :

$$p(\hat{X}|\hat{Z}) = \frac{p(\hat{X} \wedge \hat{Z})}{p(\hat{Z})} = \frac{p^Y + \mathcal{O}(p^2)}{p^Y + p^Z + \mathcal{O}(p^2)} = \frac{1}{2} + \mathcal{O}(p). \quad (\text{C1})$$

Consequently, the probability of having an  $\hat{X}$  error at the specific location that was heralded by the decoding of the X-type syndrome increases the probability from  $2p/3$  to  $\approx 1/2$ . Here,



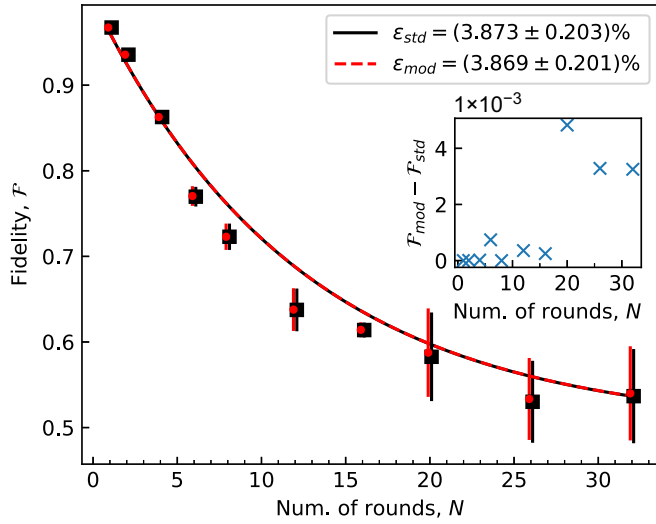


FIG. 8. Comparison of the MWPM decoding fidelity of the  $\hat{Z}_L$  basis with initial state  $|0\rangle_L$  using the standard weights (std, black) of syndrome weight up to two, not taking into account  $\hat{Y}$  errors, and the modified weight (mod, red) where the decoding is repeated once and weights are updated using conditional probabilities based on the inferred  $\hat{Y}$  error signatures. For better visibility, the error bars are scaled by a factor 10. The inset shows the differences in fidelity of the two MWPM decoding variants. The employed weights were calculated based on the same data used for calculating the shown fidelities.

the value of  $1/2$  is the coin flip probability between having found a genuine  $\hat{Z}$  error or a  $\hat{Y}$  error that in turn would carry the  $\hat{Z}$  error in addition to  $\hat{X}$ . As a next step, the decoding of the Z-type syndrome can be performed on a graph where all edges but one carry the weight  $w = -\ln 2p/3$  and the edge that corresponds to an  $\hat{X}$  correction has a weight of  $w = -\ln 1/2$ , due to the adapted probability.

## 2. Correlated decoding under circuit-level noise

For circuit-level noise and inferred error probabilities, we update a dressed probability by replacing it with the conditional probability for having an effective  $\hat{X}$  ( $\hat{Z}$ ) error. This update is conditioned on finding the complementary effective  $\hat{Z}$  ( $\hat{X}$ ) error from the previous decoding outcome. The concrete update signature and conditional probabilities depend on the circuitry and how  $\hat{Y}$  errors propagate. To illustrate this with an example, we assume circuit-level  $\hat{Y}$  errors with probability  $p_{(\sigma_m^{Z,A_i}, \sigma_{m'}^{Z,A_j}, \sigma_{m''}^{X,A_k})}^Y$  that trigger the Z-type stabilizers ( $A_i, A_j$ ) at times ( $m, m'$ ) and the X-type stabilizer  $A_k$  at time  $m''$ . These elementary errors are additional to genuine  $\hat{X}$  and  $\hat{Z}$  errors that have a signature one syndrome type. For simplicity, we are going to use the notation  $\alpha = (A_i, m)$  to indicate the space-time coordinate of a syndrome element. We denominate the corresponding elementary error as  $\hat{Y}_v$ , which causes the syndrome signature  $v = (\sigma_\alpha^Z, \sigma_\beta^Z, \sigma_\gamma^X)$  and occurs by probability  $p_{(\sigma_\alpha^Z, \sigma_\beta^Z, \sigma_\gamma^X)}^Y$ . We assume now a setting where we perform an initial uncorrelated MWPM decoding on the X-type syndrome graph by which the matching suggests that a  $\hat{Z}_v$  error has happened. Such interpretation between matching and error suggestion relies on the notion of elementary errors.

For the latter, we assume that an edge that is included in the matching attributes a most likely error on the data qubits that can be taken as the correction. Considering the concrete matching example, the flip of the single syndrome element  $\sigma_\gamma^X$  is matched to a boundary, which corresponds to a  $\hat{Z}$  error at a location specified by  $v$ . Having obtained this information, we would like to properly adapt the error probability  $\tilde{p}_{(\sigma_\alpha^Z, \sigma_\beta^Z)}^X$  of an  $\hat{X}$ -type likeliest error that is attributed to the edge  $(\alpha, \beta)$ . The weight of this edge would be updated for a round of correlated bit flip decoding. This is done by replacing  $\tilde{p}_{(\sigma_\alpha^Z, \sigma_\beta^Z)}^X$  with the conditional probability of finding an  $\hat{X}_v$  error given a  $\hat{Z}_v$  error is present (known from decoding):

$$p(\hat{X}_v | \hat{Z}_v) = \frac{p(\hat{X}_v \wedge \hat{Z}_v)}{p(\hat{Z}_v)} \approx \frac{p_{(\sigma_\alpha^Z, \sigma_\beta^Z, \sigma_\gamma^X)}^Y}{\sum_{\alpha', \beta'} p_{(\sigma_{\alpha'}^Z, \sigma_{\beta'}^Z, \sigma_\gamma^X)}^Y + \tilde{p}_{\sigma_\gamma^X}^Z}. \quad (\text{C2})$$

It is important to stress that this update only takes place if and only if the matching correction, here  $\hat{Z}_v$ , is suggested by the complementary decoder. To explain the conditional probability of Eq. (C2), first we note that it is a leading-order approximation in terms of the error rate. The numerator is given by the probability of having both an  $\hat{X}$  and a  $\hat{Z}$  error at the appropriate space-time coordinate. To leading order, this corresponds to the inferred  $\hat{Y}$  error probability. In the next order, terms like  $p^X \cdot p^Z$  would contribute as well but are omitted in this work. In the denominator of Eq. (C2), we find the leading-order probability of exhibiting a  $\hat{Z}_{\alpha_i}$  error, which can emerge due to a proper  $\hat{Z}$  error with probability  $\tilde{p}_{\sigma_\gamma^X}^Z$  or by means of a  $\hat{Y}$  error. With regard to the latter, there might exist a set of  $\hat{Y}$  errors that are degenerate with respect to the signature on the X-type syndrome. That is, there are multiple  $\hat{Y}$  errors that cause the same X-type syndrome but different Z-type syndromes. Over the corresponding probabilities, we perform a summation  $\sum_{\alpha', \beta'}$ . The adaption of the weight in this manner can be done for the Z-type syndrome decoding as well given a previous X-type syndrome decoding serving as the flag for it to be applied. The weight update in this way can also be applied for  $\hat{Y}$  error propagations that have weight-two or weight-four signatures on the joint syndrome. The correct weight update depends on these propagations that in turn are dictated by the stabilizer readout circuitry.

## 3. Application to the experimental data and interpolation approach

We perform weight updates that are compatible with the syndrome signatures of all possible  $\hat{Y}$  error propagations based on the circuitry in Appendix H. Note that the syndrome according to a  $\hat{Y}$  error propagation needs not to be contained in one measurement cycle but can also span between two consecutive rounds of stabilizer readout depending on the gate order. As mentioned in the main text, given an update rule, the decoding of both subgraphs can be repeated until no change in the decoding is observed. Practically, it is sufficient to only repeat the decoding once, taking the modified weights for the experimental data analyzed in this work. The conditional probability, used for modifying the error probabilities, comes with a statistical error as it is a function of quantities inferred from the experimentally observed syndrome statistics. This

uncertainty is genuinely larger than the uncertainty of the unaltered probability. We observe that this can have a deteriorating effect on the decoding result. Therefore, we relax the update of the probability by interpolating between the old and the new error probability. For the previous example update, that means to replace the standard weight by

$$w'_{(\sigma_\alpha^Z, \sigma_\beta^Z)} = -(1 - \gamma) \ln(\tilde{p}_{(\sigma_\alpha^Z, \sigma_\beta^Z)}^X) - \gamma \ln\left(\frac{p_{(\sigma_\alpha^Z, \sigma_\beta^Z, \sigma_\gamma^X)}^Y}{\sum_{\alpha', \beta'} p_{(\sigma_{\alpha'}^Z, \sigma_{\beta'}^Z, \sigma_\gamma^X)}^Y + \tilde{p}_{\sigma_\gamma^X}^Z}\right), \quad (\text{C3})$$

where  $\gamma \in [0, 1]$  is an interpolation parameter that we set to 0.09 for the decoding results shown in Fig. 8. We observe that for larger values of  $\gamma$ , the fidelity is not constantly improved for each number of measurement rounds; see Fig. 9(a). Already for values slightly above  $\gamma = 0.09$ , the decoding performance worsens for small and larger number of rounds. For only few syndrome readouts, this decreased fidelity persists also for larger values of  $\gamma$ . As this maximal  $\gamma$  value for obtaining a good performance in decoding is relatively low, we conjecture that at this level of statistical fluctuations of the weights, the influence of these fluctuations outweighs the improvement of including  $\hat{Y}$  error information such that minimizing the uncertainty is prioritized to improve the decoding.

#### 4. Influence of statistical fluctuations

To investigate the effect of the statistical error on the decoding performance qualitatively, we re-run the modified and unmodified decoding on a simulated dataset with larger size and therefore lower statistical fluctuations compared with the experimental data. We perform simulations of the distance-three surface code with identical circuitry under depolarizing circuit-level noise. For this model, each circuit element exhibits a depolarizing error with probability  $p$  after the execution of the circuit element. Of course, such simplistic one-parameter noise model is not intended to fully capture the actual experiment of Ref. [4] but shall only serve as a tool to understand the effect of statistical fluctuations of the weights on the decoding performance. We perform such simulation for a fixed error rate of  $p = 10^{-3}$  and a sample size that has about 12 times more syndrome readouts as compared to the experimental data. These readouts distribute into  $10^6$  shots per bin of number of readout rounds. Note that in the experimental data, especially for a larger number of readout rounds, the datasets are much smaller.

We expect a lowered statistical uncertainty for this numerically generated dataset compared with the experimental data, and we infer weights from the sampled data and perform both the uncorrelated and the modified MWPM. We indeed find that now much larger values of  $\gamma$  correspond to a consistent performance increase. We find that this improvement of the modified over the uncorrelated MWPM decoder only breaks down for values as large as  $\gamma > 0.8$ , see Fig. 9(b), instead of  $\gamma > 0.09$  as for the experimental data. Another intriguing difference for the behavior of the relative improvement comparing experimental and simulated data is the smoothness that can be observed in the latter.

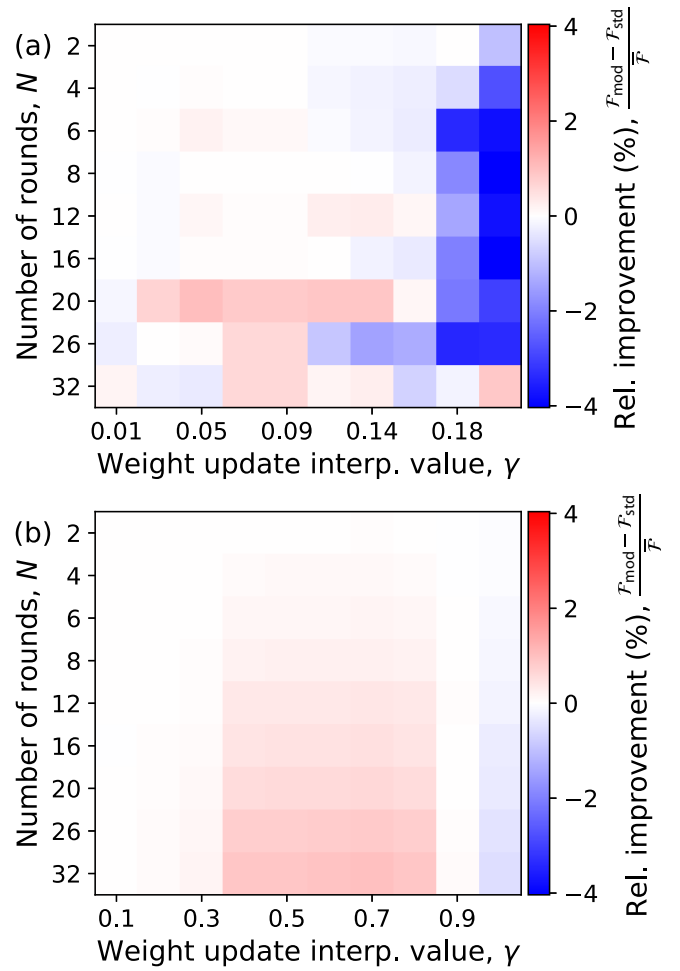


FIG. 9. The relative improvement in fidelity of repeating the MWPM once after updating the weights as a function of the number of error correction cycles  $N$  and the update strength  $\gamma$  for (a) the experimental data and (b) simulated data under depolarizing circuit-level noise. The relative improvement is defined as  $\frac{\mathcal{F}_{\text{mod}} - \mathcal{F}_{\text{std}}}{\bar{\mathcal{F}}}$ , where  $\bar{\mathcal{F}} = \frac{\mathcal{F}_{\text{mod}} + \mathcal{F}_{\text{std}}}{2}$  is the average fidelity.

To investigate this further, we also simulate the surface-17 code under a heterogeneous error model where the error rates for two-qubit gates and single-qubit gates are drawn from normal distributions  $\mathcal{N}(1.5\%, \Delta \cdot 1\%)$  and  $\mathcal{N}(0.09\%, \Delta \cdot 0.04\%)$ , respectively. The variable  $\Delta$  scales the width of the normal distribution from which the error rates are drawn. The numerical values mimic now for  $\Delta = 1$  the noise level in the experiment (compare to Ref. [4]). Note that the error rate for each gate is only drawn once and then kept fixed for all gate executions. The underlying idea for simulating a heterogeneous error model is that the resulting inferred weights will break degeneracies in the matching problem that would only be broken in correlated matching if one considered a homogenous error model where all error rates are identical. Following this line of reasoning, a more realistic and heterogeneous error model might reduce the improvement of the correlated matching. The parameter  $\Delta$  can be used to control the degree of heterogeneity of the error model. In Fig. 10, the relative improvement is plotted for a fixed number of 16 stabilizer measurements by varied  $\Delta$  and

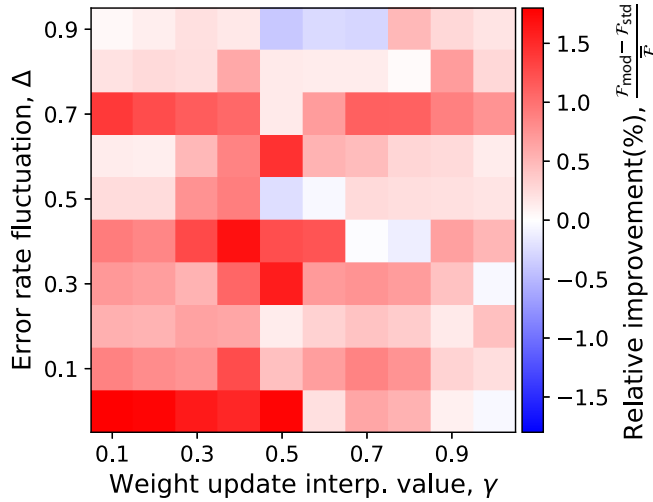


FIG. 10. The relative improvement in fidelity of correlated MWPM compared to uncorrelated MWPM for simulated data with a heterogeneous error model. The scale of fluctuations of gate error rates  $\Delta$  is shown on the vertical axis, while the strength of the weight update interpolation  $\gamma$  is shown on the horizontal axis.

interpolation strength  $\gamma$ . The dataset size of this simulation is chosen as for the homogeneous error model simulation. Each value of  $\Delta$  corresponds to a new set of simulated data. It is observable that for smaller  $\Delta$ , stronger updates should be performed. Overall, we recover the nonsmooth behavior of the improvement metric, which we also found for the experimental data in Fig. 9(a) as a result of increased statistical fluctuations.

#### APPENDIX D: QUANTUM DEVICE

The quantum processor is fabricated from a 150-nm-thin niobium film on a silicon substrate, from which the transmon islands, resonators, and control lines are patterned using optical lithography. To realize coplanar waveguide crossovers and to connect different parts of the ground plane, we fabricate airbridges from an aluminum-titanium-aluminum trilayer. The Josephson junctions are fabricated from aluminum using electron beam lithography and shadow evaporation. Each of the 17 qubits (yellow) has an individual microwave drive line (pink) and flux line (green); see Fig. 11. The qubits are capacitively coupled via a coplanar waveguides (cyan) to achieve an average qubit-qubit coupling rate of  $J/2\pi \approx 6$  MHz and a mean interaction time for a dynamic flux pulse-based CZ gate [33] of 68 ns. For dispersive readout, we couple the qubits to readout resonators (red), which are coupled to four frequency-multiplexed feedlines (purple) via individual Purcell filters (blue) to suppress qubit decay and readout crosstalk [69]. The qubit and resonator frequencies, anharmonicities, and coherence properties are given in Table II. The readout resonator frequencies (6.769–7.554 GHz) are above the auxiliary qubits, which are biased to their upper flux-insensitive frequencies (5.885–6.192 GHz), and the data qubits biased to their lower flux-insensitive frequencies (3.740–4.143 GHz). The exception to the above is the auxiliary qubit X1, which is biased to its lower flux-insensitive frequency to avoid interactions

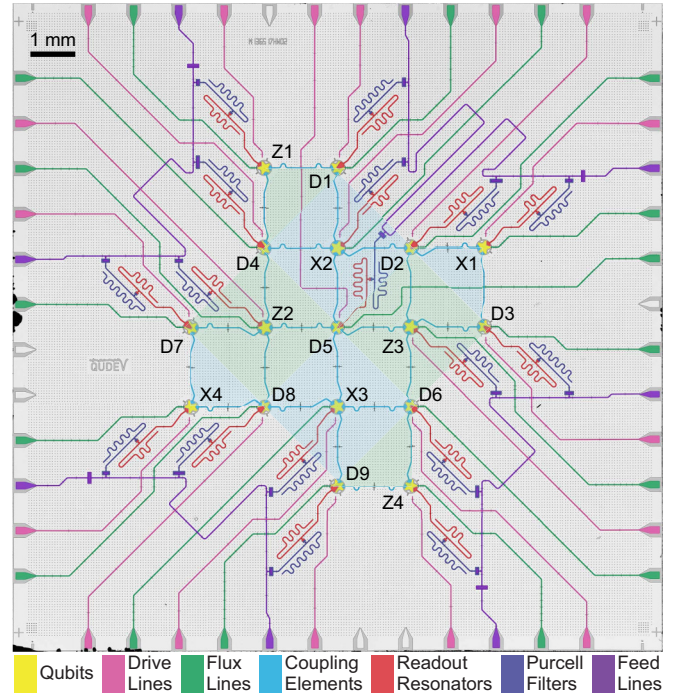


FIG. 11. False-color optical micrograph of the 17-qubit quantum processor used in this work. See text for details about the device. Figure adapted from Ref. [4].

with a strongly coupled defect near its nominal bias frequency. We find a mean single-qubit gate error of  $0.09\% \pm 0.04\%$  in randomized benchmarking [70], and a mean CZ gate error of  $1.5\% \pm 1.0\%$  in interleaved randomized benchmarking [71,72].

#### APPENDIX E: DERIVATION OF THE GENERAL ERROR PROBABILITY FORMULA

In this Appendix, we derive Eq. (10) for calculating the probability  $p_{i_1, \dots, i_n}$  that an error that flips  $n$  syndrome elements  $i_1$  to  $i_n$  occurs, given the correlations between the observed syndrome elements  $\sigma_i$ .

We start by introducing the notation for the random variables involved in the problem. Let  $P_{i_1, \dots, i_n}$  be the random variable that indicates whether an independent error that flips the syndrome elements  $\{i_1, \dots, i_n\}$  occurred ( $P_{i_1, \dots, i_n} = -1$ ) or not ( $P_{i_1, \dots, i_n} = +1$ ). Note that because the random variables take values  $\pm 1$ , its square is always one:  $P_{i_1, \dots, i_n}^2 = 1$ . The probability we are looking for is then expressed as the expectation value  $p_{i_1, \dots, i_n} = (1 - \langle P_{i_1, \dots, i_n} \rangle)/2$ .

To gain some intuition about this variable, let us write out the expressions for all possible syndrome element correlations in terms of these random variables in the case of a small number of total syndrome elements. Again, we use the notation  $\tilde{\sigma}_i = 1 - 2\sigma_i = \pm 1$  for the syndrome element at node  $i$ , for brevity, and consider the index  $i$  to include both the qubit index and the detection time of the syndrome element. We have the following sets of equations for one syndrome element:

$$\tilde{\sigma}_1 = P_1, \quad \langle \tilde{\sigma}_1 \rangle = \langle P_1 \rangle,$$

TABLE II. Qubit parameters, coherence properties, and single-qubit performance for the nine data qubits (top) and the eight auxiliary qubits (bottom). We also provide, for relevant quantities, the averaged value across the device in column  $\bar{Q}$ . RO stands for readout and RB stands for randomized benchmarking.

Parameter	D1	D2	D3	D4	D5	D6	D7	D8	D9
Qubit idle frequency, $\omega_Q/2\pi$ (GHz)	3.885	3.994	3.952	3.878	3.895	3.740	4.056	3.993	4.143
Qubit anharmonicity, $\alpha/2\pi$ (MHz)	-184	-183	-183	-184	-186	-184	-181	-183	-181
Lifetime, $T_1$ ( $\mu$ s)	31.1	29.0	69.9	55.5	32.7	59.1	33.2	25.8	30.3
Ramsey decay time, $T_2^*$ ( $\mu$ s)	36.9	14.3	36.3	87.5	45.7	16.8	47.6	37.8	46.5
Echo decay time, $T_2^e$ ( $\mu$ s)	47.2	48.6	46.5	87.8	54.2	24.1	49.9	45.9	51.7
Readout frequency, $\omega_{RO}/2\pi$ (GHz)	6.769	6.979	6.880	7.120	7.180	7.032	6.910	7.075	6.868
Qb. freq. during RO, $\omega'_Q/2\pi$ (GHz)	5.321	4.750	5.275	4.250	4.420	5.130	4.395	3.993	5.000
Qubit-RO res. coupling, $g_{Q,RR}/2\pi$ (MHz)	244	269	241	241	238	244	267	265	260
Single-qubit RB error, $\epsilon_{1Q}$ (%)	0.06	0.07	0.04	0.04	0.06	0.06	0.04	0.08	0.06
Two-state readout error, $\epsilon_{RO}^{(2)}$ (%)	0.7	0.6	0.5	2.7	1.9	1.0	2.0	0.8	0.4
Three-state readout error, $\epsilon_{RO}^{(3)}$ (%)	6.2	1.7	8.0	5.4	3.0	2.4	3.4	4.1	1.1
Parameter	X1	X2	X3	X4	Z1	Z2	Z3	Z4	$\bar{Q}$
Qubit idle frequency, $\omega_Q/2\pi$ (GHz)	4.429	5.885	6.022	6.049	6.328	6.192	5.956	6.037	4.849
Qubit anharmonicity, $\alpha/2\pi$ (MHz)	-181	-174	-170	-170	-163	-168	-171	-170	-177
Lifetime, $T_1$ ( $\mu$ s)	17.8	15.3	18.6	16.3	21.3	45.4	29.1	19.2	32.3
Ramsey decay time, $T_2^*$ ( $\mu$ s)	21.6	20.3	21.9	27.8	37.8	34.2	49.8	25.8	35.8
Echo decay time, $T_2^e$ ( $\mu$ s)	30.1	30.3	15.6	31.2	38.7	27.6	52.7	36.2	42.3
Readout frequency, $\omega_{RO}/2\pi$ (GHz)	7.372	7.554	7.258	7.461	7.316	7.502	7.200	7.412	7.170
Qb. freq. during RO, $\omega'_Q/2\pi$ (GHz)	5.900	5.885	6.022	6.049	6.328	6.192	5.956	6.037	5.347
Qubit-RO res. coupling, $g_{Q-RR}/2\pi$ (MHz)	167	168	167	168	171	170	167	171	213
Single-qubit RB error, $\epsilon_{1Q}$ (%)	0.16	0.13	0.17	0.14	0.10	0.09	0.07	0.16	0.09
Two-state readout error, $\epsilon_{RO}^{(2)}$ (%)	0.8	1.2	0.9	0.6	0.9	0.5	0.5	0.6	1.0
Three-state readout error, $\epsilon_{RO}^{(3)}$ (%)	1.3	3.2	3.5	1.5	2.1	1.6	5.2	1.3	3.2

two syndrome elements:

$$\begin{aligned}\tilde{\sigma}_1 &= P_1 P_{12}, & \langle \tilde{\sigma}_1 \rangle &= \langle P_1 \rangle \langle P_{12} \rangle, \\ \tilde{\sigma}_2 &= P_2 P_{12}, & \langle \tilde{\sigma}_2 \rangle &= \langle P_2 \rangle \langle P_{12} \rangle, \\ \tilde{\sigma}_1 \tilde{\sigma}_2 &= P_1 P_2 P_{12}^2 = P_1 P_2, & \langle \tilde{\sigma}_1 \tilde{\sigma}_2 \rangle &= \langle P_1 \rangle \langle P_2 \rangle,\end{aligned}$$

and three syndrome elements:

$$\tilde{\sigma}_1 = P_1 P_{12} P_{13} P_{123}, \quad \langle \tilde{\sigma}_1 \rangle = \langle P_1 \rangle \langle P_{12} \rangle \langle P_{13} \rangle \langle P_{123} \rangle, \quad (\text{E1a})$$

$$\tilde{\sigma}_2 = P_2 P_{12} P_{23} P_{123}, \quad \langle \tilde{\sigma}_2 \rangle = \langle P_2 \rangle \langle P_{12} \rangle \langle P_{23} \rangle \langle P_{123} \rangle, \quad (\text{E1b})$$

$$\tilde{\sigma}_3 = P_3 P_{13} P_{23} P_{123}, \quad \langle \tilde{\sigma}_3 \rangle = \langle P_3 \rangle \langle P_{13} \rangle \langle P_{23} \rangle \langle P_{123} \rangle, \quad (\text{E1c})$$

$$\tilde{\sigma}_1 \tilde{\sigma}_2 = P_1 P_2 P_{12}^2 P_{13} P_{23} P_{123}^2 = P_1 P_2 P_{13} P_{23}, \quad \langle \tilde{\sigma}_1 \tilde{\sigma}_2 \rangle = \langle P_1 \rangle \langle P_2 \rangle \langle P_{13} \rangle \langle P_{23} \rangle, \quad (\text{E1d})$$

$$\tilde{\sigma}_1 \tilde{\sigma}_3 = P_1 P_3 P_{12} P_{13}^2 P_{23} P_{123}^2 = P_1 P_3 P_{12} P_{23}, \quad \langle \tilde{\sigma}_1 \tilde{\sigma}_3 \rangle = \langle P_1 \rangle \langle P_3 \rangle \langle P_{12} \rangle \langle P_{23} \rangle, \quad (\text{E1e})$$

$$\tilde{\sigma}_2 \tilde{\sigma}_3 = P_2 P_3 P_{12} P_{13} P_{23}^2 P_{123}^2 = P_2 P_3 P_{12} P_{13}, \quad \langle \tilde{\sigma}_2 \tilde{\sigma}_3 \rangle = \langle P_2 \rangle \langle P_3 \rangle \langle P_{12} \rangle \langle P_{13} \rangle, \quad (\text{E1f})$$

$$\tilde{\sigma}_1 \tilde{\sigma}_2 \tilde{\sigma}_3 = P_1 P_2 P_3 P_{12}^2 P_{13}^2 P_{23}^2 P_{123}^3 = P_1 P_2 P_3 P_{123}, \quad \langle \tilde{\sigma}_1 \tilde{\sigma}_2 \tilde{\sigma}_3 \rangle = \langle P_1 \rangle \langle P_2 \rangle \langle P_3 \rangle \langle P_{123} \rangle. \quad (\text{E1g})$$

We observe that the expression for correlation  $\langle \tilde{\sigma}_{j_1} \cdots \tilde{\sigma}_{j_m} \rangle$  is given by the product of the expectation values  $\langle P_{k_1, \dots, k_q} \rangle$  that have an overlap in the indices at an odd number of places; i.e.,  $|\{j_1, \dots, j_m\} \cap \{k_1, \dots, k_q\}|$  is odd. This stems from the fact

that processes that flip an even number of syndrome elements in  $\tilde{\sigma}_{j_1} \cdots \tilde{\sigma}_{j_m}$  do not change the value of the correlation.

Next, we consider the product of the expectation values of all possible odd correlations  $A$ , and of all possible even



correlations  $B$ :

$$A = \prod_{\substack{\{j_1, \dots, j_m\} \subseteq \{1, \dots, r\} \\ m \text{ is odd}}} \langle \tilde{\sigma}_{j_1} \cdots \tilde{\sigma}_{j_m} \rangle,$$

$$B = \prod_{\substack{\{j_1, \dots, j_m\} \subseteq \{1, \dots, r\} \\ m \text{ is even}}} \langle \tilde{\sigma}_{j_1} \cdots \tilde{\sigma}_{j_m} \rangle,$$

where  $r$  is the total number of syndrome elements in the system. As an explicit example, for  $r = 3$ , we get

$$A = \langle \tilde{\sigma}_1 \rangle \langle \tilde{\sigma}_2 \rangle \langle \tilde{\sigma}_3 \rangle \langle \tilde{\sigma}_1 \tilde{\sigma}_2 \tilde{\sigma}_3 \rangle,$$

$$B = \langle \tilde{\sigma}_1 \tilde{\sigma}_2 \rangle \langle \tilde{\sigma}_1 \tilde{\sigma}_3 \rangle \langle \tilde{\sigma}_2 \tilde{\sigma}_3 \rangle.$$

We will analyze how many times an expectation value  $\langle P_{k_1, \dots, k_q} \rangle$  is present in either of the products  $A$  and  $B$ .

First, if  $\{k_1, \dots, k_q\}$  is not the full set of indices in the system  $\{1, \dots, r\}$ , then there exists a syndrome element  $k' \notin \{k_1, \dots, k_q\}$ . We can then, for each correlation factor  $\langle \tilde{\sigma}_{j_1} \cdots \tilde{\sigma}_{j_m} \rangle$  in  $A$ , find a partner in  $B$ , such that it only differs by the presence of the syndrome element  $k'$ . These two factors will both either include or not include  $\langle P_{k_1, \dots, k_q} \rangle$ , as the addition or removal of the syndrome element  $k'$  will not influence the parity of the overlap  $|\{j_1, \dots, j_m\} \cap \{k_1, \dots, k_q\}|$ . In the  $r = 3$  example above, we find that the term  $\langle P_{12} \rangle$  has the same power in both  $A$  and  $B$ , since we can pair the factors such that they only differ by the presence of  $\tilde{\sigma}_3$ . From Eq. (E1), we see that both  $\langle \tilde{\sigma}_1 \rangle$  from  $A$  and  $\langle \tilde{\sigma}_1 \tilde{\sigma}_3 \rangle$  from  $B$  have the factor  $\langle P_{12} \rangle$ . Similarly, both  $\langle \tilde{\sigma}_2 \rangle$  and  $\langle \tilde{\sigma}_2 \tilde{\sigma}_3 \rangle$  have  $\langle P_{12} \rangle$ , and neither  $\langle \tilde{\sigma}_1 \tilde{\sigma}_2 \tilde{\sigma}_3 \rangle$  nor  $\langle \tilde{\sigma}_1 \tilde{\sigma}_2 \rangle$  have  $\langle P_{12} \rangle$ .  $\langle \tilde{\sigma}_3 \rangle$  from  $A$  would pair with the expectation value of no syndrome elements in  $B$ , neither of which has that factor  $\langle P_{12} \rangle$ .

Second, in the case of  $\{k_1, \dots, k_q\} = \{1, \dots, r\}$ , the parity of the overlap is uniquely determined by the parity of the correlation factor, which is odd for all factors in  $A$  and even for all factors in  $B$ . Therefore,  $B$  will not have the factor  $\langle P_{1, \dots, r} \rangle$ , and  $A$  will have it as many times as many different ways we can choose an odd number of indices from a set containing  $r$  indices. This can be calculated using the binomial theorem as

$$\sum_{\substack{m \in \{1, \dots, r\} \\ m \text{ is odd}}} \binom{r}{m} = \sum_{m=0}^r \binom{r}{m} 1^{r-m} \frac{1^m - (-1)^m}{2}$$

$$= ((1+1)^r - (1-1)^r)/2 = 2^{r-1}.$$

In the  $r = 3$  example above, each of the  $2^{r-1} = 4$  factors of  $A$  contains the factor  $\langle P_{123} \rangle$ , while none of the factors of  $B$  do; see again Eq. (E1).

Combining the results that  $A$  and  $B$  have the same number of factors  $\langle P_{k_1, \dots, k_q} \rangle$  if  $\{k_1, \dots, k_q\} \neq \{1, \dots, r\}$  and that only  $A$  has  $\langle P_{1, \dots, r} \rangle$  with a power of  $2^{r-1}$ , we get the following result:

$$\langle P_{1, \dots, r} \rangle^{2^{r-1}} = \frac{A}{B} = \frac{\prod_{\substack{\{j_1, \dots, j_m\} \subseteq \{1, \dots, r\} \\ m \text{ is odd}}} \langle \tilde{\sigma}_{j_1} \cdots \tilde{\sigma}_{j_m} \rangle}{\prod_{\substack{\{j_1, \dots, j_m\} \subseteq \{1, \dots, r\} \\ m \text{ is even}}} \langle \tilde{\sigma}_{j_1} \cdots \tilde{\sigma}_{j_m} \rangle}$$

$$= \prod_{\{j_1, \dots, j_m\} \subseteq \{1, \dots, r\}} \langle \tilde{\sigma}_{j_1} \cdots \tilde{\sigma}_{j_m} \rangle^{(-1)^{m-1}}.$$

To find an expression for the expectation value for  $P_{i_1, \dots, i_n}$  in the case where  $\{i_1, \dots, i_n\}$  is not the entire set of syndrome elements in the system, we consider the combined processes  $P'_{k_1, \dots, k_q}$  with  $\{k_1, \dots, k_q\} \subseteq \{i_1, \dots, i_n\}$  that have the same effect on syndrome elements  $\{i_1, \dots, i_n\}$  as  $P_{k_1, \dots, k_q}$ :

$$P'_{k_1, \dots, k_q} = \prod_{\{j_1, \dots, j_m\} \subseteq \{1, \dots, r\} \setminus \{i_1, \dots, i_n\}} P_{k_1, \dots, k_q, j_1, \dots, j_m}.$$

As the effects due to processes on the additional syndrome elements have been absorbed into the combined processes, the previous derivation of decomposing  $P_{1, \dots, r}$  as a product of even and odd syndrome moments holds the same for decomposing  $P'_{i_1, \dots, i_n}$  as even and odd syndrome moments from the set  $\{i_1, \dots, i_n\}$ . We therefore get the result

$$\langle P'_{i_1, \dots, i_n} \rangle^{2^{n-1}} = \prod_{\{j_1, \dots, j_m\} \subseteq \{i_1, \dots, i_n\}} \langle \tilde{\sigma}_{j_1} \cdots \tilde{\sigma}_{j_m} \rangle^{(-1)^{m-1}}.$$

Writing out the definition of the combined process that flips all the syndrome elements  $\{i_1, \dots, i_n\}$

$$P'_{i_1, \dots, i_n} = \prod_{\{j_1, \dots, j_m\} \subseteq \{1, \dots, r\} \setminus \{i_1, \dots, i_n\}} P_{i_1, \dots, i_n, j_1, \dots, j_m}$$

$$= P_{i_1, \dots, i_n} \prod_{\{j_1, \dots, j_m\} \supset \{i_1, \dots, i_n\}} P_{j_1, \dots, j_m},$$

we have derived the full form of Eq. (10):

$$\langle P_{i_1, \dots, i_n} \rangle = \frac{\prod_{\substack{\{j_1, \dots, j_m\} \\ \subseteq \{i_1, \dots, i_n\}}} \langle \tilde{\sigma}_{j_1} \cdots \tilde{\sigma}_{j_m} \rangle^{(-1)^{m-1} 2^{-(n-1)}}}{\prod_{\substack{\{j_1, \dots, j_m\} \\ \supset \{i_1, \dots, i_n\}}} \langle P_{j_1, \dots, j_m} \rangle},$$

$$P_{i_1, \dots, i_n} = \frac{1}{2} - \frac{1}{2} \frac{\prod_{\substack{\{j_1, \dots, j_m\} \\ \subseteq \{i_1, \dots, i_n\}}} \langle \tilde{\sigma}_{j_1} \cdots \tilde{\sigma}_{j_m} \rangle^{(-1)^{m-1} 2^{-(n-1)}}}{\prod_{\substack{\{j_1, \dots, j_m\} \\ \supset \{i_1, \dots, i_n\}}} (1 - 2p_{j_1, \dots, j_m})}.$$

For reference to the reader, we will write out the explicit formulas assuming that there are no errors with signatures on more than two syndrome elements, recovering the formulas derived in Ref. [23]:

$$p_{ij} = \frac{1}{2} - \frac{1}{2} \sqrt{\frac{\langle \tilde{\sigma}_i \rangle \langle \tilde{\sigma}_j \rangle}{\langle \tilde{\sigma}_i \tilde{\sigma}_j \rangle}} = \frac{1}{2} - \sqrt{\frac{1}{4} + \frac{\langle \sigma_i \rangle \langle \sigma_j \rangle - \langle \sigma_i \sigma_j \rangle}{1 - 2\langle \sigma_i \oplus \sigma_j \rangle}}, \quad (\text{E2a})$$

$$p_i = \frac{1}{2} - \frac{1}{2} \frac{\langle \tilde{\sigma}_i \rangle}{\prod_{j \neq i} (1 - 2p_{ij})} = \frac{1}{2} - \frac{1/2 - \langle \sigma_i \rangle}{\prod_{j \neq i} (1 - 2p_{ij})}, \quad (\text{E2b})$$

with  $\oplus$  representing addition modulo 2.

Similarly, assuming no errors with signatures on more than four syndrome elements, we get the explicit formulas

$$p_{ijkl} = \frac{1}{2} - \frac{1}{2} \sqrt{\frac{\langle \tilde{\sigma}_i \rangle \langle \tilde{\sigma}_j \rangle \langle \tilde{\sigma}_k \rangle \langle \tilde{\sigma}_l \rangle \langle \tilde{\sigma}_i \tilde{\sigma}_j \tilde{\sigma}_k \rangle \langle \tilde{\sigma}_i \tilde{\sigma}_j \tilde{\sigma}_l \rangle \langle \tilde{\sigma}_i \tilde{\sigma}_k \tilde{\sigma}_l \rangle \langle \tilde{\sigma}_j \tilde{\sigma}_k \tilde{\sigma}_l \rangle}{\langle \tilde{\sigma}_i \tilde{\sigma}_j \rangle \langle \tilde{\sigma}_i \tilde{\sigma}_k \rangle \langle \tilde{\sigma}_i \tilde{\sigma}_l \rangle \langle \tilde{\sigma}_j \tilde{\sigma}_k \rangle \langle \tilde{\sigma}_j \tilde{\sigma}_l \rangle \langle \tilde{\sigma}_k \tilde{\sigma}_l \rangle \langle \tilde{\sigma}_i \tilde{\sigma}_j \tilde{\sigma}_k \tilde{\sigma}_l \rangle}}, \quad (\text{E3a})$$

$$p_{ijk} = \frac{1}{2} - \frac{1}{2} \sqrt{\frac{\langle \tilde{\sigma}_i \rangle \langle \tilde{\sigma}_j \rangle \langle \tilde{\sigma}_k \rangle \langle \tilde{\sigma}_i \tilde{\sigma}_j \tilde{\sigma}_k \rangle}{\langle \tilde{\sigma}_i \tilde{\sigma}_j \rangle \langle \tilde{\sigma}_i \tilde{\sigma}_k \rangle \langle \tilde{\sigma}_j \tilde{\sigma}_k \rangle}} \prod_{l \notin \{i,j,k\}} \frac{1}{1 - 2p_{ijkl}}, \quad (\text{E3b})$$

$$p_{ij} = \frac{1}{2} - \frac{1}{2} \sqrt{\frac{\langle \tilde{\sigma}_i \rangle \langle \tilde{\sigma}_j \rangle}{\langle \tilde{\sigma}_i \tilde{\sigma}_j \rangle}} \prod_{k \notin \{i,j\}} \left( \frac{1}{1 - 2p_{ijk}} \prod_{l \notin \{i,j,k\}} \frac{1}{1 - 2p_{ijkl}} \right), \quad (\text{E3c})$$

$$p_i = \frac{1}{2} - \frac{1}{2} \langle \tilde{\sigma}_i \rangle \prod_{j \neq i} \left( \frac{1}{1 - 2p_{ij}} \prod_{k \notin \{i,j\}} \left( \frac{1}{1 - 2p_{ijk}} \prod_{l \notin \{i,j,k\}} \frac{1}{1 - 2p_{ijkl}} \right) \right). \quad (\text{E3d})$$

Finally, we also numerically validated Eq. (10) on artificial datasets, where correlated errors triggering up to 12 syndromes exist; see Fig. 12.

#### APPENDIX F: BIAS IN CYCLE ERROR CALCULATION

We present an example of how not accounting for highly correlated error signatures can bias the extracted probabilities of errors with lower-weight signatures. Let us consider three syndrome elements  $\sigma_1$ ,  $\sigma_2$ , and  $\sigma_3$ , which are affected by three independent error processes. The first process flips the syndrome element  $\sigma_1$  with probability  $p_1 = 3\%$ , the second flips both syndrome elements  $\sigma_1$  and  $\sigma_2$  with probability  $p_{12} = 2.5\%$ , and the third process flips all syndrome elements with probability  $p_{123} = 1\%$ . The expectation values for the correlations between syndrome elements are given in Table III. If we use Eq. (10) to calculate the error probabilities from the correlations, we recover the original probabilities  $p_1$ ,  $p_{12}$ , and  $p_{123}$ . However, if we use the simplified formula (E3) that does not account for the highly correlated error process  $p_{123}$ , we get erroneous probabilities  $\check{p}$ ; see Table IV. We see that two-way correlations are overestimated by roughly  $p_{123}$ , because the three-way error process flips all pairs of syndromes. However,

the single-syndrome-flipping error probabilities are underestimated by  $p_{123}$ , because the two-way error probabilities that are effectively subtracted by the denominator in Eq. (E3b) were overestimated. We see that in some cases ( $\check{p}_2$  and  $\check{p}_3$ ), the extracted probabilities can even appear negative if some error channels are not accounted for.

#### APPENDIX G: SIMULATION MODEL

To obtain the simulation data shown in Fig. 3(b), we conduct a Clifford simulation using the PECOS [73] package, where we implement the circuit presented in Appendix H. Single- and two-qubit depolarizing noise is implemented by applying a random single- or two-qubit Pauli operator to the state with probability  $p$ . We use a uniform noise model, where operations on each qubit are subject to the same error probabilities. We apply single-qubit depolarizing noise during each idling step of duration  $t_{\text{IDL}}$  with probability  $p_{\text{IDL}} = (1 - e^{-t_{\text{IDL}}/\bar{T}})/4$ , where  $\bar{T} = 35 \mu\text{s}$  is approximately the mean  $T_1$  and  $T_2$  time of all the qubits. For single-qubit gates, we apply a single-qubit depolarizing channel with probability  $p_{1Q} = 0.0009$ . That is the average of single-qubit gate errors from randomized benchmarking. Similarly, for two-qubit gates, we apply a two-qubit depolarizing channel with probability  $p_{2Q} = 0.015$ , the average gate error from interleaved randomized benchmarking; see Appendix D. Finally, to model readout errors, we apply an  $\hat{X}$  gate before each readout with probability  $p_{\text{RO}} = 0.0116$ , which corresponds approximately to the average two-state readout error  $\epsilon_{\text{RO}}^{(2)}$ ; see Table II.

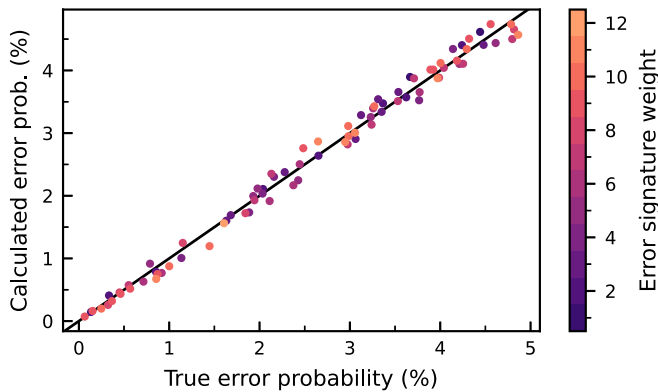


FIG. 12. Error probability as calculated according to Eq. (10) as a function of the true simulated error probability for 83 random error channels with signatures on up to 12 nodes. The simulation was done with 100 000 shots and took about 30 min on a laptop computer.

TABLE III. Syndrome correlations for the three-node example.

Correlation	Equation	Num.
$\langle \tilde{\sigma}_1 \rangle$	$(1 - 2p_1)(1 - 2p_{12})(1 - 2p_{123})$	0.875
$\langle \tilde{\sigma}_2 \rangle$	$(1 - 2p_{12})(1 - 2p_{123})$	0.931
$\langle \tilde{\sigma}_3 \rangle$	$1 - 2p_{123}$	0.980
$\langle \tilde{\sigma}_1 \tilde{\sigma}_2 \rangle$	$1 - 2p_1$	0.940
$\langle \tilde{\sigma}_2 \tilde{\sigma}_3 \rangle$	$1 - 2p_{12}$	0.950
$\langle \tilde{\sigma}_1 \tilde{\sigma}_3 \rangle$	$(1 - 2p_1)(1 - 2p_{12})$	0.893
$\langle \tilde{\sigma}_1 \tilde{\sigma}_2 \tilde{\sigma}_3 \rangle$	$(1 - 2p_1)(1 - 2p_{123})$	0.921

TABLE IV. Error probabilities  $\check{p}$  inferred using Eq. (E3) for the three-node example.

Inferred probability	Equation	Num.	Error
$\check{p}_1$	$(p_1 - p_{123})/(1 - 2p_{123})$	0.020	-0.010
$\check{p}_2$	$-p_{123}/(1 - 2p_{123})$	-0.010	-0.010
$\check{p}_3$	$-p_{123}/(1 - 2p_{123})$	-0.010	-0.010
$\check{p}_{12}$	$p_{12} + p_{123} - 2p_{12}p_{123}$	0.034	0.010
$\check{p}_{23}$	$p_{123}$	0.010	0.010
$\check{p}_{13}$	$p_{123}$	0.010	0.010

## APPENDIX H: FULL CIRCUIT DIAGRAM

A diagram of an example circuit four-cycle-long logical state preservation experiment for the  $|+\rangle_L$  state is shown in Fig. 13. The experimental data presented in this work were acquired using an equivalent 16-cycle-long circuit, preparing either  $|0\rangle_L$ ,  $|1\rangle_L$ ,  $|+\rangle_L$ , or  $|-\rangle_L$ . During the first stabilizer measurement cycle, we omit measuring the X-type (Z-type) stabilizers when preparing an eigenstate of the  $\hat{X}_L$  ( $\hat{Z}_L$ ) operator.

## APPENDIX I: EXAMPLE ANALYSIS OF VARYING ERROR RATES

As the simplest example that shows how the changes in the underlying error rates of the system during the data acquisition time can lead to apparent correlated errors, we consider the following system. There are two syndrome elements that independently flip with some probability. In the first scenario, let the error probability be constantly  $p$  throughout the data-gathering time. In this case, the syndrome correlations are

$$\langle \tilde{\sigma}_1 \rangle = \langle \tilde{\sigma}_2 \rangle = 1 - 2p, \quad (\text{I1a})$$

$$\langle \tilde{\sigma}_1 \tilde{\sigma}_2 \rangle = (1 - 2p)^2, \quad (\text{I1b})$$

and the error probabilities extracted using Eq. (10) are

$$p_{12} = \frac{1}{2} - \frac{1}{2} \sqrt{\frac{\langle \tilde{\sigma}_1 \rangle \langle \tilde{\sigma}_2 \rangle}{\langle \tilde{\sigma}_1 \tilde{\sigma}_2 \rangle}} = 0, \quad (\text{I2a})$$

$$p_1 = \frac{1}{2} - \frac{1}{2} \langle \tilde{\sigma}_1 \rangle = p, \quad (\text{I2b})$$

$$p_2 = \frac{1}{2} - \frac{1}{2} \langle \tilde{\sigma}_2 \rangle = p. \quad (\text{I2c})$$

In a second scenario, we set the individual error probabilities to zero for the first half of the data gathering and to  $2p$  for the second half, yielding syndrome correlations

$$\langle \tilde{\sigma}_1 \rangle = \langle \tilde{\sigma}_2 \rangle = 1/2 + (1 - 4p)/2 = 1 - 2p, \quad (\text{I3a})$$

$$\langle \tilde{\sigma}_1 \tilde{\sigma}_2 \rangle = 1/2 + (1 - 4p)^2/2 = (1 - 2p)^2 + 4p^2. \quad (\text{I3b})$$

We now find a nonzero probability for an error simultaneously flipping the two syndrome elements:

$$p_{12} = \frac{1}{2} - \frac{1}{2} \sqrt{\frac{\langle \tilde{\sigma}_1 \rangle \langle \tilde{\sigma}_2 \rangle}{\langle \tilde{\sigma}_1 \tilde{\sigma}_2 \rangle}} = \frac{1}{2} - \frac{1}{2} \sqrt{\frac{(1 - 2p)^2}{(1 - 2p)^2 + 4p^2}} \neq 0, \quad (\text{I4a})$$

$$p_1 = \frac{1}{2} - \frac{1}{2} \frac{\langle \tilde{\sigma}_1 \rangle}{1 - 2p_{12}} \neq p, \quad (\text{I4b})$$

$$p_2 = \frac{1}{2} - \frac{1}{2} \frac{\langle \tilde{\sigma}_2 \rangle}{1 - 2p_{12}} \neq p. \quad (\text{I4c})$$

In this extreme example, where the errors are fully off for half of the experiment, the apparent correlated error probability is equal to  $p^2$  in the limit of small  $p$ . For a smaller change in the error rates of  $\varepsilon$ , that is, error probabilities  $(1 - \varepsilon)p$  for half of the data gathering time and  $(1 + \varepsilon)p$  for the second half, we would find a correlated error probability proportional to  $\varepsilon^2 p^2$ . Due to the quadratic scaling in both  $\varepsilon$  and  $p$ , the effect is significant only for large changes in the error rates.

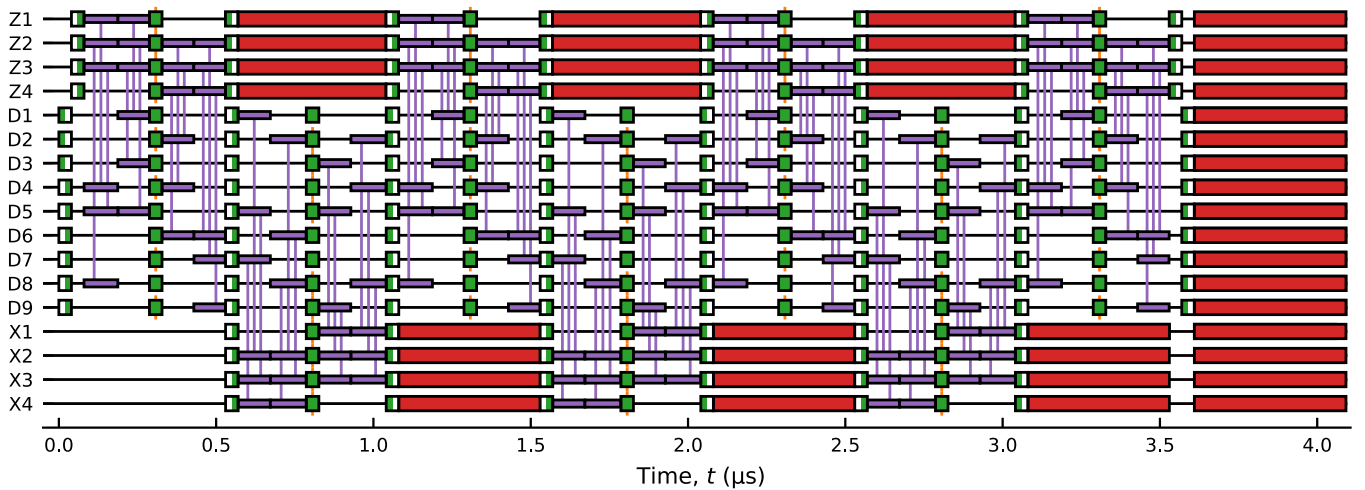


FIG. 13. Circuit diagram for a four-cycle-long logical state preservation experiment for the  $|+\rangle_L$  state. Green boxes denote single-qubit rotations around  $\hat{Y}$ , with the filling indicating the rotation angle: Left half filled is  $-\pi/2$ , right half filled is  $\pi/2$ , and fully filled is  $\pi$ . Orange vertical lines indicate virtual  $\pi$  rotations around  $\hat{Z}$ , which are implemented by flipping the phase of all following single-qubit gates. Red boxes indicate readout and pairs of purple boxes connected by a vertical line denote CZ gates.

- [1] M. Kjaergaard, M. E. Schwartz, J. Braumüller, P. Krantz, J. I.-J. Wang, S. Gustavsson, and W. D. Oliver, Superconducting qubits: Current state of play, *Annu. Rev. Condens. Matter Phys.* **11**, 369 (2020).
- [2] A. Blais, A. L. Grimsmo, S. M. Girvin, and A. Wallraff, Circuit quantum electrodynamics, *Rev. Mod. Phys.* **93**, 025005 (2021).
- [3] Z. Chen *et al.*, Exponential suppression of bit or phase errors with cyclic error correction, *Nature (London)* **595**, 383 (2021).
- [4] S. Krinner, N. Lacroix, A. Remm, A. D. Paolo, E. Genois, C. Leroux, C. Hellings, S. Lazar, F. Swiadek, J. Herrmann, G. J. Norris, C. K. Andersen, M. Müller, A. Blais, C. Eichler, and A. Wallraff, Realizing repeated quantum error correction in a distance-three surface code, *Nature (London)* **605**, 669 (2022).
- [5] Y. Kim, A. Eddins, S. Anand, K. X. Wei, E. van den Berg, S. Rosenblatt, H. Nayfeh, Y. Wu, M. Zaletel, K. Temme, and A. Kandala, Evidence for the utility of quantum computing before fault tolerance, *Nature (London)* **618**, 500 (2023).
- [6] Google Quantum AI and Collaborators, Quantum error correction below the surface code threshold, *Nature (London)* **638**, 920 (2024).
- [7] F. Arute *et al.*, Quantum supremacy using a programmable superconducting processor, *Nature (London)* **574**, 505 (2019).
- [8] H.-S. Zhong *et al.*, Quantum computational advantage using photons, *Science* **370**, 1460 (2020).
- [9] Y. Wu *et al.*, Strong quantum computational advantage using a superconducting quantum processor, *Phys. Rev. Lett.* **127**, 180501 (2021).
- [10] R. N. Tazhigulov, S.-N. Sun, R. Haghsheenas, H. Zhai, A. T. K. Tan, N. C. Rubin, R. Babbush, A. J. Minnich, and G. K.-L. Chan, Simulating models of challenging correlated molecules and materials on the Sycamore quantum processor, *PRX Quantum* **3**, 040318 (2022).
- [11] P. W. Shor, Scheme for reducing decoherence in quantum computer memory, *Phys. Rev. A* **52**, R2493 (1995).
- [12] E. Knill, R. Laflamme, and W. H. Zurek, Resilient quantum computation, *Science* **279**, 342 (1998).
- [13] B. M. Terhal, Quantum error correction for quantum memories, *Rev. Mod. Phys.* **87**, 307 (2015).
- [14] N. Lacroix *et al.*, Scaling and logic in the color code on a superconducting quantum processor, *Nature (London)* **645**, 614 (2025).
- [15] A. Paetzniak *et al.*, Demonstration of logical qubits and repeated error correction with better-than-physical error rates, [arXiv:2404.02280](https://arxiv.org/abs/2404.02280).
- [16] D. Gottesman, Stabilizer codes and quantum error correction, Ph.D. thesis, California Institute of Technology, 1997.
- [17] F. Battistel, C. Chamberland, K. Johar, R. W. Overwater, F. Sebastiano, L. Skoric, Y. Ueno, and M. Usman, Real-time decoding for fault-tolerant quantum computing: Progress, challenges and outlook, *Nano Futures* **7**, 032003 (2023).
- [18] N. Sundaresan, T. J. Yoder, Y. Kim, M. Li, E. H. Chen, G. Harper, T. Thorbeck, A. W. Cross, A. D. Córcoles, and M. Takita, Demonstrating multi-round subsystem quantum error correction using matching and maximum likelihood decoders, *Nat. Commun.* **14**, 2852 (2023).
- [19] V. Sivak, M. Newman, and P. Klimov, Optimization of decoder priors for accurate quantum error correction, *Phys. Rev. Lett.* **133**, 150603 (2024).
- [20] E. H. Chen, T. J. Yoder, Y. Kim, N. Sundaresan, S. Srinivasan, M. Li, A. D. Córcoles, A. W. Cross, and M. Takita, Calibrated decoders for experimental quantum error correction, *Phys. Rev. Lett.* **128**, 110504 (2022).
- [21] Google Quantum AI, Suppressing quantum errors by scaling a surface code logical qubit, *Nature (London)* **614**, 676 (2023).
- [22] E. Takou and K. R. Brown, Estimating decoding graphs and hypergraphs of memory QEC experiments, [arXiv:2504.20212](https://arxiv.org/abs/2504.20212).
- [23] S. T. Spitz, B. Tarasinski, C. W. J. Beenakker, and T. E. O'Brien, Adaptive weight estimator for quantum error correction in a time-dependent environment, *Adv. Quantum Technol.* **1**, 1800012 (2018).
- [24] R. Blume-Kohout and K. Young, Estimating detector error models from syndrome data, [arXiv:2504.14643](https://arxiv.org/abs/2504.14643).
- [25] Y. Tomita and K. M. Svore, Low-distance surface codes under realistic quantum noise, *Phys. Rev. A* **90**, 062320 (2014).
- [26] K. C. Miao *et al.*, Overcoming leakage in quantum error correction, *Nat. Phys.* **19**, 1780 (2023).
- [27] J. F. Marques, H. Ali, B. M. Varbanov, M. Finkel, H. M. Veen, S. L. M. van der Meer, S. Valles-Sanclemente, N. Muthusubramanian, M. Beekman, N. Haider, B. M. Terhal, and L. DiCarlo, All-microwave leakage reduction units for quantum error correction with superconducting transmon qubits, *Phys. Rev. Lett.* **130**, 250602 (2023).
- [28] N. Lacroix, L. Hofele, A. Remm, O. Benhayoune-Khadraoui, A. McDonald, R. Shillito, S. Lazar, C. Hellings, F. Swiadek, D. Colao-Zanuz, *et al.*, Fast flux-activated leakage reduction for superconducting quantum circuits, *Phys. Rev. Lett.* **134**, 120601 (2025).
- [29] M. McEwen *et al.*, Resolving catastrophic error bursts from cosmic rays in large arrays of superconducting qubits, *Nat. Phys.* **18**, 107 (2022).
- [30] D. C. McKay, C. J. Wood, S. Sheldon, J. M. Chow, and J. M. Gambetta, Efficient Z gates for quantum computing, *Phys. Rev. A* **96**, 022330 (2017).
- [31] F. W. Strauch, P. R. Johnson, A. J. Dragt, C. J. Lobb, J. R. Anderson, and F. C. Wellstood, Quantum logic gates for coupled superconducting phase qubits, *Phys. Rev. Lett.* **91**, 167005 (2003).
- [32] L. DiCarlo, J. M. Chow, J. M. Gambetta, L. S. Bishop, B. R. Johnson, D. I. Schuster, J. Majer, A. Blais, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, Demonstration of two-qubit algorithms with a superconducting quantum processor, *Nature (London)* **460**, 240 (2009).
- [33] V. Negîrmeac, H. Ali, N. Muthusubramanian, F. Battistel, R. Sagastizabal, M. S. Moreira, J. F. Marques, W. J. Vlothuizen, M. Beekman, C. Zachariadis, N. Haider, A. Bruno, and L. DiCarlo, High-fidelity controlled-Z gate with maximal intermediate leakage operating at the speed limit in a superconducting quantum processor, *Phys. Rev. Lett.* **126**, 220502 (2021).
- [34] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Surface codes: Towards practical large-scale quantum computation, *Phys. Rev. A* **86**, 032324 (2012).
- [35] R. Versluis, S. Poletto, N. Khammassi, B. Tarasinski, N. Haider, D. J. Michalak, A. Bruno, K. Bertels, and L. DiCarlo, Scalable quantum circuit and control for a superconducting surface code, *Phys. Rev. Appl.* **8**, 034021 (2017).
- [36] D. Greenbaum and Z. Dutton, Modeling coherent errors in quantum error correction, *Quantum Sci. Technol.* **3**, 015007 (2018).



- [37] S. Bravyi, M. Englbrecht, R. König, and N. Peard, Correcting coherent errors with surface codes, *npj Quantum Inf.* **4**, 55 (2018).
- [38] S. Krinner, S. Lazar, A. Remm, C. K. Andersen, N. Lacroix, G. J. Norris, C. Hellings, M. Gabureac, C. Eichler, and A. Wallraff, Benchmarking coherent errors in controlled-phase gates due to spectator qubits, *Phys. Rev. Appl.* **14**, 024042 (2020).
- [39] G. Duclos-Cianci and D. Poulin, Fast decoders for topological quantum codes, *Phys. Rev. Lett.* **104**, 050504 (2010).
- [40] M. Herold, M. J. Kastoryano, E. T. Campbell, and J. Eisert, Cellular automaton decoders of topological quantum memories in the fault tolerant setting, *New J. Phys.* **19**, 063012 (2017).
- [41] N. Delfosse, V. Londe, and M. E. Beverland, Toward a union-find decoder for quantum LDPC codes, *IEEE Trans. Inf. Theory* **68**, 3187 (2022).
- [42] N. Delfosse, A. Paetznick, J. Haah, and M. B. Hastings, Splitting decoders for correcting hypergraph faults, *arXiv:2309.15354*.
- [43] O. Higgott, T. C. Bohdanowicz, A. Kubica, S. T. Flammia, and E. T. Campbell, Improved decoding of circuit noise and fragile boundaries of tailored surface codes, *Phys. Rev. X* **13**, 031007 (2023).
- [44] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, Topological quantum memory, *J. Math. Phys.* **43**, 4452 (2002).
- [45] P. Iyer and D. Poulin, Hardness of decoding quantum stabilizer codes, *IEEE Trans. Inform. Theory* **61**, 5209 (2015).
- [46] S. Bravyi, M. Suchara, and A. Vargo, Efficient algorithms for maximum likelihood decoding in the surface code, *Phys. Rev. A* **90**, 032326 (2014).
- [47] S. C. Smith, B. J. Brown, and S. D. Bartlett, Local predecoder to reduce the bandwidth and latency of quantum error correction, *Phys. Rev. Appl.* **19**, 034050 (2023).
- [48] A. G. Fowler, Minimum weight perfect matching of fault-tolerant topological quantum error correction in average  $O(1)$  parallel time, *Quantum Inf. Comput.* **15**, 145 (2015).
- [49] O. Higgott, PyMatching: A python package for decoding quantum codes with minimum-weight perfect matching, *ACM Trans. Quantum Comput.* **3**, 1 (2022).
- [50] T. E. O'Brien, B. Tarasinski, and L. DiCarlo, Density-matrix simulation of small surface codes under current and projected experimental noise, *npj Quantum Inf.* **3**, 39 (2017).
- [51] J. Edmonds, Paths, trees, and flowers, *Can. J. Math.* **17**, 449 (1965).
- [52] O. Higgott and C. Gidney, Sparse blossom: Correcting a million errors per core second with minimum-weight matching, *Quantum* **9**, 1600 (2025).
- [53] Y. Wu and L. Zhong, Fusion blossom: Fast MWPM decoders for QEC, *arXiv:2305.08307*.
- [54] A. Paler and A. G. Fowler, Pipelined correlated minimum weight perfect matching of the surface code, *Quantum* **7**, 1205 (2023).
- [55] J. Bausch, A. W. Senior, F. J. H. Heras, T. Edlich, A. Davies, M. Newman, C. Jones, K. Satzinger, M. Y. Niu, S. Blackwell, G. Holland, D. Kafri, J. Atalaya, C. Gidney, D. Hassabis, S. Boixo, H. Neven, and P. Kohli, Learning high-accuracy error decoding for quantum processors, *Nature (London)* **635**, 834 (2024).
- [56] A. G. Fowler, Coping with qubit leakage in topological codes, *Phys. Rev. A* **88**, 042308 (2013).
- [57] B. M. Varbanov, F. Battistel, B. M. Tarasinski, V. P. Ostroukh, T. E. O'Brien, L. DiCarlo, and B. M. Terhal, Leakage detection for a transmon-based surface code, *npj Quantum Inf.* **6**, 102 (2020).
- [58] H. Ali, J. Marques, O. Crawford, J. Majaniemi, M. Serra-Peralta, D. Byfield, B. Varbanov, B. M. Terhal, L. DiCarlo, and E. T. Campbell, Reducing the error rate of a superconducting logical qubit using analog readout information, *Phys. Rev. Appl.* **22**, 044031 (2024).
- [59] D. Aharonov, A. Kitaev, and J. Preskill, Fault-tolerant quantum computation with long-range correlated noise, *Phys. Rev. Lett.* **96**, 050504 (2006).
- [60] P. Jouzdani, E. Novais, I. S. Tupitsyn, and E. R. Mucciolo, Fidelity threshold of the surface code beyond single-qubit error models, *Phys. Rev. A* **90**, 042315 (2014).
- [61] M. McEwen *et al.*, Removing leakage-induced correlated errors in superconducting quantum error correction, *Nat. Commun.* **12**, 1761 (2021).
- [62] C. J. Wood and J. M. Gambetta, Quantification and characterization of leakage errors, *Phys. Rev. A* **97**, 032306 (2018).
- [63] K. Serniak, M. Hays, G. de Lange, S. Diamond, S. Shankar, L. D. Burkhardt, L. Frunzio, M. Houzet, and M. H. Devoret, Hot nonequilibrium quasiparticles in transmon qubits, *Phys. Rev. Lett.* **121**, 157701 (2018).
- [64] A. G. Fowler, Optimal complexity correction of correlated errors in the surface code, *arXiv:1310.0863*.
- [65] N. Delfosse and J.-P. Tillich, A decoding algorithm for CSS codes using the X/Z correlations, in *2014 IEEE International Symposium on Information Theory* (Honolulu, HI, USA, 2014), pp. 1071–1075.
- [66] Y. Yuan and C.-C. Lu, A modified MWPM decoding algorithm for quantum surface codes over depolarizing channels, *arXiv:2202.11239*.
- [67] J. Fujisaki, K. Maruyama, H. Oshima, S. Sato, T. Sakashita, Y. Takeuchi, and K. Fujii, Quantum error correction with an Ising machine under circuit-level noise, *Phys. Rev. Res.* **5**, 043261 (2023).
- [68] M. Cain, C. Zhao, H. Zhou, N. Meister, J. Ataides, A. Jaffe, D. Bluvstein, and M. D. Lukin, Correlated decoding of logical algorithms with transversal gates, *Phys. Rev. Lett.* **133**, 240602 (2024).
- [69] J. Heinsoo, C. K. Andersen, A. Remm, S. Krinner, T. Walter, Y. Salathé, S. Gasparinetti, J.-C. Besse, A. Potočník, A. Wallraff, and C. Eichler, Rapid high-fidelity multiplexed readout of superconducting qubits, *Phys. Rev. Appl.* **10**, 034040 (2018).
- [70] E. Magesan, J. M. Gambetta, and J. Emerson, Scalable and robust randomized benchmarking of quantum processes, *Phys. Rev. Lett.* **106**, 180504 (2011).
- [71] E. Magesan, J. M. Gambetta, B. R. Johnson, C. A. Ryan, J. M. Chow, S. T. Merkel, M. P. da Silva, G. A. Keefe, M. B. Rothwell, T. A. Ohki, M. B. Ketchen, and M. Steffen, Efficient measurement of quantum gate error by interleaved randomized benchmarking, *Phys. Rev. Lett.* **109**, 080505 (2012).
- [72] A. D. Córcoles, J. M. Gambetta, J. M. Chow, J. A. Smolin, M. Ware, J. Strand, B. L. T. Plourde, and M. Steffen, Process verification of two-qubit quantum gates by randomized benchmarking, *Phys. Rev. A* **87**, 030301(R) (2013).
- [73] C. Ryan-Anderson, Quantum algorithms, architecture, and error correction, *arXiv:1812.04735*.