PHYSICAL REVIEW RESEARCH **6**, 013177 (2024)

Dynamical subset sampling of quantum error-correcting protocols

Sascha Heußen,[*] Don Winter,[†] Manuel Rispler,[‡] and Markus Müller[§]
Institute for Quantum Information, RWTH Aachen University, 52056 Aachen, Germany
and Institute for Theoretical Nanoelectronics (PGI-2), Forschungszentrum Jülich, 52425 Jülich, Germany

(Received 13 October 2023; accepted 8 January 2024; published 20 February 2024)

Quantum error-correcting (QEC) stabilizer codes enable protection of quantum information against errors during storage and processing. Simulation of noisy QEC codes is used to identify the noise parameters necessary for advantageous operation of logical qubits in realistic quantum computing architectures. Typical quantum error-correction techniques contain intermediate measurements and classical feedback that determine the actual noisy circuit sequence in an instance of performing the protocol. Dynamical subset sampling enables efficient simulation of such nondeterministic quantum error-correcting protocols for any type of quantum circuit and incoherent noise of low strength. As an importance sampling technique, dynamical subset sampling allows one to effectively make use of computational resources to only sample the most relevant sequences of quantum circuits in order to estimate a protocol's logical failure rate with well-defined error bars. We demonstrate the capabilities of dynamical subset sampling with examples from fault-tolerant (FT) QEC. We show that, in a typical stabilizer simulation with incoherent Pauli noise of strength $p = 10^{-3}$, our method can reach a required sampling accuracy on the logical failure rate with two orders of magnitude fewer samples than direct Monte Carlo simulation. Furthermore, dynamical subset sampling naturally allows for efficient simulation of realistic multi-parameter noise models describing faulty quantum processors. It can be applied not only for QEC in the circuit model but any noisy quantum computing framework with incoherent fault operators including measurement-based quantum computation and quantum networks.

DOI: 10.1103/PhysRevResearch.6.013177

I. INTRODUCTION

All quantum hardware inherently suffers from noise and thus will continue to be far from perfect for the upcoming decades [1]. Today's physical architectures ranging from superconducting transmon qubits over photonic systems to ion trap and neutral atom platforms are suitable to implement noisy quantum algorithms for storage, communication, or manipulation of quantum information and quantum simulation. Investigating the effect of noisy components via numerical simulation provides a path to practically realize useful quantum applications, equally important for circuit- and measurement-based quantum computation [2,3] as well as quantum networks [4,5].

Quantum algorithms typically consist of sequences of faulty quantum circuits and measurements, visualized as a tree structure in Fig. 1(a). Figure 1(b) shows that, crucially, the effect of random faults occurring in a given circuit is not known *a priori* but can subsequently lead to random measurement outcomes, which determine the next circuit of the sequence until the protocol terminates.

Widely used numerical simulation methods that perform advantageously in certain regimes of physical fault rates are summarized in Ref. [6]: For direct Monte Carlo sampling, every ideal circuit operation is followed by stochastically drawn fault operators [7,8]. It can be used at relatively large fault rates. Alternatively, one may exhaustively iterate all possible fault events and determine if each one leads to failure of the algorithm. This way, the protocol failure rate can be reconstructed, which is only feasible at very low physical fault rates [9,10]. In Ref. [6] the authors introduce a Metropolis-type technique specific to the surface code to estimate intermediate to small physical fault rates where convergence and thus a reliable confidence interval on the result is not guaranteed.

Instead of iterating *all* possible fault events at low physical fault rates, one may estimate the effect of noise by selectively sampling faults belonging to distinct subsets in a Monte Carlo-type procedure using a *finite* number of samples and approximately reconstruct failure rates with a finite sampling uncertainty. This is the concept of subset sampling, which can be employed for any quantum circuit of a fixed size [11–16]. However, this technique cannot deal with adaptive execution of noisy quantum circuits, which are typically run sequentially as part of a quantum algorithm.

A general numerical technique to efficiently simulate protocols with small physical fault rates as in experiments today or with expected future improvements on experimental

---

[*] sascha.heussen@rwth-aachen.de
[†] don.winter@rwth-aachen.de
[‡] rispler@physik.rwth-aachen.de
[§] m.mueller@physik.rwth-aachen.de

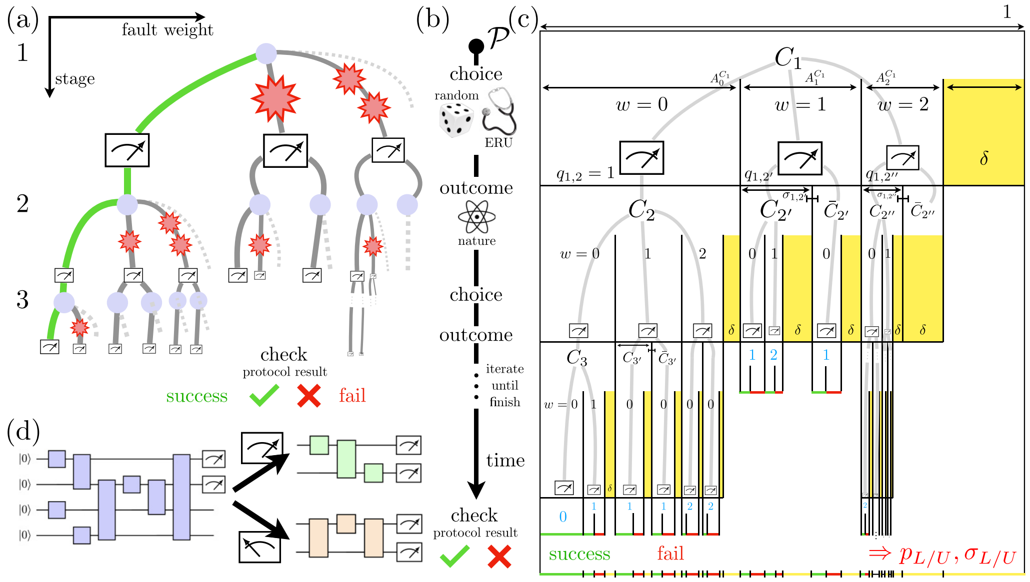


FIG. 1. Protocol illustrations for dynamical subset sampling. (a) Schematic event tree. Purple nodes symbolize individual quantum circuits. For each circuit one can choose to apply a fault with weight $w \in \mathbb{N}_0$ (red stars), which increases from left to right. In the vertical direction, measurement outcomes determine which circuit is run as the next part of a protocol. We assume, for simplicity, that any measurement leads to a binary decision which circuit is run next. After termination, we check whether a failure has occurred. The fault-free path is highlighted in green. For applications outside of QEC, multiple fault-free paths might exist. (b) Protocol sampling consists of alternating steps of choosing a fault subset for the given circuit ("choice") and evaluating the measurement at the end of that faulty circuit ("outcome") to determine the next circuit until the protocol finishes. The choice can be taken at random or by using a diagnostic criterion such as the expected reduction of uncertainty (ERU) that we propose in Sec. II. Measurement outcomes are stochastically determined by "nature". (c) Box representation of the tree in (a) containing circuits C and measurements that cause branchings. The horizontal width of any box, labeled with the fault weight w, symbolizes its binomial factor A_w. The failure rate lower bound p_L is evaluated as the sum over all individual path failure rates (red portion of lower horizontal edge). The true failure rate is underestimated by at most all boxes that were not explored (yellow), i.e., the cutoff error δ, which allows calculation of an upper bound to the failure rate p_U (red + yellow portion of lower horizontal edge). Both bounds have respective sampling uncertainties $\sigma_{L/U}$ that are rooted in the sampling uncertainties σ_i of the branching ratios q_i. The fault-free path never leads to failure. All quantities are introduced in Sec. II. (d) Example of a generic dynamical sequence of two circuits. After measurement of the first two qubits in the first circuit (blue), either the upper (green) or the lower (orange) circuit is run on the last two qubits depending on the intermediate measurement result.

capabilities to get reliable estimates of protocol failure rates is missing.

A. Results summary and paper structure

We introduce *dynamical subset sampling* as an importance sampling method to numerically simulate noisy quantum protocols at low noise strength building on previous works on subset sampling [11–16]. Dynamical subset sampling allows one to estimate protocol failure rates from few, most relevant, fault processes. A well-defined confidence interval on the failure rate estimator is maintained at all times during the sampling process. When dynamical subset sampling is employed at a given maximal noise strength $p = p_{\max}$ to obtain a protocol failure rate estimator $\hat{p}(p_{\max})$, all values for $p \rightarrow 0$ can be extracted analytically while keeping the confidence interval on $\hat{p}$ tight. This is especially useful if the true failure rate $p^* \propto p^{t+1}$ scales to zero fast as $p \rightarrow 0$ with $t \geqslant 1$ as for example in the context of fault-tolerant (FT) quantum error correction (QEC) [17]. Any noise model consisting of incoherent faults can be treated via dynamical subset sampling even for circuits that contain non-Clifford gates. Multiparameter noise models are included naturally in the subset sampling approach so it can accommodate the physical processes predominant in the given hardware architecture [13]. With the python package `qsample` [18], we provide an openly accessible numerical implementation of dynamical subset sampling [19].

This paper is structured as follows. In the remaining part of this introduction we review the required background of subset sampling. Next, we outline our dynamical generalization of the subset sampling technique in Sec. II in the context of FT QEC. In Sec. III we present illustrative examples on how dynamical subset sampling can be utilized to obtain failure rates for GHZ state preparation [20,21] and FT initialization of a logical qubit in the Steane code via flag circuits [16,22–26]. Both examples are highly relevant protocols for FT QEC and have been used in several experiments recently demonstrating fault-tolerant qubit initialization and stabilizer readout [27,28]. We conclude and provide an outlook on future work in Sec. V.

B. Review: Subset sampling

Direct Monte Carlo (MC) simulations are implemented numerically by traversing a quantum circuit and deciding with probability p whether or not to place a fault operator at a circuit location [29]. Fault operators are drawn randomly from a distribution representing the noise model. Repeated realizations of the noisy quantum circuit ("samples" or "shots") yield different output states. The MC estimator for the failure rate is then given by the number of samples that result in a protocol failure divided by the total number of samples

$$\hat{p} = \frac{\#\text{protocol failures}}{\#\text{MC samples}}. \tag{1}$$

This is the standard method to investigate QEC code performance where all components are considered noisy. The handiness of MC comes at the cost of lacking efficiency when failures happen less frequently, i.e., at lower physical fault rates p. To illustrate this, observe that for instance at a physical fault rate of $p = 2 \times 10^{-3}$ in a circuit of $g = 50$ gates, the circuit will be sampled without any fault at all $(1-p)^g \approx 90\%$ of the time. Provided further knowledge about properties of the circuit to be simulated, the contrast can become even more stark. For example, when simulating a fault-tolerant circuit no single fault can cause a failure so MC sampling leads to a trivial result $(1-p)^{50} + 50p(1-p)^{50-1} \approx 99.5\%$ of the time. When postprocessing of an MC run is needed to determine if the result is a failure or not, those 99.5% of all samples would be discarded—a very inefficient strategy [11,14]. The lower the physical fault rate $p \to 0$ the larger this ratio of trivial results will become for a given circuit [6].

The number of faults w happening during circuit execution can be used to uniquely label distinct *subsets* in the space of sampled circuits. For sufficiently low physical fault rate $p = p_{\text{phys}}$ the "largest" subset (in terms of sampling probability) is the 0-fault subset as the above example illustrates. In stark contrast to direct MC, subset sampling never samples in the 0-fault subset thus saving the more computational resources the larger the 0-fault subset, i.e., the lower p_{phys}. Within any w-fault subset, the probability for a specific w fault to occur is the probability to have exactly w faulty locations p_{phys}^w multiplied by the probability to have exactly $N_c - w$ nonfaulty locations $(1-p_{\text{phys}})^{N_c - w}$ simultaneously. There are $\binom{N_c}{w}$ possibilities to choose a w-fault so the size of the w-fault subset is given by the binomial factor

$$A_w^c(p_{\text{phys}}) = \binom{N_c}{w} p_{\text{phys}}^w (1-p_{\text{phys}})^{N_c - w} \tag{2}$$

where N_c is the number of circuit locations, which can cause a fault, and w is the fault weight, which we discuss now. Faults are drawn randomly from a circuit-level Pauli noise model meaning that all single-qubit gates, two-qubit gates, initializations and measurements are performed ideally but followed by a fault operator according to the noise model under investigation. A fault configuration of weight w belongs to the w-fault subset where w is the number of faulty locations in the circuit. The 1-fault subset contains all faults of order p_{phys}, the 2-fault subset contains all faults of order p_{phys}^2 and so on [30] The fault weight w is different from the Pauli weight of an operator: the number of qubits on which it has nontrivial support, i.e., is not the identity I. For example, a $X \otimes Z$ fault on a CNOT gate (and all other circuit elements fault-free) is a 1-weight fault of Pauli weight 2. An illustration of the subsets is given in Fig. 2. The fault-free subset of a circuit c is the largest when the physical fault rate is below the inverse number of potentially faulty circuit locations,

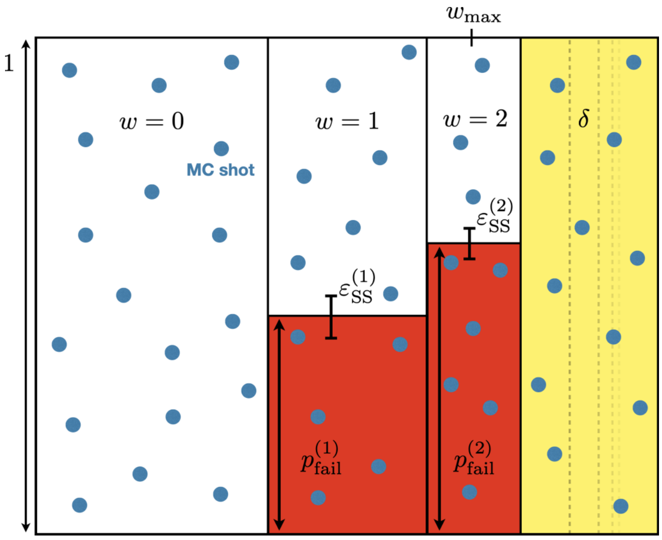


FIG. 2. Representation of w-fault subsets for a single circuit. Whereas direct MC would sample uniformly in the whole w-space (blue dots, most shots in $w = 0$) subset sampling selectively determines the subset failure rates $p_{\text{fail}}^{(w)}$ (ratio of vertical rectangle colored red) up to a weight cutoff $w_{\max}$. This is the key motivation for the subset sampling technique. Here, by splitting the sampling space we avoid the 0-fault subset by design. The upper and lower bound on the true failure rate p_c differ by the cutoff subsets illustrated by the yellow rectangle labeled δ and fainting dashed lines for $w > w_{\max}$.

$$A_0^c(p_{\text{phys}}) > A_1^c(p_{\text{phys}}) \Leftrightarrow p_{\text{phys}} < \frac{1}{N_c + 1}. \tag{3}$$

The failure rate p_c of the circuit is given by the subset failure rates $p_{\text{fail}}^{(w)}$ for each w-fault subset independently as

$$p_c = \sum_{w=0}^{N_c} A_w^c(p_{\text{phys}}) p_{\text{fail}}^{(w)}. \tag{4}$$

Each subset failure rate $p_{\text{fail}}^{(w)}$ can be estimated numerically by randomly drawing $N_{\text{SS}}^{(w)}$ fault operators of weight w according to the noise model under investigation and evaluating the ratio of w-weight-fault realizations that lead to a failure.

Notably, the p_{phys} dependence in Eq. (4) is completely covered by the binomial factors A, which are known analytically. The subset failure rates $p_{\text{fail}}^{(w)}$ are independent of the physical fault rate p_{phys} in this expression. As a consequence, the subset failure rates only need to be sampled once for a fixed numerical value $p_{\text{phys}} = p_{\max}$, so the physical fault rate is naturally removed from the sampling in this formalism. The functional behavior $p_c(p_{\text{phys}})$ can be reconstructed analytically after the sampling, which makes it easy to extract scalings. This is the major advantage of subset sampling, which becomes even clearer when we deal with fault-tolerant circuits. A QEC code with distance d can correct $t = \lfloor \frac{d-1}{2} \rfloor$ errors and an FT circuit can tolerate t faults without failing so that $p_{\text{fail}}^{(w)} = 0 \; \forall w \leqslant t$. For the FT $d = 3$ protocols considered in this paper, we thus know that, by the definition of fault tolerance, $p_{\text{fail}}^{(w=0)} = 0$ and $p_{\text{fail}}^{(w=1)} = 0$.

We can bound the true failure rate p_c by the sampled subset failure rates from above and below by assuming either $p_{\text{fail}}^{(\bar{w})} = 0$ or $p_{\text{fail}}^{(\bar{w})} = 1$ for all $\bar{w} > w_{\max}$ in the circuit. These bounds on the true failure rate read

$$\sum_{w=0}^{w_{\max}} A_w p_{\text{fail}}^{(w)} \leqslant p_c \leqslant \sum_{w=0}^{w_{\max}} A_w p_{\text{fail}}^{(w)} + \overbrace{\sum_{w=w_{\max}+1}^{N_c} A_w}^{\equiv\delta}, \tag{5}$$

where we have simplified the notation of $A_w^c(p_{\text{phys}})$ by A_w since the circuit c and the physical fault rate are fixed quantities here. Using the normalization of the binomial distribution, the cutoff error δ for subset sampling, which is the rightmost sum in Eq. (5), can be expressed as

$$\delta = \delta(p_{\text{phys}}) = 1 - \sum_{w=0}^{w_{\max}} \binom{N_c}{w} p_{\text{phys}}^{w} (1 - p_{\text{phys}})^{N_c - w}. \tag{6}$$

The bounds in Eq. (5) tighten as p_{phys} goes to zero, sharpening the advantage of subset sampling over MC at low p_{phys}. For larger p_{phys} the cutoff $w_{\max}$ needs to be larger in order to keep the cutoff error $\delta(p_{\max})$ below a desired numerical value at fixed $p_{\text{phys}} = p_{\max}$. Eventually, a large $w_{\max}$ will require sampling in a large number of subsets such that employing direct MC will become the more efficient strategy to obtain the failure rate.

We now compare the sampling errors in subset sampling and MC when estimating the failure rate. The sampling error, i.e., the standard deviation, for MC sampling

$$\varepsilon_{\text{MC}} \sim \sqrt{\frac{\hat{p}(1-\hat{p})}{N_{\text{MC}}}} \qquad (\text{as } N_{\text{MC}} \to \infty) \tag{7}$$

falls off slowly when N_{MC} is the number of samples used to estimate $\hat{p}$. For subset sampling, the uncertainty stemming from sampling within a subset analogously reads

$$\varepsilon_{\text{SS}}^{(w)} \sim \sqrt{\frac{p_{\text{fail}}^{(w)}\left(1 - p_{\text{fail}}^{(w)}\right)}{N_{\text{SS}}^{(w)}}} \qquad \left(\text{as } N_{\text{SS}}^{(w)} \to \infty\right) \tag{8}$$

where $N_{\text{SS}}^{(w)}$ is the number of samples used to estimate a single subset failure rate $p_{\text{fail}}^{(w)}$. These subset sampling errors are suppressed by the binomial factors for the total sampling error on the failure rate

$$\varepsilon_{\text{SS}} = \sqrt{\sum_{w=1}^{w_{\max}} \left[A_w^c(p_{\text{phys}}) \varepsilon_{\text{SS}}^{(w)}\right]^2} \tag{9}$$

since the 0-fault subset has the largest binomial factor but is never sampled ($\varepsilon_{\text{SS}}^{(w=0)} = 0$). The 68%-confidence interval $[\hat{p} - \varepsilon, \hat{p} + \varepsilon]$ around the sampled quantity $\hat{p}$ given above is called the Wald interval. At extreme but finite subset failure rates $p_{\text{fail}}^{(w)} \approx 0$ or $p_{\text{fail}}^{(w)} \approx 1$ it is known that the Wald interval suffers from erratic behavior in the output variance ε^2 when the numerical input values only change slightly [31,32]. In this case it is advantageous to use the more stable Wilson interval instead in order to calculate variances ε^2 [33]. More details are given in Appendix C 2.

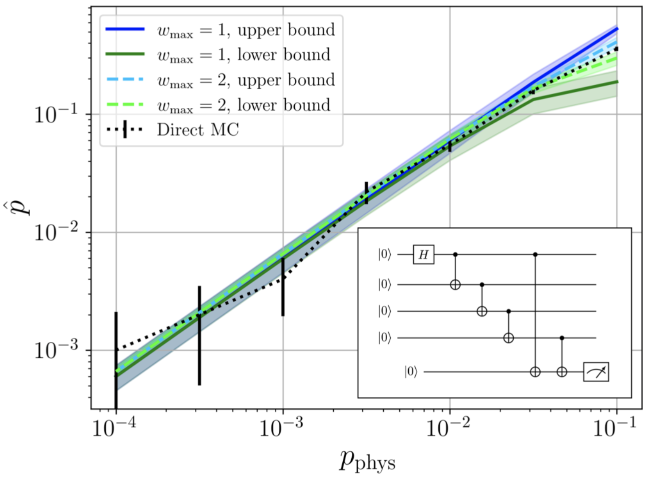


FIG. 3. Behavior of the upper and lower bound on the true failure rate p_c for two different weight cutoffs $w_{\max} = 1$ (solid) and $w_{\max} = 2$ (dashed) for subset sampling the FT four-qubit GHZ preparation circuit (inset, see Sec. III B) where failure corresponds to measuring -1 on the auxiliary qubit; The cutoff error δ is determined at constant $p_{\max}$. The sampling error ε_{SS} (shaded areas around bounds) is independent of p_{phys}. Upper and lower bounds on the failure rate tighten when lowering p_{phys}. Uncertainty intervals for direct Monte Carlo (MC) sampling with at most 1000 shots grow larger when lowering p_{phys}.

Provided that the binomial factors $A_w^c(p_{\text{phys}})$ in Eq. (9) are small (which is always the case for low p_{phys}) we need much fewer samples $\sum_w N_{\text{SS}}^{(w)} \ll N_{\text{MC}}$ to estimate a failure rate with a given variance using subset sampling compared to direct MC. When considering circuits with small N_c it might even be advantageous to exhaustively place all possible 1-faults and get the exact subset failure rate $p_{\text{fail}}^{(w=1)}$ instead of sampling in the 1-fault subset. As a result, the subset sampling error $\varepsilon_{\text{SS}}^{(w=1)}$ vanishes.

The approach can be extended to multiparameter noise models by distinguishing different physical fault rates in Eq. (2) and replace $A_w \to A_{\vec{w}}$ as described in [13]. Furthermore, studies, which have used subset sampling, include Refs. [12,15] for surface code implementations in ion traps and Floquet code implementations in superconducting Majorana platforms respectively.

C. Example: Single shot GHZ state preparation

An example for bounds on the failure rate from Eq. (5) and the combined sampling and cutoff error from Eqs. (6) and (9) is given in Fig. 3. We show the bounds and errors for the inset quantum circuit that prepares the four qubit GHZ state under depolarizing circuit-level noise (see Appendix D) verified by an additional flag auxiliary qubit [21] (more detail on this procedure is provided in Sec. III B). Measuring the flag qubit as -1 is interpreted as a protocol failure in this example. The tightening upper and lower bounds on the true failure rate $p_c(p_{\text{phys}})$ for low p_{phys} are clearly visible. For larger $w_{\max} = 2$ the bounds are narrowing faster, equivalently the cutoff error δ at $p_{\text{phys}} = p_{\max} = 10^{-2}$ is smaller than for $w_{\max} = 1$. The sampling errors $\varepsilon_{\text{SS}}^{(w)}$ only depend on the weight w and the number of samples $N_{\text{SS}}^{(w)}$; they are *independent of the physical* *fault rate*, which is not the case for direct MC. In Fig. 3, the size of the direct MC sampling error grows larger as we lower p_{phys} while keeping N_{MC} fixed because the relative error $\varepsilon_{\text{MC}}/\hat{p} \sim \hat{p}^{-1/2}$ as $\hat{p} \to 0$.

II. DYNAMICAL SUBSET SAMPLING

With dynamical subset sampling (DSS) we present an approach that extends the capabilities of subset sampling by including only the most relevant fault-weight subsets when estimating the failure rate of a *nondeterministic* quantum protocol. DSS can be employed to maximize the gain of accuracy on the failure rate estimator with every shot while maintaining a well-defined uncertainty interval throughout the simulation. In the following discussion we do not explicitly distinguish between the true protocol failure rate p^* from the numerical estimator of the failure rate $\hat{p}$ where there is no danger of confusion and only stress the difference explicitly when important.

Executing a quantum protocol $\mathcal{P}$ under the influence of noise can be viewed as running a nondeterministic sequence $\mathcal{C}$ of quantum circuits as an instance of the protocol $\mathcal{P}$, illustrated as an event tree in Fig. 1. A single element of $\mathcal{C}$ is

$$\mathcal{C}_i = (c_i, w_i), \tag{10}$$

a tuple that specifies a quantum circuit c_i and a w_i-fault subset. A subset sampling approach samples a w_i-fault subset of a circuit $c_i \in \mathcal{C}$ run as part of the protocol $N_{\mathcal{C}_i}$ times and determines the transition rate

$$q_i \equiv \mathbb{P}(\mathcal{C}_i \to c_{i+1} \,|\, \mathcal{C}_{<i}) \tag{11}$$

from $\mathcal{C}_i$ to a next circuit c_{i+1} of the sequence given all previous circuits $\mathcal{C}_{<i}$. Figure 1(c) depicts the sequential branching into subsets of different relevance in the vertical direction, quantified by the binomial factors $A(\mathcal{C}_i) = A^{c_i}_{w_i}$ and the transition rates q_i in the horizontal direction. The latter have a sampling uncertainty of $\sigma_i = \sqrt{\text{Var}(q_i)} = \sqrt{V_i}$ due to the finite number of samples [34], given by the Wilson score interval $V_i = \frac{N_{\mathcal{C}_i} q_i(1-q_i)+1/4}{(1+N_{\mathcal{C}_i})^2}$ [see Eq. (C12)].

Locally, i.e., at stage $i \in \mathbb{N}$, it is impossible to determine by qubit measurements whether or not the execution of a circuit c_i with fault weight w_i will lead to a failure after execution of the whole sequence $\mathcal{C}$. This is not only because the particular fault is drawn randomly but also because its effect on subsequent circuits cannot be determined *a priori* at stage i. We may only assume an expected value for a protocol failure from previous runs of the whole protocol and update that expectation value dynamically when acquiring additional information, i.e., recording a measurement result of circuit c_i with weight w_i. These transition rates are estimated by a finite number $N_{\mathcal{C}_i}$ of observed measurement outcomes of circuit subsets $\mathcal{C}_i$ within the protocol $\mathcal{P}$.

After termination of the whole protocol, one can determine a posteriori whether a protocol failure has occurred as a result of the specific faulty circuit sequence that was realized in this particular shot, symbolized by a red outcome at the end of a tree path in Fig. 1(c). The factors along this path determine the total contribution of this single failure event to the overall failure rate p_L. In total, the protocol failure rate is the weighted sum over all paths P_{fail} that lead to failures in this tree

$$p_L = \sum_{P_{\text{fail}}} \prod_{i \in P_{\text{fail}}} A(\mathcal{C}_i) q_i \tag{12}$$

where $q_l \equiv \mathbb{P}(\mathcal{C}_{\leqslant l} \to \text{FAIL})$ when l is the length of a path P_{fail}.

The quantity p_L is a lower bound to the true failure rate p^* in the sense that we can estimate that

$$p^* \in [p_L - \sigma_L,\ p_U + \sigma_U] \tag{13}$$

lies within a confidence interval of lower and upper bound estimators p_L and p_U. Since they are always estimated from a finite number of shots, we can quantify their sampling uncertainty as

$$\sigma_{L/U} = \sqrt{\text{Var}(p_{L/U})}. \tag{14}$$

We determine our estimator $\hat{p}$ for the failure rate to lie in the center between upper and lower bound so that

$$\hat{p} \equiv \frac{p_L + p_U}{2}. \tag{15}$$

Note that $\hat{p} = p_L$ would also yield an unbiased estimator (see Appendix A for details).

The upper bound p_U to the true failure rate p^* can be set by considering all neglected subsets in a finite tree, illustrated in Fig. 1(c) as yellow boxes. The worst case error made by neglecting these subsets is that they all exclusively produce failures all the time. At any circuit node $\mathcal{C}_i$ of the tree, its cutoff error $\delta_{\mathcal{C}_i}$ can be quantified, analogously to Eq. (6), as the complement of the binomial factors of the non-neglected subsets like

$$\delta_{\mathcal{C}_i} = 1 - \sum_{w_i=0}^{w_{i,\max}} A(\mathcal{C}_i) \tag{16}$$

[see, for example, the upper right yellow box width in Fig. 1(c)]. The upper bound p_U is then given by adding the total value of all $\delta_{\mathcal{C}_i}$ to the lower bound

$$p_U = p_L + \delta = p_L + \sum_{\mathcal{C}_i} \left(\prod_{k=1}^{i} A(\mathcal{C}_k) q_k \right) \delta_{\mathcal{C}_i} \tag{17}$$

where one must take into account the path product that leads to any $\delta_{\mathcal{C}_i}$ in the tree representation of Fig. 1(c).

The variances of p_L and p_U can be calculated using standard rules for calculating sums of products of random variables

$$\text{Var}(p_L) = \text{Var}\left(\sum_{P_{\text{fail}}} \prod_{i \in P_{\text{fail}}} A(\mathcal{C}_i) q_i \right), \tag{18}$$

$$\text{Var}(p_U) = \text{Var}(p_L + \delta), \tag{19}$$

which we lay out explicitly in Appendix B with examples. The central ingredient here is that two transition rates q_i and q_j where one is subsequent to the other in the same path are uncorrelated, i.e., $\text{Cov}(q_i, q_j) = 0$. At the branchings denoted with the measurement symbol in Fig. 1(c), the branching ratios [35] q_i are maximally anticorrelated since $\text{Cov}(q_i, 1 - q_i) = -\text{Var}(q_i)$.

As mentioned above, it is not *a priori* clear which specific circuit sequence will be realized as an instance of the quantum protocol for a particular shot due to both the randomness in realizing a particular fault operator and measurement outcomes. In the following, we suggest two methods how to sensibly choose a fault-weight subset in any stage i as illustrated in Fig. 1(b) (and provide more detail in Appendix C 3).

A simple method is to draw a random fault weight w_i according to the distribution of the binomial factors $A(\mathcal{C}_i)$ of the current circuit c_i. For a protocol that produces long circuit sequences, this method will produce small total fault weights $w_{\text{tot}} = \sum_i w_i$ for a single shot if p_{phys} is not too large. If p_{phys} is also not too small, the resulting w_{tot} of any single shot will eventually be larger than zero and thus protocol failures can be expected. For a protocol that consists of just a single circuit or repeatedly runs the same circuit until the measurement result is accepted so that faults cannot propagate from one circuit run to the next, we can exclude drawing the fault-free subset completely. Note that in principle any distribution can be used to draw fault operators. In Ref. [15], a presampling step is used to guess a distribution of most relevant subsets for a single circuit before actually performing the sampling procedure.

A more sophisticated method systematically decreases the total uncertainty of the failure rate estimation with each sample. It relies on the fact that at any point during the sampling procedure, we can determine expectation values of branching ratios q_i and, thus, failure rates. Now, we may assume how these expectation values change given one additional sample in the current stage i when the w_i-fault subset is selected to sample in. Especially, the total estimation uncertainty

$$\eta \equiv \sigma_L + \sigma_U + \delta \tag{20}$$

changes when the subset w_i is chosen if either a positive (+) or a negative (−) measurement outcome would be observed as a consequence in the next sample. We express the *expected* uncertainty after the next sample as the average over the two possible measurement outcomes $q_i\eta|_{w_i\to +} + (1-q_i)\eta|_{w_i\to -}$. The two expected uncertainties $\eta|_{w_i\to\pm}$ can be calculated by replacing the respective affected transition rate q_i in the tree like

$$q_i \to q_i^{\pm} = \begin{cases} \left(q_i + \frac{1}{N_{C_i}}\right)\frac{N_{C_i}}{N_{C_i}+1}, & + \\ q_i \frac{N_{C_i}}{N_{C_i}+1}. & - \end{cases} \tag{21}$$

as we show in Appendix C 3 b. We can then choose to sample in the subset w_i^* where the *expected reduction of uncertainty* $\Delta(w_i)$ is maximized ("ERU criterion") with the next sample compared to the current state, i.e., the total uncertainty coming from the currently known event tree

$$\Delta(w_i) = \eta - [q_i\eta|_{w_i\to +} + (1-q_i)\eta|_{w_i\to -}] \tag{22}$$

$$w_i^* = \text{argmax}_{w_i}(\Delta(w_i)). \tag{23}$$

It is sufficient to calculate $\Delta(w_i)$ for the w_i-fault subsets that are already in the tree ($w_i = 0, \ldots, w_{i,\max}$) and the largest subset yet unsampled ($w_i = w_{i,\max} + 1$) since sampling in this subset will decrease the uncertainty more than any other higher weight subset, assuming that $A_w > A_{w+1}\ \forall w \geqslant 0$, which is fulfilled for small enough p_{phys}. We consider the computational cost of this calculation further in Sec. IV.

Algorithm 1. Algorithm to dynamically estimate the failure rate p^* of a nondeterministic QEC protocol $\mathcal{P}$ including feedback from classical measurement data using the ERU criterion.

```
Input: maximum number of samples $N_{\max}$ or target uncertainty $\eta_{\max}$, noise model $\mathcal{E}$, maximum noise strength $p_{\max}$
Output: failure rate confidence interval $[p_L - \sigma_L,\ p_U + \sigma_U]$
while $N < N_{\max}$ and $\eta > \eta_{\max}$ do
    calculate $\Delta(w_i)$ for $w_i = 0, \ldots, w_{i,\max} + 1$
    set $w_i \leftarrow \text{argmax}_{w_i}(\Delta(w_i))$
    run $\mathcal{C}_i$ with $w_i$-fault drawn from $\mathcal{E}$
    determine $c_{i+1}$ from measurement outcome of $\mathcal{C}_i$ according to $\mathcal{P}$
    update branching ratio $q_i$
    if $c_{i+1} \neq$ "finish" than
        $i \leftarrow i + 1$ go to next circuit
    else
        check for failure
        update $p_L,\ p_U,\ \sigma_L$ and $\sigma_U$
        $N \leftarrow N + 1$ go to next shot
    end if
end while
```

The full dynamical subset sampling procedure is expressed in pseudocode in Algorithm 1 and as a flow chart in Fig. 4. The numerical values of $p_L,\ p_U,\ \sigma_L$, and σ_U are obtained at a fixed value of $p = p_{\max}$ and the regime of low physical fault rates $\hat{p}(p < p_{\max})$ is accessible analytically via Eq. (12) because the dependency on p lies entirely in the binomial factors $A(\mathcal{C})$,

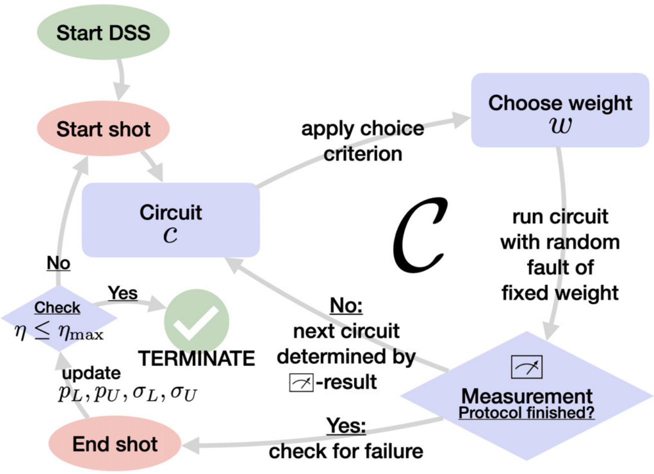


FIG. 4. Circular visualization of the choice-measurement process from Fig. 1(b). A single shot of dynamical subset sampling (DSS) consists of alternating between choosing the next circuit c_i to sample and choosing the fault-weight subset w_i to sample this circuit in. The circuit sequence $\mathcal{C}$ is determined by the underlying protocol $\mathcal{P}$ and the stochastic intermediate measurement results. The fault-weight subset can be chosen according to the binomial factor criterion or the ERU criterion (see main text). Then a random fault instance within that subset is drawn uniformly. Repeated execution of the inner cycle generates circuit sequences $\mathcal{C}$, see Eq. (10), and the event tree in Fig. 1 by sequentially running circuits that form an instance of a protocol. A new shot is run if the result of the previous shot does not lead to a resulting η – the total estimation uncertainty given in Eq. (20)—below the target uncertainty $\eta_{\max}$ or the maximum number of shots $N_{\max}$ is not reached yet.

which themselves do not depend on the sampled transition rates q_i.

In the next section we discuss examples that illustrate the advantage of dynamical subset sampling over MC simulation in the context of fault-tolerant quantum error correction.

III. APPLICATION TO PROTOCOLS FOR FT QEC

The numerical simulations in this section have been performed using our new python package `qsample` [19]. We first discuss the simple example of a flag circuit to fault-tolerantly prepare a GHZ state [20,21]. GHZ states are commonly used as resource states in QEC protocols, e.g., to extract syndrome information for fault-tolerant stabilizer measurements in Shor-type error correction [21,36,37]. As a second example, we investigate a protocol to fault-tolerantly encode the zero code word $|0\rangle_L$ in the Steane code via adaptive execution of either one or more quantum circuits [16,22,24]. The third example is $|0\rangle_L$ state preparation by sequential stabilizer measurements with flag circuits where up to eight circuits can be run in total [25,26,38].

In our simulations we apply the standard circuit-level depolarizing noise model (see Appendix D). We, however, stress again that any noise model consisting of incoherent faults can be covered by dynamical subset sampling even for circuits that contain non-Clifford gates.

A. Background: Simulating noisy QEC

Using QEC to reduce the occurrence of dangerous faults that corrupt the quantum data can allow one to push protocol failure rates below the threshold that enables advantageous operation of encoded blocks compared to physical qubits [39]. Estimating failure rates on the logical level through stabilizer simulation [40,41] is a way to evaluate the performance of stabilizer codes. These performance evaluations are needed for fault-tolerant quantum error-correcting protocols that may be used in hardware implementations. The useful operating regime is below the break-even point where the failure rate becomes lower than the failure rate of the unencoded qubit [42]. If this break-even point converges to a finite value in the infinite system size limit, this is known as the code's threshold. Depending on the code and noise model under investigation, finite thresholds for typical stabilizer codes can—if existent—range over several orders of magnitude (10^{-9} to 10^{-2} faults per operation) [38]. Thus different numerical techniques may be employed to find these thresholds, e.g., for common stabilizer codes such as surface codes [43,44], color codes [45,46], or qLDPC codes [47].

Stabilizer simulations enable the simulation of quantum codes on classical computers in polynomial time for circuits that only contain Clifford operations, which is known as the Gottesman-Knill theorem [41,48]. When investigating noisy stabilizer codes, efficiently simulating faulty circuit operations poses numerical challenges beyond circuit depth or large numbers of qubits due to the large number of MC shots needed to get reliable failure rate estimators (see Sec. I B).

Here we consider uncorrelated circuit-level Pauli noise in stabilizer simulations: Pauli fault [49] operators are stochastically placed on all individual circuit elements. The Pauli matrices form a generator basis of all n-qubit operators that may act as noise in a realistic quantum processor architecture. While other noise models such as phenomenological noise [50,51] or code capacity noise [52] do not take the microscopic details of noisy circuitry for syndrome readout into account, circuit-level noise captures experimental conditions more closely because all circuit elements are modelled as prone to noise [53]. Taking into account the effect of noise to data qubits only thus cannot be expected to accurately model experimental conditions (for a summary of various noise models, see e.g., [54]). We restrict ourselves to uncorrelated circuit-level noise. All types of incoherent noise can be analyzed in this framework of circuit-level Pauli noise, e.g., depolarizing noise [7,55,56] or crosstalk [57,58]. Furthermore, non-Clifford noise can be approximated in this way, e.g., for the amplitude damping channel [59]. To investigate coherent noise or circuits containing non-Clifford gates full wavefunction simulations can be employed at the price of exponentially growing simulation time and memory requirements [60,61].

Placing Pauli faults in a quantum circuit stochastically in a direct Monte Carlo (MC) simulation and determining whether or not these faults cause logical failure has become the standard method to obtain failure rates, thresholds and compare the performance of QEC protocols [7,53]. Recall from Sec. I B that at low physical fault rates however, direct MC is highly inefficient since most of the time the ideal fault-free circuit is sampled (no faults are placed at all). Subset sampling can obtain high accuracy on failure rates with well-defined error bars when the circuit is a *deterministic* sequence of operations [13,15]. However, common QEC protocols such as Shor-type error correction or flag fault-tolerant stabilizer readout are *nondeterministic* because they make use of intermediate measurement information ("feed forward") [21,25].

In nondeterministic QEC protocols, these "in-sequence" measurements determine at runtime which circuit to execute next. Due to the stochastic nature of quantum measurements the resulting circuit sequence performed as an instance of the QEC protocol is not fixed a priori, e.g., for adaptive execution of state preparation circuits as in [62] or FT QEC cycles as demonstrated in [27] where the circuits used for syndrome readout depend on in-sequence measurement information. Nondeterministic protocols are abundant in QEC. They include fault-tolerant initialization of logical qubits using flag schemes [16,24] or teleportation elements where gates, which are applied conditioned on classical measurement data, play a role in the implementation of a fault-tolerant universal gate set [28,63].

In the following, we apply dynamical subset sampling to nondeterministic QEC with single- and multiparameter noise models. Multiparameter noise models can distinguish between different sources of errors and thus provide a more realistic description of the faulty hardware to simulate, e.g., different noise strengths on certain locations, crosstalk [57,58] or biased Pauli noise [64–69].

B. GHZ state preparation

GHZ states are commonly used resource states in quantum algorithms, for instance for FT syndrome readout in Shor-type

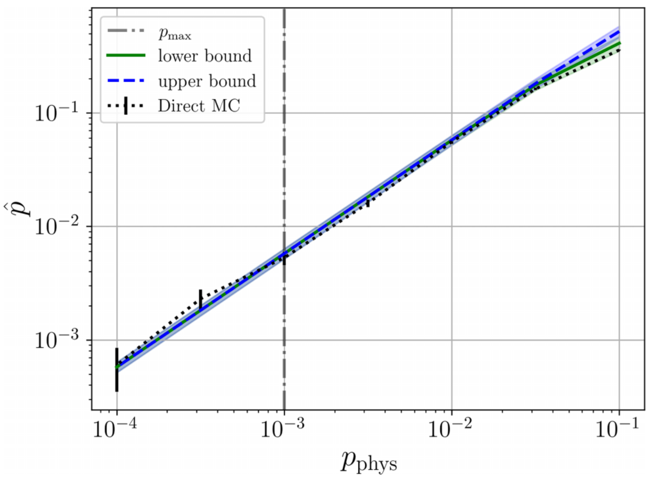


FIG. 5. Upper and lower bound on the flag rate for GHZ state preparation are tight when using subset sampling at $p_{\max} = 10^{-3}$. All rates for smaller p_{phys} can be extracted analytically while the uncertainty intervals of the direct MC estimation grow larger for smaller p_{phys}.

QEC [21,36,37] or quantum teleportation schemes [70]. The preparation procedure of a four-qubit GHZ state

$$|\text{GHZ}\rangle = \frac{|0000\rangle + |1111\rangle}{\sqrt{2}}$$

as shown in the inset of Fig. 3 is considered successful when the flag measurement yields $+1$. Here the first four gates of the circuit prepare the state on the upper four qubits while the fifth qubit is the flag qubit that signals a fault in the preparation through a -1 measurement result, i.e., when being measured in the computational $|1\rangle$ state. This additional flag qubit verification makes the preparation procedure fault-tolerant towards any 1-weight fault in the circuit ("FT-1"), meaning that given a $+1$ measurement result, the prepared state is indeed $|\text{GHZ}\rangle$ up to a single correctable error unless at least two faults have happened. However, single faults can lead to a -1 measurement of the flag qubit.

Figure 5 shows the subset sampling estimation of the flag rate over the range of physical fault rates $p_{\text{phys}} \in [10^{-4}, 10^{-1}]$. Subset sampling is performed with 100 shots at $p_{\max} = 10^{-3}$ and all other values of the failure rate are extracted analytically. We compare the subset sampling result to direct MC estimation, where 10^4 shots are needed for comparable accuracy at $p_{\max} = 10^{-3}$ and relative uncertainties are even larger for lower values of p_{phys}.

This is also reflected in Fig. 6. This advantage of subset sampling over direct MC simulation becomes even larger the lower we choose $p_{\max}$. For every shot, we choose the most "relevant" subset via the ERU criterion. The values of the ERU Δ for the selected subset w^* for each shot are plotted in Fig. 7. Here we observe that the fault-free circuit is chosen exactly once. After the very first shot there is no improvement expected from sampling the fault-free circuit. Then the 1-fault subset is selected repeatedly. The 2-fault subset is only sampled once, in the 21st shot, after the sampling uncertainty in the 1-fault subset is so small that there is more gain to the overall accuracy to reduce the cutoff error by A_2^{GHZ}. Afterwards, it is again preferential to keep sampling in the 1-fault subset. When we use the larger $p_{\max} = 10^{-2}$, we observe in Fig. 8 that the 2-fault subset is chosen more frequently since its binomial factor is now larger than at $p_{\max} = 10^{-3}$ and thus the contribution of the 2-fault subset to the logical failure rate and the total estimation uncertainty η is deemed more relevant.

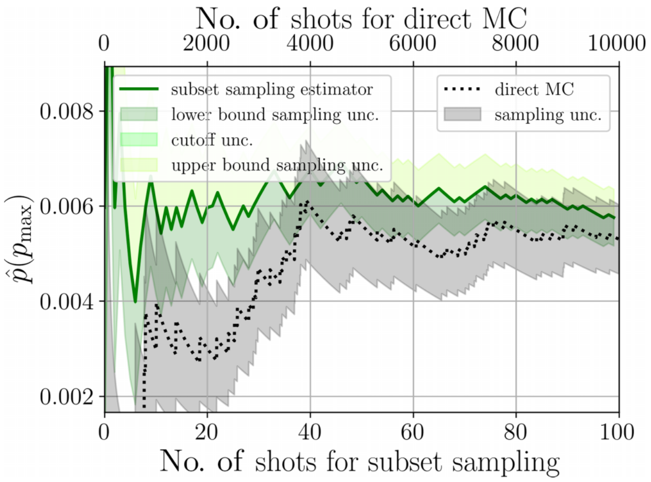


FIG. 6. At $p_{\max} = 10^{-3}$, direct MC (black dotted) needs 10^4 shots to reach a comparable accuracy as subset sampling (green solid) after 10^2 shots. Note that two different horizontal axes are used in the plot. The total estimation uncertainty (unc.) is similar after the sampling for both methods but the advantage of DSS over direct MC is two orders of magnitude in the number of samples.

C. Deterministic FT state preparation

We now present results on sampling the failure rate of a protocol [16] to fault-tolerantly and deterministically prepare the $|0\rangle_L$ state in the Steane code. For our stabilizer simulations

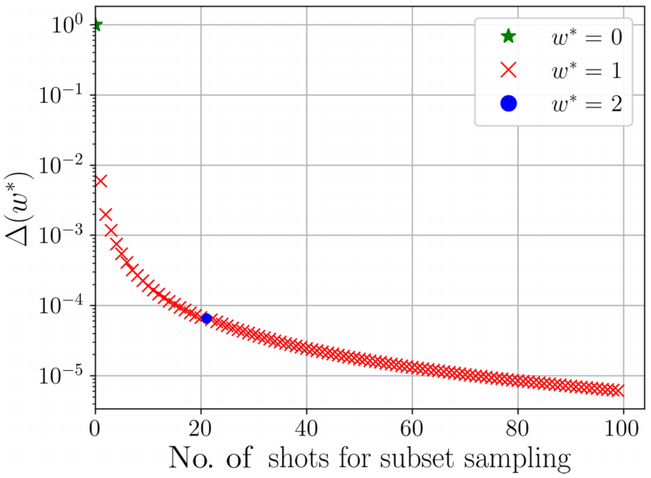


FIG. 7. The ERU criterion, at $p_{\max} = 10^{-3}$, chooses to open the next larger fault-weight subset when decreasing the sampling uncertainty σ does not improve the total uncertainty η more than decreasing the cutoff error δ.

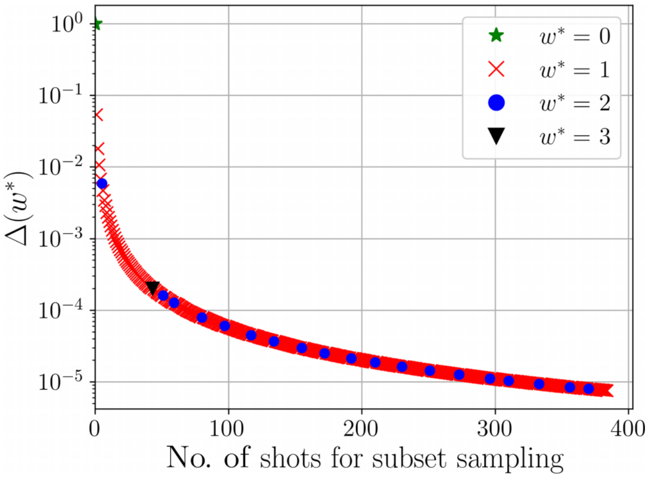


FIG. 8. The ERU criterion, at $p_{\max} = 10^{-2}$, mostly chooses the 1-fault subset. Various times, the 2-fault subset is sampled. The 3-fault subset is chosen once. The larger-weight subsets are more relevant as compared to Fig. 7 because their binomial factors A_2 and A_3 are larger at $p_{\max} = 10^{-2}$ than at $p_{\max} = 10^{-3}$.

of the Steane code we use the stabilizer generators

$$\begin{aligned} K_1^X &= X_4X_5X_6X_7 \quad & K_1^Z &= Z_4Z_5Z_6Z_7 \\ K_2^X &= X_1X_3X_5X_7 \quad & K_2^Z &= Z_1Z_3Z_5Z_7 \\ K_3^X &= X_2X_3X_6X_7 \quad & K_3^Z &= Z_2Z_3Z_6Z_7, \end{aligned} \tag{24}$$

which are illustrated in Fig. 9, and the according static look up table decoder. The $|0\rangle_L$ state is prepared fault-tolerantly in one, two or three steps: first, prepare the state nonfault-tolerantly and measure the operator $Z_3Z_5Z_6$ with a single flag qubit. If the flag qubit is measured as $+1$, we measure all data qubits in the Z basis and the protocol terminates. Otherwise, the stabilizer $K_1^Z K_2^Z K_3^Z = Z_1Z_2Z_4Z_7$ is measured with a second auxiliary qubit. The correction X_7 is applied if both auxiliary qubits are measured as -1. Then, the state preparation protocol terminates. We then measure all data qubits in the Z basis to determine whether a logical failure has incurred.

Provided that at most a weight-1 fault happens during the overall protocol, the $|0\rangle_L$ state is prepared correctly up to a correctable weight-1 error. The result is a logical failure if the minimum Hamming distance of the measured bitstring m to the bitstrings that label all eight basis states that make up the state

$$\begin{aligned} |0\rangle_L = \frac{1}{\sqrt{8}}(&|0000000\rangle + |1010101\rangle + |0110011\rangle \\ &+ |1100110\rangle + |0001111\rangle + |1011010\rangle \\ &+ |0111100\rangle + |1101001\rangle) \end{aligned} \tag{25}$$

is larger than one. This is possible because all weight-2 Z errors are of distance 1 to a logical Z operator in the Steane code.

For an FT protocol, at least two faults are necessary to cause failure and these faults can happen in different circuits. We must ensure that our knowledge about subsets that can never lead to failure is accounted for correctly in the calculation of branching uncertainties σ_i and the cutoff error δ. In particular, we must remove all paths of total weight $w_{\text{tot}} \leqslant 1$ from the cutoff error. This can be done by assuming a maximum sequence length L for such paths (see Appendix B 4 for details). Then, any FT path has a total path product smaller then $LA_1(1 - M_0)$ where $M_0 = \min_c A_0^c$ is the minimal binomial factor at $p_{\text{phys}} = p_{\max}$ out of all circuit's binomial factors and A_1 is the binomial weight of the circuit where a single fault was realized.

To illustrate dynamical subset sampling for a multiparameter noise model, in the following we estimate failure rates for a two-parameter noise model with $\vec{p} = (p_1,\ p_2)$ where p_1 is the physical single-qubit gate fault rate and p_2 is the physical two-qubit gate fault rate. The binomial factor

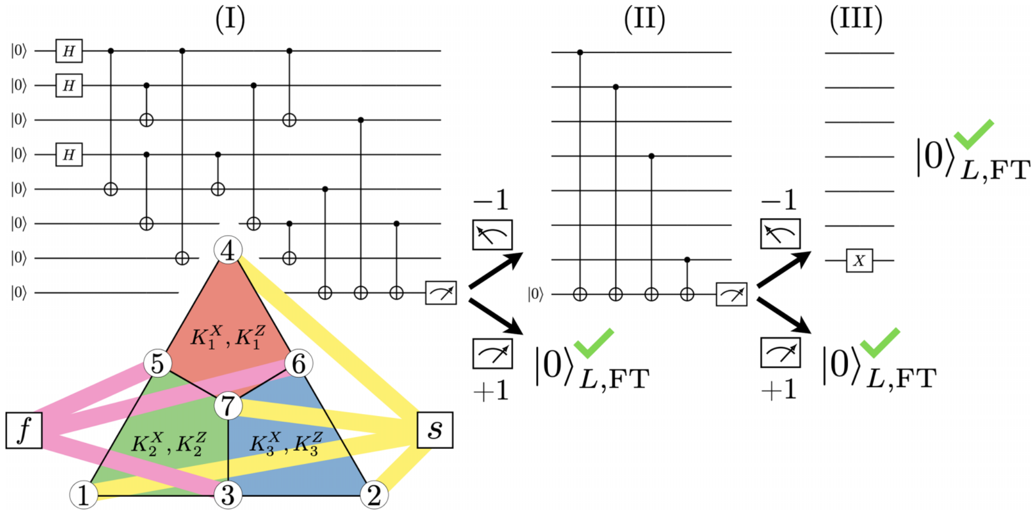


FIG. 9. Possible circuit sequences for deterministic FT state preparation of $|0\rangle_L$ in the Steane code (triangle). If the flag (pink connections to f) is triggered in step (I), the stabilizer $K_1^Z K_2^Z K_3^Z = Z_1Z_2Z_4Z_7$ is measured (yellow connections to s) in step (II). Then, if the next measurement also yields -1, the correction X_7 is applied as step (III). Otherwise the protocol terminates. At any point of termination, $|0\rangle_L$ is prepared fault-tolerantly.

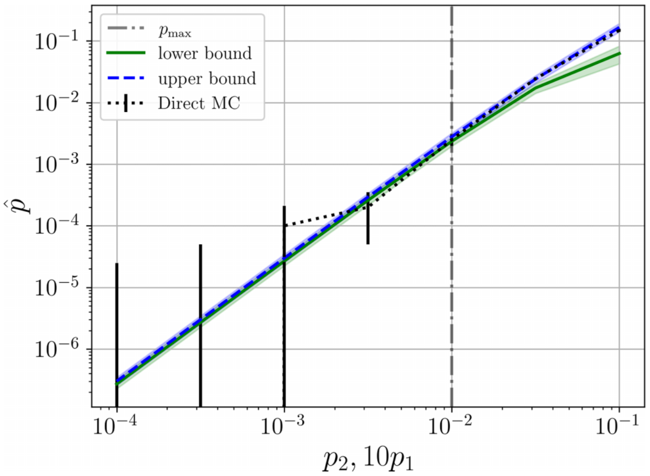


FIG. 10. For the three-step $|0\rangle_L$ state preparation protocol, direct MC achieves a reasonable estimation for $p_2 > 10^{-2}$ (and $p_1 > 10^{-3}$). The direct MC estimator lies in between the upper and lower bounds estimated via subset sampling, which diverge for growing $\vec{p}$. At low p_2 (and p_1) however, direct MC records few or no failures after $N_{\mathrm{MC}} = 10^4$ shots but the analytical scaling of subset sampling yields tight uncertainty intervals around the estimated failure rate bounds.

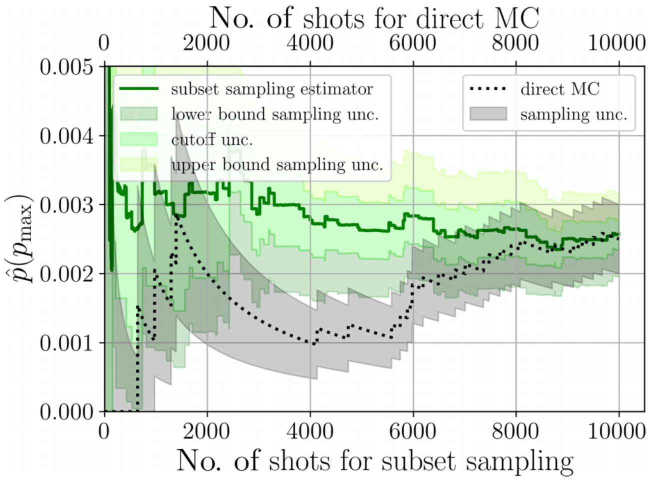


FIG. 11. The subset sampling estimator to the logical failure rate $\hat{p}$ has an uncertainty interval η (shades of green, solid) composed of the upper and lower bound sampling uncertainties σ and the cutoff uncertainty δ. After 10 000 shots at $\vec{p}_{\max} = (10^{-3}, 10^{-2})$, uncertainties are similar for both the direct MC estimation (black dotted) and dynamical subset sampling. For this example, we drew the fault weights randomly according to the circuits' binomial factors.

of Eq. (2) is extended by the now two-dimensional vector $\vec{w} = (w_1, w_2)$ labeling the fault-weight subsets for single-qubit and two-qubit faults to read

$$A^c_{\vec{w}}(p_1, p_2) = \binom{N^{(1)}_c}{w_1}\binom{N^{(2)}_c}{w_2} p_1^{w_1} p_2^{w_2} \times (1-p_1)^{N^{(1)}_c - w_1}(1-p_2)^{N^{(2)}_c - w_2} \qquad (26)$$

where $N^{(1)}_c$ and $N^{(2)}_c$ are the number of single- and two-qubit gates in the circuit c respectively. We employ dynamical subset sampling at $\vec{p}_{\max} = (10^{-3}, 10^{-2})$ and choose fault-weight subsets randomly according to the distribution of binomial factors as described in Sec. II.

In Fig. 10 we observe that the bounds estimated via dynamical subset sampling are tight in the relevant range for the two parameters $p_1 < p_{1,\max}$ and $p_2 < p_{2,\max}$ that are scaled together uniformly. Due to the quadratic scaling of the failure rate estimator $\hat{p} \sim p_i^2$ of the FT protocol, the uncertainty intervals on the direct MC estimators after a fixed number of shots at each $\vec{p}$-value are even larger than for the GHZ state preparation example. At low p_1, p_2, very few or even no logical failures are registered anymore by direct MC.

Although the total uncertainty interval η of dynamical subset sampling is comparable to the direct MC interval after 10^4 shots, as we can see in Fig. 11, we stress that many more MC shots would be required to use direct MC at lower values than $\vec{p}_{\max}$ while the DSS estimator can be extracted analytically without taking new shots.

D. Flag-FT stabilizer measurements

Another way to initialize a logical qubit into the Steane code's $|0\rangle_L$ state is performing sequential stabilizer measurements utilizing flag fault-tolerant circuits as introduced in Ref. [26] and recently demonstrated experimentally [27,71]. Here the data qubit state is projected from $|0\rangle^{\otimes 7}$ onto the simultaneous +1-eigenstate of the X stabilizers. The flag protocol achieves fault-tolerance towards 1-faults (FT-1) through extending the Steane EC look up table decoder by additional correction operators that may be applied depending on the measurement result of flag qubits. An example circuit to fault-tolerantly read out the K^X_1 stabilizer is shown in Fig. 12. A nontrivial −1 measurement outcome for this stabilizer only (while K^X_2 and K^X_3 yield trivial +1 measurements) would lead to a Z_4 Pauli correction (compare inset) that fixes the K^X_1-eigenspace. Note that the Z stabilizers are already fulfilled by initializing all data qubits in the $|0\rangle$ state. The X-stabilizer measurements cannot cause uncorrectable weight-2 X errors without triggering a flag as illustrated in Fig. 12.

The only possible dangerous errors in this scheme follow from faults on the bare readout auxiliary qubits. There, only Pauli-X faults can spread to the data qubits through the subsequent CNOT gates used to measure the X stabilizers. Dangerous faults on the auxiliary qubits that may propagate to weight-2 errors on the data qubits (and thus break FT-1) are accounted for by repeating syndrome measurements on the seven data qubits sequentially using the flag circuits for K^X_1, K^X_2 and K^X_3 until one of the following terminating cases are fulfilled:

(1) In two consecutive rounds a and b, the same syndrome is observed and no flag was triggered. In this case we apply the Z-Pauli correction according to the Steane look up table.

(2) In two consecutive rounds a and b, two different syndromes are observed but no flag was triggered. In this case we run a third round of nonfault-tolerant syndrome readout and apply the Z-Pauli correction according to the Steane look up table using the syndrome that is obtained in the third round of measurement.

(3) At any point a nontrivial flag measurement is observed. In this case we abort the flag circuit readout procedure and continue with a full nonfault-tolerant syndrome readout:

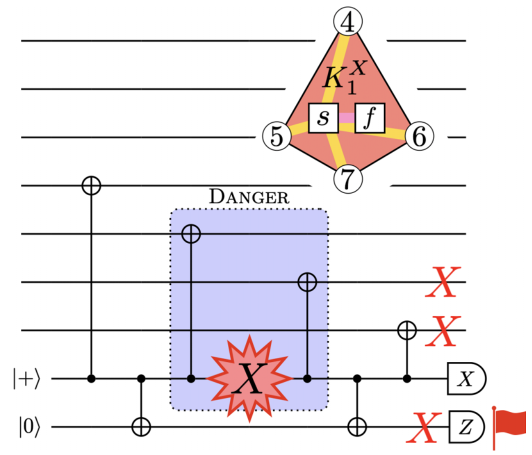


FIG. 12. The stabilizer plaquette $K_1^X = X_4X_5X_6X_7$ of the Steane code can be readout fault-tolerantly using a measurement auxiliary qubit and a flag qubit. Dangerous faults may happen on the two highlighted middle CNOT gates; an example X fault is marked, which propagates to trigger the flag for causing a weight-2 Pauli error on the data qubits. In the subsequent nonfault-tolerant stabilizer readout this weight-2 error results in the syndrome $\{+1, -1, +1\}$ and according to case 3(a) of the protocol the correction X_6X_7 is (correctly) applied. (Inset) Visualization of the qubits that belong to the red plaquette K_1 of the Steane code [according to Eqs. (24)] and the syndrome readout qubit s, which is connected to the four data qubits (yellow) and the flag qubit f, which is only connected to s (pink).

(a) If there exists an X error that is caused by one fault that triggers a flag and agrees with the observed syndrome we apply the X error a second time to correct it. (The set of all errors caused by one fault that triggers the flag is called the flag error set, see example in Fig. 12.)

(b) If no such X error exists we apply the Z-Pauli correction according to the Steane look up table.

The longest possible circuit sequence that may result from applying this protocol is drawn in Fig. 13. Actual instances of the noisy protocol can be much shorter: for instance, when the first circuit flags ("$K_{1,\mathrm{f_a}}^X$"), we will directly continue to run the non-FT syndrome extraction circuit ("K_{nf}^{XZ}") and the protocol terminates. The branching events in this protocol are determined by the measurement outcomes of the current flag circuit: Either the next flag circuit is run or one jumps ahead to the non-FT syndrome readout ("K_{nf}^{XZ}"). The variety of branching options makes dynamical subset sampling suitable to effectively explore subsets that contribute most significantly to the overall failure rate. Low-weight subsets are sampled preferentially by choosing the fault weights, again, randomly according to the distribution of their binomial factors.

Figure 14 shows that the direct MC estimator lies in between the DSS bounds, which diverge for large p_{phys}. Since the failure rate scales quadratically $\hat{p} \sim p_{\mathrm{phys}}^2$ as $p_{\mathrm{phys}} \to 0$, direct MC cannot produce reliable estimations at low p_{phys} with $N_{\mathrm{MC}} = 10^4$ shots. Dynamical subset sampling is used at $p_{\mathrm{max}} = 10^{-3}$ and yields upper and lower bounds on the true failure rate of the flag-FT protocol that do not diverge as $p_{\mathrm{phys}} \to 0$ with the same number of shots. The bound is not as tight as for the previous protocols. This indicates that paths, which are not realized by employing the binomial factor criterion in the sampling procedure and are thus contained in the cutoff error δ, have a relatively large product of binomial factors, given the chosen p_{phys}. Additionally, we assume a maximum FT path length $L = 8$ (6 flag circuits + 1 syndrome readout circuit + 1 circuit that contains only noisy measurements of all data qubits) in our simulations, which can lead to an overestimation of the upper bound (see Appendix B 4 for details). As a result of fixing $L = 8$, we might overestimate the relevance of long circuit sequences to the cutoff error δ if, at a

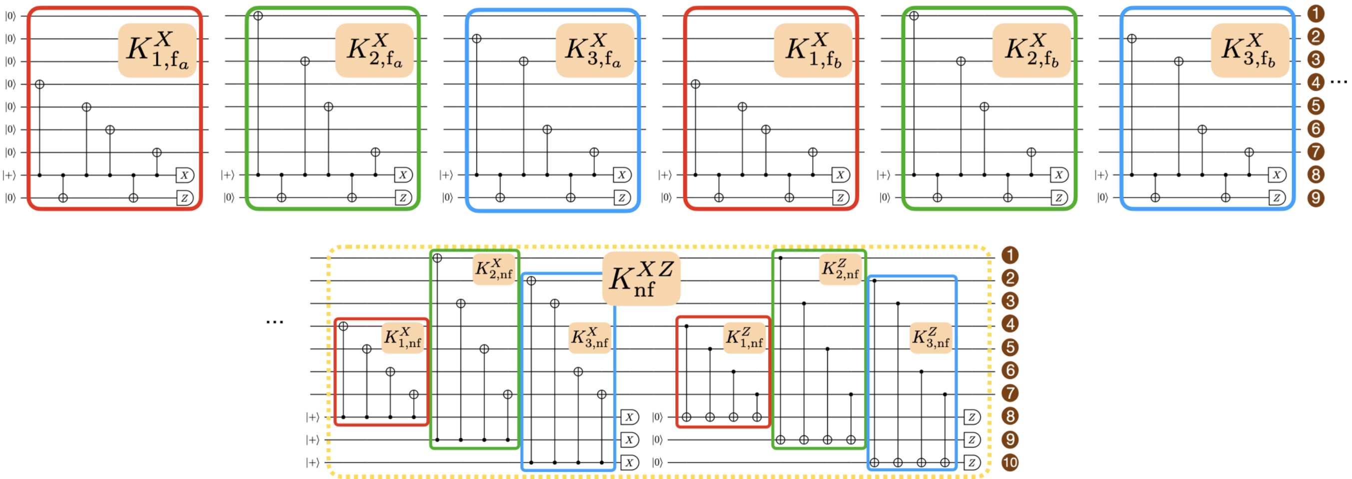


FIG. 13. Longest possible circuit sequence that may occur when using the flag fault-tolerant protocol to encode the $|0\rangle_L$ state. RGB-colored boxes mark the respective Steane code plaquette K_i^σ readout with a flag circuit ($\cdot_{\mathrm{f}}$) or a nonflag circuit ($\cdot_{\mathrm{nf}}$). Whenever the flag qubit (qubit number 9, protocol case 3) is measured as -1 or whenever two consecutive flagged measurements of the same stabilizers $K_{i,\mathrm{f}_a}^\sigma$ and $K_{i,\mathrm{f}_b}^\sigma$ disagree (qubit number 8, protocol case 2), the circuit run is interrupted and continues at the yellow boxed circuit. The nonfault-tolerant syndrome measurement (dashed yellow box "K_{nf}^{XZ}") in the end is only omitted if all six previous stabilizer readouts agree and no flag has been triggered (protocol case 1).

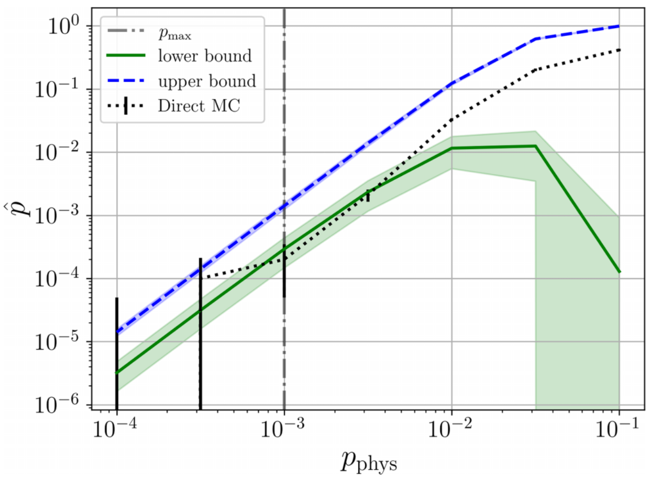


FIG. 14. Logical failure rate estimation for the state preparation protocol via flag-FT stabilizer measurements. Although a relatively large cutoff error remains at $p_{\max} = 10^{-3}$, the total uncertainty on $\hat{p}$ stays bound as $p_{\mathrm{phys}} \to 0$. The direct MC estimators are practically useless in this regime due to too few shots or, equivalently, logical failure rates being too low for the limited number of $N_{\mathrm{MC}} = 10^4$ shots. The cutoff error can be reduced by increasing $p_{\max}$ as in Fig. 26.

given $p_{\max}$, the typical sequence length is much shorter. Both aspects are not fundamental problems of dynamical subset sampling but can be improved in a future version of `qsample` (and we show a tighter bound obtained from sampling at larger p_{phys} in Appendix E).

IV. REMARKS ON EFFICIENCY AND RUN TIME

Dynamical subset sampling requires fewer shots than direct Monte Carlo simulation (and can thus be considered more efficient) when, analogous to Eq. (3), the paths that never lead to failure have the largest weight at a given $p_{\max}$ and/or these paths can be excluded by a suitable choice criterion.

DSS does not limit the execution time beyond the intrinsic limitations of the used stabilizer or statevector simulator backend and the length of the protocol sequence.

Calculation of the total estimation uncertainty η from a large event tree might take a relatively long time compared to running a single shot of a protocol. Parallelization of the sampling procedure for dynamical subset sampling can be employed twofold: Firstly, during the numerical evaluation of the expected reduction of uncertainty $\Delta(w_i)$ at any stage i within the protocol, one may continue to run further DSS shots, using the subset choices from a previous ERU evaluation or the binomial factor criterion, in order to keep lowering the sampling uncertainty, while waiting for the ERU criterion to testify. Secondly, the evaluations of $\Delta(w_i)$ at stage i can themselves be parallelized for each of the $w_{i,\max} + 1$ subsets under consideration and also for both the assumed positive or negative measurement outcome to be probed. Anyways, for practical use of DSS, any $w_{i,\max}$ should be relatively small.

V. CONCLUSIONS AND OUTLOOK

In this paper we have presented an importance sampling technique to efficiently estimate failure rates with well-defined error bars for QEC protocols with in-sequence measurements that make the realized circuit sequence nondeterministic while suffering from incoherent noise at low physical fault rates. Our dynamical subset sampling technique allows for a significant reduction in the number of samples needed to achieve a given accuracy on the estimated failure rate compared to direct Monte Carlo sampling. It can take dependencies between circuits that typically occur in QEC protocols, i.e., circuit execution conditioned on a measurement result, into account. The examples provided allow insight into useful modes of operation for dynamical subset sampling. The first is to avoid the need for post-selecting on a specific measurement result. All circuit sequences resulting from any measurement result can be accounted for. The second is to avoid sampling fault-free circuits, which does not provide useful information but takes up most of the computational resources in direct Monte Carlo simulation at low physical fault rates. In order to choose which fault-weight subset we should sample, we introduce two criteria.

While they proved effective for the examples provided in this paper, it is worth exploring whether sampling efficiency could be increased even further, e.g., by introducing a more elaborate criterion. For instance, it would be interesting to explore the possibility to let an artificial neural network choose fault-weight subsets with a cost function that punishes sampling fault-free or fault-tolerant paths but rewards sampling low-weight paths that can in fact cause failure of a protocol. On a more speculative note, one may be able to employ non-Markovian sampling algorithms that are specifically designed to sample in a nonuniform energy landscape. In a manner similar to simulated annealing [72], fault-weight subsets could be chosen such that deviating just little from the fault-free or fault-tolerant paths is preferential over large fluctuations.

The need for efficient simulation techniques of protocols with intermediate branching options becomes evident in the light of progressing fault-tolerant quantum hardware realizations [27,28,73]. Further study on dynamical subset sampling should include protocols that can have multiple fault-free paths instead of just one. Furthermore, the bound used for the FT cutoff error can be improved by, instead of using the maximum FT sequence length L, inferring the longest possible "rest of sequence" $l \leqslant L$ that can still be run at any given point of an individual sample. Also, we conjecture that a generalization of the FT cutoff error to QEC codes of arbitrary distance is possible by straightforward combinatorics.

All codes used for data analysis are available from the corresponding author upon reasonable request.

The python package `qsample`, which was used to generate all data in this manuscript, is publicly available at GitHub [18]. Within the documentation, also accessible via GitHub [19], we provide a tutorial on how to use `qsample` for modeling quantum protocols and determine failure rates by means of dynamical subset sampling and direct Monte Carlo simulation. The protocols for the examples of Sec. III are also given therein.

ACKNOWLEDGMENTS

S.H. would like to thank Friederike Butt, Lorenzo Cardarelli, Eliana Fiorelli, David Locher, Josias Old, Ciarán Ryan-Anderson, and Seyong Kim for fruitful discussions that led to the development of DSS and Thomas Botzung for also providing useful feedback on the manuscript. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – Cluster of Excellence Matter and Light for Quantum Computing (ML4Q) EXC 2004/1 – 390534769. We gratefully acknowledge support by the EU Quantum Technology Flagship grant under Grant Agreement No. 820495 (AQTION), the U.S. Army Research Office through Grant No. W911NF-21-1-0007 and the ERC Starting Grant QNets through Grant No. 804247. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation thereon. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the view of the U.S. Army Research Office. This research is also part of the Munich Quantum Valley (K-8), which is supported by the Bavarian state government with funds from the Hightech Agenda Bayern Plus. Furthermore, we receive funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101114305 ("MILLENION-SGA1" EU Project). We gratefully acknowledge computing time provided to us at the NHR Center NHR4CES at RWTH Aachen University (Project No. p0020074). This is funded by the Federal Ministry of Education and Research and the state governments participating on the basis of the resolutions of the GWK for national high performance computing at universities.

S.H. developed the DSS scheme, performed the numerical simulations and analyzed the data. D.W. authored the code for `qsample` under supervision of S.H. and all authors contributed to theory modeling and writing of the manuscript. M.M. supervised the project.

APPENDIX A: BIAS OF THE PROTOCOL ESTIMATOR

Subset sampling yields unbiased estimators of protocol failure rates, as we state for Eq. (15), in the following sense. Consider as the bias the difference between the expectation value obtained by an estimator $\hat{p}$ and the true value p^* of the failure rate. The bias of $\hat{p}$ to p^* [74] is

$$\begin{aligned}\mathrm{Bias}(\hat{p}, p^*) &= \mathbb{E}_{\mathrm{Tree}}(\hat{p}) - p^* = \mathbb{E}_{\mathrm{Tree}}\left(\frac{p_L + p_U}{2}\right) - p^* \\ &= \mathbb{E}_{\mathrm{Tree}}\left(p_L + \frac{\delta}{2}\right) - p^* \end{aligned} \tag{A1}$$

where $\mathbb{E}_{\mathrm{Tree}}(.)$ denotes the expectation value deduced from an event tree created by running a finite number of samples. For a single circuit c and a single noise parameter p_{phys}, we can explicitly plug in Eqs. (4) and (6) to express the bias as

$$\begin{aligned}\mathrm{Bias}(\hat{p}, p^*) = {}& \sum_{w=1}^{w_{\max}} A_w^c(p_{\mathrm{phys}})\hat{p}_{\mathrm{fail}}^{(w)} + \frac{1}{2}\left(1 - \sum_{w=0}^{w_{\max}} A_w^c(p_{\mathrm{phys}})\right) \\ & - p^* \to 0. \end{aligned} \tag{A2}$$

In the limit of large numbers of shots in every subset, convergence of the subset failure rate estimators $\hat{p}_{\mathrm{fail}}^{(w)}$ to their true values $p_{\mathrm{fail}}^{*(w)}$ is assured by the central limit theorem and the law of large numbers. The cutoff error contribution $\delta \to 0$ in the limit of taking all subsets into account $w_{\max} \to N_c$, which is equivalent to the limit of $p_{\mathrm{phys}} \to 0$. Thus, it is clear that our choice $\hat{p} = \frac{p_L + p_U}{2}$ is an unbiased estimator of the true failure rate p^*.

APPENDIX B: DISCUSSION OF THE FAILURE RATE UNCERTAINTY INTERVAL

In this Appendix, we refer to uncertainty intervals as "errors" since there is no danger of confusion with Pauli operators corrupting a QEC code in the following. The lower and upper bound to the logical failure rate estimator and the cutoff error, as defined in Eqs. (12), (16), and (17) in Sec. II, are given by

$$p_L = \sum_{P_{\mathrm{fail}}} \prod_i A(w_i) q_i \tag{B1}$$

$$\begin{aligned}\delta &= 1 - \sum_{P} \prod_i A(w_i) q_i \\ &= \sum_{s=1}^{L} \sum_{C_s} \left[\prod_i^{s} A(w_i) q_i \left(1 - \sum_{w_{C_s}=0}^{w_{\max}^{(C_s)}} A(w_{C_s})\right)\right] \end{aligned} \tag{B2}$$

$$p_U = p_L + \delta = 1 - \sum_{P_{\text{no-fail}}} \prod_i A(w_i) q_i. \tag{B3}$$

The quantities p_L and p_U can be used to bound the true failure rate p^* of a protocol in the sense that

$$p^* \in [p_L - \sigma_L,\ p_U + \sigma_U] \tag{B4}$$

where $\sigma_{L/U}$ is the respective sampling error for the estimators p_L and p_U that results from estimating $p_{L/U}$ from a finite number of shots and the true failure rate p^* likely lies between the upper and lower bound. In this section we derive these confidence intervals $[p_L - \sigma_L, p_U + \sigma_U]$ with

$$\sigma_{L/U} = \sqrt{\mathrm{Var}(p_{L/U})} \tag{B5}$$

from the individual branching uncertainties σ_i of an event tree.

The two equalities in Eq. (B2) are two equivalent ways to calculate the cutoff error δ in a subtractive or additive way respectively, which we elaborate further in the subsequent sections. The additive cutoff error (right-hand side) contains the sum over all stages s of an event tree of length L and all circuits C_s within a stage s with their respective binomial factors $A(w_{C_s})$. The subtractive cutoff error (left-hand side) is just calculated via the sum over all paths P in the tree. The summation index i refers to the respective path P, which the binomial factor $A(w_i)$ is part of.

1. Calculating event tree variances

In order to calculate the variances in Eq. (B5), we first show general properties of variances that are useful to calculate expressions of the form of $\mathrm{Var}(p_L)$, which contain both sums and products of random variables. Then, we explicitly calculate illustrative examples of generic event trees.

a. General formulas

Sums of random variables. For a sum of two random variables X and Y, each multiplied by constants a and b respectively, we calculate its variance as

$$\mathrm{Var}(aX+bY)=a^2\,\mathrm{Var}(X)+b^2\,\mathrm{Var}(Y)+2ab\,\mathrm{Cov}(X,Y) \tag{B6}$$

where

$$\mathrm{Cov}(X,Y)\equiv\mathbb{E}((X-\mathbb{E}(X))(Y-\mathbb{E}(Y))) \tag{B7}$$

is the covariance between X and Y. Generalizing to a sum of n variables, the total variance is

$$\mathrm{Var}\left(\sum_i^n a_iX_i\right)=\sum_i^n a_i^2\,\mathrm{Var}(X_i)+2\sum_{j=1}^n\sum_{i=1}^{j-1}a_ia_j\,\mathrm{Cov}(X_i,X_j). \tag{B8}$$

A useful observation is that

$$\mathrm{Cov}(X,1-X)=\mathrm{Cov}(X,-X)=-\mathrm{Var}(X). \tag{B9}$$

More generally, a relation that is useful at tree branchings is

$$\begin{aligned}&\mathrm{Var}[aX+b(1-X)]\\&\quad=a^2\,\mathrm{Var}(X)+b^2\,\mathrm{Var}(1-X)+2ab\,\mathrm{Cov}(X,1-X)\\&\quad=(a-b)^2\,\mathrm{Var}(X)\leqslant(a^2+b^2)\,\mathrm{Var}(X).\end{aligned} \tag{B10}$$

From this we can see directly that the variance of two subtrees that branch off a node with transition rates $q=X$ and $1-q=1-X$ is always overestimated if one were to ignore the covariance term.

Products of random variables. We can also calculate the variance of a product of two independent variables

$$\mathrm{Var}(XY)=\mathrm{Var}(X)\mathrm{Var}(Y)+\mathrm{Var}(X)\mathbb{E}(Y)^2+\mathrm{Var}(Y)\mathbb{E}(X)^2, \tag{B11}$$

which generalizes to

$$\mathrm{Var}\left(\prod_i a_iX_i\right)=\prod_i a_i^2(\mathrm{Var}(X_i)+\mathbb{E}(X_i)^2)-\prod_i a_i^2\mathbb{E}(X_i)^2 \tag{B12}$$

as long as the X_i and X_i^2 are independent [75,76]. We also refer to Eq. (B12) as Goodman's formula. Calculating a path variance with Goodman's formula requires an exponential number of additions. For a path with L branching ratios, the first product creates L factors of two-termed sums and only the term consisting of the product of all q_i^2 is subtracted again,

$$\mathrm{Var}\left(\prod_{i=1}^L q_i\right)=\prod_{i=1}^L\left[\mathrm{Var}(q_i)+q_i^2\right]-\prod_{i=1}^L q_i^2, \tag{B13}$$

so we have 2^L-1 terms in the sum that we need to calculate for a single path variance. For a tree that contains a number of paths W, calculation of the variance will take $W\times(2^L-1)$ additions and L multiplications per summand.

Covariance. The covariance of two subsequent random variables q_j and q_k of a path in an event tree is identical with zero since the latter variable q_k is conditioned on the former q_j. In general, for four dependent random variables X, Y, U, V that follow a normal distribution with expectation values x, y, u, v, it holds that

$$\begin{aligned}&\mathrm{Cov}(XY,UV)\\&\quad=xu\,\mathrm{Cov}(Y,V)+xv\,\mathrm{Cov}(Y,U)\\&\qquad+yu\,\mathrm{Cov}(X,V)+yv\,\mathrm{Cov}(X,U)\\&\qquad+\mathrm{Cov}(X,U)\mathrm{Cov}(Y,V)+\mathrm{Cov}(X,V)\mathrm{Cov}(Y,U).\end{aligned} \tag{B14}$$

It also follows directly from the definition of the covariance if X, Y and U, V are each independent that

$$\begin{aligned}\mathrm{Cov}(XY,UV)&=\mathrm{Cov}(X,U)\mathrm{Cov}(Y,V)\\&\quad+yv\mathrm{Cov}(X,U)+xu\mathrm{Cov}(Y,V).\end{aligned} \tag{B15}$$

Also, we have the special case

$$\begin{aligned}\mathrm{Cov}(XY,XV)&=\mathrm{Var}(X)\,\mathrm{Cov}(Y,V)+yv\,\mathrm{Var}(X)\\&\quad+x^2\,\mathrm{Cov}(Y,V).\end{aligned} \tag{B16}$$

Let us remark that a covariance term can be upper bounded using the Cauchy-Schwarz inequality as

$$|\mathrm{Cov}(X,Y)|\leqslant\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}. \tag{B17}$$

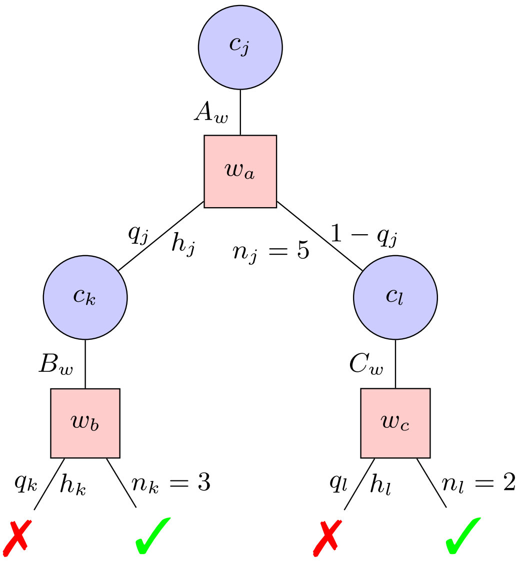


FIG. 15. Example event tree with two stages and five samples [cf. Fig. 1(c)]. Circuit nodes (blue circles) are labeled c_j, c_k, and c_l. Subset nodes (red squares) are labeled w_a, w_b, and w_c. The branching ratios q_j and q_k as well as q_j and q_l (and also q_k and q_l) are uncorrelated.

b. Examples for event tree covariance contributions

Subsequent branching rates are uncorrelated. We now demonstrate, by an example, that subsequent branching rates q_i in our event trees are indeed uncorrelated. Assume that the branching with q_k in Fig. 15 happens after, i.e., further down in the tree and in the same path, as the branching with q_j. At the transition labeled with q_j there is a list that keeps the measurement history h_j, e.g., $h_j=[0,0,1,0,1]$ meaning for instance that, for the first shot, the measurement result for the transition rate q_j was 0 and, for the fifth shot, the measurement result for the transition rate q_j was 1. In this example, our current estimate of q_j after 5 shots would then be $q_j = 3/5$. Let us assume that we branch off to c_k whenever the measurement result was 0 and to another circuit node c_l when the measurement result was 1. For q_k then, there is another list h_k, e.g., $h_k = [1, 0, 1]$. It must have the same length as the number of zeros in h_j since only those measurement results lead us to circuit node c_k. At circuit node c_l there will be a different history resulting from the measurement results 1 at c_j, e.g., $h_l = [0, 1]$. We leave out the binomial factors for this discussion for simplicity. Now obviously, the sample covariance

$$\mathrm{Cov}(X, Y) = \frac{1}{N}\sum_{i=1}^{N}[X_i - \mathbb{E}(X)][Y_i - \mathbb{E}(Y)] \tag{B18}$$

is zero since we do not plug in h_j and h_k but only the sublist where all elements are zero $h_j^{(0)} = [0, 0, 0]$. Then

$$\mathbb{E}\big(h_j^{(0)}\big) = 0, \quad \mathbb{E}\big(h_j^{(1)}\big) = 1, \quad \mathbb{E}(h_k) = \tfrac{2}{3}, \quad \mathbb{E}(h_l) = \tfrac{1}{2}, \tag{B19}$$

$$\mathrm{Cov}\big(h_j^{(0)}, h_k\big) = \tfrac{1}{3}\big[(0-0)\big(1-\tfrac{2}{3}\big) + (0-0)\big(0-\tfrac{2}{3}\big) + (0-0)\big(1-\tfrac{2}{3}\big)\big] = 0, \tag{B20}$$

$$\mathrm{Cov}\big(h_j^{(1)}, h_l\big) = \tfrac{1}{2}\big[(1-1)\big(0-\tfrac{1}{2}\big) + (1-1)\big(1-\tfrac{1}{2}\big)\big] = 0. \tag{B21}$$

We now consider different types of minimal example trees to illustrate the relevance of nonzero covariance terms. We will see, as stated above, that covariance terms for branchings from subset nodes are negative. Covariance terms of branchings from circuit nodes (into fault-weight subsets) can in general be non-negative.

Subset node branching covariance is negative. For a sum of products, we in fact calculate variances of correlated variables. This is because branching to the left, e.g., in Fig. 15, is weighted with q_j and branching to the right is weighted with $1 - q_j$ [cf. Eq. (B9)]. So for a generic branching we have terms like

$$\begin{aligned}\mathrm{Var}[Bq_jq_k + C(1-q_j)q_l] &= B^2\,\mathrm{Var}(q_jq_k) + C^2\,\mathrm{Var}[(1-q_j)q_l]\\ &\quad + 2BC\,\mathrm{Cov}[q_jq_k, (1-q_j)q_l].\end{aligned} \tag{B22}$$

From Eq. (B15) we can see that

$$\mathrm{Cov}[q_jq_k, (1-q_j)q_l] = -q_kq_l\mathrm{Var}(q_j). \tag{B23}$$

with $X = q_j, Y = q_k, U = 1 - q_j, V = q_l$ since $\mathrm{Cov}(q_k, q_l) = 0$.

Denoting $\mathrm{Var}(q_j) = V_j$, we can continue simplifying Eq. (B22) like

$$\begin{aligned}\mathrm{Var}[Bq_jq_k + C(1-q_j)q_l] &= B^2\big(q_j^2V_k + q_k^2V_j + V_jV_k\big)\\ &\quad + C^2\big[q_l^2V_j + (1-q_j)^2V_l + V_jV_l\big] - 2BCq_kq_lV_j\end{aligned} \tag{B24}$$

$$\leqslant B^2\big(q_j^2V_k + q_k^2V_j + V_jV_k\big) + C^2\big[q_l^2V_j + (1-q_j)^2V_l + V_jV_l\big] \tag{B25}$$

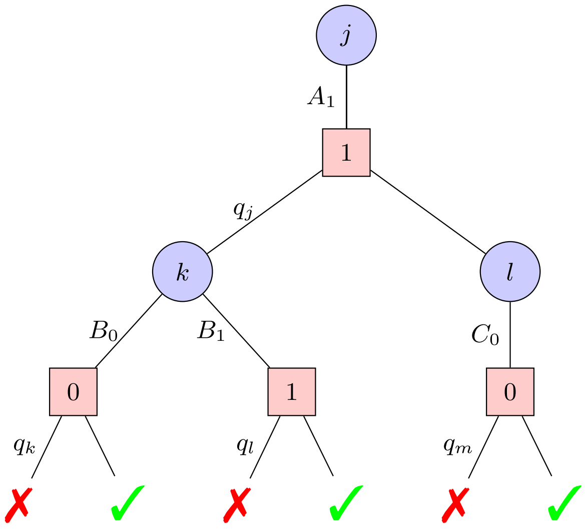


FIG. 16. Event tree example with nonzero covariance term between q_k and q_l because they both share the random variable q_j in their path. We reduced the indexing of the nodes for better readability. Circuit nodes are labeled j, k, and l. Subset nodes are labeled 0 and 1 for this example.

to obtain an expression that is symmetric under simultaneous exchange of $q_j \leftrightarrow 1 - q_j$ and $q_k, V_k \leftrightarrow q_l, V_l$ if $B = C$ as expected. Since that last term of Eq. (B24) contains only positive factors but a negative sign, we surely increase the value of the whole expression by neglecting it.

Note that we could also use the Cauchy-Schwarz inequality (B17) to obtain a (worse) bound to Eq. (B22) like

$$\begin{aligned}&\mathrm{Var}[Bq_jq_k + C(1-q_j)q_l]\\ &\quad = B^2\,\mathrm{Var}(q_jq_k) + C^2\,\mathrm{Var}[(1-q_j)q_l]\\ &\qquad + 2BC\,\mathrm{Cov}[q_jq_k, (1-q_j)q_l]\\ &\quad \leqslant B^2\,\mathrm{Var}(q_jq_k) + C^2\,\mathrm{Var}[(1-q_j)q_l]\\ &\qquad + 2BC\sqrt{\mathrm{Var}(q_jq_k)\mathrm{Var}[(1-q_j)q_l]},\end{aligned} \tag{B26}$$

which we just leave as a general comment.

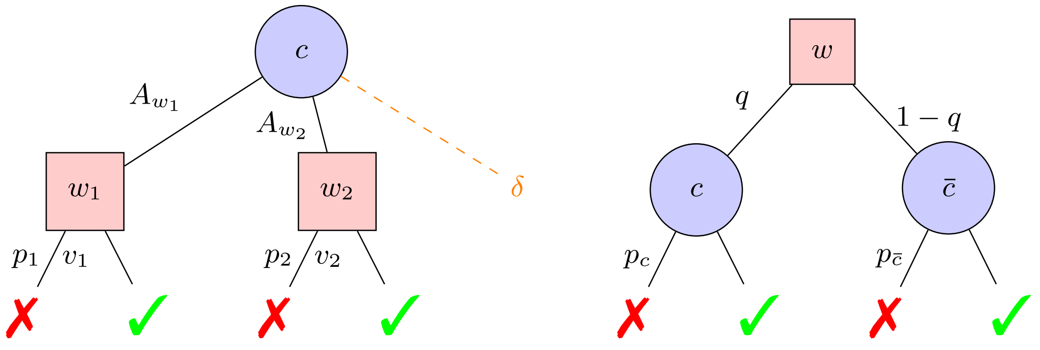
Circuit node branching covariance can be non-negative. In Fig. 15, we only looked at covariances at subset nodes. There is also a contribution from circuit nodes. Consider the tree in Fig. 16, where we also included a weight-1 node, which will cause a non-negative Cov term as we show in the following. The reason is that the branching does not involve two nodes that contain the branching ratios with opposite sign as above but both contributions are positive (B_0 and B_1 in the tree below).

The failure rate from Fig. 16 is

$$p_L = A_1(q_j(B_0q_k + B_1q_l) + (1-q_j)C_0q_m) \tag{B27}$$

and the variance

$$\begin{aligned}\mathrm{Var}(p_L) &= A_1^2\mathrm{Var}[q_j(B_0q_k + B_1q_l) + (1-q_j)C_0q_m]\\ &= A_1^2[\mathrm{Var}(q_j(B_0q_k + B_1q_l)) + \mathrm{Var}((1-q_j)C_0q_m)\\ &\quad + 2\,\mathrm{Cov}(q_j(B_0q_k + B_1q_l), (1-q_j)C_0q_m)]\\ &= A_1^2[\mathrm{Var}(q_jB_0q_k) + \mathrm{Var}(q_jB_1q_l)\end{aligned}$$



(a) $c.p = A_{w_1} p_1 + A_{w_2} p_2 \; (+\delta)$ (b) $w.p = q \times p_c + (1-q) \times p_{\bar{c}}$

FIG. 17. Any event tree can be constructed recursively by branchings from (a) circuit nodes c to subset nodes w_1, w_2, ... and (b) subset nodes w to circuit nodes c and $\bar{c}$. At higher levels, e.g., if c is the root node, the indicated failures correspond to coarse-grained subtree failure rates. If any w does not have a circuit node successor but is followed by termination of the protocol, the base case of recursion is reached. Failure rates and variances of upper bounds can be calculated by including the cutoff error δ to a circuit node's failure rate. Its value must be determined according to the rules described at the end of Appendix B 4.

$$\begin{aligned}
&+\ 2\,\mathrm{Cov}(q_j B_0 q_k, q_j B_1 q_l) + C_0^2 \mathrm{Var}((1-q_j)q_m) \\
&+ 2\,\mathrm{Cov}(q_j(B_0 q_k + B_1 q_l), (1-q_j)C_0 q_m)] \\
= {}& A_1^2\big[B_0^2\,\mathrm{Var}(q_j q_k) + B_1^2\,\mathrm{Var}(q_j q_l) \\
&+ 2B_0B_1\,\mathrm{Cov}(q_j q_k, q_j q_l) + C_0^2 \mathrm{Var}((1-q_j)q_m) \\
&+ 2\,\mathrm{Cov}(q_j(B_0 q_k + B_1 q_l), (1-q_j)C_0 q_m)\big]
\end{aligned} \tag{B28}$$

contains the negative Cov term of the form shown before but there is also the term

$$\mathrm{Cov}(q_j q_k, q_j q_l) = q_k q_l\,\mathrm{Var}(q_j) \tag{B29}$$

for which we used Eq. (B16). It clearly is a non-negative contribution that must be taken into account. We could use the Cauchy-Schwarz inequality (B17) to estimate

$$\begin{aligned}
&B_0^2\,\mathrm{Var}(q_j q_k) + B_1^2\,\mathrm{Var}(q_j q_l) + 2B_0B_1\,\mathrm{Cov}(q_j q_k, q_j q_l) \\
&\quad \leqslant \big(B_0\sqrt{\mathrm{Var}(q_j q_k)} + B_1\sqrt{\mathrm{Var}(q_j q_l)}\big)^2.
\end{aligned} \tag{B30}$$

However, the appearance of such terms can be prevented by algebraic manipulation of the variance terms so that we never need to explicitly take them into account, e.g., by recursively calculating subtree variances starting at the tree's leaves.

2. Lower bound uncertainty

It is straightforward to calculate the standard error σ_L on the lower bound p_L from Eq. (B1) using the general variance of linear combinations of random variables in Eq. (B8) and Goodman's formula for the variance of products in Eq. (B12). An iterative algorithmic procedure to calculate $\mathrm{Var}(p_L)$ will be given in Appendix C.

Let us also remark that, alternatively, failure rates and variances may also be calculated neatly by exploiting the recursive structure of the tree, namely by *recursively* "propagating" these quantities up from the leaf nodes of a tree until its root node. Any event tree can be built from the two components shown in Fig. 17. For any subset node w, we define its node rate $w.p$ (think of p as a property or an attribute of w), which can be determined as

$$w.p = w.q \times c.p + (1 - w.q) \times \bar{c}.p \tag{B31}$$

where $w.q$ $(1 - w.q)$ is the transition rate q from w to its subsequent circuit node c $(\bar{c})$ or to a "fail" ("success") outcome if w is a leaf node, i.e., w has no subsequent circuit node but the protocol terminates. Analogously, we define the node rate $c.p$ of a circuit node c as the coarse-grained failure rate, or the total failure rate of the subtrees below c. It can be calculated as $c.p = \sum_{w \in c} A_w\, w.p$ where w are all subset nodes directly subsequent to c and A_w is their respective binomial factor. The variance of $c.p$ could be calculated analogously via $c.v = \sum_{w \in c} A_w^2\, w.v$. The variance of any subset node w is obtained by Eqs. (B8) and (B12) as

$$\begin{aligned}
w.v = {}& \mathrm{Var}(w.p) = \mathrm{Var}[w.q \times c.p + (1 - w.q) \times \bar{c}.p] \\
= {}& \mathrm{Var}(w.q \times c.p) + \mathrm{Var}[(1 - w.q) \times \bar{c}.p] \\
&- \ 2\, c.p \times \bar{c}.p\,\mathrm{Var}(w.q) \\
= {}& w.q^2 \times c.v + \mathrm{Var}(w.q)\, c.p^2 + \mathrm{Var}(w.q)\, c.v \\
&+ \ (1 - w.q)^2\, \bar{c}.v + \mathrm{Var}(w.q)\, \bar{c}.p^2 \\
&+ \ \mathrm{Var}(w.q)\, \bar{c}.v - 2\, c.p \times \bar{c}.p\,\mathrm{Var}(w.q) \\
= {}& \mathrm{Var}(w.q)[c.p^2 + c.v + \bar{c}.p^2 + \bar{c}.v - 2\, c.p \times \bar{c}.p] \\
&+ \ w.q^2 \times c.v + (1 - w.q)^2\, \bar{c}.v \\
= {}& \mathrm{Var}(w.q)[(c.p - \bar{c}.p)^2 + c.v + \bar{c}.v] \\
&+ \ w.q^2 \times c.v + (1 - w.q)^2\, \bar{c}.v
\end{aligned} \tag{B32}$$

if w is not a leaf node and $w.v = \mathrm{Var}(w.q)$ if w is a leaf node.

This procedure can also be applied to calculate the upper bound p_U and its variance by including the single-circuit cutoff error $c.\delta = 1 - \sum_{w=0}^{w_{\max}} A_w^c$ into $c.p$ such that $c.p = \sum_{w \in c} A_w\, w.p + c.\delta$. Special cases for the value of δ for FT protocols or small number of shots are discussed in Appendix B 4.

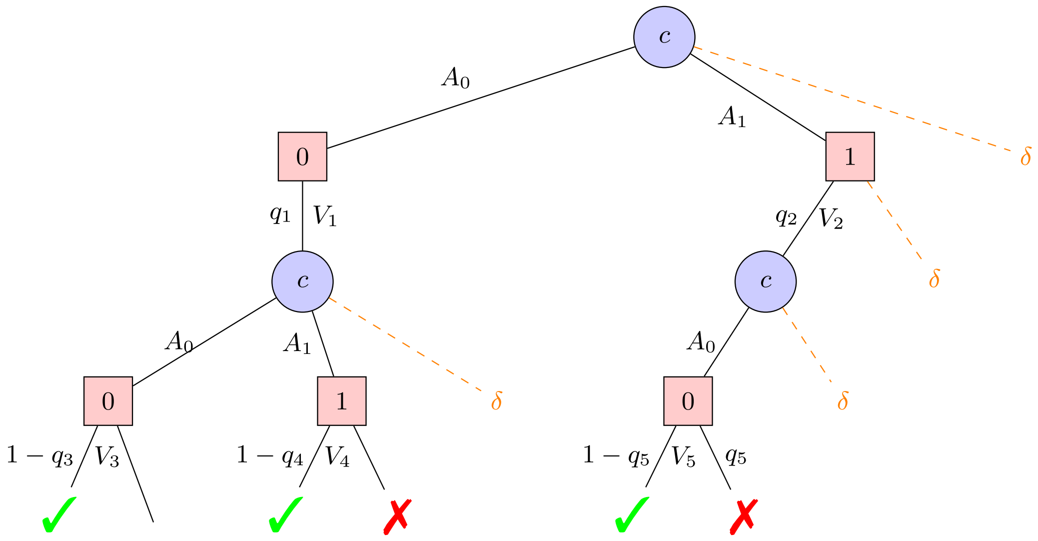


FIG. 18. Example tree structure from running the circuit c two consecutive times. No failure can happen at the end of the fault-free path, so $q_3 = V_3 = 0$. The paths indicated with dashed orange lines are yet unexplored but could potentially lead to failures and thus contribute to the cutoff error δ.

3. Upper bound uncertainty

Calculating σ_U iteratively requires careful consideration, namely, all unexplored branchings in an event tree that could potentially lead to failures when running more samples must be incorporated. As an example, take the event tree in Fig. 18 that could result from a protocol where the circuit c can be repeatedly run. Additionally, let us assume that after the first run of c, in principle, also another circuit $\bar{c}$ could be run. Not only do all unsampled subset nodes contribute to the cutoff error but also the incomplete branching with ratio q_2. Here we assume that, so far, only the branching to the left has been realized so that $q_2 = 1$ but the other branching possibility to the right, i.e., running another circuit $\bar{c}$, has not been realized yet due to a finite number of samples.

The cutoff error must be calculated as

$$\begin{aligned}\delta = {} & 1 - A_0 - A_1 + A_0 q_1(1 - A_0 - A_1)\\ & + A_1[q_2(1 - A_0) + (1 - q_2)]\end{aligned} \tag{B33}$$

and the upper bound reads

$$\begin{aligned}p_L + \delta = {} & 1 - A_0 - A_1 + A_0 q_1(1 - A_0 - A_1(1 - q_4))\\ & + A_1[q_2(1 - A_0(1 - q_5)) + (1 - q_2)],\end{aligned} \tag{B34}$$

where we explicitly have added the term $1 - q_2 = 0$: Although it does not contribute to δ, the finite variance $V_2 > 0$ leads to a contribution for the upper bound σ_U. We now calculate the variance of the upper bound

$$\begin{aligned}&\mathrm{Var}\{1 - A_0 - A_1 + A_0 q_1[1 - A_0 - A_1(1 - q_4)]\\ &\quad + A_1[q_2[1 - A_0(1 - q_5)] + (1 - q_2)]\} \equiv T_0 + T_1\end{aligned} \tag{B35}$$

$$\begin{aligned}T_0 &= A_0^2\,\mathrm{Var}\{q_1[1 - A_0 - A_1(1 - q_4)]\}\\ &= A_0^2\{(1 - A_0)^2 V_1 + A_1^2\,\mathrm{Var}[q_1(1 - q_4)]\\ &\quad - 2(1 - A_0)A_1(1 - q_4)V_1\} \sim p^2\end{aligned} \tag{B36}$$

$$\begin{aligned}T_1 &= A_1^2\,\mathrm{Var}\{q_2[1 - A_0(1 - q_5)] + (1 - q_2)\}\\ &= A_1^2(\mathrm{Var}\{q_2[1 - A_0(1 - q_5)]\}\\ &\quad + V_2 - 2[1 - A_0(1 - q_5)]V_2) \sim p^2\end{aligned} \tag{B37}$$

and observe that the standard error of the upper bound $\sigma_U = \sqrt{\mathrm{Var}(p_U)}$ scales linearly just as the failure rate estimator. Note that $V_1 = V_3 = 0$ since we assume that there can never be a branching off the fault-free path. If Fig. 18 represented an FT protocol, we would also have $q_5 = V_5 = 0$ since the weight-1 path can never lead to failure. However, $V_2 \neq 0$, which applies to QEC in the sense that the protocol can deal with different errors by realizing different circuit sequences. This means that we must properly deal with cutoff errors that arise from not having explored both possible circuit node successors from a subset node.

4. Cutoff error for FT protocols

As mentioned in Secs. I and III, protocols can be designed to have fault tolerance guarantees, in practice, e.g., FT-1 means that the occurrence of one single fault anywhere in the protocol cannot lead to failure. While the consequential improvement of the failure rate scaling p^* from $\mathcal{O}(p)$ to $\mathcal{O}(p^2)$ is exactly what we want from an FT theory perspective, it poses the following challenge here: For any protocol, the cutoff error must upper bound the worst-case possible subset failure rates that could occur within so far unsampled parts of the protocol's event tree. When left as above, the cutoff error calculation will be ignorant to the FT guarantee, i.e., in the calculation of δ we will assume that all unexplored paths lead to failure, which will give a scaling as $\delta = \mathcal{O}(p)$. But this would render the estimator overly pessimistic and the relative error would diverge for small p. However, we can exploit the fact that also the unexplored paths may contain cases where failure can be excluded when provided an FT guarantee.

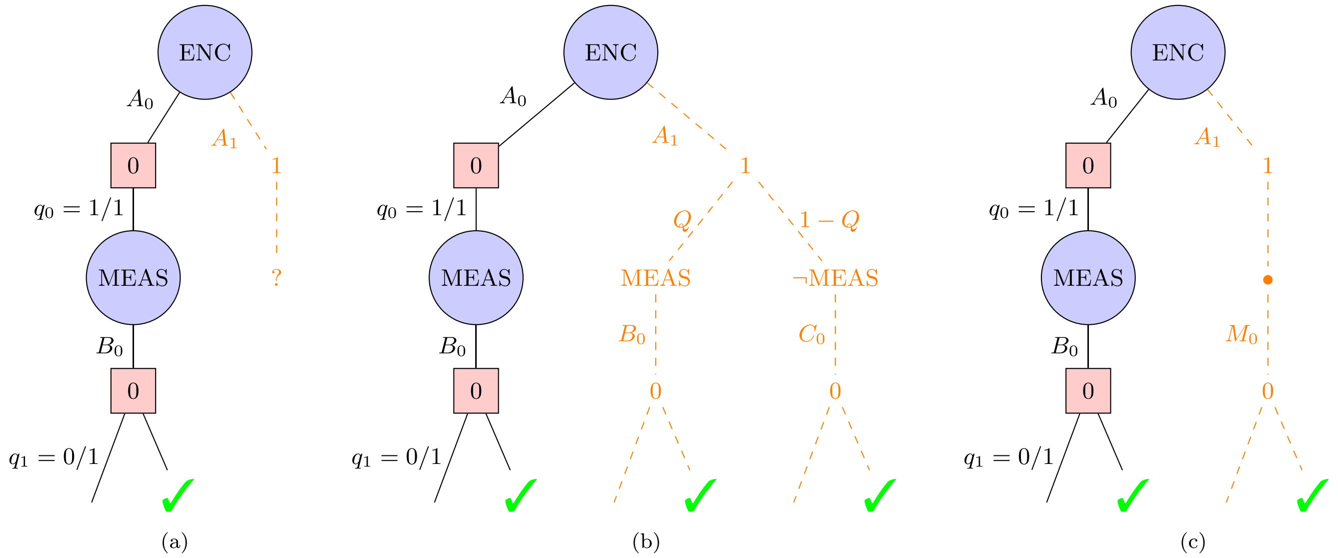


FIG. 19. Example trees illustrate the cutoff error (orange, no frames) after a single shot of a protocol containing the circuits ENC and MEAS has been run (black frames). It is assumed that the protocol consists of two stages, which is not the case in general. (a) What is the worst case contribution of the orange 1-subset to the failure rate?, (b) Two possible circuit branchings within the unsampled tree have an FT path, and (c) The FT paths can be upper bounded by a single node • with binomial factors $A^{\bullet}_w = M_w$.

In the following, we show in two steps how to choose the upper bound such that its scaling is of the same order in p as the failure rate, i.e., the relative error stays constant when scaling p. We only consider the special case of fault tolerance towards a single fault in a QEC code of distance $d = 3$ explicitly. The first step considers the branching ratios at a single stage within the tree. The second step then deals with circuit sequences, i.e., the worst case tree depth.

Figure 19(a) shows the challenge, which is special to an FT protocol. While for a non-FT protocol the cutoff error contribution at a circuit node called ENC would easily be calculated as $1 - A_0$, we now do not want to assume anymore that all paths branching off into the 1-subset of ENC lead to failure in the worst case. This is because we know beforehand that, since the protocol is FT, no 1-path can ever lead to failure.

Two possible paths are shown in Fig. 19(b). Their total weights are A_1QB_0 and $A_1(1-Q)C_0$. Reversely, if we do not exclude these paths from our cutoff error, we would estimate $\delta \sim 1 - A_0 \sim p$. Although this would be a correct worst case bound to state, it unnecessarily overestimates the a priori known worst case and causes the relative error δ/p_L of $p_L \sim p^2$ to diverge at $p \to 0$ since then $\delta \sim p$. At the orange subset node "1", we cannot know a priori what the branching ratio Q will be exactly without sampling (again, note that the point of δ is to make a statement about unsampled parts of the tree). It can also not be queried from the protocol since this would require testing all possible weight-1 faults at this stage. Also, here we are interested in analytically stating the worst case error. We do not even know what the two possible resulting circuits after the branching are since we cannot determine beforehand how the protocol deals with every individual fault. As an example, let us for the moment assume, the subsequent circuit could either be MEAS or ¬MEAS (a circuit other than MEAS). The cutoff error contribution of the orange subtree would then be $A_1(Q(1-B_0) + (1-Q)(1-C_0))$. One of the two binomial factors of the subsequent circuits will be smaller than the other. Here, let us assume for illustration (and without loss of generality) that $B_0 > C_0$. Then we can upper bound the cutoff error contribution as

$$\begin{aligned} &A_1(Q(1-B_0) + (1-Q)(1-C_0)) \\ &\quad \leqslant A_1(Q(1-C_0) + (1-Q)(1-C_0)) = A_1(1-C_0), \end{aligned} \tag{B38}$$

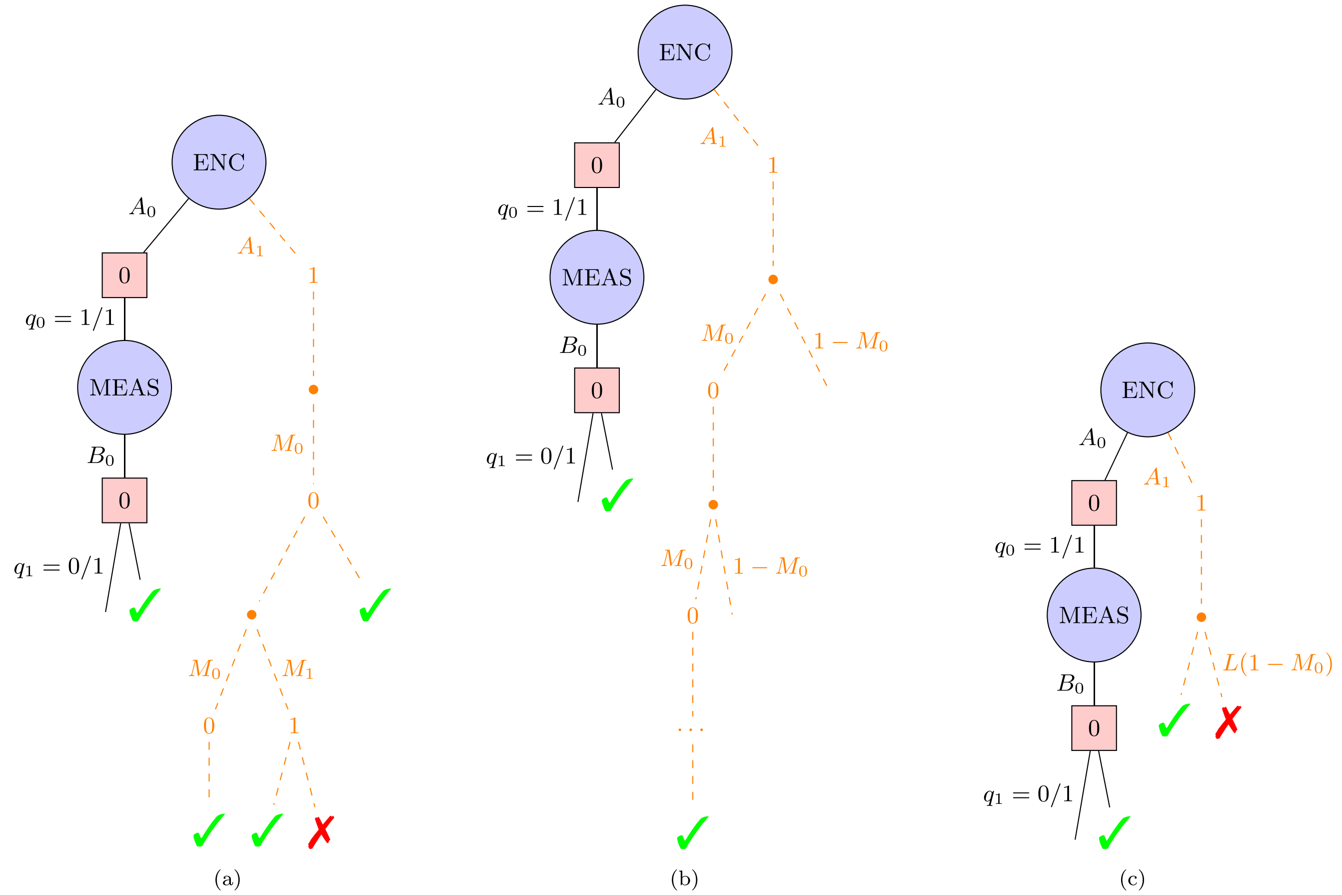
which has the desired effect to remove the unknown branching ratio Q from our estimation. In practice, we know the smallest possible binomial factor independently from sampling. Thus, as shown in Fig. 19(c), we can replace the branching with Q by a single node with a binomial factor

$$M_0 = \min_{c \in \mathcal{C} \in \mathcal{P}} \left(A^c_0\right). \tag{B39}$$

Now, as a second step, we need to consider longer circuit sequences, as for example in Fig. 19(c). There could follow another circuit at the left leaf in the orange subtree, where, again, a second 1-subset could lead to failure. This scenario is depicted in Fig. 20(a). Although the weight sequence 1-0 can never lead to failure after termination of the protocol, the single fault could cause another branching so that a third circuit will be run. Then, the weight sequence 1-0-1 is not FT anymore and *can* lead to failure. We must consider this case in our cutoff error.

This event is covered by the tree in Fig. 20(b). There is the cutoff contribution $A_1M_0(1-M_0)$ that contains this particular example. Now, consider the fictitious case that the orange subtree could potentially be infinitely deep. Then, the cutoff contribution would be

$$\begin{aligned} &A_1[(1-M_0) + M_0(1-M_0) + M_0^2(1-M_0) + \cdots] \\ &= A_1(1-M_0)\sum_{k=0}^{\infty} M_0^k = A_1(1-M_0)\frac{1}{1-M_0} = A_1 \sim p, \end{aligned} \tag{B40}$$



(a) (b) (c)

FIG. 20. Example trees illustrate the cutoff error contributions from longer circuit sequences. The actual circuit sequence is unimportant and can be bounded by M_0 as shown in Fig. 19. Only the maximum length of all FT paths L is needed. (a) A contribution to the failure rate can come from a deeper stage within the orange subtree (1-0-1), further down than another FT path (1-0), (b) All cutoff error contributions are of the form $1 - M_0$ within the orange subtree illustrated by the two branchings to the right, and (c) The worst case contribution of the orange 1-subset to the failure rate is $LA_1(1 - M_0)$.

which destroys our initial goal to formulate a cutoff error that scales like p_L. However, we know beforehand that the tree can actually never be infinitely deep. All FT paths are finite because no real protocol would go on forever without dangerous faults and since the protocol was actively designed, its maximum length L in case only correctable errors occur, is always known. This changes the cutoff contribution to read

$$\begin{aligned} &A_1((1 - M_0) + M_0(1 - M_0) + M_0^2(1 - M_0) \\ &\quad + \cdots + M_0^{L-1}(1 - M_0)) \\ &= A_1(1 - M_0)\sum_{k=0}^{L-1} M_0^k \leqslant LA_1(1 - M_0) \sim p^2, \end{aligned} \tag{B41}$$

as we aimed for.

The maximum FT protocol lengths for the protocols considered here are given in Table I alongside with the circuits that belong to the protocol.

There are different cases, illustrated in Fig. 21, for how to consider the (FT) cutoff error and calculate the variance of the upper bound to the failure rate for a general tree:

(i) If the total path weight up to a subset node is 0, we add no additional branching circuit because we assume there is only one fault-free path.

(ii) Only for FT protocol: If the total path weight up to a subset node is 1, and the branching ratio q of the subset node is equal to 1 and there is only one circuit node below the subset node but the other one is still unknown, then we have to add a complementary branching with factor $1 - q$ and a δ node with value $L(1 - M_0)$. While this node does not contribute directly to the cutoff error, it will contribute to the variance of the upper bound via $\mathrm{Var}(1 - q) \neq 0$.

(iii) If the total path weight up to a subset node is 2 (for an FT protocol) / 1 (for a non-FT protocol) or larger and the branching ratio q of the subset node is equal to 1 and there is only one circuit node below the subset node but the other one is still unknown, then we have to add a complementary branching with factor $1 - q$ and a δ node with value 1 below the complementary branching with $1 - q$.

(iv) Only for FT protocol: If the total path weight up to a circuit node is 0 and if the circuit node does not have a 1-subset, we have to add a virtual one. The virtual 1-subset has a δ node of value $L(1 - M_0)$. The δ node of the circuit node then has the usual value.

As a concluding remark, we note that to practically calculate the cutoff error within a given stage i, we can determine the factor $l = L - i$ to account for the longest possible "rest of the sequence" that could still follow from the current pro-

TABLE I. Maximum FT protocol lengths for the protocols considered in Sec. III.

Protocol $\mathcal{P}$	max FT length L	circuits $c \in \mathcal{C} \in \mathcal{P}$
GHZ state preparation	1	GHZ
Deterministic $\lvert 0\rangle_L$	4	ENC, SZ, X7, MEAS
Stabilizer $\lvert 0\rangle_L$	8	SX1a, SX2a, SX3a, SX1b, SX2b, SX3b, NFS, MEAS
Repeat until success	2+1	ENC, MEAS

tocol stage, where i circuits have been executed already. In Fig. 20(c) we may replace with L with l. However, `qsample` always uses the fixed value L for the cutoff error calculation.

There is one extra peculiarity to a "repeat until success" protocol. Here we mean, for instance, a protocol that consists of repeating an encoding circuit ENC until a verification measurement signals success and then measuring all qubits as represented by the node MEAS (see Refs. [16,28] for examples). Since all qubits are always reset upon execution of the encoding circuit, we can only count its maximum length (of 2) from the last reset on. For example, a path ENC-1-ENC-0-MEAS-0 is possible to realize. However, it is not possible to realize an FT path longer than $2 + t$, where $t = 1$. For example, a path ENC-1-ENC-1-ENC-0-MEAS-0 would be possible but it is of order p^2 and thus covered by the cutoff error contribution $A_1(1 - M_0)$ in Fig. 20(b). For a "repeat until success" protocol, one may achieve tighter bounds to the true logical failure rate by adjusting which leaf node variances are set to zero and which values are assigned to δ-nodes based on the last qubit reset.

APPENDIX C: NUMERICAL IMPLEMENTATION: `QSAMPLE`

In this Appendix, we describe the numerical implementation of dynamical subset sampling in the python package `qsample`, which is deployed together with this paper at GitHub [18].

1. Numerical calculation of estimators

We model a quantum *protocol* as a graph containing quantum circuits at its vertices and transition rules as edges, which define the order of circuit execution. At the beginning of sampling, the protocol graph starts at a predefined unique node, *root*, and is subsequently traversed by choosing transitions dynamically during runtime conditioned on the measurement results of all circuits in the sequence of the current shot as visualized in Fig. 1(d). Naturally, as our goal is to obtain failure rates of a protocol, we check at the end of the circuit sequence whether a failure has occurred. Protocol terminations that do not lead to failure are not explicitly represented as a node, i.e., there is no *success* node in a `qsample` protocol.

Individual samples of a protocol are efficiently tracked in a tree data structure, analogous to Fig. 1(c), the *sample tree*. Repeatedly executing a noisy protocol thus nondeterministically yields circuit sequences, modelled as paths of the sample tree, which each start at a unique root node and end at a *leaf* node. We restrict the maximum number of successors in the protocol graph to at most two, since we assume that any quantum protocol can be cast in such a form. By this constraint, we can model the transition of a noisy circuit to the next by a Bernoulli random variable X, which can take on the values 0,1 corresponding to which of two circuits is chosen. The transition rates to either circuit are then estimated by $\mathbb{E}[X]$ and $1 - \mathbb{E}[X]$, respectively. Additionally, the binomial factors A_w

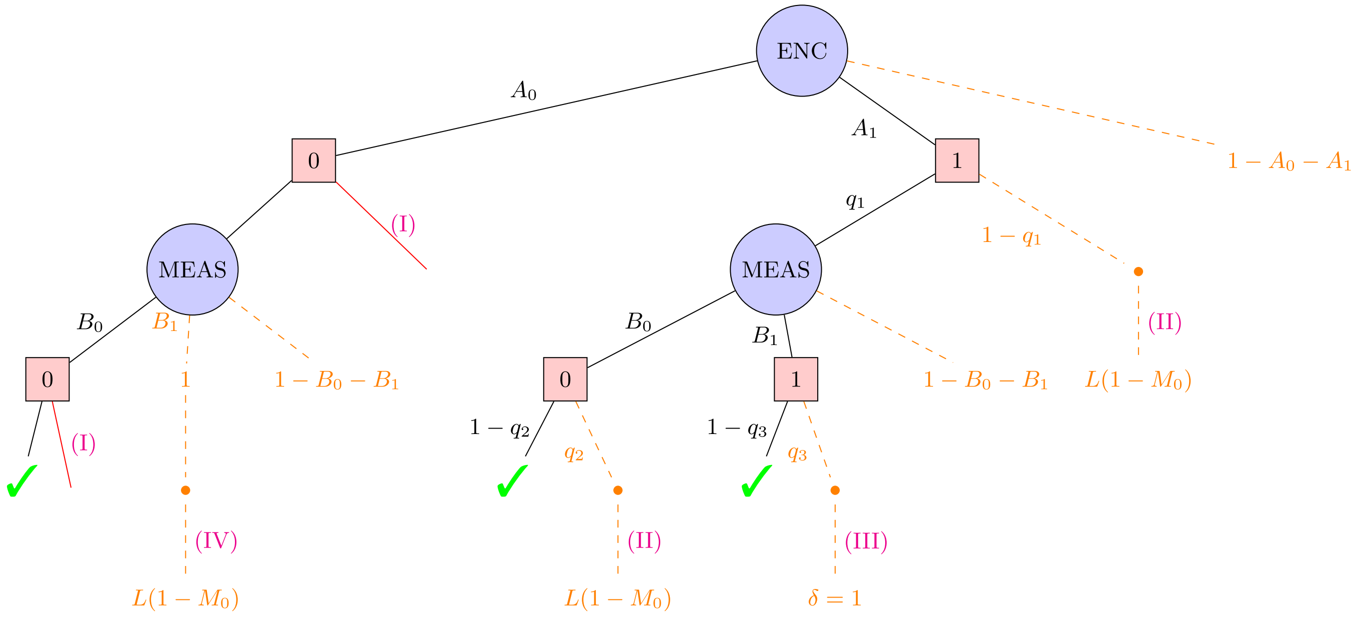


FIG. 21. An FT protocol tree illustrating the four cases of the FT cutoff error. If the protocol were non-FT, the cases (II) and (IV) would be replaced with instances of case (III).

and cutoffs δ, determined as the four cases in Appendix B 4 prescribe, are included in the sample tree, branching off from any circuit node [see Fig. 1(c)]. Thus, we define a path $P \in \mathcal{P}$ as a unique sequence of circuits and binomial factors (one per circuit), represented by circuit nodes and subset nodes in the tree, ending on one of three terminal nodes, *fail*, *no-fail*, or δ. As the fail and no-fail leaves represent the binary event whether or not a failure occurred, we denote a path ending on a fail leaf as *fail path* $P_f \in \mathcal{P}_f$. Furthermore, we define a path ending on a δ node as $P_\delta \in \mathcal{P}_\delta$. The lower bound p_L is then implemented as the sum over all fail paths $\mathcal{P}_f$ as

$$p_L = \mathbb{E}\left[\sum_{P_f \in \mathcal{P}_f} \prod_{n \in P_f} A_w^n X_n\right] = \sum_{P_f \in \mathcal{P}_f} \prod_{n \in P_f} A_w^n \mathbb{E}[X_n], \tag{C1}$$

where we denote the binomial factor for a subset of weight w corresponding to the circuit node n as A_w^n. The second equality of Eq. (C1) follows from independence of X_n within the same path. The upper bound p_U additionally includes the sum over all path products $\mathcal{P}_\delta$, where the product is over all circuit transition rates $\mathbb{E}[X]$ and binomial factors A_w in a path, i.e.,

$$p_L + \delta = p_L + \sum_{P_\delta \in \mathcal{P}_\delta} \prod_{n \in P_\delta} A_w^n f(n), \tag{C2}$$

where $f(n) = \mathbb{E}[X_n]$, except for δ nodes at which the value $f(n) = \delta_n$ is determined by the cutoff error of the particular circuit node [see a single yellow box in Fig. 1(c) and Eq. (16)]. The uncertainty on the failure rate bounds is calculated in a similar way as

$$\mathrm{Var}[p_L] = \mathrm{Var}\left[\sum_{P_f \in \mathcal{P}_f} \prod_{n \in P_f} A_w^n X_n\right], \tag{C3}$$

$$\mathrm{Var}[p_L + \delta] = \mathrm{Var}\left[\sum_{P_f \in \mathcal{P}_f} \prod_{n \in P_f} A_w^n X_n + \sum_{P_\delta \in \mathcal{P}_\delta} \prod_{n \in P_\delta} A_w^n X_n\right]. \tag{C4}$$

If we express the result of a path product as a random variable $P_i = \prod_{n \in P} A_w^n X_n$, we have for the variance of the sum

$$\begin{aligned}\mathrm{Var}\left[\sum_i P_i\right] &= \sum_{i,j} \mathrm{Cov}[P_i, P_j] \\ &= \underbrace{\sum_i \mathrm{Var}[P_i]}_{\text{path variance}} + \underbrace{2 \sum_{i<j} \mathrm{Cov}[P_i, P_j]}_{\text{overlap variance}},\end{aligned} \tag{C5}$$

where we denote the first sum as the *path variance* and the second as *overlap variance*. It is worth mentioning that we have to consider the possibility that any path can potentially end in a failure. Even though a failure may not have been sampled for a path yet, its variance can be nonzero. These *virtual* paths have contributions to both the path and overlap variance. Exempt from this are the fault-free path and any path with a total fault weight of at most 1 for FT-1 protocols.

Algorithm 2. Variance of a path.

```
Input: Path $P$
Output: Variance of $P$, $\mathrm{Var}[P]$
for node $n$ in path $P$ do
    if $n$ is a subset node then
        $A \leftarrow A \times A_w^n$
    else if $n$ is a circuit node then
        $\mathbb{E}[P]^2 \leftarrow \mathbb{E}[P]^2 \times \mathbb{E}[X_n]^2$
        $\mathbb{E}[P^2] \leftarrow \mathbb{E}[P^2] \times (\mathrm{Var}[X_n] + \mathbb{E}[X_n]^2)$
    end if
end for
return $A^2(\mathbb{E}[P^2] - \mathbb{E}[P]^2)$
```

The path variances are calculated via Goodman's formula. We first consider the variance of a path $P \in \mathcal{P}$. Due to independence of all X_i within a path P we get

$$\begin{aligned}\mathrm{Var}[P] &= \prod_i \left(A_w^i\right)^2 \mathbb{E}\left[X_i^2\right] - \prod_i \left(A_w^i\right)^2 \mathbb{E}[X_i]^2 \\ &= \left(\prod_i A_w^i\right)^2 \left(\prod_i \left(\mathrm{Var}(X_i) + \mathbb{E}[X_i]^2\right) - \prod_i \left(\mathbb{E}[X_i]^2\right)\right).\end{aligned} \tag{C6}$$

It is straightforward to implement Eq. (C6) on our tree data structure as listed in Algorithm 2. Here, the value of a circuit node (subset node) is its associated transition rate (binomial factor). The values of δ nodes are determined according to Appendix B 4. The variance of a circuit node is determined by the Wilson score interval [see Eq. (C12)] and the variance of a subset node is zero.

The covariance terms are harder to obtain. We note that only overlapping paths have a nonzero covariance and that overlapping paths share one unique node, below which the paths split and above which all nodes are common. Further, we can distinguish two situations in which the last common node is either a circuit node so that the subsequent branching is determined by the random variable X or a subset node so that the subsequent branching is determined by binomial factor A_w. Both situations are depicted in Fig. 22.

Consider first the case where paths split at a circuit node. If two paths $P_0 = UA_0 \sum_i V_i$ and $P_1 = UA_1 \sum_i W_i$ split at X_m

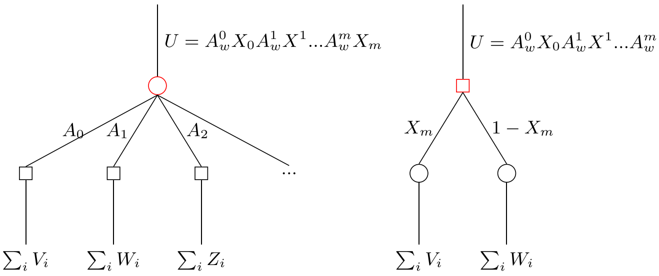


FIG. 22. Paths intersecting at a circuit node (circle) on the left and at subset node (rectangle) on the right marked in red. The common *upstream* path contains all common random variables X_i and constants A_w^i. Its path product can be interpreted as a new random variable U. The random variables V_i, W_i, Z_i represent the rest of those paths *downstream*.

with common path U, the overlap variance reads

$$\mathrm{Cov}[P_0, P_1] = \underbrace{\mathbb{E}\left[A_0 \sum_i V_i\right] \mathbb{E}\left[A_1 \sum_i W_i\right]}_{\text{subtree sums } \mathcal{S}_w} \mathrm{Var}[U]. \tag{C7}$$

As the downstream paths V_i and W_i are all unique they are independent of each other, thus their covariance is fully determined by the variance of the upstream path scaled by the weights of its downstream *subtrees*, i.e., the sum of all paths starting at X_m and ending on, e.g., failure nodes. By Eq. (C5), we have to consider all unique combinations of paths overlapping at this node, which leads to the total contribution to the overlap variance from paths diverging at circuit nodes,

$$2\sum_{i<j} \mathrm{Cov}[P_i, P_j] = 2\sum_{i<j} \mathbb{E}\left[A_w^i \sum_k V_k^i\right] \mathbb{E}\left[A_w^j \sum_l W_l^j\right] \mathrm{Var}[U_i] \tag{C8}$$

In case paths split at subset nodes, there are always only two downstream paths. However, as the last common random variable in both paths are complements of each other, we obtain for the two paths $P_0 = UX_m \sum_i V_i$ and $P_1 = U(1 - X_m) \sum_i W_i$ the covariance

$$\begin{aligned}\mathrm{Cov}[P_0, P_1] &= \underbrace{\mathbb{E}\left[\sum_i V_i\right] \mathbb{E}\left[\sum_i W_i\right]}_{\text{subtree sums } \mathcal{S}_k} \mathrm{Cov}(UX_m, U(1 - X_m)) \\ &= \mathbb{E}\left[\sum_i V_i\right] \mathbb{E}\left[\sum_i W_i\right] \\ &\quad \times (\mathrm{Var}[U]\mathbb{E}[X_m] - \mathrm{Var}(UX_m)).\end{aligned} \tag{C9}$$

Thus, the total contribution to the overlap variance for paths overlapping at a subset node is

$$\begin{aligned}2\sum_{i<j} \mathrm{Cov}[P_i, P_j] &= 2\sum_{i<j} \mathbb{E}\left[\sum_k V_k^i\right] \mathbb{E}\left[\sum_l W_l^j\right] \\ &\quad \times \left(\mathrm{Var}[U_i]\mathbb{E}\left[X_m^i\right] - \mathrm{Var}\left(U_i X_m^i\right)\right).\end{aligned} \tag{C10}$$

In the python package `qsample` the calculation of variances of p_L and $p_L + \delta$ is performed as described in pseudocode in Algorithm 3.

Algorithm 3. Variance on lower bound p_L (upper bound $p_L + \delta$) to failure rate.

```
Input: Sample tree $\mathcal{T}$
Output: Variance $\mathrm{Var}[p_L]$ ($\mathrm{Var}[p_L + \delta]$)
Get $\mathcal{P} = \mathcal{P}_f$ ($\mathcal{P}_f \cup \mathcal{P}_\delta$) from $\mathcal{T}$
Calculate sum of path variances $V_1 \leftarrow \sum_{P \in \mathcal{P}} \mathrm{Var}[P]$
Get overlaps $O \leftarrow \{O_i = \max_d (P \cap P') \forall P, P' \in \mathcal{P}\}$
  with node depth $d$
for $O_i \in O$ do
    if $O_i$ is a circuit node then
        Calculate subtree sums $\mathcal{S}_w = A_w^m \sum_j \mathbb{E}[X_j^w]$ for $j > m$
        Calculate upstream path variance $\mathrm{Var}[O_i]$
        $V_2^i \leftarrow 2\mathrm{Var}[O_i] \sum_{w<w'} \mathcal{S}_w \mathcal{S}_{w'}$
    else if $O_i$ is a subset node then
        Calculate subtree sums $\mathcal{S}_k = \sum_j \mathbb{E}[X_j^k]$ for $j > m$
        Calculate upstream path variance $\mathrm{Var}[O_i]$
        Calculate path variance $\mathrm{Var}[O_i X_m]$
        Get expectation value $\mathbb{E}[X_m]$
        $V_2^i \leftarrow 2(\mathrm{Var}[O_i]\mathbb{E}[X_m] - \mathrm{Var}[O_i X_m]) \prod_k \mathcal{S}_k$
    end if
    $V_2 \leftarrow V_2 + V_2^i$
end for
return $V_1 + V_2$
```

2. Uncertainty intervals

The sampling error for MC sampling can be estimated by the Wald interval

$$\varepsilon_{\mathrm{MC}} = \sqrt{\frac{\hat{p}(1 - \hat{p})}{N}} \tag{C11}$$

so that for a large number of samples $N \to \infty$ the true failure rate p^* is likely to be found in the 68% confidence interval $[\hat{p} - \varepsilon_{\mathrm{MC}}, \ \hat{p} + \varepsilon_{\mathrm{MC}}]$. It is known that for $\hat{p}$ estimations that are close to or equal to zero or one after a finite but potentially small number of samples the Wald interval suffers from irregularities. These can be prevented using the Wilson score interval [33] instead, which is bounded by

$$p_\pm = \frac{1}{1 + \frac{z_{\alpha/2}^2}{N}} \left(\hat{p} + \frac{z_{\alpha/2}^2}{2N} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{N} + \frac{z_{\alpha/2}^2}{4N^2}} \right) \tag{C12}$$

at confidence level $1 - \alpha$ where z is the quantile function of the normal distribution. The sampling error for all numerical simulations in `qsample` is given as the Wilson score interval (C12) at a confidence level of 68% ($z_{\alpha/2} = 1$) in a symmetric form $[\hat{p} - \frac{p_+ - p_-}{2}, \hat{p} + \frac{p_+ - p_-}{2}]$.

3. Selection criterion

a. Binomial-factor-based choice

The default procedure in `qsample` simply selects the subsets randomly based on the binomial distribution $A_w(p_{\max})$ of subsets for each circuit at run-time. For each circuit, we draw a random number $r \in [0, 1]$ and choose the weight w for which

$$\sum_{i=0}^{w-1} A_i \leqslant r \leqslant \sum_{i=0}^{w} A_i \tag{C13}$$

where the left sum is understood to be equal to zero if $w = 0$.

Furthermore, sampling of the weight-0 subset of a circuit can be prohibited if the circuit under fault-free conditions always yields the same deterministic outcome, for instance for a "repeat until success" protocol. In this case, the sums in Eq. (C13) should start from $i = 1$ while each term is rescaled such that the overall subset selection probability remains unity.

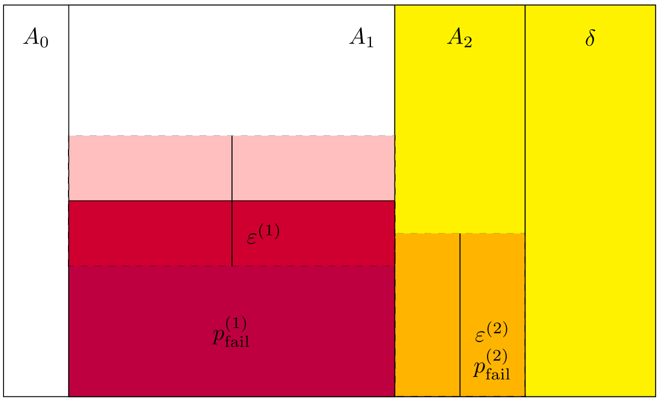


FIG. 23. Box representation of different fault-weight subsets for a protocol that consists of a single circuit with binomial factors A_i. The subset failure rate $p_{\text{fail}}^{(1)}$ has a sampling uncertainty $\varepsilon^{(1)}$. Opening the 2-fault subset amounts to reducing the cutoff error δ by the binomial factor A_2 but in turn increasing the overall sampling uncertainty. This is because an initial estimate from few shots of $p_{\text{fail}}^{(2)}$ comes with a large uncertainty $\varepsilon^{(2)}$. (The 0-fault subset is assumed to be the largest subset, but we draw it on the side for better visibility of the higher-weight subsets.)

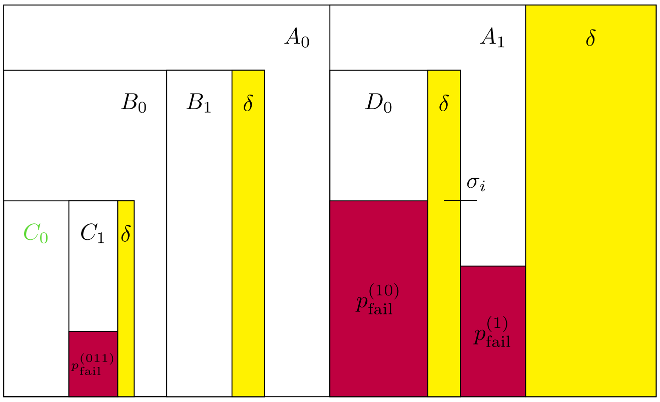


FIG. 24. Circuit sequences can be represented by nested boxes as in Fig. 1(c). A subbox can take up at most all the space provided by the enclosing box. The fault-free path has the total binomial factor $A_0B_0C_0$ here. Within the subbox labeled by the green C_0, no failure can occur [compare to the green path in Fig. 1(a)].

b. ERU choice

We implement the ERU criterion whose defining feature is that it always chooses the subset so that the next shot maximizes the expected reduction of the total uncertainty of $\hat{p}$. It balances the trade-off between continuing to sample in a subset that is already known—thereby decreasing the sampling error—or going to a new subset where, initially, we would end up with a relatively large sampling uncertainty but we get to reduce the cutoff error δ.

It requires calculation of the shifted transition rates $q_i^{\pm}$ as given by Eq. (21), where it is assumed that the next sample yields a ± measurement outcome. Let us assume that any q_i is determined by

$$q_i = \frac{m_{\mathcal{C}_i}}{N_{\mathcal{C}_i}} \tag{C14}$$

where $m_{\mathcal{C}_i}$ counts the numbers of positive measurement outcomes and $N_{\mathcal{C}_i}$ is the total number of samples of the particular node q_i belongs to. Then

$$\begin{aligned} q_i^+ &\equiv \frac{m_{\mathcal{C}_i}+1}{N_{\mathcal{C}_i}+1} = \frac{m_{\mathcal{C}_i}+1}{N_{\mathcal{C}_i}}\frac{N_{\mathcal{C}_i}}{N_{\mathcal{C}_i}+1} = \left(\frac{m_{\mathcal{C}_i}}{N_{\mathcal{C}_i}} + \frac{1}{N_{\mathcal{C}_i}}\right)\frac{N_{\mathcal{C}_i}}{N_{\mathcal{C}_i}+1} \\ &= \left(q_i + \frac{1}{N_{\mathcal{C}_i}}\right)\frac{N_{\mathcal{C}_i}}{N_{\mathcal{C}_i}+1} \quad \text{and} \end{aligned} \tag{C15}$$

$$q_i^- \equiv \frac{m_{\mathcal{C}_i}}{N_{\mathcal{C}_i}+1} = q_i\frac{N_{\mathcal{C}_i}}{N_{\mathcal{C}_i}+1} \tag{C16}$$

as stated in Eq. (21).

The expected reduction of uncertainty (ERU) at a node i is defined by Eq. (22). An illustration is given in Fig. 23. It depicts a known subset failure rate $p_{\text{fail}}^{(1)}$ in the 1-subset with the yellow marked cutoff error δ, which includes A_2. $p_{\text{fail}}^{(1)}$ has a sampling uncertainty of $\varepsilon^{(1)}$ so that the total length of the uncertainty interval η about $\hat{p}$, i.e., the difference between upper and lower bound according to Eq. (20), is $2\varepsilon^{(1)}+\delta$ [note that the estimator $\hat{p}$ is centered between upper and lower bound according to Eq. (15) and that, since we only consider a single circuit, the sampling uncertainties of the upper and lower bound are identical so that $2\varepsilon^{(1)} = \sigma_L + \sigma_U$ in this example]. If we were to spend the next shot in the 2-subset, the total uncertainty would change to read $2\varepsilon^{(1)} + 2\varepsilon^{(2)} + \delta - A_2$. Another shot in the 1-subset would just decrease $\varepsilon^{(1)}$ and keep δ unchanged. The choice that reduces the *expected* total uncertainty the most, is the one that is taken according to the ERU criterion. In this example we have

$$\begin{aligned} \Delta(2) &= (2\varepsilon^{(1)} + \delta) - \big[p_{\text{fail}}^{(2)} \cdot (2(\varepsilon^{(1)} + \varepsilon^{(2)}|_+) + \delta - A_2) \\ &\quad + \big(1 - p_{\text{fail}}^{(2)}\big) \cdot (2(\varepsilon^{(1)} + \varepsilon^{(2)}|_-) + \delta - A_2)\big] \\ &= A_2 - 2\varepsilon^{(2)}|_+, \end{aligned} \tag{C17}$$

$$\begin{aligned} \Delta(1) &= (2\varepsilon^{(1)} + \delta) - \big[p_{\text{fail}}^{(1)} \cdot (2\varepsilon^{(1)}|_+ + \delta) \\ &\quad + \big(1 - p_{\text{fail}}^{(1)}\big) \cdot (2\varepsilon^{(1)}|_- + \delta)\big] \\ &= 2\varepsilon^{(1)} - \big[p_{\text{fail}}^{(1)} \cdot 2\varepsilon^{(1)}|_+ + \big(1 - p_{\text{fail}}^{(1)}\big) \cdot 2\varepsilon^{(1)}|_-\big]. \end{aligned} \tag{C18}$$

Note that, although there is no current value for $p_{\text{fail}}^{(2)}$ yet, this is not a problem for evaluating Eq. (C17) since we have $\varepsilon^{(2)}|_+ = \varepsilon^{(2)}|_-$, which is always the case when we open a new subset.

The more complicated situation of circuit sequences is illustrated in Fig. 24. When choosing a weight subset at a given stage, we must consider the total uncertainty of all the contained boxes (and the cutoff error of the given stage). Choosing 0 in the outermost box, for example, will be done under consideration of $p_{\text{fail}}^{(011)}$ and contributions to δ from all stages. For choosing the 1-subset, we must consider the total uncertainty of failure in all subboxes of A_1. This especially now includes the uncertainty about the size of the subboxes, as indicated by σ_i, the uncertainty of the branching ratio q_i that leads from A_1 to the circuits with binomial factors D_w. Note that this uncertainty does not exist for the 0-subsets since the fault-free path is deterministic. The splitting is further illustrated in Fig. 25.

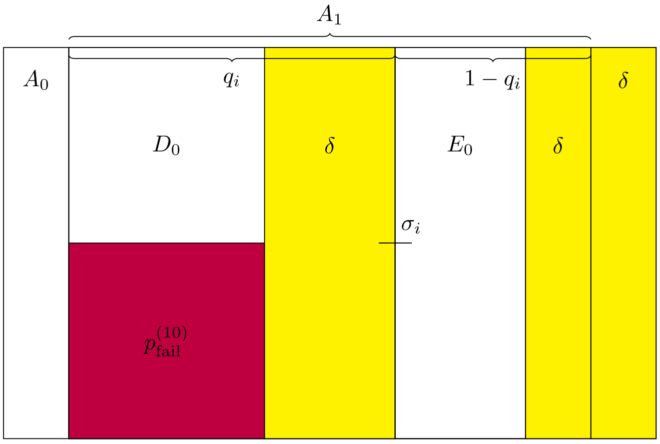


FIG. 25. Subboxes of the A_1 box, labeled D and E, are weighted by the branching ratio q_i or $1 - q_i$ respectively [cf. Fig. 1(c)]. Since the branching ratio estimate q_i has an uncertainty σ_i, the size of any subbox is not determined with total confidence.

When probing a new subset, i.e., reducing the total cutoff error, for the calculation of the ERU, it is reasonable to assume that a newly opened subset will lead to failure with probability $1/2$. Since nothing is known about the subtree below this subset node and we do not get a measurement outcome that would reveal some new information about the tree, we can only assume failure with a finite probability, e.g., $1/2$. Note that while assuming a failure probability of 0 is unrealistic for a non-FT path, assuming a failure probability of 1 would cause the cutoff error to not change at all compared to the current tree value. This is different from the single-circuit case discussed above since there we can immediately evaluate the effect of the failure or no-failure outcome after choosing a subset, which cannot be done for a circuit sequence. It is not strictly necessary to assume an initial failure rate of $1/2$ but any finite value is fine. It could, for example, be set as the current average value of all subset failure rates $\frac{1}{N_w}\sum_w p_{\rm fail}^{(w)}$. This is not yet implemented in `qsample`.

APPENDIX D: NOISE MODEL

For the examples that we show in Sec. III, we employ circuit-level depolarizing noise. We apply the fault-operators

$$F_1 \in \{\sigma_i, \forall i \in [1, 2, 3]\}, \tag{D1}$$

$$F_2 \in \{\sigma_i \otimes \sigma_j, \forall i, j \in [0, 1, 2, 3]\} \setminus I \otimes I, \tag{D2}$$

where σ_i are the Pauli matrices and $\sigma_0 = I$ is the identity operation. Single-qubit faults F_1 are applied uniformly with probability $p_1/3$ and are not applied with probability $1 - p_1$. The two-qubit faults F_2 are applied uniformly with probability $p_2/15$ and are not applied with probability $1 - p_2$. This means that

(1) a single-qubit gate is followed by a Pauli fault F_1 drawn uniformly and independently from $\{X, Y, Z\}$ with probability $p_1/3$,

(2) a two-qubit gate is followed by a two-Pauli fault F_2 drawn uniformly and independently from $\{I, X, Y, Z\}^{\otimes 2} \backslash I \otimes I$ with probability $p_2/15$,

(3) qubit initialization is flipped (e.g., $|0\rangle \rightarrow |1\rangle$, $|+\rangle \rightarrow |-\rangle$) with probability $2p_1/3$ and

(4) qubit measurements yield a flipped result ($\pm 1 \rightarrow \mp 1$) with probability $2p_1/3$.

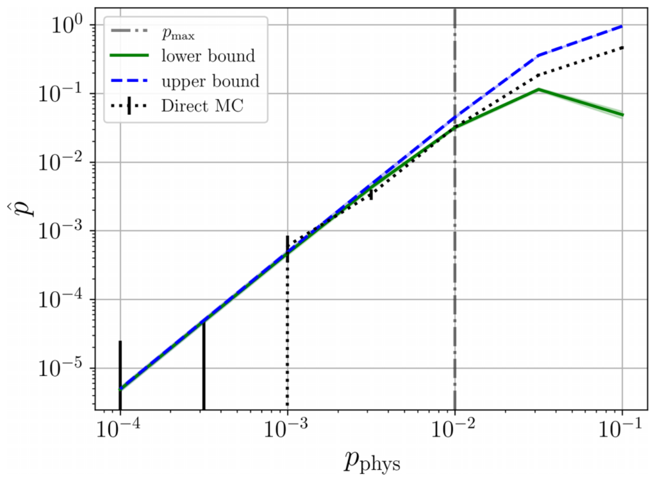


FIG. 26. Improved bounds for the logical failure rate estimation for the state preparation protocol via flag-FT stabilizer measurements. The only difference to Fig. 14 is that we sampled at an increased $p_{\rm max} = 10^{-2}$ (gray vertical dash-dotted line). Here, the cutoff error vanishes in the limit $p_{\rm phys} \rightarrow 0$ since 10^4 shots with DSS and binomial factor choice suffice to sample all relevant subsets.

APPENDIX E: TIGHTER BOUNDS FOR FLAG-FT STABILIZER MEASUREMENTS

The relatively large cutoff error that we found for the flag-FT stabilizer measurement example protocol in Sec. III D can be reduced in practice without changing the dynamical subset sampling algorithm. In Fig. 26 we show a much smaller cutoff error that we obtain by increasing the maximal physical fault rate to $p_{\rm max} = 10^{-2}$ instead of sampling at $p_{\rm max} = 10^{-3}$. While the cutoff error remains large at $p_{\rm max} = 10^{-2}$, this increase has the effect that more different fault-weight subsets are chosen by the binomial factor choice criterion. As a consequence, the space of possible paths is more thoroughly explored and the bounds to the true protocol failure rate tighten as we lower $p_{\rm phys}$, in agreement with the direct MC data points, which have a growing uncertainty interval as in Fig. 14.

[1] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[2] H. J. Briegel, D. E. Browne, W. Dür, R. Raussendorf, and M. Van den Nest, Measurement-based quantum computation, Nat. Phys. **5**, 19 (2009).

[3] R. Jozsa, An introduction to measurement based quantum computation, in *Quantum Information Processing—From Theory to Experiment*, NATO Science Series, III: Computer and Systems Sciences, edited by D. G. Angelakis, M. Christandl, A. Ekert, A. Kay, and S. Kulik, Vol. 199 (IOS Press, Amsterdam, 2006), pp. 137–158.

[4] L.-M. Duan and C. Monroe, Colloquium: Quantum networks with trapped ions, Rev. Mod. Phys. **82**, 1209 (2010).

[5] W. Kozlowski and S. Wehner, Towards large-scale quantum networks, in *Proceedings of the Sixth Annual ACM International Conference on Nanoscale Computing and Communication* (Association for Computing Machinery, New York, 2019), p. 1.

[6] S. Bravyi and A. Vargo, Simulation of rare events in quantum error correction, Phys. Rev. A **88**, 062308 (2013).

[7] A. M. Steane, Overhead and noise threshold of fault-tolerant quantum error correction, Phys. Rev. A **68**, 042322 (2003).

[8] R. Raussendorf, J. Harrington, and K. Goyal, Topological fault-tolerance in cluster state quantum computation, New J. Phys. **9**, 199 (2007).

[9] A. V. Rynbach, A. Muhammad, and J. Kim, A quantum computing performance simulator based on circuit failure probability and fault path counting, ACM J. Emer. Tech. Comput. Syst. **14**, 1 (2018).

[10] A. G. Fowler, Analytic asymptotic performance of topological codes, Phys. Rev. A **87**, 040301(R) (2013).

[11] M. Li, M. Gutiérrez, S. E. David, A. Hernandez, and K. R. Brown, Fault tolerance with bare ancillary qubits for a [7, 1, 3] code, Phys. Rev. A **96**, 032341 (2017).

[12] C. J. Trout, M. Li, M. Gutiérrez, Y. Wu, S.-T. Wang, L. Duan, and K. R. Brown, Simulating the performance of a distance-3 surface code in a linear ion trap, New J. Phys. **20**, 043038 (2018).

[13] M. Gutiérrez, M. Müller, and A. Bermúdez, Transversality and lattice surgery: Exploring realistic routes toward coupled logical qubits with trapped-ion quantum processors, Phys. Rev. A **99**, 022330 (2019).

[14] M. Li, Fault-tolerance on near-term quantum computers and subsystem quantum error correcting codes, Ph.D. thesis, Georgia Institute of Technology, 2020.

[15] A. Paetznick, C. Knapp, N. Delfosse, B. Bauer, J. Haah, M. B. Hastings, and M. P. da Silva, Performance of planar floquet codes with majorana-based qubits, PRX Quantum **4**, 010310 (2023).

[16] S. Heußen, L. Postler, M. Rispler, I. Pogorelov, C. D. Marciniak, T. Monz, P. Schindler, and M. Müller, Strategies for a practical advantage of fault-tolerant circuit design in noisy trapped-ion quantum computers, Phys. Rev. A **107**, 042422 (2023).

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

[18] D. Winter and S. Heußen, qsample, https://github.com/dpwinter/qsample (2023).

[19] https://dpwinter.github.io/qsample/.

[20] D. M. Greenberger, M. A. Horne, and A. Zeilinger, Going beyond bell's theorem, in *Bell's Theorem, Quantum Theory and Conceptions of the Universe* (Springer, New York, 1989), p. 69.

[21] P. W. Shor, Fault-tolerant quantum computation, in *Proceedings of 37th Conference on Foundations of Computer Science* (IEEE, Piscataway, NJ, 1996), p. 56.

[22] A. Steane, Multiple-particle interference and quantum error correction, Proc. R. Soc. London A **452**, 2551 (1996).

[23] A. M. Steane, Error correcting codes in quantum theory, Phys. Rev. Lett. **77**, 793 (1996).

[24] H. Goto, Minimizing resource overheads for fault-tolerant preparation of encoded states of the steane code, Sci. Rep. **6**, 19578 (2016).

[25] R. Chao and B. W. Reichardt, Quantum error correction with only two extra qubits, Phys. Rev. Lett. **121**, 050502 (2018).

[26] C. Chamberland and M. E. Beverland, Flag fault-tolerant error correction with arbitrary distance codes, Quantum **2**, 53 (2018).

[27] C. Ryan-Anderson, J. G. Bohnet, K. Lee, D. Gresh, A. Hankin, J. P. Gaebler, D. Francois, A. Chernoguzov, D. Lucchetti, N. C. Brown, T. M. Gatterman, S. K. Halit, K. Gilmore, J. A. Gerber, B. Neyenhuis, D. Hayes, and R. P. Stutz, Realization of real-time fault-tolerant quantum error correction, Phys. Rev. X **11**, 041058 (2021).

[28] L. Postler, S. Heußen, I. Pogorelov, M. Rispler, T. Feldker, M. Meth, C. D. Marciniak, R. Stricker, M. Ringbauer, R. Blatt *et al.*, Demonstration of fault-tolerant universal quantum gate operations, Nature (London) **605**, 675 (2022).

[29] We use the term "location" to refer to an operation (a gate, a measurement or an initialization) at a specific position in the circuit.

[30] The fault weight w is different from the Pauli weight of an operator: the number of qubits on which it has non-trivial support, i.e., is not the identity I. For example, a $X \otimes Z$ fault on a CNOT gate (and all other circuit elements fault-free) is a 1-weight-fault of Pauli weight 2.

[31] L. D. Brown, T. T. Cai, and A. DasGupta, Interval estimation for a binomial proportion, Stat. Sci. **16**, 101 (2001).

[32] A. Dauphin, M. Müller, and M. A. Martin-Delgado, Efficient algorithm to compute the berry conductivity, New J. Phys. **16**, 073016 (2014).

[33] E. B. Wilson, Probable inference, the law of succession, and statistical inference, J. Am. Stat. Assoc. **22**, 209 (1927).

[34] Since we assume there exists only one fault-free path [see Fig. 1(a)], the uncertainties to branch off from this path are identical to zero.

[35] We use the terms "branching ratio" and "transition rate" interchangeably.

[36] D. P. DiVincenzo and P. W. Shor, Fault-tolerant error correction with efficient quantum codes, Phys. Rev. Lett. **77**, 3260 (1996).

[37] M. B. Plenio, V. Vedral, and P. L. Knight, Conditional generation of error syndromes in fault-tolerant error correction, Phys. Rev. A **55**, 4593 (1997).

[38] S. J. Devitt, W. J. Munro, and K. Nemoto, Quantum error correction for beginners, Rep. Prog. Phys. **76**, 076001 (2013).

[39] D. Aharonov and M. Ben-Or, Fault-tolerant quantum computation with constant error rate, SIAM J. Comput. **38**, 1207 (2008).

[40] S. Aaronson and D. Gottesman, Improved simulation of stabilizer circuits, Phys. Rev. A **70**, 052328 (2004).

[41] C. Gidney, Stim: a fast stabilizer circuit simulator, Quantum **5**, 497 (2021).

[42] D. Gottesman, Quantum fault tolerance in small experiments, arXiv:1610.03507.

[43] A. Y. Kitaev, Fault-tolerant quantum computation by anyons, Ann. Phys. **303**, 2 (2003).

[44] 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).

[45] H. Bombin and M. A. Martin-Delgado, Topological quantum distillation, Phys. Rev. Lett. **97**, 180501 (2006).

[46] A. G. Fowler, Two-dimensional color-code quantum computation, Phys. Rev. A **83**, 042310 (2011).

[47] N. P. Breuckmann and J. N. Eberhardt, Quantum low-density parity-check codes, PRX Quantum **2**, 040101 (2021).

[48] D. Gottesman, The heisenberg representation of quantum computers, arXiv:quant-ph/9807006.

[49] We distinguish between faults, errors and failures: A fault is an instance of noise on a circuit location, e.g., a single Pauli operator applied after an ideal operation. An error is the result of a faulty location at the end of the circuit, e.g., a Pauli fault propagated to a larger weight error. A failure is a wrongly decoded state as the result of an error of weight $w > \lfloor \frac{d-1}{2} \rfloor$ where d is the code distance.

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

[51] T. Ohno, G. Arakawa, I. Ichinose, and T. Matsui, Phase structure of the random-plaquette z2 gauge model: accuracy threshold for a toric quantum memory, Nucl. Phys. B **697**, 462 (2004).

[52] H. G. Katzgraber, H. Bombin, and M. A. Martin-Delgado, Error threshold for color codes and random three-body ising models, Phys. Rev. Lett. **103**, 090501 (2009).

[53] R. Raussendorf and J. Harrington, Fault-tolerant quantum computation with high threshold in two dimensions, Phys. Rev. Lett. **98**, 190504 (2007).

[54] A. J. Landahl, J. T. Anderson, and P. R. Rice, Fault-tolerant quantum computing with color codes, arXiv:1108.5738.

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

[56] A. Bermudez, X. Xu, R. Nigmatullin, J. O'Gorman, V. Negnevitsky, P. Schindler, T. Monz, U. G. Poschinger, C. Hempel, J. Home, F. Schmidt-Kaler, M. Biercuk, R. Blatt, S. Benjamin, and M. Müller, Assessing the progress of trapped-ion processors towards fault-tolerant quantum computation, Phys. Rev. X **7**, 041061 (2017).

[57] D. M. Debroy, M. Li, S. Huang, and K. R. Brown, Logical performance of 9 qubit compass codes in ion traps with crosstalk errors, Quantum Sci. Technol. **5**, 034002 (2020).

[58] P. Parrado-Rodríguez, C. Ryan-Anderson, A. Bermudez, and M. Müller, Crosstalk suppression for fault-tolerant quantum error correction with trapped ions, Quantum **5**, 487 (2021).

[59] M. Gutiérrez, L. Svec, A. Vargo, and K. R. Brown, Approximation of realistic errors by clifford channels and pauli measurements, Phys. Rev. A **87**, 030302(R) (2013).

[60] M. B. Plenio and P. L. Knight, The quantum-jump approach to dissipative dynamics in quantum optics, Rev. Mod. Phys. **70**, 101 (1998).

[61] T. Jones, A. Brown, I. Bush, and S. C. Benjamin, Quest and high performance simulation of quantum computers, Sci. Rep. **9**, 10736 (2019).

[62] M. Foss-Feig, A. Tikku, T.-C. Lu, K. Mayer, M. Iqbal, T. M. Gatterman, J. A. Gerber, K. Gilmore, D. Gresh, A. Hankin *et al.*, Experimental demonstration of the advantage of adaptive quantum circuits, arXiv:2302.03029.

[63] C. Chamberland and A. W. Cross, Fault-tolerant magic state preparation with flag qubits, Quantum **3**, 143 (2019).

[64] D. K. Tuckett, S. D. Bartlett, and S. T. Flammia, Ultrahigh error threshold for surface codes with biased noise, Phys. Rev. Lett. **120**, 050505 (2018).

[65] D. K. Tuckett, A. S. Darmawan, C. T. Chubb, S. Bravyi, S. D. Bartlett, and S. T. Flammia, Tailoring surface codes for highly biased noise, Phys. Rev. X **9**, 041031 (2019).

[66] 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).

[67] J. P. Bonilla Ataides, D. K. Tuckett, S. D. Bartlett, S. T. Flammia, and B. J. Brown, The xzzx surface code, Nat. Commun. **12**, 2172 (2021).

[68] A. S. Darmawan, B. J. Brown, A. L. Grimsmo, D. K. Tuckett, and S. Puri, Practical quantum error correction with the xzzx code and kerr-cat qubits, PRX Quantum **2**, 030345 (2021).

[69] Q. Xu, N. Mannucci, A. Seif, A. Kubica, S. T. Flammia, and L. Jiang, Tailored xzzx codes for biased noise, Phys. Rev. Res. **5**, 013035 (2023).

[70] D. Gottesman and I. L. Chuang, Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations, Nature (London) **402**, 390 (1999).

[71] J. Hilder, D. Pijn, O. Onishchenko, A. Stahl, M. Orth, B. Lekitsch, A. Rodriguez-Blanco, M. Müller, F. Schmidt-Kaler, and U. G. Poschinger, Fault-tolerant parity readout on a shuttling-based trapped-ion quantum computer, Phys. Rev. X **12**, 011032 (2022).

[72] S. Kirkpatrick Jr., C. D. Gelatt Jr., and M. P. Vecchi, Optimization by simulated annealing, Science **220**, 671 (1983).

[73] S. Krinner, N. Lacroix, A. Remm, A. Di Paolo, E. Genois, C. Leroux, C. Hellings, S. Lazar, F. Swiadek, J. Herrmann *et al.*, Realizing repeated quantum error correction in a distance-three surface code, Nature (London) **605**, 669 (2022).

[74] J. Soch, The book of statistical proofs, https://statproofbook.github.io (2023).

[75] L. A. Goodman, On the exact variance of products, J. Am. Stat. Assoc. **55**, 708 (1960).

[76] L. A. Goodman, The variance of the product of k random variables, J. Am. Stat. Assoc. **57**, 54 (1962).