

Magic State Injection with Erasure Qubits

Shoham Jacoby^{1,2,*} Yotam Vaknin^{1,2} Alex Retzker^{1,2} and Arne L. Grimsmo¹

¹*AWS Center for Quantum Computing, Pasadena, California 91125, USA*

²*Racah Institute of Physics, The Hebrew University of Jerusalem, Givat Ram, 91904 Jerusalem, Israel*



(Received 27 April 2025; revised 21 July 2025; accepted 4 September 2025; published 4 November 2025)

Erasure qubits constitute a promising approach for tackling the daunting resources required for fault-tolerant quantum computing. By heralding of erasure errors, both the error-correction threshold and the sub-threshold scaling of the logical error rate are significantly improved. While previous research has focused primarily on fault-tolerant quantum memories, we extend this investigation to magic state injection—a critical yet resource-intensive component of fault-tolerant quantum computation. We show that, after postselection on erasures, the logical error rate of the injected magic state is set by the residual Pauli error, while the space-time overhead is only marginally increased as compared with the case of non-erasure qubits with a similar noise strength. These conclusions hold both for injection into the surface code and for injection and cultivation on the color code. For the former, we show that most of the gains can be achieved by the use of just three strategically placed erasure qubits in the surface code patch, independent of the patch size. For the latter, in contrast, it is beneficial to have all the qubits in the cultivation patch be erasure qubits. Our results for cultivation suggest that algorithmically relevant logical error rates may be within reach *without* magic state distillation for erasure rates $\lesssim 4 \times 10^{-3}$ and residual Pauli error rates of $\sim 10^{-4}$.

DOI: [10.1103/5q1v-ry3y](https://doi.org/10.1103/5q1v-ry3y)

I. INTRODUCTION

One of the foremost challenges in realizing universal quantum computation is the reliable implementation of non-Clifford operations, such as the T gate, within quantum error correction codes [1,2]. Magic-state-based approaches are a leading strategy for enabling these crucial gates in fault-tolerant architectures [3,4]. However, magic states can often not be created fault-tolerantly and instead must be injected into the code and distilled to higher fidelity [3–7].

The fidelity with which these states are created has a large impact on the resource cost of the subsequent distillation process. For example, in the 15-to-1 distillation process [4,7], an injection error rate p leads to an output error rate of $\sim p^3$ after distillation. This means that an order-of-magnitude reduction in the injection infidelity results in a 3-orders-of-magnitude reduction in distillation fidelity. In particular, the difference in overhead for magic state injection with $p \sim 10^{-3}$ vs $p \sim 10^{-4}$ can be striking,

because it may determine whether one or multiple distillation rounds are necessary for near-term fault-tolerant applications [7]. Similar conclusions hold for the recently introduced “magic state cultivation” protocol, where physical error rates of $\sim 10^{-4}$ bring near-term applications within reach *without* distillation [8].

Magic state injection protocols have been improved over a series of studies [5,8–14]. Nevertheless, injection is fundamentally limited by the error rate of the underlying physical qubits. This limitation motivates the search for new qubit platforms that naturally support higher-fidelity operations, or, as is the case for erasure qubits, offer mechanisms to detect errors more efficiently.

Erasure qubits have emerged as a promising candidate for the implementation of fault-tolerant quantum computation. In erasure-based architectures, qubits can signal whether an error has occurred, effectively converting unknown errors into heralded erasure errors. This additional information enables more targeted error-correction strategies, leading to higher thresholds for fault-tolerant memories [15,16]. Potential realizations of erasure qubits include neutral atoms [15], trapped ions [17], and superconducting qubits [16,18], with recent experimental demonstrations [19–22].

In this work, we extend magic state injection circuits on the surface [9–11,23–25] and color code [8,13] to erasure qubit implementations. We present theoretical arguments

*Contact author: shohamjac@gmail.com

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

backed by extensive numerical analysis. Many of our conclusions are of a general nature and should also apply to other injection protocols. When post-selection is done on erasure events, the logical injection error is independent of the erasure error rate, and is set by the (assumed much lower) residual Pauli error rate within the computational subspace of the erasure qubits. We show that this conclusion continues to hold to a good approximation for noisy erasure detection at realistic error rates.

Moreover, there is only a marginal reduction in the injection acceptance rate if we compare erasure qubits with non-erasure qubits and set the erasure rate of the former equal to the Pauli error rate of the latter. In other words, the conversion of Pauli errors to heralded erasure errors comes at minimal cost in terms of the number of retries before a magic state is accepted. To support our analytical argument, we conduct a broad numerical study. A second benefit comes from the well-known improved error-correction performance associated with erasure qubits [15,16], which reduces the cost of expansion to a larger code after the magic state has been injected, making it practical to perform injection on smaller distance codes.

In contrast to quantum error correction, where frequent erasure checks and resets are required to gain a substantial advantage over non-erasure qubits [26], the post-selected nature of state injection means we can check for erasures only at the end of the injection. Alternatively, a practical choice which we use in our numerical analysis is to check once per syndrome round during the injection, in parallel with ancillary measurements. Less frequent erasure checks may relax hardware requirements, as erasure detection and reset can be slow in practice. The challenges associated with rapid erasure detection can perhaps be avoided, while still retaining very large overhead reduction for magic state factories, by the use of specialized patches with erasure qubits for injection, with magic states subsequently transported to patches of non-erasure qubits.

We show that almost all the benefit of using erasure qubits in the injection step is retained by placing only a small, constant number of erasure qubits at strategic locations in the code patch. For surface code injection, as few as three erasure qubits are required, independent of the desired size of the final encoded magic state. For the “cultivation” step following the initial injection in the recently introduced “magic state cultivation” protocol [8], the situation is a bit different. There, it does pay to make most of the qubits in the patch into erasure qubits, as we show in Sec. V.

We demonstrate the advantages of erasure qubits through numerical simulations using the injection protocol presented in Refs. [10,11] on the surface code, as well as the recently introduced injection and cultivation protocol on the color code [8]. We compare erasure qubits with differing erasure rates with non-erasure qubits. The results reveal that even at relatively high erasure rates and with

noisy erasure detection, there is a significant advantage for injection using erasure qubits.

II. ERASURE QUBITS

Erasure qubits are a type of qubit that converts a specific noise mechanism to a heralded error. Erasure can be detected by measuring the qubit in a state outside its computational subspace [15,16,21,22], or by use of, an ancillary qubit, that flags erasure events while the data qubits remain in the computational subspace [18].

Erasure qubits become beneficial when erasures are the dominant error mechanism in the system and undetectable errors are heavily suppressed. In the limit where erasure dominates, the surface code error-correcting threshold is close to 5% erasure per controlled NOT (CNOT) gate [15,16,26], and the sub-threshold logical error rate has better scaling with distance [15,26]. More concretely, for small physical error rates we expect the logical error to be exponentially suppressed with distance as [26–29]

$$p_L \propto q^{\alpha d}, \quad (1)$$

where q approaches a linear function of the physical error rates in the low-noise regime ($q \ll 1$) and $q = 1$ marks the threshold. While for non-erasure qubits we have $\alpha = 1/2$, the scaling α is generally increased for erasure qubits, and approaches 1 in the limit where erasures dominate [26].

Multiple experiments have demonstrated that one can engineer a significant bias where undetectable errors within the computational subspace are far less likely than erasure errors. For superconducting transmon qubits, the recently introduced dual-rail qubit [16,22] suffers mostly from energy-relaxation noise, which is converted to erasure. Meanwhile, dephasing, which manifests itself as dephasing error, is suppressed by a stable energy gap due to a capacitive coupling of the two transmons comprising the dual rail. Superconducting cavity dual-rail qubits [18,30,31], which are motivated by the intrinsically low dephasing of the cavity’s linear oscillator modes, allow the detection of cavity decay, the dominant source of noise in the system. Lastly, for certain cold atom platforms, such as ^{171}Yb and ^{88}Sr , decay from the Rydberg state is significantly biased toward identifiable states, allowing the detection of more than 90% of decay events during a two-qubit (2Q) gate [15,32].

III. MAGIC STATE INJECTION

The Eastin-Knill theorem [33] states that universal quantum computation cannot be implemented transversely on a quantum error-correcting code. To overcome this limitation, researchers have developed various schemes, including code switching and gate teleportation.

Gate teleportation requires specific input states that typically cannot be prepared fault-tolerantly. In practice,

this challenge is addressed through a two-step process: (1) Magic state injection [9–11,34], where many noisy encoded states are generated in a non-fault-tolerant manner, and (2) magic state distillation [3], where noisy states are distilled into fewer high-fidelity states suitable for gate teleportation.

Magic state injection involves encoding a magic state in a low-distance (d_1) code and subsequently expanding it to a code with a larger distance (d_2). This method is particularly vulnerable to errors during the initial stages of the process. These early errors can propagate through the system and persist, regardless of the final code size.

Here we consider the surface code magic state injection protocol described in Refs. [9–11], but we emphasize that many of our conclusions are general and should apply to other injection schemes as well. Color code cultivation is presented in Sec. V. The protocol proceeds as follows:

- (1) A noisy magic state is prepared in a d_1 surface code. We exemplify this step by implementing both the scheme of Lao and Criger [10], building on previous work by Li [9], which we refer to as “Lao-Criger injection,” and the scheme of Gidney [11], which we refer to as “hook injection.” Lao-Criger injection allows the injection of an arbitrary state, while hook injection allows the injection of states confined to the X - Y plane ($\langle \psi | Z | \psi \rangle = 0$). We present both injection protocols to emphasize that our results are not specific to a particular injection scheme, but apply to magic state injection in general. This step includes initialization of the data qubits and performing one round of stabilizer extraction. A subset of stabilizers are deterministic in the absence of errors, and the process is restarted if any of these give an unexpected outcome.
- (2) Stabilizer measurements are performed on the distance d_1 surface code for another $r - 1$ rounds. If any stabilizer measurement flips in subsequent rounds, the state is discarded and the process restarts. In earlier work $r = 2$ was typically used [9,10], but in Ref. [11] it was shown that for large values of d_1 , increasing r is beneficial.
- (3) The surface code is finally expanded to a larger distance d_2 , followed by d_2 rounds of standard error correction. The final distance, d_2 , is chosen to be large enough to support the desired error rate of the subsequent distillation protocol [7].

The logical error rate has a contribution from undetected low-weight errors in the first two steps, and topologically non-trivial errors in the third step. In Fig. 1 we illustrate the first step of the hook injection protocol for the case of a $d_1 = 2$ surface code, and we present a full analysis of the two injection strategies in Appendix B.

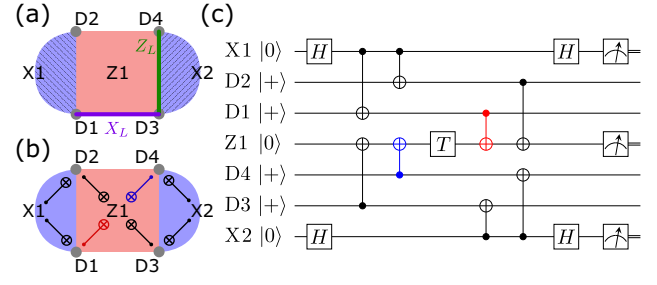


FIG. 1. First step of the hook injection circuit from Ref. [11] for a rotated surface code of distance $d_1 = 2$. (a) Physical layout of the four data qubits D1–D4, the two X stabilizers (blue), and the Z stabilizer (pink). The hatched stabilizers are deterministic in the first round. The logical operators Z_L and X_L are depicted in green and purple, respectively. (b) CNOT gate locations over the surface code patch. (c) Circuit implementing the first round of the hook injection. Some physical Pauli errors on the T gate and blue and red CNOT gates can flip the logical operator without being detected, and thus their error probability has a linear contribution to the logical error rate. The T gate translates to a rotation of the Z_L logical operator, leading to the preparation of a $|T\rangle = T|+\rangle$ state (see Appendix B2). Injection distance $d_1 = 2$ was used for visual clarity; we note, however, that hook injection can be performed at any surface code distance.

During the first step, single fault locations can cause an undetected logical error (see, e.g., the T gate and the blue and red CNOT gates in Fig. 1) [35]. Consequently, the logical error rate has a linear contribution from the Pauli error probability of one-qubit (1Q) gates p_1 , 2Q gates p_2 , and initialization p_{IN} .

To quantify these contributions, we approximate the logical error rate achieved after the first two steps as follows:

$$p_{L,\text{inj}}(\bar{p}) \approx ap_1 + bp_2 + cp_{\text{IN}} + O(p^2). \quad (2)$$

with $\bar{p} \equiv (p_1, p_2, p_{\text{IN}})$. Here $a, b, c \in \mathbb{R}$ are coefficients derived from counting the number of undetected fault locations that flip the logical operators. Coefficients for hook injection and Lao-Criger injection are given in Appendix B. For a noise model with $p_{\text{IN}} = p_2 = p$, $p_1 = p/10$, and injection of the $|S\rangle = (|0\rangle + i|1\rangle)/\sqrt{2}$ state [which can be simulated by Clifford operations, unlike the magic state $|T\rangle = (|0\rangle + e^{i\pi/4}|1\rangle)/\sqrt{2}$] the logical error rate for hook injection $p_{\text{HI}} = 7/30p$ and for Lao-Criger injection $p_{\text{LI}} = 46/30p$, showing the favorable performance of hook injection.

A. Injection with erasure qubits

By implementing magic state injection using erasure qubits [16], we can use the additional information received from the erasures to improve the post-selection.

Erasure qubits split the noise into two parts: undetectable Pauli errors that contribute to the injection fidelity

in linear order, and erasure errors that are identifiable by use of erasure detection. Useful erasure qubits should convert dominant noise processes to erasures. We therefore assume that the probabilities e_1 , e_2 , and e_{IN} of erasure errors during 1Q gates, 2Q gates, and initialization, respectively, obey

$$e_2, e_1, e_{\text{IN}} \gg p_2, p_1, p_{\text{IN}} \quad (3)$$

for erasure qubits.

This allows us to reduce the logical error by discarding the state for every detected erasure in the first two steps, at the expense of lowering the injection acceptance rate. For perfect erasure detection, such that every erasure in the first round is identifiable, the logical error rate of the injection is strictly given only by the undetectable errors, i.e.,

$$p_{L,\text{inj}}(\bar{e}, \bar{p}) = p_{L,\text{inj}}(\bar{p}) \quad (4)$$

independent of $\bar{e}$ and with $p_{L,\text{inj}}(\bar{p})$ defined as before [see Eq. (2)]. This manifests itself a key advantage of using erasure qubits for magic state injection, as the logical error rate does not depend on the erasure rate.

On the other hand, discarding of states due to erasure errors directly affects the acceptance rate. However, the impact is not very large, if our comparison is with non-erasure qubits with a similar noise rate, as we show in the following.

The acceptance rate of the protocol can be broken down into two parts: the probability that no erasure has occurred $R_e(\bar{e})$, and the probability that no Pauli error was detected $R_p(\bar{p})$. The overall acceptance rate of an injection circuit discarding erasure errors is then the product

$$R(\bar{e}, \bar{p}) = R_e(\bar{e}) R_p(\bar{p}). \quad (5)$$

It is natural to compare the performance of a protocol with erasure qubits with that of a protocol using non-erasure qubits, where the Pauli error rate of the non-erasure qubits is comparable in magnitude to the erasure rate. This is motivated by the observation that physical implementations of erasure qubits “convert” the dominant physical error processes, which would normally lead to undetectable errors, into detectable erasures [19–22].

For clarity, we denote the Pauli error rates when using non-erasure qubits by $\bar{p}_n$. Numerical analysis suggests that most mechanisms contributing to Pauli errors are detected (Appendix D), and as such we get that for $\bar{p}_n = \bar{e}$ the acceptance rates due to erasure errors and due to Pauli errors are nearly identical, i.e., $R_e(\bar{e}) \approx R_p(\bar{p}_n)$. In addition, in the limit of interest where the residual Pauli error rate of the erasure qubits is very small [Eq. (3)], only a small fraction of the discarded states originates from residual Pauli errors, i.e., $1 - R_p(\bar{p}) \ll 1 - R_e(\bar{e})$. Consequently, the acceptance rate of erasure qubits is expected

to be nearly the same as that of non-erasure qubits when $\bar{e} = \bar{p}_n$:

$$R(\bar{e}, \bar{p}) \approx R_p(\bar{p}_n = \bar{e}). \quad (6)$$

In practice, implementation of erasure qubits may come at the cost of increased complexity compared with implementation of non-erasure qubits in the same physical platform, making an apples-to-apples comparison challenging. In particular, erasure qubits may come at the cost of additional hardware, more frequent measurements and resets, and slower gates. Although the resulting erasure errors are more benign, this added complexity can lead to higher error rates relative to those for their non-erasure counterparts. For example, in dual-rail systems, the excitation number averaged over the computational basis states is increased by a factor of roughly 2 as compared with that for a qubit encoded in the zero- and one-excitation states of a single mode [36]. The effective erasure rate is therefore approximately twice the decay rate of a conventional, single-mode qubit when the same hardware is used.

In this work, we consider three scenarios for comparison: an almost perfect “conversion” of Pauli errors to erasure errors $\bar{e} = \bar{p}_n$, and two values for non-perfect “conversion,” $\bar{e} = 4\bar{p}_n$ and $\bar{e} = 10\bar{p}_n$. We note that certain architectures, such as tunable transmons and cold atoms, may be able to achieve $\bar{e} \approx \bar{p}_n$ [15,16], depending on the gate implementation. Additionally, we assume that erasure qubits retain some residual Pauli error p , which we take to be an order of magnitude smaller than for non-erasure qubits in our numerics, $p = p_n/10$.

Figure 2 presents a comparison of injection acceptance rates and logical error rates for non-erasure and erasure qubits, at different ratios of $\bar{p}_n$ to $\bar{e}$. We use a noise model where reset, measurement, and CNOT gates carry a probability p_n (p) of inducing a Pauli error on the participating non-erasure (erasure) qubits, while other single-qubit gates and idling carry a reduced error probability of $p_n/10$ ($p/10$). For the erasure qubits, we assume equal erasure probability for reset, measurement, and CNOT operations e , while other single-qubit operations have a reduced erasure probability $e/10$. Idling occurs at every stabilizer extraction round and affects all qubits not currently engaged in measurement. Erasure detection is performed at the end of every stabilizer extraction round in parallel with ancilla measurements. An erased qubit that interacts with another qubit in a two-qubit gate induces a fully depolarizing channel on the other qubit. See Appendix A for further details of the noise model.

From Fig. 2, we see that for the $e = p_n$ scenario, the injection acceptance rate with erasure qubits is reduced only marginally compared with that with non-erasure qubits, as we anticipated in Eq. (6). However, the logical error rate reduces significantly. As expected the logical rate depends only on the residual error rate p when erasure

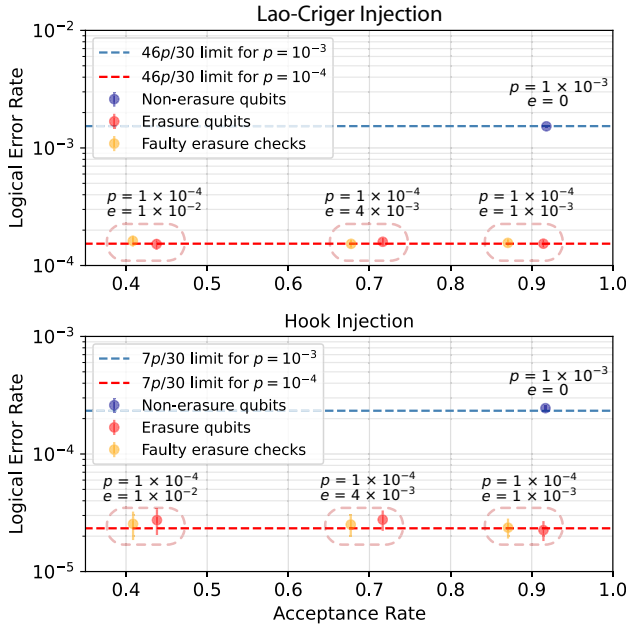


FIG. 2. Logical error rate and acceptance rate for ungrown injection with $d_1 = 3$ for Lao-Criger injection and hook injection of an $|S\rangle = S|+\rangle$ state. Points represent non-erasure ($e = 0, p_n = 10^{-3}$), almost perfect erasure conversion ($e = 10^{-3}, p = 10^{-4}$), and two values of non-perfect erasure conversion rates ($e = 4 \cdot 10^{-3}, p = 10^{-4}$ and $e = 10^{-2}, p = 10^{-4}$). The theoretical limits [Eq. (2)] for non-erasure qubits and erasure qubits are depicted by the dashed blue and red lines, respectively. Each erasure scenario is calculated for perfect erasure detection and for erasure detection error rate $e_{FP} = e_{FN} = 10^{-2}$ for false-positive and false-negative detections, respectively.

detection is perfect [Eq. (4)]. This observation essentially indicates that post-selection on an event results in an error rate that is independent of the probability of that event [14,37].

1. Faulty erasure detection

For imperfect erasure detection, false positive and false negative detections contribute differently. False positive erasure detections lead only to unnecessary discards of potentially valid states, reducing the protocol's acceptance rate. On the other hand, false negative detections can contribute to the logical error rate. However, the latter is mitigated by three factors: first, this is a second-order effect where an erasure first happens and then goes undetected, such that the probability of false negatives is already quite low; second, these undetected erasure errors generate depolarizing noise on both themselves and interacting qubits in our model, making them likely to trigger a stabilizer detection event; third, they must evade detection across all r rounds.

To balance the competing effects between false positive and false negative erasure detections, we choose to discard the state only if we detect two erasures. This approach

reduces unnecessary discards from false positives, as a single erroneous detection will not trigger abortion. While this could potentially allow some true erasure errors to go undetected (false negatives), we find that this is unlikely in practice, as undetected erasure errors are likely to trigger stabilizer detection events as they stay erased and inject noise into the system over multiple rounds. We have found in our numerical analysis that this trade-off is favorable, as the reduction in false-positive-induced discards outweighs the minimal increase in undetected errors.

As demonstrated in Fig. 2, introduction of significant false negative and false positive rates $e_{FN} = e_{FP} = 10^{-2}$ with this scheme results in a minimal effect on the logical error rate, and only a marginal reduction of the acceptance rate. Equations (4) and (6) are thus reasonably good approximations even for noisy erasure detection.

B. Code expansion

In the third step we expand the code from distance d_1 to distance d_2 , where d_2 is the desired final distance of the surface code. The overall error rate of the injection protocol has contributions from errors in all three steps, which can be approximately expressed as

$$p_L \approx p_{L,\text{inj}} + p_{L,\text{exp}}, \quad (7)$$

with $p_{L,\text{inj}}$ and $p_{L,\text{exp}}$ denoting the logical error rates in the injection and expansion steps, respectively.

The logical error rate of the injection process is limited by the linear contributions from low-weight errors in the initial step [Eq. (2)]. The expansion step should ideally exhibit a significantly lower error rate than the injection process.

During the first stabilizer extraction round in the expansion process, the code still exhibits fault distance of the smaller distance d_1 code, and we thus expect the logical error $p_{L,\text{exp}}$ to be dominated by errors happening in this round. For low physical error rates, we thus expect $p_{L,\text{exp}}$ to follow Eq. (1) with $d = d_1$ and where q is a linear combination of e and p (see Appendix C for numerical fits to this Ansatz).

By simulating an expansion of a perfectly prepared state (i.e., one with no noise before the expansion round), we can isolate and quantify the logical error rate contributed solely by the expansion step, $p_{L,\text{exp}}$. Figure 3(a) illustrates these simulation results across various parameter regimes, demonstrating how expansion-induced errors scale with code distance d_1 and physical error rates, with fixed $d_2 = 15$. To compare erasure and non-erasure qubits, we use the same noise model as in Ref. III A. Near-perfect erasure conversion leads to significantly better scaling of the logical error rate with code distance, while imperfect conversions mitigate this improvement by bringing the effective error rate closer to the threshold (Appendix C).

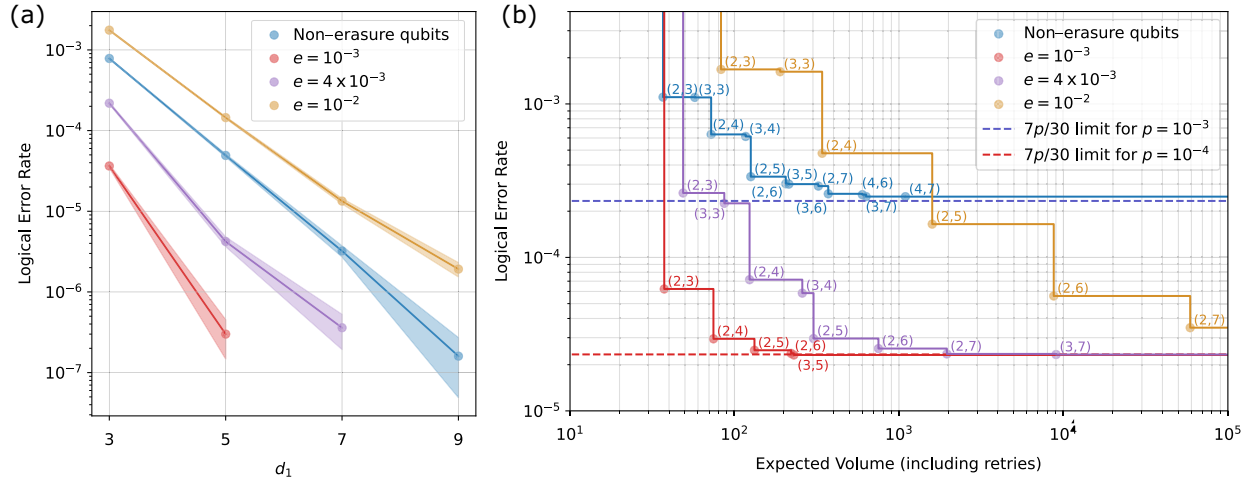


FIG. 3. (a) Logical error rate of the expansion from a distance d_1 to a $d_2 = 15$ surface code. The encoded d_1 surface code state is prepared noiselessly, allowing examination of the error contributions from the expansion step. The same erasure conversion scenarios as in Fig. 2 are analyzed: non-erasure qubits ($e = 0, p_n = 10^{-3}$), near-perfect erasure conversion ($e = 10^{-3}, p = 10^{-4}$), and two values of non-perfect erasure conversions ($e = 4 \cdot 10^{-3}, p = 10^{-4}$ and $e = 10^{-2}, p = 10^{-4}$). (b) Pareto front of the entire injection and expansion protocol for the same scenarios as in (a) (same color key). The points, each labeled with its corresponding (r, d_1) values, are obtained by scanning combinations of the number of post-selected rounds r and the injection distance d_1 , where $r \in \{2, 3, 4\}$ and $d_1 \in \{3, \dots, 7\}$. The theoretical limits [Eq. (2)] for non-erasure qubits and erasure qubits are depicted by the dashed blue and red lines, respectively.

To evaluate the effectiveness of the complete injection process, we simulate the entire protocol as described in Ref. [11]. The simulation begins with hook injection of an $|S\rangle$ state into a surface with code distance d_1 , followed by $r - 1$ rounds of stabilizer extraction, a single round of expansion to distance d_2 , and d_2 memory rounds, and ends with in-place Y basis measurement [38] to determine the injection success. In this section, we assume perfect erasure detection after every gate in order to simplify the decoding process. We refer the reader to Refs. [26,39] for detailed analysis of how sparse and non-perfect erasure detection can affect the results of a quantum memory protocol.

In Fig. 3(b), we plot the error rate and the expected volume of the entire injection circuit, including expansion. The expected volume is calculated as

$$V(d_1, r) = \frac{Q_{d_1} r}{R}, \quad (8)$$

with Q_{d_1} being the number of data and ancillary qubits in the distance d_1 surface code and R being the acceptance rate of the injection. The error rate has contributions from all of the injection steps, and therefore we can see that for lower d_1 values, the expansion is limited by the errors in the expansion round. For higher values of d_1 , the logical rate approaches the ideal logical error rate of $7/30p$ [Eq. (2)].

When comparing the logical error rates shown in Fig. 3(a) for the expansion step alone and Fig. 3(b) for the full injection process, we see that the limiting factor

depends on which stage contributes more significantly to the total error. If the dominant contribution arises from the expansion step because d_1 is not sufficiently large, then no matter how much the injection stage is improved, the final logical error rate remains constrained by the expansion process itself. On the other hand, if the injection step is the primary source of errors, the overall logical error rate quickly saturates at the theoretical limit of $7/30p$ [Eq. (2) and Appendix B].

IV. HYBRID ERASURE AND NON-ERASURE SYSTEM

In certain hardware implementations, such as superconducting qubits, realizing erasure-enabled qubits can be more expensive than realizing non-erasure qubits [22]. Motivated by this, we investigate a hybrid strategy that integrates both erasure and non-erasure qubits within the same surface code patch, deploying erasure qubits only at those circuit locations where they provide the most benefit.

The injection circuit is subject to two distinct types of fault location: undetectable faults that add directly to the logical error rate and detectable faults that cause the state to be discarded and the injection procedure to be restarted, thereby lowering the overall acceptance rate without affecting the logical error rate. In both the Lao-Criger scheme and the hook injection scheme, increasing the injection distance d_1 does not add more undetectable faults [10,11]. By introducing erasure qubits to circuit locations that contribute to the first (undetectable) error type, and using non-erasure qubits elsewhere, one can retain the reduced

logical error rate of erasure qubit injection while minimizing the number of required erasure qubits. Furthermore, in scenarios in which erasure qubits exhibit a higher error rate than non-erasure qubits, $e > p_n$, the reduced error rate associated with the second (detectable) type of error mechanism increases the protocol's acceptance rate by reducing the frequency of discards. In this section, we add a subscript to the residual Pauli error rates of the erasure qubits, p_e , to better distinguish them from the Pauli error rate of the non-erasure qubits in the same code patch, which we denote by p_n .

The resulting error rate injection on the hybrid patch preserves the linear contributions from p_e , as it is driven by the residual Pauli error of the erasure qubits participating in the “sensitive” circuit locations. Additionally, there is an extra contribution proportional to p_n^2 , which is negligible as long as $p_n^2 \ll p_e$:

$$p_{L,\text{inj,hybrid}}(p_e, p_n) = p_{L,\text{inj}}(p_e) + O(p_n^2). \quad (9)$$

For example, in the hook injection circuit depicted in Fig. 1, our proposed approach is implemented by selecting D1, D3, and Z1 to be erasure qubits, as they participate in the colored CNOT gates and in the T gate. These three qubits covers all fault locations that contribute linearly to the logical error rate, independent of the size of the surface code patch. These qubits cover all weight-1 fault locations regardless of the patch size (see Ref. [11] for details), thereby making our hybrid approach viable for any injection distance d_1 .

In some architectures, and specifically superconducting qubits, erasure qubits exhibit longer gate times, requiring the non-erasure qubits to remain idle during this period. We mimic such a scenario by comparison with an alternative noise model, inspired by the relaxation times of the superconducting dual-rail qubits and transmons (see Appendix E). This alternative noise model has a CNOT Pauli error rate for non-erasure qubits given by $p_2^* = (p_n + e_2)/2$. This alternative noise model is labeled “hybrid (with idling)” in Fig. 4.

We compare numerical results for distance-3 hook injection with non-erasure qubits, erasure qubits, and the hybrid approach—where only three qubits are erasure qubits—in Fig. 4. In Fig. 4(a), we show the acceptance rate for the different scenarios. In Fig. 4(b), we show the corresponding logical error rate as a function of the erasure conversion ratio e/p_n . It can be seen that the hybrid architecture shows a significant acceptance rate increase over all-erasure patch, while the logical error rate is essentially unaffected as long as $p_n^2 \ll p_e$, as expected. The improvement in acceptance rate becomes more pronounced as the erasure rate increases relative to the non-erasure Pauli rate, and is negligible when $e = p_n$, since the acceptance rates of the full-erasure and non-erasure patches are nearly equal, as explained in Sec. III A.

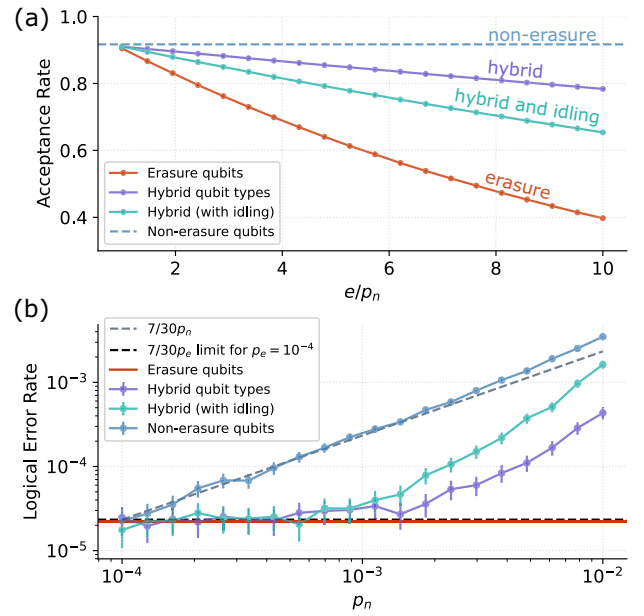


FIG. 4. (a) Acceptance rates and (b) logical error rates for distance-3 qubit patches for different architectures: all erasure qubits (red), all non-erasure qubits (blue), a hybrid design combining erasure and non-erasure qubits (purple), and a hybrid design with different gate times taken into account (green). A non-erasure error rate of $p_n = 10^{-3}$ is used in (a), and the residual Pauli error rate is fixed at $p_e = 10^{-4}$ throughout. An erasure rate of $e = 10^{-3}$ is used for (b). The results demonstrate that for practical non-erasure error rates ($p_n \sim 10^{-3}$), hybrid architectures maintain low logical error rates comparable to the erasure-only case, while achieving higher acceptance rates.

Thus, the hybrid approach is a particularly cost-effective way to reduce magic state overhead in architectures where high-quality erasure qubits can be realized, but are either more expensive or exhibit significantly higher erasure rates than non-erasure qubits.

V. INJECTION AND CULTIVATION ON THE COLOR CODE

Zero-level distillation [13,40–45] realizes a high-fidelity magic state by repeated measurements. Gidney *et al.* [8] recently introduced a novel zero-level distillation approach, dubbed “magic state cultivation.” This procedure comprises three steps: injection, cultivation, and grafting, where the latter is a code expansion step from a small color code to a hybrid color and surface code. Compared with the injection and expansion protocols discussed in the preceding two sections, the cultivation step is a fundamentally new ingredient. Vakinin *et al.* [46] demonstrated that the use of erasure qubits can increase cultivation efficiency. In this study, we analyze the benefits of the hybrid construction for cultivation.

The injection step prepares a magic state in a small color code, similarly to the surface code injection protocols discussed in Sec. III. We focus here on the “unitary injection”

from Ref. [8], which shares similarities with hook injection for the surface code, consisting of a state creation phase via unitary operation followed by a round of stabilizer measurement. This step results in a magic state that has a linear physical error rate contribution to the logical error rate:

$$p_{L,\text{color},\text{inj}} = \frac{2}{15}p_2 + \frac{1}{3}p_1, \quad (10)$$

with use of the same notation for one-qubit and two-qubit locations as previously.

The magic state cultivation protocol's primary advantage, however, emerges during the second step, cultivation, where we perform a fault-tolerant measurement of the logical $H_{XY} \equiv (X + Y)/\sqrt{2}$ operator on the color code. This critical measurement detects low-order errors that would otherwise dominate the logical error rate, substantially improving the overall protocol's effectiveness. By identifying and flagging these errors, the fault-detection mechanism significantly increases the fidelity of the resulting magic state beyond what the initial injection alone can achieve.

The arguments presented in Sec. III A, which lead to the conclusion that the logical fidelity is set by the residual Pauli error rate of the erasure qubits, Eq. (4), and that the acceptance rate is roughly equal to that of non-erasure qubits with the same noise strength, Eq. (6), are applicable to the first two steps, injection and cultivation, of the overall cultivation procedure. While the last step in cultivation uses gap-based post-selection, the first two steps are post-selected on any detector firing. In our erasure-based protocol, this corresponds to post-selection on all erasures. Accordingly, we simulate these steps and present the results in Fig. 5. In this section, for better comparison with Ref. [8], we adopt the same uniform noise model as used there. In this model, each operation [1Q gate, 2Q gate, and state preparation and measurement (SPAM)] is subject to a Pauli error with probability p and an erasure error with probability e . While these simulations do not incorporate erasure measurement errors, our analysis in Sec. III A suggests that failed erasure measurements would likely be mitigated by subsequent erasure and stabilizer measurements in the protocol. We anticipate this limitation would manifest itself only as a small reduction in overall performance [26].

In the final grafting step, the code is expanded into a larger-distance hybrid color and surface code. In this step, an abort criterion is determined by gap decoding [8]. As shown in Ref. [46], erasure qubits will give an advantage to this step as well, since erasure information can be used to refine the soft abortion criterion. On the other hand, in architectures where erasure qubits are costly compared with non-erasure qubits, it may also be pertinent to consider a hybrid approach where a small color code patch of erasure qubits is grafted into a larger patch of non-erasure

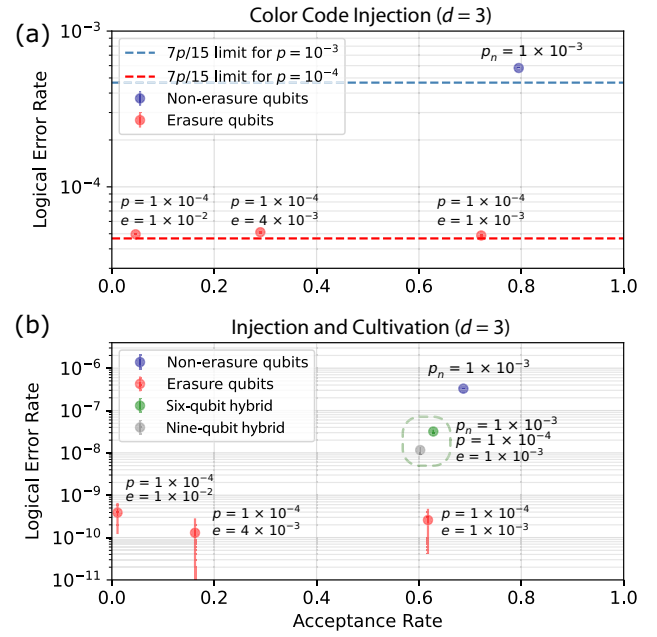


FIG. 5. Logical error rate and acceptance rate for (a) unitary injection and (b) unitary injection and cultivation for distance-3 color code, with use of the protocol from Ref. [8]. Points are for non-erasure ($e = 0, p_n = 10^{-3}$), almost perfect erasure conversion ($e = 10^{-3}, p = 10^{-4}$), and two values of non-perfect erasure conversions ($e = 4 \cdot 10^{-3}, p = 10^{-4}$ and $e = 10^{-2}, p = 10^{-4}$). Hybrid setups with, respectively, six erasure qubits and nine erasure qubits ($e = 10^{-3}, p_n = 10^{-3}, p = 10^{-4}$) are also presented. The theoretical limits [Eq. (10)] for non-erasure qubits and erasure qubits are depicted by the dashed blue and red lines, respectively, in (a).

qubits. We leave these more complex questions for future work.

To devise a hybrid erasure and non-erasure approach, we need to identify the fault locations that contribute most to logical errors. Unlike injection, the cultivation circuit does not have any single fault location capable of undetectably causing a logical error. Instead, logical errors arise through combinations of faults, each involving at least d_1 fault locations. For $d_1 = 3$, a computerized search reveals hundreds of thousands of such fault combinations, making it unfeasible to fully suppress errors by converting a small subset of qubits to erasure qubits, as was done for injection in Sec. IV. Nevertheless, we find that conversion of just six qubits into erasure qubits can suppress at least one fault in most of these combinations, resulting in more than a ten-fold reduction in the logical error rate. Further reduction is achieved by conversion of nine qubits into erasure qubits, fully covering at least one fault in every fault combination and reducing the error rate by a factor of 30. Both hybrid schemes are illustrated in Fig. 5 as the “six-qubit hybrid” and “nine-qubit hybrid” data points. However, the benefit

of these hybrid schemes is modest, as in contrast conversion of all 15 qubits to erasure qubits yields a significantly greater improvement—by a factor of roughly 10^3 .

VI. CONCLUSION

We have demonstrated that erasure qubits can significantly enhance magic state injection and magic state cultivation, by providing additional information that enables more effective post-selection. Specifically, the logical error rate can drastically reduce with minimal reduction in the acceptance rate, if dominant errors can be converted into detectable erasure errors. This improvement persists even when erasure rates are modestly higher than error rates in non-erasure qubits, and erasure detection itself is noisy.

For surface code injection, we moreover demonstrate that almost all the benefits of using erasure qubits can be achieved with the use of a hybrid system with only three strategically placed erasure qubits at the most impactful locations, offering an efficient alternative when full erasure qubit implementation proves cost-prohibitive or introduces significant performance overhead.

For color code cultivation, on the other hand, our analysis shows that it pays to make all the qubits of the injection and cultivation process into erasure qubits. In this case, for cultivation on a $d = 3$ color code, we find that erasure qubits with erasure rate $e \gtrsim 10^{-3}$ and residual error rate $p = 10^{-4}$ have an approximately 1000-fold reduction in logical error rate over non-erasure qubits with physical error rate $p_n = 10^{-3}$, at a marginal cost in acceptance rate.

The benefits of using erasure qubits become even more pronounced when we consider subsequent code expansion, which we study for injection into the surface code. We speculate that large gains can also be had for the “grafting” step of magic state cultivation—see also a discussion in Ref. [46]—but leave a quantitative study of this for future work. Moreover, the method presented here can be directly implemented to cultivation methods that utilize non-local connectivity [46,47]. Our results for the injection and cultivation steps are, however, indicative that extremely low logical error rates relevant for early fault-tolerant applications may be within reach with erasure rates of $\sim 10^{-3}$ and residual undetectable errors at the $\sim 10^{-4}$ level.

ACKNOWLEDGMENT

We thank the staff from across the AWS Center for Quantum Computing that enabled this project.

DATA AVAILABILITY

The data and the scripts used in this work are available in Ref. [48]. The data that support the findings of this article are openly available [46].

APPENDIX A: NOISE MODEL

In our simulations, we use a noise model in which two-qubit gates (CNOT gates) and SPAM operations carry a probability p of inducing a Pauli error, and non-SPAM one-qubit operations (idling and single-qubit gates) carry a probability $p/10$ of inducing a Pauli error. Idling occurs at every stabilizer extraction round in parallel with measurements, and affects all qubits not being measured. All Pauli errors are considered equally likely, i.e., the noise channels are fully depolarizing.

For erasure qubits, there is an additional probability e of erasure error occurring for two-qubit gates and SPAM operations, and a probability of $e/10$ for erasure during non-SPAM one-qubit operations. An erased qubit that interacts with another qubit induces a fully depolarizing channel on the other qubit. A measurement on an erased qubit gives a random outcome, while state preparation is assumed to correctly reinitialize a qubit, even if it was previously erased. In two-qubit gates, each qubit is erased with probability $1 - \sqrt{1 - e_2}$, leading to a total erasure rate of e_2 per gate.

APPENDIX B: INJECTION TYPES

1. Lao-Crigrer injection

The protocol in Ref. [10], which we refer to as “Lao-Crigrer injection,” is a version of the protocol presented by Li [9], adapted to the rotated surface code. It starts by initializing a qubit in the desired injection state, and measuring the stabilizers r times.

The initialized qubit can be either in the corner of the patch or in the middle of the patch. In this study, we focus exclusively on initializing the center qubit, as it exhibits a better error rate; see Ref. [10] for further details.

The logical error rate up to first order in the gates error is given by

$$p_{LI} = \frac{3}{5}p_2 + p_{IN} + \frac{2}{3}p_1 + O(p^2). \quad (B1)$$

This error rate results from the rate of initialization error affecting the middle qubit, a single-qubit gate error during the rotation of this qubit into the magic state, and nine particular two-qubit errors, each occurring with probability $p_2/15$ in three specific CNOT gates during the first round of stabilizer extraction.

The noise channel is expressed as

$$\begin{aligned} \mathcal{E}_{LI} = & (1 - p_{LI})[I] + (p_{IN} + \frac{2}{3}p_1)[F_L] \\ & + \frac{5}{15}p_2[X_L] + \frac{2}{15}p_2[Y_L] + \frac{2}{15}p_2[Z_L], \end{aligned} \quad (B2)$$

with $[F_L]$ being the channel that flips the state to an orthogonal state, and $[P_L]$ being the channel that applies the logical Pauli P on the state [9].

2. Hook injection

Hook injection was introduced by Gidney [11]. Instead of initializing a qubit in the desired injection state, it initializes a $|+\rangle$ state and rotates the logical operator. It does so by switching the order of the CNOT gates in the stabilizer extraction circuit, in a way that Z rotations on an ancilla rotate the logical operator.

This can be seen by propagating the Z -rotation operator to the start of the circuit (Fig. 6). The propagation itself can be derived by observing the relation

$$\begin{aligned} & \prod_{i=1}^n (\text{CNOT}_{ia}^\dagger) \cdot e^{i\theta Z_a} \cdot \prod_{j=1}^n (\text{CNOT}_{ja}) \\ &= e^{i\frac{\pi}{4} \sum_i (I_i - Z_i) \otimes (I_a - X_a)} \cdot e^{i\theta Z_a} \cdot e^{i\frac{\pi}{4} \sum_i (I_i - Z_i) \otimes (I_a - X_a)} \\ &= e^{i\theta Z_1 \dots Z_n Z_a} \end{aligned} \quad (\text{B3})$$

for a set of qubits indexed from 1 to n and ancilla indexed as a . If the ancilla is initialized in $|0\rangle$, we get the relation depicted in Fig. 6(a).

The hook injection has a lower logical error rate, contributed only from one error mechanism in the T gate application, and three error mechanisms in the following and preceding CNOT gates [blue and red in Fig. 6(b)]. This

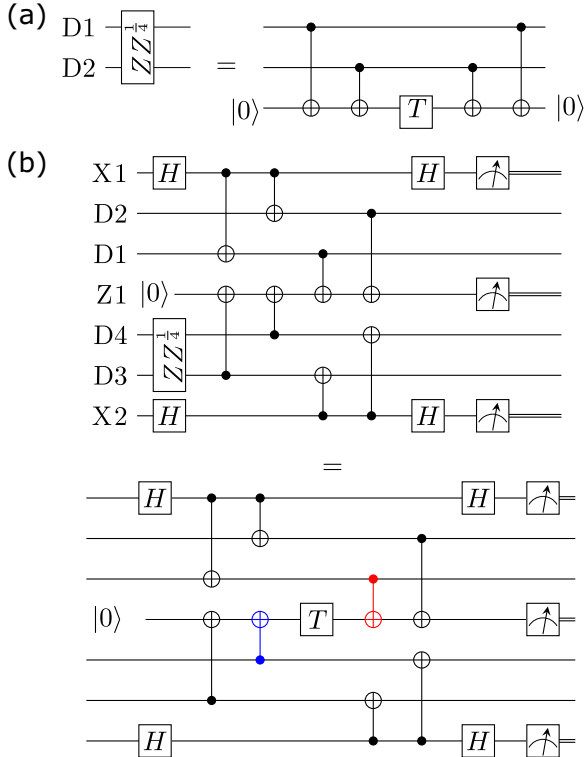


FIG. 6. (a) Identity from Eq. (B3) for $\theta = \frac{\pi}{8}$. (b) Propagation of the T gate using this identity.

gives an error rate expressed as

$$p_{\text{HI}} = \frac{1}{3}p_1 + \frac{3}{15}p_2. \quad (\text{B4})$$

The noise channel is expressed as

$$\begin{aligned} \mathcal{E}_{\text{HI}} &= (1 - p_{\text{HI}})[I] + \frac{1}{3}p_1[F_L] \\ &+ \frac{2}{15}p_2[Y_L] + \frac{1}{15}p_2[Z_L]. \end{aligned} \quad (\text{B5})$$

APPENDIX C: SIMULATING ERASURE QUBITS

In this work, we use the Stim library [49] to perform Clifford simulations. To simulate erasures we generate an erasure sequence for each run, and insert fully depolarizing noise wherever an erasure has happened, converting the erasure circuit into a matchable stabilizer circuit. We then perform minimum-weight perfect matching decoding using PyMatching [50].

Throughout this work, we inject an $|S\rangle = S|+\rangle = (|0\rangle + i|1\rangle)/\sqrt{2}$ state as a test case for injection, as it

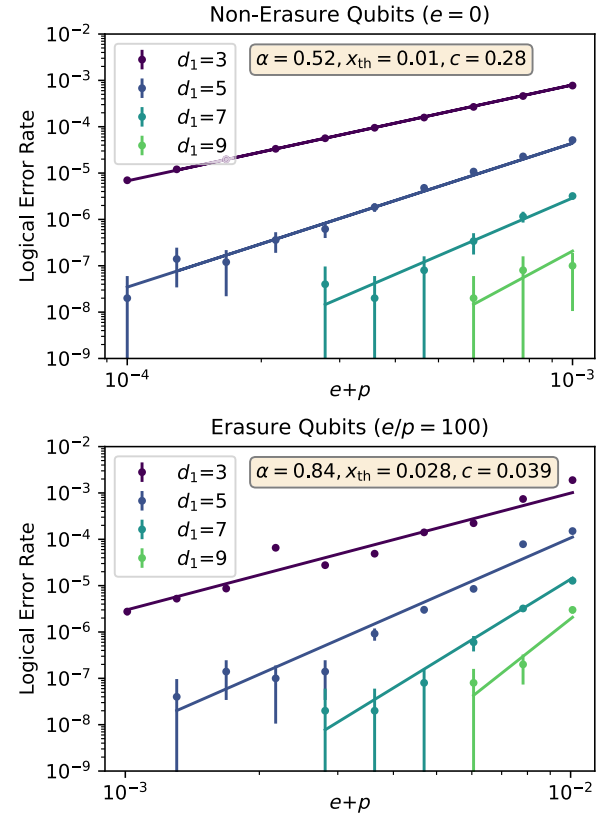


FIG. 7. Logical error rate of the noisy expansion step from a distance- d_1 code to a distance- d_2 code, with the low-distance code state being prepared noiselessly. Solid lines represent fits with Eqs. (C1) and (C2), with parameters given in the floating text box.

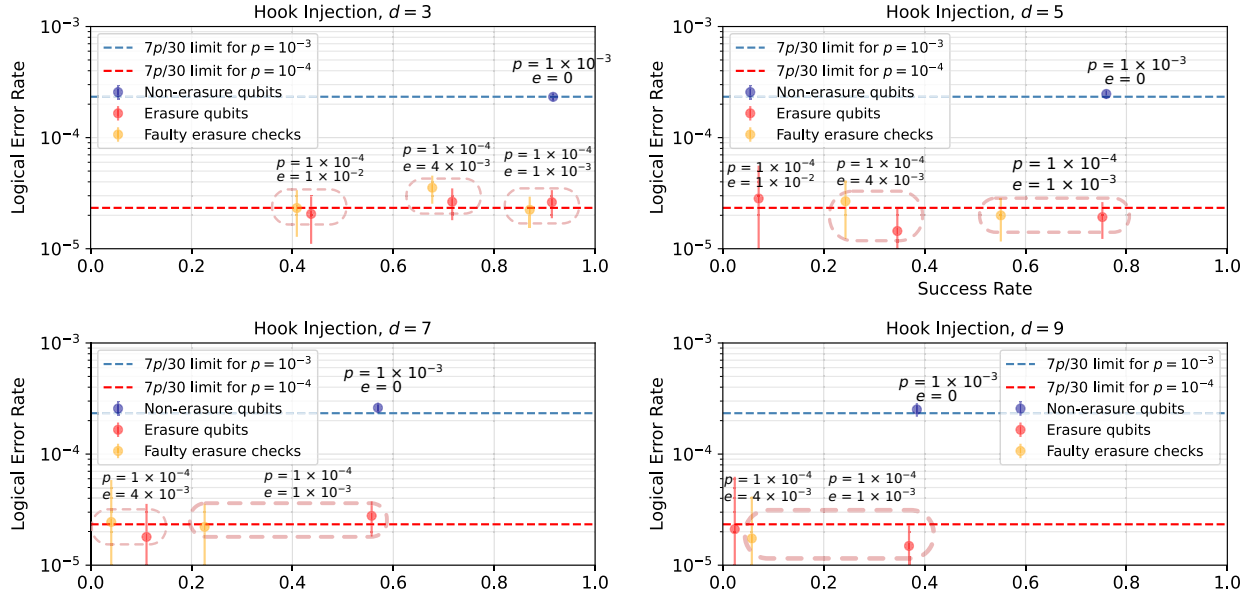


FIG. 8. Variations of Fig. 2 for distances $d_1 \in \{3, 5, 7, 9\}$. For each point, 2×10^6 Monte Carlo simulations were performed. Points with no successful samples were omitted.

includes only gates that can be simulated efficiently on classical computer [51].

In Sec. III A, we perfectly decode the resulting injected state by performing a noiseless stabilizer extraction round at the end of the circuit.

In Sec. III B, we use the circuits published by Gidney [11], which consist of r rounds of injection, a single round of code expansion, and d_2 memory rounds, followed by an in-place Y -basis measurement [38] to verify the injected state. In the sole expansion simulation [Fig. 3(a)], we sample the circuits with noise applied only during the expansion step. In the end-to-end simulations [Fig. 3(b)], noise is applied across all operations throughout the circuit.

Throughout the paper, error bars represent 95% confidence intervals.

To investigate the scaling with distance for the expansion step [see Eq. (1)], we fix the ratio of erasure errors to Pauli errors, such that $p = Re$ for a constant ρ , and define $x = e + p = (1 + \rho)e$ to have a one-parameter noise model. For small x we use an *Ansatz* from Ref. [26] for the logical error rate

$$p_{L,\text{exp}} = c \left(\frac{x}{x_{\text{th}}} \right)^{\alpha d}, \quad (\text{C1})$$

with c , x_{th} , and α fit parameters. For non-erasure qubits, a slightly better fit is achieved with

$$p_{L,\text{exp}} = c \left(\frac{x}{x_{\text{th}}} \right)^{\alpha(d+1)}. \quad (\text{C2})$$

Note that for the code expansion, we are limited by the smallest distance d_1 in the first round of the expansion, and

we thus expect the logical error rate to be approximated by these *Ansätze* for $d = d_1$.

It can be seen from Eq. (C1) that the slopes of the erasure qubit scenarios in Fig. 3(a) are governed by the term $\alpha \log(x/x_{\text{th}})$. In all those scenarios, the qubits are erasure-biased, and thus x_{th} and α remain nearly constant [16,26]. However, $x \equiv e + p$ changes between the different scenarios ($e \in \{10^{-3}, 4 \cdot 10^{-3}, 10^{-2}\}$), leading to differences in the slope.

In Fig. 7, we fit those *Ansätze* to numerical simulations of expansion where the low-distance code d_1 is prepared noiselessly.

To further strengthen our claim, we present in Fig. 8 numerical simulations as reported in Fig. 2 for injection distances $d_1 \in \{3, 5, 7, 9\}$.

We also present a variation of Fig. 2 corresponding to the case in which the protocol is aborted after a single detected erasure, rather than after two detected erasures as considered in the main text. These results are shown in Fig. 9. As observed, the success rate is significantly reduced, while the logical error rate remains largely unaffected.

APPENDIX D: ACCEPTANCE RATE

The probability of not having any erasure is given by

$$R_e(\bar{e}) = (1 - e_{\text{SPAM}})^{x_{\text{SPAM}}} (1 - e_1)^{x_1} (1 - e_2)^{x_2}, \quad (\text{D1})$$

where x_1 (x_2, x_{SPAM}) is the number of 1Q (2Q, SPAM) gates in the circuit. The probability of not having any detected Pauli error is more complex, as for each gate some Pauli combinations will be detected and some will not be

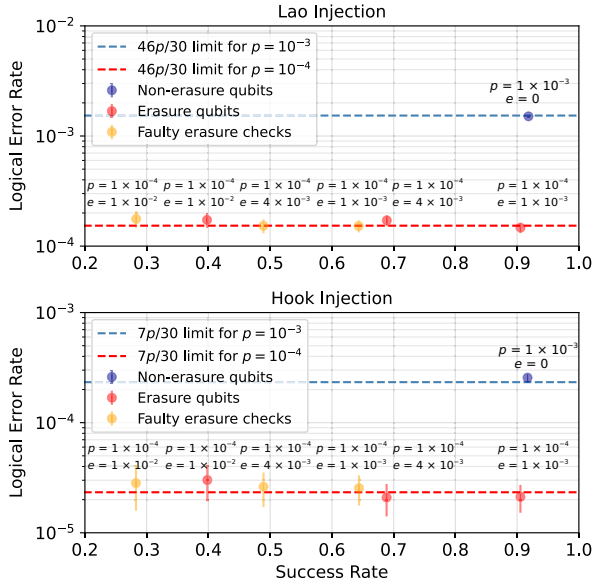


FIG. 9. Variation of Fig. 2 for the case where the protocol is aborted after a single erasure detection. The erasure detection error rates are set to $e_{FP} = e_{FN} = 10^{-2}$ for false positives and false negatives, respectively.

detected. Denoting the probability of a detected Pauli error on gate g to be $p_g^D < p_g$, with p_g the error rate of the gate, we can approximate the acceptance rate:

$$R_p(\bar{p}) \approx \prod_{g \in G} (1 - p_g^D), \quad (\text{D2})$$

with G being the set of all gates in the circuit. This expression approximates the acceptance rate by considering only first-order error probabilities, presuming that higher-order terms have a negligible impact on the overall probability.

To evaluate the probability for a Pauli error to be detected, we show a numerical search of the probability for detection per gate p_g^D for the $d_1 = 3$ hook injection circuit

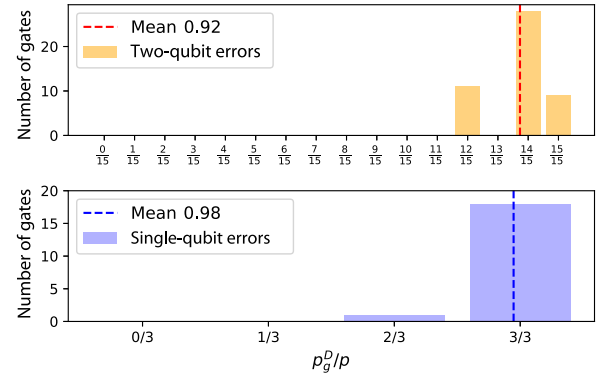


FIG. 10. Histogram of detection rate of errors in two-qubit gates (top) and single-qubit gates (bottom) for a distance-3 hook injection circuit with two stabilizer extraction rounds $r = 2$. SPAM errors are all detected and thus not displayed. The average p_g^D is depicted as dashed lines.

in Fig. 10. Each error location is analyzed for the number of error mechanisms (i.e., the number of different Pauli combinations that are detected) and the normalized probability for each gate is counted in the histogram. It can be seen that for both 2Q gates and 1Q gates, most Pauli error mechanisms are detected. SPAM errors are all detected and thus not displayed in Fig. 10.

This observation—that most Pauli errors are detected—aligns with the design of the surface code, which is tailored to detect and correct errors. The weight-1 errors that are not detected are either: (1) logical faults introduced as part of the injection process or (2) fault locations that act trivially on the state, such as the error depicted in Fig. 12. When the surface code is scaled, the number of type-2 fault locations grows proportionally with the total number of fault locations in the circuit, while the number of type-1 faults remains constant [11]. As a result, the detection rate is expected to remain roughly unchanged with increasing code distance. To further showcase this, we present the average value of p_g^D for distances up to 15 in Fig. 11.

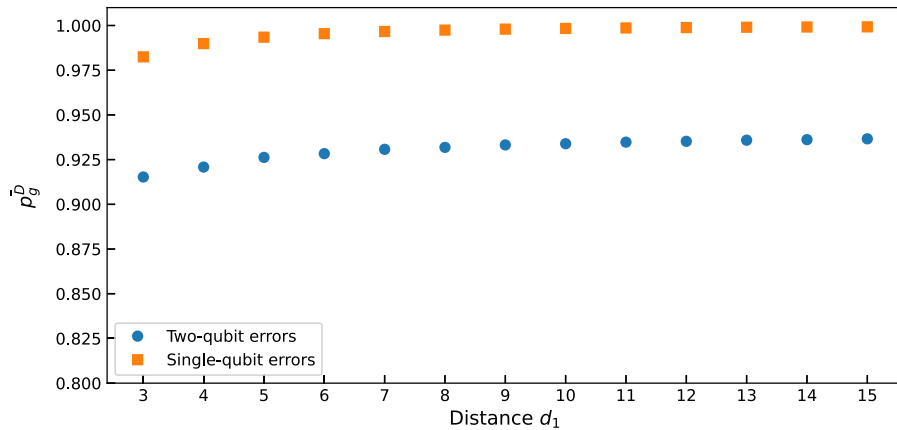


FIG. 11. Average p_g^D/p for various distances, showing that the percentage of detected errors rises only as the distance increases.

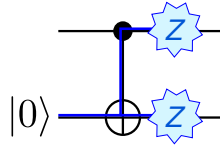


FIG. 12. A ZZ error occurring after the CNOT gate acts trivially on the state and would not trigger an abort. However, if this error is converted into an erasure, it would lead to the protocol being aborted, reducing the success rate. Our analysis suggests that such errors are rare, and that most Pauli errors indeed lead to an abort.

To provide additional support for our findings in Sec. IV, we extend our analysis by showing variations of Fig. 4 using different distances ($d_1 = 5$ and $d_1 = 7$) in Fig. 13.

APPENDIX E: HYBRID DUAL-RAIL AND TRANSMON ARCHITECTURE

Here we develop an error model for the hybrid non-erasure and erasure qubit protocol from Sec. IV. We base this on a scenario where the non-erasure qubits are transmons and the erasure qubits are dual-rail qubits. Since dual-rail qubits and transmons have different gate times, to implement the hybrid architecture, we need to account for idling of some qubits while waiting for other gates to finish. In this appendix, we offer a general analysis of the idling errors for the hybrid architecture.

In the hybrid patch, we have three types of two-qubit gates, (1) between two transmons, (2) between a transmon and a dual-rail qubit, and (3) between two dual-rail qubits. Each of those has a different gate time: $t_{2\text{dr}}$ between two dual-rail qubits, $t_{2\text{tr}}$ between two transmons, and $t_{\text{tr-dr}}$ between a transmon and a dual-rail qubit. Dual-rail qubit gates have longer gate times [16]; thus, in each step of the stabilizer extraction circuit, transmons participating in a transmon-transmon gate need to idle while waiting for the dual-rail qubit gates to finish.

Our assumptions are as follows:

- (1) Since the dual-rail qubits and transmons are fabricated on the same chip, we assume they share the same decay time T_1 .
- (2) We assume equal dephasing and decay times $T_\phi = T_1$ for the transmon and ignore dephasing for the dual-rail qubit.
- (3) The error rates of those gates are dictated only by the decay and dephasing rates of the qubits. This assumption is not exact; however, due to the balance between gate fidelity and gate time, it is often appropriate [52].

Under the twirling approximation, the Pauli error rate for transmons is expressed as [53]

$$p(t) = 3/4 - \frac{1}{4}e^{-t/T_1} - \frac{1}{2}e^{-3t/2T_1} \approx t/T_1. \quad (\text{E1})$$

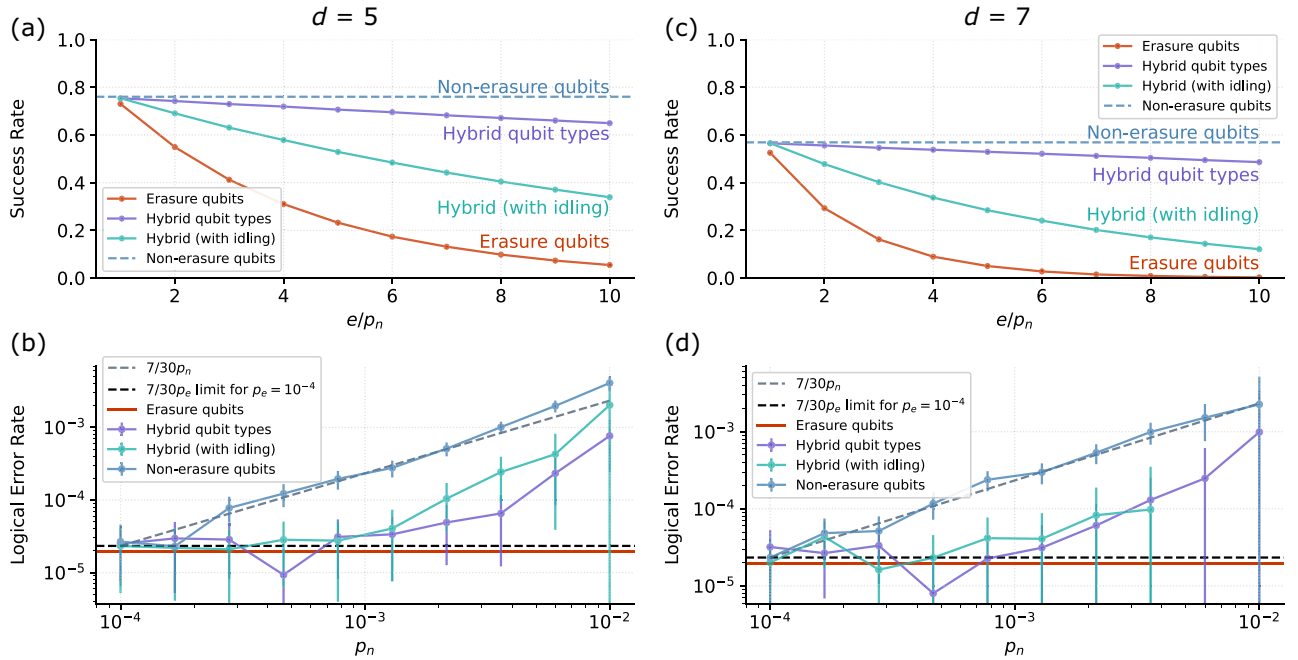


FIG. 13. Figure 4 replotted for two distances, $d_1 = 5$ and $d_1 = 7$. (a),(c) Success probability as a function of the ratio e/p_n , with $p_n = 10^{-3}$ and $p_e = 10^{-4}$. (b),(d) Logical error rate vs p_n for fixed $e = 10^{-3}$ and $p_e = 10^{-4}$. Note that as p_n increases, the number of successful shots decreases, which results in larger error bars at the upper end of the scale.

The erasure rate of dual-rail qubits can be expressed by the decay rate of the two transmons it is composed of. Since dual-rail encoding maintains exactly one excitation distributed between the two transmons [16], the erasure rate is expressed as

$$e(t) = 1 - e^{-t/T_1} \approx t/T_1. \quad (\text{E2})$$

During idle periods, it can be assumed that the transmons undergo dynamic decoupling, reducing most dephasing impact. Thus, when a transmon remains idle for a duration t_{idle} , it experiences an error rate given by

$$p_{\text{idle}}(t_{\text{idle}}) = t/2T_1. \quad (\text{E3})$$

In each time step of the stabilizer extraction circuit, the transmons are subject to two distinct error sources: the two-qubit gate operations ($p_{n,2}$) and the subsequent idling period ($p_{n,\text{idle}}$) while waiting for the dual-rail operations to finish. The combined effective error rate $p_{n,2}^*$ can be expressed as

$$\begin{aligned} p_{n,2}^* &= p_{n,2} + p_{n,\text{idle}} \\ &= p(t_{2\text{dr}}) + \frac{t_{2\text{dr}} - t_{2\text{tr}}}{2T_1} = \frac{e_2 + p_{n,2}}{2}. \end{aligned} \quad (\text{E4})$$

SPAM operation can be delayed or advanced to address the different gate time issue, and one-qubit gates are assumed to require negligible idling time as they are shorter, with smaller error rates.

-
- [1] J. Preskill, Reliable quantum computers, *Proc. R. Soc. London. Series A: Math. Phys. Eng. Sci.* **454**, 385 (1998).
 - [2] B. M. Terhal, Quantum error correction for quantum memories, *Rev. Mod. Phys.* **87**, 307 (2015).
 - [3] S. Bravyi, and A. Kitaev, Universal quantum computation with ideal Clifford gates and noisy ancillas, *Phys. Rev. A* **71**, 022316 (2005).
 - [4] S. Bravyi, and J. Haah, Magic-state distillation with low overhead, *Phys. Rev. A—At., Mol. Opt. Phys.* **86**, 052329 (2012).
 - [5] D. Horsman, A. G. Fowler, S. Devitt, and R. V. Meter, Surface code quantum computing by lattice surgery, *New J. Phys.* **14**, 123011 (2012).
 - [6] 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).
 - [7] D. Litinski, Magic state distillation: Not as costly as you think, *Quantum* **3**, 205 (2019).
 - [8] C. Gidney, N. Shutty, and C. Jones, Magic state cultivation: growing t states as cheap as CNOT gates, [ArXiv:2409.17595](https://arxiv.org/abs/2409.17595).
 - [9] Y. Li, A magic state's fidelity can be superior to the operations that created it, *New J. Phys.* **17**, 023037 (2015).
 - [10] L. Lao, and B. Criger, in *Proceedings of the 19th ACM International Conference on Computing Frontiers* (Association for Computing Machinery, New York, United States, 2022), p. 113.
 - [11] C. Gidney, Cleaner magic states with hook injection, [ArXiv:2302.12292](https://arxiv.org/abs/2302.12292).
 - [12] S. Singh, A. S. Darmawan, B. J. Brown, and S. Puri, High-fidelity magic-state preparation with a biased-noise architecture, *Phys. Rev. A* **105**, 052410 (2022).
 - [13] C. Chamberland, and K. Noh, Very low overhead fault-tolerant magic state preparation using redundant ancilla encoding and flag qubits, *Npj Quantum Inf.* **6**, 91 (2020).
 - [14] H. Bombín, M. Pant, S. Roberts, and K. I. Seetharam, Fault-tolerant postselection for low-overhead magic state preparation, *PRX Quantum* **5**, 010302 (2024).
 - [15] Y. Wu, S. Kolkowitz, S. Puri, and J. D. Thompson, Erasure conversion for fault-tolerant quantum computing in alkaline earth Rydberg atom arrays, *Nat. Commun.* **13**, 4657 (2022).
 - [16] A. Kubica, A. Haim, Y. Vaknin, H. Levine, F. Brandão, and A. Retzker, Erasure qubits: Overcoming the t_1 limit in superconducting circuits, *Phys. Rev. X* **13**, 041022 (2023).
 - [17] M. Kang, W. C. Campbell, and K. R. Brown, Quantum error correction with metastable states of trapped ions using erasure conversion, *PRX Quantum* **4**, 020358 (2023).
 - [18] J. D. Teoh, P. Winkel, H. K. Babla, B. J. Chapman, J. Claes, S. J. de Graaf, J. W. O. Garmon, W. D. Kalfus, Y. Lu, A. Maiti, K. Sahay, N. Thakur, T. Tsunoda, S. H. Xue, L. Frunzio, S. M. Girvin, S. Puri, and R. J. Schoelkopf, Dual-rail encoding with superconducting cavities, *Proc. Natl. Acad. Sci.* **120**, e2221736120 (2023).
 - [19] S. Ma, G. Liu, P. Peng, B. Zhang, S. Jandura, J. Claes, A. P. Burgers, G. Pupillo, S. Puri, and J. D. Thompson, High-fidelity gates and mid-circuit erasure conversion in an atomic qubit, *Nature* **622**, 279 (2023).
 - [20] P. Scholl, A. L. Shaw, R. B.-S. Tsai, R. Finkelstein, J. Choi, and M. Endres, Erasure conversion in a high-fidelity Rydberg quantum simulator, *Nature* **622**, 273 (2023).
 - [21] A. Koottandavida, I. Tsioutsios, A. Kargioti, C. R. Smith, V. R. Joshi, W. Dai, J. D. Teoh, J. C. Curtis, L. Frunzio, R. J. Schoelkopf *et al.*, Erasure detection of a dual-rail qubit encoded in a double-post superconducting cavity, *Phys. Rev. Lett.* **132**, 180601 (2024).
 - [22] H. Levine *et al.*, Demonstrating a long-coherence dual-rail erasure qubit using tunable transmons, *Phys. Rev. X* **14**, 011051 (2024).
 - [23] S. B. Bravyi, and A. Y. Kitaev, Quantum codes on a lattice with boundary, [arXiv:quant-ph/9811052](https://arxiv.org/abs/quant-ph/9811052).
 - [24] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, Topological quantum memory, *J. Math. Phys.* **43**, 4452 (2002).
 - [25] A. Y. Kitaev, Fault-tolerant quantum computation by anyons, *Ann. Phys.* **303**, 2 (2003).
 - [26] S. Gu, Y. Vaknin, A. Retzker, and A. Kubica, Optimizing quantum error correction protocols with erasure qubits, [ArXiv:2408.00829](https://arxiv.org/abs/2408.00829).
 - [27] A. G. Fowler, Analytic asymptotic performance of topological codes, *Phys. Rev. A* **87**, 040301 (2013).
 - [28] S. Bravyi, and A. Vargo, Simulation of rare events in quantum error correction, *Phys. Rev. A* **88**, 062308 (2013).

- [29] F. H. E. Watson, and S. D. Barrett, Logical error rate scaling of the toric code, *New J. Phys.* **16**, 093045 (2014).
- [30] N. Mehta, J. D. Teoh, T. Noh, A. Agrawal, A. Anderson, B. Birdsall, A. Brahmabhatt, W. Byrd, M. Cacioppo, A. Cabrera *et al.*, Bias-preserving and error-detectable entangling operations in a superconducting dual-rail system, *ArXiv:2503.10935*.
- [31] S. J. de Graaf, S. H. Xue, B. J. Chapman, J. D. Teoh, T. Tsunoda, P. Winkel, J. W. Garmon, K. M. Chang, L. Frunzio, S. Puri *et al.*, A mid-circuit erasure check on a dual-rail cavity qubit using the joint-photon number-splitting regime of circuit qed, *Npj Quantum Inf.* **11**, 1 (2025).
- [32] P. Scholl, A. L. Shaw, R. B.-S. Tsai, R. Finkelstein, J. Choi, and M. Endres, Erasure conversion in a high-fidelity Rydberg quantum simulator, *Nature* **622**, 273 (2023).
- [33] B. Eastin, and E. Knill, Restrictions on transversal encoded quantum gate sets, *Phys. Rev. Lett.* **102**, 110502 (2009).
- [34] D. Horsman, A. G. Fowler, S. Devitt, and R. Van Meter, Surface code quantum computing by lattice surgery, *New J. Phys.* **14**, 123011 (2012).
- [35] This can be understood by consideration of the initialization of the magic state as some physical pulse with specific time and amplitude that can in principle rotate to any arbitrary angle. Any error in the rotation angle would result in some overlap of the state on a non-magic state that cannot be identified without distillation.
- [36] In dual-rail qubits, erasure decay can occur from both logical states $|0_L\rangle$ and $|1_L\rangle$, unlike in a transmon, where decay occurs only when the qubit is in $|1\rangle$. This means that when averaging is done over the computational basis, the effective erasure rate is larger relative to that for a single-mode qubit with the same physical decay rates.
- [37] E. Knill, Fault-tolerant postselected quantum computation: Threshold analysis, *ArXiv:quant-ph/0404104*.
- [38] C. Gidney, Inplace access to the surface code Y basis, *Quantum* **8**, 1310 (2024).
- [39] K. Chang, S. Singh, J. Claes, K. Sahay, J. Teoh, and S. Puri, Surface code with imperfect erasure checks, *ArXiv:2408.00842*.
- [40] E. Dennis, Toward fault-tolerant quantum computation without concatenation, *Phys. Rev. A* **63**, 052314 (2001).
- [41] H. Goto, Step-by-step magic state encoding for efficient fault-tolerant quantum computation, *Sci. Rep.* **4**, 7501 (2014).
- [42] H. Goto, Minimizing resource overheads for fault-tolerant preparation of encoded states of the Steane code, *Sci. Rep.* **6**, 19578 (2016).
- [43] T. Itogawa, Y. Takada, Y. Hirano, and K. Fujii, Efficient magic state distillation by zero-level distillation, *PRX Quantum* **6** (2025).
- [44] Y. Hirano, T. Itogawa, and K. Fujii, in *2024 IEEE International Conference on Quantum Computing and Engineering (QCE)* (IEEE, 2024), pp. 843–853.
- [45] R. S. Gupta, N. Sundaresan, T. Alexander, C. J. Wood, S. T. Merkel, M. B. Healy, M. Hillenbrand, T. Jochym-O'Connor, J. R. Wootton, T. J. Yoder *et al.*, Encoding a magic state with beyond break-even fidelity, *Nature* **625**, 259 (2024).
- [46] Y. Vaknin, S. Jacoby, A. Grimsmo, and A. Retzker, Efficient magic state cultivation on the surface code, *ArXiv:2502.01743*.
- [47] Z.-H. Chen, M.-C. Chen, C.-Y. Lu, and J.-W. Pan, Efficient magic state cultivation on $\mathbb{RP}^2$, *ArXiv:2503.18657*.
- [48] S. Jacoby, Y. Vaknin, A. Retzker, and A. L. Grimsmo, Data for “Magic state injection with erasure qubits”, 2025, <https://doi.org/10.5281/zenodo.15874462>.
- [49] C. Gidney, Stim: a fast stabilizer circuit simulator, *Quantum* **5**, 497 (2021).
- [50] O. Higgott, PyMatching: A Python package for decoding quantum codes with minimum-weight perfect matching, *ACM Trans. Quantum Comput.* **3**, 1 (2022).
- [51] E. Knill, R. Laflamme, and W. H. Zurek, Resilient quantum computation, *Science* **279**, 342 (1998).
- [52] F. Motzoi, J. M. Gambetta, P. Rebentrost, and F. K. Wilhelm, Simple pulses for elimination of leakage in weakly nonlinear qubits, *Phys. Rev. Lett.* **103**, 110501 (2009).
- [53] J. Ghosh, A. G. Fowler, and M. R. Geller, Surface code with decoherence: An analysis of three superconducting architectures, *Phys. Rev. A* **86**, 062318 (2012).