Towards a realistic GaAs-spin qubit device for a classical error-corrected quantum memory

Manuel Rispler

QuTech, Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands
and JARA Institute for Quantum Information, Forschungszentrum Jülich GmbH, 52428 Jülich, Germany

Pascal Cerfontaine ⬤

JARA Institute for Quantum Information, RWTH Aachen University, 52074 Aachen, Germany

Veit Langrock

JARA Institute for Quantum Information, Forschungszentrum Jülich GmbH, 52428 Jülich, Germany

Barbara M. Terhal

QuTech, Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands;
JARA Institute for Quantum Information, Forschungszentrum Jülich GmbH, 52428 Jülich, Germany;
and EEMCS, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands

Based on numerically optimized real-device gates and parameters we study the performance of the phase-flip (repetition) code on a linear array of gallium arsenide (GaAs) quantum dots hosting singlet-triplet qubits. We first examine the expected performance of the code using simple error models of circuit-level and phenomenological noise, reporting, for example, a circuit-level depolarizing noise threshold of approximately 3%. We then perform density-matrix simulations using a maximum-likelihood and minimum-weight matching decoder to study the effect of real-device dephasing, readout error, and quasistatic as well as fast gate noise. Considering the tradeoff between qubit readout error and dephasing time (T_2) over measurement time, we identify a subthreshold region for the phase-flip code which lies within experimental reach.

I. INTRODUCTION

While the theoretical cornerstones of robust quantum computing have been established in the two previous decades in the form of a theory of quantum error correction (QEC) and fault tolerance (see, e.g., [1] and references therein), actual experiments putting this body of research to use are a matter of active research. Until now, the use of a repetition code has been demonstrated in various platforms, such as superconducting qubits [2], and the use of error detection via the four-qubit code has been demonstrated in trapped-ion [3] and superconducting qubits [4–6]. At the same time, the use of bosonic quantum error correction has led to qubits with enhanced lifetimes [7].

In this paper, we model an experimental setup with the aim of demonstrating error suppression of Pauli-Z errors in a spin qubit architecture that holds the prospect of scalability. As is well known, current state-of-the-art experiments on spin qubits allow for one-dimensional (1D) connectivity, while two-dimensional (2D) connectivity and control are more challenging. Indeed, a big challenge in scaling up spin qubits is the "fanout" problem: each quantum dot needs several control lines for defining the dot potential and operating the qubit. This requirement severely limits the capability of tightly packing many qubits onto a chip, especially in a two-dimensional fashion (see, e.g., [8] for discussion). The 2D connectivity required for the surface code is highly nontrivial [9]: promising realizations of long-range two-qubit gates by electron shuttling [10,11] or mediated via superconducting resonators [12] are under active but incomplete development.

Identifying dephasing as the dominant qubit noise process leads us to suggest a one-dimensional layout on which we can operate the phase-flip repetition code. In itself this will not give us a full logical qubit, as it will enhance the X-error rate of the encoded qubit while lowering the Z-error rate. However, it would allow a demonstration of decoding as needed for the surface code. In addition, the tolerable noise threshold error rate is more relaxed as compared to the surface code [1]. This milestone—making a 1D array which realizes the repetition code and demonstrating that the logical error decreases with increasing code distance—has been reached for superconducting devices in [2]. While challenging, we will show that such a device would be currently experimentally feasible for singlet-triplet qubits, even in GaAs, in terms of fabrication and operation, assuming the two-qubit device performance as investigated by [13], without the need for additional elements. We show this by numerically running decoders on repeated error-correcting cycles of the phase-flip code, subject to phenomenological and circuit-level noise, as well as realistic device-specific noise, which we simulate using

full-density-matrix simulations. Previously, such density-matrix simulations and fine-tuned decoders have been very informative in analyzing the performance of small surface codes, such as SURFACE-17 [14] in superconducting devices [15]. Even though our paper is focused on GaAs qubits, a very similar analysis could be done for Si-based qubits: we review and compare some of the elementary-component specifications for Si-based qubits in Sec. II F.

We note that previous work exploring the use of a 1D quantum dot array for making a logical qubit has been done in [16]. However, in this previous work the focus was partially on the four-qubit code and its concatenation, which, executed with purely 1D connectivity, has a depressingly low threshold ($\approx 10^{-4}$). Here instead we focus solely on the repetition code with natural noise (i.e., no artificially inserted errors) and examine logical error rates with growing array size.

In Sec. II we review the singlet-triplet qubit in GaAs and our modeling of relevant hardware components, including the proposed device in Sec. II E. In Sec. III we review aspects of the phase-flip code and discuss decoding methods and results using simple phenomenological and circuit-level error models. This sets the stage for the study and interpretation of the effects of real noise modeled using full-density-matrix simulations in Sec. IV. In Sec. IV we study the tradeoff between readout error and an effective dephasing error rate, the effect of enhanced gate noise, and the effect of stochastifying the noise via the Pauli twirl approximation, and we discuss leakage and leakage reduction (not included in numerical simulations). In Sec. IV E we also numerically present and discuss the enhanced X-error rate. In Sec. V we provide a qualitative discussion about going beyond the 1D phase-flip code: a phase-flip repetition code could be the starting point for a surface code architecture, akin to using the phase-flip code as the bottom code in [17,18]. We will argue, however, that the advantage of this approach is not immediate, but depends on various current unknowns such as the noise bias and fidelity of long- or short-range two-qubit gates.

II. THE SINGLET-TRIPLET QUBIT IN GALLIUM ARSENIDE

The idea of using the spin degree of freedom of a single electron is among the first proposed physical realizations of a qubit [19]. One suitable environment to store, control, and read out single electrons with all-electric control is offered by the platform of gate-defined quantum dots in semiconductor heterostructures. The heterostructure is built such that a two-dimensional electron gas emerges, from which we can load single electrons into quantum dots that are formed by defining a confining potential through applying voltages across gates. Two prominent materials are silicon [20] and gallium arsenide (GaAs). Since device fabrication in GaAs is straightforward and we have validated noise models at hand, we focus on that implementation. We comment on the (close) relation to silicon spin qubits in Sec. II F. We will use the singlet-triplet qubit encoding, where the qubit states are chosen as the $m_z = 0$ subspace of two electron spins residing in two adjacent quantum dots [21]. This has several advantages. The encoding allows the qubit subspace to be less sensitive to the nuclear spin background present in the host material

TABLE I. Time duration of gates and measurement on GaAs singlet-triplet qubits. The last column indicates whether the parameter is fixed or varied in the numerical simulation. Note that the readout time is nominally fixed but the relevant scale is T_2/t_{readout}, i.e., it is sufficient to vary T_2 (see Sec. II D).

Time	Values (range)	Num. simulation
T_1	≈ 2 ms [26]	Fixed
T_2^*	100 ns [23]	Lower limit on T_2
T_2	870 μs [27]	Varied vs t_{readout}
$t_{\mathrm{1q-gate}}$	20 ns [13]	Fixed
$t_{\mathrm{2q-gate}}$	50 ns [13]	Fixed
t_{readout}	1 μs [28]	Fixed

(which is especially prominent in GaAs; it can be mitigated in silicon by using purified ^{28}Si). The singlet-triplet encoding furthermore makes all-electrical qubit control possible by virtue of the Pauli exclusion principle, which is the workhorse behind initialization, gate operations via exchange interaction (provided a finite magnetic-field gradient maintained by other means, see Sec. II B), and readout via Pauli spin blockade. Since these operations are also what introduces noise to the system, we will briefly discuss these operations in this section. The states $|0\rangle$ and $|1\rangle$ of this qubit—expressed in the usual spin-up and -down notation—are the $|0\rangle \equiv |\uparrow\downarrow\rangle$ and $|1\rangle \equiv |\downarrow\uparrow\rangle$ states in the $m_z = 0$ subspace of the two-electron spin-wave function. The two remaining triplet states ($T_+ = |\uparrow\uparrow\rangle$ and $T_- = |\downarrow\downarrow\rangle$) are energetically Zeeman split by applying a static global magnetic field. They can, to a good approximation, be neglected [22], but play a role in qubit leakage (see Sec. IV F). Control over this qubit is established via the exchange interaction between the two electrons in the adjacent dots, which stems from a virtual hopping process between the two dots. This virtual hopping is only allowed for the singlet due to the Pauli principle, creating a small energy difference between the two states, which gives an effective Heisenberg-type interaction. The strength of this interaction can be varied by shifting the dot potentials (i.e., voltages) with respect to another (the "detuning"). The exchange coupling strength of the effective Hamiltonian [see Eq. (1)] can be changed rapidly on the order of nanoseconds through the use of arbitrary waveform generators. The second single-qubit control axis is given by magnetic-field gradients between neighboring dots. In the case of GaAs, the nuclear background field can be locally polarized by so-called dynamic nuclear polarization (DNP) [23], creating the desired gradient field. Fast control of the exchange interaction allows for gate operations with fidelities above 99% by numerical optimization of the pulse sequence [24,25]. We go into more detail on gate operations in Sec. II B. Qubit relaxation is quite strongly suppressed as the qubit frequency sits at a relatively low density of states in the environment, thus leading to a high T_1 time (see Table I). In contrast to this, the bare coherence T_2^* time can be lower than 100 ns, but this can be remedied by echoing techniques (Hahn echo, Carr-Purcell-Meiboom-Gill), which lead to effective T_2 times on the order of μs with recent experimental results going as high as 870 μs [27] (see Table I). The language of coherence and relaxation times is

TABLE II. Error rates of GaAs singlet-triplet qubits. Here $\mathcal{F}$ is the average gate and readout fidelity, respectively. The single-qubit gate numbers are based on numerical models and experimental randomized benchmarking, while the two-qubit gate numbers are based on numerical models so far. Note that we do not vary the fidelity itself but the detailed Hamiltonian model underlying the gate (see Sec. II B).

Component	Infid. $1 - \mathcal{F}$	Num. simul.
Single-qubit gate	0.1–5.0×10^{-3} [13,25]	Varied
Two-qubit gate	0.1–5.0×10^{-3} [13]	Varied
Qubit measurement	5×10^{-3}–1×10^{-1} [28,29]	Varied
Qubit initialization	3×10^{-2}	Varied

common when characterizing qubits experimentally, which is why we stick to this language. Let us, however, remark that ultimately the parameters relevant to error correction are the error rates [cf. Eq. (5) and Table II]. While the two are closely related, the exact relationship depends on the underlying noise model. For T_1 and T_2 this is a Markovian model, which implicitly assumes an exponentially decaying state fidelity, thereby potentially overestimating error rates when non-Markovian low-frequency noise is present (and not completely filtered out by the respective echoing technique), such that in some regime the state fidelity decays slower than exponentially [26]. For a detailed discussion see, e.g., [30]. In the other extreme, very high coherence times are limited by the gate noise induced by the gates that are employed for the echoing pulses themselves. Two-qubit gates are the focus of the current research effort and recent results of accurately modeling the two-qubit operation analogous to the single-qubit case suggest operability at the same high fidelities as the single-qubit gates [13].

A. Qubit initialization

A natural qubit initialization state is the singlet. Using the usual charge notation referring to the number of electrons in the respective dots, we start in a (1,1) charge configuration at small detuning. The dots are then largely detuned so that an electron will tunnel from one dot to the other irrespective of the spin state, yielding a (2,0) charge configuration. Next, the dot potentials are set to a configuration where a triplet state would exchange one electron with the reservoir, resulting in a (1,0) charge state, whereas a singlet state would remain in (2,0). This in effect makes sure we end up in a singlet (2,0) state, which can thus be initialized by fast electron exchange with the lead by waiting in this configuration for a time on the order of tens of nanoseconds [21,31]. For gate operations, one typically moves to smaller detuning. The gates done in [13], which will be explained in the subsequent section, assume an incoming small detuning baseline at the start of the pulse sequence in order to deal with pulse transients caused by the finite bandwidth of the voltage pulses used for qubit control (for details see [13]). In order to reach this qubit operation point after the singlet initialization, one can adiabatically decrease the detuning, such that the singlet hybridizes with the triplet state $|T_0\rangle = \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle)$ into the state $|\uparrow\downarrow\rangle = |0\rangle$ (or, respectively, $|\downarrow\uparrow\rangle$ depending on the direction of the magnetic field gradient). Adiabatic ramping takes on the order of 200 ns and the likely error type is accidentally initializing the qubit state $|1\rangle$, which we model as a bit-flip channel with varying error probability in the few percent range. An alternative way would be to diabatically move to small detuning, which could be done in about 1 ns, preserving and thus preparing the singlet state $|S\rangle = \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle) \equiv \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$.

The gate simulations we use in the present paper and their associated experimental results always employ(ed) adiabatic ramping (due to the pulse transients as mentioned above). In contrast to this, diabatic ramping could potentially be more appealing due to its rapidity, furthermore already preparing the ancilla in a superposition state needed for the syndrome extraction we employ for the error-correcting code. Since the effect of diabatically moving to the qubit operation point lacks experimental characterization, we cover both methods in our simulations, as follows.

We fix the initialization time in the circuit model (see Figs. 2 and 6) to 20 ns; any longer initialization time can be subsumed under the measurement time, which will be a variable in the simulation results, thereby leaving the choice
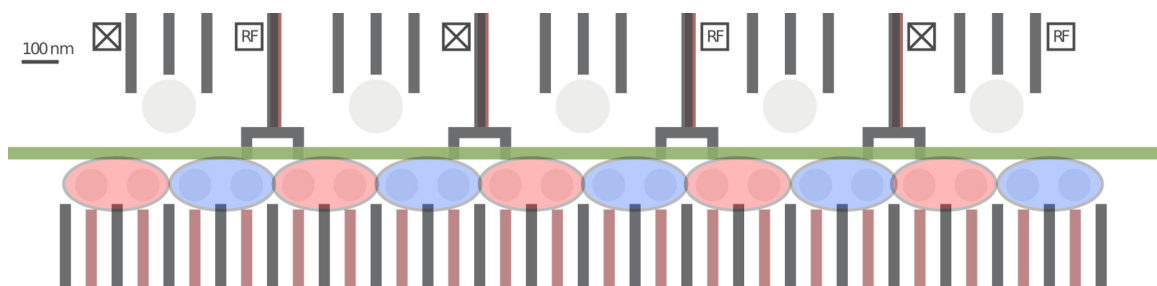

FIG. 1. Top-view layout of a device hosting a phase-flip code in a one-dimensional array of quantum dots. The qubits are in the center of the device, indicated by alternating red (data qubits) and blue (ancilla qubits) ellipsoids each encompassing two quantum dots, since we are using singlet-triplet qubits. The gate indicated by the central horizontal green line isolates the quantum dots from the electron reservoirs situated above. The pitchfork gates can be used to couple each ancilla and qubit to an electron reservoir, which is essential for fast initialization and readout. For the latter, the single-electron transistors situated between the pitchfork gates can be used to detect the change of electron number in the dot. The lines from below are the gates used for defining and operating the quantum dots (see [43]). Note that the rightmost ancilla qubit is not needed for the phase-flip code but useful for leakage reduction (see Sec. IV F.)
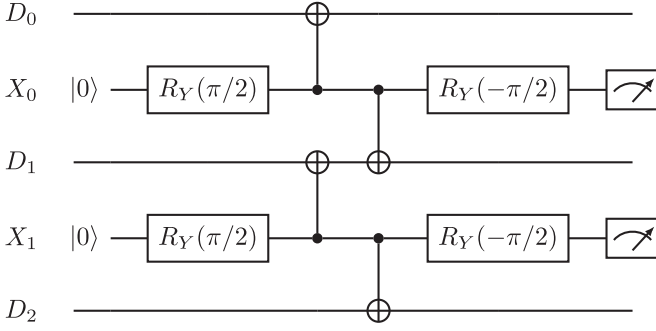
FIG. 2. Circuit diagram for one error-correction cycle of the distance-3 phase-flip code. The XX-parity checks are implemented by coupling to ancillas, which are subsequently measured to give the syndrome information.

of initialization method open to experimental characterization and preference. Simulationwise, both initialization methods, adiabatic ramping followed by a $R_Y(\pi/2)$-rotation gate and directly initializing a singlet, are equivalent; the missing bit flip for initializing a $|+\rangle$ state instead of the resulting singlet state $|-\rangle$ from diabatic ramping can be absorbed into the interpretation of the ancilla measurement by flipping the measurement bit outcome. Likewise for the varying initialization time, for the purpose of this paper, the initialization error will be subsumed under readout error, since for all circuits used in this paper the (bit-flip) error can be propagated through the quantum circuit to the measurement location (see Fig. 2). This can be seen by the fact that the $R_Y(\pi/2)$ rotations convert between X and Z (modulo prefactors), so that a bit-flip error X converts into a Z error which commutes with the action of the controlled-NOT (CNOT) gate, and then turns back into a X error before reaching the measurement location.
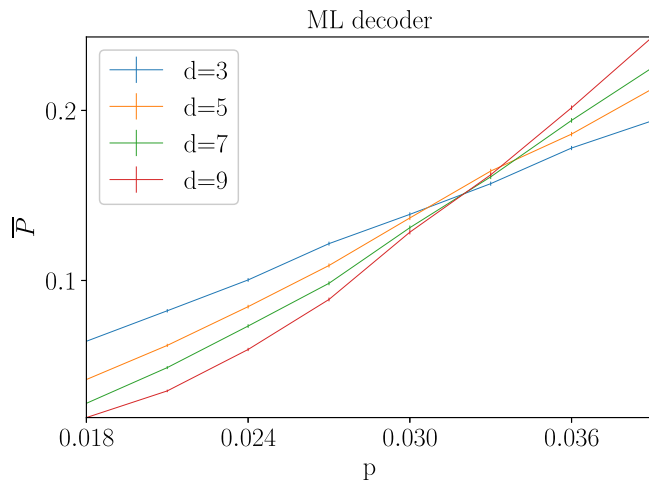

FIG. 3. Logical vs physical error rate of the phase-flip code for circuit-level depolarizing noise, using maximum-likelihood decoding and plotting for increasing code distance d. The threshold is estimated as $p_c = 3.3\%$, below which the logical $\overline{Z}$-error rate $\overline{P}$ decreases with increasing code distance. For every d, we run d QEC cycles N times with $N = 10^4$.
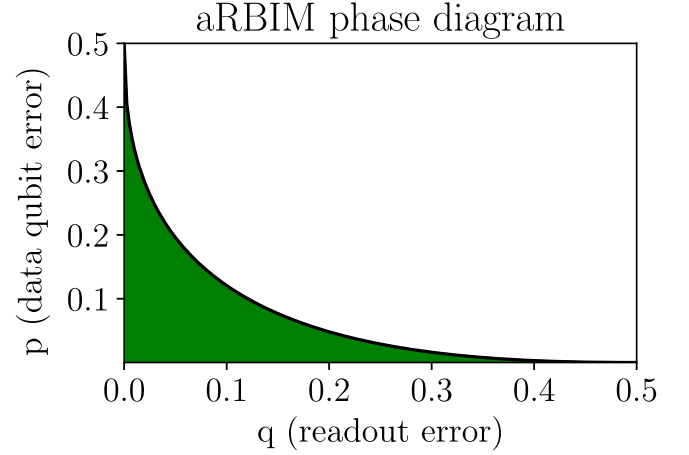

FIG. 4. Phase diagram of the anisotropic random bond Ising model according to the Takeda-Nishimori conjecture [53] given by Eq. (10). The green region marks the ferromagnetic phase, which corresponds to the correctable region of the phase-flip code for anisotropic phenomenological noise.

B. Gate operations

Most of the properties of the singlet-triplet qubit that we consider can be grasped by an effective Hamiltonian that describes the spin dynamics of two qubits, that is, four dots, each hosting a single electron with spin vector $\boldsymbol{\sigma}^{(i)}$, $i = 1, \ldots, 4$, given by

$$H(\boldsymbol{\epsilon}, \boldsymbol{b}) = \frac{1}{4} \sum_{j=1}^{3} J_{j,j+1}(\epsilon_{j,j+1})\boldsymbol{\sigma}^{(j)} \cdot \boldsymbol{\sigma}^{(j+1)} + \frac{1}{2} \sum_{i=1}^{4} B_i(\boldsymbol{b})\sigma_z^{(i)}.$$

(1)

Let us explain the terms entering this Hamiltonian. The first term is the Heisenberg interaction Hamiltonian where $\epsilon_{ij} = V_i - V_j$ denotes the dot detuning between the neighboring dots i and j. We denote all detunings together as $\boldsymbol{\epsilon} = (\epsilon_{12}, \epsilon_{23}, \epsilon_{34})$. As alluded to in Sec. II, this detuning leads to an effective
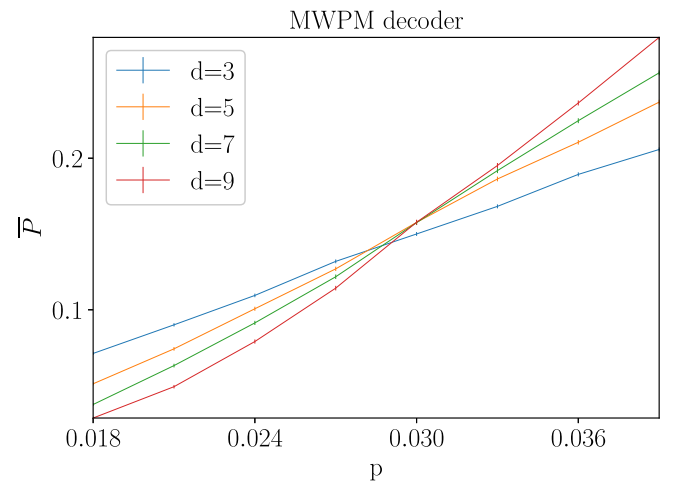

FIG. 5. Logical vs physical error rate of the phase-flip code for the circuit-level depolarizing noise channel for increasing code distance d and minimum-weight decoding. The threshold with this decoder is slightly below MLD at circa $p_c = 3\%$.
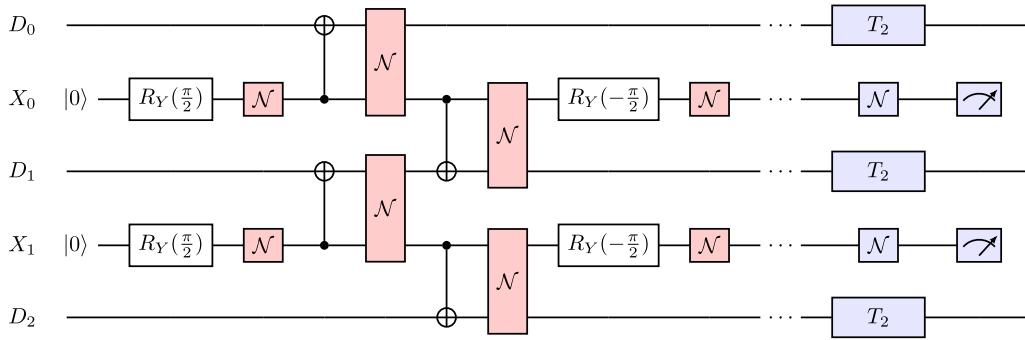
FIG. 6. Circuit diagram for one noisy cycle for the distance-3 phase-flip code. We use red and blue to discriminate between errors happening during parity check gate executions and errors happening during the measurement of the ancilla qubits. The red boxes $\mathcal{N}$ denote the gate noise channels on the single- and two-qubit gates and the initialization error; blue boxes indicate the noise processes during readout. Time is not to scale, indicated by dotted qubit wires: the measurement takes much longer than the gate execution; during this whole time dephasing happens on the data qubits indicated by the blue gate T_2. The measurement is faulty, modeled by a bit-flip channel preceding the measurement. Our simulation contains the effect of T_2 over the full time between gate executions of consecutive QEC cycles (readout time + gate times), however the readout time is dominant, since all data qubits in the bulk otherwise only idle during initialization (20 ns, see Sec. II A) and the Y rotations of the ancillas (2×20 ns). The two boundary qubits idle an additional 50 ns during the CNOT gates. Qubit relaxation is not indicated for readability since it is negligible but it is included in the simulation.

exchange interaction between neighboring dots (hence the sum over nearest neighbors), which is described in terms of the Pauli matrices $\boldsymbol{\sigma}^{(i)}$ and the interaction strength J_{ij}. The second term in the Hamiltonian Eq. (1) is a typical Zeeman Hamiltonian describing the coupling of the electrons to the magnetic fields B_i. By writing $B_i(\boldsymbol{b})$, we allude to the fact that the relevant variables are the magnetic-field gradients $b_{ij} = B_i - B_j$, again using boldface to denote the three gradients between neighboring dots: $\boldsymbol{b} = (b_{12}, b_{23}, b_{34})$. The loss of one variable in moving from site variables (the on-site fields) to link variables (the gradients between neighbors) can be seen by noting that the sum of all fields is irrelevant for the qubit dynamics, since it is a term coupling to $\sum_i \sigma_z^i$, which is zero on any qubit state and thus drops out of the dynamics. We give the explicit change of basis in Appendix B. The dynamics of this Hamiltonian allow for universal control, i.e., arbitrary single- and two-qubit gates, provided that the magnetic field gradients $\boldsymbol{b}$ are finite. Note that we only describe the dynamics of the dots involved in the respective quantum gate, which implicitly assumes negligible cross-talk to other qubits, i.e., we can execute gates in parallel on disjoint sets of qubits. Finding optimal pulse sequences is a challenging optimization problem; we refer the reader to [13], where the average gate fidelity was used as the target function of a numerical optimization routine in order to find good composite pulse sequences for single- and two-qubit gates. The gate unitary coming out of such a pulse sequence is given by the time-evolution operator over a time duration, namely, the gate execution time t_{gate}, which is divided into N intervals of length Δt. During each individual interval the couplings are kept approximately constant, such that the total time-evolution operator is of the form

$$U = \prod_{m=1}^{N} \exp[-iH(\boldsymbol{\epsilon}_m, \boldsymbol{b})\Delta t]. \quad (2)$$

As mentioned in Sec. II, the finite magnetic-field gradients $\boldsymbol{b}$ necessary for universal gates are implemented by DNP of

the background nuclear field (in silicon this is implemented instead by, e.g., the use of micromagnets due to the absence of the nuclear field [32]). This poses a potential problem, since during DNP the dot cannot host a qubit while being used for polarizing the nuclear field. Reaching sufficient levels of polarization takes on the order of hundreds of milliseconds. For the purpose of running the phase-flip code we imagine first using all dots to perform DNP, then initialize the qubits, execute the rounds of error correction for the phase-flip code, and finally measure all the qubits (see, e.g., Fig. 2). For the small distances of the error-correcting code in question in the present paper, DNP does not have to be repeated during the circuit execution, because the runtime is on the order of tens of microseconds (number of QEC cycles times measurement integration time, see Table I) and one round of DNP enables subsequent qubit operations for timescales of at least milliseconds (as on these timescales the nuclear field is stable) before another round of DNP is necessary. The noise characteristics of these gates are captured by the noise affecting the control parameters: the dot detuning and the magnetic-field gradient ("control noise"). We denote the detuning noise by $\delta\epsilon(t)$ and the noise on the magnetic-field gradient by δb, changing the Hamiltonian in Eq. (1) to one which includes the effective noise processes by changing

$$J_{ij}(\epsilon_{ij}) \rightarrow J_{ij}(\epsilon_{ij} + \delta\epsilon_{ij}), \quad (3)$$

$$b_{ij} \rightarrow b_{ij} + \delta b_{ij}. \quad (4)$$

The detuning noise constitutes variations on the voltages, which suffer from random telegraph noise, which is explained by charge traps in the vicinity of the quantum dot, that are periodically loaded and unloaded, leading to small spikes in the voltage signal. While still under debate, this model agrees well with experimental observation and yields $1/f$-type noise (in reference to the power spectrum) [26]. For this detuning noise we use a model that encompasses fast and slow components compared to the timescale of the execution of a gate. The

magnetic-field noise δb is dominated by low-frequency noise, the underlying mechanism being nuclear spin diffusion, which is well described by slowly varying (quasistatic) noise [23,33].

The effect of control noise on the level of gate execution is that we are implementing a Hamiltonian the parameters of which are slightly offset, in turn leading to the actual unitary gate applied to the qubit being slightly different from the intended gate. If these control parameters are changing fast compared to gate operation times, then such noise can be well described by its quantum channel which is obtained by averaging over many noise realizations. In contrast to this, noise parameters which change slowly with respect to gate operation times lead to the computation being subject to the same systematic error over its entire time. In terms of simulating these operations, the fast noise channel can be simulated once and then acts as the same channel at any location. To model slow noise, we will instead draw noise parameters per circuit run instance and only do the averaging after sampling from this distribution on the level of the entire circuit simulation. The details of this model are given in Appendix A. Another error channel besides control noise is leakage, which we briefly discuss here. When looking at the joint Hilbert space of two neighboring qubits, we can see that the $m_z = 0$ subspace is six-dimensional, since it contains the two additional states $|T_+T_-\rangle$ and $|T_-T_+\rangle$. These can be reached from a computational two-qubit state by exchanging the spin between the two dots in the middle (e.g., $|\uparrow\downarrow\uparrow\downarrow\rangle \rightarrow |\uparrow\uparrow\downarrow\downarrow\rangle$), a transition which is not gapped by the global Zeeman field. However, these transitions can be avoided by making such a spin flip energetically costly by imposing that at all times $b_{23} = B_3 - B_2 \gg J_{23}$ [see Eq. (1)]. Detailed two-qubit gate simulations [13] predict very low leakage rates ($\approx 10^{-4}$), which is why we focus on control noise errors in the present paper, deferring leakage to Sec. IV F, where we explain how leakage can be incorporated into error correction via leakage reduction.

C. Readout

A feature of the singlet-triplet qubit is that it grants access to a high fidelity readout mechanism in the form of spin-to-charge conversion. Here one exploits the Pauli exclusion principle, which forbids the two electrons in the triplet state to be in the same dot. By going to a high detuning, it becomes energetically favorable for the singlet to have both electrons in the same dot. This way we can discriminate the qubit states by measuring the number of electrons in either of the dots, which can be done using a single-electron transistor (SET). Readout fidelity is a bottleneck in current spin qubit implementations, both in terms of the measurement integration time, i.e., the time one needs to discriminate the signal on the SET, as well as in terms of the fidelity of this measurement. While current readout numbers can be as bad as milliseconds and 80% fidelity [34], we will see that even these numbers are not necessarily fatal for our purposes, since the code at hand can in principle tolerate a high readout error in a certain parameter regime. On the other hand, there have been recent proposals pushing the envelope of spin qubit readout to the 99% fidelity regime and integration times on the order of microseconds by using a latched readout [28]. A dispersive qubit readout

TABLE III. Comparative table of average gate fidelities and operation times for silicon qubits found in the literature. Since there are several qubit types, the type is indicated as LD (Loss–DiVincenzo) for the bare single-electron spin qubit and as ST_0 for the singlet-triplet qubit. Note that the two-qubit gate infidelity for LD in silicon is the experimental number so far and can be expected to decrease, at least on the basis of theoretical models.

Component	Values (range)	Type
Gate times	$\approx 1\ \mu s$ [36]	LD
2-qu. gate $1 - \mathcal{F}$ (Exp.)	2×10^{-2} [36]	LD
2-qu. gate $1 - \mathcal{F}$ (Th.)	1×10^{-4} [13]	ST_0
T_2	1 ms [37]	LD
T_1	>1 s [38]	LD
Measurement time	1–6 μs [35,39]	ST_0
Meas. error rate	2×10^{-2}–3×10^{-3} [35,39]	ST_0
Meas. error rate	2×10^{-1} [34]	LD

using an on-chip resonator, taking 6 μs with fidelity 98%, has been achieved for Si qubits [35]. In Tables I and II we summarize some of these numbers for the GaAs singlet-triplet qubit. In Table III and Sec. II F we discuss similar numbers for Si qubits.

D. Qubit dephasing and noise tradeoff

So far, we have described gate errors and readout errors. In a typical quantum error-correction protocol (including the one employed in this paper) [1], to protect the information encoded in the data qubits, one extracts error information by using ancilla qubits, which are subsequently measured. During the measurement integration time, the data qubits of the error-correcting code have to wait before we can continue with the next cycle. As the measurement time is $O(100)$ times slower than the individual gate times in Table I, dephasing of the data qubits during measurement by far dominates dephasing-induced Z or X errors during the rest of the QEC cycle.

This slow measurement bottleneck can in principle be circumvented by supplying fresh ancillas for new QEC cycles while the current ancillas are still being measured (see, e.g., the analysis in [40]), or by avoiding measurement all together [41]. However, we choose to treat qubits as a scarce resource, which seems more reasonable in the near term, and thus reuse the measurement qubits to keep the number of qubits minimal. As described above, the dominant qubit noise component is dephasing, which means that during the idling the data qubits suffer a phase-flip error rate [42]:

$$p_Z = \frac{1 - \exp(-t_{\text{readout}}/T_2)}{2}, \tag{5}$$

where t_{readout} is the readout time and T_2 is the dephasing time which can vary—depending on dynamical decoupling pulses—from T_2^* to the best values reported with dynamical decoupling during the readout.

All in all, the efficacy of phase-flip error correction will depend on (1) the intensity of Z errors due to the CNOT gates, (2) the ancilla qubit measurement error rate, and (3) the ratio t_{readout}/T_2. Naturally, the latter quantity can be experimentally

reduced by either shortening the readout time or lengthening the dephasing time by more intensive dynamical decoupling (although there is a limit on how much dynamical decoupling can help as the single-qubit gates themselves introduce additional Z errors). We identify the tradeoff between the readout quality (2) and qubit dephasing (3) as the main tradeoff of the proposed experiment.

E. Proposed linear dot array

To operate the phase-flip code, we imagine a one-dimensional array of alternating data and ancilla qubits, which we realize by a linear array of quantum dots as shown in Fig. 1. This layout is modified with respect to previous linear designs used in several groups [43–45] to allow singlet-triplet qubit operation with DNP [23].

To this end, our proposal first simplifies the fanout of the large number of electrostatic surface gates by moving to the third dimension and stacking three gate layers vertically, isolated from each other by dielectric layers in between. This enlarges the space between adjacent gates and thus lowers the demands on the fabrication process. In Fig. 1, the lowest layer is indicated in red, the middle layer is in dark gray, and the top layer is in green.

Similar to previous designs, the dots are defined by a long horizontal gate (on the topmost layer and thus shown in green) and so-called barrier and plunger gates coming in from the bottom (dark gray and red, respectively). The barrier gates predominantly control the tunnel barriers to adjacent dots, and to the electron reservoirs for the leftmost and rightmost dots. The plunger gates mainly control the dot potentials. By using bias tees, the dc potential of these gates can be set using stable voltage sources while fast control with roughly 300-MHz bandwidth (for manipulation, initialization and readout) can be coupled in simultaneously using arbitrary waveform generators.

In this array, qubits are encoded in two separate but tunnel-coupled quantum dots (using the singlet-triplet encoding described in Sec. II). Experiments [43] have shown that such an array can reach sufficiently high tunnel couplings [31] on the order of tens of μeV. In order to sense the charge state of each dot in the linear chain, sensing dots (SETs) are defined by gates coming in from the top of the diagram. Each SET is confined by the green horizontal gate and three additional gates. The SETs are used for readout of the two qubits next to them. They are connected to electron reservoirs via Ohmic contacts (square boxes) which are in turn connected to a ground potential (black crosses) or a rf readout circuit [46].

In order to allow straightforward singlet-triplet qubit operation, we add the black "pitchfork" gates coming in from the top. These gates can be used to couple each qubit to an electron reservoir by metallic screening of the confinement potential of the green gate. This coupling can be used for fast initialization in a singlet state, which is essential for performing DNP. The black pitchfork gates are screened by lower-lying gates indicated in red so that the electron reservoir surrounding the SETs is not affected by their presence. This allows the potential of the pitchfork gates to be changed without a detrimental effect on the SET potentials.

Direct initialization of each qubit is useful for initializing a well-defined qubit state, and for performing DNP to control magnetic-field gradients between adjacent dots (see Sec. II B). Sufficiently high magnetic-field gradients can suppress leakage to noncomputational T_- and T_+ states during two-qubit gates (see Sec. II B and [13]).

F. Silicon-based qubits

While we picked gallium arsenide for this particular realization, our results are for the most part also informative for silicon qubits, with the silicon ST_0 qubit being the closest cousin to our setup. The gate simulations that were done in [13] also discuss this relationship and conclude that similar gate fidelities can be expected in silicon. We list some typical parameters for silicon qubits in Table III for comparison (note that there are different types of silicon qubits, see references). It is apparent that the noise bias towards decoherence T_2 vs relaxation T_1 is equally prominent in silicon. The main reason for using (purified) silicon is the absence of the nuclear field, which allows for substantially higher relaxation and decoherence times. However, this has the drawback of losing the possibility for dynamic nuclear polarization, which is typically replaced by mounting micromagnets on the sample, which still poses questions on scalability. Gate times are substantially longer, to some extent remedied by advantages in relaxation and decoherence times. An advantage towards dealing with leakage is that single-electron spins can be used as qubits; this has, however, the major drawback that the readout is not of high fidelity. This suggests that a hybrid device could be used with ancilla qubits being singlet-triplet qubits and data qubits embodied by single spins.

III. PHASE-FLIP CODE

The phase-flip code is simply the classical repetition code with n qubits: instead of correcting bit flips we protect against phase flips by rotating the stabilizer checks. The phase-flip code can be viewed as the natural stepping stone or a testbed for implementing the surface code, as its manner of decoding is similar (see also Sec. V).

A. Parity checks, logical operators, preparation, and measurement

The parity checks on n qubits (with n odd) on a line are the nearest-neighbor checks $S_i = X_i \otimes X_{i+1}$ for $i = 0, \dots, n-2$. The logical operators are $\overline{Z} = \bigotimes_{i=0}^{n-1} Z_i$ and $\overline{X} = X_i$ for any i. These act on the logical qubit states defined as $|\overline{0}\rangle \equiv \frac{1}{\sqrt{2}}(|++\dots+\rangle + |--\dots-\rangle)$, $|\overline{1}\rangle \equiv \frac{1}{\sqrt{2}}(|++\dots\rangle - |--\dots-\rangle)$, and $|\overline{+}\rangle = |++\dots+\rangle$ and $|\overline{-}\rangle = |--\dots-\rangle$. The code has distance $d = n$ with respect to Z errors and distance 1 with respect to X errors. The parity check measurements are implemented by coupling both qubits to an ancilla through a CNOT gate, which can be done in two time steps, in parallel on all even and odd qubits. The circuit diagram is given in Fig. 2. Here, the ancilla preparation and measurement in the $\pm$ basis is realized by using rotations $R_Y(\pm\pi/2)$.

A destructive logical $\overline{Z}$ measurement for the phase-flip code corresponds to measuring all qubits in the Z basis and

taking the product of all individual Z_i outcomes. Note that each individual Z_i outcome will be random, but the product of the expectation value is 1 (-1) when applied to $|\overline{0}\rangle$ ($|\overline{1}\rangle$). This measurement is very sensitive to measurement error as its outcome depends on all the individual outcomes. A destructive logical $\overline{X}$ measurement is the measurement of all qubits in the X basis and taking the majority of the answers, and hence is very robust to error. The preparation of $|\overline{+}\rangle$ ($|\overline{-}\rangle$) is simple as it is the preparation of each qubit in the $|+\rangle$ state ($|-\rangle$). The preparation of $|\overline{0}\rangle$ can be done by starting all qubits in the $|00\ldots0\rangle$ state, (also) an eigenstate of $\overline{Z}$, and measuring the S_i checks. These S_i checks will have random outcomes and thus require single-qubit Z corrections (to set their eigenvalues to $+1$) which does not affect the eigenvalue of $\overline{Z}$. In our simulations we start the data qubits in the circuit in Fig. 2 in the state $|++\ldots+\rangle$, and, after repeated rounds of quantum error-correction cycles, we measure each qubit in the X basis. The circuit for code size n is an alternating array of n data qubits with $n-1$ ancilla qubits and otherwise identical QEC cycles as in Fig. 2.

B. Decoding

The decoding problem is an inference problem in which we use error information, obtained by stabilizer measurements, to infer whether a logical error has happened on the encoded qubit. Let us first make a few very general comments on decoding which apply well beyond the phase-flip code studied in this paper.

We can describe the decoding problem as a two-player setup in which one person is the experimenter preparing an encoded input state, performing the QEC cycles and finally deciding whether to measure the encoded qubit in the logical $\overline{X}$ or $\overline{Z}$ basis. The other player is "the decoder" who does not know the input state, nor what final measurement is chosen, but only gets all parity check measurement data **M** acquired during the QEC cycles. The goal of the decoder is to give the experimenter two bits $b_z = 0, 1$ and $b_x = 0, 1$, based on knowing **M**. The experimenter uses these bits as follows. If she measures $\overline{Z}$ and $b_z = 0$, she accepts the outcome of $\overline{Z}$ as the true outcome; if $b_z = 1$ she flips the outcome. Similarly, if she does the $\overline{X}$ measurement, she uses the bit b_x to flip the outcome or not. A slightly more general formulation is one in which the decoder also gets all the last destructive measurement data M_{final}, and tells the experimenter how to read this final logical measurement using M_{final} and the record **M**. The reason to think about the decoder and the experimenter as different identities is that the decoder should not be able to enhance her performance by knowing what state the experimenter started with, or by knowing in advance what measurement the experimenter will perform. In addition, she may be asked to perform the computational task of decoding *on-line*, meaning that there should be no time delay due to her computation lagging behind when the experimenter finishes doing all measurements and wants to know the answer to the final logical measurement.

The most powerful decoder could simulate the entire noisy quantum computation using a full-density-matrix simulation given the best possible noise model. Naturally, this form of decoding is not scalable and defeats the purpose of quantum

computing. In addition, in the two-player setup, the computational power of such decoder has to go beyond that of quantum computing, as she has to postselect her simulation based on the measurement data that she obtains from the experimenter. We will refer to this form of decoding as maximum-likelihood decoding (MLD), as it is the best possible way of decoding.

One can consider the performance of such a maximum-likelihood decoder for a code family of growing distance $n = d$, such as the phase-flip code against Z errors. To compare the decoder's performance for different d, we let the experimenter execute d QEC cycles before executing a final logical measurement [47].

How do we assess the performance of the decoder? Assessing its performance first of all assumes that the noise model that we employ is an accurate description of the physical setup, that is, we have a fairly accurate description of the full-density-matrix evolution, and we can thus compare the decoder's decision with the maximum-likelihood decision.

In our case we are first interested in the occurrence of a logical $\overline{Z}$ error as the aim of the code is to reduce this, hence the experimenter will finally measure $\overline{X}$, i.e., measure all qubits in the $\pm$ basis. In addition, in assessing the performance of the decoder, we will imagine the last measurement step of $\overline{X}$ by the experimenter to be error free, simply so that it does not count towards the logical error rate which is a function of the number of cycles n.

Let us first give the success probability for maximum-likelihood decoding. In principle, the experimenter starts the qubits in some arbitrary encoded unknown state σ. After n rounds of parity check measurements, let the output state of the data qubits be $\rho_{\mathbf{M}}^{\sigma}$ where **M** is a multi-index label for the parity check outcomes, i.e., **M** has (no. of stabilizers) $\times$ (no. of QEC cycles) entries. Now the simplest maximum-likelihood decoder assumes that the input is, say, $|\overline{+}\rangle$ (instead of σ) and runs a density-matrix simulation and outputs "flip," i.e., $b_x = 1$, when in her simulation $\mathbb{P}(\overline{X} = -1|\mathbf{M}, \overline{+}) > \mathbb{P}(\overline{X} = 1|\mathbf{M}, \overline{+})$. The failure probability, *if the state of the experimenter was indeed* $|\overline{+}\rangle$, is then

$$\overline{P}_{\text{MLD}}$$
$$= \sum_{\mathbf{M}} \mathbb{P}(\mathbf{M}) \times \min(\mathbb{P}(\overline{X} = -1|\mathbf{M}, \overline{+}), \mathbb{P}(\overline{X} = 1|\mathbf{M}, \overline{+})).$$
$$(6)$$

We will take this probability as a proxy for the MLD logical failure probability for $\overline{Z}$ for arbitrary input states σ for simplicity, as we do not expect that starting with $|\overline{-}\rangle$ would give the decoder a very different decision, or that arbitrary inputs will fare very differently. This is exactly correct when the noise model is that of depolarizing errors which act in a completely state-independent way, but it is an assumption when we implement more general noise in Sec. IV. Similarly, if we would evaluate MLD decoding for a $\overline{X}$ error, we would prepare a $|\overline{0}\rangle$ state and the decoder decides $b_z = 1$ when $\mathbb{P}(\overline{Z} = -1|\mathbf{M}, \overline{0}) > \mathbb{P}(\overline{Z} = 1|\mathbf{M}, \overline{0})$ with corresponding logical failure probability.

To test maximum-likelihood decoding we apply it to a standard noise model of circuit-level depolarizing noise. This model is as follows. For a single qubit, the (symmetric)

depolarizing channel is the Pauli channel:

$$\mathcal{E}(\rho) = (1-p)\rho + \frac{p}{3}\sum_i P_i\rho P_i. \qquad (7)$$

Circuit-level noise means that we apply this noise channel after every single-qubit element (initialization, single-qubit gate, readout) in the circuits such as Fig. 2. For a two-qubit gate, we apply the two-qubit version of the symmetric depolarizing channel which leaves the state alone with probability $1-p$ and applies any one of 15 two-qubit Pauli channels with probability $p/15$. We use full-density-matrix simulations using QUANTUMSIM software available at [48] to simulate the noisy circuits, although for the noise models in this section simpler stabilizer simulations are possible. By evaluating Eq. (6) for different code distances and error probability p, we find a threshold value of the depolarizing circuit noise of $p_c = 0.033$ in Fig. 3 using maximum-likelihood decoding. The threshold here is taken to be the point where the curves for different d cross, which captures the asymptotic performance of the code.

To put these results in perspective and discuss the computationally efficient method of minimum-weight matching decoding, let us take a step back and first consider some simpler phenomenological noise models. The simplest one is to assume that all parity check measurements are perfect and in each QEC round every qubit is hit by a phase-flip error channel:

$$\mathcal{E}(\rho) = (1-p)\rho + pZ\rho Z. \qquad (8)$$

After a single QEC round one can mark which stabilizers are flipped as defects and match these defects to each other or to the outside boundary, choosing the matching which minimizes the total distance between the matched defects. The matching produces a logical error when the number of phase-flip errors is at least $(d+1)/2$ (i.e., the majority is in error and—unbeknownst to the decoder—choosing the complementary matching would have been correct). Thus the logical failure probability is

$$\overline{P} = \sum_{k=(d+1)/2}^{d} \binom{d}{k} p^k (1-p)^{d-k}, \qquad (9)$$

which for $d \to \infty$ tends towards zero for any $p < 0.5$ giving the well-known threshold value $p = 0.5$ for the phase-flip code with perfect syndrome measurements.

A next model is that of phenomenological noise. In this noise model, each qubit undergoes a Z flip with error probability p in each QEC round and each parity check is perfect except for the outcome of the ancilla measurement being flipped with probability q. In this noise model we have the additional effect that defects do not necessarily correspond to actual data qubit errors anymore but can instead stem from an inaccurate measurement itself. To handle this, one needs to process syndrome measurement outcomes from one round to the next. From the record **M** the decoder creates a syndrome *defect* record by placing a defect between cycle t and $t+1$ when the stabilizer measurement outcome changes from cycle t to $t+1$. A measurement error at cycle t will thus lead to a pair of defects at time $(t-1, t)$ and $(t, t+1)$ and pairing these defects in decoding means that we interpret it as

such. An incoming qubit phase-flip error in round $t+1$ will lead to two neighboring defects at time $t+1$ in the bulk of the lattice, which can be matched. On the boundary there is only one defect (which could be matched to the boundary). A minimum-weight matching decoder now takes this defect record and matches all defects with the goal of minimizing the total distance between the matched defects, where the distance is taken as a function of the error probability p and q. The algorithm of minimum-weight perfect matching (MWPM) is the efficient Edmonds blossom algorithm [49].

If we choose the measurement error rate q equal to the Z error rate p, then, except for the marginal details on the boundary, the decoding problem is known to be equivalent to the decoding problem of the toric code with perfect measurements [50]. The equivalence can be understood by seeing that the time direction of the repetition code plays the role of the second spatial direction of the toric code. This toric code with perfect measurements has the well-known threshold value of $p = 0.11$ [51] under maximum-likelihood decoding. The minimum-weight matching decoder performs close to optimal with a threshold of $p = 0.105$ [50,52].

The case $p \neq q$ and the optimally achievable maximum-likelihood threshold are relevant in understanding the numerical data for full-density-matrix simulations in Sec. IV as they effectively feature the same tradeoff between measurement error (error rate q) vs incoming error (error rate p). Takeda *et al.* [53] have conjectured the phase boundary separating the below-threshold to above-threshold region to lie at the following line:

$$H(p) + H(q) = 1, \qquad (10)$$

where $H(x)$ denotes the Shannon entropy $H(x) = -x\log_2(x) - (1-x)\log_2(1-x)$; see Fig. 4. When $p = q$, the condition $H(p_c) = 1/2$ implies $p_c \approx 11\%$. The threshold boundary shape in Fig. 4 provides a guidance for understanding the results in the next section, e.g., Fig. 7. Even though the noise is more involved as it includes some level of gate noise, the main features of a tradeoff between readout error vs phase-flip error are indeed present. In Appendix C we review the underlying theory behind maximum-likelihood decoding for the repetition code.

Let us now present our results on applying MWPM on the *circuit-level depolarizing* noise model to see the discrepancy with ML decoding. In order to decide how to best match up the syndrome defects, we assign equal weight to both timelike (ancilla) and spacelike (data) errors, i.e., for decoding we essentially assume a phenomenological noise model with equal measurement as incoming error rate $p = q$. We then perform the matching of a given syndrome graph with the weights given by this noise model by using a standard implementation of the blossom algorithm in the PYTHON package NETWORKX [54]. As shown in Fig. 5, the logical error rates cross at a value of $p = 3\%$, which shows that MWPM with a rather simple noise model performs close to optimal for circuit-level depolarizing noise. This good performance might be not too surprising given the low (equal to 2) weight parity checks of the phase-flip code. The inclusion of so-called vertical hooks [2,50] for the matching algorithm might bring the MWPM threshold closer to the optimal MLD threshold, however we chose not to investigate this slight difference
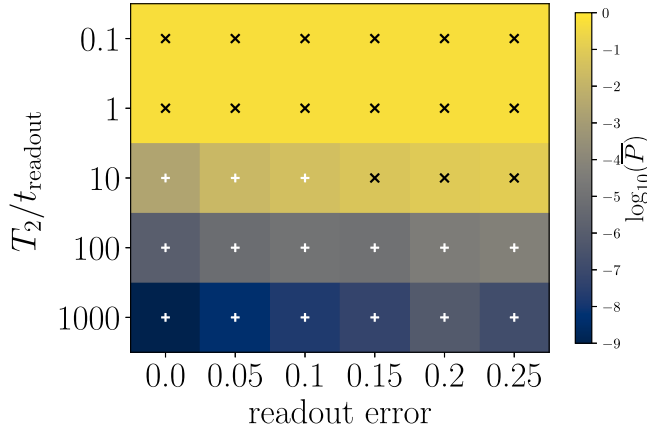
FIG. 7. Phase diagram of the error enhancement to error suppression transition of the logical Z error of the phase-flip code for lowering readout error (x axis) and lowering qubit errors (y axis), using an MWPM decoder. The criterion of error enhancement (black $\times$, upper region) and error suppression (white $+$, lower region) is determined by a full circuit simulation of the real phase-flip code circuit with noisy components. This strongly indicates the feasibility of showing error suppression in the spin qubit device depicted in Fig. 1. The logical error rate is shown as a heat map (corresponding to the distance-9 code at each parameter tuple).

further. For the surface code, the question of the timing of errors (i.e., at which point in the parity check circuit they occur) is much more sensitive due to the higher-weight parity checks and the fact that X and Z checks are acting on the same qubits, which can lead to problematic error patterns (notably single ancilla errors spreading to several data qubit errors) [50]. In a phase-flip code the CNOT gates cannot propagate ancilla errors to data qubit errors due to the structure of the parity checks.

IV. DEVICE-SPECIFIC FULL-DENSITY-MATRIX SIMULATIONS

A. Full-density-matrix circuit simulation

Using the QUANTUMSIM framework [48] we build a numerical simulation of the quantum circuit outlined in Fig. 6, where all the components are faulty, i.e., the gates are given by realistic noisy channels, the qubits decohere, and the initialization and measurement can be imperfect, according to Tables I and II. In particular, the gates are derived from the qubit Hamiltonian $H(\boldsymbol{\epsilon} + \delta\boldsymbol{\epsilon}, \mathbf{b} + \delta\mathbf{b})$ with the magnetic field gradients $\mathbf{b}$ and the detunings $\boldsymbol{\epsilon}$ between the dots, Eq. (1), to which both fast and slow noise are added according to an accurate noise model (Appendix A). Relaxation and decoherence times for the qubits are used during measurements and idling steps. Initialization and measurement errors are modeled by a bit-flip channel preceding the measurements, that is, they are jointly subsumed under readout error.

B. Dephasing vs readout: Parameter exploration

The goal of our proposal is to run the device in Fig. 1 at increasing code distances and gather statistics on the logical error rates in order to test whether current technology is good

TABLE IV. Gate metrics for the gate sequences we use in this paper. The (in-)fidelities are the entanglement fidelities computed by the pulse sequence optimization developed in separate work by one of us (P.C.) (see the related paper [13]). The χ matrix was computed from these gate sequences as part of the present paper in order to estimate the error rates of the gates. As a rough estimate we give the error rate as $1 - \chi_{00}$.

	CNOT	$R_Y(\pi/2)$
$\mathcal{F}$	99.74%	99.89%
$1 - \mathcal{F}_{slow}$	9×10^{-4}	1.5×10^{-4}
$1 - \mathcal{F}_{fast}$	1.9×10^{-3}	1×10^{-3}
$p \approx 1 - \chi_{00}$	2.8×10^{-3}	1.2×10^{-3}

enough to achieve error suppression in a spin qubit experiment, i.e., the device can operate below the error-correction threshold of the code at hand.

Looking at the circuit components, we first estimate the error rate of the involved single- and two-qubit gates, which we do by averaging the channel over many (10^3) quasistatic noise realizations and then turning this channel into a Pauli channel by averaging over the Pauli group, which is known as the Pauli twirling approximation (PTA). For a channel given by $\mathcal{E}(\rho) = \sum_{m,n} \chi_{mn} P_m \rho P_n$ where P_m are Pauli channels and $P_0 = I$, the PTA takes the approximation $\mathcal{E}_{\text{PTA}} = \sum_m \chi_{mm} P_m \rho P_m$ with $1 - \chi_{00}$ the corresponding error rate. The gate fidelity is experimentally accessible through randomized benchmarking and is also the metric for which the pulse sequences are optimized, i.e., we list both $1 - \chi_{00}$ as well as the gate (in-)fidelities computed by the pulse sequence finder designed by one of us (P.C.) in a different related project [13] in Table IV [55]. We note that these fidelities and gate error estimates do not enter in our simulations as we simulate full gate dynamics, but the numbers are meant to provide guidance of expected performance and check whether gate noise is biased towards particular Pauli errors. We have observed that the other entries on the diagonal of the χ matrix for the gates are not very biased towards a particular error with entries not varying over more than two orders of magnitude.

Comparing these numbers to the thresholds we found in Fig. 3, it is reasonable to believe that gate operations are below threshold for the phase-flip code. As described above, one of the bottlenecks at least in current and near term implementations of spin qubits is their long measurement times. While waiting for the measurement, the data qubits are idling, which as a noise process is strongly biased towards dephasing. Thus T_2 and T_2^* are relevant figures of merit. We therefore undertake a numerical parameter study for a range of realistic estimates for integration time and readout fidelities. We explore the range between bare T_2^* times and the most optimistic timescales for echoing, by choosing $T_2/t_{\text{readout}} = \{0.1, 1, 10, 100, 1000\}$, since we deem it reasonable to believe that the multidot device will be able to operate within this range (see Sec. II E). The readout time t_{readout} is yet another parameter which is to be determined, however it is typically on the order of 1 μs, in which case the y axis of Fig. 7 and related figures directly correspond to T_2 times.

Since the readout fidelity is still to be characterized to high accuracy, we let this parameter range over a wide, possibly pessimistic, spectrum by allowing up to 25% readout error. For every parameter pair we construct the circuit for several code distances $d = 3, 5, 7, 9$ and take 10^3–10^4 samples, which we decode with both the maximum-likelihood decoder and a minimum-weight decoder, where we adapt the weights by choosing a spatial weight $w_s = -\log(p)$ with the error rate given by the dephasing time and a temporal weight $w_t = -\log(q)$ with the measurement error rate q. For a code of distance d we execute d QEC cycles as in Sec. III B. We observe that this MWPM decodes close to optimal (MLD) for our parameter ranges, but we do not plot the MLD data here.

For every parameter tuple, we determine whether the logical error rate is monotonically decreasing with increasing code distance d. If this criterion is fulfilled, we mark the tuple with a white $+$ symbol; if it is not fulfilled, we mark it with a black $\times$ symbol. Furthermore, to estimate the code performance we overlay the plot with a heatmap of the logical error rate of the distance-9 code at the corresponding parameter value. The result is shown in Fig. 7. We observe that for $T_2 = 10\,\mu$s with a given measurement integration time of $1\,\mu$s the code could tolerate a readout infidelity of up to 15%. From the next order of magnitude in $T_2 > 100\,\mu$s the study suggests an extremely high tolerance to readout errors. This can be understood by looking back at the phenomenological anisotropic noise model (see Fig. 4). While the noise in our simulation is in principle not that simple since it occurs at the circuit level and is spatially and temporally correlated, this suggests that when the gates have low error rates we can indeed tolerate very high readout noise.

C. The effect of increasing gate noise

In our gate simulations so far, we took the experimentally measured values for the width of the distribution of slow b field $\sigma_{\delta b}$ and exchange noise $\sigma_{\delta\epsilon}$ (Appendix A). Since this noise has to this point never been characterized in a device of more than four dots, extrapolating these values might be too optimistic. To safeguard against possible increases in the noise distribution, we studied the effect of a wider distribution by performing the simulation described in the previous paragraph (Sec. IV B) with a standard deviation of the noise distribution, which is twice and four times as large for both the charge noise as well as the magnetic-field fluctuations. The results of doubling the noise strength can be seen in Fig. 8: the region of error suppression stays intact but the logical error rate increases. In contrast to this, for a fourfold increase of the noise strength we have observed that the structure of the parameter region is lost and we do not observe a contingent region that would indicate error suppression in the studied parameter range.

D. Coherence of noise and Pauli twirling

The question we want to address in this section is whether the coherence of the noise and the temporal and spatial correlations of slow noise are relevant in our numerics. The
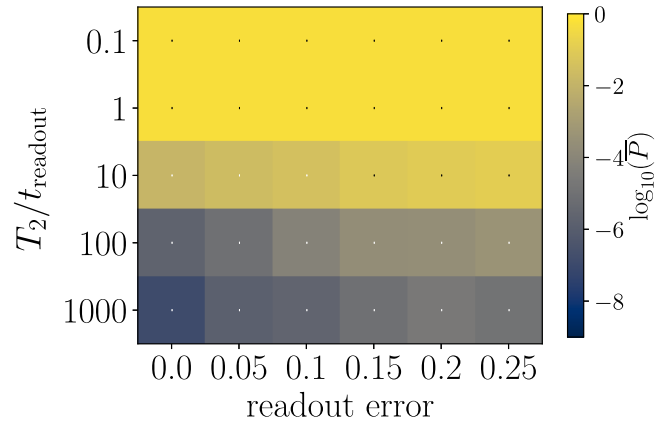

FIG. 8. Cf. Fig. 7: Here we doubled the noise distribution width $\tilde{\sigma}_{\delta b, \delta\epsilon} = 2\sigma_{\delta b, \delta\epsilon}$. The logical error rates increase by one order of magnitude, but the essential structure is maintained: the phase-flip code provides error suppression of phase noise for a wide range of parameters. (Error enhancement and suppression markers are not shown; they are the same as in Fig. 7.)

impact of applying a Pauli twirl approximation converting a general noise channel to a depolarizing noise channel and enabling scalable numerical simulations of Clifford circuits, instead of full-density-matrix simulations, has already been studied in the literature (see, e.g., [56]). Results on small codes suggest that Pauli twirling overestimates the logical error rate, thus providing to some extent an upper bound on the logical error rate [57]. It has also been reported that Pauli twirling is a good approximation for incoherent noise models and worse for coherent errors [58].

Since we use density-matrix simulations we can compare full noise simulations with those in which we remove slow temporal correlations in the noise as well as apply the PTA approximation for each gate in the circuit. We compare the full circuit simulation reported above with a simulation where we replace every noisy gate by its perfect incarnation followed by an asymmetric depolarizing channel, which is obtained by twirling the true noisy gate error channel (sampled sufficiently often so the channel has a convergent representation).

As shown in Figs. 9 and 10, the shape of the phase boundary stays the same and also the logical error rate differs only slightly. Our numerical results thus suggest that Pauli twirling does not dramatically alter the predictions of simulating a small phase-flip code in our regime of parameters. This has several explanations: probably the dominating noise is not gate noise, which is quite far below threshold in itself, no matter which error model is used. Furthermore, it is also not completely coherent; the contribution from slow noise is comparable to the contribution from fast noise. Another contribution, which has also been observed in other studies, is the projective nature of the ancilla qubit measurement model, projecting out superpositions of Pauli errors onto specific Pauli errors on the data qubits.

All in all we conclude that Pauli twirling seems to be a very acceptable numerical approximation in our regime, showing that costly full-density-matrix simulations are in fact not needed to estimate the logical code performance for our parameter regime.
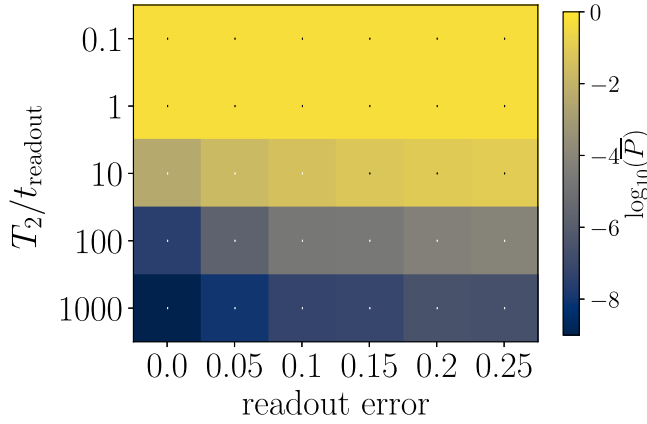
FIG. 9. The effect of Pauli twirling: Analogous to Fig. 7, we plot the phase diagram of the error enhancement to error suppression transition of the phase-flip code (using MWPM decoding), but here instead of simulating the full noise dynamics we replace every noise process by its (slow and fast) noise averaged and then twirled version. (Error enhancement and suppression markers are not shown; they are the same as in Fig. 7.)

E. The enhanced logical X-error rate

Of course, dealing with a phase-flip code, we have no logical protection against X errors. While the qubit noise is very biased towards phase flips, the gates in the parity check circuit invariably introduce X errors. If we start the QEC cycles in the state $|\overline{0}\rangle$ and finally destructively measure $\overline{Z}$, it gives us an estimate for the logical $\overline{X}$-error rate. We can roughly estimate this error rate by computing the probability that there are an odd number of X errors on the final qubits (including errors in the final measurement) after d QEC cycles. Let $p_{\text{eff},x}$ be the effective X error of a qubit per QEC cycle. This error rate is induced by the CNOT gate and X errors induced by T_1 relaxation during the measurement. Note that bit-flip errors
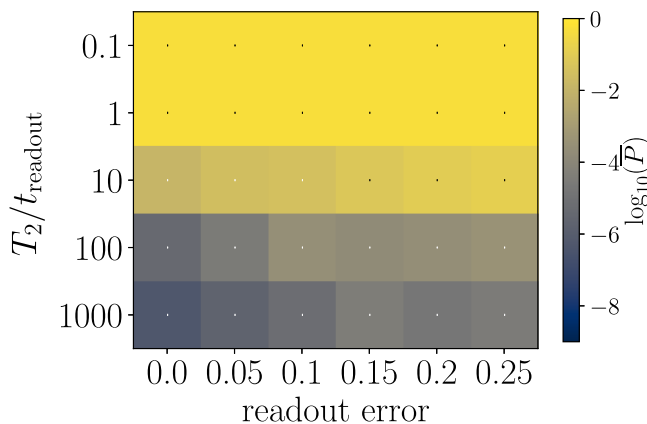

FIG. 10. The effect of Pauli twirling: Analogous to Fig. 8, we increase the noise strength of slow noise by a factor of 2 in the distribution width $\tilde{\sigma}_{\delta b,\delta\epsilon} = 2\sigma_{\delta b,\delta\epsilon}$ and plot the phase diagram of the error enhancement to error suppression transition of the phase-flip code, but here instead of simulating the full noise dynamics we replace every noise process by its twirled and noise averaged version. (Error enhancement and suppression markers are not shown; they are the same as in Fig. 7.)
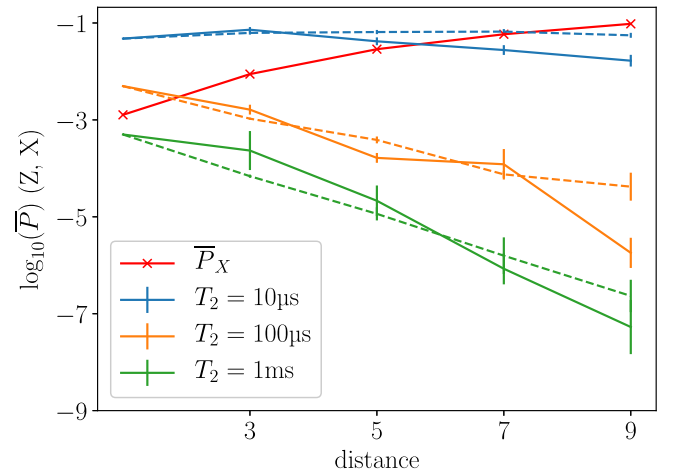

FIG. 11. Logical error rate tradeoff of the phase-flip code. Solid $\overline{P}_Z$ is for readout error 5% and dashed is for 15%. The rising line is the X-error rate; the constant and falling lines are Z-error rates (ordering analogous to the legend.) For sufficiently long T_2 times (with $t_{\text{readout}} = 1\ \mu\text{s}$), by encoding in the phase-flip code, the logical Z-error rate is exponentially suppressed. In contrast to this, the logical X-error rate increases with larger encoding circuit. The distance-3 code has a logical X-error rate at the value of the surface code threshold, which upon noise optimization could allow for a subthreshold error rate for a code concatenation.

coming out of a $R_Y(\pm\pi/2)$ gate do not enter the X-error rate, since they propagate to an even number of data qubits. In total, the logical qubit suffers an X error if there are an odd number of X errors in a space-time volume of size d^2 (composed of d QEC cycles on d qubits), i.e.,

$$\overline{P}_X(p_{\text{eff},x}) = \sum_{k\ \text{odd}} \binom{d^2}{k} p_{\text{eff},x}^k (1 - p_{\text{eff},x})^{d^2-k}$$

$$= \frac{1}{2}[1 - (1 - 2p_{\text{eff},x})^{d^2}]. \tag{11}$$

Neglecting the effect of T_1 errors, we estimate that the CNOT introduces an X error on the data qubit with error rate $p_{\text{eff},x} \approx 0.0013$. For comparison, we compute the logical X-error rate in the circuit simulation by sending the state $|\overline{0}\rangle = \frac{1}{\sqrt{2}}(|++\ldots+\rangle + |--\ldots-\rangle)$ through the circuit. The relative sign of this state is flipped by X errors and thus we can determine the logical X-error rate by determining the probability that the final state has negative relative sign. The values of this are in good agreement with the phenomenological formula in Eq. (11) at $p_{\text{eff},x} \approx 0.0013$.

We show the behavior of the logical X-error rate in Eq. (11) for increasing code distance compared to the logical Z-error rate for the same distance, which we plot for two settings of readout error (5 and 15%) in Fig. 11. As expected, the logical X error is growing for large distances, however the distance-3 code is not hopeless. It sits close to the surface code threshold at 1% [51]. Note that the logical X error is extremely sensitive to the X-error rate of the CNOT gate. The gate sequences could be tailored to specifically avoid these types of errors by adapting the target function of the pulse sequence optimization routine, thereby potentially pushing

the logical X-error rate below surface code threshold (see also Sec. V).

F. Leakage and leakage reduction

The qubits employed in the envisioned architecture are encoded in the $\{|S\rangle, |T_0\rangle\}$ subspace of two electrons in a double dot. The remaining triplet states $\{|T_+\rangle, |T_-\rangle\}$ comprise noncomputational leakage states and the most relevant process for leakage in our setup is faulty gate operation [24]. Continued leakage accumulation poses a threat to error correction [59–61], but the problem can be addressed using so-called leakage reduction units (LRUs). These units return leaked qubits to the computational subspace, potentially introducing qubit errors, which can be handled by the error-correcting code [62]. For singlet-triplet qubits, such a LRU can be implemented by a swap if leaked (SIL) procedure [62], swapping the concerned (data) qubit with a fresh ancilla qubit only in case the data qubit is in a leakage state. A compact way of representing the desired operation (see Sec. II for qubit and leakage states) is a truth table specifying the required mapping for such a SIL procedure [62]:

$$|S^D \Psi_0^A\rangle \rightarrow |S^D \Psi_1^A\rangle, \quad |T_0^D \Psi_0^A\rangle \rightarrow |T_0^D \Psi_1^A\rangle,$$
$$|T_-^D \Psi_0^A\rangle \rightarrow |\Psi_{r,-}^D T_-^A\rangle, \quad |T_+^D \Psi_0^A\rangle \rightarrow |\Psi_{r,+}^D T_+^A\rangle,$$

where $|\Psi_0^A\rangle$ and $|\Psi_1^A\rangle$ are initialization and output states in the computational ancilla subspace $\{|S^A\rangle, |T_0^A\rangle\}$ and $|\Psi_{r,-}^D\rangle$ and $|\Psi_{r,+}^D\rangle$ are reset states in the $\{|S^D\rangle, |T_0^D\rangle\}$ subspace, all of which may be chosen arbitrarily. We have used D for data qubit and A for ancilla qubit, i.e., we are swapping leakage from data to ancilla qubit. This gadget is a two-qubit interaction, which can be implemented analogously to the two-qubit gates we employ elsewhere in this paper, so that it is reasonable to assume that the SIL sequence can be similar in both duration (50 ns, see Table I) as well as fidelity. A neat feature of the proposed chip design (Fig. 1) is that it allows for fast energy selective reset into a singlet state via coupling to a nearby electron reservoir, such that it is possible to execute the leakage reduction sequence including ancilla reset on a timescale of tens to hundreds of nanoseconds depending on the initialization method of $|\Psi_0^A\rangle$ (see Sec. II A).

For the chosen qubit type, material system, and architecture, leakage induction will be dominantly due to execution of gates (with, e.g., the single-qubit gate leakage rate measured to be 0.13% [25]). Therefore, most of the leakage will be accumulated during the echoing sequence executed on the data qubits while the ancilla qubits are being measured. These echoing techniques are necessary to generate the high effective T_2 used throughout the paper, but need of the order of ten single-qubit gates to reach T_2 times that ensure sufficient qubit coherence after the measurement time has passed (e.g., in [63] 16 gates were needed to reach 200 μs). We therefore place the leakage reduction block directly before the stabilizer gates (two-qubit interactions) between data and ancilla qubits (see Fig. 13), allowing us to remove leakage acquired during the echoing sequence from the data qubits before reaching the stabilizer block. Conversion of leakage to errors in the computational subspace allows for the error-correcting circuit to detect and correct them. This is essential, since having
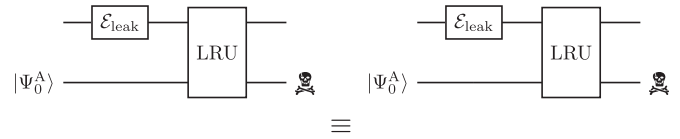

FIG. 12. The error channel of a data qubit undergoing a leakage process is converted into an effective qubit channel by using an ancilla that is swapped in by the LRU gate in case of leakage. The ancilla qubit is afterwards discarded (or reset and used as an ancilla qubit for QEC), which we represent by the ☠ symbol.

leaked states as input to the two-qubit gates of the parity checks can cause checks to flip or spread errors (depending on experimentally and theoretically unexplored details). Performing an LRU operation has the effect that a leaked data qubit state is reinitialized in the computational qubit subspace, which from a quantum information perspective means that we can effectively ignore the ancilla (tracing it out) and can describe the effective single-*qubit* channel (emphasis on qubit, i.e., not leakage), as shown in Fig. 12.

Let us describe the leakage process by some incoming leakage error rate p_{leak} with a corresponding channel:

$$\mathcal{E}_{\text{leak}}(\rho) = (1 - p_{\text{leak}})\rho + p_{\text{leak}}\Sigma, \tag{12}$$

where ρ is some qubit state, but importantly Σ is some state in the leakage space, i.e., some mixture of $|T_+\rangle$ and $|T_-\rangle$. The effect of the LRU is to transform Σ into a reset state inside the computational subspace. This means that the leakage channel
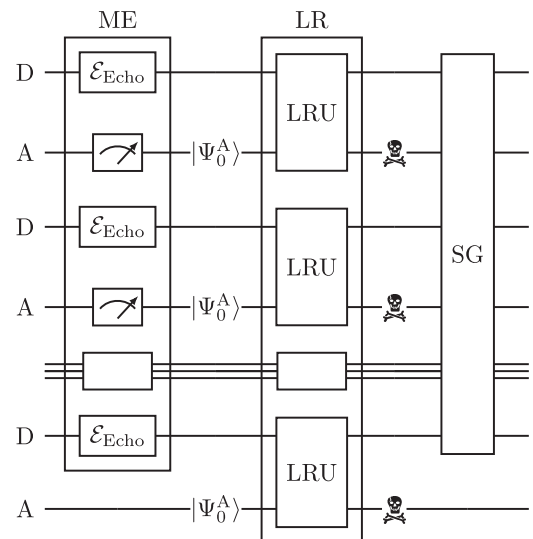

FIG. 13. Circuit arrangement of the operation blocks to sustain the logical qubit including leakage reduction. During the measurement and echo (ME) block (of duration > 1 μs) syndromes are collected from the ancillas, while the data qubits are subject to echo sequences, which are assumed to be the main culprit for inducing leakage (see text). The subsequent leakage reduction (LR) converts the leakage into qubit errors in the computational subspace, which are passed to the following stabilizer gate (SG) block where the two-qubit interactions take place. As each data qubit has to be paired up with an ancilla partner during the LR block, one extra ancilla qubit is added to the array.

is transformed into the effective channel

$$\mathcal{E}_{\text{eff}}(\rho) = \mathcal{E}_{\text{LRU}}[\mathcal{E}_{\text{leak}}(\rho)] = (1 - p_{\text{leak}})\rho + p_{\text{leak}}\rho_{\text{reset}}, \quad (13)$$

where ρ_{reset} is some state supported in the space spanned by $|T_0\rangle$ and $|S\rangle$. Here we are ignoring higher-order contributions, such as the LRU itself being a noisy process. If this error channel produces Z errors, they can be detected by subsequent QEC cycles of the phase-flip code. We therefore propose to perform a leakage reduction procedure directly before the error-correction unit, as shown in Fig. 13. It should be noted that both the effective coherence time T_2 during the echo sequence as well as the accumulated leakage in the data qubits increase with the number of gates (π pulses) executed during the echo sequence, as described in the previous paragraph. This implies that it might be a bit naive to strive for T_2 times as large as possible, since this comes with the cost of increased gate-induced error rates. The larger the number of gates in the echo sequence, the more errors of the Pauli-X type accumulate (see Sec. IV E), which are not detectable by the QEC code at hand. This again implies a tradeoff that probably makes very high T_2 times less desirable than they might seem at first glance. Let $p_{\text{gate},x}$ denote the X-error rate on a data qubit induced by a decoupling sequence. Then by using n gates the induced X error per data qubit scales as $p_{\text{eff},x} = n p_{\text{gate},x} + \mathcal{O}(p_{\text{gate},x}^3)$. It is striking that [27] uses up to on the order of hundreds of Pauli-X gates (π pulses) to reach 0.87-ms T_2 time, while [63] only needed 16 X gates to demonstrate T_2 around 200 μs, which indicates that the scaling of the number of pulses with achieved T_2 times is superlinear (this approximate argument should be taken with care, e.g., the mentioned numbers are experimental results and were reached with π pulses that are themselves imperfect). As we have elaborated in the main part of this paper, when assuming readout times on the order of microseconds, hundreds of microseconds of T_2 would comfortably put us in a regime of Z-error suppression (Fig. 7). For a rough estimate, our study of induced X errors in Sec. IV E shows that an effective induced X error of 10^{-3} would be borderline tolerable, when trying to stay below the surface code threshold under code concatenation. This implies that even the moderate number of tens of π pulses needed in the decoupling of [63] would already elevate the X-error rate to a level on the order of the surface code threshold of 10^{-2}, when assuming $p_{\text{gate},x} \approx 10^{-4}$–$10^{-3}$ (Sec. IV B). We therefore note that while our paper focuses on the effective Z errors induced by the natural noise bias of spin qubits towards dephasing as opposed to relaxation one should not neglect X errors when looking beyond this. We imagine that this problem can be solved by making do with moderate numbers of π pulses (on the order of tens of pulses) and combining this with an idea of improving the gates that are used in the decoupling sequences. Given that in the scenario described here errors induced by the π pulses are much more harmful when of the X-error type than when of the Z-error type, one could change the cost function in the gate optimization routines (see details of the gate noise model in Appendix A) to optimize not for overall gate infidelity but specifically for minimal X-error rate, which could potentially push the induced error rates below the surface code threshold.

V. DISCUSSION AND OUTLOOK: TOWARDS THE SURFACE CODE?

Here we discuss several open questions relating to the feasibility of moving beyond a small phase-flip code.

A layout based on a linear chain of quantum dots is mostly likely not easily scalable to the large number of qubits required for building a practically useful quantum computer. While we mitigated the fanout problem by introducing extra gate layers, this issue will reappear as more qubits are added. In addition, connecting a large number of gates to off-the-shelf room-temperature electronics via dedicated wires is also a serious obstacle for scaling to large numbers of qubits. In the near future, both of these problems could be solved by integrated control electronics which can be connected to the qubit chip via flip-chip bonding, thus reducing the fanout and the need for wires [64].

In addition to issues with wiring and fanout, control cross-talk will complicate the operation of closely packed qubits. Since true two-qubit cross-talk caused by unwanted two-qubit interactions is likely not an issue for singlet-triplet qubits [13], compensating linear control cross-talk between adjacent control lines and qubits does not present a fundamental obstacle. However, calibrating cross-talk in large arrays of tightly packed qubits could still prove difficult if the cross-talk does not fall off quickly enough with the distance between qubits.

In addition, it could be necessary to consider the effects of second-order tunneling or mediated exchange [65,66] if the exchange interaction between several dots is turned on simultaneously. Furthermore, DNP [23] has not been realized in large quantum dot arrays so far, but should also not pose a fundamental obstacle. However, coherence times are unlikely to be longer than demonstrated for isolated qubits as additional error mechanisms and decay channels are added.

If a 2D device architecture were available, one may ask whether the use of the phase-flip code would still be of interest. For example, one could directly use the surface code with limited "cross-bar" control as envisioned in [9] or consider faster, measurement-free, use of small surface codes [41]. One could imagine using a clustered 2D layout with each cluster representing a logical qubit encoded by the phase-flip code connected by long-range CNOT gates with other clusters or ancilla qubits, while short-range CNOT gates are used inside the cluster.

The answer to this "best architecture" question cannot fully be given at this stage and depends on various considerations, for example, to what extent the noise is biased towards Z errors. Let us give some considerations.

For GaAs singlet-triplet qubits, even when performing dynamical decoupling, a factor of 10 between the X-error rate and Z-error rate may be expected. Similarly, idling Si-based qubits have a strong noise bias towards T_2 (see Table III). If two-qubit gates are of very high fidelity (with negligible error) and noise is dominated by phase-flip errors during slow, noisy measurements, then, instead of using the phase-flip code as bottom code, one might be able to directly benefit from using a modified surface code in which one measures X and Y checks, considered in [67]. With an (incoming) phenomenological Z-error rate $p_Z \approx p$ and X- and Y-error rate equal to $p_X = p_Y \approx p/200$ (noise

bias $\eta = 100$) and a measurement error rate equal to p, the reported threshold in [67] is $p_c = 5\%$. For the GaAs numbers in Tables I and II, $p_Z \approx t_{\text{readout}}/T_2 = 1.25 \times 10^{-3}$ and $p_X \approx t_{\text{readout}}/T_1 = 5 \times 10^{-4}$, putting one safely below this 5% threshold.

However, this picture is not realistic since the two-qubit gate error rates are in fact not negligible and biased-noise or bias-preserving gates have not been the focus of previous research. For example, doing four CNOT gates in a surface code QEC cycle, each CNOT gate would have to have X-error rate at least below $\frac{5\%}{800} = 0.625 \times 10^{-4}$ in order to get below this 5% threshold of [67]. It might be interesting to develop such noise-bias preserving CNOT with dominant Z noise as has been done for superconducting devices in [68].

Given that the noise bias is unlikely to be a direct advantage in surface code decoding a la [67], we can further consider using the phase-flip code as the bottom code to equalize the X- and Z-error rates as we have numerically explored in Sec. IV E. Thus the phase-flip repetition code could then be concatenated, in principle, with the surface code.

If a strongly biased-noise CZ gate (with dominant Z errors) was available for spin qubits, one could imagine following an approach advocated in [17] with the phase-flip code as bottom code. Note, however, that the logical CNOT gadget for the phase-flip code from [18] which aims to preserve this noise bias is nonideal as it heavily uses slow qubit measurements and additional qubit overhead. So we will not focus on obtaining a CNOT in this way. Let us now discuss the possible drawbacks and variations of this concatenation scheme.

First, let us imagine encoding each qubit of the surface code in the phase-flip code, including surface code ancilla qubits. One then imagines alternating three or more cycles of phase-flip error correction with applying logical gates, e.g., the CNOT gates to implement the parity check circuits of the surface code as in Figs. 14 and 15. As the phase-flip repetition code is a Calderbank-Shor-Steane code, the logical CNOT gate is a transversal gate, meaning that it can be realized by applying copies of the basic CNOT gate between two code blocks [69].

Now consider the logical error rate induced by such transversal CNOT gate if we use the distance-3 phase-flip code. Since the phase-flip code has *no* correction against X errors, we have a logical failure on one of the encoded qubits when any one of the three CNOT gates produces an outgoing X error. Said differently, the use of the three-qubit phase-flip code would make us suffer from a CNOT-induced error rate which is three times higher than if we were to use no phase-flip encoding but use the surface code "straight-up." Given that the surface code cannot tolerate an error rate of more than 1% of circuit-level depolarizing noise [70], such form of concatenation puts some very strong demands on the CNOT X-error rate.

One simplification of this scheme is to use bare ancilla qubits to measure the toric code checks instead of phase-flip encoded ancilla qubits (see the circuits in Figs. 14 and 15). This is advantageous since these ancilla qubits are only short lived so preparation and measurement errors are dominant sources of errors. In particular, the logical preparation of $|\overline{0}\rangle$ and logical measurement $\overline{Z}$ are cumbersome in the phase-flip
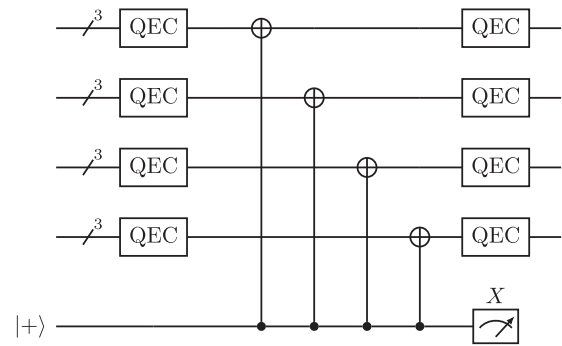

FIG. 14. Surface code X-parity check circuit, using data qubits each encoded with the three-bit phase-flip code suggested by the number 3 on the lines. Each QEC unit denotes three or more cycles of phase-flip error correction and is repeated continuously, interspersed by CNOT gates. The surface code ancilla could be chosen to be encoded, so that $|+\rangle$ denotes a logical $|\overline{+}\rangle$ and X denotes the logical $\overline{X}$ measurement for this ancilla. The preparation of $|\overline{+}\rangle = |+++\rangle$ is simple and the logical $\overline{X}$ measurement is robust by majority-vote taking. However, since the circuit is short in duration, using a single qubit prepared in $|+\rangle$ as ancilla will most likely have better performance. Note that each CNOT in the circuit has to be done transversally, between triples of qubits when the ancilla is encoded, or between a triple and a single ancilla qubit when the ancilla qubit is unencoded.

code. If we consider Fig. 15, a transversal CNOT gate between encoded data qubit and encoded ancilla then gets replaced by three CNOT gates between encoded data qubits and the single-qubit ancilla. These three CNOT gates propagate any X error on the data qubits (a logical $\overline{X}$ for the phase-flip code) to the ancilla and make it flip and thus detectable. If two out of the three qubits have an X error, it is equivalent to a stabilizer and the ancilla will not flip. For the X-parity circuit, even though one could use an encoded ancilla, as the preparation of $|\overline{+}\rangle = |+++\rangle$ is straightforward and its measurement robust, its performance will be no better than using a simple $|+\rangle$ and measuring in the X basis. So again, we apply three CNOT gates between an encoded data qubit and the
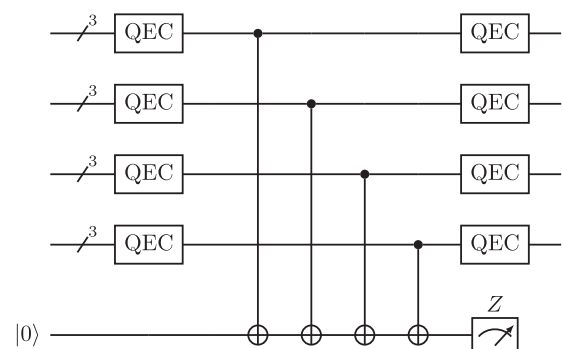

FIG. 15. Surface code Z-parity check circuit on phase-flip encoded triples of qubits. The ancilla could be replaced by its phase-flip encoded version, but the logical $\overline{Z}$ measurement is sensitive to measurement errors and the state $|\overline{0}\rangle$ is not simple to prepare. It is thus certainly better to use a single ancilla qubit started in $|0\rangle$ and a direct Z measurement.

$|+\rangle$ ancilla. For any odd number of Z errors on a data qubit, either the logical $\overline{Z} = Z_1 Z_2 Z_3$ or Z_i will thus make the ancilla flip from $|+\rangle$ to $|-\rangle$. This has to be taken into account in the decoding of the toric code, i.e., information about dephasing errors on the phase-flip code has to be combined with toric code information.

However, now that we consider using single uncorrected ancillas, one can also see the whole scheme as one big code. This code has local adjacent XX checks on the clusters (due to the phase-flip code) which are measured continuously in the EC units in Figs. 14 and 15. In addition, the code has (high) weight-12 Z checks for surface code X-error correction as well as the usual weight-4 X checks (or alternatively, weight 12 as well) for Z-error correction. It then seems that this scheme can only be of interest when short-range and long-range CNOT have extremely low error rates so that the gains of the phase-flip code are not undone by the losses involved in having to do many CNOT gates.

All in all, we see that the achievable quality and bias of the two-qubit gate is crucial in setting the optimal use of quantum error correction beyond the scheme presented in this paper. Theoretical explorations of noise-bias preserving gates and their use will be worthwhile.

APPENDIX A: DETAILS OF THE NOISE MODEL USED IN NUMERICAL SIMULATION

Noisy gates can be modeled by perfect gates instantaneously followed by a noise channel describing the deviation of the actual noisy gate from the perfect (targeted) gate. We model the single- and two-qubit gate similarly to how it was done in [13], decomposing the noise into slow and fast components, where in our simulations we additionally account for the fact that we do not only care about a single gate but about the performance of many QEC cycles. As described in Sec. II B, the dominant noise processes can be understood as variations to the control signals that comprise the execution of the desired gate. To remind the reader, these are the dot detunings ϵ and the magnetic-field gradients b, such that the relevant noise parameters are random fluctuations around these values, which we denote by $\delta\epsilon$ and δb. In case of a two-qubit gate these are thus three variables each, for all the exchange couplings and gradients between neighboring dots. The pulse sequence takes the time t_{gate}, during which we modulate the detuning quickly. Let us now write this time dependence explicitly as $\epsilon(t)$. This is in principle a continuous analog signal, however to very good approximation described by a discretization into time steps, i.e., a discrete time series

of detunings. Adding noise to this signal amounts to adding another time series on top of the intended one for each individual ϵ_{ij} and b_{ij} (the intended one comes out of a numerical optimization as also stated in Sec. II B). The noise signal is therefore a time series which can be described by the methods of time-series analysis typically done in terms of the so-called power spectrum of the autocorrelation function of the signal (see, e.g., [71]). This analysis describes the noise signal in the frequency domain, which lets us decompose the noise into fast and slow components with respect to the gate execution time. The slow parts are variations to the signal that do not change over the duration of the gate; we thus model them as a random value set once and kept over the gate execution time. These random variables can be efficiently obtained by drawing from a normal distribution with a standard deviation that comes out of experimental observations (in our case $\sigma_{\delta\epsilon} = 8\,\mu\text{V}$ [26] and $\sigma_{\delta b} = 0.3\,\text{mT}$ [72]). That is, we draw as many uncorrelated random variables as there are detunings in the setup. This constitutes a noise instance of slow noise, which we can add to the detuning time sequence. The same analysis holds for the magnetic-field gradients in complete analogy; moreover for magnetic-field noise it turns out that the noise is dominated by slow components [33], such that the above method is exhaustive. For charge noise on the detuning, we have to go further. Modeling the fast noise component is slightly more involved, since here we have to take the variations during the gate sequence duration into account. We generate noise instances of fast noise by imitating the power spectrum that has been characterized experimentally for our system (GaAs): the high-frequency spectrum follows $S_{\epsilon,\alpha}(f) \propto S_0/f^\alpha$ with $\alpha = 0.7$ [26]. Since this spectrum has not been measured above a few MHz, which are still relevant frequencies for our gates, we extrapolate with $\alpha = 0.7$ such that $S_0 = 4 \times 10^{-20}\text{V}^2/\text{Hz}$ at 1 MHz [26]. With the ability to generate fast and slow noise instances, we now turn to how we obtain the effective noise channels that enter our simulations as described in Fig. 6. On the level of an individual noise instance, a noise signal just has to be added to the intended pulse sequence and magnetic-field gradients before integrating the Schrödinger equation [Eq. (2)], which gives an instance of a unitary

$$\tilde{U} = \prod_{m=1}^{N} \exp[-iH(\epsilon_m + \delta\epsilon, b + \delta b)\Delta t] \qquad (A1)$$

that is the noisy version of a perfect unitary U. In order to capture the noisy part of this gate, we undo the ideal part and give the noisy unitary part the name E:

$$E = U^\dagger \tilde{U}. \qquad (A2)$$

Whether E describes slow or fast noise only alters what we put into $\delta\epsilon$ and δb. For slow noise, we draw these parameters $\delta\epsilon_{\text{slow}}$ and δb_{slow} once per realization of the whole QEC circuit, such that all gates during the circuit execution experience the same instance of slow noise (the same offset of the detuning and the magnetic-field gradients). The averaging over slow noise is then performed by averaging over many instances of noisy circuit realizations.

Since we work with a full-density-matrix simulation, we express this as a unitary quantum channel:

$$\mathcal{N}_{\text{slow}}(\rho) = E_{\text{slow}} \rho E_{\text{slow}}^{\dagger}. \tag{A3}$$

In contrast to this, when we want to describe fast noise processes, we generate many ($\Omega = 10^4$) realizations of the random signal $\delta\epsilon_{\text{fast}}$, each time integrating the Schrödinger equation to obtain a unitary instance E_k of Eq. (A2), where the index runs over the number of realizations. We then describe the fast noise process as the average quantum channel by averaging over these instances, i.e.,

$$\mathcal{N}_{\text{fast}}(\rho) = \frac{1}{\Omega} \sum_{k=1}^{\Omega} E_k \rho E_k^{\dagger}. \tag{A4}$$

The combined noise process $\mathcal{N}$ is then described by concatenating slow and fast noise:

$$\mathcal{N} = \mathcal{N}_{\text{fast}} \circ \mathcal{N}_{\text{slow}}. \tag{A5}$$

We implemented these channels for the gates needed for the phase-flip code (Fig. 6) in QUANTUMSIM. Amplitude damping and phase damping are described in the Kraus representation $\mathcal{E}(\rho) = \sum_i A_i \rho A_i^{\dagger}$ by the following Kraus channels:

$$A_0^{\text{AD}} = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix}, \quad A_1^{\text{AD}} = \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix}, \tag{A6}$$

$$A_0^{\text{PD}} = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\eta} \end{pmatrix}, \quad A_1^{\text{PD}} = \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{\eta} \end{pmatrix}, \tag{A7}$$

where γ and η are the amplitude and phase damping rates. These rates are typically expressed as times T_1 (relaxation or amplitude damping time), T_φ (pure dephasing time, phase damping), and T_2 (decoherence time), for which it holds that $e^{-t/T_1} = 1 - \gamma$, $e^{-t/T_\varphi} = \sqrt{(1-\eta)}$, and $e^{-t/T_2} = \sqrt{(1-\gamma)(1-\eta)}$. Since in our setup, $T_1 \gg T_2$, decoherence acts like pure dephasing noise $T_2 \approx T_\varphi$, which as a noise channel can be also written (as seen, e.g., by exploiting the unitary freedom in the Kraus representation) as a phase-flip channel $\mathcal{S}(\rho) = (1 - p_Z)\rho + p_Z Z \rho Z$ with error rate

$$p_Z = \frac{1 - e^{-t/T_\varphi}}{2} \approx \frac{1 - e^{-t/T_2}}{2}. \tag{A8}$$

APPENDIX B: ZEEMAN HAMILTONIAN WITH MAGNETIC-FIELD GRADIENTS

For completeness we here show the Hamiltonian in Eq. (1) expressed in terms of the magnetic-field gradients. The Zeeman Hamiltonian is simply $H_{\text{Zeeman}} = \frac{1}{2} \sum_i B_i \sigma_z^i$. The aim is a change of variables $\boldsymbol{B} \to (B_{\text{G}}, \boldsymbol{b})$ with $b_{ij} = B_j - B_i$ and $B_{\text{G}} = \frac{1}{4} \sum_i B_i$. This implies a change-of-basis matrix:

$$R = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix}. \tag{B1}$$

Evidently the Pauli-Z matrices have to be transformed with $(R^{-1})^{\text{T}}$, such that the total Hamiltonian in Eq. (1) is expressed as

$$H(\boldsymbol{\epsilon}, \boldsymbol{b}) = \frac{1}{2} B_{\text{G}} \sum_{i=1}^{4} \sigma_z^{(i)} + \frac{b_{12}}{8} \left[-3\sigma_z^{(1)} + \sigma_z^{(2)} + \sigma_z^{(3)} + \sigma_z^{(4)} \right]$$

$$+ \frac{b_{23}}{4} \left[-\sigma_z^{(1)} - \sigma_z^{(2)} + \sigma_z^{(3)} + \sigma_z^{(4)} \right]$$

$$+ \frac{b_{34}}{8} \left[-\sigma_z^{(1)} - \sigma_z^{(2)} - \sigma_z^{(3)} + 3\sigma_z^{(4)} \right]$$

$$+ \frac{1}{4} \sum_{\langle ij \rangle} J_{ij}(\boldsymbol{\epsilon}) \boldsymbol{\sigma}^{(i)} \cdot \boldsymbol{\sigma}^{(j)}, \tag{B2}$$

where the term involving B_{G} can be ignored for the qubit dynamics (since $\sum_{i=1}^{4} \sigma_z^{(i)}$ is zero on the qubit subspace).

APPENDIX C: DATA VS READOUT NOISE: ANISOTROPIC RANDOM BOND ISING MODEL

As established in [50,51], the maximum-likelihood decoding problem for the toric code subjected to phenomenological noise can be mapped onto the evaluation of the partition function of a classical spin model. The spin model is disordered, depending on quenched randomness which is due to the random outcomes of syndrome measurements. As mentioned in the main text, the decoding of the repetition code with noisy parity checks with phenomenological error rate $q = p$ is isomorphic to the surface code with data qubit error rate p and perfect readout. Both models map onto the 2D random bond Ising model (RBIM), which has the Hamiltonian

$$H = -\sum_{\langle ij \rangle} J_{ij} S_i S_j \tag{C1}$$

with classical spin variables $S_i = \pm 1$ on a square 2D lattice. The sign of the coupling is drawn from a bimodal probability distribution:

$$J_{ij} = +J \quad \text{with } 1 - p, \quad J_{ij} = -J \quad \text{with } p. \tag{C2}$$

This entails antiferromagnetic bond defects with probability p (p is usually called the bond concentration). Antiferromagnetic bonds form strings, the end points of which are called Ising vortices. These end points correspond to the syndrome defects as defined earlier and antiferromagnetic (AFM) bond strings correspond to possible error strings.

The decoding problem can then be phrased as finding recovery strings with the same end points as the error string and asking whether they fall in the same homology class or not. Let us imagine some error string the qubits have suffered. This string is in actuality hidden from us; the syndrome only reveals the end points of strings. We (the decoder) construct "candidate" recovery strings to explain the syndrome that is seen and recover from the error the qubits have suffered. The combination of error string plus recovery string necessarily forms a loop; the question of logical correctness is whether and with what likelihood this loop is homologically nontrivial.

This argument is then turned around; namely, if the probability of being in the same homology class goes to unity in the thermodynamic limit, the likelihood of making a logical

error goes to zero. To make the connection to the classical spin model, the recovery strings can be viewed as being created by thermal excitations. As long as error string plus recovery string form a loop with trivial homology, they are the boundary of a region, which is a domain wall. The phase diagram of the spin model has two axes, the bond concentration p and the temperature $T = \frac{1}{\beta}$. Since the error string (setting the bond defects) and the recovery string are actually drawn from the same distribution, the Boltzmann weight and the bond concentration have to be identified as

$$e^{2\beta J} = \frac{p}{1-p}, \tag{C3}$$

which is called the Nishimori line in the two-dimensional β-p plane. The correctability condition is thus—in the statistical mechanics picture—translated into the condition that domain walls remain localized, which is the case in the ferromagnetic phase of the random bond Ising model. The threshold of the code is given by the point where the phase boundary of the ferromagnetic to paramagnetic transition crosses the Nishimori line (at $p_c \approx 0.11$ [51]).

Anisotropic RBIM: Takeda-Nishimori conjecture

We review the lesser known case $p \neq q$, corresponding to the fact that timelike AFM bonds are created with a different probability than spacelike bonds. This maximum-likelihood decoding problem can then be mapped to the evaluation of the partition function of an anisotropic random bond Ising model:

$$H = - \sum_{\langle ij \rangle \in \text{vertical}} J_{ij}^v S_i S_j - \sum_{\langle ij \rangle \in \text{horizontal}} J_{ij}^h S_i S_j \tag{C4}$$

with

$$J_{ij}^v = J^v \quad \text{with } 1-p, \qquad J_{ij}^h = J^h \quad \text{with } 1-q, \tag{C5}$$

$$J_{ij}^v = -J^v \quad \text{with } p, \quad J_{ij}^h = -J^h \quad \text{with } q \tag{C6}$$

and the Nishimori conditions

$$e^{2\beta J^h} = \frac{p}{1-p}, \quad e^{2\beta J^v} = \frac{q}{1-q}. \tag{C7}$$

This model has been analyzed by Takeda *et al.* [53]. The spin model has four free parameters, the two bond concentrations p and q and the two coupling strengths (J^h and J^v). Enforcing the two Nishimori conditions projects the four-dimensional configuration space down onto a two-dimensional plane with the bond concentrations p and q as free parameters. This plane is called the Nishimori sheet (analogous to the well-known Nishimori line), on which the criticality condition, being a relation between p and q, corresponds to a line, the critical line marking the phase boundary between error suppressing (ferromagnetic) and error enhancing (paramagnetic) bond concentrations (see Fig. 4 in the main text).

[1] B. M. Terhal, Quantum error correction for quantum memories, Rev. Mod. Phys. **87**, 307 (2015).

[2] J. Kelly, R. Barends, A. G. Fowler, A. Megrant, E. Jeffrey, T. C. White, D. Sank, J. Y. Mutus, B. Campbell, Yu Chen, Z. Chen, B. Chiaro, A. Dunsworth, I. C. Hoi, C. Neill, P. J. J. O'Malley, C. Quintana, P. Roushan, A. Vainsencher, J. Wenner, A. N. Cleland, and J. M. Martinis, State preservation by repetitive error detection in a superconducting quantum circuit, Nature (London) **519**, 66 (2015).

[3] N. M. Linke, M. Gutierrez, K. A. Landsman, C. Figgatt, S. Debnath, K. R. Brown, and C. Monroe, Fault-tolerant quantum error detection, Sci. Adv. **3**, e1701074 (2017).

[4] M. Takita, A. W. Cross, A. D. Córcoles, J. M. Chow, and J. M. Gambetta, Experimental Demonstration of Fault-Tolerant State Preparation with Superconducting Qubits, Phys. Rev. Lett. **119**, 180501 (2017).

[5] C. Vuillot, Is error detection helpful on IBM 5Q chips? Quantum Inf. Comput. **18**, 949 (2018).

[6] R. Harper and S. T. Flammia, Fault-Tolerant Logical Gates in the IBM Quantum Experience, Phys. Rev. Lett. **122**, 080504 (2019).

[7] N. Ofek, A. Petrenko, R. Heeres, Ph. Reinhold, Z. Leghtas, B. Vlastakis, Y. Liu, L. Frunzio, S. M. Girvin, L. Jiang, M. Mirrahimi, M. H. Devoret, and R. J. Schoelkopf, Extending the lifetime of a quantum bit with error correction in superconducting circuits, Nature (London) **536**, 441 (2016).

[8] D. P. Franke, J. S. Clarke, L. M. K. Vandersypen, and M. Veldhorst, Rent's rule and extensibility in quantum computing, Microprocess. Microsyst. **67**, 1 (2019).

[9] R. Li, L. Petit, D. P. Franke, J. P. Dehollain, J. Helsen, M. Steudtner, N. K. Thomas, Z. R. Yoscovits, K. J. Singh, S. Wehner, L. M. K. Vandersypen, J. S. Clarke, and M. Veldhorst, A crossbar network for Silicon quantum dot qubits, Sci. Adv. **4**, eaar3960 (2018).

[10] J. M. Taylor, W. Dür, P. Zoller, A. Yacoby, C. M. Marcus, and M. D. Lukin, Solid-State Circuit for Spin Entanglement Generation and Purification, Phys. Rev. Lett. **94**, 236803 (2005).

[11] T. Fujita, T. A. Baart, Chr. Reichl, W. Wegscheider, and L. M. K. Vandersypen, Coherent shuttle of electron-spin states, npj Quantum Inf. **3**, 1 (2017).

[12] S. P. Harvey, C. G. L. Bøttcher, L. A. Orona, S. D. Bartlett, A. C. Doherty, and A. Yacoby, Coupling two spin qubits with a high-impedance resonator, Phys. Rev. B **97**, 235409 (2018).

[13] P. Cerfontaine, R. Otten, M. A. Wolfe, P. Bethke, and H. Bluhm, A high-fidelity gateset for exchange-coupled singlet-triplet qubits, Phys. Rev. B **101**, 155311 (2020).

[14] Y. Tomita and K. M. Svore, Low-distance surface codes under realistic quantum noise, Phys. Rev. A **90**, 062320 (2014).

[15] T. E. O'Brien, B. Tarasinski, and L. DiCarlo, Density-matrix simulation of small surface codes under current and projected experimental noise, npj Quantum Inf. **3**, 39 (2017).

[16] 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).

[17] P. Aliferis, F. Brito, D. P. DiVincenzo, J. Preskill, M. Steffen, and B. M. Terhal, Fault-tolerant computing with biased-noise

superconducting qubits: A case study, New J. Phys. **11**, 013061 (2009).

[18] P. Aliferis and J. Preskill, Fault-tolerant quantum computation against biased noise, Phys. Rev. A **78**, 052331 (2008).

[19] D. Loss and D. P. DiVincenzo, Quantum computation with quantum dots, Phys. Rev. A **57**, 120 (1998).

[20] M. A. Fogarty, K. W. Chan, B. Hensen, W. Huang, T. Tanttu, C. H. Yang, A. Laucht, M. Veldhorst, F. E. Hudson, K. M. Itoh, D. Culcer, T. D. Ladd, A. Morello, and A. S. Dzurak, Integrated Silicon qubit platform with single-spin addressability, exchange control and single-shot singlet-triplet readout, Nat. Commun. **9**, 4370 (2018).

[21] J. R. Petta, A. C. Johnson, J. M. Taylor, E. A. Laird, A. Yacoby, M. D. Lukin, C. M. Marcus, M. P. Hanson, and A. C. Gossard, Coherent manipulation of coupled electron spins in semiconductor quantum dots, Science **309**, 2180 (2005).

[22] R. Hanson, L. P. Kouwenhoven, J. R. Petta, S. Tarucha, and L. M. K. Vandersypen, Spins in few-electron quantum dots, Rev. Mod. Phys. **79**, 1217 (2007).

[23] H. Bluhm, S. Foletti, D. Mahalu, V. Umansky, and A. Yacoby, Enhancing the Coherence of A Spin Qubit by Operating it As a Feedback Loop that Controls its Nuclear Spin Bath, Phys. Rev. Lett. **105**, 216803 (2010).

[24] P. Cerfontaine, T. Botzem, D. P. DiVincenzo, and H. Bluhm, High-Fidelity Single-Qubit Gates for Two-Electron Spin Qubits in GaAs, Phys. Rev. Lett. **113**, 150501 (2014).

[25] P. Cerfontaine, T. Botzem, J. Ritzmann, S. S. Humpohl, A. Ludwig, D. Schuh, D. Bougeard, A. D. Wieck, and H. Bluhm, Closed-loop control of a GaAs-based singlet-triplet spin qubit with 99.5% gate fidelity and low leakage, Nat. Commun. **11**, 4144 (2020).

[26] O. E. Dial, M. D. Shulman, S. P. Harvey, H. Bluhm, V. Umansky, and A. Yacoby, Charge Noise Spectroscopy Using Coherent Exchange Oscillations in A Singlet-Triplet Qubit, Phys. Rev. Lett. **110**, 146804 (2013).

[27] F. K. Malinowski, F. Martins, P. D. Nissen, E. Barnes, Ł. Cywiński, M. S. Rudner, S. Fallahi, G. C. Gardner, M. J. Manfra, C. M. Marcus, and F. Kuemmeth, Notch filtering the nuclear environment of a spin qubit, Nat. Nanotechnol. **12**, 16 (2016).

[28] T. Nakajima, M. R. Delbecq, T. Otsuka, P. Stano, S. Amaha, J. Yoneda, A. Noiri, K. Kawasaki, K. Takeda, G. Allison, A. Ludwig, A. D. Wieck, D. Loss, and S. Tarucha, Robust Single-Shot Spin Measurement with 99.5% Fidelity in A Quantum dot Array, Phys. Rev. Lett. **119**, 017701 (2017).

[29] C. Barthel, D. J. Reilly, C. M. Marcus, M. P. Hanson, and A. C. Gossard, Rapid Single-Shot Measurement of a Singlet-Triplet Qubit, Phys. Rev. Lett. **103**, 160503 (2009).

[30] Ł. Cywiński, R. M. Lutchyn, C. P. Nave, and S. Das Sarma, How to enhance dephasing time in superconducting qubits, Phys. Rev. B **77**, 174509 (2008).

[31] T. Botzem, M. D. Shulman, S. Foletti, S. P. Harvey, O. E. Dial, P. Bethke, P. Cerfontaine, R. P. G. McNeil, D. Mahalu, V. Umansky, A. Ludwig, A. Wieck, D. Schuh, D. Bougeard, A. Yacoby, and H. Bluhm, Tuning Methods for Semiconductor Spin Qubits, Phys. Rev. Appl. **10**, 054026 (2018).

[32] K. Takeda, A. Noiri, J. Yoneda, T. Nakajima, and S. Tarucha, Resonantly Driven Singlet-Triplet Spin Qubit in Silicon, Phys. Rev. Lett. **124**, 117701 (2020).

[33] D. J. Reilly, J. M. Taylor, E. A. Laird, J. R. Petta, C. M. Marcus, M. P. Hanson, and A. C. Gossard, Measurement of Temporal Correlations of the Overhauser Field in a Double Quantum Dot, Phys. Rev. Lett. **101**, 236803 (2008).

[34] T. F. Watson, S. G. J. Philips, E. Kawakami, D. R. Ward, P. Scarlino, M. Veldhorst, D. E. Savage, M. G. Lagally, M. Friesen, S. N. Coppersmith, M. A. Eriksson, and L. M. K. Vandersypen, A programmable two-qubit quantum processor in Silicon, Nature (London) **555**, 633 (2018).

[35] G. Zheng, N. Samkharadze, M. L. Noordam, N. Kalhor, D. Brousse, A. Sammak, G. Scappucci, and L. M. K. Vandersypen, Rapid gate-based spin read-out in Silicon using an on-chip resonator, Nat. Nanotechnol. **14**, 742 (2019).

[36] W. Huang, C. H. Yang, K. W. Chan, T. Tanttu, B. Hensen, R. C. C. Leon, M. A. Fogarty, J. C. C. Hwang, F. E. Hudson, K. M. Itoh, A. Morello, A. Laucht, and A. S. Dzurak, Fidelity benchmarks for two-qubit gates in Silicon, Nature (London) **569**, 532 (2019).

[37] J. Yoneda, K. Takeda, T. Otsuka, T. Nakajima, M. R. Delbecq, G. Allison, T. Honda, T. Kodera, S. Oda, Y. Hoshi, N. Usami, K. M. Itoh, and S. Tarucha, A quantum-dot spin qubit with coherence limited by charge noise and fidelity higher than 99.9%, Nat. Nanotechnol. **13**, 102 (2017).

[38] A. Hollmann, T. Struck, V. Langrock, A. Schmidbauer, F. Schauer, K. Sawano, H. Riemann, N. V. Abrosimov, D. Bougeard, and L. R. Schreiber, Large and Tunable Valley Splitting in ^{28}Si /SiGe Quantum Dots, Phys. Rev. Applied **13**, 034068 (2020).

[39] S. Schaal, I. Ahmed, J. A. Haigh, L. Hutin, B. Bertrand, S. Barraud, M. Vinet, C.-M. Lee, N. Stelmashenko, J. W. A. Robinson, J. Y. Qiu, S. Hacohen-Gourgy, I. Siddiqi, M. F. Gonzalez-Zalba, and J. J. L. Morton, Fast Gate-Based Readout of Silicon Quantum Dots using Josephson Parametric Amplification, Phys. Rev. Lett. **124**, 067701 (2020).

[40] D. P. DiVincenzo and P. Aliferis, Effective Fault-Tolerant Quantum Computation with Slow Measurements, Phys. Rev. Lett. **98**, 020501 (2007).

[41] H. E. Ercan, J. Ghosh, D. Crow, V. N. Premakumar, R. Joynt, M. Friesen, and S. N. Coppersmith, Measurement-free implementations of small-scale surface codes for quantum-dot qubits, Phys. Rev. A **97**, 012318 (2018).

[42] P. K. Sarvepalli, A. Klappenecker, and M. Rötteler, Asymmetric quantum codes: constructions, bounds and performance, Proc. R. Soc. A **465**, 1645 (2009).

[43] C. Volk, A. M. J. Zwerver, U. Mukhopadhyay, P. T. Eendebak, C. J. van Diepen, J. P. Dehollain, T. Hensgens, T. Fujita, C. Reichl, W. Wegscheider, and L. M. K. Vandersypen, Loading a quantum-dot based Qubyte register, npj Quantum Inf. **5**, 29 (2019).

[44] T. Ito, T. Otsuka, S. Amaha, M. R. Delbecq, T. Nakajima, J. Yoneda, K. Takeda, G. Allison, A. Noiri, K. Kawasaki, and S. Tarucha, Detection and control of charge states in a quintuple quantum dot, Sci. Rep. **6**, 39113 (2016).

[45] D. M. Zajac, T. M. Hazard, X. Mi, E. Nielsen, and J. R. Petta, Scalable Gate Architecture for a One-Dimensional Array of Semiconductor Spin Qubits, Phys. Rev. Appl. **6**, 054013 (2016).

[46] D. J. Reilly, C. M. Marcus, M. P. Hanson, and A. C. Gossard, Fast single-charge sensing with a RF quantum point contact, Appl. Phys. Lett. **91**, 89 (2007).

[47] The reason not to examine the decoder's performance where the number of phase-flip cycles is fixed to a constant c, while the number of qubits n in the code and its distances increases, is that one can show that in that case the decoder will certainly be performing worse for increasing n. The reason is that she is more likely to interpret a connected string of qubit errors of weight $>2c$ as two separate sequences of measurement errors which terminate in any unknown future measurement record. In other words, she will not infer any corrections for such errors whereas these errors will become prominent with increasing n. If the decoder performs worse with increasing $n = d$, it implies that there is no gain in using codes with larger distance d.

[48] B. Tarasinski and V. Ostroukh, https://gitlab.com/quantumsim/quantumsim.

[49] J. Edmonds, Paths, trees, and flowers, Canadian J. Math. **17**, 449 (1965).

[50] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, Topological quantum memory, J. Math. Phys. **43**, 4452 (2002).

[51] Ch. Wang, J. Harrington, and J. Preskill, Confinement-Higgs transition in a disordered gauge theory and the accuracy threshold for quantum memory, Ann. Phys. (NY) **303**, 31 (2003).

[52] N. Kawashima and T. Aoki, Zero-temperature critical phenomena in two-dimensional spin glasses, arXiv:cond-mat/9911120.

[53] K. Takeda, T. Sasamoto, and H. Nishimori, Exact location of the multicritical point for finite-dimensional spin glasses: A conjecture, J. Phys. A: Math. Gen. **38**, 3751 (2005).

[54] A. A. Hagberg, D. A. Schult, and P. J. Swart, Exploring network structure, dynamics, and function using networkX, in *Proceedings of the Seventh Python in Science Conference*, edited by G. Varoquaux, T. Vaught, and J. Millman (NetworkX, Pasadena, CA, USA, 2008), pp. 11–15.

[55] Note that the infidelity components do not quite add up to the total infidelity in Table IV. We overestimate by concatenating the noise processes, because fidelity is not a linear function and averaging over slow and fast noise together is not necessarily the same as averaging over them separately and concatenating the two processes. Using averaged noise channels for fast gates saves computational effort; the price is that we are potentially overestimating the noise compared to joint averaging.

[56] S. Bravyi, M. Englbrecht, R. König, and N. Peard, Correcting coherent errors with surface codes, npj Quantum Inf. **4**, 1 (2018).

[57] A. Katabarwa and M. R. Geller, Logical error rate in the Pauli twirling approximation, Sci. Rep. **5**, 1 (2015).

[58] M. Gutiérrez, C. Smith, L. Lulushi, S. Janardan, and K. R. Brown, Errors and pseudothresholds for incoherent and coherent noise, Phys. Rev. A **94**, 042338 (2016).

[59] P. Aliferis and B. M. Terhal, Fault-tolerant quantum computation for local leakage faults, Quantum Inf. Comput. **7**, 139 (2007).

[60] A. G. Fowler, Coping with qubit leakage in topological codes, Phys. Rev. A **88**, 042308 (2013).

[61] B. M. Varbanov, F. Battistel, B. M. Tarasinski, V. P. Ostroukh, T. E. O'Brien, L. DiCarlo, and B. M. Terhal, Leakage detection for a transmon-based surface code, arXiv:2002.07119 (2020).

[62] S. Mehl, H. Bluhm, and D. P. DiVincenzo, Fault-tolerant quantum computation for singlet-triplet qubits with leakage errors, Phys. Rev. B **91**, 085419 (2015).

[63] H. Bluhm, S. Foletti, I. Neder, M. Rudner, D. Mahalu, V. Umansky, and A. Yacoby, Dephasing time of GaAs electron-spin qubits coupled to a nuclear bath exceeding 200 μs, Nat. Phys. **7**, 109 (2010).

[64] L. Geck, A. Kruth, H. Bluhm, S. van Waasen, and S. Heinen, Control electronics for semiconductor spin qubits, Quantum Sci. Technol. **5**, 015004 (2020).

[65] S. Mehl, H. Bluhm, and D. P. DiVincenzo, Two-qubit couplings of singlet-triplet qubits mediated by one quantum state, Phys. Rev. B **90**, 045404 (2014).

[66] T. A. Baart, T. Fujita, Ch. Reichl, W. Wegscheider, and L. M. K. Vandersypen, Coherent spin-exchange via a quantum mediator, Nat. Nanotech. **12**, 26 (2017).

[67] 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).

[68] S. Puri, L. St-Jean, J. A. Gross, A. Grimm, N. E. Frattini, P. S. Iyer, A. Krishna, S. Touzard, L. Jiang, A. Blais, S. T. Flammia, and S. M. Girvin, Bias-preserving gates with stabilized cat qubits, arXiv:1905.00450 (2019).

[69] One can understand this by tracking the XX stabilizers through the CNOT as well as the logical operators.

[70] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Surface codes: Towards practical large-scale quantum computation, Phys. Rev. A **86**, 032324 (2012).

[71] E. Paladino, Y. M. Galperin, G. Falci, and B. L. Altshuler, $1/f$ noise: Implications for solid-state quantum information, Rev. Mod. Phys. **86**, 361 (2014).

[72] P. Cerfontaine, T. Botzem, S. S. Humpohl, D. Schuh, D. Bougeard, and H. Bluhm, Feedback-tuned noise-resilient gates for encoded spin qubits, arXiv:1606.01897 (2016).