Supplemental Material: "Error Suppression for Arbitrary-Size Black Box Quantum Operations"

I. NOISE MODEL

We assume we are given an apparatus that is meant to carry out the unitary U but instead carries out a CPTP channel $\mathcal{U}(\cdot)$, with $R+1$ Kraus operators $\{K_0, K_1, ..., K_R\}$, where K_0 is the Kraus operator closest to U. Now, in our setup we append ancillas to this apparatus. Let I_A be the collective identity operation on all parts of the system that aren't part of the apparatus. Then, assuming the ancillas are noiseless, a reasonable way to extend the noise model to the rest of the system is to use Kraus operators

$$K_i^{(s)} = I_A \otimes K_i, i = 0, 1, 2, ..., R. \tag{S1}$$

This assumption is justified when the cross-talk between the apparatus and the rest of the system is small. Subsequently, we will suppress the I_A and (S) superscript, writing K_i to be understood in the above sense.
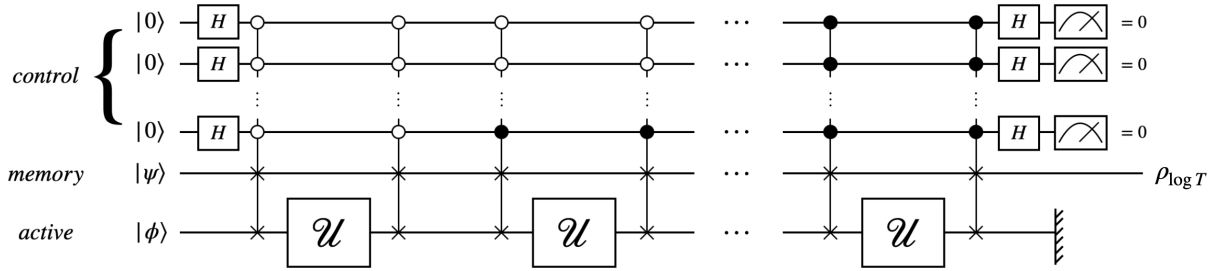
II. SMALL PARAMETER CASE

FIG. S1: General circuit for gate-based EF that makes some number T calls to an apparatus that implements the noisy process $\mathcal{U}$ as follows: (1) Prepare the intended input state $|\psi\rangle$ in one register, and the null state $|\phi\rangle$ in another. We refer to the former as the *memory register* and the latter as the *active register*. (2) Prepare $\log T$ ancillas in the *control register* in the equal superposition state $|++...+\rangle$. (3) For each branch $|i\rangle$ in the computational basis (i.e. $|0...00\rangle, |0...01\rangle$, and so forth), swap the state $|\psi\rangle$ into the active register. We then run the active register through our apparatus, given by noisy process $\mathcal{E}$, before swapping the registers back for the same branch. Do this for every branch of the superposition. (4) Finally, perform $\log T$ parallel measurements on the control register in the X basis, and post-select for every measurement being $+1$. This corresponds to projecting the control register back onto the equal superposition state $|++...+\rangle$.

We begin by deriving the scaling behaviour of gate-based EF in the primary case of interest – that is, we are given an apparatus that is already pretty good, and everything else in the circuit is noiseless. Formally, we say that there exists a Kraus operator K_0 that is ε-close to the ideal unitary U, such that we can write $K_0 = U - \varepsilon\xi$ for some small ε and some suitably normalized operator ξ with $\|\xi\| = 1$. Given such a form for K_0, it follows that $K_i \sim O(\sqrt{\varepsilon})$ for $i \neq 0$. We will also assume that $1 - F_0 \sim O(\varepsilon)$, where F_0 is the native fidelity of the apparatus prior to error fitration. Note that this assumption is not compatible with $K_0 = V$ where V is some coherent unitary error, as we will see in Sec. II C, where we consider those cases separately.

A. $T = 2$ case

We will first work out the $T = 2, \log T = 1$ case. We index all the relevant density matrices and fidelities with $\log T$. The unnormalized output of Fig. S1 when $\log T = 1$ is the density matrix

$$\tilde{\rho}^{(1)} = \frac{1}{2}\mathcal{U}(|\psi\rangle\langle\psi|) + \frac{1}{2}\sum_{i=0,j=0}^{R} K_i|\psi\rangle\langle\psi|K_j^\dagger \text{Tr}\left(\rho_\phi K_i^\dagger K_j\right), \tag{S2}$$

where we note that $F_0 = \langle\psi|U^\dagger\mathcal{U}(|\psi\rangle\langle\psi|)U|\psi\rangle$. The conclusions of this section are independent of the state $|\phi\rangle$, so we can choose any $|\phi\rangle$.

The probability of success is given by the trace of this density matrix,

$$P_s^{(1)} = \frac{1}{2} + \frac{1}{2} \sum_{i=0,j=0}^{R} \langle \psi | K_j^\dagger K_i | \psi \rangle \text{Tr}\left(\rho_\phi K_i^\dagger K_j \right). \tag{S3}$$

Recall our assumption that $K_0 = U - \varepsilon \xi$. By the arguments in the previous section, the terms in the sum where $i, j \neq 0$ are $O(\varepsilon^2)$. As such, when computing the fidelity we only have to keep terms where either $i, j = 0$. We first show that the probability of success is close to 1:

$$P_s^{(1)} = \text{Tr}\tilde{\rho}^{(1)} = \frac{1}{2} + \frac{1}{2} \langle \psi | K_0^\dagger K_0 | \psi \rangle \text{Tr}(\rho_\phi K_0^\dagger K_0) + \frac{1}{2} \sum_{i \neq 0} \left(\langle \psi | K_0^\dagger K_i | \psi \rangle \text{Tr}(\rho_\phi K_i^\dagger K_0) + h.c. \right) + O(\varepsilon^2)$$

$$= 1 - \frac{1}{2} \varepsilon \left(\langle \psi | \xi^\dagger U | \psi \rangle + \langle \psi | U^\dagger \xi | \psi \rangle + \text{Tr}(\rho_\phi \xi^\dagger U) + \text{Tr}(\rho_\phi U^\dagger \xi) \right)$$

$$+ \frac{1}{2} \sum_{i \neq 0} \left(\langle \psi | U^\dagger K_i | \psi \rangle \text{Tr}(\rho_\phi K_i^\dagger U) + h.c. \right) + O(\varepsilon^2)$$

which is ε-close to 1. Similarly, working out the inner product with the ideal state,

$$\langle \psi | U^\dagger \tilde{\rho}^{(2)} U | \psi \rangle = \frac{1}{2} F_0 + \frac{1}{2} \langle \psi | U^\dagger K_0 | \psi \rangle \langle \psi | K_0^\dagger U | \psi \rangle \text{Tr}\left(\rho_\phi K_0^\dagger K_0 \right)$$

$$+ \frac{1}{2} \sum_{i=1}^{R} \left(\langle \psi | U^\dagger K_i | \psi \rangle \langle \psi | K_0^\dagger U | \psi \rangle \text{Tr}\left(\rho_\phi K_i^\dagger K_0 \right) + \langle \psi | U^\dagger K_0 | \psi \rangle \langle \psi | K_i^\dagger U | \psi \rangle \text{Tr}\left(\rho_\phi K_0^\dagger K_i \right) \right) + O(\varepsilon^2)$$

$$= \frac{1}{2} F_0 + \frac{1}{2} - \frac{1}{2} \varepsilon \left(\langle \psi | U^\dagger \xi | \psi \rangle + \langle \psi | \xi^\dagger U | \psi \rangle + \text{Tr}\left(\rho_\phi \xi^\dagger U \right) + \text{Tr}\left(\rho_\phi U^\dagger \xi \right) \right)$$

$$+ \frac{1}{2} \sum_{i=1}^{R} \left(\langle \psi | U^\dagger K_i | \psi \rangle \text{Tr}\left(\rho_\phi K_i^\dagger U \right) + h.c. \right) + O(\varepsilon^2)$$

Dividing and expanding the denominator in ε, we see that all the remaining $O(\varepsilon)$ terms cancel out, and we are left with

$$(1 - F)_1 = 1 - \frac{\langle \psi | U^\dagger \tilde{\rho}^{(2)} U | \psi \rangle}{\text{Tr}\tilde{\rho}^{(2)}} \tag{S4}$$

$$= \frac{1}{2}(1 - F)_0 + O(\varepsilon^2), \tag{S5}$$

i.e. the infidelity is halved with a single ancilla qubit.

B. General T case

Consider a particular trajectory for a state $|\psi\rangle$ that goes through the circuit with $\log T$ ancilla qubits. At each call of the apparatus it gets a different Kraus operator applied to it, which we write $K_{i_1}, K_{i_2}, \ldots, K_{i_T}$. Whether this operator is applied to the state $|\psi\rangle$ or $|\phi\rangle$ depends on the branch of the ancilla. For the $|t\rangle$ branch of the ancilla, K_{i_t} is applied to $|\psi\rangle$ whereas the rest are applied to $|\phi\rangle$. We can write this as,

$$|t\rangle|\psi\rangle|\phi\rangle \mapsto |t\rangle K_{i_t}|\psi\rangle K_{i_T} K_{i_{T-1}} \ldots K_{i_{t+1}} K_{i_{t-1}} \cdots K_{i_1}|\phi\rangle. \tag{S6}$$

for $t = 0, 1, 2, \ldots, T-1$, with $|t\rangle$ corresponding to the binary representation of t. We can think of $\mathbf{i}$ as a vector with T components $i_j \in \{0, 1, \ldots, R\}$. The entire vector $\mathbf{i}$ indexes one possible outcome of the circuit.

For ease of notation we will define,

$$\overline{K_{i_t}} = K_{i_T} K_{i_{T-1}} \ldots K_{i_{t+1}} K_{i_{t-1}} \cdots K_{i_1} \tag{S7}$$

To get the full density matrix, we must sum over all possible vectors $\mathbf{i} \in \{0, 1, \ldots, R\}^{\otimes T}$. This gives the density matrix after the T applications of the black box as,

$$\sum_{\mathbf{i}} \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} |t\rangle K_{i_t}|\psi\rangle \overline{K}_{i_t}|\phi\rangle \right] \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} |t\rangle K_{i_t}|\psi\rangle \overline{K}_{i_t}|\phi\rangle \right]^\dagger$$

After measuring the control register, post-selecting on $|+\rangle^{\otimes \log T}$ and tracing out the active register,

$$\tilde{\rho}^{(\log T)} = \frac{1}{T^2} \sum_{\mathbf{i}} \sum_{t=1}^{T} \sum_{q=1}^{T} K_{i_t} |\psi\rangle \langle\psi| K_{i_q}^\dagger \mathrm{Tr}\rho_\phi \overline{K}_{i_q}^\dagger \overline{K}_{i_t}$$

First we consider the terms where $t = q$. These give

$$\frac{1}{T^2} \sum_{t=1}^{T} \sum_{\mathbf{i}} K_{i_t} |\psi\rangle \langle\psi| K_{i_t}^\dagger \mathrm{Tr}\rho_\phi \overline{K}_{i_t}^\dagger \overline{K}_{i_t}$$

where the sum over $\mathbf{i}$ can be refactored into

$$\sum_{\mathbf{i}} K_{i_t} |\psi\rangle \langle\psi| K_{i_t}^\dagger \mathrm{Tr}\rho_\phi \overline{K}_{i_t}^\dagger \overline{K}_{i_t} = \sum_{j_1=0}^{R} K_{j_1} |\psi\rangle \langle\psi| K_{j_1}^\dagger \sum_{j_T=0}^{R} \cdots \sum_{j_2=0}^{R} \mathrm{Tr}\rho_\phi K_{j_{2^T}}^\dagger \cdots K_{i_2}^\dagger K_{i_2} \cdots K_{i_{2^T}}$$

$$= \sum_{j_1=0}^{R} K_{j_1} |\psi\rangle \langle\psi| K_{j_1}^\dagger$$

$$= \mathcal{U}(|\psi\rangle\langle\psi|),$$

where we have used completeness property of the Kraus channel. Hence the density matrix can be rewritten

$$\tilde{\rho}^{(\log T)} = \frac{1}{T} \mathcal{U}(|\psi\rangle\langle\psi|) + \frac{1}{T^2} \sum_{\mathbf{i}} \sum_{t=1}^{T} \sum_{q\neq t} \left(K_{i_t} |\psi\rangle \langle\psi| K_{i_q}^\dagger \mathrm{Tr}\rho_\phi \overline{K}_{i_q}^\dagger \overline{K}_{i_t} \right)$$

1. General Success Probability

The success probability is given by

$$P_s^{(\log T)} = \mathrm{Tr}\tilde{\rho}^{(T)} = \frac{1}{T} + \frac{1}{T^2} \sum_{\mathbf{i}} \sum_{t=1}^{T} \sum_{q\neq t} \left(\langle\psi| K_{i_q}^\dagger K_{i_t} |\psi\rangle \mathrm{Tr}\rho_\phi \overline{K}_{i_q}^\dagger \overline{K}_{i_t} \right) \tag{S8}$$

In the general case, we can still get some idea of how the probability of success scales with T. Generically we expect the probability to go down with each error and the number of errors to scale with T, so we can immediately write down

$$P_S^{(T)} \sim 1 - O(T)\varepsilon \tag{S9}$$

Working through the above expressions more carefully, one finds that the exact scaling at small ε can be rigorously lower bounded by

$$P_S^{(T)} \geq 1 - T\varepsilon. \tag{S10}$$

As long as $T \ll \frac{1}{\varepsilon}$, we will have $P_S^{(T)}$ is ε-close to 1, we can proceed to calculate the infidelity. In fact, much tighter bounds can be given for $P_S^{(T)}$ under some conditions on the action of the Kraus operators K_i on the state $|\phi\rangle$.

2. Success probability under special favourable conditions

Now, suppose at least one of the following three conditions hold:

1. The Kraus operators are all mutually commuting, i.e. $[K_i, K_j] = 0$ for all $i, j = 0, 1, ..., R$.

2. $|\phi\rangle$ is stationary under the channel, i.e., $K_i|\phi\rangle \propto |\phi\rangle$ for all $i = 0, 1, ..., R$.

3. The circuit has a noise bias, and furthermore is bias preserving, such that we can choose some $|\phi\rangle$ to be immune to the dominant error (e.g. if the circuit only has phase flips, we can choose $|\phi\rangle = |0\rangle$).

Under any of these conditions, we will find that

$$\frac{1}{T^2}\sum_{\mathbf{i}}\sum_{t=1}^{T}\sum_{q>t}^{T}\left(\langle\psi|K_{i_q}^{\dagger}K_{i_t}|\psi\rangle\mathrm{Tr}\rho_{\phi}\overline{K}_{i_q}^{\dagger}\overline{K}_{i_t}+h.c.\right)$$

$$=\frac{T-1}{T}\sum_{i,j}\langle\psi|K_i^{\dagger}K_j|\psi\rangle\langle\phi|K_j^{\dagger}K_i|\phi\rangle$$

To make further progress, we note that not all trajectories containing Kraus operators where $i\neq 0$ or $j\neq 0$ are necessarily rejected by the protocol. Hence assuming that all such trajectories are rejected will give us a lower bound on the success probability, given by

$$P_S^{(\log T)}\geq\frac{1}{T}+\frac{T-1}{T}\langle\psi|K_0^{\dagger}K_0|\psi\rangle\langle\phi|K_0^{\dagger}K_0|\phi\rangle$$

$$\geq 1-4\varepsilon+\frac{4\varepsilon}{T}$$

$$\geq 1-4\varepsilon$$

3. General fidelity scaling

Having established that in the small parameter case, we can get $1-P_s\sim O(\varepsilon)$ either under the conditions of the previous section, or when T is not too large, we can now proceed to compute the fidelity associated with having $\log T$ control qubits.

In terms of $\tilde{\rho}^{(\log T)}, P_S^{\log T}$,

$$(1-F)_{\log T}=\frac{P_S^{\log T}-\langle U\psi|\tilde{\rho}^{(\log T)}|U\psi\rangle}{P_S^{\log T}}, \tag{S11}$$

where we can expand,

$$\langle U\psi|\tilde{\rho}^{(\log T)}|U\psi\rangle=\frac{1}{T}(1-F)_0+\frac{1}{T^2}\sum_{\mathbf{i}}\sum_{t=1}^{T}\sum_{q\neq t}\left(\langle\psi|U^{\dagger}K_{i_t}|\psi\rangle\langle\psi|K_{i_q}^{\dagger}U|\psi\rangle\mathrm{Tr}\rho_{\phi}\overline{K}_{i_q}^{\dagger}\overline{K}_{i_t}\right)$$

Since we are calculating quantities only to $O(\varepsilon)$, in the numerator we only need to keep compute terms for which either all the applied Kraus operators are K_0, or all the applied Kraus operators but one are K_0 (we can loosen this assumption when dealing with the bounded probability case).

First, we can combine the two terms in the numerator by factoring out the ϕ associated terms:

$$P_S^{\log T}-\langle U\psi|\tilde{\rho}^{(\log T)}|U\psi\rangle$$

$$=\frac{1}{T}(1-F)_0+\frac{1}{T^2}\sum_{\mathbf{i}}\sum_{t=1}^{T}\sum_{q\neq t}\left(\langle\psi|K_{i_q}^{\dagger}K_{i_t}|\psi\rangle-\langle\psi|U^{\dagger}K_{i_t}|\psi\rangle\langle\psi|K_{i_q}^{\dagger}U|\psi\rangle\right)\mathrm{Tr}\rho_{\phi}\overline{K}_{i_q}^{\dagger}\overline{K}_{i_t}$$

Consider the terms in the brackets. For the terms in the sum where all Kraus operators applied are K_0, we have

$$\langle\psi|K_0^{\dagger}K_0|\psi\rangle-\langle\psi|U^{\dagger}K_0|\psi\rangle\langle\psi|K_0^{\dagger}U|\psi\rangle$$

$$=1-\varepsilon\langle\psi|U^{\dagger}\xi|\psi\rangle-\varepsilon\langle\psi|\xi^{\dagger}U|\psi\rangle-1+\varepsilon\langle\psi|U^{\dagger}\xi|\psi\rangle+\varepsilon\langle\psi|\xi^{\dagger}U|\psi\rangle+O(\varepsilon^2)$$

$$=O(\varepsilon^2),$$

so these terms don't contribute.

For the terms where $i_q=0$ and $i_t\neq 0$, we have

$$\langle\psi|K_0^{\dagger}K_{i_t}|\psi\rangle-\langle\psi|U^{\dagger}K_{i_t}|\psi\rangle\langle\psi|K_0^{\dagger}U|\psi\rangle$$

$$=\langle\psi|U^{\dagger}K_{i_t}|\psi\rangle-\langle\psi|U^{\dagger}K_{i_t}|\psi\rangle+O(\varepsilon^{3/2})$$

$$=O(\varepsilon^{3/2}).$$

Now, we can neglect the $O(\varepsilon^{3/2})$ term, since multiplying by the remaining ϕ-associated factor $O(\varepsilon^{1/2})$, it will be $O(\varepsilon^2)$. A similar cancellation occurs for the terms where $i_q \neq 0$ and $i_t = 0$. We are left with

$$P_S^{\log T} - \langle U\psi|\tilde{\rho}^{(\log T)}|U\psi\rangle = \frac{1}{T}(1-F)_0 + O(\varepsilon^2)$$

Finally, since $(1-F)_0 \sim O(\varepsilon)$ and $P_S^{\log T} \sim 1 - O(\varepsilon)$,

$$
\begin{aligned}
(1-F)_{\log T} &= \frac{P_S^{\log T} - \langle U\psi|\tilde{\rho}^{(\log T)}|U\psi\rangle}{P_S^{\log T}} \\
&= \frac{\frac{1}{T}(1-F)_0}{1-O(\varepsilon)} + O(\varepsilon^2) \\
&= \frac{1}{T}(1-F)_0 + O(\varepsilon^2),
\end{aligned}
\tag{S12}
$$

which gives Eqn. [3] of the main text.

C.   Unitary errors

This calculation comes with one significant caveat. The way we have written $K_0 = U - \varepsilon\xi$ looks extremely general, and one is tempted to conclude that this error suppression works for any such Kraus channel. However, we note that the behaviour for unitary error channels, where $K_0 = V$ is a pure unitary is exceptional and must be considered separately. To see why that is the case, suppose $V = U - \varepsilon\xi$ is unitary. Then

$$
\begin{aligned}
VV^\dagger &= (U - \varepsilon\xi)(U^\dagger - \varepsilon\xi^\dagger) \\
&= 1 - \varepsilon(U\xi^\dagger + \xi U^\dagger) + O(\varepsilon^2).
\end{aligned}
$$

In other words, for V to be unitary to first order we require $U\xi^\dagger + \xi U^\dagger = 0$. Now, the base fidelity of such a channel is,

$$
\begin{aligned}
(1-F)_0 &= 1 - \langle\psi|U^\dagger V|\psi\rangle\langle\psi|V^\dagger U|\psi\rangle \\
&= \varepsilon(U\xi^\dagger + \xi U^\dagger) + O(\varepsilon^2) \\
&= O(\varepsilon^2).
\end{aligned}
$$

Since $(1-F)_0 \sim O(\varepsilon^2)$, the calculation of the previous section tells us nothing about the error suppression of such unitary error channels.

In fact, coherent unitary errors cannot be suppressed by gate-based EF. Suppose $\mathcal{U}(\cdot) = V(\cdot)V^\dagger$. It is not difficult to see that the outcome of gate-based EF for any number of ancillas is always $\rho_{\log T} = V|\psi\rangle\langle\psi|V^\dagger$, hence $F_{\log T} = F_0$. We conclude that gate-based EF works to suppress stochastic errors only, leaving unitary errors alone. Fortunately, this still encompasses the vast majority of error channels commonly considered.

III.   GUARANTEEING FIDELITY ENHANCEMENT WITH A SINGLE CONTROL QUBIT

We claim that error filtration can enhance fidelity with a single qubit as long as the base fidelity F_0 of the apparatus for the input state $|\psi\rangle$ satisfies $F_1 > \frac{1}{2}$. This can be done by setting $|\phi\rangle = |\psi\rangle$. The final state ρ_1 with a single control qubit is

$$\rho_1 = \frac{1}{2}\rho_0 + \frac{1}{2}\rho_0^2$$

where $\rho_0 = \mathcal{U}(|\psi\rangle\langle\psi|)$.

The expression for the fidelity is

$$F_1 = \frac{\langle U\psi|\rho_1|U\psi\rangle}{\mathrm{Tr}\rho_1} = \frac{F_0 + \langle U\psi|\rho_0^2|U\psi\rangle}{1 + \mathrm{Tr}\rho_0^2}. \tag{S13}$$

The condition $F_1 > F_0$ then reduces to

$$\langle U\psi|\rho_0^2|U\psi\rangle - F_0\mathrm{Tr}\rho_0^2 \tag{S14}$$

To evaluate this, let us write

$$\rho_0 = F_0|U\psi\rangle\langle U\psi| + (1 - F_0)\sigma, \tag{S15}$$

where σ is a density matrix satisfying $\langle U\psi|\sigma|U\psi\rangle = 0$. The relevant quantities become

$$\langle U\psi|\rho_0^2|U\psi\rangle = F_0^2 + (1 - F_0)^2\langle U\psi|\sigma^2|U\psi\rangle \tag{S16}$$
$$\mathrm{Tr}\rho_0^2 = F_0 + (1 - F_0)^2\mathrm{Tr}\sigma^2 \tag{S17}$$

Then,

$$\begin{aligned}
\langle U\psi|\rho_0^2|U\psi\rangle - F_0\mathrm{Tr}\rho_0^2 &= (1 - F_0)\left[F_0^2 + (1 - F_0)^2\left(\langle U\psi|\sigma^2|U\psi\rangle - F_0\mathrm{Tr}\sigma^2\right)\right] \\
&\geq (1 - F_0)F_0(2F_0 - 1) > 0,
\end{aligned} \tag{S18}$$

where we have used $\langle U\psi|\sigma^2|U\psi\rangle \geq 0, \mathrm{Tr}\sigma^2 \leq 1$. The above condition is satisfied whenever $F_0 > 1/2$, with $F_0 = F_1$ for a coherent channel. Hence, a channel improvement is guaranteed as long as the initial fidelity is greater than $1/2$, representing a minimal condition for which gate-based EF yields some improvement.

IV. ERROR FLOOR FOR BOUNDED FAILURE PROBABILITY

As noted above the expressions do not lend themselves to easy simplification of an arbitrary $|\phi\rangle$, and we are unable to make strong statements about the behaviour of the gate-based EF circuit when $T \gg 1/\varepsilon$ in the general case. However, in the case of channels which admit a $|\phi\rangle$ for which we can write $K_i|\phi\rangle \propto |\phi\rangle$, not only is the probability of failure bounded (following conditions 2. and 3. from the previous section), one can show that the state approaches a well-defined limit as $T \to \infty$. We will show this implies that the $O(\varepsilon^2)$ terms do not blow up as $T \to \infty$, and represent an error floor to the scheme.

A. Relation to vacuum extension

The assumption we make for this section is closely related to the concept of a *vacuum extension*, introduced and defined Sec. (2c) of [1]. The vacuum extension provides a consistent way to define a superposition of channels. As [1] points out, the original error filtration scheme [2] is such an example of a superposition of channels. Hence, one might take the point of view of error filtration as the operationalization of the superposition of channels to suppress errors. In gate-based EF, as demonstrated in our main results, we have loosened this demand, and do not require our scheme to strictly carry out a superposition of channels in order to succeed.

We note that any vacuum extension will yield the favourable probability scaling in gate-based EF. Furthermore, there are various physical systems in which a vacuum extension arises naturally, most notably in optical systems, where the vacuum extension of a system tends to be the actual electromagnetic vacuum. For instance, one might encode qubits with polarized light $|0(1)\rangle = |H(V)\rangle$. Since most apparatuses act trivially on the electromagnetic vacuum, i.e. no input state, one can then define the vacuum extension to a channel using the actual electromagnetic vacuum.

For our purposes of the calculations in this section, we will not make full use of a rigorous definition of the vacuum extension. Instead, it will be sufficient for us to define a state $|\phi\rangle$ such that,

$$K_i|\phi\rangle = \sqrt{q_\phi^{(i)}}|\phi\rangle, \tag{S19}$$

where $q_\phi^{(i)}$ can be interpreted as the probability of the Kraus operator K_i occurring given $|\phi\rangle$ as an input state to the apparatus. For convenience, We will refer to any state $|\phi\rangle$ with these properties relative to our channel of interest as a *pseudo-vacuum state*.

B. Mixed unitary error channels

1. Two unitaries

We first consider the simplest possible model where $K_0 = \sqrt{1-p}U$ and $K_1 = \sqrt{p}V$, where U, V are unitary operators. We assume the existence of a psuedo-vacuum state $|\phi\rangle$ relative to this channel,

$$K_0|\phi\rangle = \sqrt{1-p}|\phi\rangle, \qquad K_1|\phi\rangle = \sqrt{p}|\phi\rangle. \tag{S20}$$

Now suppose we are given an input state $|\psi\rangle$. To be as restrictive as possible, we can pick the minimum fidelity state $|\psi\rangle$ in the input space, but this will not matter for our calculation. We can express the action of V on $|\psi\rangle$ as:

$$V|\psi\rangle = e^{i\nu}\cos\theta|U\psi\rangle + \sin\theta|U\psi^{\perp}\rangle. \tag{S21}$$

with ν chosen so that $\theta \in [0, \pi/2]$. Note that we have to keep this phase around because it affects the outcome of gate-based EF when $T > 1$. $|U\psi^{\perp}\rangle$ is simply the part of the state that is orthogonal to $|U\psi\rangle$.

The fidelity for this state is

$$F_0 = 1 - p + p\cos^2\theta = 1 - p\sin^2\theta. \tag{S22}$$

Now we want to evaluate F_∞, which is the fidelity approached by gate-based EF as $T \to \infty$. As $T \to \infty$, the output density matrix turns out to approach a pure state:

$$|\psi_\infty\rangle \propto (1-p)U|\psi\rangle + pV|\psi\rangle \tag{S23}$$

$$= (1 - p + pe^{i\nu}\cos\theta)|U\psi\rangle + p\sin\theta|U\psi^{\perp}\rangle. \tag{S24}$$

To see why this is the case, consider the density matrix for arbitrary T. Consider a single error configuration, where for the some set of branches I_0 where $|I_0| = m$, the black box carries out K_0 and in the rest of the $T - m$ branches it carries out K_1. Prior to measuring the control register, this corresponds to the state

$$\frac{1}{\sqrt{T}}\sum_{t \in I_0}|t\rangle K_0|\psi\rangle K_0^{m-1}K_1^{T-m}|\phi\rangle + \frac{1}{\sqrt{T}}\sum_{t \notin I_0}|t\rangle K_1|\psi\rangle K_0^m K_1^{T-m-1}|\phi\rangle \tag{S25}$$

$$= \frac{1}{\sqrt{T}}(1-p)^{m/2}p^{(T-m)/2}\left(\sum_{t \in I_0}|i\rangle U|\psi\rangle + \sum_{t \notin I_0}|i\rangle V|\psi\rangle\right)|\phi\rangle. \tag{S26}$$

Tracing out the vacuum register and measuring the control register, we are left with

$$\frac{1}{\sqrt{2T}}(1-p)^{m/2}p^{(T-m)/2}\left(mU|\psi\rangle + (T-m)V|\psi\rangle\right) \tag{S27}$$

We see that since the action on the pseudo-vacuum state is trivial, this state only depends on the number of times K_0, K_1 are applied in this particular error configuration. Thus all error configurations with the same $m, T - m$ distribution of Kraus operators coherently combine in the final density matrix, giving a T choose m enhancement to this state, such that the final (unnormalized) density matrix is

$$\rho_T = \sum_{m=1}^{T}\frac{1}{2T}\binom{T}{m}(1-p)^m p^{T-m}\left(mU|\psi\rangle + (T-m)V|\psi\rangle\right)\left(mU|\psi\rangle + (T-m)V|\psi\rangle\right)^{\dagger} \tag{S28}$$

Now, we know from the binomial distribution that as $T \to \infty$ that when viewed as a function of m, $\binom{T}{m}(1-p)^m p^{T-m}$ is sharply peaked around $m = (1-p)T$. Thus for $T \to \infty$,

$$\rho_\infty \simeq \left((1-p)U|\psi\rangle + pV|\psi\rangle\right)\left((1-p)U|\psi\rangle + pV|\psi\rangle\right)^{\dagger} \tag{S29}$$

which is the pure state $|\psi_\infty\rangle$. This provides a rigorous sense in which we are converting a mixture of states $U|\psi\rangle, V|\psi\rangle$ into a superposition. The trace and fidelity of this state are

$$\mathrm{Tr}\rho_\infty = (1 - p + pe^{i\nu}\cos\theta)^2 + p^2\sin^2\theta, \tag{S30}$$

$$\langle U\psi|\rho_\infty|U\psi\rangle = (1 - p + pe^{i\nu}\cos\theta)^2, \tag{S31}$$

$$F_\infty = \frac{(1 - p + pe^{i\nu}\cos\theta)^2}{(1 - p + pe^{i\nu}\cos\theta)^2 + p^2\sin^2\theta}. \tag{S32}$$

We are interested in two questions. First, when does this represent an improvement on the base fidelity – in other words, when is $F_0 < F_\infty$? Second, in the small parameter case, what does this mean for the $O(\varepsilon^2)$ terms?

First, if ν is known, then we can simply write down $F_0 < F_\infty$ in terms of p, ν as

$$1 - p + p\cos^2\theta < \frac{(1 - p + pe^{i\nu}\cos\theta)^2}{(1 - p + pe^{i\nu}\cos\theta)^2 + p^2\sin^2\theta}. \tag{S33}$$

Note that expression, unlike the single qubit control case, does not furnish a condition on F_0, since F_0 contains no information about ν. It will be useful for us to pare down this requirement even more to a condition stated only in terms of p, since $p, 1 - p$ are easily determined.

It is instructive to consider the two extremal values $\nu = 0, \pi$. $\nu = 0$ corresponds to constructive interference in the final state and $\nu = \pi$ corresponds to destructive inteference. The $\nu = 0$ case showcases how phase coherence can amplify the effects of error filtration. However, we will make the pessimistic assumption that $\nu = \pi$ in order to obtain a minimal and single-variable success condition, since

$$F(|\psi_\infty\rangle) > F((1 - p - p\cos\theta)|U\psi\rangle + p\sin\theta|U\psi^\perp\rangle) \equiv F'_\infty, \tag{S34}$$

where $F(\cdot)$ refers to the fidelity of the state $(\cdot)$. With this, the pessimistic success condition is

$$1 - p + p\cos^2\theta = F_1 < F'_\infty = \frac{(1 - p - p\cos\theta)^2}{(1 - p - p\cos\theta)^2 + p^2\sin^2\theta}. \tag{S35}$$

To simplify, we write $\cos\theta$ in terms of F_1 and solve for p. Subject to the constraints $0 \leq p \leq 1, 1 - p \leq F_1 < 1$, this yields

$$p < \frac{1}{4}, \quad \text{or} \tag{S36}$$

$$\frac{1}{4} < p < \frac{1}{2} \text{ and } 1 - p < F < \frac{1}{4p}. \tag{S37}$$

The first condition is easy to interpret. The inequality $p < \frac{1}{4}$ asserts that the probability of V is sufficiently small that the final state resembles $|U\psi\rangle$ despite possible destructive interference due to ν. The second condition provides for the possibility that the interference can be quite possibly be quite large, given both a destructive phase and a larger probability of V. However, the upper bound on F in that case serves to bound the size of $\cos\theta$ and thus the size of the destructive interference.

In terms of F_1, this establishes two regions of success, although F_1 alone cannot yield a sufficient condition over this entire range. As we will see shortly, the first condition generalizes more readily to more complicated channels.

Finally, let us consider where the small parameter regime fits in this picture. Let $2p = \varepsilon \ll 1$ so that K_0 is of the form $U - \varepsilon A + O(\varepsilon^2)$. Then we can expand F_∞ as

$$F_\infty = \frac{1 - \varepsilon(1 + e^{i\nu}\cos\theta) + \frac{1}{4}\varepsilon^2(1 + e^{i\nu}\cos\theta)^2}{1 - \varepsilon(1 + e^{i\nu}\cos\theta) + \frac{1}{4}\varepsilon^2[(1 + e^{i\nu}\cos\theta)^2 + \sin^2\theta]}$$
$$= 1 - \frac{1}{4}\sin^2\theta\varepsilon^2 + O(\varepsilon^2). \tag{S38}$$

This reaffirms the small parameter scaling we derived earlier, as $F_{T\to\infty} \to 1 - O(\varepsilon^2)$. Furthermore, the coefficient of ε^2 as $T \to \infty$ approaches a T independent limit and does not blow up. Hence, we can interpret it as an error floor to gate-based EF under the assumptions in this section.

2. Many unitaries, with K_0 perfect

Now we consider the case of a mixed unitary error channel with $K_0 \propto U$ and multiple unitary errors possible, so that

$$K_0 = \sqrt{1 - p}U, \qquad K_{i\neq 0} = \sqrt{p\lambda_i}V_i, \tag{S39}$$

where $\sum_i \lambda_i = 1$. We write the action of V on $|\psi\rangle$ as

$$V_i|\psi\rangle = e^{i\nu_i}\cos\theta_i|U\psi\rangle + \sin\theta_i|U\psi_i^\perp\rangle. \tag{S40}$$

Again, we choose ν such that $\theta \in [0, \pi/2]$. The base fidelity is

$$F_0 = 1 - p + p \sum_i \lambda_i \cos^2 \theta_i. \tag{S41}$$

Following the same argument as the two unitary case, we find

$$|\psi_\infty\rangle = (1-p)U|\psi\rangle + p \sum_i \lambda_i V_i |\psi\rangle = (1 - p + p \sum_i \lambda_i e^{i\nu_i} \cos \theta_i)|U\psi\rangle + p \sum_i \lambda_i \sin \theta_i |U\psi_i^\perp\rangle \tag{S42}$$

To derive a minimal success condition in the same spirit as before, we note that the fidelity is minimized when $\nu_i = -\pi$ for all i. This corresponds to constructive interference between all the $|U\psi_i^\perp\rangle$, which is greatest when they are all parallel, i.e. $|U\psi_i^\perp\rangle = |U\psi^\perp\rangle$ for all i. Let

$$|\psi'_\infty\rangle = (1 - p - p \sum_i \lambda_i \cos \theta_i)|U\psi\rangle + p \sum_i \lambda_i \sin \theta_i |U\psi^\perp\rangle. \tag{S43}$$

Then

$$F(|\psi_\infty\rangle) > F(|\psi'_\infty\rangle) \equiv F'_\infty, \tag{S44}$$

and it will be sufficient for us to say $F_1 < F'_\infty$. We calculate F'_∞:

$$F'_\infty = \frac{(1 - p - p \sum_i \lambda_i \cos \theta_i)^2}{(1 - p - p \sum_i \lambda_i \cos \theta_i)^2 + p^2 \left(\sum_i \lambda_i \sin \theta_i \right)^2}. \tag{S45}$$

The $F_0 < F_\infty$ condition now reads:

$$1 - p + p \sum_i \lambda_i \cos^2 \theta_i < \frac{(1 - p - p \sum_i \lambda_i \cos \theta_i)^2}{(1 - p - p \sum_i \lambda_i \cos \theta_i)^2 + p^2 \left(\sum_i \lambda_i \sin \theta_i \right)^2}. \tag{S46}$$

So far the above still contains p, θ_i, λ_i. We'd like to reduce it to just p. To do so, we will wrestle it into the form we've already solved for the two unitary channel. To do so, we note the trivial bound

$$F_1 = 1 - p + p \sum_i \lambda_i \cos^2 \theta_i < F'_1 = 1 - p + p \cos^2 \theta_{max} \tag{S47}$$

where $\cos \theta_{max}$ is the maximum possible value $\cos \theta_i$ can take. F_∞ is similarly lower bounded by

$$F((1 - p - p \sum_i \lambda_i \cos \theta_i)|U\psi\rangle + p \sum_i \lambda_i \sin \theta_i |U\psi^\perp\rangle) > F((1 - p - p \cos \theta_{max})|U\psi\rangle + p \sin \theta_{max}|U\psi^\perp\rangle) \tag{S48}$$

where $\sin \theta_{max}$ is the minimum possible value of $\sin \theta_i$. Then our condition will be fulfilled if we have

$$1 - p + p \cos^2 \theta_{max} < \frac{(1 - p - p \cos \theta_{max})}{(1 - p - p \cos \theta_{max})^2 + p^2 \sin^2 \theta_{max}} \tag{S49}$$

which reduces to the two unitary case. Hence, applying the results of the previous section,

$$p < \frac{1}{4} \tag{S50}$$

is a minimal single parameter success condition.

3. Many unitaries, with K_0 imperfect

Our next generalization is to allow K_0's action on $|\psi\rangle$ to have a perpendicular component $|U\psi^\perp\rangle$, such that

$$K_0|\psi\rangle = \sqrt{1-p} \cos \theta_0 |U\psi\rangle + \sqrt{1-p} \sin \theta_0 |U\psi^\perp\rangle. \tag{S51}$$

We assume that $\cos\theta_0 > 1/2$, otherwise one can hardly call the apparatus something that attempts to carry out the operation U. In particular, $\cos\theta_0 < 1/2$ would imply $F_1 < 1/2$, a situation which we will try to avoid entirely.

The action of $K_{i\neq0}$ on $|\psi\rangle$ is defined as before. Again, we will deal with the two operator case first, for which

$$F_0 = (1-p)\cos^2\theta_0 + p\cos^2\theta_1. \tag{S52}$$

To derive a condition on $\mathcal{U}$, we will bound the limiting fidelity $F_\infty^{(K)}$ associated with $\mathcal{U}$ with the limiting fidelity $F_\infty^{(L)}$ associated with a different channel $\mathcal{L}$. Consider such a Kraus channel $\mathcal{L}$ with three operators L_0, L_0', L_1, whose action on $|\psi\rangle$ are defined by

$$L_0|\psi\rangle = \sqrt{(1-p)\cos\theta_0}|U\psi\rangle \tag{S53}$$

$$L_0'|\psi\rangle = \sqrt{(1-p)\sin\theta_0}|U\psi_0^\perp\rangle \tag{S54}$$

$$L_1|\psi\rangle = K_1|\psi\rangle \tag{S55}$$

where L_0, L_0' are proportionate to unitaries, so that their action on the pseudo-vacuum extension is, by assumption,

$$L_0|\phi\rangle = \sqrt{(1-p)\cos\theta_0}, \tag{S56}$$

$$L_0'|\phi\rangle = \sqrt{(1-p)\sin\theta_0}, \tag{S57}$$

$$L_1|\phi\rangle = K_1|\phi\rangle. \tag{S58}$$

First let's check if this channel is physical. We have

$$\mathrm{Tr}\,(L_0|\psi\rangle\langle\psi|L_0) = (1-p)\cos^2\theta_0 \tag{S59}$$

$$\mathrm{Tr}\,(L_0'|\psi\rangle\langle\psi|L_0') = (1-p)\sin^2\theta_0 \tag{S60}$$

$$\mathrm{Tr}\,(L_1|\psi\rangle\langle\psi|L_1) = p \tag{S61}$$

which add up to 1. Hence, by defining the action of L_i on the rest of the Hilbert space appropriately, we can construct a valid Kraus channel with these properties.

The fidelity of $\mathcal{L}$ on $|\psi\rangle$ is

$$F_0^{(L)} = (1-p)\cos^2\theta_0 + p\cos^2\theta_1 = F_1 \tag{S62}$$

The $T \to \infty$ limits for $\mathcal{U}, \mathcal{L}$ before measurement of the control register are given by the pure unnormalized states,

$$|\psi_\infty^{(K)}\rangle = \left[(1-p)\cos\theta_0 + pe^{i\nu_1}\cos\theta_1\right]|U\psi\rangle + (1-p)\sin\theta_0|U\psi_0^\perp\rangle + p\sin\theta_1|U\psi_1^\perp\rangle$$
$$|\psi_\infty^{(L)}\rangle = \left[(1-p)\cos^2\theta_0 + pe^{i\nu_1}\cos\theta_1\right]|U\psi\rangle + (1-p)\sin^2\theta_0|U\psi_0^\perp\rangle + p\sin\theta_1|U\psi_1^\perp\rangle \tag{S63}$$

We observe that if $-\pi/2 < \nu < \pi/2$ implies that

$$F(|\psi_\infty^{(K)}\rangle) > F(|\psi_\infty^{(L)}\rangle) \tag{S64}$$

since $\cos\theta_0 - \cos^2\theta_0 > \sin\theta_0 - \sin^2\theta_0 > 0$, so $|\psi_\infty^{(L)}\rangle$ has a relatively larger component in $|U\psi\rangle$. If $\pi/2 < \nu < 3\pi/2$, then we just have to ensure $(1-p)\cos\theta_0 > p\cos\theta_1$ for the same thing to be true, which can be ensured by $1-p > p$.

Thus we only have to show that the condition holds for the channel specified by Kraus operators L_i. We have done this already, and re-using the minimal condition we found earlier with $(1-p) \mapsto (1-p)\cos^2\theta_0$, we have the condition

$$(1-p)\cos^2\theta_0 > \frac{3}{4}, \tag{S65}$$

which straightforwardly generalizes to as Kraus operators $K_{i\neq0}$ as we want.

C. Channels with non-unitary Kraus operators

Finally, we deal with the most general case, where K_i can be non-unitary and K_0 can be imperfect. In that case the probability of K_i is state dependent. The key insight here is to realize that as $T \to \infty$, the Kraus operators have well-defined probability given by

$$q_\phi^{(i)} \equiv \langle\phi|K_i^\dagger K_i|\phi\rangle, \tag{S66}$$

because at each time step, we feed the apparatus $T-1$ copies of $|\phi\rangle$ and only a single copy of $|\psi\rangle$. Note that K_i is effectively unitary on $|\phi\rangle$.

To generalize our previous results, we simply have to replace $1-p$ with $q_\phi^{(0)}$ and $p\lambda_i$ with $q_\phi^{(i)}$ in the expression for F_∞. If $q_\phi^{(0)} = 1-p$ as with the unitary case, then $q_\phi^{(0)} > 3/4$ will suffice. Since $K_0|\psi\rangle$ has the largest component of $|U\psi\rangle$, if $q_\phi^{(0)} > 1-p$ can only increase F_∞. Thus our condition becomes

$$q_\phi^{(0)} > q_\psi^{(0)} > \frac{3}{4\cos^2\theta_0}. \tag{S67}$$

V. ISOMORPHISM WITH ERROR FILTRATION FOR QUANTUM COMMUNICATION

In the event we have a psuedo-vacuum state $|\phi\rangle$ available to us (see Sec. IV A), our scheme is explicitly isomorphic to the original error filtration proposal in the ideal case [2, 3]. We will characterize this in the $T=2$ case, with the more general T following straight-forwardly. For simplicity, suppose our ideal unitary is the identity, as is the case for a quantum communication task. Let the state we want to transmit be

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle. \tag{S68}$$

The effect of multiplexing in error filtration is to encode this in a W-state [3]. For $T=2$, this becomes

$$\alpha\hat{a}_0^\dagger + \beta\hat{a}_1^\dagger \mapsto \frac{1}{\sqrt{2}}\alpha\left(\hat{a}_0^\dagger + \hat{b}_0^\dagger\right) + \frac{1}{\sqrt{2}}\beta\left(\hat{a}_1^\dagger + \hat{b}_1^\dagger\right), \tag{S69}$$

where $\hat{a}_i, \hat{b}_i$ refer to some kind of encoding in a photonic mode.

Explicitly writing this out in the fock basis, we have

$$\frac{1}{\sqrt{2}}\alpha|1,0,0,0\rangle + \frac{1}{\sqrt{2}}\beta|0,1,0,0\rangle + \frac{1}{\sqrt{2}}\alpha|0,0,1,0\rangle + \frac{1}{\sqrt{2}}\beta|0,0,0,1\rangle \tag{S70}$$

In gate-based EF, we pair the state $|\psi\rangle$ with $|+\rangle$ and attach $|\phi\rangle$ states, which looks like

$$\frac{1}{\sqrt{2}}|0\rangle\left(\alpha|0\rangle + \beta|1\rangle\right)|\phi\rangle + \frac{1}{\sqrt{2}}|1\rangle\left(\alpha|0\rangle + \beta|1\rangle\right)|\phi\rangle \tag{S71}$$

$$\mapsto \frac{1}{\sqrt{2}}\alpha|00\rangle|\phi\rangle + \frac{1}{\sqrt{2}}\alpha|10\rangle|\phi\rangle + \frac{1}{\sqrt{2}}\beta|01\rangle|\phi\rangle + \frac{1}{\sqrt{2}}\beta|11\rangle|\phi\rangle \tag{S72}$$

Written in this way one can clearly see the one-to-one correspondence of kets in either scheme. Due to the principle of superposition, we can deal with each ket individually. If the error operators only act trivially on the state $|\phi\rangle$ (in gate-based EF) and the vacuum (in EF for quantum communication), then we can carefully map the effective error on each ket in gate-based EF to an error on each ket in error filtration.

More formally it was observed in Eqn. (3.1) of [1], there exists a unitary operation that connects a spatially separated quantum superposition of paths into a quantum superposition of paths over an additional register:

$$U(|0\rangle \otimes |\psi\rangle) = |\psi\rangle \otimes |\Omega\rangle \tag{S73}$$
$$U(|1\rangle \otimes |\psi\rangle) = |\Omega\rangle \otimes |\psi\rangle. \tag{S74}$$

By appending the state $|\phi\rangle$, one can embed this in an isomorphism U' that maps gate-based EF states to error filtration states.

VI. ANCILLA ERRORS

A. Effect of ancilla errors

Finally, we work out the effect of ancilla errors. We will consider X and Z type errors, or bit flips and phase flips, on the control register and their effects on the fidelity and success probability of gate-based EF.
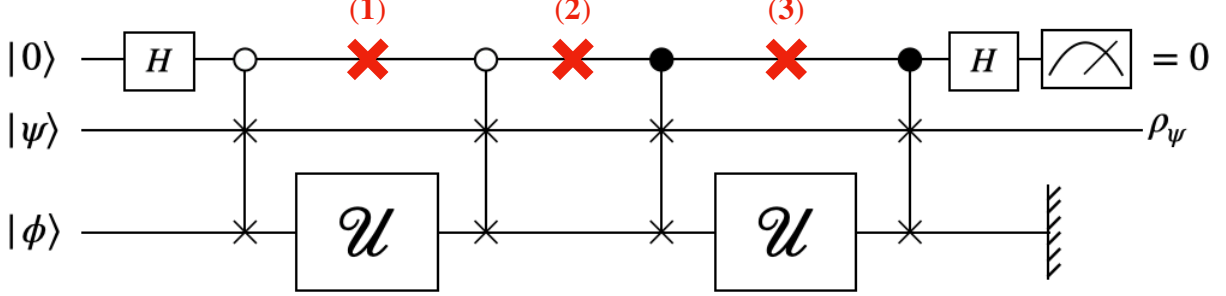
FIG. S2: Gate-based EF with $T = 2$, with potential control fault locations highlighted in red.

1. Bit flip errors

A bit flip error on the ancilla can potentially ruin the whole query. To see why this happens, we consider the $T = 2$ case, for which the circuit is depicted in Fig. S2. We consider the 3 potential fault locations highlighted in red. For simplicity, suppose the bit flip errors occur in a run where the black box happens to apply the perfect unitary U each time. We will show that a bit flip in any of these indicated fault locations in general leads to the failure of the query.

1. A bit flip in location (1) leads to the state (before measurement and tracing out the active register)

$$|+\rangle \otimes U|\phi\rangle \otimes U|\psi\rangle \tag{S75}$$

such that the final state is $U|\phi\rangle$, which in general is unrelated to the state we want. Since the control register is in the $|+\rangle$ state, an error in this location is maximally bad, as the final measurement will post-select $U|\phi\rangle$ with unit probability.

2. A bit flip in location (2) leads to a double query of each of the participating states:

$$\frac{1}{\sqrt{2}}|0\rangle|\psi\rangle U^2|\phi\rangle + \frac{1}{\sqrt{2}}|1\rangle U^2|\psi\rangle|\phi\rangle, \tag{S76}$$

which does not contain our desired state $U|\psi\rangle$ at all.

3. A bit flip in location (3) leads to the same final state as (1).

To estimate the effect of such errors on the final scheme, suppose the black box takes a time τ_U to run, and the error rate of each control ancilla is ε_{bf}. We assume that τ_U is much longer than the idle time of the control ancilla between the controlled-SWAPs on different branches, as well before and after the first and final queries.

For larger values of T, we can assume any such error of the type described by fault locations $(1), (3)$ have a 0 contribution to the fidelity, and do not affect the post-selection probability. To first order in ε_C we expect the final infidelity to account for this as

$$(1 - F)_{\log T} \simeq \frac{1}{T}(1 - F)_0 + \tau_U \varepsilon_{bf} T \log T. \tag{S77}$$

Due to the disastrous potential of bit flip errors, we emphasize that gate-based EF should only be applied when the error hierarchy in the main text is respected with regard to bit flip errors on the control register. The above expression allows us to concretely estimate the error hierarchy required to handle bit flip errors. For gate-based EF to yield an overall error suppression, we require

$$\frac{1}{T}(1-F)_0 + \tau_U \varepsilon_{bf} T \log T < (1-F)_0 \tag{S78}$$

We can expressing the error hierarchy as a ratio between $\tau_U \varepsilon_{bf}$ and $(1-F)_0$, we have

$$\frac{\tau_U \varepsilon_{bf}}{(1-F)_0} < \frac{T-1}{T^2 \log T} \simeq \frac{1}{T \log T}. \tag{S79}$$

This tells us that minimally, for $T = 2$, we need the error rate on ancilla qubits to be less than $1/2$ the base infidelity of the apparatus. Furthermore, this indicates that for a set-up with ancilla errors, there is an optimal T beyond which gate-based EF no longer yields any advantage. This optimal T is depicted for the case of QRAM in [FIG].

2. Phase flip errors

Since phase flip errors on the control register commute with any controlled-SWAP operation, a phase flip error on fault location $(1), (2)$ or (3) will have the same effect. Let ε_{pf} be the phase flip error rate on each control qubit. The presence of a phase flip error anywhere can be commuted to the end of the circuit, where the Hadamard transforms it into a bit flip error, upon which such an error causes the output of the circuit to be rejected.

This means that to first order in $\varepsilon, \varepsilon_{pf}$, phase flips *do not affect the fidelity at all*. Since phase flip errors are detected, we expect there to be a decrease in success probability, to first order given by

$$\Delta P_{pf} \simeq \tau_U \varepsilon_{pf} T \log T \tag{S80}$$

In conclusion, the effect of phase flip and bit flip errors are well-characterized. Knowing the quantities $\tau_U, \varepsilon_{bf}, \varepsilon_{pf}, F_0$ then allows one to see whether the error hierarchy is respected and make a decision as to whether to carry out gate-based EF, as well as to what extent T.

3. Memory errors

A final category of ancilla error to consider are memory errors. These are most likely to happen to the memory register while the active register is querying the apparatus. Any error during this time is deleterious to the final query, as it will result in an error on $|\psi\rangle$ or $U|\psi\rangle$ in the $T - 1$ inactive branches of the query. Such errors contribute in the same way to the fidelity and error hierarchy as bit flips.

B. Numerical expansion of error suppressed QRAM with ancilla errors

Unfortunately, to perform a full simulation of both the gate-based EF applied to QRAM with errors in both ancilla and QRAM qubits is extremely computationally taxing. Instead, to observe the qualitative performance of gate-based EF under ancilla errors, it will suffice to perform a first order expansion of the infidelity and failure probabilities.

To do so, we assume the infidelity and failure probabilities have the form

$$(1-F)_{\log_2 T} \simeq (1 - N_{loc} \varepsilon')(1-F)^{(0)}_{\log_2 T} + \varepsilon' \sum_{\eta \in \text{err. loc.}} (1-F)^{(1),\eta}_{\log_2 T}, \tag{S81}$$

$$P_{\log_2 T,\text{fail}} \simeq (1 - N_{loc} \varepsilon') P^{(0)}_{\log_2 T,\text{fail}} + \varepsilon' \sum_{\eta \in \text{err. loc.}} P^{(1),\eta}_{\log_2 T,\text{fail}}, \tag{S82}$$

where N_{loc} is the number of ancilla error locations in the circuit, ε' is the probability of error per ancilla error location. The superscript (0) indicates the bare infidelity and probability of failure without ancilla errors, as calculated in the main text. The superscript $(1), \eta$ requires a little more explanation. η indexes ancilla error locations in the circuit, and the superscript $(1), \eta$ refers to the infidelity/probability of failure given an error in location η but nowhere else.
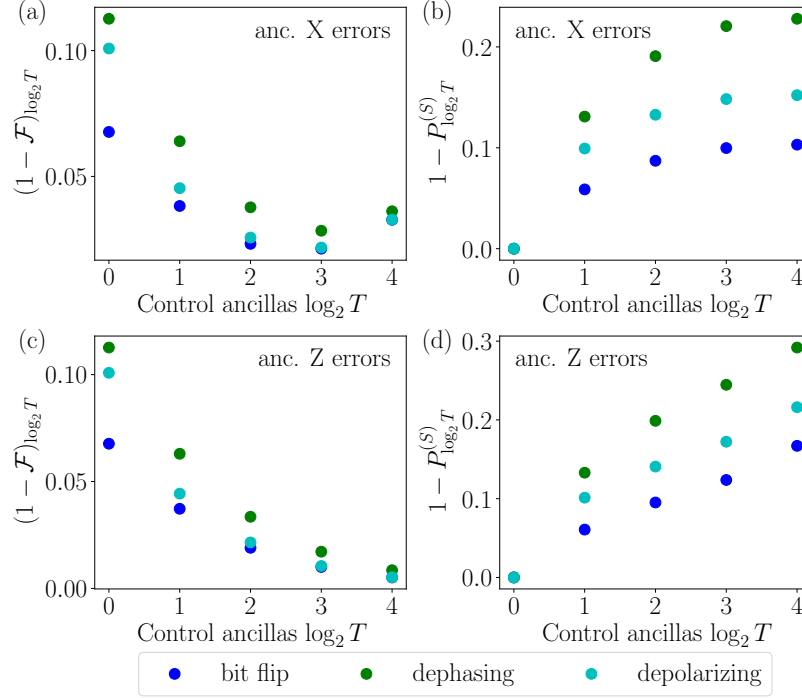
FIG. S3: Gate-based EF applied to QRAM with ancilla errors, for QRAM depth $\log N = 2$. (a, b) Numerically estimated infidelity and probability of failure against number of control ancillas subject to X errors at a rate $\varepsilon' = 0.001$. QRAM errors occur at rate 0.01 per qubit per time step. (c, d) Same as (a, b) but for control ancilla subject to Z errors.

As such, the above expressions account for the first order expansion in ancilla errors, by discarding the possibility of errors occurring in both ancilla and QRAM at the same time. $(1 - F)_{\log_2 T}^{(1),\eta}$, $P_{\text{fail}}^{(1),\eta}$ are then calculate numerically.

In this manner, expanding with X, Z errors in the ancillas yields the plots in Fig. S3. Note the qualitative features of these plots – (1) Ancilla X errors eventually overcome the error suppression from gate-based EF, and we have an optimal T for which the overall infidelity is lowest. (2) Probability of failure still appears to plateau with ancilla X errors. (3) Ancilla Z errors do not to first order affect the infidelity scaling. However, this comes at the cost of increased failure probabilities.

To illustrate the idea that there is an optimal T for a given cSWAP and QRAM error rate, we plot infidelity against both ancilla bit flip rate (since only X errors affect infidelity) and the number of control ancillas in Fig. S4a. In Fig. S4b, we make a qualitative version plot of the "EF advantage", where a data point is depicted as green if adding one control ancilla improves the overall fidelity for a given ancilla bit flip rate. We observe that for each error rate, there is an optimal working point for T, given by the right-most green entry.

Using the perturbative expression, we can gain some insight into two questions. Given ancilla bit flip errors: (1) When does gate-based EF give an advantage with a single control qubit? (2) What is the maximum T for which we still expect gate-based EF to yield better fidelities than the original black box operation?

For (1), we require that $(1 - F)_2 > (1 - F)_0$. In this case, $N_{\text{loc}} \sim O(1)$, and we write this as a constant $C = N_{\text{loc}}$, and $(1 - F)_2^{(0)} \simeq (1 - F)_0/2$. Conservatively, we can bound $(1 - F)_{\log_2 T}^{(1),\eta} < 1$. Substituting these expressions into Eq. (S81), we require

$$(1 - F)_0 > (1 - C\varepsilon') \frac{1}{2}(1 - F)_0 + C\varepsilon'. \tag{S83}$$

Solving for ε', we find that we require an ancilla bitflip rate

$$\varepsilon' < \frac{1}{2C}(1 - F)_0 + O((1 - F)_0^2), \tag{S84}$$

which demonstrates that we only need ε' to be less than the infidelity of the black box up to a constant factor.

For (2), we require that $(1 - F)_{\log_2 T} > (1 - F)_0$. In this case, the only difference from the above is that we have

$N_{\text{loc}} \sim O(T \log_2 T)$. Performing the same calculation, we find that we require

$$\varepsilon' < \frac{1}{O(T \log_2 T)}(1 - F)_0 + O((1 - F)_0)^2 + O\left(\frac{1}{T}\right). \tag{S85}$$

Note that the above gives the condition for which having control ancillas still gives an improvement on the original black box fidelity. However, we only need to satisfy Eq. (S84) for gate-based EF to give some advantage.
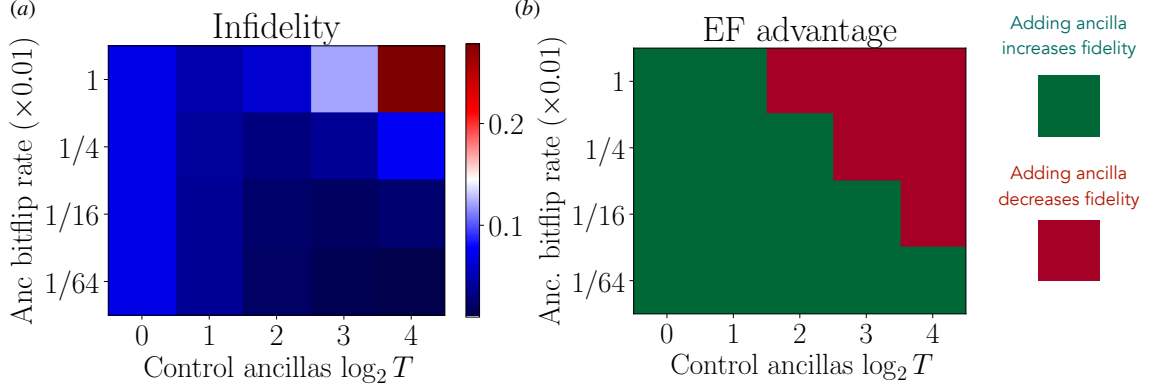

FIG. S4: (a) Infidelity for a depth 2 QRAM with gate-based EF against number of control ancillas (horizontal axis) and ancilla bitflip rate (vertical axis). (b) Plot of "EF advantage" for the same data. The entry is green if adding the ancilla increases the fidelity for the given error rate, and red otherwise.
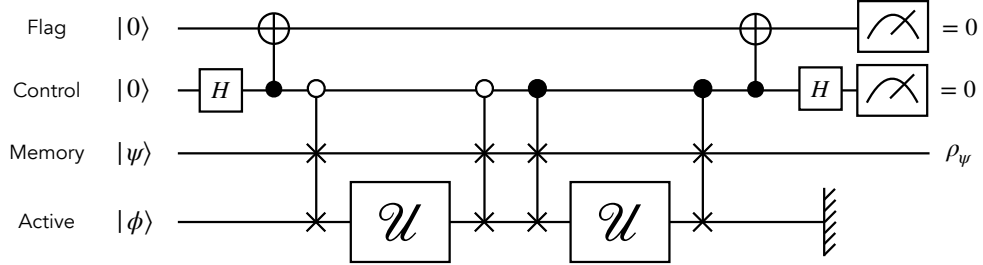
C.   Mitigating bit flip errors with a flag qubit

FIG. S5: Gate-based EF with a single control qubit ($T = 2$) and a flag qubit to detect bit flip errors in the ancilla.

In the preceding discussion, we showed that bit flip errors on the control qubit can be problematic for gate-based EF. Fortunately, the "control" part of the cSWAP acting on the ancilla commutes with the Z operator. Hence, we can use a single flag qubit to detect bit flip errors in the ancilla [4]. In Fig. S5, the first hadamard and cNOT encode the flag and control in the $ZZ = +1$ subspace. The final cNOT allows us to measure the parity of the flag and control qubit. If the parity is odd, the flag measurement will yield $|1\rangle$, and we will know that a bit flip error must have occurred on either the flag or control qubit. In that case, we can discard the outcome and restart the procedure. This reduces the impact of bit flip errors on the fidelity quadratically (since two bit flip errors must occur for the bit flips to be undetectable), at the cost of a decreased probability of success. The flag qubit assisted scheme above generalizes easily to $\log T$ control qubits, by similarly appending a flag to each control qubit. This provides a way to mitigate the effect of bit flip errors on the ancilla with relatively low overhead.

VII.   OPTICAL IMPLEMENTATIONS OF THE CONTROL REGISTER

In this section, we point out several optical schemes for implementing the cSWAP operation crucial for gate-based EF. First, we note that the control register essentially acts as an addressing register in either a QRAM or Quantum Read-Only Memory (QROM) [5], with the memory register acting as the data to be queried, hence proposals (such as provided in the original BB-QRAM proposal [6]) for such quantum memories may be adapted to carry out the cSWAPs required for gate-based EF. Furthermore, several proposals for optical implementation of the cSWAP operations required for QRAM have been put forth since [6], which show that it is possible with multiple different encodings. For instance, [7] proposes an implementation using frequency-encoded optical qubits, and [8] proposes an implementation using fock state encoded qubits.

Finally, for concreteness, we highlight here an additional scheme by which one can carry out an optical cSWAP. The cSWAP operation can be decomposed into a fifty-fifty beam-splitter operation on the active and memory register, followed by a controlled-Z between the control and active register, followed by the inverse beam-splitter operation on the active and memory register (see Fig. 13b of [9]). To achieve such an operation, we use optical polarization qubits for the active and memory registers, where the $|0\rangle, |1\rangle$ logical states are encoded in the horizontal and vertical polarization states $|H\rangle, |V\rangle$. The beam-splitter operations can then be carried out using linear optics. Finally, to get a controlled-Z operation, we can use the first half of the Duan-Kimble scheme [10] by encoding the control qubit in a three-level atom, which will allow us to carry out a controlled-Z between the control qubit and the polarization-encoded memory register.

[1] G. Chiribella and H. Kristjánsson, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **475**, 20180903 (2019), publisher: Royal Society, URL https://royalsocietypublishing.org/doi/10.1098/rspa.2018.0903.

[2] N. Gisin, N. Linden, S. Massar, and S. Popescu, Physical Review A **72**, 012338 (2005), publisher: American Physical Society, URL https://link.aps.org/doi/10.1103/PhysRevA.72.012338.

[3] M. K. Vijayan, A. P. Lund, and P. P. Rohde, Quantum **4**, 303 (2020), ISSN 2521-327X, URL https://doi.org/10.22331/q-2020-08-03-303.

[4] R. Chao and B. W. Reichardt, Phys. Rev. Lett. **121**, 050502 (2018), URL https://link.aps.org/doi/10.1103/PhysRevLett.121.050502.

[5] C. T. Hann, G. Lee, S. Girvin, and L. Jiang, PRX Quantum **2**, 020311 (2021), publisher: American Physical Society, URL https://link.aps.org/doi/10.1103/PRXQuantum.2.020311.

[6] V. Giovannetti, S. Lloyd, and L. Maccone, Physical Review Letters **100**, 160501 (2008), publisher: American Physical Society, URL https://link.aps.org/doi/10.1103/PhysRevLett.100.160501.

[7] K. C. Chen, W. Dai, C. Errando-Herranz, S. Lloyd, and D. Englund, PRX Quantum **2**, 030319 (2021), ISSN 2691-3399, arXiv: 2103.07623, URL http://arxiv.org/abs/2103.07623.

[8] F.-Y. Hong, Y. Xiang, Z.-Y. Zhu, L.-z. Jiang, and L.-n. Wu, Phys. Rev. A **86**, 010306 (2012), URL https://link.aps.org/doi/10.1103/PhysRevA.86.010306.

[9] B. J. Chapman, S. J. de Graaf, S. H. Xue, Y. Zhang, J. Teoh, J. C. Curtis, T. Tsunoda, A. Eickbusch, A. P. Read, A. Koottandavida, et al., PRX Quantum **4**, 020355 (2023), URL https://link.aps.org/doi/10.1103/PRXQuantum.4.020355.

[10] L.-M. Duan and H. J. Kimble, Phys. Rev. Lett. **92**, 127902 (2004), URL https://link.aps.org/doi/10.1103/PhysRevLett.92.127902.