Tailoring quantum error correction to spin qubits

Bence Hetényi ^{®*} and James R. Wootton IBM Quantum, IBM Research Zurich, CH-8803 Rüschlikon, Switzerland

(Received 28 July 2023; accepted 27 February 2024; published 26 March 2024)

Spin qubits in semiconductor structures bring the promise of large-scale two-dimensional integration, with the possibility to incorporate the control electronics on the same chip. In order to perform error correction on this platform, the characteristic features of spin qubits need to be accounted for. For example, qubit readout involves an additional qubit which necessitates careful reconsideration of the qubit layout. The noise affecting spin qubits has further peculiarities such as the strong bias towards dephasing. In this work we consider state-of-the-art error correction codes that require only nearest-neighbor connectivity and are amenable to fast decoding via minimum-weight perfect matching. Compared to the surface code, the XZZX code, the reduced-connectivity surface code, the XYZ² matching code, and the Floquet code all bring different advantages in terms of error threshold, connectivity, or logical qubit encoding. We present the spin-qubit layout required for each of these error correction codes, accounting for reference qubits required for spin readout. The performance of these codes is studied under circuit-level noise accounting for distinct error rates for gates, readout, and qubit decoherence during idling stages.

DOI: 10.1103/PhysRevA.109.032433

I. INTRODUCTION

Fault-tolerant quantum computation requires large-scale quantum processors, where quantum states can be encoded in a noise-free subspace of a myriad of noisy qubits. Spin qubits in semiconductor quantum dots offer critical features compatible with having millions of qubits locally connected in a two-dimensional (2D) lattice, making them auspicious candidates for fault-tolerant quantum computing [1-3]. Additionally, the half-century-long experience of the semiconductor industry is believed to provide a key advantage over other quantum computing platforms [2]. Spin-qubit platforms to date have demonstrated single- and two-qubit operations and spin readout above 99% fidelity [4–6] in devices with up to six qubits [7].

Provided that the error rates of the physical qubits are below a certain threshold value, quantum error correction (QEC) codes can suppress errors exponentially with the number of qubits [8-10]. One of the most famous QEC codes is the surface code, which requires only nearest-neighbor interaction between qubits on a square grid [11,12]. Alongside the modest connectivity requirements, the popularity of the surface code lies on the high threshold error rate and the availability of fast and high-performance classical decoding schemes such as minimum-weight perfect matching (MWPM) [13]. Furthermore, numerous schemes have been developed for the surface code that enable universal quantum computation in a faulttolerant way [14]. In recent years, several QEC codes have been proposed that fulfill the above-listed criteria [15–20]. Some of them even overtake the surface code in terms of error threshold [17] and connectivity requirements [18–20].

The threshold error rate is an important target for qubit platforms, however, its value depends strongly on the noise model assumed to calculate it. In circuit-level error models, individual physical gate errors, qubit readout errors and decoherence during idling are all taken into account [21]. To simplify the model, these errors are assumed to happen with the same probability which makes the resulting numbers hard to compare with physical scenarios where some error mechanisms are more pronounced than others. Moreover, hardware-specific constraints, like noise biases, can significantly change the quantum circuit introducing additional noise channels in the model [22].

Here we provide a detailed quantitative analysis for a wide range of gate and readout fidelities, decoherence, and measurement times. Simple formulas are derived to help future experiments estimating their device-specific thresholds and the qubit overhead required for fault tolerance. Our work serves to determine important experimental details required to perform quantum memory experiments. In particular we find fault-tolerant circuits for syndrome measurements and logical state preparation and readout, adapted to each QEC code considered and tailored to the needs of spin-qubit readout.

Moreover, in present readout schemes for spin qubits, defined in single quantum dots, two qubits are required for the readout, one of which acts as a reference spin [23]. In such a setting only one qubit of the readout system may be used to encode information. Considering multiple QEC codes, qubit layouts are proposed in this work that acknowledge this constraint. Some codes can even benefit from this peculiar feature, by reaching a higher effective connectivity, or a sparser grid of qubits that relaxes some requirements for the large-scale fabrication techniques.

This paper is organized as follows in Sec. II the basic concepts of QEC are introduced followed by the derivation of the noise model for spin qubits and some simple formulas

^{*}bence.hetenyi@ibm.com

TABLE I. Spin-qubit connectivity (i.e., average number of two-qubit connections in a 2D lattice) and the parameters of the linearized threshold surface appearing in Eq. (13). The total number of qubits and ancillas are given by $N_{q,a} = v_{q,a}d^2$ for a code distance *d*. The overall performance of the fitting is further characterized by the maximum and the mean deviation from the numerical value, i.e., $\delta p^{\text{th}} = p^{\text{th}} - p^{\text{num,th}}$. Thresholds are given at $\eta_T = 20$ and $\eta_G = 1$. Asterisk denotes syndrome measurement via Bell-state preparation.

	Connectivity	v_q/v_a	p_G^{th} (%)	p_T^{th} (%)	p_R^{th} (%)	$\sqrt{\left\langle \left(\frac{\delta p^{\mathrm{th}}}{p^{\mathrm{th}}} ight)^2 \right\rangle}$
Surface code	$3\frac{1}{3}$	1/2	0.82	3.94	14.5	0.08
code	$3\frac{1}{3}$	1/2	0.37	15.1	12.9	0.11
3-CX surface code	$2\frac{2}{3}$	1/2	0.65	4.1	8.7	0.1
XYZ^2 code	4	2/2	0.465	4.2	16.5	0.12
Floquet color code*	$2\frac{1}{4}$	1.5/4.5	0.48	0.7	1.41	0.005
Honeycomb Floquet code*	$2\frac{1}{4}$	1.5/4.5	0.43	1.16	0.99	0.015

are presented relating the parameters of our error model to the experimental figures of merit. In the last subsection we derive the optimal measurement time for QEC applications accounting for qubit decoherence during midcircuit measurements. These concepts are applied to the surface code in Sec. III revealing some of the hitherto unexplored parts of the noise-parameter space. Different QEC codes and the corresponding spin-qubit layouts are presented in Sec. IV, comparing the performance of different codes supplying valuable information about the connectivity qubit-overhead tradeoff for future experiments. The main results from which the error threshold for a device-specific noise model can be deduced are summarized in Table I. An outlook for future work and concluding remarks are contained in Sec. V and VI.

II. ERROR CORRECTION WITH SPIN QUBITS

A. A brief introduction to quantum error correction

In this section we introduce only the basic concepts and methods in QEC to motivate our choice of the error correction codes and to familiarize the reader with the terminology used in our work. Rather than providing precise mathematical definitions, we instead focus on qualitative explanations for the sake of conciseness. The interested reader may consult a more exhaustive review in the topic such as Refs. [24,25].

Our goal with quantum error correction is to encode logical qubits, i.e., qubits with arbitrarily suppressed error rates, in a low-dimensional subspace of several physical qubits, such that single- and two-qubit gates can be applied between the encoded logical qubits which are read out at the end of the fault-tolerant quantum circuit. One of the prime candidates for such an encoding is the surface code, where it has been shown that all the necessary ingredients can be implemented assuming entangling gates only between nearest-neighbor physical qubits. The rigorous description of some of these methods is beyond the scope of this paper, therefore, we resort to quantum memory experiments. In a quantum memory experiment a single logical qubit is encoded in the logical X or Z, left idling, while mutually commuting observables are repeatedly being measured, and then read out. Logical errors can be revealed indirectly from the collected observables without having measured the logical qubit.

Since only one qubit of information is needed to be preserved, taking a connected lattice of N physical

qubits, N - 1 independent multiqubit measurements can be performed, provided that the measurement statistics of the logical subspace remains unaffected. In *stabilizer codes*, a set of mutually commuting operators, called *stabilizer operators*, define these measurements. Logical operators commute with all stabilizers and anticommute with each other. Stabilizer operators are defined as Pauli strings acting on ≥ 2 physical qubits. Logical operators act on $\ge d$ qubits, where d is called the *code distance*. The code distance defines the smallest number of independent errors that cannot be detected by stabilizer operators since it is the smallest number of Pauli operators required to perform a logical gate.

Stabilizer measurements ensure that in every cycle the system is projected back into the logical subspace. After several rounds of syndrome measurements, the logical qubit may be read out by measuring every qubit on which the logical operator has a support in the appropriate basis. Since physical errors may or may not have changed the logical information a *decoder* is employed, a decoder is a classical algorithm that uses all the stabilizer measurement outcomes (called *syndromes*) to guess whether the physical errors have changed the outcome of the logical qubit measurement compared to the encoded information.

Fault tolerance can be achieved if the decoder has better and better success rate as the number of qubits used for the encoding is increased. The conditions under which this is possible are formulated in the *threshold theorem*. The theorem can be formulated in multiple ways, using different assumptions on the noise processes [8–10]. Let us phrase the statement in the following way: *there exists a finite physical error probability pth of local errors, below which a certain accuracy* ϵ *of the quantum computation of depth D can be ensured by encoding logical qubits in polylog*(D/ϵ) *physical qubits.* Summarizing the arguments above, the threshold error rate p^{th} of the physical qubits depends on three factors: (i) the error correction code itself, (ii) the decoding algorithm, (iii) and the error model.

Here we focus on a special class of stabilizer codes which is defined by the topological properties of the error syndromes which is easiest to understand in the Hamiltonian formalism. A Hamiltonian can be defined from the negative sum of all stabilizer operators. The ground state of this Hamiltonian (i.e., the mutual +1 eigenstate of all stabilizers) is twofold degenerate. Logical operators act on this subspace as gauge operators. Single-qubit errors change some of the syndromes, creating pointlike excitations above the ground state. If the possible excitations in the model are composed of two bosonic paticles which are their own antiparticle and have nontrivial braiding statistics, the model is called a $D(\mathbb{Z}_2)$ anyon model [26]. Throughout this work we will refer to this set of codes as the *surface code family* after its most prominent representative, the surface code.

The choice of the decoder has a large impact on the error threshold. The maximum-likelihood decoder is shown to result in an optimal decoding, but has an exponential computational complexity which makes it unfeasible for system sizes large enough for fault-tolerant quantum computations especially if the syndrome information needs to be processed in real time [27]. For members of the surface code family, error syndromes always appear pairwise due to the general properties of the underlying anyon model. This crucial property makes these codes amenable for decoding by minimum-weight perfect matching (MWPM) which is almost linear computational complexity [13]. Together with the qualitative similarities to the well-established surface code, this motivated our choice to focus on the surface code family and use the same MWPM decoding scheme throughout our work.

The error model incorporates probability of different type of errors, as well as spatial and temporal correlations between error events. In the next subsections we will consider the physical origin of noise in spin-qubit devices and derive an error model that acknowledges the special properties of this platform. The error model needs to be efficiently simulable, such that one can make predictions about the resources required for fault-tolerant quantum computing.

B. Noise model of spin qubits

In this section we consider some common features and constraints of spin-qubit devices and use them to establish the error model to be used for the calculation of the threshold surface of the surface code and its comparison to other QEC codes of the same family. There are multiple ways to encode qubits in the spin and orbital degrees of freedom of semiconductor quantum dots [3]. Here we focus on the type of spin qubits, where the qubit states correspond to the spin projections of a single electron or hole in a quantum dot that is split by a magnetic field according to the original proposal of Loss and DiVincenzo [1].

Arbitrary single-qubit rotations as well as the two-qubit CX gate can be implemented natively in with spin qubits [7]. Gate errors are often characterized with a single number, the gate fidelity \mathcal{F}_G . As opposed to gate tomography, the fidelity does not give detailed information about the noise channels. Since different spin-qubit platforms can have very different noise channels, we will focus on the gate fidelity and assume that single- and two-qubit gates are followed by single- and two-qubit depolarizing noise, i.e., the *m*-qubit density matrix becomes $\rho_m \rightarrow (1 - p_{Gm})\rho_m + 2^{-m}p_{Gm}\mathbb{1}_{2^m}$, where the probabilities are determined by the fidelity as

$$p_{G1} = 1 - \mathcal{F}_{G1} \equiv p_G \frac{2}{1 + \eta_G},$$
 (1a)

$$p_{G2} = 1 - \mathcal{F}_{G2} \equiv p_G \frac{2\eta_G}{1 + \eta_G},$$
 (1b)

with p_G being the average gate fidelity and the error bias between single- and two-qubit gates reads as

$$\eta_G = \frac{p_{G2}}{p_{G1}}.$$
 (2)

An idling spin qubit loses its phase coherence at a higher rate than the change in population of its basis states. Dephasing can come from the low-frequency fluctuations of the qubit splitting that changes the precession frequency of the spin around the axis of the magnetic field. Such low-frequency noise often originates from the slow dynamics of nuclear spins, or 1/f charge noise coupling to the spin splitting via spin-orbit interaction [28,29]. Relaxation, on the other hand, is dominated by phonon emission which is typically suppressed by the low phonon density of states at the energy of the qubit splitting [30].

Relaxation can be treated in the Bloch-Redfield approximation [31], which yields an exponential decay of the diagonal elements of the density matrix. This is equivalent to Pauli *X* and *Y* errors [i.e., $\rho \rightarrow (1 - p_{T1})\rho + \frac{p_{T1}}{2}(X\rho X + Y\rho Y)$] with probability [32]

$$p_{T_1} = \frac{1 - e^{-\tau_i/T_1}}{4} \approx \frac{\tau_i}{4T_1}$$
(3)

with T_1 being the relaxation time and τ_i is the idling time. On the other hand, dephasing due to low-frequency noise is better described in the filter function formalism [3], leading to a Gaussian decay of the off-diagonal element of the density matrix with a timescale T_2 . This process corresponds to Pauli-Z errors [i.e., $\rho \rightarrow (1 - p_{T_2})\rho + p_{T_2}Z\rho Z$] with probability

$$p_{T_2} = \frac{1 - e^{-\tau_i^2/T_2^2}}{2} - \frac{1 - e^{-\tau_i/T_1}}{4} \approx \frac{\tau_i^2}{2T_2^2} - \frac{\tau_i}{4T_1}.$$
 (4)

The difference in decoherence rates can be quantified with the noise bias

$$\eta_T = \frac{p_{T_2}}{p_{T_1}} \approx \frac{2\tau_i T_1}{T_2^2} - 1,$$
(5)

while the total idling error rate reads as $p_T = p_{T_1} + p_{T_2}$. However, we note that $\eta_T \ll T_2/T_1$ (since $\tau_i \ll T_1, T_2$), meaning that the noise bias can be substantially lower than the naive expectation of $\eta_T \sim T_2/T_1$. Furthermore, one might apply an arbitrary dynamical decoupling pulse sequence during idling further improving the T_2 dephasing time that enters the above equations. Using these considerations we expect $\eta_T \sim O(10)$ [5].

Readout of spin qubits is typically carried out via spin-tocharge conversion and charge sensing. Here we only consider those conversion schemes which do not require a reservoir connected to the quantum dot accommodating the qubit. The conversion of spin to charge involves two spin qubits in close proximity, one in a known state and another one to be measured. Reducing the tunnel barrier between the two quantum dots gives rise to a spin-selective tunneling, i.e., Pauli-spin blockade, after which the (change in) charge state of the quantum dot can be detected. Alternatively, one can exploit the strong spin-photon coupling in a setup where a single particle with strong spin-orbit interaction is situated in a double quantum dot that is coupled to a resonator. From an architecture point of view, both of these approaches require twice the space of a single qubit. Therefore, in the following we assume that readout involves pair of qubits in the qubit layouts to be presented.

The equipment for charge sensing imposes further restrictions on the qubit architecture. Charge sensing can be carried out using single-electron transistors (SET) in a radiofrequency reflectometry setup, which is a popular method for small-scale devices. However, this requires an SET in the proximity of the spin qubits which has a substantial footprint compared to the quantum dot dimensions. Scalable architectures can potentially employ gate-dispersive sensing techniques with signal multiplexing, similar to that of superconducting qubits, relaxing the spatial requirements of the sensor near the qubits [33].

An important aspect of the readout is the compromise to be made between measurement time τ_R and fidelity. The fidelity of charge sensing is exponentially improved by increasing the measurement time [33]. That does not mean that the longer one measures, the better it gets. Measurement time is limited by relaxation processes (deep in the Pauli blockade for charge sensing) or Landau-Zener tunneling (in the case of a dispersive readout). The readout error can be described (qualitatively) as

$$p_R(\tau_R) = 1 - \mathcal{F}_R(\tau_R) = 1 - (1 - e^{-\tau_R/\tau_{\min}})e^{-\tau_R/T_{1R}}, \quad (6)$$

where T_{1R} is the decoherence time of the qubit being read out and τ_{\min} is the minimum integration time, i.e., time needed to achieve a signal-to-noise ratio of 2 [33]. The maximum readout fidelity is then achieved at the measurement time

$$\tau_R^* = \tau_{\min} \ln \left(1 + \frac{T_{1R}}{\tau_{\min}} \right). \tag{7}$$

In our error model readout errors are two-qubit depolarizing errors followed by a classical bit flip on the measurement outcome (e.g., infidelity of the charge detector). Both of these lead to a faulty syndrome bit with a joint probability p_R .

Finally, we note that within the Pauli noise model, reset errors during the stabilizer measurements generate the same syndrome as readout errors. Therefore, we merge these error rates into a single error parameter, the reset-readout error rate

$$p_{\rm RR} = p_{\rm res}[1 - p_R(\tau_R)] + p_R(\tau_R)(1 - p_{\rm res}),$$
 (8)

where p_{res} is the probability that a faulty initialization flips the measurement outcome [e.g., for depolarizing noise $p_{\text{res}} = \frac{8}{15}(1 - \mathcal{F}_{\text{init}})$].

C. Error-threshold surface and resource estimation

If the ratio of different probabilities is kept constant, one obtains a single-parameter error model where p = 0 corresponds to no errors. Therefore, according to the threshold theorem a finite error threshold p^{th} can be found, the value of which will depend on the ratio of error probabilities. Asymptotically, for $p \ll p^{\text{th}}$ the logical failure rate can be approximated as

$$P_L \sim w \left(\frac{p}{p^{\text{th}}}\right)^{d/2},$$
 (9)

where w is a prefactor that depends on how many length-d/2 paths can lead to logical failure. Considering different ratios

of the error parameters, i.e., different directions in the errorparameter space, and calculating the corresponding threshold values maps out a threshold surface in the parameter space \mathbf{p} that encloses the origin ($\mathbf{p} = 0$). Inside the threshold surface the logical error rate can be decreased arbitrarily by increasing the number of qubits.

Fixing the two error bias parameters η_G and η_T , we are left with a three-dimensional space of (p_G, p_T, p_{RR}) . In the simplest case, when the threshold surface is a plane and determined by three points $(p_G^{\text{th}}, 0, 0), (0, p_T^{\text{th}}, 0)$, and $(0, 0, p_{RR}^{\text{th}})$, it is straightforward to show that

$$\frac{p}{p^{\text{th}}} = \frac{p_G}{p_G^{\text{th}}} + \frac{p_T}{p_T^{\text{th}}} + \frac{p_{\text{RR}}}{p_{\text{RR}}^{\text{th}}},\tag{10}$$

where $p_G^{\text{th}} = p_G^{\text{th}}(\eta_G)$ and $p_T^{\text{th}} = p_T^{\text{th}}(\eta_T)$. This formula is in correspondence with Ref. [34], where $p/p^{\text{th}} \propto \Lambda^{-1}$, and the observed nonlinearities of the threshold surface corroborate with the effects observed in the experiment (see Supplementary Information of [34]). In that case Eq. (13) implies that the isotropic circuit level noise ($p_G = p_T = p_{\text{RR}} = p$) threshold can be recovered as $p_{\text{ic}}^{-1} = (p_G^{\text{th}})^{-1} + (p_T^{\text{th}})^{-1} + (p_{\text{RR}}^{\text{th}})^{-1}$, whereas the threshold for the phenomenological noise model ($p_T = p_{\text{RR}} = p$) is $p_{\text{ph}}^{-1} = (p_T^{\text{th}})^{-1} + (p_{\text{RR}}^{\text{th}})^{-1}$. In Table I we summarize the parameters of the linearized threshold surface introduced in Eq. (13) for the six QEC codes studied in the upcoming sections.

Assuming that a given error configuration is below the error threshold according to Eq. (13) one can estimate the number of qubits required for a single logical qubit with practical logical error rate. Let us fix a target logical error rate, i.e., $P_L/w = 10^{-12}$. The required code distance then becomes

$$d = \left\lceil -\frac{24}{\ln(p/p^{\text{th}})} \right\rceil,\tag{11}$$

implying a total qubit count of $N_{\text{tot}} = (v_q + v_a)d^2$ for a single logical qubit, where $v_{q(a)}$ is the number of qubits (ancillas) in a unit cell of the QEC code. For example, for $p/p^{\text{th}} = 0.5$ one needs $N_{\text{tot}} = 6400(v_q + v_a) \sim O(10^4)$ qubits. Less qubits are required as noise decreases. For example, $N_{\text{tot}} = 576(v_q + v_a) \sim O(10^3)$ qubits suffice for $p/p^{\text{th}} = 0.1$.

D. Tradeoff between measurement time and fidelity

The maximal-fidelity readout time τ_R^* in Eq. (7) gives the optimal readout error rate if every qubit is measured simultaneously, e.g., at the end of the circuit. However, in quantum error correction data qubits are idling during the measurement of the ancillas, therefore, one could expect that it is worthwhile to move away from the maximal readout fidelity point (by reducing the measurement time) in order to improve the idling error rates.

We have seen in Eq. (6) how the readout error rate depends on the integration time τ_R . Furthermore, neglecting relaxation for simplicity, the idling error rate during readout and reset reads as

$$p_{T_2}(\tau_R) \approx \frac{1 - e^{-(\tau_R + \tau_{\rm res})^2/T_2^2}}{2},$$
 (12)

where τ_{res} is the reset timescale. Using $p_{RR}(\tau_R)$ and $p_{T_2}(\tau_R)$ one can derive the readout time that takes the readout-reset and idling error rates furthest away from the threshold surface.

Sticking to our example of a linear threshold surface characterized by p_G^{th} , p_T^{th} , and p_R^{th} , the distance from the threshold surface is given by

$$\frac{p}{p^{\text{th}}} = \frac{p_G}{p_G^{\text{th}}} + \frac{p_T(\tau_R)}{p_T^{\text{th}}} + \frac{p_{\text{RR}}(\tau_R)}{p_{\text{RR}}^{\text{th}}}.$$
 (13)

In order to reduce the qubit overhead, we need to minimize this ratio with respect to τ_R . Assuming that $\tau_R \ll T_2$, T_{1RO} the readout time that is optimal for QEC applications becomes

$$\tau_{\rm QEC}^* = \tau_R^* - \tau_{\rm min} \ln \left(1 + \frac{p_R^{\rm th}}{p_T^{\rm th}} \frac{(\tau_{\rm QEC}^* + \tau_{\rm res}) T_{1R}}{(1 - 2p_{\rm res}) T_2^2} \right).$$
(14)

This equation can easily be solved recursively, starting from $\tau_{\text{QEC}}^* = \tau_R^*$ on the right-hand side. As we will see later on the example of the surface code, it is even possible that the maximal fidelity readout time falls outside the threshold surface, while τ_{OEC}^* is deep inside it.

E. Surface code with circuit-level noise

Let us consider Kitaev's surface code with rotated boundaries [12]. In this QEC code, d^2 data qubits are arranged in a square grid and the group of stabilizer operators contain (d - d) $(1)^2$ plaquettes with products of Pauli-X and Pauli-Z operators acting on four nearest-neighbor qubits. At the boundaries of the lattice one needs 2d - 2 stabilizers with support on only two qubits each yielding $d^2 - 1$ constraints for the d^2 degrees of freedom. Measuring all the stabilizers is then equivalent to completing four- and two-body parity measurements on neighboring qubits. In what follows, we consider the standard circuit representation of the surface code and consider individual errors during a quantum memory experiment that can lead to nonlocal error chains and show a way to neutralize them. Afterwards, a possible experimental protocol of initializing and reading out a logical qubit is discussed, assuming noisy ingredients throughout the entire process.

Provided that every qubit can be read out individually, the four-body stabilizer measurements of the surface code can be achieved by adding ancilla qubits at the center of every plaquette, leading to denser square grid of qubits [11]. Ancillas are reset to the state $|0\rangle$ before every stabilizer round and therefore do not add degrees of freedom to the system. The *Z*-plaquette measurements are then implemented via four controlled-not (CX) gates controlled on corresponding data qubits and targeted on the ancilla. The controlled gates will flip the state of the ancilla an even or an odd number of times depending on the parity of the four data qubits. Measuring the ancilla then yields the eigenvalue of the *Z* stabilizer.

Assuming any gate in the stabilizer measurement circuit can induce an error in the corresponding part of the circuit, errors can propagate from the ancilla qubit to the data qubits via the CX gates (e.g., X errors spread from the control to the target, and Z errors from the target to the control). The worst-case-scenario errors, called *hook errors* in the literature, are the X (Z) errors in the X-stabilizer (Z-stabilizer) circuits that occur before the third CX gate because this error will spread to two of the data qubits [11]. Four (three) data qubit PHYSICAL REVIEW A 109, 032433 (2024)

error), that is less harmful than two data-qubit errors. In the rotated surface code, the X (Z) logical operator is a Pauli string that acts on a column (row) of qubits. Depending on the schedule of CX gates, hook errors can be a distance-2 substring of a logical operator, meaning that half as many of them are required to induce an undetectable logical error, i.e., reducing the effective code distance to $\lceil d/2 \rceil$. With careful scheduling, however, it can be ensured that no single error event in the circuit can induce errors that reduce the code distance in the surface code [11].

Noisy logical initialization and readout

A quantum memory experiment consists of the initialization of the logical qubit in the Z(X) basis followed by several rounds of stabilizer measurements and the final readout of the logical qubit on the same basis. Afterwards a decoder is used to infer from the syndrome data if a logical X(Z) error happened. Comparing the initialized logical eigenvalue to the final corrected logical readout of the simulation allows one to calculate logical error rates for different physical error rates and code distances to determine the error threshold.

The initialization of the logical qubit is often carried out assuming noise-free initial and final stabilizer measurements as well as perfect initialization and measurement of the data qubits. Here we briefly review this protocol to show that it leads to an unphysical solution for the threshold surface, and compare this result with a fault-tolerant protocol we used in our work, where every qubit and quantum gate is subject to the noise model introduced in Sec. II B.

In the case of the ideal logical initialization and readout protocols one prepares every qubit in the $|0\rangle$ ($|+\rangle$) state which is an eigenstate of the logical Z(X) operator. Performing the first round of stabilizer measurements, the X(Z) stabilizer outcomes will be random even in the absence of errors since the physical qubit resets did not prepare a surface code state. Since stabilizers commute with the logical operator, the logical eigenvalue is still intact after the first round of stabilizers, but the system is projected into a subspace of some given X(Z) stabilizer eigenvalues. From the second round, every stabilizer measurement would return the same syndrome as in the first one, so one can start inserting errors and the decoder will have enough information to deal with the logical correction. The final measurement of the logical follows a similar logic: performing a noise-free round of stabilizer measurements before reading out the physical qubits that reveal the encoded logical eigenvalue. It is easy to see that the logical eigenvalue cannot be corrupted in this protocol if only readout or reset errors happen during the noisy stabilizer rounds.

In a real experiment, however, the initialization and final readout are noisy processes as well. If the errors are strongly polarized towards readout errors, the logical readout will be heavily affected by errors, implying a finite threshold error value against readout errors. The preparation of the initial state can be performed fault tolerantly with noisy operations only as well. If a code contains pure Pauli-X and pure Pauli-Z stabilizers and logical operators, only the Z-stabilizer (X-stabilizer) outcomes are needed to determine whether a logical Z(X) error happened since X(Z) errors do not affect the



FIG. 1. (a) Qubit layout and connectivity of a bulk section of the surface code. Gray filled nodes represent qubits that are connected to ancillas (empty nodes) via single links representing connections via two-qubit gates. Ancillas can be read out pairwise or swapped with each other (double link). X and Z stabilizer plaquettes are shown in red and blue, respectively. (b) Circuit representation of qubits q_1 and q_2 taking part in an X and a Z stabilizer measurement. The qubits shown are also part of two additional stabilizer measurements each (implied by faint gates). The CX schedule and the connectivity necessitate three (one) SWAP gates for the X (Z) stabilizer measurements. Red, cyan, and blue colored elements in the circuit represent single- and two-qubit depolarizing noise (or bit flip on classical lines) for readout, single-, and two-qubit gates, respectively. During the measurement qubits experience X - Y (dark green) and Z errors (light green) with different rate.

X stabilizers (Z stabilizers). We can, therefore, prepare every qubit in the $|0\rangle$ ($|+\rangle$) state initially and read out every qubit on the Z (X) basis in the end. The initial data qubit resets ensure that all Z- (X-) stabilizer eigenvalues are known before the first round, and final measurements can be used to infer the relevant syndromes after the last round of stabilizer measurements, thereby allowing for errors to be injected at any point in the process [35].

III. SURFACE CODE WITH SPIN QUBITS

In the previous section we have seen that the surface code can be conveniently embedded in a square lattice of qubits. However, since spin-qubit readout requires two qubits, the square-grid qubit layout needs to be modified. An example for such a modified lattice was provided by Ref. [2] and shown in Fig. 1(a). Data qubits are connected to four neighboring ancillas via two-qubit gates but do not participate in a readout pair. Ancillas, on the other hand, always come in pairs where they can be swapped, reset, or read out in a single step.

A pair of ancillas can be initialized directly in the $|00\rangle$ state with a fidelity \mathcal{F}_{init} and their Z parity can be read out in a partially destructive process [36,37] we denote as a ZZ^* measurement box. Keeping one of the ancillas in the $|0\rangle$ state as a reference, the Z-stabilizer measurement circuits can be realized using a single SWAP gate as shown for the upper ancilla pair on Fig. 1(b). The X-stabilizer measurements, on the other hand, require in total three SWAPs [see the lower ancilla pair of Fig. 1(b)]. Without the additional SWAP gates, both X and Z plaquettes are bound to have the same (or equivalent) CX schedules. Consequently, X-error pairs and Z-error pairs are injected in the same direction, reducing the code distance for either of the logical operators regardless of the choice of boundary conditions. Alternative stabilizer measurement schemes, trying to leverage the second ancilla qubit, can be found in Appendix A.

If we restrict our attention to memory experiments, it is tempting to propose a qubit connectivity where the ancilla pairs are placed such that the the number of SWAP gates enforced by the layout is minimized. At the same time, moving towards fault-tolerant quantum computing with multiple logical qubits, such a hard-wiring of the CX schedule in the physical qubit layout is not possible. For example, twist defects require an on-demand change in the checkerboard pattern of plaquette operators [14].

Fault-tolerant logical-state preparation and readout requires data qubits to be initialized and read out. This can be done by adding a single pair of ancillas to the surface code lattice, allowing one to assign one ancilla pair to every code qubit. Initialization is done by resetting the ancillas, swapping the data qubit with the ancilla it is connected to, and performing a second reset on the ancillas. Similarly the final measurement can be carried out using a single layer of SWAP gates. Consequently, data qubit measurement does not require additional hardware. Further details about the CX schedule, logical initialization, and measurement can be found in Appendix B.

Calculating the logical failure rate as a function of $p = (p_T^2 + p_{RR}^2)^{1/2}$ for different code distances, the intersection of failure rates yields the threshold for a given p_{RR}/p_T . Repeating this for several distinct ratios one obtains a threshold curve in the parameter space of $(0, p_T, p_R)$. In addition, it is important to consider the logical qubit initialized both along the X and Z axes (taking the smaller threshold) since the biased idling noise can strongly favor one type of logical operator. On Fig. 2(a) we compare the threshold curves with the logical initialization and readout performed using ideal stabilizer circuits and the fault-tolerant protocol for $p_G = 0$. As expected, only the fault-tolerant protocol leads to finite threshold for readout errors only showcasing the importance of the logical initialization protocol in memory-experiment simulations.

Using the linear approximation of the threshold curve, we show that in certain cases the maximal readout fidelity is far from the optimal choice for error correction. For the parameter values on Fig. 2(b) the maximum fidelity $(\mathcal{F}_R \approx 99\%)$ is achieved at $\tau_R^* = 9.8 \,\mu$ s, whereas the optimal choice is $\tau_{\text{QEC}}^* = 5.8 \,\mu$ s for $T_2 = 50 \,\mu$ s. Using the optimal



FIG. 2. (a) Threshold curves for ideal logical state preparation (magenta) and noisy logical state preparation in a quantum memory simulation. At low idling error rates the ideal logical preparation and measurement significantly overestimates the threshold. Threshold values are obtained from code distances $d \in \{11, 13, 15, 17\}$ and T = d stabilizer cycles. (b) Readout error rate (6) for varying integration (solid black) time, threshold curve (solid blue) for $p_G = 0.1\%$, and the linearized threshold curve (dashed blue). The integration time corresponding to the maximum fidelity (τ_R^*) falls outside the threshold, while setting the integration time for the QEC optimum (τ_{QEC}^*) remains deep inside the threshold. Numbers used in this exemplary plot are $\tau_{min} = 1.5 \ \mu s$, $T_{1R} = 1 \ ms$, $\tau_{res} = 3 \ \mu s$, $p_{res} = 0.5\%$, and $T_2 = 50 \ \mu s$.

readout time, we are deep inside the fault-tolerant regime, while the integration time for the maximal readout fidelity leaves so much time for decoherence on the data qubits that the rate of success decreases with increasing system size.

The analysis so far considered an optimistic fixed value for the gate errors. Although gate errors make substantial contribution to the overall performance of error correction. In order to gain a more complete picture of the threshold surface we calculated the threshold surface in the three-dimensional parameter space of (p_G, p_T, p_R) for error biases $\eta_G = 1$ and $\eta_T = 20$. The results are presented on Fig. 3(b). The error threshold for a gate-error-dominated scenario [i.e., $\mathbf{p} \approx$ $(p_G, 0, 0)$] is substantially lower than the threshold values of the opposite limit. Even though the threshold surface does not have the shape of a plane, we may fit a plane to the data points using the three threshold parameters from Eq. (13), to allow for a simple quantitative comparison between QEC codes as well as a useful proxy for resource estimation. The quality of the fit can be described by the mean distance between the surface and the fitted plane, which can be used to estimate the error of the linear approximation (see Table I). Due to the monotonicity of the threshold surface, one expects the strongest bias dependencies in the corners, i.e., for $(0, p_T^{th}(\eta_T), 0)$ and $(p_G^{th}(\eta_G), 0, 0)$. Therefore, we focused our error-bias analysis on these two points.

From Fig. 3(c) it is apparent that the idling threshold p_T^{th} is peaked at $\eta_T = \frac{1}{2}$, that corresponds to depolarizing noise. This can be understood from the fact that for $\eta_T \ll 1$, X and Y errors occur with a probability $p_{X-Y} = p$, both contributing to logical X errors. Similarly, for the Z-biased case ($\eta_T \gg 1$)



FIG. 3. (a) Qubit layout and connectivity map for a bulk section of the rotated surface code (b) error-threshold surface in the parameter space of gate (p_G) , idling (p_T) , and readout error rates (p_R) with biases set to $\eta_T = 20$ and $\eta_G = 1$. Each point is calculated from code distances $d \in \{11, 13, 15, 17\}$ and T = d stabilizer cycles. Green and blue arrows indicate the change of $(p_G, 0, 0)$ and $(0, p_T, 0)$ corner points of the surface for a wide range of noise biases. (c) Dependence of p_G^{th} on the gate error bias η_G and p_T^{th} on the idling error bias η_T .

the probability-*p* errors contribute to logical-*Z* errors. On the other hand, for depolarizing noise, only two of the three Paulis affect a logical with a joint probability 2p/3. Indeed we observe a roughly $1.5 \times$ increase in the error threshold. We note that noise-bias-tailored decoders could take advantage of the peculiar syndrome pattern of biased noise, leading to potentially higher thresholds [22], but such an advantage has not been demonstrated yet for circuit-level error models.

Furthermore, there is a significant increase in gate threshold p_G^{th} for errors biased towards single-qubit gate errors on Fig. 3(c). This can be understood from the stabilizer measurement circuit in Fig. 1(b). Only *X*-stabilizer circuits include single-qubit gates which only contribute to faulty syndromes. Faulty syndromes being equivalent to readout errors have a high threshold.

IV. ERROR CORRECTION BEYOND KITAEV'S SURFACE CODE

Now we turn our discussion towards other members of the surface-code family. In recent years several new candidates appeared that challenge the surface code in terms of error threshold [17], required connectivity [18–20], and prospects for fault-tolerant logical gates [15,16].

Among the codes considered here, there are codes that can be obtained from the Calderbank-Shor-Steane (CSS) construction [38] using only pure X and Z type of Pauli operators, while some other codes include mixed stabilizers and logicals. As we will show, candidates from the latter class tend to perform better against biased idling errors than the CSS counterparts.

Furthermore, we analyzed two Floquet codes, where the stabilizer operators change periodically from one stabilizer round to the next [18,19]. Such a scheme facilitates the measurement of six-body stabilizer operators with very low connectivity (i.e., $2\frac{1}{4}$ two-qubit links per qubit on average).

A. The closest relatives: The XZZX and the 3-CX surface codes

The XZZX code can be simply derived from the rotated surface code, by exchanging X and Z Pauli operators along one of the diagonals for every plaquette such that every plaquette operator, from left to right and top to bottom consist of Pauli operators X, Z, Z, and X [17]. Logical operators need analogous adjustments to maintain the necessary commutation relations. See Appendix B for a more detailed example. The XZZX code requires the same connectivity as the rotated surface code [see Fig. 4(a)], but the local basis transformations necessitate extra single-qubit gates on the data qubits which increase the circuit depth and the error budget of gates. On the other hand, the XZZX code brings significant improvement in terms of idling threshold compared to the surface code as can be seen on the scale of the p_T axis on Fig. 4(b).

From the gate-bias dependence of the respective corner point [shown on Fig. 4(c)] it is clear that the XZZX code can present a real advantage over the surface code only if two-qubit gate errors are more likely. The most remarkable property of the threshold surface is the dependence of the idling threshold on η_T . Since the logicals are not pure Pauli X and Z operators as for the surface code, the maximum is not achieved for depolarizing noise $\eta_T = \frac{1}{2}$ but in the limit where $\eta_T \gg 1$, which we believe to be the relevant limit for spin qubits (i.e., $T_2 \ll T_1$).

The second candidate, the 3-CX surface code, presented recently by Ref. [20] measures surface-code stabilizers in a two-round stabilizer measurement cycle such that stabilizers are measured once per round, but the state of the data qubits only returns to the original state in every second round. In this peculiar measurement sequence, only three of the four connections are used for every data qubit [see Fig. 4(d)], reducing the required connectivity to an effective hexagonal grid. In Appendix C we briefly review the detecting region picture developed by Ref. [20] and provide the stabilizer measurement circuit for the d = 2 instance of the code.

The measurement sequence of the 3-CX surface code requires a modified spatial boundary compared to the rotated surface code [20] that we discuss in Appendix C. Moreover, the fault-tolerant logical readout suggested in Ref. [20] involves simultaneous measurement of all data and ancilla qubits. Although, we only have the hardware to read out half of the qubits in our spin-qubit lattice. In order to overcome this issue, we developed an improved final readout for which the ancilla-only readout pairs of Fig. 1(b) are sufficient.

In QEC codes, reduced connectivity often comes at the expense of a significantly reduced threshold [39], however, Figs. 4(e) and 4(f) show, in correspondence with the findings of Ref. [20], that no significant compromise was made by adapting the stabilizer measurement sequence of the surface code to a lower-connectivity qubit lattice.

B. XYZ² matching code

Hexagonal matching codes are special class of $D(\mathbb{Z}_2)$ anyon models where hexagonal plaquette stabilizers are bicolorable and host different anyon species with the required braiding statistics. Fermionic quasiparticles, combined from two anyons of different species, are confined to string stabilizers that connect the same colored plaquettes [15]. In largerscale lattices the confined-fermion property of matching codes allows for twist-defect-based logical operators without introducing five-body stabilizers as for the surface code.

The XYZ² code is a variant of the hexagonal matching codes where the string stabilizers are parallel links on the hexagonal lattice [16]. Suitable boundary conditions of the code can be found from the concatenation of a two-qubit repetition code (stabilized by ZZ) and the XZZX. In total we get six-body stabilizer operators with X, Y, Z, X, Y, and Z Paulis in clockwise direction around the hexagonal plaquettes, and ZZ link along the XY edges of the hexagonal plaquettes.

Stabilizer measurements in a spin-qubit architecture can be achieved in two rounds such that there is a pair of qubits in the face of each hexagon as shown in Fig. 5(a). The scheduling of CX gates needs to follow similar considerations to the surface-code case in order to prevent the injection of hook errors. In Appendix D we show that there is such a choice of CX schedule for the XYZ^2 code.

The price to be paid for the dense encoding is that there are less ancilla pairs than stabilizers. Meaning that for the final readout of the logical and inference of relevant stabilizers, the code qubits also need to be part of readout pairs. However, in terms of connectivity the layout still remains a regular grid



FIG. 4. (a) Qubit layout and connectivity map for a bulk section of the XZZX code (b) threshold surface for $\eta_T = 20$ and $\eta_G = 1$. (c) Dependence of p_G^{th} on the gate error bias η_G and p_T^{th} on the idling error bias η_T for XZZX code. Each point is calculated from code distances $d \in \{11, 13, 15, 17\}$ and T = d stabilizer cycles. [(d)–(f)] Similarly, qubit layout, threshold surface, and bias dependencies of the 3-CX surface code.



FIG. 5. (a) Qubit layout and connectivity map for a bulk section of the XYZ² matching code. Plaquette stabilizers are *XYZXYZ* Pauli products acting on the the six data qubits around each ancilla pair and *ZZ* link stabilizers are shown with thick blue lines. Data-qubit readout system is required for the final logical readout (see Appendix D for further details). (b), (c) Threshold surface and bias dependencies of the XYZ² matching code for $\eta_T = 20$ and $\eta_G = 1$. Each point is calculated from code distances $d \in \{11, 13, 15, 17\}$ and T = d stabilizer cycles.



FIG. 6. Qubit layout and connectivity map for a bulk section of the Floquet color code. After measuring the red-highlighted XX links, additional plaquette stabilizers are six-qubit Pauli-X stabilizers on *R*-labeled plaquettes, Pauli-Z stabilizers on *G*-labeled plaquettes, and both X and Z stabilizers on *B*-labeled plaquettes (notation is coming from the parent color code). In each time step a set of links is measured such that the six-qubit stabilizer types are exchanges between two plaquette types, indicated by yellow arrows. (b), (c) Threshold surface and bias dependencies of the Floquet color code for $\eta_T = 20$ and $\eta_G = 1$. Each point is calculated from code distances $d \in \{10, 12, 14, 16\}$ and T = d/2 stabilizer cycles (six rounds per cycle).

of qubits (as discussed in detail in Appendix D) realizable, e.g., with $2 \times N$ arrays and long-range links. Since only the spin blockade requires immediate proximity, parallel rows of readout pairs can be accommodated by bilinear arrays and connected via long-range couplers.

The threshold surface of the XYZ² code is closer to a plane than that of the previous surface-code variants [see Fig. 5(b)]. Interestingly, the readout threshold remains comparable to that of the surface code even though the stabilizers are read out in separate measurement rounds. Not being a CSS code, the idling threshold is not peaked around the depolarizing limit and due to different X and Y basis conversions in the plaquette measurements single-qubit gate errors contribute significantly to the error budget, i.e., error threshold biased towards singlequbit errors is not significantly higher than the ones biased towards two-qubit gate errors. These results are presented in Fig. 5(c).

C. Floquet codes

A new type of stabilizer codes has been proposed recently by Hastings and Haah [18], which do not have a static stabilizer group and logical operators, but they change periodically over six rounds. Their proposal was based on Kitaev's honeycomb model (which as a static stabilizer code has no logical qubits). An even more recent example is the Floquet color code. This is a CSS code that can be obtained from a color code using the anyon condensation picture of Ref. [19]. Since the results for both codes are quantitatively very similar, we will present only those for the latter in the main text. We defer a full comparison of the two to Appendix E (see Fig. 18).

A brief summary of the anyon condensation picture can be found in Appendix E, here we only consider the stabilizer group after different measurement rounds. The code can be realized on a hexagonal lattice of data qubits placing an ancilla pair to every edge of the lattice. From an architecture standpoint Floquet codes require a very sparse connectivity, only $2\frac{1}{4}$ links per data qubit leaving valuable space for wire routing and other elements of the control circuitry.

The hexagonal lattice is tricolorable such that every hexagon has neighbors with a different color. We label them as R, G, and B. In every round one measures a set of links that are matching same-labeled hexagons on some basis. For example, starting from the stabilizer group depicted in Fig. 6, where R plaquettes are hosting only a pure Pauli-X operator, G plaquettes only Pauli-Z, and B plaquettes both types: one measures green links (matching G plaquettes) on the Z basis to effectively exchange the roles of R and B while leaving Gintact. Such exchange of roles can be performed in a cyclic manner using differently colored X and Z links only to arrive in the same state after six rounds. We point out that leaving one species of plaquettes with the same stabilizers in each round is crucial for the preservation of the logical information.

In the SWAP-based syndrome measurement scheme we have used so far, one would need two CX gates and a SWAP in-between to read out the link stabilizers. However, another stabilizer measurement protocol can also be employed that avoids using SWAP gates, thereby reducing the gate-error budget of the noise model (see also Appendix A). If the ancilla state is prepared in the $(|00\rangle + |11\rangle)/\sqrt{2}$ Bell state, CXs can target the two ancillas independently, such that the measurement of the ancilla pair remains deterministic. For the results presented in Figs. 6(b) and 6(c) we employed this Bell-state protocol.

The threshold surface of the Floquet color code is very well approximated by a plane [see Fig. 6(b)]. This property could be attributed to (i) the fact that during a link measurement only a single-data-qubit error can be injected, (ii) the lack of fault-tolerant logical initialization and readout. The bias dependencies on Fig. 6(c) of the gate and idling thresholds show the expected behavior for CSS codes.

Even though proposals exist for the spatial boundaries of both types of Floquet codes [19,40], here we only studied these codes with toric spatial boundary conditions for simplicity. Likewise, we omitted the problem of fault-tolerant initialization and readout of the logical qubit since the multiround stabilizer measurement protocol gives a finite (and relatively low) readout-reset threshold.

D. Comparison of QEC codes

As discussed previously, a linearized threshold surface could, through simple formulas, yield valuable insight about the device-specific threshold and the optimal measurement time for midcircuit measurements. Here we focus on the corner points of the linearized threshold surface obtained by fitting a plane to the simulated data. The characterization of the surface with only three numbers aids simple estimates for present and future experiments determining optimal readout parameters for error correction as well as estimating the device-specific error threshold and qubit overhead for different spin-qubit architectures from high to low connectivity.

Our results are summarized in Table I. The three threshold parameters allow us to quantify some trends for spin-qubit devices expected from the QEC literature. For example, error thresholds rapidly decrease by reducing the connectivity, the limiting case being linear qubit chains (connectivity = 2) that are shown to have thresholds $p^{\text{th}} = 10^{-5}-10^{-4}$ for multiple linear QEC codes [39]. Further, we see that gate thresholds are up to an order of magnitude smaller than the other two threshold parameters in agreement with the results on the surface code where thresholds for circuit-level noise are significantly lower than that of simplified noise models.

An example how our results may be utilized is by considering a hypothetical scenario, when good single- and two-qubit gates are available ($p_G \sim 0.1\%$), but the readout is very limited, e.g., during the minimum integration time of the ancilla readout, idling errors on data qubits reach up to a few percent probability. This would exclude the surface code and many more from our comparison in Table I but the XZZX code, featuring high idling threshold parameter, can still be utilized. Having identified the qubit layout to be built, the integration time can be set according to Eq. (14) in order to reach the best error correcting performance.

The QEC codes studied are all defined in a twodimensional qubit layout with local two-qubit connections. This implies that the number of data qubits *n* scales with the square of the code distance, i.e., $n = v_q d^2$ [41]. The total number of qubits required including both data and ancilla qubits $N_{\text{tot}} = (v_q + v_a)d^2$ can also be accessed from Table I. The lowest qubit overhead corresponds to the closest relatives of the rotated surface code coming at $N_{\text{tot}} = 3d^2$. Exploiting the double ancillas required by the spin-blockade readout, the XYZ² code uses only $4d^2$ qubits to measure six-body stabilizer operators as opposed to $6d^2$ of the Floquet codes.

Finally, we make some quantitative comparisons to values found in the literature for the different QEC codes. disregarding the finite idling bias, for the rotated surface code in the presence of isotropic circuit level noise we get $p_{\rm ic} = 0.65\%$ and for the phenomenological noise model $p_{\rm ph} = 3.1\%$ is obtained, which are in line with the expectations [21,22]. For the XZZX code we obtain $p_{\rm ic} = 0.35\%$ and $p_{\rm ph} = 6.9\%$, where the latter is in good agreement with Ref. [17]. Some deviations are to be expected for these threshold values due to the differences in the stabilizer measurement protocol, e.g., the use of pair-wise ancilla readout and the additional SWAP gates.

V. DISCUSSION

Some important characteristic properties of spin-qubit architectures are taken into account in our model and the threshold surface analysis allows one to tailor the results to system-specific features. However, some assumptions on the noise model need further improvement for more accurate device-level threshold estimates. In particular, gate errors as well as readout errors were modeled here as depolarizing noise, which may have tendencies to be more like either bitflip or phase errors in a specific hardware. Also, correlated noise due to crosstalk can be taken into account within the Pauli-error model, but it requires a high-level characterization of the envisioned quantum processor. Finally, some types of errors like coherent errors (systematic rotation of every qubit with the same angle) are not captured in our model and need to be analyzed with separate methods [42].

Restricting the focus regarding the qubit layout and the expected noise model would help to develop noise-tailored decoders [22] with improved performance compared to the one utilized in our work. Fault-tolerant quantum computation in scalable architectures will also require real-time decoding, the performance of which is likely to be compromised by the available time budget. A detailed study is required to identify the optimal decoding strategy which provides the highest thresholds for the given noise model and available decoding time.

Since the Bell-state protocol introduced for the link measurements of Floquet codes leads to a reduced number of SWAP gates, it is tempting to try a similar strategy for all other codes considered. However, note that some of the SWAP gates on the ancillas are necessitated by the specific CX schedule to avoid errors that would reduce the code distance. These SWAP gates are still required in the latter protocol. Furthermore, in the Bell-state protocol, it is the reset errors of the ancillas that would propagate to the data qubits as opposed to gate errors in the SWAP-based method, reducing the readout-reset threshold in the former case.

Our findings can support the development of new QEC codes tailored to spin-qubit architectures. One possibility is to combine XZZX stabilizers with the measurement sequence of the 3-CX surface code to obtain a lower connectivity error correction code with high idling threshold error rate at strongly biased noise. Floquet codes also provide promising prospects for future codes with low connectivity.

Finally, the best error correction code will be the one that provides the lowest qubit overhead en route to fault-tolerant quantum computing. Beyond the quantum memory experiments considered here, fault-tolerant logical gate schemes using twist defects or lattice surgery [14] also need to be revisited under concrete noise models to find the lowest space-time overhead for a given device design.

For the Clifford-circuit simulations we used STIM [43]. For each code distance and physical error rate we took up to 300 000 shots unless 30 000 logical failures are encountered before that. For the decoding we used PYMATCHING [13].

VI. CONCLUSION

Taking into account the common features of spin-qubit platforms we have derived an error model accounting for different error rates for gate and readout errors as well as decoherence during midcircuit measurements. This helped us to quantify the tradeoff between fast and accurate qubit measurements for error correction applications. We derived a formula for the optimal integration time that can be used for the calibration of the qubit readout given the noise parameters of the device and QEC code of interest.

Considering state-of-the-art error correction codes that are compatible with locally connected 2D architectures, we proposed four different qubit layouts required for quantum memory experiments which offer the spin-qubit community different options to find balance between experimental feasibility and low-overhead execution of QEC codes. Furthermore, we analyzed the threshold surface in a multidimensional parameter space facilitating back-of-the-envelope estimates for the error threshold and qubit overhead involving on the gate and readout fidelities as well as the decoherence rates for a given experimental setup.

All the scripts used for the threshold calculations as well as the plotted data are available at [44].

ACKNOWLEDGMENTS

We thank B. Srivastava and the spin-qubit team at IBM Research, Zurich, for the useful discussions, B. J. Brown for providing access to the script on the Floquet color code, and J. Asbóth for pointing out some relevant references. The authors acknowledge support from the NCCR SPIN, a National Centre of Competence in Research, funded by the Swiss National Science Foundation (Grant No. 51NF40-180604).

APPENDIX A: SPIN-QUBIT READOUT AND STABILIZER MEASUREMENTS

In order to discuss the prospects of stabilizer measurements with spin qubits we turn to the example of the Z plaquettes of the surface code and present three options for performing stabilizer measurements using an ancilla pair instead of the conventional single ancilla that can be read out directly. Excluding readout protocols that require a lead coupled to the quantum dot hosting the qubit, we focus on readout via Pauli spin blockade.

In the Pauli spin blockade, the two-qubit subspace corresponds to the four low-energy two-particle eigenstates of a symmetric double quantum dot with one particle residing in each dot. When the energy of one of the dots is increased, a particle tunneling from the higher-energy dot into the lowerenergy one is only allowed into the spin-singlet state (triplets involve a higher-energy orbital due to the exclusion principle). This protocol converts one of the states into a doubly occupied charge state that is measurable via a charge sensor. The rest of the states can only end up in the measurable charge configuration by relaxation [33].

The above-described effect can be utilized for readout in two different regimes: (i) destructive parity measurement, or (ii) single-state discrimination [36,37]. In the destructive parity measurement protocol, one exploits the fact that one of the three blocked states relaxes much faster than the others, and thereby the charge measurement distinguishes two of the basis states from the rest of the states. By destructive measurement we mean that the entanglement is lost within a subspace, e.g., $a|00\rangle + b|11\rangle \rightarrow |00\rangle$ after observing an even outcome. We model this measurement protocol as a ZZ parity measurement followed by an initialization of the qubits. Single-state discrimination refers to the case when the three blocked states remain in the original charge configuration and distinguished from the fourth state. For simplicity we take the discriminated state to be the $|11\rangle$ state and thereby the readout is effectively performed on the eigenbasis of the CZ operator.

The main limitation of pairwise readout is that only one bit of information can be obtained from the two-qubit subspace. In the following we present three options to overcome this limitation and leverage the presence of the additional ancilla qubit.

The most straightforward way to overcome the limitations of the pairwise readout mechanisms of spin qubits is to utilize one of the two qubits as an ancilla in the usual sense and preserve the other one in the $|0\rangle$ state as a reference qubit for the readout as shown in Fig. 7 (top) [2]. For most qubit layouts with low connectivity ($g \leq 4$) this necessitates additional SWAP gates between the ancillas, which in turn increases the susceptibility of the syndrome measurement to gate errors. However, we note that this protocol can be performed using *CZ* measurements and is also straightforwardly extendable to the case when a single particle is occupying the two dots and read out in the circuit QED protocol [3].

In order to reduce the number of additional SWAP gates in the measurement protocol, one could prepare the ancilla pair in the Bell state

$$|\psi\rangle = (|00\rangle + |11\rangle)/\sqrt{2},\tag{A1}$$

and map the two halves of the plaquette operator to the two ancillas separately. As shown in Fig. 7 (middle) this protocol allows for stabilizer measurements without SWAP gates on the ancillas. One might expect that the same protocol would work with the preparation of the ancilla pair in the $|00\rangle$ state, but such a protocol is not fault tolerant as we will see later when discussing the measurements in the detecting region picture. The main limitation of the Bell-state approach is the fact that ancilla-initialization errors create two-qubit errors which typically significantly reduce the readout threshold. Furthermore, SWAP gates are still necessary in either *X*- or *Z*-type stabilizer measurements to avoid hook-error injection (discussed in Appendix B).

Our third option of measuring stabilizers is relying on the concept of flag qubits [45]. Using an additional qubit as a flag qubit one can acquire additional information from the success of the stabilizer measurement process by monitoring Z errors in the part of the circuit where they can lead to hook errors. This flagging procedure needs to happen in the middle of the circuit by inserting CNOT gates between the flag and the ancilla qubits. Since we need to extract two bits of information from the ancilla pair, we need three subsequent (nondestructive) CZ



FIG. 7. Alternative stabilizer measurement protocols with two ancilla qubits and pairwise readout. Top: SWAP-based protocol with destructive parity measurement (ZZ) which is replaceable by single-state discrimination (CZ). Middle: Bell-state measurement protocol reducing the SWAP gate count. Bottom: second ancilla is utilized as a flag qubit requiring subsequent measurements with single-state discrimination. Overlay: detecting regions expanding from the reset-measurement pair, blue and green denote Z-type regions, red denotes an X-type region.

measurements to distinguish the stabilizer and flag outcomes [see Fig. 7 (bottom)]. Additionally, the flagged region cannot include SWAP gates because two-qubit error on the ancillas can induce a hook error and mute the flag at the same time. Due to these limitations and the increased circuit depth, we did not include the flag-qubit approach in our numerical analysis.

The syndrome measurement protocols can also be viewed in the detecting region picture (see Appendix C or Ref. [20] for further information). If we start from the $|00\rangle$ state, two detecting regions must appear, one corresponding to Z_{a0} and one to Z_{a1} (or any combination of these two). Let us consider a circuit without SWAPgates. Working with Z_{a0} and $Z_{a0}Z_{a1}$ regions, the $Z_{a0}Z_{a1}$ region can be used to extend a four-body stabilizer as the blue region in the top right of Fig. 7, but the region Z_{a0} will expand only to the qubits on the upper end leading to a semiplaquette stabilizer which does not commute with some other bulk stabilizers. On the other hand, working with a Bell pair implies the two ancilla stabilizers $X_{a0}X_{a1}$ and $Z_{a0}Z_{a1}$, where the X region will not be able to expand and therefore leads to the desired expansion of a plaquette stabilizer without any side effects.

Similarly the SWAP syndrome measurement can be understood as a way of trying to avoid the expansion of the second stabilizer Z_{a1} , by swapping it with the other ancilla and not letting it participate in a two-qubit operation with the code



FIG. 8. Connectivity and schedule required for the simultaneous plaquette stabilizer measurements for a distance-3 surface code. Light (dark) gray nodes represent the code qubits (ancillas). Red (blue) circles denote qubits that participate in the X (Z) logical operator with the corresponding single-qubit gate. X (Z) plaquettes are shown in red (blue). For the final readout of the logical observable, an additional ancilla pair is required which can be placed next to qubit 0 or 8.

qubits. The flag qubit approach extends the former concept by utilizing the second region to gather some error information from the syndrome ancilla during the circuit. This is achieved by expanding an X region to the syndrome ancilla which cannot expand to the code qubits but is sensitive to those Z errors that would create a hook error by spreading to two of the code qubits.

APPENDIX B: SURFACE CODE AND THE XZZX CODE

Here we describe the properties of the surface code and the XZZX code and discuss some details of the numerical calculation. The surface code with the boundary conditions used in this work [12] is often referred to as the rotated rotated surface code which is a $[[d^2, 1, d]]$ CSS stabilizer code [38]. The XZZX code, on the other hand, is a variant of the surface code with locally rotated stabilizers featuring exceptionally high threshold against biased noise [17].

Improvements can be made in both cases, such as using a tailored decoder for biased noise for the surface code improving idling thresholds for high- η_T bias [22], or choosing the boundary conditions of the XZZX code such that the high-rate errors have a longer path to logical failure [17]. In this work, however, we did not explore such possibilities as they would have made the fair comparison of different codes difficult.

1. Stabilizers and logical operators

For both the surface code and the XZZX code the stabilizers are four- and two-body operators that can be divided into two groups as shown in red and blue on Fig. 8. For the regular surface code red (blue) plaquettes consist of only Pauli X (Pauli Z) while for the XZZX code each type of plaquette contains two X and two Z Paulis in the same orientation, e.g., $X_1Z_4Z_0X_3$ in the notation of Fig. 8.

Logical operators are shown in Fig. 8 for the d = 3 surface code. Note that the X(Z) logical share a single qubit with only one two-body X(Z) plaquette on the boundary and two qubits

with (d - 1)/2 boundary plaquettes of the *Z* (*X*) type. Analogously, the logical *X* and *Z* operators in the XZZX code would be $\mathcal{X}^{(3)} = X_0 Z_1 X_2$ and $\mathcal{Z}^{(3)} = Z_0 X_3 Z_6$ in the d = 3 case.

2. Stabilizer measurements and CNOT scheduling

The order in which we apply the CNOT gates is important for two reasons: (i) plaquette measurements have to commute and therefore if two plaquettes share two qubits, the schedule has to be such that the first plaquette gathers the parity information of the two qubits before a CNOT gate of the second plaquette would act on either of the qubits; (ii) the CNOT schedule of differently colored plaquettes must take the effect of hook errors into account [11].

In the surface code a hook error is a pair of Pauli-X(Z) errors propagating from a faulty ancilla to the code qubits. In the case of X(Z) plaquettes this can reduce the code distance of the logical X(Z) operator depending on which qubits are affected [11]. In order to avoid injection of hook errors, the schedule of the CNOT gates is chosen according to Fig. 8. Even though pairs of Pauli errors can still be injected by faulty ancillas, it still takes (d+1)/2 error events, to cause a logical error, as opposed to $\lceil (d+1)/4 \rceil$ for the schedule that induces hook errors.

The schedule displayed in Fig. 8 satisfies both the commutation and the hook-error criteria. Note, however, that the layout does not favor the schedule for the red plaquettes. SWAP gates need to be applied on the respective ancilla pairs after round 0 and round 2 (which was not taken into account in Ref. [2]). Since SWAP gates on the ancillas are inevitable due to the connectivity and the CNOT schedule, we used the SWAPbased protocol presented in Appendix A. The Z-plaquette measurement is shown in the top left of Fig. 7. This only merely reduced the gate-error threshold, but allowed for a significant improvement for the readout threshold compared to the Bell-state approach.

3. Fault-tolerant logical initialization and readout

Here we briefly discuss the protocol for the fault-tolerant initialization and readout of an eigenstate of the logical Z operator of the surface code. The X observable of the surface code can be obtained by applying Hadamard gates on every code qubit in the beginning and before the final measurement, whereas the same protocol applies for the observables of the XZZX code with appropriate local basis transformations.

The key observation is that only X and Y errors can flip the eigenvalue of the logical Z operator, that are both detected by Z plaquettes. Therefore, it suffices to know the eigenvalue of every Z plaquette in the first (last) round to ensure fault tolerance during the initialization (readout) of the logical observable. If every qubit is initialized in the $|0\rangle$ state in the beginning, the eigenvalue of the Z logical is +1 and every Z plaquette has eigenvalue +1. This means that one can already detect X and Y errors in the first round of stabilizer measurements by considering the Z plaquettes only. X plaquettes, on the other hand, will have nondeterministic outcome in the first round, so they can only be compared with the previous round starting from the second measurement round.



FIG. 9. Connectivity and schedule required for the simultaneous plaquette stabilizer measurements for the 3-CX surface code (distance 3). Light (dark) gray nodes represent the code qubits (ancillas). Red (blue) circles denote qubits that participate in the X (Z) logical operator with the corresponding single-qubit gate. Red lines, representing CNOT gates in round 0, are covered by the purple lines in most cases.

The final measurement in the Z basis can be performed on a similar footing. After the last round of stabilizer measurements, every qubit needs to be measured on the Z basis, from which one can infer the logical eigenvalue as well as every stabilizer eigenvalue to see if an error had happened right after the last round of stabilizer measurements that flipped the logical observable. For our spin-qubit lattice only ancillas can be read out directly, therefore, we need to swap every qubit with one of the nearby ancillas and measure the ancilla pairs afterwards. Note that in the case of the surface code, the final measurement requires only one extra pair of ancillas regardless of code distance.

The fault-tolerant initialization and readout of the logical observable is crucial not only for fault-tolerant computation on multiple logical qubits, but also in quantum memory experiments with a single logical qubit. If initialization and the final readout are assumed to be perfect, no logical error can appear on the logical observable in the presence of measurement errors only. This limit is clearly nonphysical and has significant impact on the overall error threshold if the errors are biased towards readout errors as shown in Fig. 2(a) for the case of the surface code.

APPENDIX C: THE 3-CX SURFACE CODE

The 3-CX surface code is a recently proposed variant of the surface code that relies on the concept of detecting regions in order to alleviate the connectivity requirements of the surface code [20]. The spin-qubit connectivity graph is shown in Fig. 9, where the hexagons and the squares of the surface-code case are merged by removing one link per data qubit in the bulk.

1. Stabilizers and logical operators

The stabilizer measurements are done in two half-cycles, where the CNOT schedule of the first half-cycle is shown on Fig. 9. The second half-cycle uses the reversed schedule of the first one. Each of the half-cycles measure $d^2 - 1$ stabilizers.

These are indeed the stabilizers of the surface code, but the X and Z type of bulk stabilizers are exchanged between the two half-cycles and the boundary stabilizers are adapting correspondingly (see Figs. 8.1 and 8.7 in Ref. [20]). This choice also allows one to define the logical operators in the same way as for the rotated surface code.

Another important difference is the boundary of the physical qubit lattice, which is different from that of the rotated surface code, in order to preserve the code distance [20]. A simple way to see this is trying to extend the structure in Fig. 9 downwards or to the left while preserving the connectivity pattern. This would mean adding ancilla pairs that are only connected to a single data qubit, preventing them from measuring multibody stabilizers that are essential to protect the logical qubit.

2. Stabilizer measurements in the detecting region picture

A detecting region is a closed region of the circuit diagram originating from a reset (or set of resets) and terminating on a measurement (or set of measurements). A pure X(Z) detecting region, one that only contains CNOT gates, can be defined by the property that a single Pauli Y or Z(X or Y) error changes the total parity of the involved measurements.

Detecting regions can be terminated in multiple ways on later measurements. This arbitrariness is the consequence of the fact that detecting regions at a given time are related to the generators of the space-time stabilizer group at the corresponding time. As a simple example consider three qubits, where the parity of qubits q_1 and q_2 is measured via the ancilla a. The two data qubits are initially in the eigenstate of the Z_1Z_2 parity operator, which is the only stabilizer in the space-time stabilizer group at t = 0. Resetting the ancilla creates a single Z stabilizer Z_a that commutes with the parity operator and therefore gets admitted in the stabilizer group at t = 1 [46]. The choice of stabilizer generators is ambiguous, we can use, e.g., $\{Z_a, Z_1Z_2\}$ or $\{Z_a, Z_aZ_1Z_2\}$. We then apply a CNOT, controlled on q_1 and targeted on a, which extends the ancilla stabilizer as $Z_a \rightarrow Z_a Z_1$. For the two, mathematically equivalent, sets of stabilizer generators we get $\{Z_a, Z_1Z_2\} \rightarrow$ $\{Z_a Z_1, Z_1 Z_2\}$ and $\{Z_a, Z_a Z_1 Z_2\} \rightarrow \{Z_a Z_1, Z_a Z_2\}$. One can regard a specific choice of generators as detecting regions. In the latter case the second detecting region is said to have contracted since its weight is reduced by one. Sticking to this example, after yet another CNOT between q_2 and a we get $\{Z_a Z_1 Z_2, Z_a\}$. Finally, measuring the ancilla removes Z_a from the group as an element as well as from every element that contains it. Note that starting from a certain stabilizer Z_1Z_2 we contracted the corresponding detecting region and measured the parity. At the same time we expanded another region which took over the place of the original parity operator. Having a large number of qubits and stabilizers, the ambiguity of assigning resets to detecting regions creates several options for the tiling of detecting regions. One general design principle can be to maintain the \mathbb{Z}_2 anyon character of regions by ensuring that every qubit (in the bulk of the lattice) at every time is covered by a pair of X and a pair of Z detecting regions [20].

We have collected some effective rules for the construction of detecting regions in the case of the 3-CX surface code: (i) Z- (X-) type regions can only originate (terminate) on Z-basis (X-basis) resets and measurements.

(ii) If a Z- (X-) type region is incident only on the target (control) qubit of a CNOT gate, it expands to the control (target) qubit.

(iii) If the same Z- (X-) type region is incident on both the control and the target qubit of a CNOT gate the region contracts to the target (control) qubit.

For further information and rigor about detecting regions we refer the reader to Refs. [20,46].

Between every measurement and reset the detecting regions represent the stabilizers of the surface code. The type of the boundary stabilizers does not change so the logical operators remain the same throughout the circuit. In the example of Fig. 10 the circuit of the d = 2 version is shown at full width, using single (individually measurable) ancillas. We can see this as having a four-body X plaquette and two two-body Z plaquettes after the first measurement, but after the second measurement we see a four-body Z plaquette expanding that will be the subject of our final measurement. The two-body X plaquettes are not shown, but using the rules above they can be straightforwardly identified as being located on in q_0q_1 and q_2q_3 edges, respectively.

3. Fault-tolerant initialization and readout of the logical qubit

For the initialization we can follow directly Ref. [47], but we need to modify the final measurement part since the solution provided by McEwen *et al.* measures every qubit (both data and ancilla) in the half-cycle state. This is not possible in spin-qubit architectures since qubits can only be measured pairwise, and only ancilla types have reference qubits for the measurements. In the following we present a scheme for the final measurement round in which only the data qubits need to be measured. For spin qubits this can be done in the same way as for the rotated surface code, by swapping one of the ancillas from each pair with a neighboring data qubit and measuring the ancilla-data pair.

As described in Ref. [20] each detecting region has to be terminated on an appropriate measurement. However, as for the conventional rotated surface code, only those stabilizers play a role in the decoding which are sensitive to errors that affect the outcome of the logical measurement. Consequently, one can measure every data qubit in the Z basis right after the last round of syndrome measurements as in the upper part of Fig. 10. From this the Z eigenvalue of the logical qubit can be inferred and every Z-type detecting region is terminated. The syndromes obtained for each detecting region can be compared to the appropriate values of the last syndrome detection round the same way as if the final measurement would have been round (T+1).

For comparison the final measurement scheme of Ref. [47] is shown in the lower part of Fig. 10. After the last two CNOT layers, the stabilizer group is extended with a trivial element Z_{a1} and a nontrivial element $Z_{a0}Z_{q_0}Z_{q_3}$ necessitating the measurement of ancillas a_0 and a_1 . In this scheme in total $2d^2 - 1 - \lceil (d-1)/2 \rceil$ measurements are required [of which only O(d) are trivial], but in the protocol we used in this work only d^2 simultaneous measurements are needed in the final round.



FIG. 10. Measurement circuit of the detecting regions and final measurement of the 3-CX surface code for d = 2 with the required connectivity (top right). The two distinct detection rounds ending with ancilla measurements are alternating in the detection cycles. Top figure: an X-type detecting region in a regular detection round (red), and the final measurement of the last Z-type detecting region (blue) with d^2 measurements that is compatible with spin qubits. Bottom figure: two Z-type regions in a regular detection round (green and purple), and final detection round with three regions (blue and green) terminated on $\sim 2d^2$ measurements as in Ref. [47]. Dashed lines separate different layers of CNOT gates, simplifying the comparison with the time-slice picture of Ref. [20].

APPENDIX D: THE XYZ² MATCHING CODE

The XYZ² code is a $[[2d^2, 1, d]]$ stabilizer code [16] defined on a hexagonal lattice. This code is a special case of the hexagonal matching code family [15], with two-body link operators and appropriate boundary conditions. The connectivity required for this code can be embedded in a square grid of spin qubits, where ancilla pairs are located in the middle of each hexagon, giving the highest encoding density ~1 qubit or ancilla among the considered error correction codes.

1. Stabilizers and logical operators

In the hexagonal matching code there are two types of stabilizers, plaquettes, and links. There are d^2 link operators which are products of Pauli-Z operators for two neighboring qubits, where the neighbors are to be taken in the same direction for every pair (see dashed green lines on Fig. 11). Plaquettes in the bulk are six-body Pauli operators of alternating type, i.e., starting with a Pauli Z on a qubit of the hexagon where a link is incident, followed by X, Y, Z, X, and Y operators in a clockwise direction along the hexagon. In total there are $(d - 1)^2$ bulk plaquettes (blue and red hexagons on Fig. 11). At the boundary of the lattice one finds 2d - 2 semiplaquettes which are three-body Pauli operators defined by the same rule as their bulk counterparts [16].

As in the case of the surface code, plaquette operators can be divided into two groups as shown in Fig. 11. Such a division will be useful for the CNOT scheduling of plaquette measurements as well as for the matching decoder developed for this code.

Logical operators are defined on the edges of the diamondshaped boundary and contain both X, Y, and Z Paulis as shown on Fig. 11. This choice differs from that of Ref. [16] only by stabilizer operators. The motivation behind defining them on the boundary comes from the similarity to the surface code which allows one to perform the decoding in an analogous way.

2. Stabilizer measurements and CNOT scheduling

Measurements of plaquette stabilizers can be performed via a pair of ancillas sitting in the middle of each plaquette as in Fig. 11, resulting in a connectivity requirement of a simple square lattice. Link measurements reuse the ancilla pairs of the plaquettes and therefore performed in separate measurement rounds, after the plaquette measurements.

The specific choice for the layout in Fig. 11 ensures that link stabilizers can be measured without using any SWAP gates. On the other hand, plaquette measurements do require SWAP



FIG. 11. Connectivity and schedule required for the simultaneous plaquette stabilizer measurements for a distance-3 XYZ^2 matching code. Light (dark) gray nodes represent the code qubits (ancillas). Green dashed line shows the qubit pairs along which the link stabilizers are measured. Red (blue) circles denote qubits that participate in the Y (Z) logical operator with the corresponding single-qubit gate.



FIG. 12. Top (bottom) figure presents the stabilizer circuit of the blue (red) plaquettes of Fig. 11 requiring one (three) SWAP gates on the ancilla pairs. The labeling of the data and ancilla qubits starts with the qubit that is acted on in round 0 for both types of plaquettes, and continues clockwise.

gates (one for blue plaquettes three for reds) as shown on Fig. 12 and a CNOT schedule that allows for neighboring plaquette measurements to commute and avoids injecting hook errors. In this work the schedule was chosen according to Fig. 11. Further details on the error chains created by single error events and the corresponding syndromes will be discussed in the next section.

3. Decoding

Even with the careful scheduling of the CNOT gates, single errors on the ancillas can lead to syndromes with up to five detection events. For example, a *Y* error on the ancilla (not the reference qubit) flips the measurement outcome, which flips back again in the next round (2-detection events). If such an error happens after round 1 or round 3 it induces Pauli errors

on two data qubits which change the eigenvalue of one link and two plaquette stabilizers (see examples in Fig. 13).

However, same-colored plaquettes, combined with every link stabilizer, can be gathered into subgraphs and decoded separately. For a given subgraph, the single-error syndromes become analogous to the case of the surface code, with each syndrome containing at most four detection events. This is the direct consequence of the matching code family being a $D(\mathbb{Z}_2)$ anyon model [15], where the plaquette operators host anyons of the corresponding color and links host fermions, i.e., a pair of red and blue anyons. When we separate the two subgraphs, the link fermions are divided into two anyons.

Four-body detection events can be decomposed into a timelike pair, i.e., measurement error, and a spacelike pair, i.e., data qubit error or error chain. Note that spacelike pairs do not necessarily appear at the same time, due to the scheduling of plaquette measurements and the separately measured links. Nevertheless, timelike pairs can be separated easily since they are always local in space.

In the decoding graph nodes correspond to possible detection events (i.e., comparison of subsequent stabilizer measurements) and edges connect two events that can appear simultaneously as a result of a single error. We can take advantage of the decomposition of timelike and spacelike syndrome pairs by assigning a weight to every edge based on the joint probability p of all the single-qubit errors that result in the corresponding pair of syndromes. The weights are calculated from the probabilities as $w = \ln[p/(1-p)]$. Given a weighted decoding graph, a perfect matching decoder can be used to find an error configuration that is consistent with the syndromes.

In this work we assigned the error probabilities in the following way:

$$p_{Z}^{1q} = \sum_{i+j+k+l+m \in \text{odd}} {\binom{2}{i}} p_{T2}^{i} (1-p_{T2})^{2-i} {\binom{4}{j}} (p_{G1}/3)^{j} (1-p_{G1}/3)^{4-j} {\binom{4}{k}} (4p_{G2}/15)^{k} (1-4p_{G2}/15)^{4-k} \times {\binom{N_{2}}{l}} (8p_{G2}/15)^{l} (1-8p_{G2}/15)^{N_{2}-l} (4p_{G2}/15)^{m} (1-4p_{G2}/15)^{1-m}$$
(D1)

is the probability of a Z error on a single qubit since every qubit is subject to two idling phase during measurements, four single-qubit gates (from two plaquettes), and four two-qubit gates (three from plaquettes, one from a link), and finally a two-qubit gate error inducing a Y or Z ancilla error that propagates to the qubit as a Z error (can happen via one of the plaquettes for every qubit or via the link for every second qubit). In the second row N_2 denotes the total number of two-qubit gates in round 0 and round 4 of the plaquette measurement, i.e., $N_2 = 2$ for red plaquettes $N_2 = 4$ for blues (see also in Fig. 12). A Pauli-Z error on a data qubit creates a pair of detection events on two neighboring plaquettes of the same type.

Similarly the probability of an *X* or *Y* error is

$$p_{XY}^{1q} = \sum_{i+j+k \in \text{odd}} \binom{2}{i} p_{T1}^{i} (1-p_{T1})^{2-i} \binom{4}{j} (2p_{G1}/3)^{j} (1-2p_{G1}/3)^{4-j} \binom{4}{k} (8p_{G2}/15)^{k} (1-8p_{G2}/15)^{4-k}.$$
(D2)

Such an error induces detection events on a link and on two plaquettes of different types.

Furthermore, the probability of a timelike error, i.e., measurement error, is

$$p^{m} = \sum_{i+j+k \in \text{odd}} p_{R}^{i} (1-p_{R})^{1-i} {\binom{2}{j}} (8p_{R}/15)^{j} (1-8p_{R}/15)^{2-j} {\binom{N_{2}}{k}} (8p_{G2}/15)^{k} (1-8p_{G2}/15)^{N_{2}-k},$$
(D3)



FIG. 13. Schematic figure of a d = 5 matching code with the scheduling chosen according to Fig. 11. *Y* and *Z* errors on the ancilla qubit lead to one-, two-, and three-qubit errors. These error chains correspond to pairs of detection events in the two sublattices. Two- and three-qubit errors propagate parallel to the respective edge, therefore cannot induce logical errors. Green *X*, *Y*, and *Z* denote the type of error occurring on the data qubits. Plaquette detection events on the two different types of plaquettes are shown as blue and red nodes, while link detection events are shown as a pair of blue and red nodes, referring to their role in both decoding problems. Gray nodes show stabilizers where no error was detected.

where the first term is the readout assignment error and the second corresponds to an error in the measurement or the initialization of the qubit. This time N_2 is the number of all two-qubit gates ($N_2 = 7$ for red plaquettes and $N_2 = 9$ for blue plaquettes and $N_2 = 2$ for links). The set of errors that occur with this probability create two detection events at the same plaquette or link in two subsequent measurement cycles.

Finally, the probability of a faulty ancilla injecting two or three Pauli errors is more straightforward to calculate as there are only two possibilities for an ancilla to pick up a Y or Z error (after round 1 or round 3 for the two-qubit case and during or after round 2 for the three-qubit case). Therefore, we get

$$p^{2q} = p^{3q} = 2(8p_{G2}/15)(1 - 8p_{G2}/15)$$
 (D4)

for the two- and three-qubit error-chain probabilities. The syndromes induced by such error chains are less straightforward, therefore, we summarize them in Fig. 13. Importantly none of these reduce the code distance and both of them create a pair of (relevant) detection events.

Close to the boundaries, single detection events can appear that must be connected to a boundary node. There are in total four boundary nodes, two for each type of plaquette. If the single detection event is associated to that boundary where one of the logical operators is defined, the graph edge connecting the single node to the boundary node will also imply a logical error.

4. Fault-tolerant initialization and readout of the logical qubit

The hexagonal matching code is the only candidate in our work that utilizes one qubit per ancilla for the encoding. Due to the additional constraint on the readout (pairwise, destructive), it is a nontrivial task to initialize and read out the logical qubit while inferring sufficiently many stabilizer eigenval-



FIG. 14. Full connectivity map, extended with d ancillas (labeled 37 and 38) for the logical initialization and readout. Qubits are divided into overlapping groups of five, of which two pairs of qubits can be read out simultaneously (denoted by thick light and dark gray links).

ues to maintain fault tolerance. Here we describe in detail the scheme for the logical initialization and readout used in this work.

Since every qubit is participating in one link only, one shall consider all the links and distribute them into two groups such that red (blue) links are the ones matching two nonneighboring blue (red) plaquettes on their corners. Regarding the red (blue) logical, i.e., logical Y(Z), the relevant stabilizer eigenvalues for decoding are the plaquettes colored red (blue) on Fig. 11 as well as every link. To initialize or measure all these stabilizer eigenvalues and the logical qubit on the Y(Z)basis, we initialize or measure red (blue) links on the XY - YXbasis and blue (red) links on the ZI-IZ basis. Knowing both m_{XY} and m_{YX} eigenvalues at a given step (initial or final), the eigenvalue of the corresponding Z link is just the product of the two, i.e., $m_{ZZ} = m_{XY}m_{YX}$ (similarly for the ZI-IZ case) providing every link stabilizer eigenvalue. Furthermore, the relevant plaquette eigenvalues can be constructed from products of four known eigenvalues (m_{XY} and m_{YX} on the neighboring links as well as m_{ZI} and m_{IZ} on the incident links). Finally, one can easily check that the logical eigenvalue is also determined by the appropriate product of the aforementioned link eigenvalues (see Fig. 14).

In order to perform the required *XY-YX* and *ZI-IZ* link measurements, we append *d* additional ancillas and regroup the data and ancilla qubits as shown in Fig. 14. Despite the different shapes (triangle or tetragon) and types (red or blue), every group has at least the following connectivity: two qubits, q_0 and q_1 , are connected to the same ancilla a_0 which can be read out or initialized together with another ancilla a_1 ; a third qubit q'_0 (that is in state $|0\rangle$ or acts as q_0 in a neighboring group) can be read out or initialized together with q_1 . Figures 15 and 16 show the circuits for the initialization and readout of such link operators.

APPENDIX E: FLOQUET CODES

Floquet codes on a honeycomb lattice can be derived from the color code in the anyon condensation picture [19]. The color code is defined on the honeycomb lattice, where the vertices correspond to the data qubits, and the stabilizers are six-body X and Z plaquette operators on the hexagons that are divided into three color groups such that each hexagon is



FIG. 15. Initialization of the XY - YX link operator in the mutual +1 eigenstate. Initialization of the ZI - IZ link is trivial after initialization of every qubit pair in the $|00\rangle$ state.

surrounded by contrasting colors. In this model nine bosonic excitations can be defined labeled as red, green, blue and x, y, z. These bosons have nontrivial braiding statistics with each other if they have different color and Pauli labels, thus, we refer to them as *anyons*.

In the Floquet codes under consideration, the nine bosons of the color code are condensed into two sets of bosons that braid nontrivially with one another [19]. The resulting code is therefore part of the surface code family, i.e., a $D(\mathbb{Z}_2)$ anyon model. By periodically changing which anyon is condensed, an arbitrary single-qubit error can be detected on the honey-comb lattice while the logical qubit remains well defined. The change in the anyon condensation should always be done in a way that one of the *e* (*m*) anyons from the old code remains an *e* (*m*) anyon in the new condensed code.

The main difference between the two Floquet codes considered in this work, the Floquet color code [19] and the honeycomb Floquet code [18], is which of the condensed phases are used and how they are transformed to one another in subsequent measurement rounds. A noteworthy difference is that the Floquet color code has been constructed such that it is a CSS code.



FIG. 16. Top: final readout circuit for the XY-ZZ (or equivalently XY-YX) link measurement. Bottom: ZI-IZ link measurement. The SWAP gate between q_0 and a_0 in both circuits ensures that q'_0 (i.e., q_0 of the neighboring group) is in the $|0\rangle$ state before the pairwise readout.



FIG. 17. Connectivity required by the Floquet codes (d = 2), with tricoloring of the hexagons and the links according to the color code. Convention X: red, Y: green, Z: blue.

1. Stabilizers and logical operators

As discussed above, on the abstract level a condensed color code at a given time is equivalent to a surface code. However, on the microscopic level the relation between detection cells (detection cells in the terminology of Ref. [19] are detecting regions where the stabilizer measurements are considered as a single step) and anyons is different. This is due to the fact that the condensed color code has two different Pauli stabilizers on a given plaquette, as opposed to the $D(\mathbb{Z}_2)$ anyon models we studied before. Stabilizer measurements in the Floquet code are performed by measuring a set of XX, YY, or ZZ links (in the Floquet color code only XX and ZZ) that are connecting same colored plaquettes in subsequent measurement rounds. The links can be colored as well according to the plaquettes they connect.

Let us consider the point in time when X links are measured on the red edges (see Fig. 17). Each link will anticommute with the red Z plaquettes, which we therefore must have measured in the previous round to ensure fault tolerance (every region lives for four rounds in the six-round-long measurement cycle). The link measurements allow us to infer the green and blue X plaquettes, which marks the end of the green X detection cells and the beginning of a blue X cells. In total we have four detecting regions after the red X link measurements: blue X, blue Z, red X, and green Z. If a Z error happens somewhere a red and a green X region will report it. Similarly, an X error shows up on green and blue Z regions. Therefore, we can identify red and a blue X plaquettes with nontrivial outcome in this round with e anyons, while red and green Z plaquettes will host m anyons.

Logical operators also transform together with the stabilizer group. The transformation of the logicals can be viewed as follows: having a logical operator defined in a given phase, e.g., a string of Pauli operators, one can define an equivalent logical operator by multiplying it with all the link stabilizers that have support on the logical, thereby shifting the logical or changing its Pauli type. One can choose the next phase such that it is compatible with the one of the equivalent versions of the logical operators ensuring the conservation of the logical subspace [19].



FIG. 18. (a) Threshold surface of the Floquet color code for $\eta_T = 20$ and $\eta_G = 1$. (b) Dependence of p_G^{th} on the gate error bias η_G and p_T^{th} on the idling error bias η_T for Floquet color code. Each point is calculated from code distances $d \in \{10, 12, 14, 16\}$ and T = d/2 stabilizer cycles (three rounds per cycle). (c), (d) Similarly, threshold surface and bias dependencies of the honeycomb Floquet code.

2. Details for the spin-qubit version

For both of the Floquet codes we considered periodic spatial and open temporal boundary conditions. Since the plaquette stabilizers are not measured in a single round, the effectively perfect logical initialization and measurement does not lead to such a side effect as for the static stabilizer codes. For both versions given a code distance d we performed d/2 measurement cycles, that is, 3d measurement rounds.

Furthermore, we used the Bell-state protocol of Appendix A for the link measurements. Since the link measurements use two CNOT gates, removing the SWAP gate from the ancillas considerably increases the gate error threshold (from about 0.35% to ~0.45%). On the other hand, the multiround measurement cycle results in a low readout threshold ($p_R^{\text{th}} \sim 1\%$) for any of the stabilizer measurement protocols, therefore, the link-error injection due to faulty ancilla initialization does not lead to such a significant compromise as for the static stabilizer codes.

Finally, we note that only one logical observable was used for the threshold calculations, but we tested two different axes for the idling bias to avoid getting an overly optimistic threshold for a given logical, as typical for CSS codes.

3. Comparison of the Floquet color code and the honeycomb Floquet code

A side-by-side comparison of the threshold parameters for the Floquet color code and the honeycomb Floquet code is shown in Fig. 18. Both threshold surfaces are very well approximated by a plane and have similar threshold parameters. The honeycomb code, as opposed to the Floquet color code, has a remarkably small sensitivity to noise biases both for gate errors and idling errors.

APPENDIX F: DETAILS OF THE NUMERICS

For the Clifford-circuit simulations we used STIM [43]. For each code distance and physical error rate we took up to 300 000 shots unless 30 000 logical failures are encountered before that. Decoding graphs are built based on the respective STIM detector error models, except for the XYZ² code which is detailed in Appendix D. The reason for the latter was that stim was unable to decompose some of the hyperedges in the decoding graph. The decoding was performed with PYMATCH-ING in each case [13].

- D. Loss and D. P. DiVincenzo, Quantum computation with quantum dots, Phys. Rev. A 57, 120 (1998).
- [2] M. Veldhorst, H. G. J. Eenink, C. H. Yang, and A. S. Dzurak, Silicon CMOS architecture for a spin-based quantum computer, Nat. Commun. 8, 1766 (2017).
- [3] G. Burkard, T. D. Ladd, A. Pan, J. M. Nichol, and J. R. Petta, Semiconductor spin qubits, Rev. Mod. Phys. 95, 025003 (2023).
- [4] X. Xue, M. Russ, N. Samkharadze, B. Undseth, A. Sammak, G. Scappucci, and L. M. K. Vandersypen, Quantum logic with spin qubits crossing the surface code threshold, Nature (London) 601, 343 (2022).
- [5] A. Noiri, K. Takeda, T. Nakajima, T. Kobayashi, A. Sammak, G. Scappucci, and S. Tarucha, Fast universal quantum gate above the fault-tolerance threshold in silicon, Nature (London) 601, 338 (2022).
- [6] G. A. Oakes, V. N. Ciriano-Tejel, D. F. Wise, M. A. Fogarty, T. Lundberg, C. Lainé, S. Schaal, F. Martins, D. J. Ibberson, L. Hutin, B. Bertrand, N. Stelmashenko, J. W. A. Robinson, L. Ibberson, A. Hashim, I. Siddiqi, A. Lee, M. Vinet, C. G. Smith, J. J. L. Morton *et al.*, Fast high-fidelity single-shot readout of spins in silicon using a single-electron box, Phys. Rev. X 13, 011023 (2023).
- [7] S. G. J. Philips, M. T. Madzik, S. V. Amitonov, S. L. de Snoo, M. Russ, N. Kalhor, C. Volk, W. I. L. Lawrie, D. Brousse, L. Tryputen, B. P. Wuetz, A. Sammak, M. Veldhorst, G. Scappucci, and L. M. K. Vandersypen, Universal control of a six-qubit quantum processor in silicon, Nature (London) 609, 919 (2022).
- [8] E. Knill, R. Laflamme, and W. H. Zurek, Resilient quantum computation, Science 279, 342 (1998).
- [9] D. Aharonov and M. Ben-Or, Fault-tolerant quantum computation with constant error rate, SIAM J. Comput. 38, 1207 (2008).
- [10] A. Y. Kitaev, Quantum computations: Algorithms and error correction, Russ. Math. Surv. 52, 1191 (1997).
- [11] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, Topological quantum memory, J. Math. Phys. 43, 4452 (2002).
- [12] H. Bombin and M. A. Martin-Delgado, Optimal resources for topological two-dimensional stabilizer codes: Comparative study, Phys. Rev. A 76, 012305 (2007).
- [13] O. Higgott, Pymatching: A python package for decoding quantum codes with minimum-weight perfect matching, ACM Trans. Quantum Comput. 3, 1 (2022).
- [14] B. J. Brown, K. Laubscher, M. S. Kesselring, and J. R. Wootton, Poking holes and cutting corners to achieve clifford gates with the surface code, Phys. Rev. X 7, 021029 (2017).
- [15] J. R. Wootton, A family of stabilizer codes for anyons and majorana modes, J. Phys. A: Math. Theor. 48, 215302 (2015).
- [16] B. Srivastava, A. Frisk Kockum, and M. Granath, The XYZ² hexagonal stabilizer code, Quantum 6, 698 (2022).
- [17] J. P. Bonilla Ataides, D. K. Tuckett, S. D. Bartlett, S. T. Flammia, and B. J. Brown, The XZZX surface code, Nat. Commun. 12, 2172 (2021).
- [18] M. B. Hastings and J. Haah, Dynamically generated logical qubits, Quantum 5, 564 (2021).
- [19] M. S. Kesselring, J. C. M. de la Fuente, F. Thomsen, J. Eisert, S. D. Bartlett, and B. J. Brown, Anyon condensation and the color code, PRX Quantum 5, 010342 (2024).
- [20] M. McEwen, D. Bacon, and C. Gidney, Relaxing hardware requirements for surface code circuits using time-dynamics, Quantum 7, 1172 (2023).

- [21] R. Raussendorf, J. Harrington, and K. Goyal, Topological faulttolerance in cluster state quantum computation, New J. Phys. 9, 199 (2007).
- [22] D. K. Tuckett, S. D. Bartlett, S. T. Flammia, and B. J. Brown, Fault-tolerant thresholds for the surface code in excess of 5% under biased noise, Phys. Rev. Lett. **124**, 130501 (2020).
- [23] Current rectification by pauli exclusion in a weakly coupled double quantum dot system, Science 297, 1313 (2002).
- [24] J. R. Wootton, Quantum memories and error correction, J. Mod. Opt. 59, 1717 (2012).
- [25] B. M. Terhal, Quantum error correction for quantum memories, Rev. Mod. Phys. 87, 307 (2015).
- [26] A. Kitaev, Fault-tolerant quantum computation by anyons, Ann. Phys. 303, 2 (2003).
- [27] F. Battistel, C. Chamberland, K. Johar, R. W. J. 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).
- [28] W. A. Coish and D. Loss, Hyperfine interaction in a quantum dot: Non-Markovian electron spin dynamics, Phys. Rev. B 70, 195340 (2004).
- [29] D. Culcer, X. Hu, and S. Das Sarma, Dephasing of Si spin qubits due to charge noise, Appl. Phys. Lett. 95, 073102 (2009).
- [30] C. Tahan and R. Joynt, Relaxation of excited spin, orbital, and valley qubit states in ideal silicon quantum dots, Phys. Rev. B 89, 075302 (2014).
- [31] F. Bloch, Generalized theory of relaxation, Phys. Rev. **105**, 1206 (1957).
- [32] Y. Tomita and K. M. Svore, Low-distance surface codes under realistic quantum noise, Phys. Rev. A 90, 062320 (2014).
- [33] F. Vigneau, F. Fedele, A. Chatterjee, D. Reilly, F. Kuemmeth, M. F. Gonzalez-Zalba, E. Laird, and N. Ares, Probing quantum devices with radio-frequency reflectometry, Appl. Phys. Rev. 10, 021305 (2023).
- [34] Z. Chen, K. J. Satzinger, J. Atalaya, A. N. Korotkov, A. Dunsworth, D. Sank, C. Quintana, M. McEwen, R. Barends, P. V. Klimov, S. Hong, C. Jones, A. Petukhov, D. Kafri, S. Demura, B. Burkett, C. Gidney, A. G. Fowler, A. Paler, H. Putterman *et al.*, Exponential suppression of bit or phase errors with cyclic error correction, Nature (London) **595**, 383 (2021).
- [35] A. G. Fowler and C. Gidney, Low overhead quantum computation using lattice surgery, arXiv:1808.06709.
- [36] A. E. Seedhouse, T. Tanttu, R. C. C. Leon, R. Zhao, K. Y. Tan, B. Hensen, F. E. Hudson, K. M. Itoh, J. Yoneda, C. H. Yang, A. Morello, A. Laucht, S. N. Coppersmith, A. Saraiva, and A. S. Dzurak, Pauli blockade in silicon quantum dots with spin-orbit control, PRX Quantum 2, 010303 (2021).
- [37] D. J. Niegemann, V. El-Homsy, B. Jadot, M. Nurizzo, B. Cardoso-Paz, E. Chanrion, M. Dartiailh, B. Klemt, V. Thiney, C. Bäuerle, P.-A. Mortemousque, B. Bertrand, H. Niebojewski, M. Vinet, F. Balestro, T. Meunier, and M. Urdampilleta, Parity and singlet-triplet high fidelity readout in a silicon double quantum dot at 0.5 K, PRX Quantum 3, 040335 (2022).
- [38] A. R. Calderbank and P. W. Shor, Good quantum errorcorrecting codes exist, Phys. Rev. A 54, 1098 (1996).
- [39] C. Jones, M. A. Fogarty, A. Morello, M. F. Gyure, A. S. Dzurak, and T. D. Ladd, Logical qubit in a linear array of semiconductor quantum dots, Phys. Rev. X 8, 021058 (2018).

- [40] J. Haah and M. B. Hastings, Boundaries for the honeycomb code, Quantum 6, 693 (2022).
- [41] S. Bravyi, D. Poulin, and B. Terhal, Tradeoffs for reliable quantum information storage in 2D systems, Phys. Rev. Lett. 104, 050503 (2010).
- [42] S. Bravyi, M. Englbrecht, R. König, and N. Peard, Correcting coherent errors with surface codes, npj Quantum Inf. 4, 55 (2018).
- [43] C. Gidney, Stim: A fast stabilizer circuit simulator, Quantum 5, 497 (2021).
- [44] https://github.com/ibm/qec-with-spin-qubits.
- [45] C. Chamberland and M. E. Beverland, Flag fault-tolerant error correction with arbitrary distance codes, Quantum 2, 53 (2018).
- [46] D. Gottesman, Opportunities and challenges in fault-tolerant quantum computation, arXiv:2210.15844.
- [47] https://github.com/strilanc/midout.