Noise resilience of variational quantum compiling

Sharma, Kunal
Khatri, Sumeet
Cerezo de la Roca, Marco Vinicio Sebastian
Coles, Patrick Joseph

PAPER • OPEN ACCESS

Noise resilience of variational quantum compiling

To cite this article: Kunal Sharma *et al* 2020 *New J. Phys.* **22** 043006

View the article online for updates and enhancements.

New Journal of Physics

The open access journal at the forefront of physics

PAPER

Noise resilience of variational quantum compiling

Kunal Sharma[1,2] , Sumeet Khatri[2], M Cerezo[1,3] and Patrick J Coles[1]

[1] Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, United States of America
[2] Hearne Institute for Theoretical Physics and Department of Physics and Astronomy, Louisiana State University, Baton Rouge, LA United States of America
[3] Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM, United States of America

E-mail: kunals2693@gmail.com

Abstract

Variational hybrid quantum-classical algorithms (VHQCAs) are near-term algorithms that leverage classical optimization to minimize a cost function, which is efficiently evaluated on a quantum computer. Recently VHQCAs have been proposed for quantum compiling, where a target unitary U is compiled into a short-depth gate sequence V. In this work, we report on a surprising form of noise resilience for these algorithms. Namely, we find one often learns the correct gate sequence V (i.e. the correct variational parameters) despite various sources of incoherent noise acting during the cost-evaluation circuit. Our main results are rigorous theorems stating that the optimal variational parameters are unaffected by a broad class of noise models, such as measurement noise, gate noise, and Pauli channel noise. Furthermore, our numerical implementations on IBM's noisy simulator demonstrate resilience when compiling the quantum Fourier transform, Toffoli gate, and W-state preparation. Hence, variational quantum compiling, due to its robustness, could be practically useful for noisy intermediate-scale quantum devices. Finally, we speculate that this noise resilience may be a general phenomenon that applies to other VHQCAs such as the variational quantum eigensolver.

1. Introduction

Obtaining accurate answers from near-term quantum computers is a challenge with major scientific and technological implications. In these so-called noisy intermediate-scale quantum (NISQ) computers [1], errors arise, for example, due to decoherence processes, gate noise, and measurement noise. Clearly, error mitigation techniques will be necessary to make use of NISQ devices. Several promising error mitigation strategies have recently emerged, including zero-noise extrapolation [2], quasi-probability decomposition [2], post-selection [3, 4], noise-aware compiling [5], and machine learning for circuit-depth compression [6]. Let us consider two other strategies for error mitigation in what follows.

Hybridizing a quantum algorithm by pushing some of the complexity onto a classical computer allows one to only run a portion of the computation on the (error-prone) quantum computer. Excellent examples of this strategy are variational hybrid quantum-classical algorithms (VHQCAs) [7]. VHQCAs only employ a quantum computer to evaluate a cost function that depends on the parameters of a quantum gate sequence and then leverage a classical optimization routine to minimize the cost and hence train the parameters. The most famous VHQCA is the variational quantum eigensolver (VQE) [8], where the cost function is the energy for some Hamiltonian and hence the goal is to prepare the ground state. VHQCAs have been proposed for many other applications [9–22].

Another strategy for error mitigation is to find quantum circuits or quantum algorithms that are inherently noise resilient. Circuits for quantum error correction [23, 24], of course, have this property of inherent noise resilience, and in fact, such circuits are resilient to all types of noise on a subset of the qubits. More generally, one could ask whether a circuit is resilient to a particular kind of noise process. Hence, for every circuit, which aims to compute some quantity, one could ask what noise models do not affect the output of the circuit.

The two strategies just mentioned have an interesting intersection: researchers have observed that some VHQCAs have some inherent noise resilience. McClean *et al* [7] noted that coherent errors (e.g., systematic gate biases) can lead to a situation where the formal unitary $V(\alpha)$ specified by the parameters α is different from the actual unitary that is physically implemented $\widetilde{V}(\alpha)$. This error is correctable if there exists a vector β such that one can physically implement the unitary $\widetilde{V}(\alpha + \beta)$ within one's ansatz, with the condition that $\widetilde{V}(\alpha + \beta) = V(\alpha)$. If this condition is satisfied, then one could still physically achieve the minimum value of the cost function, where the minimum value would be associated with different parameters than one would have in the noiseless case. We refer to this kind of noise resilience as *Cost Value Resilience*, since the value of the cost function at the global minimum is unaffected by the noise. Cost Value Resilience is important, e.g. if one is interested in estimating the ground state energy of a Hamiltonian with VQE.

In this work, we report on a different kind of noise resilience for VHQCAs. Instead of considering Cost Value Resilience, we consider the case where the optimal parameters are noise resilient, which we call *Optimal Parameter Resilience*. While Cost Value Resilience is related to coherent noise, we find that Optimal Parameter Resilience holds for certain kinds of incoherent noise, such as decoherence processes and readout errors. For certain applications, obtaining the correct optimal parameters is more important than obtaining the correct value of the cost function.

Quantum compiling [25–27] is one of these applications. Compiling refers to transforming a high-level algorithm into a low-level machine code. For quantum compiling, it is crucial to do this transformation optimally, i.e. to keep the low-level code as short as possible, since errors accumulate with circuit depth. VHQCAs offer a promising framework for (optimal) quantum compiling. Three recent works introduced VHQCAs for quantum compiling, henceforth referred to as variational quantum compiling (VQC) [19–21]. In VQC one trains the parameters α of a short-depth gate sequence $V(\alpha)$ such that it is close to a target unitary U. Here, some distance measure between $V(\alpha)$ and U serves as the cost function and is efficiently evaluated on a quantum computer, while a classical optimizer adjusts the parameters α to minimize the cost. VQC could be an important tool for NISQ computing since it could optimally shrink the depth of quantum circuits. However, a potential issue is that one needs to put the target unitary U on the NISQ device, and hence the target itself is noisy or defective. Furthermore, there are noise sources in other parts of the cost-evaluation circuit. All of these may lead to a defective optimal $V(\alpha)$, with the noise effectively compiled into $V(\alpha)$.

Addressing these concerns, our main results are rigorous theorems stating that many different types of noise during cost evaluation do not affect the optimal $V(\alpha)$. For example, we show that VQC is resilient to measurement noise (readout error). We also show resilience to incoherent gate noise and decoherence processes, such as Pauli channels and non-unital Pauli channels, acting at specific times during the cost-evaluation circuit. In addition to these analytical results, we implement VQC on IBM's noisy quantum simulator [28] (which simulates their quantum hardware) for several quantum gates: quantum Fourier transform, Toffoli, and W-state preparation. In each case, we observed significant noise resilience (even more resilience than what is explained by our theorems) such that we effectively learned the true optimal values of α despite the noise.

Finally, we speculate that the resilience phenomenon that we demonstrate for VQC may be more general, potentially applying to other VHQCAs. For example, we discuss the potential for seeing this resilience for VQE, and as a warm-up for the reader, we give a simple example in the next section where VQE exhibits Optimal Parameter Resilience. We also establish in the Discussion section that VQC is a special case of VQE, and hence our main results can be viewed as being relevant to VQE.

2. Warm-up: simple VQE example

Here we show that VQE [8] exhibits Optimal Parameter Resilience (OPR) to uncorrelated measurement noise for a special class of Hamiltonians. VQE may exhibit OPR more generally, although the proof would certainly be more involved. Hence we consider here this special case for illustration and leave the more general case for future work.

Consider a Hamiltonian that is a sum of local Pauli operators

$$H = -\sum_{j=1}^{n} c^{(j)} \sigma_{w(j)}^{(j)}, \qquad (1)$$

where $\sigma_{w(j)}^{(j)} = U_{w(j)}^{(j)} \sigma_z^{(j)} (U_{w(j)}^{(j)})^\dagger$ is a local operator on qubit j that is unitarily equivalent to the Pauli z operator $\sigma_z^{(j)}$. Physically, this Hamiltonian arises for a system of n non-interacting spin-1/2 particles in a non-uniform (i.e. j-dependent) magnetic field. Without loss of generality, one can take the $c^{(j)}$ coefficients to be non-negative (i.e. absorb any negativity into the definition of the Pauli operator). The ground state $|\psi_0\rangle$ of H has a tensor product form: $|\psi_0\rangle = \bigotimes_{j=1}^{n} |w(j)_+\rangle$, where $|w(j)_+\rangle$ is the eigenvector of $\sigma_{w(j)}^{(j)}$ with the +1 eigenvalue.

Now suppose there is measurement noise in the cost-evaluation circuit. In the ideal case, one measures $\langle H \rangle = \sum_j c^{(j)} \langle \sigma_{w(j)}^{(j)} \rangle = \sum_j c^{(j)} \langle U_{w(j)}^{(j)} \sigma_z^{(j)} (U_{w(j)}^{(j)})^\dagger \rangle$ by applying $(U_{w(j)}^{(j)})^\dagger$ on the jth qubit and measuring it on the standard basis to estimate $\langle \sigma_{w(j)}^{(j)} \rangle$. Then, by performing classical post-processing we compute the weighted sum in $\langle H \rangle$. However, with measurement noise, the $\sigma_z^{(j)}$ operator gets replaced by $\widetilde{\sigma}_z^{(j)} = (p_{00}^{(j)} - p_{10}^{(j)})|0\rangle \langle 0| - (p_{11}^{(j)} - p_{01}^{(j)})|1\rangle \langle 1|$. Here, $p_{kl}^{(j)}$ is the probability to obtain the k outcome when feeding in the $|l\rangle$ state on the jth qubit. Hence, instead of measuring $\langle \sigma_{w(j)}^{(j)} \rangle$, one measures $\langle \widetilde{\sigma}_{w(j)}^{(j)} \rangle$ with $\widetilde{\sigma}_{w(j)}^{(j)} = U_{w(j)}^{(j)} \widetilde{\sigma}_z^{(j)} (U_{w(j)}^{(j)})^\dagger$. In other words, the Hamiltonian H gets replaced by an effective Hamiltonian:

$$\widetilde{H} = -\sum_{j=1}^{n} c^{(j)} \widetilde{\sigma}_{w(j)}^{(j)}. \tag{2}$$

The ground state of $\widetilde{H}$ is a tensor product of one-qubit states that are the eigenvectors of $\widetilde{\sigma}_{w(j)}^{(j)}$ with the largest eigenvalue. Suppose we assume that $p_{00}^{(j)} + p_{11}^{(j)} > p_{01}^{(j)} + p_{10}^{(j)}$ for all j, which means that the probability of getting the correct outcome is greater than the probability for getting the wrong outcome. With this assumption, the largest eigenvalue of $\widetilde{\sigma}_z^{(j)}$ is associated with the $|0\rangle$ state, and hence the largest eigenvalue of $\widetilde{\sigma}_{w(j)}^{(j)}$ is associated with $|w(j)_+\rangle$. Therefore, despite the measurement noise, one still finds that the ground state is $|\psi_0\rangle = \bigotimes_{j=1}^{n} |w(j)_+\rangle$. This implies that one would still learn the correct optimal parameters of the state-preparation circuit if one implemented VQE for this Hamiltonian.

3. Background: variational quantum compiling

Let us now move on to variational quantum compiling (VQC). VQC was first introduced in [19], under the name of quantum-assisted quantum compiling (QAQC). Two later works further investigated VQC [20, 21] with slightly different approaches. Since we are attempting to unite these works [19–21] under one umbrella, we are proposing the name VQC (instead of QAQC) as a unifying term.

There are two overarching approaches to VQC. One is to compile the full unitary matrix U by considering the action of U on all input states (or an informationally complete set of states) [19, 21]. The other is to compile only a particular column of the matrix U by considering the action of U on a fixed input state [19, 20]. The benefit of the first approach is that it is fully general, applying even when one does not know what the input state to U will be (for example, if U occurs in the middle of one's quantum algorithm). The benefit of the second approach is that, when the input state is known, it could lead to a shorter-depth compilation since it does not require compilation of the entire unitary matrix.

3.1. Full unitary matrix compiling
Full unitary matrix compiling (FUMC) was treated in detail in [19]. This work introduced cost functions based on the entanglement fidelity and proposed quantum circuits to quantify the cost based on the overlap between maximally entangled states. A slightly different but equivalent approach was employed in [21]. We focus on the approach of [19] in what follows.

Two cost functions were considered in [19]. One cost function C_{HST} quantifies the Hilbert–Schmidt inner product between the target unitary U and the trainable gate sequence V, as follows:

$$C_{\text{HST}} = 1 - F_{\text{HST}}, \quad \text{with} \quad F_{\text{HST}} = |\text{Tr}(V^\dagger U)|^2 / d^2, \tag{3}$$

where $d = 2^n$ is the Hilbert-space dimension and n is the number of qubits that U acts on, and where we write V instead of $V(\alpha)$ for simplicity. The circuit for computing C_{HST} is called the Hilbert–Schmidt Test (HST) and is shown in figure 1(a). First, one prepares a maximally entangled state $|\Phi\rangle^{AB}$ by acting with a depth-two circuit E, then one applies U followed by $V^\dagger$ on half of this maximally entangled state. Finally one measures the overlap with the original maximally entangled state $|\Phi\rangle^{AB}$ by applying $E^\dagger$ and quantifying the probability of the all-zeros measurement outcome. One can verify that this probability is equal to $F_{\text{HST}} = |\text{Tr}(V^\dagger U)|^2 / d^2$. This cost function is operationally meaningful since it is equivalent to the average fidelity $\overline{F}(U, V) = \int |\langle \psi | V^\dagger U | \psi \rangle|^2 \mathrm{d}\psi$ between states acted upon by U versus those acted upon by V, as follows [29, 30]:

$$C_{\text{HST}} = \frac{d+1}{d}(1 - \overline{F}(U, V)). \tag{4}$$

Note that C_{HST} is faithful in that $C_{\text{HST}} = 0$ iff $V = U$ (up to a global phase).

An alternative cost function [19] is given by

$$C_{\text{LHST}} = 1 - F_{\text{LHST}}, \quad \text{with} \quad F_{\text{LHST}} = \frac{1}{n}\sum_{j=1}^{n} F_{\text{LHST}}^{(j)}, \tag{5}$$
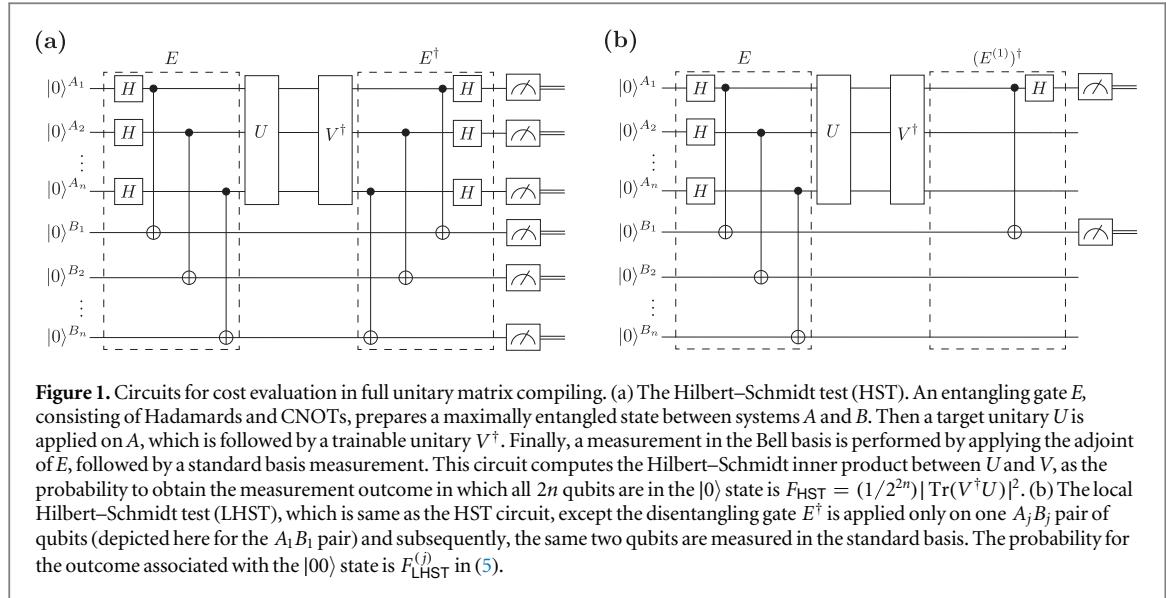
Figure 1. Circuits for cost evaluation in full unitary matrix compiling. (a) The Hilbert–Schmidt test (HST). An entangling gate E, consisting of Hadamards and CNOTs, prepares a maximally entangled state between systems A and B. Then a target unitary U is applied on A, which is followed by a trainable unitary $V^\dagger$. Finally, a measurement in the Bell basis is performed by applying the adjoint of E, followed by a standard basis measurement. This circuit computes the Hilbert–Schmidt inner product between U and V, as the probability to obtain the measurement outcome in which all $2n$ qubits are in the $|0\rangle$ state is $F_{\mathsf{HST}} = (1/2^{2n})|\operatorname{Tr}(V^\dagger U)|^2$. (b) The local Hilbert–Schmidt test (LHST), which is same as the HST circuit, except the disentangling gate $E^\dagger$ is applied only on one $A_j B_j$ pair of qubits (depicted here for the $A_1 B_1$ pair) and subsequently, the same two qubits are measured in the standard basis. The probability for the outcome associated with the $|00\rangle$ state is $F_{\mathsf{LHST}}^{(j)}$ in (5).

where $F_{\mathsf{LHST}}^{(j)}$ is the probability of the 00 measurement outcome in the local Hilbert–Schmidt test (LHST), which is the circuit shown in figure 1(b). Note that F_{HST} is the entanglement fidelity for the quantum channel defined by $V^\dagger U$. On the other hand, $F_{\mathsf{LHST}}^{(j)}$ is the entanglement fidelity for the quantum channel obtained from feeding into $V^\dagger U$ the maximally mixed state on $\overline{A}_j$ and then tracing over $\overline{A}_j$, where $\overline{A}_j$ consists of all qubits in A other than A_j. As shown in [19]

$$C_{\mathsf{LHST}} \leqslant C_{\mathsf{HST}} \leqslant n C_{\mathsf{LHST}}, \tag{6}$$

which implies that C_{LHST} is also a faithful cost function, i.e. $C_{\mathsf{LHST}} = 0$ iff $V = U$ (up to a global phase).

The overall cost function proposed by [19] was a convex combination of C_{HST} and C_{LHST}:

$$C(q) = q C_{\mathsf{HST}} + (1 - q) C_{\mathsf{LHST}}. \tag{7}$$

Here, q is a free parameter with $0 \leqslant q \leqslant 1$. The definition of $C(q)$ was motivated in [19] by the fact that C_{HST} has a direct operational meaning (equation (4)) but it becomes difficult to train for large n due to a vanishing gradient [31], whereas C_{LHST} is trainable but does not have a direct operational meaning. Hence one can take a weighted average of these two functions, where for small n one can choose $q \approx 1$, while for large n one can choose $q \approx 0$.

3.2. Compiling with a fixed input state

Fixed input state compiling (FISC) of a unitary matrix was introduced in [20, 19] and treated in significant detail in [20]. In this case, the goal is to train a gate sequence V so that it has the same effect as a target unitary U when acting on a given input state $|\psi_0\rangle$. For simplicity and due to its technological relevance, we will consider the case where $|\psi_0\rangle = |\mathbf{0}\rangle$ is the all-zero state, so that we are interested in training V to satisfy (up to a global phase):

$$U|\mathbf{0}\rangle = V|\mathbf{0}\rangle, \quad \text{or equivalently} \quad W|\mathbf{0}\rangle = |\mathbf{0}\rangle, \tag{8}$$

with $W = V^\dagger U$. To quantify how far $W|\mathbf{0}\rangle$ is from the state $|\mathbf{0}\rangle$, one can define the cost function

$$C_{\mathsf{LET}} = 1 - G_{\mathsf{LET}}, \tag{9}$$

where G_{LET} is the fidelity $F(\rho, \sigma) = (\operatorname{Tr}[\sqrt{\sqrt{\rho}\,\sigma\sqrt{\rho}}])^2$ between these two states:

$$G_{\mathsf{LET}} = F(|\mathbf{0}\rangle\langle\mathbf{0}|, W|\mathbf{0}\rangle\langle\mathbf{0}|W^\dagger) = |\langle\mathbf{0}|W|\mathbf{0}\rangle|^2 = \operatorname{Tr}[P_{\mathbf{0}} W|\mathbf{0}\rangle\langle\mathbf{0}|W^\dagger], \tag{10}$$

with $P_{\mathbf{0}} = |\mathbf{0}\rangle\langle\mathbf{0}|$ the projector onto the all-zero state. We employed the LET subscript here since we refer to the circuit used to quantify (9) and (10) as the Loschmidt echo test (LET), shown in figure 2(a). The Loschmidt echo [32] refers to a forward and backward time evolution with the intent of recovering the initial state. This is analogous to the circuit in figure 2(a) where one first evolves forward with U and then attempts to undo that evolution with $V^\dagger$, to recover the initial state $|\mathbf{0}\rangle$. Hence the probability of the all-zero measurement outcome in figure 2(a) is precisely G_{LET}.

One can see that compiling with a fixed input state leads to more freedom and hence more solutions than full unitary matrix compiling. Note that $C_{\mathsf{HST}} = 0$ iff $W = e^{i\phi}\mathbb{1}$ where ϕ is a global phase factor. On the other hand, $C_{\mathsf{LET}} = 0$ iff $|\langle\mathbf{0}|W|\mathbf{z}\rangle| = |\langle\mathbf{z}|W|\mathbf{0}\rangle| = \delta_{\mathbf{z},\mathbf{0}}$ for all bit strings $\mathbf{z}$. Hence, for W that achieve $C_{\mathsf{LET}} = 0$, the $(n-1) \times (n-1)$ unitary principal submatrix of W with matrix elements $\langle\mathbf{z}|W|\mathbf{z}'\rangle$ (such that $\mathbf{z}, \mathbf{z}' \neq \mathbf{0}$) remains completely arbitrary. This degeneracy of optima can simplify the optimization of V as any of these optima will lead to $C_{\mathsf{LET}} = 0$.
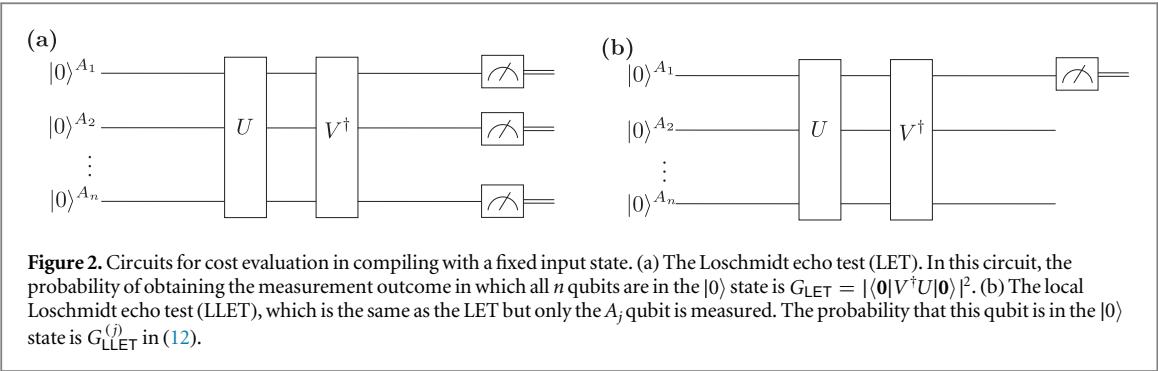
Figure 2. Circuits for cost evaluation in compiling with a fixed input state. (a) The Loschmidt echo test (LET). In this circuit, the probability of obtaining the measurement outcome in which all n qubits are in the $|0\rangle$ state is $G_{\text{LET}} = |\langle\mathbf{0}|V^{\dagger}U|\mathbf{0}\rangle|^2$. (b) The local Loschmidt echo test (LLET), which is the same as the LET but only the A_j qubit is measured. The probability that this qubit is in the $|0\rangle$ state is $G_{\text{LLET}}^{(j)}$ in (12).

Analogous to the LHST cost for full unitary matrix compiling, one can define a cost function for fixed input state compiling that involves local observables:

$$C_{\text{LLET}} = 1 - G_{\text{LLET}} = 1 - \frac{1}{n}\sum_{j=1}^{n} G_{\text{LLET}}^{(j)}, \qquad \text{with} \qquad G_{\text{LLET}}^{(j)} = \text{Tr}[(P_0^{A_j} \otimes \mathbb{1}^{\overline{A}_j})W|\mathbf{0}\rangle\langle\mathbf{0}|W^{\dagger}]. \tag{11}$$

Here, $P_0^{A_j}$ is the projector onto the zero state on the A_j qubit, and $\mathbb{1}^{\overline{A}_j}$ denotes the identity on all qubits except A_j and n is the number of qubits. We call the circuit used to compute C_{LLET} the local Loschmidt echo test (LLET), and this circuit is shown in figure 2(b). Note that

$$G_{\text{LLET}}^{(j)} = \text{Tr}_{A_j}[P_0^{A_j}\rho^{(j)}] = \langle 0|\rho^{(j)}|0\rangle = F(|0\rangle\langle 0|, \rho^{(j)}), \tag{12}$$

where $\rho^{(j)} = \text{Tr}_{\overline{A}_j}[W|\mathbf{0}\rangle\langle\mathbf{0}|W^{\dagger}]$. Hence $G_{\text{LLET}}^{(j)}$ corresponds to the probability of the zero outcome for the circuit in figure 2(b). With a proof similar to that of (6) one can show that

$$C_{\text{LLET}} \leqslant C_{\text{LET}} \leqslant nC_{\text{LLET}}, \tag{13}$$

and hence $C_{\text{LLET}} = 0$ iff $C_{\text{LET}} = 0$. Furthermore, one can define an overall cost function analogous to $C(q)$ in (7)

$$C'(q) = qC_{\text{LET}} + (1-q)C_{\text{LLET}}, \tag{14}$$

which again is motivated by the fact that C_{LET} has a direct operational meaning but is difficult to train for large n, whereas the opposite is true for C_{LLET}. Hence one can take $q \approx 1$ for small n and $q \approx 0$ for large n.

4. Noise processes

In this work, we consider three different types of noise [33, 34]: (1) decoherence noise, (2) gate noise, and (3) measurement noise. We now discuss how we mathematically model these three types of noise.

Let us start with decoherence. Physical models of decoherence often refer to T_1 and T_2 processes, which respectively pertain to thermal relaxation (energy dissipation) and dephasing (loss of phase coherence). These processes are typically modeled as local quantum channels acting independently on individual qubits. However, mathematically it is easier to deal with classes of quantum channels that act globally on sets of qubits (which can contain the independent local channels as a special case). In what follows, we define three types of global quantum channels: depolarizing noise, Pauli noise, and non-unital Pauli noise. It is worth noting that Pauli noise includes T_2 processes as a special case (i.e. the dephasing channel is a Pauli channel), and non-unital Pauli noise includes T_1 processes as a special case (i.e. the amplitude damping channel is a non-unital Pauli channel). Consider the following precise definitions.

Definition 1. We define depolarizing noise (DN) as a completely positive trace-preserving (CPTP) map that maps an n-qubit state ρ to the state $p\rho + (1-p)\mathbb{1}/(2^n)$.

Definition 2. We define Pauli Noise (PN) as a CPTP map $\mathcal{P}$ whose superoperator is diagonal in the Pauli basis. In other words, its action on a Pauli operator $X^{\boldsymbol{l}}Z^{\boldsymbol{k}} := X^{l_1}Z^{k_1} \otimes \ldots \otimes X^{l_n}Z^{k_n}$ is given by $\mathcal{P}(X^{\boldsymbol{l}}Z^{\boldsymbol{k}}) = c_{\boldsymbol{lk}}X^{\boldsymbol{l}}Z^{\boldsymbol{k}}$, where $c_{\boldsymbol{00}} = 1$. Furthermore, we assume that $c_{\boldsymbol{lk}} \geqslant 0$ for all $\boldsymbol{l}$ and $\boldsymbol{k}$, where $l_1, \ldots, l_n, k_1, \ldots, k_n \in \{0, 1\}$.

Definition 3. We define non-unital Pauli noise (NUPN) as a CPTP map $\mathcal{P}_{\text{NU}}$ whose action on the identity is $\mathcal{P}_{\text{NU}}(\mathbb{1}) = \mathbb{1} + \sum_{(\boldsymbol{l},\boldsymbol{k})\neq(\boldsymbol{0},\boldsymbol{0})}d_{\boldsymbol{lk}}X^{\boldsymbol{l}}Z^{\boldsymbol{k}}$, and whose action on all other Pauli operators $X^{\boldsymbol{l}}Z^{\boldsymbol{k}}$ with $(\boldsymbol{l}, \boldsymbol{k}) \neq (\boldsymbol{0}, \boldsymbol{0})$ is given by $\mathcal{P}_{\text{NU}}(X^{\boldsymbol{l}}Z^{\boldsymbol{k}}) = c_{\boldsymbol{lk}}X^{\boldsymbol{l}}Z^{\boldsymbol{k}}$. Furthermore, we assume that $c_{\boldsymbol{lk}} \geqslant 0$ for all $\boldsymbol{l}$ and $\boldsymbol{k}$.

Next, we consider gate noise. While gate noise can involve coherent errors such as systematic gate bias, such errors are hardware-specific, and hence we focus on incoherent gate noise. We consider a simple model for gate noise in which every time a gate is implemented, a Pauli channel acts both before and after this gate. Furthermore, for generality, we allow these Pauli channels to act globally on all qubits, which serves as a model for cross-talk (where gates affect qubits on which they are intended to act trivially).

Definition 4. We define Pauli gate noise (PGN) as a simple noise model in which all gates are preceded and followed by global Pauli channels. In other words, for a gate G, instead of its action on a state ρ being $G\rho G^\dagger$, we model its action as $\mathcal{P}'(G\mathcal{P}(\rho)G^\dagger)$ where $\mathcal{P}$ and $\mathcal{P}'$ are Pauli channels. Note that these Pauli channels act on all qubits, including qubits on which G acts trivially.

Finally, we consider measurement noise, also known as readout error. For a single qubit, we model measurement noise as a classical bit-flip channel, where feeding in the standard basis state $|l\rangle$ leads to the k outcome with probability p_{kl}. We allow for asymmetry in that one can have $p_{01} \neq p_{10}$, which is an important generality, e.g. when T_1 noise occurs during the measurement process. For multiple qubits, our measurement noise model is a tensor product of the aforementioned bit-flip channels, corresponding to uncorrelated measurement noise.

Definition 5. We define measurement noise (MN) as a modification of the standard-basis *POVM* elements, which are $\{P_0 = |0\rangle\langle 0|, P_1 = |1\rangle\langle 1|\}$ for a noiseless single qubit. With measurement noise, this *POVM* gets replaced by $\{\widetilde{P}_0, \widetilde{P}_1\}$, with $\widetilde{P}_0 = p_{00}|0\rangle\langle 0| + p_{01}|1\rangle\langle 1|$ and $\widetilde{P}_1 = p_{10}|0\rangle\langle 0| + p_{11}|1\rangle\langle 1|$, where $p_{00} + p_{10} = 1$, $p_{01} + p_{11} = 1$, and p_{kl} is the probability of getting the k outcome given the l input. Furthermore we assume that $p_{kk} > p_{kl}$ for $l \neq k$. Hence, for an n-qubit standard-basis measurement with measurement noise, we write the *POVM* element associated with the bit string $\boldsymbol{z} = (z_1, \ldots, z_n)$ as

$$\widetilde{P}_{\boldsymbol{z}} = \bigotimes_{j=1}^{n} (p_{z_j 0}^{(j)}|0\rangle\langle 0| + p_{z_j 1}^{(j)}|1\rangle\langle 1|), \tag{15}$$

with $\sum_{z_j} p_{z_j 0}^{(j)} = 1$ and $\sum_{z_j} p_{z_j 1}^{(j)} = 1$, and we assume that $p_{z_j z_j}^{(j)} > p_{z_j l}^{(j)}$ for $l \neq z_j$.

5. Main results

Before proceeding to the main results we first define two versions of optimal parameter resilience (OPR), i.e. of learning the correct gate sequence V despite various sources of noise, which we refer to as strong-OPR and weak-OPR.

Definition 6. Let $\mathbb{V}_d$ be the set of $d \times d$ unitary matrices. Let $C_{\mathsf{QC}}(V)$ be a cost function of V with $V \in \mathbb{V}_d$, and suppose that $C_{\mathsf{QC}}(V)$ can be evaluated using a quantum circuit denoted QC. Let $\widetilde{C}_{\mathsf{QC}}(V)$ denote the noisy version of $C_{\mathsf{QC}}(V)$, i.e. the corresponding function whenever the circuit QC is run in the presence of some noise process $\mathcal{N}$. Let $\mathbb{V}_d^{\mathrm{opt}}$ and $\widetilde{\mathbb{V}}_d^{\mathrm{opt}}$ respectively denote the sets of unitaries that optimize $C_{\mathsf{QC}}(V)$ and $\widetilde{C}_{\mathsf{QC}}(V)$, i.e.

$$\mathbb{V}_d^{\mathrm{opt}} = \{V' \in \mathbb{V}_d : C_{\mathsf{QC}}(V') = \min_{V \in \mathbb{V}_d} C_{\mathsf{QC}}(V)\}, \tag{16}$$

$$\widetilde{\mathbb{V}}_d^{\mathrm{opt}} = \{V' \in \mathbb{V}_d : \widetilde{C}_{\mathsf{QC}}(V') = \min_{V \in \mathbb{V}_d} \widetilde{C}_{\mathsf{QC}}(V)\}. \tag{17}$$

We say that $C_{\mathsf{QC}}(V)$ exhibits strong-OPR to $\mathcal{N}$ if $\widetilde{\mathbb{V}}_d^{\mathrm{opt}} = \mathbb{V}_d^{\mathrm{opt}}$. We say that $C_{\mathsf{QC}}(V)$ exhibits weak-OPR to $\mathcal{N}$ if $\widetilde{\mathbb{V}}_d^{\mathrm{opt}} \subseteq \mathbb{V}_d^{\mathrm{opt}}$.

5.1. Noise resilience of full unitary matrix compiling

Let us begin with full unitary matrix compiling (FUMC). Figure 3 shows the two noise models that we will consider for FUMC. As shown in this figure, τ_1 and τ_2 are respectively defined as the times just before and just after the application of $V^\dagger U$. We note that the noise models considered in figure 3 capture fairly well the physical noise that is present in, e.g. superconducting-qubit quantum computers, with the exception that only depolarizing noise is allowed during the action of $V^\dagger U$. We make this simplification for ease of analysis, although our numerics in section 6 relax this assumption.

Consider the following definition for the noise model depicted in figure 3(a).

Definition 7. We define noise Model 1 to be the following noise process during the *HST* circuit: (1) global depolarizing noise acting continuously throughout the circuit, (2) global Pauli noise at times τ_1 and τ_2, (3) global depolarizing noise on system A acting continuously in between τ_1 and τ_2, (4) global non-unital Pauli noise on

Figure 3. Schematic diagram of: (a) Noise Model 1 of definition 7, and (b) Noise Model 2 of definition 8. The following acronyms are employed: depolarizing noise (DN), Pauli gate noise (PGN), Pauli noise (PN), non-unital Pauli noise (NUPN), and measurement noise (MN). Red dashed boxes indicate the time period and the qubits on which the noise process acts. Time τ_1 (τ_2) corresponds to the time immediately before (after) the action of the unitary $V^\dagger U$. While both panels show the HST, these noise models are also applicable to the LHST, provided one replaces $E^\dagger$ with $(E^{(j)})^\dagger$.

system B acting continuously in between τ_1 and τ_2, (5) Pauli gate noise during E and $E^\dagger$, and (6) measurement noise. We also use the term Noise Model 1 when the same noise model acts during the LHST circuit, provided one replaces $E^\dagger$ with $(E^{(j)})^\dagger$.

We now state our first main result. The proof of this result is given in appendix D, with some useful preliminaries and lemmas given in appendices A–C.

Theorem 1. *The cost functions C_{HST} and C_{LHST} exhibit strong-OPR to Noise Model 1 in definition 7.*

Note that this theorem also implies that $C(q) = qC_{\mathsf{HST}} + (1 - q)C_{\mathsf{LHST}}$ exhibits strong-OPR to Noise Model 1, for all values of q. This is because the set $\mathbb{V}_d^{\mathsf{opt}} = \widetilde{\mathbb{V}}_d^{\mathsf{opt}}$ defined in (16) and (17) is the same for C_{HST} and C_{LHST} functions. Hence this same set is optimal for $C(q)$.

Consider the implications of theorem 1. First, this theorem implies that FUMC is resilient to the measurement noise model in definition 5. Second, FUMC is completely resilient to Pauli gate noise during the entangling and disentangling gates, E and $E^\dagger$. Note that this Pauli gate noise is global and hence accounts for cross talk. Third, FUMC is resilient to global depolarizing noise acting continuously throughout the circuit, as well as global Pauli noise acting at the specific times τ_1 and τ_2. Fourth, FUMC is resilient to depolarizing noise acting on system A and non-unital Pauli noise acting on system B, provided that each of these process act (possibly continuously) during the time interval between τ_1 and τ_2. We emphasize that Pauli noise includes dephasing channels (T_2 noise) as a special case, while non-unital Pauli noise includes the depashing channel (T_1 noise) as a special case. Importantly, theorem 1 states that FUMC is resilient to the general case where all of these noise processes occur together.

We now state our second main result (proven in appendix E), which deals with the noise model in figure 3(b).

Definition 8. We define Noise Model 2 to be the following noise process during the *HST* circuit: (1) global depolarizing noise acting continuously throughout the circuit, (2) global Pauli noise at times τ_1 and τ_2, (3) global non-unital Pauli noise on system A at time τ_1, (4) global depolarizing noise on system A acting continuously in between τ_1 and τ_2, (5) global Pauli noise on system B acting continuously in between τ_1 and τ_2, (6) Pauli gate noise during E and $E^\dagger$, and (7) measurement noise. We also use the term Noise Model 2 when the same noise model acts during the *LHST* circuit, provided one replaces $E^\dagger$ with $(E^{(j)})^\dagger$.

Theorem 2. *The cost functions C_{HST} and C_{LHST} exhibit strong-OPR to Noise Model 2 in definition 8.*

The implications of theorem 2 are similar to those of theorem 1. The main difference is that theorem 2 allows for non-unital Pauli noise on system A at time τ_1, at the expense of only allowing Pauli noise to act continuously on system B between τ_1 and τ_2. The other aspects of the noise models treated by these two theorems are identical.

The above two theorems immediately imply several corollaries below. These corollaries establish resilience to noise models that are different and in some cases more general than the noise models previously considered, at the expense of possibly specializing the form of the unitary $W = V^\dagger U$. See appendix G for the proofs of all corollaries.

Corollary 1. *The cost functions C_{HST} and C_{LHST} exhibit strong-OPR to a noise model that includes the following: (1) all noise processes in Noise Model 1, as well as (2) a noise process during the implementation of $\mathcal{W} = \mathcal{W}_k \circ \cdots \circ \mathcal{W}_1 = \mathcal{V}^\dagger \circ \mathcal{U}$ (i.e. in the time interval between τ_1 and τ_2) in which global Pauli channels $\{\mathcal{P}_1^A, \ldots, \mathcal{P}_k^A\}$ act on system A, such that the overall channel on A is $\mathcal{P}_k^A \circ \mathcal{W}_k \cdots \circ \mathcal{P}_1^A \circ \mathcal{W}_1$, provided that the following condition is satisfied:*

$$(\mathcal{P}_k^A \circ \mathcal{W}_k \cdots \circ \mathcal{P}_1^A \circ \mathcal{W}_1)(\cdot) = (\mathcal{W}_k \circ \mathcal{W}_{k-1} \cdots \circ \mathcal{W}_1 \circ \widehat{\mathcal{P}}^A)(\cdot). \tag{18}$$

Here $\widehat{\mathcal{P}}^A$ is also a Pauli channel, and the channels $\mathcal{U}$, $\mathcal{V}^\dagger$, and $\mathcal{W}$ correspond to conjugating the state by the unitaries U, $V^\dagger$, and W, respectively.

The condition in (18) implies that the overall channel consisting of global Pauli channels acting on system A during the implementation of $\mathcal{W}$ is mathematically equivalent (although physically inequivalent) to a Pauli channel followed by $\mathcal{W}$. Therefore, corollary 1 follows from theorem 1.

Consider the following implications of corollary 1. Unitaries corresponding to the Clifford group necessarily satisfy the condition in (18), as shown in appendix A. Therefore, corollary 2 below holds for any Clifford unitary W. Moreover, tensor-product unitaries satisfy this same condition provided that the noise is local depolarizing noise, and hence corollary 3 below also follows from corollary 1.

Corollary 2. *Let the $W = V^\dagger U$ gate sequence have the form $W = W_2^A W_1^A$ with W_1^A composed only of Clifford gates. Then the cost functions C_{HST} and C_{LHST} exhibit strong-OPR to a noise model that includes the following: (1) all noise processes in Noise Model 1, as well as (2) a noise process during the implementation of $\mathcal{W}_1^A = \mathcal{W}_{1,k} \circ \cdots \circ \mathcal{W}_{1,1}$, in which global Pauli channels $\{\mathcal{P}_1^A, \ldots, \mathcal{P}_k^A\}$ act on system A, such that the overall channel on A is $\mathcal{P}_k^A \circ \mathcal{W}_{1,k} \cdots \circ \mathcal{P}_1^A \circ \mathcal{W}_{1,1}$.*

Corollary 3. *Let the $W = V^\dagger U$ gate sequence have the form $W = W_2^A W_1^A$ with $W_1^A = W_1^{A'} \otimes W_1^{A''}$ being a tensor product, i.e. W is a tensor product up to a particular time. Then the cost functions C_{HST} and C_{LHST} exhibit strong-OPR to a noise model that includes the following: (1) all noise processes in Noise Model 1, as well as (2) a noise process during the implementations of $\mathcal{W}_1^{A'} = \mathcal{W}_{1,k}^{A'} \circ \cdots \circ \mathcal{W}_{1,1}^{A'}$ and $\mathcal{W}_1^{A''} = \mathcal{W}_{1,l}^{A''} \circ \cdots \circ \mathcal{W}_{1,1}^{A''}$ in which local depolarizing channels $\{\mathcal{D}_{1,1}^{A'}, \ldots, \mathcal{D}_{1,k}^{A'}\}$ and $\{\mathcal{D}_{1,1}^{A''}, \ldots, \mathcal{D}_{1,l}^{A''}\}$ act on subsystems A' and A'', respectively, such that the overall channel on $A = A'A''$ is $(\mathcal{D}_{1,k}^{A'} \circ \mathcal{W}_{1,k}^{A'} \ldots \mathcal{D}_{1,1}^{A'} \circ \mathcal{W}_{1,1}^{A'}) \otimes (\mathcal{D}_{1,l}^{A''} \circ \mathcal{W}_{1,l}^{A''} \ldots \mathcal{D}_{1,1}^{A''} \circ \mathcal{W}_{1,1}^{A''})$.*

The following corollary follows from theorem 2 and is analogous to corollary 1.

Corollary 4. *The cost functions C_{HST} and C_{LHST} exhibit strong-OPR to the following noise model: (1) all noise processes in Noise Model 2, as well as (2) a noise process during the implementation of $\mathcal{W} = \mathcal{W}_k \circ \cdots \circ \mathcal{W}_1 = \mathcal{V}^\dagger \circ \mathcal{U}$ (i.e. in the time interval between τ_1 and τ_2) in which global non-unital Pauli channels $\{\mathcal{P}_{\mathrm{NU},1}^A, \ldots, \mathcal{P}_{\mathrm{NU},k}^A\}$ act on system A such that the overall channel on A is $\mathcal{P}_{\mathrm{NU},k}^A \circ \mathcal{W}_k \cdots \circ \mathcal{P}_{\mathrm{NU},1}^A \circ \mathcal{W}_1$, provided that the following condition is satisfied:*

$$(\mathcal{P}_{\mathrm{NU},k}^A \circ \mathcal{W}_k \cdots \circ \mathcal{P}_{\mathrm{NU},1}^A \circ \mathcal{W}_1)(\cdot) = (\mathcal{W}_k \circ \mathcal{W}_{k-1} \cdots \mathcal{W}_1 \circ \widehat{\mathcal{P}}_{\mathrm{NU}}^A)(\cdot), \tag{19}$$

where $\widehat{\mathcal{P}}_{\mathrm{NU}}^A$ is also a non-unital Pauli channel.

Finally, we present a simple corollary of theorem 1 based on the ricochet property of the standard Bell state. Note that the noise model in the following corollary is fairly simple but nonetheless physically distinct from those considered in figure 3, since it allows for global non-unital Pauli noise to occur during the implementation of W.

Corollary 5. *The cost functions C_{HST} exhibits strong-OPR to the following noise model: (1) global depolarizing noise acting continuously throughout the circuit, (2) global non-unital Pauli noise on system A at a fixed time in between τ_1 and τ_2.*

5.2. Noise resilience of fixed input state compiling

Let us now consider fixed input state compiling (FISC). Recall that the cost-evaluation circuits, shown in figure 2, have less structure than the circuits in figure 1. As a result, the noise model that we consider in the FISC case is simpler than the previously considered noise models. In particular, we define the following noise model, which is depicted in figure 4. Note that, in this context, τ_1 is defined as the time just before the application of $V^\dagger U$, and there is no need to consider a noisy quantum channel occurring after $V^\dagger U$ since the measurement occurs immediately after $V^\dagger U$.
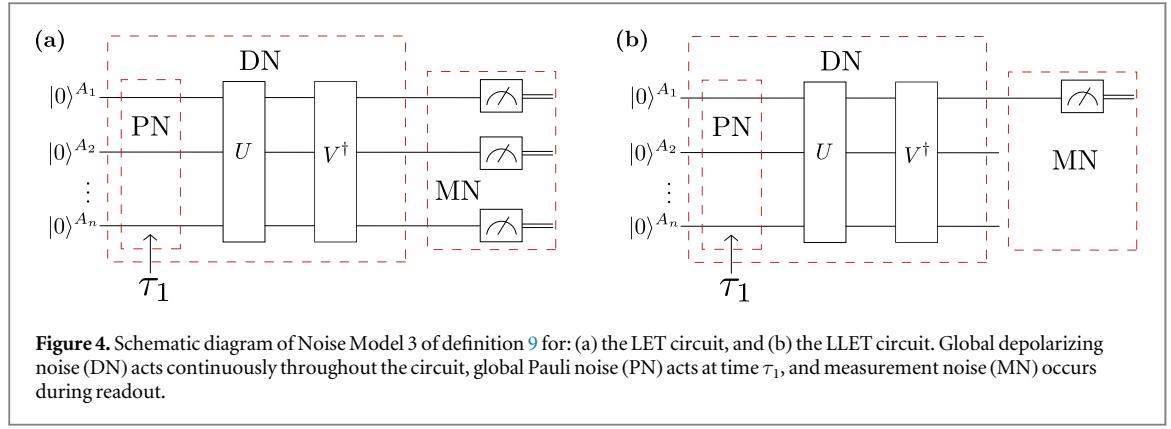
Figure 4. Schematic diagram of Noise Model 3 of definition 9 for: (a) the LET circuit, and (b) the LLET circuit. Global depolarizing noise (DN) acts continuously throughout the circuit, global Pauli noise (PN) acts at time τ_1, and measurement noise (MN) occurs during readout.

Definition 9. We define Noise Model 3 to be the following noise process during the LET or the LLET: (1) global depolarizing noise acting continuously throughout the circuit, (2) global Pauli noise acting at time τ_1, and (3) measurement noise.

We now state our main result for FISC, which is proven in appendix F.

Theorem 3. *The cost functions* C_{LET} *and* C_{LLET} *exhibit weak-OPR, as defined in definition 6, to Noise Model 3 in definition 9.*

This theorem implies that FISC is resilient to the measurement noise model in definition 5. Furthermore, it is resilient to Pauli noise acting at τ_1 and global depolarizing noise acting continuously throughout the circuit.

We remark that while FUMC exhibits strong-OPR for the noise models considered (see the previous section), here FISC exhibits weak-OPR instead. The latter arises from the fact that the optimal set of unitaries $\mathbb{V}_d^{\mathrm{opt}}$ for FISC can be highly degenerate (i.e. can contain many unitaries) and the presence of noise could in general break such degeneracy. The 'weak' term in weak-OPR is simply the fact that the number of global optima is possibly reduced by noise, not that the noise resilience itself is weak. Hence, weak-OPR should still be viewed as noise resilience, since the global optima in the presence of noise correspond to global optima in the noiseless case. This implies that training in the presence of noise will lead one to find the correct optimal parameters for $V(\boldsymbol{\alpha})$.

Under certain conditions, theorem 3 implies that $C'(q)$ defined in (14) will also exhibit weak-OPR to Noise Model 3. Let $\mathbb{V}_{d,\mathsf{LET}}^{\mathrm{opt}}$ and $\mathbb{V}_{d,\mathsf{LLET}}^{\mathrm{opt}}$ denote the sets of unitaries that optimize C_{LET} and C_{LLET}, respectively. In the absence of noise we have $\mathbb{V}_{d,\mathsf{LET}}^{\mathrm{opt}} = \mathbb{V}_{d,\mathsf{LLET}}^{\mathrm{opt}}$, while in the presence of noise, theorem 3 implies $\widetilde{\mathbb{V}}_{d,\mathsf{LET}}^{\mathrm{opt}} \subseteq \mathbb{V}_{d,\mathsf{LET}}^{\mathrm{opt}}$ and $\widetilde{\mathbb{V}}_{d,\mathsf{LLET}}^{\mathrm{opt}} \subseteq \mathbb{V}_{d,\mathsf{LLET}}^{\mathrm{opt}}$. Hence, if $\widetilde{\mathbb{V}}_{d,\mathsf{LET}}^{\mathrm{opt}} \cap \widetilde{\mathbb{V}}_{d,\mathsf{LLET}}^{\mathrm{opt}} \neq \varnothing$, then for any value of q, $C'(q) = qC_{\mathsf{LET}} + (1-q)C_{\mathsf{LLET}}$ will also exhibit weak-OPR to Noise Model 3, where the unitaries that optimize $C'(q)$ in the noisy case belong to $\widetilde{\mathbb{V}}_{d,\mathsf{LET}}^{\mathrm{opt}} \cap \widetilde{\mathbb{V}}_{d,\mathsf{LLET}}^{\mathrm{opt}}$.

Theorem 3 implies the following corollaries, which establish resilience to noise models that go beyond Noise Model 3 at the expense of specializing the form of W. Note that these corollaries are analogous to Corollaries 1–3, and corollary 6 implies Corollaries 7 and 8. See appendix G for the proofs.

Corollary 6. *The cost functions* C_{LET} *and* C_{LLET} *exhibit weak-OPR to a noise model that includes the following: (1) all noise processes in Noise Model 3, as well as (2) a noise process during the implementation of* $\mathcal{W} = \mathcal{W}_k \circ \cdots \circ \mathcal{W}_1 = \mathcal{V}^\dagger \circ \mathcal{U}$ *in which global Pauli channels* $\{\mathcal{P}_1, \dots, \mathcal{P}_k\}$ *act, such that the overall channel is* $\mathcal{P}_k \circ \mathcal{W}_k \cdots \circ \mathcal{P}_1 \circ \mathcal{W}_1$, *provided that the following condition is satisfied:*

$$(\mathcal{P}_k \circ \mathcal{W}_k \cdots \circ \mathcal{P}_1 \circ \mathcal{W}_1)(\cdot) = (\mathcal{W}_k \circ \mathcal{W}_{k-1} \cdots \circ \mathcal{W}_1 \circ \widehat{\mathcal{P}})(\cdot), \tag{20}$$

where $\widehat{\mathcal{P}}$ *is also a Pauli channel.*

Corollary 7. *Let the* $W = V^\dagger U$ *gate sequence have the form* $W = W_2^A W_1^A$ *with* W_1^A *composed only of Clifford gates. Then the cost functions* C_{LET} *and* C_{LLET} *exhibit weak-OPR to a noise model that includes the following: (1) all noise processes in Noise Model 3, as well as (2) a noise process during the implementation of* $\mathcal{W}_1^A = \mathcal{W}_{1,k} \circ \cdots \circ \mathcal{W}_{1,1}$, *in which global Pauli channels* $\{\mathcal{P}_1^A, \dots, \mathcal{P}_k^A\}$ *act on system A, such that the overall channel on A is* $\mathcal{P}_k^A \circ \mathcal{W}_{1,k} \cdots \circ \mathcal{P}_1^A \circ \mathcal{W}_{1,1}$.
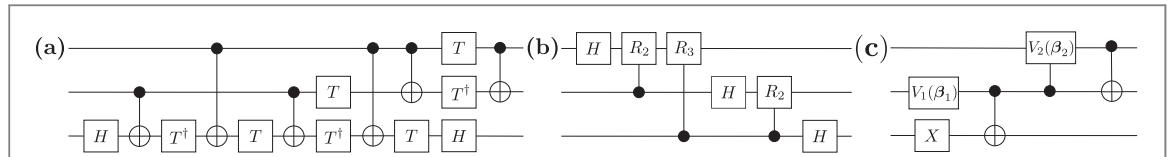
Figure 5. Quantum circuits for: (a) Toffoli Gate, (b) three-qubit quantum Fourier transform, and (c) three-qubit W-state preparation. Here, R_m stands for the controlled phase gate with a phase shift of $\phi = e^{2\pi i/2^m}$, and $V_k(\beta_k)$ is given by (21). For the three-qubit W-state preparation circuit we have $\beta_1 = (2\arccos(\sqrt{1/3}), 0, 0)$ and $\beta_2 = (\pi/2, 0, 0)$.
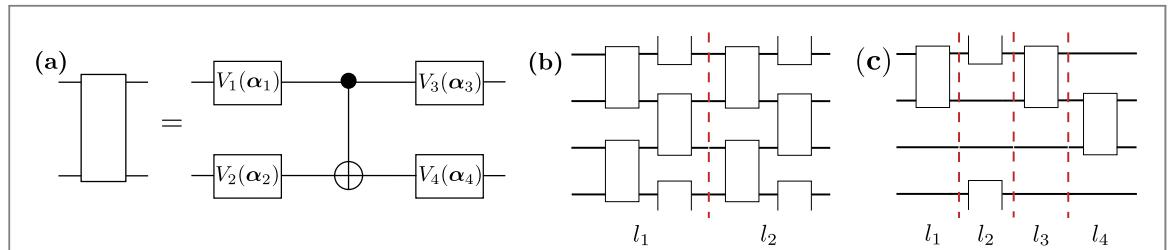
Figure 6. (a) The dressed CNOT is composed of a CNOT preceded and followed by single-qubit gates $V_k(\alpha_k)$, where $V_k(\alpha_k)$ is given by (21). (b) Two layers of the alternating-pair ansatz in the case of four qubits. Each layer is composed of dressed CNOTs acting on alternating pairs of neighboring qubits. (c) Schematic representation of the target-inspired ansatz. In this approach, the gate sequence of dressed CNOTs is obtained from the gate sequence of the target unitary U.

Corollary 8. *Let the* $W = V^\dagger U$ *gate sequence have the form* $W = W_2^A W_1^A$ *with* $W_1^A = W_1^{A'} \otimes W_1^{A''}$ *being a tensor product, i.e.* W *is a tensor product up to a particular time. Then the cost functions* C_{LET} *and* C_{LLET} *exhibit weak-OPR to a noise model that includes the following: (1) all noise processes in Noise Model 3, as well as (2) a noise process during the implementations of* $\mathcal{W}_1^{A'} = \mathcal{W}_{1,k}^{A'} \circ \cdots \circ \mathcal{W}_{1,1}^{A'}$ *and* $\mathcal{W}_1^{A''} = \mathcal{W}_{1,l}^{A''} \circ \cdots \circ \mathcal{W}_{1,1}^{A''}$ *in which local depolarizing channels* $\{\mathcal{D}_{1,1}^{A'}, ..., \mathcal{D}_{1,k}^{A'}\}$ *and* $\{\mathcal{D}_{1,1}^{A''}, ..., \mathcal{D}_{1,l}^{A''}\}$ *act on subsystems* A' *and* A'', *respectively, such that the overall channel on* $A = A'A''$ *is* $(\mathcal{D}_{1,k}^{A'} \circ \mathcal{W}_{1,k}^{A'} ... \mathcal{D}_{1,1}^{A'} \circ \mathcal{W}_{1,1}^{A'}) \otimes (\mathcal{D}_{1,l}^{A''} \circ \mathcal{W}_{1,l}^{A''} ... \mathcal{D}_{1,1}^{A''} \circ \mathcal{W}_{1,1}^{A''})$.

6. Implementations

In this section, we present the results of implementing VQC on the following three-qubit unitaries: the Toffoli gate, the three-qubit quantum Fourier transform (QFT), and a W-state preparation circuit. Each of these unitaries is of interest, e.g. the Toffoli gate when combined with the Hadamard gate provides a universal gate set for quantum computing [35], the QFT is a subroutine in Shor's algorithm [36], and W-state preparation is useful for the quantum approximate optimization algorithm [37, 9]. Figure 5 shows gate sequences corresponding to these unitaries obtained from the literature. The Toffoli gate in figure 5(a) is decomposed into a gate sequence that contains nine one-qubit gates and six CNOTs [38]. For the QFT we employ its textbook circuit [33] in figure 5(b), while the circuit for W-state preparation in figure 5(c) was derived from [39, 40].

Our VQC implementations were performed using IBM's noisy quantum simulator [28] with a noise model built from the reported noise parameters and connectivity of IBM's 14-qubit Melbourne quantum computer [41]. We remark that for VQC, we must have a target unitary U that is written as a gate sequence in the native gate language and the native connectivity of the hardware. IBM's simulator for the Melbourne device has a square lattice connectivity and native gate alphabet of CNOTs, arbitrary rotation around Z and $\pi/2$ rotation around X. Hence, transforming the gate sequences in figure 5 for the native device will typically add an overhead of additional gates. Therefore, the target gate sequences in our implementations actually correspond to IBM's compilation (with this overhead included) of the circuits in figure 5.

In IBM's noise model [28, 42], one-qubit gate errors are modeled as a single-qubit depolarizing error followed by a thermal relaxation error, where thermal relaxation refers to both T_1 and T_2 channels. Similarly, two-qubit gate errors consist of a two-qubit depolarizing error followed by single-qubit thermal relaxation errors on each qubit. Finally, the noise model includes single-qubit readout errors.

We employ two different ansatzes, shown in figure 6, and (as described below) we employ gradient-based optimization algorithms to train the gate sequence $V(\alpha)$. In figures 7–8, we plot the results of implementing VQC with IBM's noisy simulator for the three-qubit gates in figure 5. In each plot, we show the value of the noisy cost functions versus the number of iterations of the optimization algorithm. Additionally, we plot the corresponding value of the noiseless cost functions evaluated for the variational parameters α obtained from the
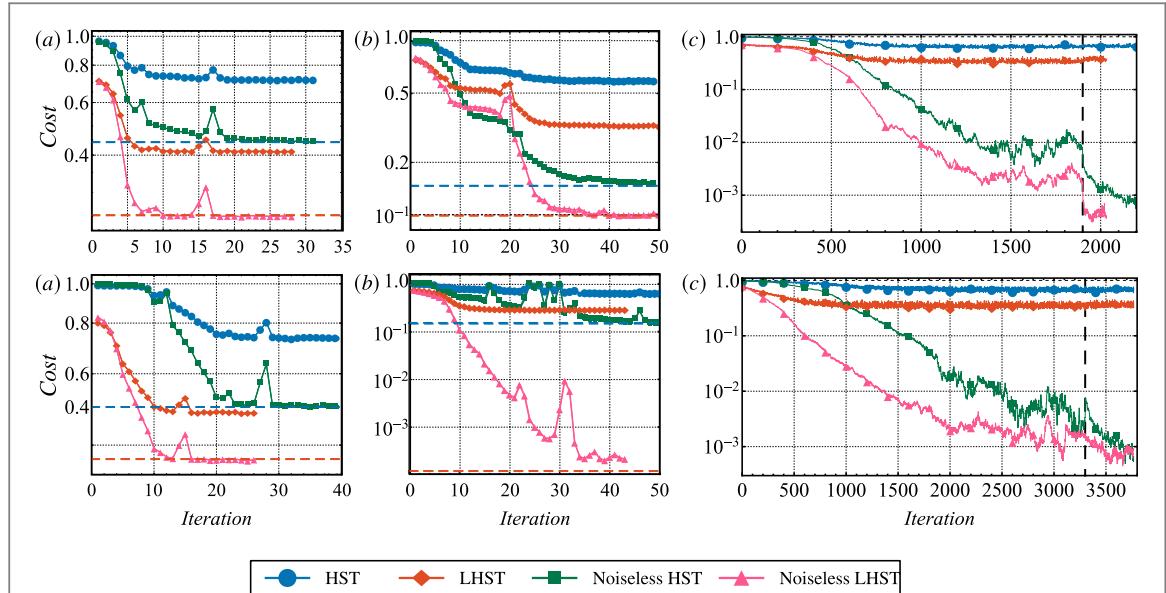
Figure 7. VQC implementations for the Toffoli gate (top) and three-qubit QFT (bottom). The ansatz for $V(\alpha)$ is: (a) one layer of the alternating-pair ansatz, (b) two layers of the alternating-pair ansatz, (c) the target-inspired ansatz. The blue and green curves respectively plot the values of $\widetilde{C}_{HST}$ and $\widetilde{C}_{LHST}$ obtained by training $V(\alpha)$ in the presence of noise. The green and pink curves respectively plot the values of C_{HST} and C_{LHST} evaluated at the variational parameters α obtained from the noisy optimization of $V(\alpha)$. Curves are plotted as a function of the number of iterations in the gradient-descent algorithm, and the y-axis is in log-scale. The blue and red dashed lines in (a) and (b) correspond to the minimum value of C_{HST} and C_{LHST}, respectively, determined by optimizing $V(\alpha)$ in a noise-free environment. Top: in both (a) and (b), the green and pink curves converge to the dashed blue and red lines, respectively. Bottom: While in (a) the green and pink curves converge to the dashed lines, in (b) the termination condition for the optimization algorithm was reached before the pink curve could achieve convergence. The number of shots per iteration was $N = 50\,000$ for (a) and (b). For (c) we employed the iCANS optimizer [44], where the total number of shots was 1.4×10^7 and the minimum number of shots per iteration was initially $N_{min} = 2$. The thick dashed vertical line in (c) indicates the point where we set $N_{min} = 250$, which helped to further reduce the cost function.
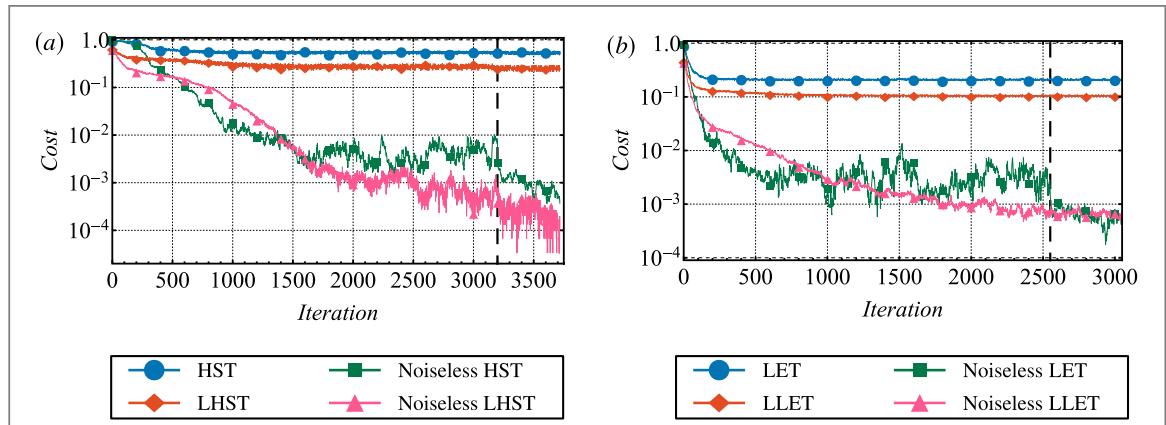
Figure 8. VQC implementations for the three-qubit W-state preparation circuit for (a) the FUMC approach, and (b) the FISC approach. The trainable gate sequence $V(\alpha)$ is given by the target-inspired ansatz. In the left (right) panel the blue and green curves plot respectively the values of $\widetilde{C}_{LET}$ ($\widetilde{C}_{LLET}$) obtained by noisy training of $V(\alpha)$. Similarly, in the left (right) panel the green and pink curves give respectively the values of C_{HST} ($\widetilde{C}_{LET}$) and C_{LHST} ($\widetilde{C}_{LLET}$) evaluated at the variational parameters α obtained from the noisy optimization of $V(\alpha)$. Curves are plotted as a function of the number of gradient-descent iterations, with the y-axis in log-scale. Via noisy training, the noiseless cost functions go down to $\sim 10^{-4}$. Initially we set $N_{min} = 2$, and the thick dashed vertical lines shows the point where we increased this value to $N_{min} = 250$. Increasing the minimum number of shots iCANS employs to compute each partial derivative leads to smaller cost function values in both cases.

noisy optimization. These results allow us to verify if the parameters obtained from the noisy optimization are indeed minimizing the noiseless cost functions. Before discussing the results, we first give details for our ansatzes and optimization methods.

6.1. Ansatzes and optimization methods
As previously mentioned, to implement VQC we consider two ansatzes for the trainable unitary $V(\alpha)$. The building block of our ansatzes is a dressed CNOT gate, which is a two-qubit gate composed of a CNOT preceded

and followed by single-qubit gates $V_k(\boldsymbol{\alpha}_k)$ acting on each qubit, as shown in figure 6(a). Each single-qubit gate $V_k(\boldsymbol{\alpha}_k)$ is decomposed (up to a global phase) into three elementary rotations parameterized by three angles in the vector $\boldsymbol{\alpha}_k = (\alpha_{k,1}, \alpha_{k,2}, \alpha_{k,3})$ as

$$V_k(\boldsymbol{\alpha}_k) = \mathrm{e}^{-\mathrm{i}\alpha_{k,3}\sigma_z/2}\mathrm{e}^{-\mathrm{i}\alpha_{k,2}\sigma_y/2}\mathrm{e}^{-\mathrm{i}\alpha_{k,1}\sigma_z/2}. \tag{21}$$

Let us now introduce our ansatzes. We note that our two ansatzes are fairly similar to the ones introduced in [19]. In our first ansatz, each layer is composed of n dressed CNOTs, where n is the number of qubits (in the special case of $n = 2$ each layer consists of one dressed CNOT), with the precise structure defined as follows.

Definition 10. We define the alternating-pair ansatz as a layered ansatz in which each layer consists of (parameterized) dressed CNOT gates acting on alternating pairs of neighboring qubits as illustrated in figure 6(b).

We remark that it is useful to distinguish between a *complete* ansatz, in which an exact compilation for U is contained inside the ansatz, versus an *incomplete* ansatz, where exact compilation is not possible. In general, a small number of layers can lead to an incomplete ansatz, where one can only reach approximate compilation. Hence, increasing the number of layers l could allow one to obtain better compilations of U. Note however that while a large number of layers can achieve a complete ansatz, it can also be harder to train and can lead to a longer-depth circuit.

The alternating-pair ansatz may not lead to the optimal depth compilation for U, particularly in the complete ansatz case. Our second ansatz attempts to fix the issue of introducing unnecessary depth by having a structure that depends on U.

Definition 11. We construct the target-inspired ansatz by taking the gate sequence for the target unitary U, expanding this gate sequence into single-qubit gates and CNOTs, removing all single-qubit gates that precede or follow a CNOT, and replacing each remaining CNOT in the gate sequence with a (parameterized) dressed CNOT. Finally, each remaining single-qubit gate is replaced by a parametrized single-qubit gate.

As schematically depicted in figure 6(c), each layer is now composed of one dressed CNOT. This ansatz will always be complete since its structure is inspired by U. While this ansatz is not useful to compress the number of CNOTs in $V(\boldsymbol{\alpha})$, it is useful as a proof-of-concept to demonstrate OPR for complete ansatzes. We remark that a simple modification of this ansatz, where the placements of the dressed CNOTs are optimized over instead of fixed, would actually be useful for circuit-depth compression. Furthermore, we have implemented this dressed CNOT placement optimization, and we find that we obtain similar noise resilience results as those for the target-inspired ansatz.

Let us now discuss the optimization methods. As previously mentioned, the trainable gate sequence $V(\boldsymbol{\alpha})$ is a function of a set of parameters $\boldsymbol{\alpha}$ corresponding to the collection of the internal gate angles in each dressed CNOT. To optimize these parameters, we employ a gradient-descent approach. This approach exploits the fact that the gradient with respect to $\boldsymbol{\alpha}$ of C_{HST}, C_{LHST}, C_{LET}, and C_{LLET} can be computed by using the circuits for HST, LHST, LET, and LLET, respectively [43, 19]. We remark that we used different gradient-based approaches for the shallow and deep ansatz cases, since the latter requires a more sophisticated and efficient optimizer.

Specifically, for the shallow ansatz cases where there are few parameters, we employ the simple gradient-based approach outlined in [19, appendix 4] . In this approach, the number of shots N per iteration is fixed. (We choose $N = 50\,000$.) On the other hand, for deep ansatzes with larger numbers of parameters, we employ a more sophisticated gradient-based approach that improves efficiency by reducing the number of shots required [44]. This approach is the individual coupled adaptive number of shots (iCANS) algorithm of [44], which is a measurement-frugal method that often outperforms other optimizers in the presence of noise. The iCANS optimizer frugally adjusts the number of shots both for a given iteration and for a given partial derivative in a stochastic gradient descent. When employing iCANS, one sets as input: (1) the total number of shots employed during the optimization, and (2) the minimum number of shots (denoted $N_{\min}$) employed to estimate the gradient for a given iteration. We set the latter to initially be $N_{\min} = 2$ and then later increase this to $N_{\min} = 250$, which empirically leads to good convergence.

6.2. Toffoli gate
The top panels in figure 7 show results of implementing VQC for the Toffoli gate. Figure 7 (top, a) corresponds to $V(\boldsymbol{\alpha})$ being given by a single layer of the alternating-pair ansatz of definition 10. Here, the noisy cost functions $\widetilde{C}_{\mathrm{HST}}$ and $\widetilde{C}_{\mathrm{LHST}}$ (blue and red curve, respectively) tend to decrease as the number of iterations increases and converge to non-zero values. We remark that the number of iterations can be different for $\widetilde{C}_{\mathrm{HST}}$

and $\widetilde{C}_{\text{LHST}}$ since the termination condition of the optimization algorithm can be reached for a different number of iterations.

Figure 7 (top, a) also depicts the cost functions C_{HST} and C_{LHST} evaluated for the variational parameters $\boldsymbol{\alpha}$ obtained from the noisy optimization (green and pink curve, respectively). These curves show that as the number of iterations increases, both C_{HST} and C_{LHST} tend to decrease too, indicating that the noisy training is *indirectly training the noiseless cost functions*, i.e. the adjustments to the parameters $\boldsymbol{\alpha}$ made by noisy training are reducing the noiseless cost functions. Note that C_{HST} and C_{LHST} do not converge to zero since a single layer of three dressed CNOTs forms an incomplete ansatz for the Toffoli gate.

In order to determine if the algorithm is reaching the minimum value achievable with just one layer, we have also implemented VQC to compile the Toffoli gate in a noise-free simulation. The minimum values achieved for C_{HST} and C_{LHST} are shown as a blue and red dashed curve, respectively. Surprisingly, the cost functions evaluated with the parameters from the noisy training (green and pink curves) converge to the dashed lines. This suggests that the optimal parameters are noise resilient since noisy training reaches the minimum value obtained by noise-free training. As a caveat, however, we note that it is not clear whether the minima reached are global or local optima.

Figure 7 (top, b) plots the VQC results for Toffoli with $V(\boldsymbol{\alpha})$ given by two layers of the alternating-pair ansatz. In this case, C_{HST} and C_{LHST} converge to values which are smaller than the ones obtained in the one-layer case. The latter indicates that two layers allow for a more complete compilation of the Toffoli gate, albeit it appears that the ansatz is not yet complete. Note that both the decomposition of the Toffoli gate in figure 5, as well as two layers of the alternating-pair ansatz, consist of six CNOTs. However, the placement of the dressed CNOTs does not seem to be optimal. Finally, let us remark that the green and pink curves converge to the dashed blue and red lines, respectively. Hence, this once again shows that the optimal parameters are noise resilient. Similar to the previous case, it is not clear whether the minima reached are global or local minima.

Figure 7 (top, c) shows results for the target-inspired ansatz of definition 11. As the number of iterations increases, all curves tend to decrease, with the green and pink curves converging to values of the order of 10^{-4}. We remark that we have verified that $W = V^{\dagger}U \approx \mathbb{1}$ for the parameters obtained. In this case, we do not plot dashed blue and red curves since the ansatz is complete and the minimum of the noiseless cost functions is zero.

These results indicate that optimizing $V(\boldsymbol{\alpha})$ in the presence of noise yields the correct variational parameters $\boldsymbol{\alpha}$, which minimize the noiseless cost function. Hence, both C_{HST} and C_{LHST} appear to exhibit OPR for the realistic noise model considered.

6.3. Quantum Fourier transform

We now discuss the VQC results for the three-qubit QFT. Figure 7 shows the results for $V(\boldsymbol{\alpha})$ consisting of: a single layer of the alternating-pair ansatz of definition 10 (bottom, a), two layers of the alternating-pair ansatz (bottom, b), and the target-inspired ansatz of definition 11 (bottom, c). As shown in these plots, most of the results for QFT are similar to the results for the Toffoli gate. In all cases the noiseless cost functions tended to decrease with iterations, indicating that noisy training indirectly trains the noiseless costs.

For the one-layer case of figure 7 (bottom, a) the green and pink curves (noiseless cost functions evaluated at the parameters obtained from noisy training) converge to the value obtained by training in a noise-free environment (dashed curve). Here, the non-zero value of the dashed curve indicates that a one-layer ansatz is incomplete. This is in contrast to figure 7 (bottom, b), where the dashed red line of C_{LHST} is of the order of 10^{-4}, implying that the ansatz is complete. Once again, in figure 7 (bottom, b), the green and pink curves approximately converge to the dashed lines (noiseless training), indicating noise resilience. Finally, figure 7 (bottom, c), shows that that both C_{HST} and C_{LHST} appear to exhibit OPR, as we can indirectly train the parameters in $V(\boldsymbol{\alpha})$ in the presence of noise.

6.4. W-state preparation

Finally, we discuss the results of implementing of VQC for both FUMC and FISC of a W-state preparation circuit. We remark here that we did not perform FISC for the Toffoli gate and the QFT since those unitaries act trivially on the $|0\rangle$ state. Moreover, we are only interested in comparing the FUMC and the FISC approach with a complete ansatz, meaning that we only considered the target-inspired ansatz of definition 11.

As shown in figure 8, all cost functions C_{HST}, C_{LHST}, C_{LET}, and C_{LLET} can be optimized indirectly via noisy training of $V(\boldsymbol{\alpha})$. Both for FUMC and FISC the cost functions go down to $\sim 10^{-4}$, while for FUMC one can even reach values of $\sim 10^{-5}$ when employing the LHST. Hence, our numerics indicate that C_{HST}, C_{LHST}, C_{LET}, and C_{LLET} appear to exhibit OPR to IBM's realistic noise model.

7. Discussion

7.1. VQC in the NISQ era
Our analytical and numerical results suggest that variational quantum compiling (VQC) could be a useful tool for near-term noisy quantum computing. While there are several intended uses for VQC [19], the main purpose is for circuit-depth compression of quantum algorithms. This depth compression arises because VQC could achieve optimal compiling, whereas classical methods for quantum compiling either scale exponentially (if they are aiming at optimal compiling) or are sub-optimal when they are restricted to local (instead of global) compiling of the circuit.

Suppose one is able to achieve depth compression with VQC. This implies that the target unitary U has a longer depth than the trained gate sequence $V(\alpha)$. Prior to our work, one may have been concerned that this depth compression might not reduce noise, because perhaps the noise occurring during U is somehow compiled into the gate sequence $V(\alpha)$. However, our work shows that this is not the case. Despite various sources of incoherent noise (e.g. see the noise model in figure 3), we find that one learns the correct optimal parameters α for $V(\alpha)$. This means that, after performing VQC, if one was to implement the gate sequence $V(\alpha)$ instead of U, then one should see that $V(\alpha)$ really does achieve less noise than U, since the depth of $V(\alpha)$ is shorter.

7.2. Summary of results
In this work, we treated two different forms of VQC: Full Unitary Matrix Compiling (FUMC) and Fixed Input State Compiling (FISC). Our main analytical results were stated in theorems 1–3. We found that both FUMC and FISC are resilient to measurement noise. In addition, they are both resilient to global depolarizing noise acting continuously throughout the circuit and global Pauli noise occurring just prior to the implementation of $W = V^{\dagger}U$.

For FUMC, we were able to prove resilience to additional sources of noise, such as Pauli gate noise during the entangling and disentangling gates as well as non-unital Pauli noise occurring at particular times in the circuit. The fact that our noise resilience results are more extensive for FUMC than for FISC may simply be due to the fact that the cost-evaluation circuit for FUMC is more complicated than that for FISC. Hence it is possible that this additional resilience is needed to make the two approaches have similar levels of noise resilience. Alternatively, it could be possible that either FUMC or FISC is more noise resilient than the other, although this remains to be established. (Note that our numerics did not see a significant difference in the noise resilience of FUMC versus FISC.)

In addition, Corollaries 1–8 stated resilience results for noise models that go beyond the noise models considered in theorems 1–3, at the expense of possibly specializing the form of the unitary $W = V^{\dagger}U$ (for example, to Clifford unitaries or tensor-product unitaries). In particular, these corollaries considered noise that occurs during the implementation of W, which is certainly practically relevant.

Our numerical results were presented in figures 7–8. Generally speaking, these numerics agreed with our theoretical expectations and hinted at resilience beyond what is stated in our theorems, which we discuss in the next subsection. We emphasize that our implementations employed the noise model of IBM's 14-qubit Melbourne device, and hence this shows that VQC exhibits resilience for currently available hardware.

7.3. Noise resilience beyond our theorems
There are two senses in which VQC might exhibit resilience beyond the results stated in our theorems. The first sense is that VQC may be resilient to more general noise models than the ones we considered. The second sense is that VQC may be resilient even for the incomplete ansatz case, on which we elaborate below. Both of these possibilities appear to be supported by our numerical implementations.

For evidence supporting the idea that VQC may be resilient to more general noise models, consider the following. The noise model associated with IBM's 14-qubit Melbourne device is more general than the noise models depicted in figures 3 and 4, and the unitaries we considered in figure 5 do not fall into the special cases (e.g. Clifford or tensor product) treated by Corollaries 1–8. For example, IBM's noise model has non-unital Pauli noise associated with each gate and hence occurring throughout the implementation of $W = V^{\dagger}U$. Thus, our theorems and corollaries do not cover all of noise processes occurring in IBM's noise model. Despite this, we were able to reduce the noiseless cost (via noisy training) to $\sim 10^{-4}$ for the Toffoli gate (figure 7 (top, c)) and QFT (figure 7 (bottom, c)), and to $\sim 10^{-5}$ for W state preparation (figure 8).

Naturally, our theorems and corollaries have a bias towards noise models that are mathematically easy to work with, such as Pauli noise or depolarizing noise, since this makes it easier to formulate proofs. It is therefore important for future work to attempt to show resilience beyond these noise models.

As noted above, VQC may also have resilience beyond the complete ansatz case. Recall that we say an ansatz for $V(\alpha)$ is complete (incomplete) if it contains (does not contain) an exact compilation of U. Our theorems and corollaries are restricted to the complete ansatz case, whereas our numerics in figure 7 also consider the

incomplete ansatz case. Interestingly, figure 7 showed that typically one can obtain the same value for the noiseless cost with either noisy or noiseless training. This surprising result suggests that perhaps the optimal values for $\boldsymbol{\alpha}$ may be resilient to noise even for the incomplete ansatz case, and future work should investigate this possibility.

In addition, it will be important to investigate the effect of noise on the parameter landscape and parameter trainability (e.g. [45]). Our work indicates that the global optimum of VQC may not change with noise, but does not address the difficulty of finding this optimum.

7.4. Coherent versus incoherent noise

In the Introduction, we emphasized the distinction between OPR and cost value resilience [7]. The latter is relevant to coherent noise, whereas OPR is relevant to incoherent noise. Intuitively, we anticipate that coherent noise (e.g. systematic gate biases) in VQC will often shift the location of the global minimum in parameter space, and hence we expect coherent noise to have a non-trivial effect on the optimal parameters in VQC. Because of this intuition, we have focused our paper and our definition of OPR solely on incoherent noise. We remark that our definition of OPR, which is stated in terms of unitaries (rather than parameters), would need to be modified if one is interested in studying parameter resilience for coherent noise. However, as noted, we do not anticipate resilience to coherent noise to hold. We also remark that other strategies exist to correct coherent noise [46]. Nevertheless, an interesting question for future work will be see whether OPR holds partially whenever both coherent and incoherent noise are present. In addition, it will be interesting to combine the ideas of OPR and cost value resilience into a single framework.

7.5. Noise resilience of VQE

Finally, let us consider VHQCAs more generally. In particular, let us revisit the variational quantum eigensolver (VQE) that we discussed in section 2. As we now show, VQC is a special case of VQE. This idea was noted for FISC in [20]. However, the argument is more subtle for the FUMC case.

The key observation is that the various cost functions can be rewritten as the expectation values for some effective Hamiltonians:

$$C_{\mathsf{LET}} = \langle \psi(\boldsymbol{\alpha})|H_{\mathsf{LET}}|\psi(\boldsymbol{\alpha})\rangle, \quad C_{\mathsf{LLET}} = \langle \psi(\boldsymbol{\alpha})|H_{\mathsf{LLET}}|\psi(\boldsymbol{\alpha})\rangle,$$
$$C_{\mathsf{HST}} = \langle \chi(\boldsymbol{\alpha})|H_{\mathsf{HST}}|\chi(\boldsymbol{\alpha})\rangle, \quad C_{\mathsf{LHST}} = \langle \chi(\boldsymbol{\alpha})|H_{\mathsf{LHST}}|\chi(\boldsymbol{\alpha})\rangle. \tag{22}$$

Here $|\psi(\boldsymbol{\alpha})\rangle \in \mathcal{H}^A$ and $|\chi(\boldsymbol{\alpha})\rangle \in \mathcal{H}^{AB}$ are n-qubit and $2n$-qubit states, respectively, given by

$$|\psi(\boldsymbol{\alpha})\rangle = V(\boldsymbol{\alpha})|\mathbf{0}\rangle, \qquad |\chi(\boldsymbol{\alpha})\rangle = (V(\boldsymbol{\alpha}) \otimes \mathbb{1}^B)|\Phi\rangle, \tag{23}$$

where $\mathcal{H}^X$ denotes the Hilbert space of system X, and $|\Phi\rangle = E|\mathbf{0}\rangle$ is the standard maximally entangled state on AB. We remark that $|\chi(\boldsymbol{\alpha})\rangle$ is simply the Choi state associated with $V(\boldsymbol{\alpha})$.

For the cost functions associated with FISC, the effective Hamiltonians are given by

$$H_{\mathsf{LET}} = \mathbb{1}^A - U|\mathbf{0}\rangle\langle\mathbf{0}|U^\dagger, \qquad H_{\mathsf{LLET}} = \mathbb{1}^A - \frac{1}{n}\sum_{j=1}^{n} U(P_0^{A_j} \otimes \mathbb{1}^{\overline{A}_j})U^\dagger, \tag{24}$$

where $P_0^{A_j}$ is the projector onto the zero state of A_j. For the cost functions associated with FUMC, the effective Hamiltonians are given by

$$H_{\mathsf{HST}} = \mathbb{1}^{AB} - (U \otimes \mathbb{1}^B)|\Phi\rangle\langle\Phi|(U^\dagger \otimes \mathbb{1}^B),$$
$$H_{\mathsf{LHST}} = \mathbb{1}^{AB} - \frac{1}{n}\sum_{j=1}^{n}(U \otimes \mathbb{1}^B)(|\Phi^{(j)}\rangle\langle\Phi^{(j)}| \otimes \mathbb{1}^{\overline{A}_j\overline{B}_j})(U^\dagger \otimes \mathbb{1}^B), \tag{25}$$

where $|\Phi^{(j)}\rangle$ is the standard maximally entangled state on A_jB_j. With these Hamiltonians, one can verify that the expressions in (22) are equal to the original cost function definitions in section 3. Hence, we have just shown that VQC is a special case of VQE, where the goal is to prepare the ground state of one of the Hamiltonians in (24) or (25).

The fact that VQC is a special case of VQE implies that, for specific Hamiltonians, VQE is noise resilient. Namely, we have shown that VQE exhibits OPR when the Hamiltonian has the form in either (24) or (25). This naturally points to the question of whether VQE is resilient more generally. It is therefore a very interesting direction for future research to extend our noise resilience to Hamiltonians other than the ones we considered.

8. Conclusions

In this work, we discovered a novel kind of noise resilience for variational hybrid quantum-classical algorithms (VHQCAs). We introduced the idea of optimal parameter resilience (OPR), where the variational parameters corresponding to the global optimum are unaffected by various types of incoherent noise. We showed that

variational quantum compiling (VQC) exhibits OPR. This paves the way for VQC to be used in the era of noisy intermediate-scale quantum computing as a tool for circuit-depth compression. Important future research directions include: (1) extending our theorems to show resilience to more general noise models than the ones we considered (which our numerics suggest may be possible), (2) exploring noise resilience for the incomplete ansatz case (which our numerics indicate may also be resilient), (3) analyzing approximate noise resilience, (4) studying the effect of noise on the parameter training process, and (5) generalizing our resilience results to other Hamiltonians for the variational quantum eigensolver and exploring resilience for other VHQCAs (for example, some evidence of noise resilience was recently reported in [47]).

Acknowledgments

Appendix A. Preliminaries

The main goal of the appendix is to provide the proofs of theorems 1–3 and Corollaries 1–8. For these proofs, we will need to first review some definitions and properties. We point readers to [33, 34] for additional background.

Pauli Basis. In our proofs, we will work in the Pauli product basis, involving a tensor product of one-qubit Pauli operators. This is a natural basis to choose, given the qubit structure of quantum computers. Let

$$X^{\boldsymbol{l}} := \sigma_x^{l_1} \otimes \sigma_x^{l_2} \otimes \cdots \otimes \sigma_x^{l_n}, \qquad Z^{\boldsymbol{k}} := \sigma_z^{k_1} \otimes \sigma_z^{k_2} \otimes \cdots \otimes \sigma_z^{k_n}, \tag{A1}$$

where $l_1, l_2, \ldots, l_n \in \{0, 1\}$, $k_1, k_2, \ldots, k_n \in \{0, 1\}$, $\boldsymbol{l} = (l_1, \ldots, l_n)$, and $\boldsymbol{k} = (k_1, \ldots k_n)$. The following properties are satisfied by the Pauli operators:

$$X^{\boldsymbol{l}_1} X^{\boldsymbol{l}_2} = X^{\boldsymbol{l}_1 \oplus \boldsymbol{l}_2}, \quad Z^{\boldsymbol{k}_1} Z^{\boldsymbol{k}_2} = Z^{\boldsymbol{k}_1 \oplus \boldsymbol{k}_2}, \quad X^{\boldsymbol{l}} Z^{\boldsymbol{k}} = (-1)^{\boldsymbol{l} \cdot \boldsymbol{k}} Z^{\boldsymbol{k}} X^{\boldsymbol{l}}, \quad \mathrm{Tr}[X^{\boldsymbol{l}} Z^{\boldsymbol{k}}] = 2^n \delta_{\boldsymbol{l},\boldsymbol{0}}, \delta_{\boldsymbol{k},\boldsymbol{0}}, \tag{A2}$$

which follow from the properties of the single-qubit Pauli operators.

Pauli group. The Pauli group of n qubits is $\mathbb{G}_n := \{\pm 1, \pm i\} \times \{I, \sigma_x, \sigma_y, \sigma_z\}^{\otimes n}$.

Clifford group. The Clifford group on n qubits is the set of unitaries that normalize the Pauli group, i.e.

$$\mathbb{C}_n := \{U : U\mathbb{G}_n U^\dagger \in \mathbb{G}_n\}. \tag{A3}$$

Maximally entangled states. In what follows, we consider the following maximally entangled states $|\Phi^+\rangle\langle\Phi^+| = |\phi^+\rangle\langle\phi^+|^{\otimes n}$, where $|\phi^+\rangle = (|0, 0\rangle + |1, 1\rangle)/\sqrt{2}$. The aforementioned tensor product of maximally entangled states can be written in the Pauli basis as follows:

$$|\Phi^+\rangle\langle\Phi^+|_{AB} = \frac{1}{2^{2n}} \sum_{\boldsymbol{l},\boldsymbol{k}} X_A^{\boldsymbol{l}} Z_A^{\boldsymbol{k}} \otimes X_B^{\boldsymbol{l}} Z_B^{\boldsymbol{k}} = \frac{1}{2^{2n}} \sum_{\boldsymbol{l},\boldsymbol{k}} Z_A^{\boldsymbol{k}} X_A^{\boldsymbol{l}} \otimes Z_B^{\boldsymbol{k}} X_B^{\boldsymbol{l}}. \tag{A4}$$

All-zero state. Noting that $|0\rangle\langle 0| = (\mathbb{1} + \sigma_z)/2$, then in the Pauli basis the all-zero state $|\boldsymbol{0}\rangle\langle\boldsymbol{0}| = |0\rangle\langle 0|^{\otimes n}$ is

$$|\boldsymbol{0}\rangle\langle\boldsymbol{0}| = \frac{1}{2^n}(\mathbb{1} + \sigma_z)^{\otimes n} = \frac{1}{2^n} \sum_{\boldsymbol{l}} Z^{\boldsymbol{l}}. \tag{A5}$$

Pauli channels. A Pauli noise channel corresponds to the action of random Pauli operators on a quantum state ρ according to a probability distribution. Let $\mathcal{P}^A$ denote an n-qubit Pauli channel acting on system $A = A_1, \ldots A_n$. Then the action of $\mathcal{P}^A$ on the state ρ is given by

$$\mathcal{P}^A(\rho) = \sum_{\boldsymbol{l},\boldsymbol{k}} p_{\boldsymbol{l},\boldsymbol{k}}^A X_A^{\boldsymbol{l}} Z_A^{\boldsymbol{k}} \rho (X_A^{\boldsymbol{l}} Z_A^{\boldsymbol{k}})^\dagger, \tag{A6}$$

where $0 \leqslant p_{\boldsymbol{l},\boldsymbol{k}}^A \leqslant 1$, and $\sum_{\boldsymbol{l},\boldsymbol{k}} p_{\boldsymbol{l},\boldsymbol{k}}^A = 1$. Using the properties in (A2), we find that

$$\mathcal{P}^A(X_A^{\boldsymbol{a}} Z_A^{\boldsymbol{b}}) = \sum_{\boldsymbol{l},\boldsymbol{k}} p_{\boldsymbol{l},\boldsymbol{k}}^A X_A^{\boldsymbol{l}} Z_A^{\boldsymbol{k}} X_A^{\boldsymbol{a}} Z_A^{\boldsymbol{b}} Z_A^{\boldsymbol{k}} X_A^{\boldsymbol{l}} = \sum_{\boldsymbol{l},\boldsymbol{k}} (-1)^{\boldsymbol{a} \cdot \boldsymbol{k}} (-1)^{\boldsymbol{b} \cdot \boldsymbol{l}} p_{\boldsymbol{l},\boldsymbol{k}}^A X_A^{\boldsymbol{a}} Z_A^{\boldsymbol{b}} = p_{\boldsymbol{a},\boldsymbol{b}}^A X_A^{\boldsymbol{a}} Z_A^{\boldsymbol{b}}, \tag{A7}$$

where $p_{\boldsymbol{a},\boldsymbol{b}}^A := \sum_{\boldsymbol{l},\boldsymbol{k}} (-1)^{\boldsymbol{a} \cdot \boldsymbol{k}} (-1)^{\boldsymbol{b} \cdot \boldsymbol{l}} p_{\boldsymbol{l},\boldsymbol{k}}^A$ and $-1 \leqslant p_{\boldsymbol{a},\boldsymbol{b}}^A \leqslant 1$ for all $\boldsymbol{a}, \boldsymbol{b} \in \{0, 1\}^n$. Similarly, the action of a global Pauli channel $\mathcal{P}^{AB}$ acting on systems $A = A_1 \cdots A_n$ and $B = B_1 \cdots B_n$, respectively, is defined as

$$\mathcal{P}^{AB}(X_A^{\boldsymbol{a_1}} Z_A^{\boldsymbol{b_1}} \otimes X_B^{\boldsymbol{a_2}} Z_B^{\boldsymbol{b_2}}) = p_{\boldsymbol{a_1},\boldsymbol{a_2},\boldsymbol{b_1},\boldsymbol{b_2}}^{AB} X_A^{\boldsymbol{a_1}} Z_A^{\boldsymbol{b_1}} \otimes X_B^{\boldsymbol{a_2}} Z_B^{\boldsymbol{b_2}}. \tag{A8}$$

Non-unital Pauli noise channels. The action of a non-unital Pauli channel $\mathcal{P}_{\mathrm{NU}}$ on an n-qubit Pauli operators is

$$\mathcal{P}_{\mathrm{NU}}(X^{\boldsymbol{a}} Z^{\boldsymbol{b}}) = c_{\boldsymbol{a},\boldsymbol{b}} X^{\boldsymbol{a}} Z^{\boldsymbol{b}} \quad \forall \; \boldsymbol{a} \neq \boldsymbol{0}, \; \boldsymbol{b} \neq \boldsymbol{0}, \tag{A9}$$

$$\mathcal{P}_{\mathrm{NU}}(X^{\boldsymbol{0}} Z^{\boldsymbol{0}}) = \mathcal{P}_{\mathrm{NU}}(\mathbb{1}) = \mathbb{1} + \sum_{(\boldsymbol{a},\boldsymbol{b}) \neq (\boldsymbol{0},\boldsymbol{0})} d_{\boldsymbol{a},\boldsymbol{b}} X^{\boldsymbol{a}} Z^{\boldsymbol{b}}. \tag{A10}$$

We now prove the following lemma based on Clifford unitaries and Pauli channels.

Lemma 1. *Let W be a Clifford unitary and let $\mathcal{P}$ be a Pauli channel. Then for any state ρ, the following holds:*

$$(\mathcal{W} \circ \mathcal{P})(\rho) = (\mathcal{Q} \circ \mathcal{W})(\rho), \tag{A11}$$

where $\mathcal{Q}$ is another Pauli channel.

Proof. From (A6) it follows that

$$\mathcal{W} \circ \mathcal{P}(\rho) = W\left(\sum_{l,k} p_{l,k} X^l Z^k \rho Z^k X^l\right) W^\dagger = \sum_{l,k} p_{l,k} (W X^l Z^k W^\dagger)(W \rho W^\dagger)(W Z^k X^l W^\dagger) \tag{A12}$$

$$= \sum_{l,k} p_{l,k} X^{\boldsymbol{m}(l,k)} Z^{\boldsymbol{n}(l,k)} W \rho W^\dagger Z^{\boldsymbol{n}(l,k)} X^{\boldsymbol{m}(l,k)} \tag{A13}$$

$$= (\mathcal{Q} \circ \mathcal{W})(\rho). \tag{A14}$$

The third equality follows from the definition of a Clifford unitary (A3), while the last equality follows from (A6). □

Appendix B. Noisy entangling and disentangling gates in FUMC

For the proofs given in appendices D–G, we will make use of some properties of the noisy versions of entangling E and disentangling $E^\dagger$ gates that appear in FUMC. Hence, it is helpful to first state these properties in this appendix. Recall that, for Pauli gate noise acting during E or $E^\dagger$, we assume that global Pauli channels act before and after each Hadamard, as well as before and after each CNOT. This noise model incorporates the case when there could be correlated Pauli noise acting on different qubits during E and $E^\dagger$. We note that the noisy entangling gate is the same for both the HST and the LHST.

Let $E = E^{AB}$ denote the ideal entangling gate, which can be split into a tensor product of two qubit entangling gates $E^{A_j B_j}$ as

$$E^{AB} = E^{A_1 B_1} \otimes E^{A_2 B_2} \otimes \cdots \otimes E^{A_n B_n} = \bigotimes_{j=1}^{n} E^{A_j B_j}. \tag{B1}$$

Moreover, each $E^{A_j B_j}$ consists of a Hadamard gate acting on A_j followed by a CNOT gate acting on both A_j and B_j. In the quantum channel notation we write this as $\mathcal{E}^{A_j B_j} = \mathcal{C}_X^{A_j B_j} \circ (\mathcal{H}^{A_j} \otimes \mathcal{I}^{B_j})$, where $\mathcal{H}^{A_j}$ are the quantum channels that implement the Hadamard gates and $\mathcal{C}_X^{A_j B_j}$ are the quantum channels that implement the CNOTs. The noisy version of $\mathcal{E}^{AB}$, which we denote by $\widetilde{\mathcal{E}}^{AB}$, is

$$\widetilde{\mathcal{E}}^{AB} := \bigotimes_{j=1}^{n} \mathcal{R}_j^{AB} \circ \mathcal{C}_X^{A_j B_j} \circ \bigotimes_{j=1}^{n} \mathcal{Q}_j^{AB} \circ (\mathcal{H}^{A_j} \otimes \mathcal{I}^{B_j}) \circ \mathcal{P}_j^{AB}, \tag{B2}$$

where $\mathcal{P}_j^{AB}, \mathcal{Q}_j^{AB}$, and $\mathcal{R}_j^{AB}$ are $2n$-qubit global Pauli channels for all $i \in \{1,\ldots,n\}$, as defined in (A8). Since both Hadamard and CNOT gates are Clifford unitaries, by using lemma 1 we find that

$$\widetilde{\mathcal{E}}^{AB} := \mathcal{M}^{AB} \circ \bigotimes_{j=1}^{n} \mathcal{C}_X^{A_j B_j} \circ \bigotimes_{j=1}^{n} (\mathcal{H}^{A_j} \otimes \mathcal{I}^{B_j}), \tag{B3}$$

where $\mathcal{M}^{AB}$ is another Pauli channel.

We now apply $\widetilde{\mathcal{E}}^{AB}$ on the all-zeros state $|\boldsymbol{0}, \boldsymbol{0}\rangle\langle \boldsymbol{0}, \boldsymbol{0}|^{AB}$. Consider the following chain of equalities:

$$\widetilde{\mathcal{E}}^{AB}(|\boldsymbol{0}, \boldsymbol{0}\rangle\langle \boldsymbol{0}, \boldsymbol{0}|^{AB}) = \widetilde{\mathcal{E}}^{AB}\left(\frac{1}{2^{2n}}\sum_{\boldsymbol{a},\boldsymbol{b}} Z_A^{\boldsymbol{a}} \otimes Z_B^{\boldsymbol{b}}\right) = \frac{1}{2^{2n}}\sum_{\boldsymbol{a},\boldsymbol{b}} m_{\boldsymbol{a},\boldsymbol{a},\boldsymbol{b},\boldsymbol{b}}^{AB} X_A^{\boldsymbol{a}} Z_A^{\boldsymbol{b}} \otimes X_B^{\boldsymbol{a}} Z_B^{\boldsymbol{b}}, \tag{B4}$$

where we used (A5), (A8), and the following identities for all $j \in \{1, \ldots n\}$:

$$(\mathcal{H}^{A_j} \otimes \mathcal{I}^{B_j})(Z_{A_j}^{a_j} \otimes Z_{B_j}^{b_j}) = X_{A_j}^{a_j} \otimes Z_{B_j}^{b_j}, \quad (\mathcal{C}_X^{A_j B_j})(X_{A_j}^{a_j} \otimes \mathbb{1}_B)$$

$$= X_{A_j}^{a_j} \otimes X_{B_j}^{a_j}, \quad (\mathcal{C}_X^{A_j B_j})(\mathbb{1}_{A_j} \otimes Z_{B_j}^{b_j}) = Z_{A_j}^{b_j} \otimes Z_{B_j}^{b_j}. \tag{B5}$$

The noisy disentangling channel for the HST is given by the adjoint of the noisy entangling channel, as defined in (B2). On the other hand, since in the LHST only two qubits $A_j B_j$ are measured for a given run of the experiment, the disentangling channel is applied only on the $A_j B_j$ pair. However, we assume that global Pauli channels act on $2n$ qubits before and after the Hadamard and CNOT gate. For each $j \in \{1, \ldots, n\}$, the disentangling channel is given by the adjoint of the following channel:

$$\widetilde{\mathcal{E}}_j'^{AB} := \mathcal{R}_j^{AB} \circ (\mathcal{C}_X^{A_j B_j} \otimes \mathcal{I}^{\overline{A_j} \overline{B_j}}) \circ \mathcal{Q}_j^{AB} \circ (\mathcal{H}^{A_j} \otimes \mathcal{I}^{B_j} \otimes \mathcal{I}^{\overline{A_j} \overline{B_j}}) \circ \mathcal{P}_j^{AB}, \tag{B6}$$

$$= \mathcal{M}_j^{AB} \circ (\mathcal{C}_X^{A_j B_j} \otimes \mathcal{I}^{\overline{A_j} \overline{B_j}}) \circ (\mathcal{H}^{A_j} \otimes \mathcal{I}^{B_j} \otimes \mathcal{I}^{\overline{A_j} \overline{B_j}}), \tag{B7}$$

where $\mathcal{P}_j^{AB}$, $\mathcal{Q}_j^{AB}$, $\mathcal{R}_j^{AB}$, and $\mathcal{M}_j^{AB}$ are $2n$-qubit global Pauli channels, as defined in (A8), and we used lemma 1. We remark that the Pauli channels are defined with a j subscript in (B7) to emphasize that for different runs of the experiment the Pauli channels that act could be different.

From arguments similar to those used to derive (B4), we find that

$$\widetilde{\mathcal{E}}_j'^{AB}(|0, 0\rangle\langle 0, 0|^{A_j B_j} \otimes \mathbb{1}_{\overline{A_j} \overline{B_j}}) = \frac{1}{2^2} \sum_{a_j, b_j = 0}^{1} m_{a_j, a_j, b_j, b_j}^{AB} (X_{A_j}^{a_j} Z_{A_j}^{b_j} \otimes X_{B_j}^{a_j} Z_{B_j}^{b_j} \otimes \mathbb{1}_{\overline{A_j} \overline{B_j}}). \tag{B8}$$

Appendix C. Measurement noise in FUMC

For the proofs given in appendices D–G, we will make use of some properties of measurement noise in FUMC. Hence, it is helpful to first state these properties in this appendix.

Let P_0 denote the POVM element associated with getting the all-zeros outcome in the noiseless HST, which can be expressed as $P_0 := |\mathbf{0}\rangle\langle\mathbf{0}| = \bigotimes_{j=1}^{2n} |0\rangle\langle 0|$. We consider the measurement noise as follows. For each qubit j, where $j \in \{1, \ldots, 2n\}$, the ideal projector $|0\rangle\langle 0|$ gets replaced by $p_{00}^{(j)} |0\rangle\langle 0| + p_{01}^{(j)} |1\rangle\langle 1|$. Moreover, we assume that for all j the following strict inequality holds: $p_{00}^{(j)} > p_{01}^{(j)}$.

Let $\widetilde{P}_0$ denote the noisy POVM element. Then the following equalities hold:

$$\widetilde{P}_0 = \bigotimes_{j=1}^{n} (p_{00}^{A_j} |0\rangle\langle 0|^{A_j} + p_{01}^{A_j} |1\rangle\langle 1|^{A_j}) \otimes \bigotimes_{j=1}^{n} (p_{00}^{B_j} |0\rangle\langle 0|^{B_j} + p_{01}^{B_j} |1\rangle\langle 1|^{B_j}) \tag{C1}$$

$$= \sum_{\mathbf{a}, \mathbf{b}} p^A(\mathbf{a}) p^B(\mathbf{b}) |\mathbf{a}, \mathbf{b}\rangle\langle \mathbf{a}, \mathbf{b}|^{AB}, \tag{C2}$$

with $p^A(\mathbf{a}) = (p_{01}^{A_1})^{a_1} \cdots (p_{01}^{A_n})^{a_n} (p_{00}^{A_1})^{1-a_1} \cdots (p_{00}^{A_n})^{1-a_n}$ and $p^B(\mathbf{b}) = (p_{01}^{B_1})^{b_1} \cdots (p_{01}^{B_n})^{b_n} (p_{00}^{B_1})^{1-b_1} \cdots (p_{00}^{B_n})^{1-b_n}$.

C.1. Effective noisy measurement operator for the HST

In the noiseless HST, the measurement is preceded by the disentangling unitary $(E^{AB})^\dagger$, where E^{AB} is defined in (B1). In the Heisenberg picture, this corresponds to the evolution of the measurement operator with respect to the unitary E^{AB}. We now derive the effective noisy POVM element as the evolution of $\widetilde{P}_0$ under the noisy entangling channel $\widetilde{\mathcal{E}}^{AB}$ (defined in section B).

Using (A5), $|\mathbf{a}, \mathbf{b}\rangle\langle \mathbf{a}, \mathbf{b}|^{AB}$ can be expressed as follows:

$$|\mathbf{a}, \mathbf{b}\rangle\langle \mathbf{a}, \mathbf{b}|^{AB} = (X_A^{\mathbf{a}} \otimes X_B^{\mathbf{b}}) \left(\frac{1}{2^{2n}} \sum_{\mathbf{l}, \mathbf{k}} Z_A^{\mathbf{l}} \otimes Z_B^{\mathbf{k}}\right) (X_A^{\mathbf{a}} \otimes X_B^{\mathbf{b}})$$

$$= \frac{1}{2^{2n}} \sum_{\mathbf{l}, \mathbf{k}} (-1)^{\mathbf{a} \cdot \mathbf{l}} (-1)^{\mathbf{b} \cdot \mathbf{k}} Z_A^{\mathbf{l}} \otimes Z_B^{\mathbf{k}}, \tag{C3}$$

where we used the properties of the Pauli operators as defined in (A2). Then, from (B4) and the linearity of quantum channels, it follows that

$$\widetilde{\mathcal{E}}^{AB}(|\mathbf{a}, \mathbf{b}\rangle\langle \mathbf{a}, \mathbf{b}|^{AB}) = \frac{1}{2^{2n}} \sum_{\mathbf{l}, \mathbf{k}} m_{\mathbf{l}, \mathbf{l}, \mathbf{k}, \mathbf{k}}^{AB} (-1)^{\mathbf{a} \cdot \mathbf{l}} (-1)^{\mathbf{b} \cdot \mathbf{k}} X_A^{\mathbf{l}} Z_A^{\mathbf{k}} \otimes X_B^{\mathbf{l}} Z_B^{\mathbf{k}}. \tag{C4}$$

Therefore, from (C2) and (C4) it follows that

$$\widetilde{\mathcal{E}}^{AB}(\widetilde{P}_{\mathbf{0}}) = \frac{1}{2^{2n}}\sum_{\boldsymbol{a},\boldsymbol{b}} m_{\boldsymbol{a},\boldsymbol{a},\boldsymbol{b},\boldsymbol{b}}^{AB} \widehat{p}_{\boldsymbol{a},\boldsymbol{b}}^{A} Z_A^{\boldsymbol{b}} X_A^{\boldsymbol{a}} \otimes Z_B^{\boldsymbol{b}} X_B^{\boldsymbol{a}}, \tag{C5}$$

where $\widehat{p}_{\boldsymbol{a},\boldsymbol{b}}^{A} = \sum_{\boldsymbol{l},\boldsymbol{k}}(-1)^{\boldsymbol{a}\cdot\boldsymbol{l}}(-1)^{\boldsymbol{b}\cdot\boldsymbol{k}}p^A(\boldsymbol{l})p^B(\boldsymbol{k})$, and $p^A(\boldsymbol{l})$ and $p^B(\boldsymbol{k})$ are probability distributions as in (C2).

C.2. Effective noisy measurement operator for the LHST

In the LHST, a noisy measurement on two qubits $A_j B_j$ is preceded by the disentangling unitary $(E^{A_j B_j})^\dagger$ acting on the same two qubits. Similar to section C.1, we now derive the effective POVM element as the evolution of the operator $Q_{00}^{(j)}$ (defined below) under the adjoint of the noisy disentangling channel, as defined in (B7). The noisy POVM for the qubits $A_j B_j$ is given by

$$\widetilde{Q}_{00}^{(j)} = \sum_{a',b'=0}^{1} p^{A_j}(a')p^{B_j}(b')|a',\,b'\rangle\langle a',\,b'|^{A_j B_j}, \tag{C6}$$

which follows from (C2). Moreover, the overall noisy POVM for the LHST is defined as

$$\widetilde{Q}_{00} = \frac{1}{n}\sum_{j=1}^{n} \widetilde{Q}_{00}^{(j)} \otimes \mathbb{1}_{\overline{A}_j \overline{B}_j}. \tag{C7}$$

By using arguments similar to those used in (C3), (C4), and (C5), we find that

$$\widetilde{\mathcal{E}}_j'^{AB}(\widetilde{Q}_{00}^{(j)} \otimes \mathbb{1}_{\overline{A}_j \overline{B}_j}) = \frac{1}{2^2}\sum_{a_j,b_j} m_{a_j,a_j,b_j,b_j}^{AB} \widehat{p}_{a_j,b_j}^{A_j} Z_A^{b_j} X_A^{a_j} \otimes Z_{B_j}^{b_j} X_{B_j}^{a_j} \otimes \mathbb{1}_{\overline{A}_j \overline{B}_j}, \tag{C8}$$

where $\widetilde{\mathcal{E}}_j'^{AB}$ is given by (B7) and $\widehat{p}_{a_j,b_j}^{A_j} = \sum_{a',b'=0}^{1}(-1)^{a_j a'}(-1)^{b_j b'}p^{A_j}(a')p^{B_j}(b')$.

Therefore, the overall effective noisy POVM for the LHST is defined as

$$\widetilde{\mathcal{E}}'^{AB}(\widetilde{Q}_{00}) = \frac{1}{2^2}\frac{1}{n}\sum_{j=1}^{n}\sum_{a_j,b_j=0}^{1} m_{a_j,a_j,b_j,b_j}^{AB} \widehat{p}_{a_j,b_j}^{A_j} Z_A^{b_j} X_A^{a_j} \otimes Z_{B_j}^{b_j} X_{B_j}^{a_j} \otimes \mathbb{1}_{\overline{A}_j \overline{B}_j}. \tag{C9}$$

Appendix D. Proof of theorem 1

Before providing a proof of theorem 1, we prove the following lemma.

Lemma 2. *Let $C_{\mathsf{QC}}(V)$ be a cost function of V with $V \in \mathbb{V}_d$, and $\mathbb{V}_d$ the set of $d \times d$ unitary matrices. Additionally suppose that $C_{\mathsf{QC}}(V)$ can be evaluated using a quantum circuit denoted QC as follows:*

$$C_{\mathsf{QC}}(V) := \mathrm{Tr}[\Lambda\mathcal{E}_V(\rho)], \tag{D1}$$

where ρ is a quantum state, Λ denotes a POVM element and $\mathcal{E}_V$ denotes the noisy unital quantum channel describing the evolution of the state throughout the computation, which depends on the unitary V. Then $\widetilde{C}_{\mathsf{QC}}(V)$ exhibits strong-OPR to a noise model composed of $\mathcal{E}_V$ and a global depolarizing channels acting continuously throughout the computation.

Proof. Without loss of generality let us decompose $\mathcal{E}_V$ as k noisy unital quantum channels: $\mathcal{E}_V = \mathcal{E}_V^k \circ \ldots \circ \mathcal{E}_V^1$. In the presence of global depolarizing noise acting throughout the computation, the cost function can now be expressed as

$$\widetilde{C}_{\mathsf{QC}}(V) = \mathrm{Tr}[\Lambda(\mathcal{D}^{k+1} \circ \mathcal{E}_V^k \circ \ldots \circ \mathcal{D}^2 \circ \mathcal{E}_V^1 \circ \mathcal{D}^1)(\rho)], \tag{D2}$$

where we have interleaved the channels $\mathcal{E}_V^i$ with global depolarizing channels $\mathcal{D}^i$. From definition 1 and from the fact that $\mathcal{E}_V^i(\mathbb{1}) = \mathbb{1}$, it follows that

$$\widetilde{C}_{\mathsf{QC}}(V) = \mathrm{Tr}[\Lambda(\mathcal{D}^{k+1}\circ\mathcal{E}_V^k \circ \ldots \circ \mathcal{D}^2 \circ \mathcal{E}_V^1\circ\mathcal{D}^1)(\rho)] = p\,\mathrm{Tr}[\Lambda(\mathcal{E}_V^k \circ \ldots \mathcal{E}_V^2 \circ \mathcal{E}_V^1)(\rho)] + (1-p)\mathrm{Tr}[\Lambda\mathbb{1}]/2^n \tag{D3}$$

$$= pC_{\mathsf{QC}}(V) + (1-p)/2^n, \tag{D4}$$

where $p = p_{k+1} \ldots p_1$. Let $\mathbb{V}_d^{\mathrm{opt}}$ denote the sets of unitaries that optimize $C_{\mathsf{QC}}(V)$ i.e.

$$\mathbb{V}_d^{\mathrm{opt}} = \{V' \in \mathbb{V}_d\colon C_{\mathsf{QC}}(V') = \min_{V \in \mathbb{V}_d} C_{\mathsf{QC}}(V)\}. \tag{D5}$$

Then, from (D4) we have that any unitary in $\mathbb{V}_d^{opt}$ will also optimize $\widetilde{C}_{\mathsf{QC}}(V)$. Hence $\widetilde{C}_{\mathsf{QC}}(V)$ exhibits strong-OPR to a noise model composed of $\mathcal{E}_V$ and a global depolarizing channels acting throughout the computation. $\square$

By means of lemma 2 we know that if we show that a quantity exhibits OPR to a noise model $\mathcal{N}$ which does not include global depolarizing noise acting continuously throughout the computation, then said quantity will also exhibit OPR if we include global depolarizing noise to $\mathcal{N}$.

We now provide a proof for theorem 1.

Theorem 1. *The cost functions C_{HST} and C_{LHST} exhibit strong-OPR to Noise Model 1 in definition 7.*

Proof. We begin by breaking up the HST circuit into three time intervals. In the first time interval, the noisy entangling channel $\widetilde{\mathcal{E}}^{AB}$ is applied. In the second time interval, the quantum channel $\mathcal{V}^{\dagger} \circ \mathcal{U}$ implements the unitaries U and $V^{\dagger}$. Finally, in the third time interval $(\widetilde{\mathcal{E}}^{AB})^{\dagger}$ is applied. We assume that the global depolarizing noise occurs on systems AB during all three time intervals and the global depolarizing noise occurs on system A during the implementation of $\mathcal{V}^{\dagger} \circ \mathcal{U}$. Moreover, suppose that two different global Pauli channels $\mathcal{Q}^{AB}$ and $\widehat{\mathcal{Q}}^{AB}$ act at times τ_1 and τ_2, respectively, and global non-unital Pauli channels act continuously on system B in between τ_1 and τ_2.

Let $\rho^{(0)}$ denotes the initial state of the HST circuit and is given by $\rho^{(0)} = |\mathbf{0}, \mathbf{0}\rangle \langle \mathbf{0}, \mathbf{0}|^{AB}$. At τ_1 the state is

$$\rho^{(1)} = \mathcal{Q}^{AB}(\mathcal{D}^{AB}_{p^{(1,k)}} \circ \widetilde{\mathcal{E}}^{AB}_k ... \mathcal{D}^{AB}_{p^{(1,1)}} \circ \widetilde{\mathcal{E}}^{AB}_1(\rho^{(0)})), \tag{D6}$$

where we have broken up the τ_1 into k time increments and $\widetilde{\mathcal{E}}^{AB}_k \circ ... \widetilde{\mathcal{E}}^{AB}_1$ is the channel that implements the noisy entangling channel $\widetilde{\mathcal{E}}^{AB}$, as defined in (B2). Moreover, each $\widetilde{\mathcal{E}}^{AB}_i$ is followed by a global depolarizing channel $\mathcal{D}^{AB}_{p^{(1,i)}}$, where $p^{(r,s)}$ denotes the depolarizing probability for the sth time increment of the rth time interval. Then $\rho^{(1)}$ reduces to

$$\rho^{(1)} = \mathcal{Q}^{AB}(\mathcal{D}^{AB}_{p^{(1,k)}} \circ \widetilde{\mathcal{E}}^{AB}_k ... \widetilde{\mathcal{E}}^{AB}_2(p^{(1,1)}\widetilde{\mathcal{E}}^{AB}_1(\rho^{(0)}) + (1 - p^{(1,1)})\mathbb{1}/2^{2n}) \tag{D7}$$

$$= p^{(1)} \mathcal{Q}^{AB} \circ \widetilde{\mathcal{E}}^{AB}(\rho^{(0)}) + (1 - p^{(1)})\mathbb{1}/d = p^{(1)}\left[\frac{1}{2^{2n}}\sum_{\boldsymbol{a},\boldsymbol{b}} \beta^{AB}_{\boldsymbol{a},\boldsymbol{b}} X^{\boldsymbol{a}}_A Z^{\boldsymbol{b}}_A \otimes X^{\boldsymbol{a}}_B Z^{\boldsymbol{b}}_B\right] + (1 - p^{(1)})\mathbb{1}/2^{2n}, \tag{D8}$$

where $p^{(1)} = p^{(1,1)} ... p^{(1,k)}$. The second equality follows from lemma 2 as $\widetilde{\mathcal{E}}^{AB}$ consists of only unitary and Pauli channels, and thus each $\widetilde{\mathcal{E}}^{AB}_i$ is a unital channel, where $i \in \{1, ..., k\}$. The last equality follows from (B4) and (A8), where $\beta^{AB}_{\boldsymbol{a},\boldsymbol{b}} = m^{AB}_{\boldsymbol{a},\boldsymbol{a},\boldsymbol{b},\boldsymbol{b}} q^{AB}_{\boldsymbol{a},\boldsymbol{a},\boldsymbol{b},\boldsymbol{b}}$.

Similarly, the state at τ_2 is given by

$$\rho^{(2)} = \widehat{\mathcal{Q}}^{AB}(\mathcal{D}^{AB}_{p^{(2,l)}} \circ \mathcal{D}^{A}_{s^{(2,l)}} \circ (\mathcal{W}_l \otimes \mathcal{P}^{B}_{\mathrm{NU},l}) ... \mathcal{D}^{AB}_{p^{(2,1)}} \circ \mathcal{D}^{A}_{s^{(2,1)}} \circ (\mathcal{W}_1 \otimes \mathcal{P}^{B}_{\mathrm{NU},1})(\rho^{(1)})). \tag{D9}$$

We first find the action of the channel $\mathcal{W}_1 \otimes \mathcal{P}^{B}_{\mathrm{NU},1}$ on $\rho^{(1)}$. Consider that

$$(\mathcal{W}_1 \otimes \mathcal{P}^{B}_{\mathrm{NU},1})(\rho^{(1)}) = \frac{1}{2^{2n}}(\mathcal{W}_1 \otimes \mathcal{P}^{B}_{\mathrm{NU},1})\left[p^{(1)}\left(\sum_{(\boldsymbol{a},\boldsymbol{b})\neq(\mathbf{0},\mathbf{0})} \beta^{AB}_{\boldsymbol{a},\boldsymbol{b}} X^{\boldsymbol{a}}_A Z^{\boldsymbol{b}}_A \otimes X^{\boldsymbol{a}}_B Z^{\boldsymbol{b}}_B\right) + \mathbb{1}_{AB}\right] \tag{D10}$$

$$= \frac{1}{2^{2n}}\left[p^{(1)}\left(\sum_{(\boldsymbol{a},\boldsymbol{b})\neq(\mathbf{0},\mathbf{0})} \beta^{AB}_{\boldsymbol{a},\boldsymbol{b}} c^{(1)}_{\boldsymbol{a},\boldsymbol{b}} W_1 X^{\boldsymbol{a}}_A Z^{\boldsymbol{b}}_A W^{\dagger}_1 \otimes X^{\boldsymbol{a}}_B Z^{\boldsymbol{b}}_B\right) + \mathbb{1}_{AB} + \sum_{(\boldsymbol{g},\boldsymbol{h})\neq(\mathbf{0},\mathbf{0})} d^{(1)}_{\boldsymbol{g},\boldsymbol{h}}\mathbb{1}_A \otimes X^{\boldsymbol{g}}_B Z^{\boldsymbol{h}}_B\right], \tag{D11}$$

where we used the definition of a non-unital Pauli channel from (A9) and (A10). We note that the terms that are independent of W_i do not affect the global optima. Therefore, the only relevant term in (D9) is

$$\widetilde{\rho}^{(2)} = \frac{p^{(2)} s^{(2)} p^{(1)}}{2^{2n}} \widehat{\mathcal{Q}}^{AB}\left(\sum_{(\boldsymbol{a},\boldsymbol{b})\neq(\mathbf{0},\mathbf{0})} \beta^{AB}_{\boldsymbol{a},\boldsymbol{b}}(\prod_{i=1}^{m} c^{(i)}_{\boldsymbol{a},\boldsymbol{b}}) W X^{\boldsymbol{a}}_A Z^{\boldsymbol{b}}_A W^{\dagger} \otimes X^{\boldsymbol{a}}_B Z^{\boldsymbol{b}}_B\right), \tag{D12}$$

where $p^{(2)} = p^{(2,1)} ... p^{(2,l)}$ and $s^{(2)} = s^{(2,1)} ... s^{(2,l)}$, and where we have used (A9) and lemma 2.

Finally, the relevant term after the action of the noisy disentangling channel is

$$\widetilde{\rho}^{(3)} = \mathcal{D}^{AB}_{p^{(3,m)}} \circ (\widetilde{\mathcal{E}}^{AB}_m)^{\dagger} ... \mathcal{D}^{AB}_{p^{(3,1)}} \circ (\widetilde{\mathcal{E}}^{AB}_1)^{\dagger}(\widetilde{\rho}^{(2)}) = p^{(3)}(\widetilde{\mathcal{E}}^{AB})^{\dagger}(\widetilde{\rho}^{(2)}) + (1 - p^{(3)})\mathbb{1}/2^{2n}, \tag{D13}$$

where $p^{(3)} = p^{(3,m)} ... p^{(3,1)}$. The last equality follows from the fact that the channel $(\widetilde{\mathcal{E}}^{AB})^{\dagger}$ consists of unitary channels and Pauli channels, and thus each $(\widetilde{\mathcal{E}}^{AB}_i)^{\dagger}$ is a unital channel. Therefore, the term that decides the global optima in the HST is given by

$$\sigma^{(3)} = (\widetilde{\mathcal{E}}^{AB})^{\dagger} \circ \widehat{\mathcal{Q}}^{AB}\left(\sum_{(\boldsymbol{a},\boldsymbol{b})\neq(\mathbf{0},\mathbf{0})} \beta^{AB}_{\boldsymbol{a},\boldsymbol{b}}(\prod_{i=1}^{m} c^{(i)}_{\boldsymbol{a},\boldsymbol{b}}) W X^{\boldsymbol{a}}_A Z^{\boldsymbol{b}}_A W^{\dagger} \otimes X^{\boldsymbol{a}}_B Z^{\boldsymbol{b}}_B\right), \tag{D14}$$

where we have omitted the scaling factors. Let $\widetilde{F}_{\mathsf{HST}}(V) \propto f(V) := \mathrm{Tr}\,[\widetilde{P}_0 \sigma^{(3)}]$. Then

$$f(V) = \mathrm{Tr}\left[(\widehat{\mathcal{Q}}^{AB} \circ \widetilde{\mathcal{E}}^{AB})(\widetilde{P}_0)\left(\sum_{(\boldsymbol{a},\boldsymbol{b})\neq(\boldsymbol{0},\boldsymbol{0})} \beta_{\boldsymbol{a},\boldsymbol{b}}^{AB}(\prod_{i=1}^{m} c_{\boldsymbol{a},\boldsymbol{b}}^{(i)}) W X_A^{\boldsymbol{a}} Z_A^{\boldsymbol{b}} W^\dagger \otimes X_B^{\boldsymbol{a}} Z_B^{\boldsymbol{b}}\right)\right] \tag{D15}$$

$$= \mathrm{Tr}\left[\sum_{\substack{(\boldsymbol{a},\boldsymbol{b})\neq(\boldsymbol{0},\boldsymbol{0}) \\ \tilde{\boldsymbol{a}},\tilde{\boldsymbol{b}}}} \kappa_{\boldsymbol{a},\tilde{\boldsymbol{a}},\boldsymbol{b},\tilde{\boldsymbol{b}}}^{AB} Z_A^{\tilde{\boldsymbol{b}}} X_A^{\tilde{\boldsymbol{a}}} W X_A^{\boldsymbol{a}} Z_A^{\boldsymbol{b}} W^\dagger \otimes Z_B^{\tilde{\boldsymbol{b}}} X_B^{\tilde{\boldsymbol{a}}} X_B^{\boldsymbol{a}} Z_B^{\boldsymbol{b}}\right] \tag{D16}$$

$$= \mathrm{Tr}_A\left[\sum_{(\boldsymbol{a},\boldsymbol{b})\neq(\boldsymbol{0},\boldsymbol{0})} \kappa_{\boldsymbol{a},\boldsymbol{a},\boldsymbol{b},\boldsymbol{b}}^{AB} Z_A^{\boldsymbol{b}} X_A^{\boldsymbol{a}} W X_A^{\boldsymbol{a}} Z_A^{\boldsymbol{b}} W^\dagger\right]. \tag{D17}$$

The second equality follows from (C5), where we set $\kappa_{\boldsymbol{a},\tilde{\boldsymbol{a}},\boldsymbol{b},\tilde{\boldsymbol{b}}}^{AB} := (1/2^{2n}) \widetilde{m}_{\tilde{\boldsymbol{a}},\tilde{\boldsymbol{a}},\tilde{\boldsymbol{b}},\tilde{\boldsymbol{b}}}^{AB} \widehat{p}_{\tilde{\boldsymbol{a}},\tilde{\boldsymbol{b}}}^A \widehat{q}_{\tilde{\boldsymbol{a}},\tilde{\boldsymbol{a}},\tilde{\boldsymbol{b}},\tilde{\boldsymbol{b}}}^{AB} \beta_{\boldsymbol{a},\boldsymbol{b}}^{AB}(\prod_{i=1}^{m} c_{\boldsymbol{a},\boldsymbol{b}}^{(i)})$. The last equality follows from (A2). Let $\mathbb{V}_d^{\mathrm{opt}}$ denote the sets of unitaries that optimize $F_{\mathsf{HST}}(V)$ (and hence $C_{\mathsf{HST}}(V)$) such that

$$\mathbb{V}_d^{\mathrm{opt}} = \{V' \in \mathbb{V}_d \colon W = (V')^\dagger U = \mathrm{e}^{\mathrm{i}\phi}\mathbb{1}, \quad \text{for some} \quad \phi \in [0, 2\pi]\}. \tag{D18}$$

We remark that this set of unitaries also optimizes $F_{\mathsf{LHST}}(V)$ (and hence $C_{\mathsf{LHST}}(V)$). Then, for $V' \in \mathbb{V}_d$ we find $f(V') = \sum_{(\boldsymbol{a},\boldsymbol{b})\neq(\boldsymbol{0},\boldsymbol{0})} \kappa_{\boldsymbol{a},\boldsymbol{a},\boldsymbol{b},\boldsymbol{b}}^{AB}$. Let

$$T(V) := \sum_{(\boldsymbol{a},\boldsymbol{b})\neq(\boldsymbol{0},\boldsymbol{0})} \sqrt{\kappa_{\boldsymbol{a},\boldsymbol{a},\boldsymbol{b},\boldsymbol{b}}^{AB}} X_A^{\boldsymbol{a}} Z_A^{\boldsymbol{b}} W^\dagger \otimes |\boldsymbol{a},\boldsymbol{b}\rangle, \qquad S(V) := \sum_{(\boldsymbol{a}',\boldsymbol{b}')\neq(\boldsymbol{0},\boldsymbol{0})} \sqrt{\kappa_{\boldsymbol{a}',\boldsymbol{a}',\boldsymbol{b}',\boldsymbol{b}'}^{AB}} W^\dagger X_A^{\boldsymbol{a}'} Z_A^{\boldsymbol{b}'} \otimes |\boldsymbol{a}',\boldsymbol{b}'\rangle. \tag{D19}$$

Consider the following inequality:

$$f(V) = |\langle S(V), T(V)\rangle| \leqslant \sqrt{\mathrm{Tr}(S(V)^\dagger S(V))}\sqrt{\mathrm{Tr}(T(V)^\dagger T(V))} = \sum_{(\boldsymbol{a},\boldsymbol{b})\neq(\boldsymbol{0},\boldsymbol{0})} \kappa_{\boldsymbol{a},\boldsymbol{a},\boldsymbol{b},\boldsymbol{b}}^{AB}, \tag{D20}$$

where we used the Cauchy–Schwarz inequality. Moreover, note that the inequality in (D20) is saturated for any matrix $V' \in \mathbb{V}_d$ if we assume that the coefficients $\kappa_{\boldsymbol{a},\boldsymbol{a},\boldsymbol{b},\boldsymbol{b}}^{AB}$ characterizing the noise satisfy $\kappa_{\boldsymbol{a},\boldsymbol{a},\boldsymbol{b},\boldsymbol{b}}^{AB} \geqslant 0$. Therefore, the set of unitaries that optimize $F_{\mathsf{HST}}(V)$ (and hence $C_{\mathsf{HST}}(V)$) is $\widetilde{\mathbb{V}}_d^{opt} = \mathbb{V}_d^{opt}$. According to definition 6, the latter means that C_{HST} exhibits strong-OPR to Noise Model 1 in definition 7.

We now show that the cost function C_{LHST} exhibits strong-OPR to Noise Model 1. The LHST corresponds to the optimization of the following function:

$$\widetilde{F}_{\mathsf{LHST}}(V) \propto g(V) = \mathrm{Tr}\left[(\widehat{\mathcal{Q}}^{AB} \circ \widetilde{\mathcal{E}}'^{AB})(\widetilde{Q}_{00})\left(\sum_{(\boldsymbol{a},\boldsymbol{b})\neq(\boldsymbol{0},\boldsymbol{0})} \beta_{\boldsymbol{a},\boldsymbol{b}}^{AB}(\prod_{i=1}^{m} c_{\boldsymbol{a},\boldsymbol{b}}^{(i)}) W X_A^{\boldsymbol{a}} Z_A^{\boldsymbol{b}} W^\dagger \otimes X_B^{\boldsymbol{a}} Z_B^{\boldsymbol{b}}\right)\right], \tag{D21}$$

where we replaced the disentangling and measurement channels in (D15) with (C9). Consider the following:

$$g(V) = \mathrm{Tr}\left[\left(\frac{1}{2^2}\frac{1}{n}\sum_{j=1}^{n}\sum_{a_j',b_j'=0}^{1} \widetilde{m}_{a_j',a_j',b_j',b_j'}^{AB} \widehat{p}_{a_j',b_j'}^{A_j} \widehat{q}_{a_j',a_j',b_j',b_j'} Z_{A_j}^{b_j'} X_{A_j}^{a_j'} \otimes Z_{B_j}^{b_j'} X_{B_j}^{a_j'} \otimes \mathbb{1}_{\overline{A_j}\overline{B_j}}\right)\right.$$

$$\left. \times \left(\sum_{(\boldsymbol{a},\boldsymbol{b})\neq(\boldsymbol{0},\boldsymbol{0})} \beta_{\boldsymbol{a},\boldsymbol{b}}^{AB}(\prod_{i=1}^{m} c_{\boldsymbol{a},\boldsymbol{b}}^{(i)}) W X_A^{\boldsymbol{a}} Z_A^{\boldsymbol{b}} W^\dagger \otimes X_B^{\boldsymbol{a}} Z_B^{\boldsymbol{b}}\right)\right] \tag{D22}$$

$$= \mathrm{Tr}\left[\sum_{j=1}^{n}\sum_{(\boldsymbol{a},\boldsymbol{b})\neq(\boldsymbol{0},\boldsymbol{0})}\sum_{a_j',b_j'=0}^{1} \xi_{\boldsymbol{a},a_j',\boldsymbol{b},b_j'}^{(j)}(Z_{A_j}^{b_j'} X_{A_j}^{a_j'} \otimes \mathbb{1}_{\overline{A_j}}) W X_A^{\boldsymbol{a}} Z_A^{\boldsymbol{b}} W^\dagger \otimes Z_{B_j}^{b_j'} X_{B_j}^{a_j'} X_{B_j}^{a_j} Z_{B_j}^{b_j} X_{\overline{B_j}}^{\overline{a}_j} Z_{\overline{B_j}}^{\overline{b}_j}\right] \tag{D23}$$

$$= \mathrm{Tr}_A\left[\sum_{j=1}^{n}\sum_{(\boldsymbol{a},\boldsymbol{b})\neq(\boldsymbol{0},\boldsymbol{0})}\sum_{a_j',b_j'=0}^{1} \xi_{\boldsymbol{a},a_j',\boldsymbol{b},b_j'}^{(j)}(Z_{A_j}^{b_j'} X_{A_j}^{a_j'} \otimes \mathbb{1}_{\overline{A_j}}) W X_A^{\boldsymbol{a}} Z_A^{\boldsymbol{b}} W^\dagger \otimes \mathrm{Tr}_{B_j}(Z_{B_j}^{b_j'} X_{B_j}^{a_j'} X_{B_j}^{a_j} Z_{B_j}^{b_j}) \mathrm{Tr}_{\overline{B_j}}(X_{\overline{B_j}}^{\overline{a}_j} Z_{\overline{B_j}}^{\overline{b}_j})\right] \tag{D24}$$

$$= \mathrm{Tr}_A\left[\sum_{j=1}^{n}\sum_{(a_j,b_j)\neq(0,0)} \xi_{a_j,a_j,b_j,b_j}^{(j)}(Z_{A_j}^{b_j} X_{A_j}^{a_j} \otimes \mathbb{1}_{\overline{A_j}})(W(X_{A_j}^{a_j} Z_{A_j}^{b_j} \otimes \mathbb{1}_{\overline{A_j}}) W^\dagger)\right] \tag{D25}$$

$$\leqslant \sum_{j=1}^{n}\sum_{(a_j,b_j)\neq(0,0)} \xi_{a_j,a_j,b_j,b_j}^{(j)}, \tag{D26}$$

where in (D24) we have split Tr_B into a contribution from qubit B_j and a contribution on all qubits except B_j, and where $\xi^{(j)}_{\boldsymbol{a},a'_j,\boldsymbol{b},b'_j} = (1/4n)\widetilde{m}^{A,B}_{\boldsymbol{a}'_j,a'_j,b'_j,b'_j}\widehat{p}^{A_j}_{a'_j,b'_j}\widehat{q}_{a'_j,a'_j,b'_j,b'_j}\beta^{AB}_{\boldsymbol{a},\boldsymbol{b}}(\prod^m_{i=1}c^{(i)}_{\boldsymbol{a},\boldsymbol{b}})$. The first equality is derived from (C9), while the inequality follows from the arguments similar to (D20).

Here we remark that the inequality (D26) is saturated for any unitary matrix in the set of unitaries that optimize $F_{\mathsf{HST}}(V)$ (and hence $C_{\mathsf{LHST}}(V)$) given by (D18). Hence, C_{LHST} exhibits strong-OPR to Noise Model 1 in definition 7 if we assume that the coefficients $\xi^{(j)}_{a_j,a_j,b_j,b_j}$ characterizing the noise satisfy $\xi^{(j)}_{a_j,a_j,b_j,b_j} \geqslant 0$. □

Appendix E. Proof of theorem 2

Theorem 2. *The cost functions C_{HST} and C_{LHST} exhibit strong-OPR to Noise Model 2 in definition 8.*

Proof. We break up the HST circuit into three time intervals similar to section D. We again assume that the global depolarizing noise occurs on system AB during all three time intervals and the global depolarizing noise occurs on system A during the implementation of $\mathcal{V}^\dagger \circ \mathcal{U}$. Moreover, suppose that a global Pauli channel $\mathcal{Q}^{AB}$ followed by a global non-unital Pauli channel $\mathcal{P}^A_{\mathrm{NU}}$ acts at time τ_1. Furthermore, a global pauli channel $\widehat{\mathcal{Q}}^{AB}$ acts at time τ_2, while a global Pauli channel acts continuously on the system B in between τ_1 and τ_2.

The state at τ_1 is given by

$$\rho^{(1)} = p^{(1)}\mathcal{P}^A_{\mathrm{NU}} \circ \mathcal{Q}^{AB}\circ\widetilde{\mathcal{E}}^{AB}(\rho^{(0)}) + (1 - p^{(1)})\mathcal{P}^A_{\mathrm{NU}}(\mathbb{1}/2^{2n}) \tag{E1}$$

$$= p^{(1)}\left[\frac{1}{2^{2n}}\sum_{(\boldsymbol{a},\boldsymbol{b})\neq(\boldsymbol{0},\boldsymbol{0})}\widetilde{\beta}^{AB}_{\boldsymbol{a},\boldsymbol{b}}X^{\boldsymbol{a}}_A Z^{\boldsymbol{b}}_A \otimes X^{\boldsymbol{a}}_B Z^{\boldsymbol{b}}_B\right] + \frac{1}{2^{2n}}\mathbb{1} + \frac{1}{2^{2n}}\sum_{(\boldsymbol{g},\boldsymbol{h})\neq(\boldsymbol{0},\boldsymbol{0})}d_{\boldsymbol{g},\boldsymbol{h}}X^{\boldsymbol{g}}_A Z^{\boldsymbol{h}}_A \otimes \mathbb{1}_B. \tag{E2}$$

The first equality follows from arguments similar to those used to derive (D6)–(D8). The last equality follows from (B4), (A9), and (A10), where $\widetilde{\beta}^{AB}_{\boldsymbol{a},\boldsymbol{b}} = m^{AB}_{\boldsymbol{a},\boldsymbol{a},\boldsymbol{b},\boldsymbol{b}}q^{AB}_{\boldsymbol{a},\boldsymbol{a},\boldsymbol{b},\boldsymbol{b}}c_{\boldsymbol{a},\boldsymbol{b}}$.

At τ_2 the state is

$$\rho^{(2)} = \widehat{\mathcal{Q}}^{AB}(\mathcal{D}^{AB}_{p^{(2,l)}}\circ\mathcal{D}^A_{s^{(2,l)}}\circ(\mathcal{W}_l \otimes \widehat{\mathcal{P}}^B_l)\ldots \mathcal{D}^{AB}_{p^{(2,1)}} \circ \mathcal{D}^A_{s^{(2,1)}}\circ(\mathcal{W}_1 \otimes \widehat{\mathcal{P}}^B_1)(\rho^{(1)})). \tag{E3}$$

The term that depends on W in (E3) is given by

$$\widetilde{\rho}^{(2)} = \frac{1}{2^{2n}}\widehat{\mathcal{Q}}^{AB}\left[p^{(2)}s^{(2)}p^{(1)}\sum_{(\boldsymbol{a},\boldsymbol{b})\neq(\boldsymbol{0},\boldsymbol{0})}\widetilde{\beta}^{AB}_{\boldsymbol{a},\boldsymbol{b}}(\prod^l_i\widehat{p}^{(i)}_{\boldsymbol{a},\boldsymbol{b}})WX^{\boldsymbol{a}}_A Z^{\boldsymbol{b}}_A W^\dagger \otimes X^{\boldsymbol{a}}_B Z^{\boldsymbol{b}}_B + \sum_{(\boldsymbol{g},\boldsymbol{h})\neq(\boldsymbol{0},\boldsymbol{0})}d_{\boldsymbol{g},\boldsymbol{h}}WX^{\boldsymbol{g}}_A Z^{\boldsymbol{h}}_A W^\dagger \otimes \mathbb{1}_B\right], \tag{E4}$$

where we used the definition of Pauli channels from (A6) and (A8). By omitting the scaling factors, the relevant term after τ_3 is given by

$$\widetilde{\rho}^{(3)} = (\widetilde{\mathcal{E}}^{AB})^\dagger \circ \widehat{\mathcal{Q}}^{AB}\left(p^{(2)}s^{(2)}p^{(1)}\sum_{(\boldsymbol{a},\boldsymbol{b})\neq(\boldsymbol{0},\boldsymbol{0})}\widetilde{\beta}^{AB}_{\boldsymbol{a},\boldsymbol{b}}(\prod^l_i\widehat{p}^{(i)}_{\boldsymbol{a},\boldsymbol{b}})WX^{\boldsymbol{a}}_A Z^{\boldsymbol{b}}_A W^\dagger \otimes X^{\boldsymbol{a}}_B Z^{\boldsymbol{b}}_B\right)$$
$$+ (\widetilde{\mathcal{E}}^{AB})^\dagger \circ \widehat{\mathcal{Q}}^{AB}\left(\sum_{(\boldsymbol{g},\boldsymbol{h})\neq(\boldsymbol{0},\boldsymbol{0})}d_{\boldsymbol{g},\boldsymbol{h}}WX^{\boldsymbol{g}}_A Z^{\boldsymbol{h}}_A W^\dagger \otimes \mathbb{1}_B\right). \tag{E5}$$

Let $\widetilde{F}_{\mathsf{HST}}(V) \propto f(V) := \mathrm{Tr}[\widetilde{P}_0\widetilde{\rho}^{(3)}]$. Then

$$f(V) = \mathrm{Tr}\left[(\widehat{\mathcal{Q}}^{AB} \circ \widetilde{\mathcal{E}}^{AB})(\widetilde{P}_0)\left(p^{(2)}s^{(2)}p^{(1)}\sum_{(\boldsymbol{a},\boldsymbol{b})\neq(\boldsymbol{0},\boldsymbol{0})}\widetilde{\beta}^{AB}_{\boldsymbol{a},\boldsymbol{b}}(\prod^l_i\widehat{p}^{(i)}_{\boldsymbol{a},\boldsymbol{b}})WX^{\boldsymbol{a}}_A Z^{\boldsymbol{b}}_A W^\dagger \otimes X^{\boldsymbol{a}}_B Z^{\boldsymbol{b}}_B\right)\right]$$
$$+ \mathrm{Tr}\left[(\widehat{\mathcal{Q}}^{AB} \circ \widetilde{\mathcal{E}}^{AB})(\widetilde{P}_0)\left(\sum_{(\boldsymbol{g},\boldsymbol{h})\neq(\boldsymbol{0},\boldsymbol{0})}d_{\boldsymbol{g},\boldsymbol{h}}WX^{\boldsymbol{g}}_A Z^{\boldsymbol{h}}_A W^\dagger \otimes \mathbb{1}_B\right)\right]. \tag{E6}$$

Moreover, for simplicity we denote

$$f_1(V) := \mathrm{Tr}\left[(\widehat{\mathcal{Q}}^{AB} \circ \widetilde{\mathcal{E}}^{AB})(\widetilde{P}_0)\left(\sum_{(\boldsymbol{a},\boldsymbol{b})\neq(\boldsymbol{0},\boldsymbol{0})}\widetilde{\beta}^{AB}_{\boldsymbol{a},\boldsymbol{b}}(\prod^l_i\widehat{p}^{(i)}_{\boldsymbol{a},\boldsymbol{b}})WX^{\boldsymbol{a}}_A Z^{\boldsymbol{b}}_A W^\dagger \otimes X^{\boldsymbol{a}}_B Z^{\boldsymbol{b}}_B\right)\right], \tag{E7}$$

$$f_2(V) := \mathrm{Tr}\left[(\widehat{\mathcal{Q}}^{AB} \circ \widetilde{\mathcal{E}}^{AB})(\widetilde{P}_0)\left(\sum_{(\boldsymbol{g},\boldsymbol{h})\neq(\boldsymbol{0},\boldsymbol{0})}d_{\boldsymbol{g},\boldsymbol{h}}WX^{\boldsymbol{g}}_A Z^{\boldsymbol{h}}_A W^\dagger \otimes \mathbb{1}_B\right)\right]. \tag{E8}$$

Let us focus on $f_1(V)$ and $f_2(V)$ individually. Consider the following:

$$
\begin{aligned}
f_1(V) &= \mathrm{Tr}\left[\sum_{\substack{(\boldsymbol{a},\boldsymbol{b})\neq(0,0) \\ \boldsymbol{a}',\boldsymbol{b}'}} \vartheta^{AB}_{\boldsymbol{a},\boldsymbol{a}',\boldsymbol{b},\boldsymbol{b}'} Z_A^{\boldsymbol{b}'} X_A^{\boldsymbol{a}'} W X_A^{\boldsymbol{a}} Z_A^{\boldsymbol{b}} W^\dagger \otimes Z_B^{\boldsymbol{b}'} X_B^{\boldsymbol{a}'} X_B^{\boldsymbol{a}} Z_B^{\boldsymbol{b}} \right] \\
&= \mathrm{Tr}\left[\sum_{(\boldsymbol{a},\boldsymbol{b})\neq(0,0)} \vartheta^{AB}_{\boldsymbol{a},\boldsymbol{a},\boldsymbol{b},\boldsymbol{b}} Z_A^{\boldsymbol{b}} X_A^{\boldsymbol{a}} W X_A^{\boldsymbol{a}} Z_A^{\boldsymbol{b}} W^\dagger \right] \\
&\leqslant \sum_{(\boldsymbol{a},\boldsymbol{b})\neq(0,0)} \vartheta^{AB}_{\boldsymbol{a},\boldsymbol{a},\boldsymbol{b},\boldsymbol{b}}.
\end{aligned}
\tag{E9}
$$

The first equality follows from (C5), where $\vartheta^{AB}_{\boldsymbol{a},\boldsymbol{a}',\boldsymbol{b},\boldsymbol{b}'} = (1/2^{2n}) \widetilde{m}^{AB}_{\boldsymbol{a}',\boldsymbol{b}',\boldsymbol{b}'} \widehat{\widetilde{p}}^A_{\boldsymbol{a}',\boldsymbol{b}'} \widehat{q}^{AB}_{\boldsymbol{a}',\boldsymbol{a}',\boldsymbol{b}',\boldsymbol{b}'} \widetilde{\beta}^{AB}_{\boldsymbol{a},\boldsymbol{b}} (\prod_i^l \widehat{p}^{(i)}_{\boldsymbol{a},\boldsymbol{b}})$. The inequality follows from the arguments similar to (D20). Here, the last inequality in (E9) is saturated for any matrix V in the set $\mathbb{V}^{\mathrm{opt}}_d$ of unitaries that optimize $F_{\mathsf{HST}}(V)$ (and hence $C_{\mathsf{LHST}}(V)$) given by (D18).

On the other hand

$$
f_2(V) = \mathrm{Tr}\left[\sum_{\substack{(\boldsymbol{g},\boldsymbol{h})\neq(0,0) \\ \boldsymbol{a}',\boldsymbol{b}'}} \varsigma^{AB}_{\boldsymbol{g},\boldsymbol{a}',\boldsymbol{h},\boldsymbol{b}'} Z_A^{\boldsymbol{b}'} X_A^{\boldsymbol{a}'} W X_A^{\boldsymbol{g}} Z_A^{\boldsymbol{h}} W^\dagger \otimes Z_B^{\boldsymbol{a}'} X_B^{\boldsymbol{b}'} \right]
\tag{E10}
$$

$$
= \mathrm{Tr}_A\left[\sum_{\substack{(\boldsymbol{g},\boldsymbol{h})\neq(0,0) \\ \boldsymbol{a}',\boldsymbol{b}'}} \varsigma^{AB}_{\boldsymbol{g},\boldsymbol{a}',\boldsymbol{h},\boldsymbol{b}'} Z_A^{\boldsymbol{b}'} X_A^{\boldsymbol{a}'} W X_A^{\boldsymbol{g}} Z_A^{\boldsymbol{h}} W^\dagger \otimes \mathrm{Tr}_B(Z_B^{\boldsymbol{a}'} X_B^{\boldsymbol{b}'}) \right]
\tag{E11}
$$

$$
= \sum_{(\boldsymbol{g},\boldsymbol{h})\neq(0,0)} \varsigma^{AB}_{\boldsymbol{g},0,\boldsymbol{h},0} \mathrm{Tr}_A(X_A^{\boldsymbol{g}} Z_A^{\boldsymbol{h}}) = 0.
\tag{E12}
$$

where $\varsigma^{AB}_{\boldsymbol{g},\boldsymbol{a}',\boldsymbol{h},\boldsymbol{b}'} = (1/2^{2n}) \widetilde{m}^{AB}_{\boldsymbol{a}',\boldsymbol{b}',\boldsymbol{b}'} \widehat{\widetilde{p}}^A_{\boldsymbol{a}',\boldsymbol{b}'} \widehat{q}^{AB}_{\boldsymbol{a}',\boldsymbol{a}',\boldsymbol{b}',\boldsymbol{b}'} d_{\boldsymbol{g},\boldsymbol{h}} \widetilde{\beta}^{AB}_{\boldsymbol{a},\boldsymbol{b}} (\prod_i^l \widehat{p}^{(i)}_{\boldsymbol{a},\boldsymbol{b}})$. From the last equality it follows that $f_2(V)$ is independent of W (and hence of V) and thus does not affect the global optima. Therefore, from (E9) it follows that the set of unitaries that optimize $\widetilde{F}_{\mathsf{HST}}(V)$ (and hence $\widetilde{C}_{\mathsf{HST}}(V)$) is $\widetilde{\mathbb{V}}^{\mathrm{opt}}_d = \mathbb{V}^{\mathrm{opt}}_d$. From definition 6 this implies that C_{HST} exhibits strong-OPR to Noise Model 2 in definition 8 if we assume that the coefficients $\vartheta^{AB}_{\boldsymbol{a},\boldsymbol{a},\boldsymbol{b},\boldsymbol{b}}$ characterizing the noise satisfy $\vartheta^{AB}_{\boldsymbol{a},\boldsymbol{a},\boldsymbol{b},\boldsymbol{b}} \geqslant 0$.

We now show that the cost function C_{LHST} exhibits strong-OPR to Noise Model 2. In particular, in the LHST we want to optimize the following function:

$$
\begin{aligned}
\widetilde{F}_{\mathsf{LHST}}(V) &\propto g(V) \\
&= \mathrm{Tr}\Big[(\widehat{\mathcal{Q}}^{AB} \circ \widetilde{\mathcal{E}}'^{AB})(\widetilde{Q}_{00}) \Big(p^{(2)} s^{(2)} p^{(1)} \sum_{(\boldsymbol{a},\boldsymbol{b})\neq(0,0)} \widetilde{\beta}^{AB}_{\boldsymbol{a},\boldsymbol{b}} (\prod_i^l \widehat{p}^{(i)}_{\boldsymbol{a},\boldsymbol{b}}) W X_A^{\boldsymbol{a}} Z_A^{\boldsymbol{b}} W^\dagger \otimes X_B^{\boldsymbol{a}} Z_B^{\boldsymbol{b}} \Big) \Big] \\
&\quad + \mathrm{Tr}\Big[(\widehat{\mathcal{Q}}^{AB} \circ \widetilde{\mathcal{E}}'^{AB})(\widetilde{Q}_{00}) \Big(\sum_{(\boldsymbol{g},\boldsymbol{h})\neq(0,0)} d_{\boldsymbol{g},\boldsymbol{h}} W X_A^{\boldsymbol{g}} Z_A^{\boldsymbol{h}} W^\dagger \otimes \mathbb{1}_B \Big) \Big],
\end{aligned}
\tag{E13}
$$

where we replaced the disentangling and measurement channels in (E6) with (C9). We now break up $g(V)$ into two different functions.

$$
g_1(V) := \mathrm{Tr}\left[(\widehat{\mathcal{Q}}^{AB} \circ \widetilde{\mathcal{E}}'^{AB})(\widetilde{Q}_{00}) \left(\sum_{(\boldsymbol{a},\boldsymbol{b})\neq(0,0)} \widetilde{\beta}^{AB}_{\boldsymbol{a},\boldsymbol{b}} (\prod_i^l \widehat{p}^{(i)}_{\boldsymbol{a},\boldsymbol{b}}) W X_A^{\boldsymbol{a}} Z_A^{\boldsymbol{b}} W^\dagger \otimes X_B^{\boldsymbol{a}} Z_B^{\boldsymbol{b}} \right) \right],
\tag{E14}
$$

$$
g_2(V) := \mathrm{Tr}\left[(\widehat{\mathcal{Q}}^{AB} \circ \widetilde{\mathcal{E}}'^{AB})(\widetilde{Q}_{00}) \left(\sum_{(\boldsymbol{g},\boldsymbol{h})\neq(0,0)} d_{\boldsymbol{g},\boldsymbol{h}} W X_A^{\boldsymbol{g}} Z_A^{\boldsymbol{h}} W^\dagger \otimes \mathbb{1}_B \right) \right].
\tag{E15}
$$

By using arguments similar to those used to derive equations (E10)–(E12) and from (C9), it follows that $g_2(V)$ is independent of W (and hence of V). Therefore, to prove the noise resilience of the LHST, we focus only on $g_1(V)$. We then get:

$$
g_1(V) = \mathrm{Tr}\left[\sum_{j=1}^n \sum_{(\boldsymbol{a},\boldsymbol{b})\neq(0,0)} \sum_{a_j',b_j'=0}^1 \tau^{(j)}_{\boldsymbol{a},a_j',\boldsymbol{b},b_j'} (Z_{A_j}^{b_j'} X_{A_j}^{a_j'} \otimes \mathbb{1}_{\overline{A}_j}) W X_A^{\boldsymbol{a}} Z_A^{\boldsymbol{b}} W^\dagger \otimes Z_{B_j}^{b_j'} X_{B_j}^{a_j'} X_{B_j}^{a_j} Z_{B_j}^{b_j} X_{\overline{B}_j}^{\overline{a}_j} Z_{\overline{B}_j}^{\overline{b}_j} \right],
\tag{E16}
$$

where $\tau^{(j)}_{\boldsymbol{a},a_j',\boldsymbol{b},b_j'} = (1/4n) \widetilde{m}^{AB}_{a_j',a_j',b_j',b_j'} \widehat{\widetilde{p}}^{A_j}_{a_j',b_j'} \widehat{q}^{AB}_{a_j',a_j',b_j',b_j'} \widetilde{\beta}^{AB}_{\boldsymbol{a},\boldsymbol{b}} (\prod_i^l \widehat{p}^{(i)}_{\boldsymbol{a},\boldsymbol{b}})$. We note that (E16) is similar to (D23). Therefore, from the proof in section D it follows that

$$g_l(V) \leqslant \sum_{j=1}^{n} \sum_{(a_j,b_j) \neq (0,0)} \tau^{(j)}_{a_j,a_j,b_j,b_j}, \tag{E17}$$

where the inequality is saturated for unitaries V' in the set $\mathbb{V}^{\text{opt}}_d$ of unitaries that optimize $F_{\text{LHST}}(V)$ (and hence $C_{\text{LHST}}(V)$) given by (D18). This further implies that

$$g(V) \leqslant g(V'), \quad \text{for all} \quad V' \in \mathbb{V}^{\text{opt}}_d = \widetilde{\mathbb{V}}^{\text{opt}}_d. \tag{E18}$$

Thus C_{LHST} exhibits strong-OPR to Noise Model 2 if we assume that the coefficients $\tau^{(j)}_{a_j,a_j,b_j,b_j}$ characterizing the noise satisfy $\tau^{(j)}_{a_j,a_j,b_j,b_j} \geqslant 0$. $\qquad\square$

Appendix F. Proof of theorem 3

Theorem 3. *The cost functions C_{LET} and C_{LLET} exhibit weak-OPR, as defined in definition 6, to Noise Model 3 in definition 9.*

Proof. Let us remark that in order to show weak-OPR to Noise Model 3 we just need to consider Pauli noise acting at τ_1 and measurement noise, since noise resilience to global depolarizing noise follows from lemma 2.

We first consider the C_{LET} cost function. From equations (A5) and (A6) we get that the action of the Pauli channel acting at time τ_1 is given by

$$\mathcal{P}(|\mathbf{0}\rangle\langle\mathbf{0}|) = \sum_{l,k} q_{l,k} X^l Z^k |\mathbf{0}\rangle\langle\mathbf{0}| Z^k X^l = \sum_l q_l |l\rangle\langle l|, \tag{F1}$$

where $q_l = \sum_k q_{l,k}$. Similarly, we can express the noisy measurement POVM from definition 5 as

$$\widetilde{P}_{\mathbf{0}} = \bigotimes_{j=1}^{n} (p^{(j)}_{00}|0\rangle\langle 0| + p^{(j)}_{01}|1\rangle\langle 1|) = \sum_i p_i |i\rangle\langle i|, \tag{F2}$$

with $\mathbf{i} = i_1 i_2 \dots i_n$ a bit string and $p_i = p^{(1)}_{0i_1} p^{(2)}_{0i_2} \dots p^{(n)}_{0i_n}$. For the present noise model we are interested in determining the optimum of the function

$$\widetilde{G}_{\text{LET}}(V) = \text{Tr}[\widetilde{P}_{\mathbf{0}}(\mathcal{W} \circ \mathcal{P})(|\mathbf{0}\rangle\langle\mathbf{0}|)], \tag{F3}$$

with $\mathcal{W} = \mathcal{V}^\dagger \circ \mathcal{U}$ the channel that implements U followed by $V^\dagger$. Then, by means of (F1) and (F2) we find

$$\widetilde{G}_{\text{LET}}(V) = \text{Tr}\left[\left(\sum_i p_i |i\rangle\langle i|\right)\left(\sum_l q_l W|l\rangle\langle l|W^\dagger\right)\right] = \sum_{i,l} p_i q_l w_{il}, \tag{F4}$$

where $w_{il} = |\langle i|W|l\rangle|^2$ are the matrix elements of a doubly stochastic matrix such that $\sum_i w_{il} = \sum_l w_{il} = 1$.

Let us now denote by $\mathbf{q}^\downarrow$ the vector with elements q_i ordered in decreasing order. Similarly, we denote by $\mathbf{p}^\downarrow$ the vector with elements p_l ordered in decreasing order. Additionally, let $\{|q_r\rangle\}$ and $\{|p_s\rangle\}$ be the basis in which $\mathbf{q}^\downarrow$ and $\mathbf{p}^\downarrow$ are ordered, respectively, i.e.

$$\mathcal{P}(|\mathbf{0}\rangle\langle\mathbf{0}|) = \sum_r q^\downarrow_r |q_r\rangle\langle q_r|, \quad \text{and} \quad \widetilde{P}_{\mathbf{0}} = \sum_s p^\downarrow_s |p_s\rangle\langle p_s|. \tag{F5}$$

Then, from the permutation inequality (or the rearrangement inequality) [48] we have

$$\widetilde{G}_{\text{LET}}(V) = \sum_{i,l} p_i q_l w_{il} \leqslant \mathbf{p}^\downarrow \cdot \mathbf{q}^\downarrow. \tag{F6}$$

The inequality in (F6) is saturated for matrices $W \in \mathbb{S}$, where $\mathbb{S}$ is the subset of the Permutation Group which maps $\{|p_s\rangle\}$ to $\{|q_r\rangle\}$. We remark here that if the vector $\mathbf{q}^\downarrow$ (or $\mathbf{p}^\downarrow$) has components of equal magnitude, then the set $\mathbb{S}$ is degenerate. Moreover, note that

$$p_{\mathbf{0}} \geqslant p_i, \quad \text{and} \quad q_{\mathbf{0}} \geqslant q_i, \quad \forall i \neq \mathbf{0}, \tag{F7}$$

where the second inequality follows from definition 2, while the first inequality always holds since $p_{\mathbf{0}} = \prod_{j=1}^{n} p^{(j)}_{00}$, and since we have assumed that $p^{(j)}_{00} > p^{(j)}_{01} \, \forall j$.

We now recall that $\mathbb{V}^{\text{opt}}_d$ denotes the set of unitaries that optimize $C_{\text{LET}}(V)$ and $C_{\text{LLET}}(V)$, i.e. $\forall V' \in \mathbb{V}^{\text{opt}}_d$ we have $W'|\mathbf{0}\rangle = (V')^\dagger U|\mathbf{0}\rangle = |\mathbf{0}\rangle$ (up to a global phase), which entails $w'_{i\mathbf{0}} = w'_{\mathbf{0}i} = \delta_{i,\mathbf{0}}$, and hence equation (F4) becomes

$$\widetilde{G}_{\text{LET}}(V') = p_{\mathbf{0}} q_{\mathbf{0}} + \sum_{i,l \neq \mathbf{0}} p_i q_l w'_{il}. \tag{F8}$$

Since $p_{\mathbf{0}} \geqslant p_i$ and $q_{\mathbf{0}} \geqslant q_i \, \forall i$ then the first term in (F8) corresponds to the first term in the summation $\mathbf{p}^\downarrow \cdot \mathbf{q}^\downarrow = \sum_r q^\downarrow_r p^\downarrow_r$. Hence, in order to saturate (F6) we now need that $W' \in \mathbb{S}$, i.e. the $(n-1) \times (n-1)$ principal submatrix of W' with matrix elements $\langle z|W'|z'\rangle$ (such that $z, z' \neq \mathbf{0}$) must map $\{|p_s\rangle\}$ to $\{|q_r\rangle\}$

(where $s \neq 0$ and $r \neq 0$). Combining this result with (F6) we have that for any matrix V in $\mathbb{V}_d$ (the set of $d \times d$ unitary matrices)

$$\widetilde{G}_{\mathsf{LET}}(V) \leqslant \boldsymbol{p}^{\downarrow} \cdot \boldsymbol{q}^{\downarrow} = \widetilde{G}_{\mathsf{LET}}(V'), \tag{F9}$$

where $V' \in \widetilde{\mathbb{V}}_d^{\mathrm{opt}}$ and where

$$\widetilde{\mathbb{V}}_d^{\mathrm{opt}} = \{V' \in \mathbb{V}_d \colon W = (V')^{\dagger}U \in \mathbb{S}\}. \tag{F10}$$

Evidently, not all matrices in $\mathbb{V}_d^{\mathrm{opt}}$ are in $\mathbb{S}$, which then entails that $\widetilde{\mathbb{V}}_d^{\mathrm{opt}} \subseteq \mathbb{V}_d^{\mathrm{opt}}$, and further means that C_{LET} exhibits weak-OPR to Noise Model 3 according to definition 6.

Let us now consider the noise resilience of LLET to Noise Model 3 of definition 9. We are now interested in the optimum of

$$\widetilde{G}_{\mathsf{LLET}}(V) = \frac{1}{n}\sum_{j=1}^{n}\mathrm{Tr}\,[(p_{00}^{(j)}|0\rangle\langle0| + p_{01}^{(j)}|1\rangle\langle1|) \otimes \mathbb{1}^{\overline{A}_j})(\mathcal{W}\circ\mathcal{P})(|\boldsymbol{0}\rangle\langle\boldsymbol{0}|)] \tag{F11}$$

$$= \frac{1}{n}\sum_{j=1}^{n}\mathrm{Tr}\left[(p_{00}^{(j)}|0\rangle\langle0| + p_{01}^{(j)}|1\rangle\langle1|) \otimes \mathbb{1}^{\overline{A}_j})(\sum_{l}q_l W|\boldsymbol{l}\rangle\langle\boldsymbol{l}|W^{\dagger})\right]. \tag{F12}$$

For any matrix $V' \in \mathbb{V}_d^{\mathrm{opt}}$ we have $W'|\boldsymbol{0}\rangle = (V')^{\dagger}U|\boldsymbol{0}\rangle = |\boldsymbol{0}\rangle$ (up to global phase) and $\sum_l q_l W'|\boldsymbol{l}\rangle\langle\boldsymbol{l}|(W')^{\dagger} = q_0|\boldsymbol{0}\rangle\langle\boldsymbol{0}| + \sum_{l\neq0}q_l W'|\boldsymbol{l}\rangle\langle\boldsymbol{l}|(W')^{\dagger}$, which leads to

$$\widetilde{G}_{\mathsf{LLET}}(V') = \frac{1}{n}\sum_{j=1}^{n}p_{00}^{(j)}q_{\boldsymbol{0}} + \frac{1}{n}\sum_{j=1}^{n}\mathrm{Tr}\left[(p_{00}^{(j)}|0\rangle\langle0| + p_{01}^{(j)}|1\rangle\langle1|) \otimes \mathbb{1}^{\overline{A}_j}))\left(\sum_{l\neq\boldsymbol{0}}q_l W'|\boldsymbol{l}\rangle\langle\boldsymbol{l}|(W')^{\dagger}\right)\right]. \tag{F13}$$

On the other hand, for any unitary matrix $V \in \mathbb{V}_d$

$$\begin{aligned}
\widetilde{G}_{\mathsf{LLET}}(V) &= \frac{1}{n}\sum_{j=1}^{n}\mathrm{Tr}\,[(p_{00}^{(j)}|0\rangle\langle0| + p_{01}^{(j)}|1\rangle\langle1|) \otimes \mathbb{1}^{\overline{A}_j})q_{\boldsymbol{0}} W|\boldsymbol{0}\rangle\langle\boldsymbol{0}|W^{\dagger}] \\
&\quad + \frac{1}{n}\sum_{j=1}^{n}\mathrm{Tr}\left[(p_{00}^{(j)}|0\rangle\langle0| + p_{01}^{(j)}|1\rangle\langle1|) \otimes \mathbb{1}^{\overline{A}_j})(\sum_{l\neq\boldsymbol{0}}q_l W|\boldsymbol{l}\rangle\langle\boldsymbol{l}|W^{\dagger})\right] \\
&\leqslant \frac{1}{n}\sum_{j=1}^{n}\left(\mathrm{Tr}\,[p_{00}^{(j)}q_{\boldsymbol{0}}\mathbb{1}W|\boldsymbol{0}\rangle\langle\boldsymbol{0}|W^{\dagger}] + \mathrm{Tr}\left[(p_{00}^{(j)}|0\rangle\langle0| + p_{01}^{(j)}|1\rangle\langle1|) \otimes \mathbb{1}^{\overline{A}_j})(\sum_{l\neq\boldsymbol{0}}q_l W|\boldsymbol{l}\rangle\langle\boldsymbol{l}|W^{\dagger})\right]\right) \\
&= \frac{1}{n}\sum_{j=1}^{n}p_{00}^{(j)}q_{\boldsymbol{0}} + \frac{1}{n}\sum_{j=1}^{n}\mathrm{Tr}\left[(p_{00}^{(j)}|0\rangle\langle0| + p_{01}^{(j)}|1\rangle\langle1|) \otimes \mathbb{1}^{\overline{A}_j})(\sum_{l\neq\boldsymbol{0}}q_l W|\boldsymbol{l}\rangle\langle\boldsymbol{l}|W^{\dagger})\right],
\end{aligned} \tag{F14}$$

where the inequality follows from the fact that $p_{00}^{(j)} > p_{01}^{(j)}$, and hence

$$(p_{00}^{(j)}|0\rangle\langle0| + p_{01}^{(j)}|1\rangle\langle1|) \otimes \mathbb{1}^{\overline{A}_j} \leqslant (p_{00}^{(j)}|0\rangle\langle0| + p_{00}^{(j)}|1\rangle\langle1|) \otimes \mathbb{1}^{\overline{A}_j} \leqslant p_{00}^{(j)}\mathbb{1}. \tag{F15}$$

We can then simplify equation (F14) as

$$\widetilde{G}_{\mathsf{LLET}}(V) \leqslant \frac{1}{n}\sum_{j=1}^{n}p_{00}^{(j)}q_{\boldsymbol{0}} + \frac{1}{n}\sum_{j=1}^{n}\sum_{l\neq\boldsymbol{0},k\neq\boldsymbol{0}}q_l p_k^{(j)}w_{kl} = \frac{1}{n}\sum_{j=1}^{n}p_{00}^{(j)}q_{\boldsymbol{0}} + \sum_{l\neq\boldsymbol{0},k\neq\boldsymbol{0}}q_l\tilde{p}_k w_{kl}, \tag{F16}$$

where we have $p_k^{(j)} = p_{00}^{(j)}$ if $k_j = 0$, and $p_k^{(j)} = p_{01}^{(j)}$ if $k_j = 1$. On the the other hand, in the second equality of (F16) we have defined $\tilde{p}_k = \frac{1}{n}\sum_{j=1}^{n}p_k^{(j)}$. Finally, the following inequality follows again from the rearrangement inequality

$$\widetilde{G}_{\mathsf{LLET}}(V) \leqslant \frac{1}{n}\sum_{j=1}^{n}p_{00}^{(j)}q_{\boldsymbol{0}} + \sum_{l\neq\boldsymbol{0},k\neq\boldsymbol{0}}q_l^{\downarrow}\tilde{p}_k^{\downarrow}, \tag{F17}$$

which is saturated for matrices $W \in \mathbb{S}'$, where $\mathbb{S}'$ is a subset of the Permutation Group such that $\sum_{l\neq\boldsymbol{0},k\neq\boldsymbol{0}}q_l\tilde{p}_k w_{kl} = \sum_{l\neq\boldsymbol{0},k\neq\boldsymbol{0}}q_l^{\downarrow}\tilde{p}_k^{\downarrow}$. Here $q^{\downarrow}$ and $\tilde{p}^{\downarrow}$ are vectors with components q_l and $\tilde{p}_k$ in decreasing order, respectively. Hence, we can define the set of matrices which saturate (F17) as

$$\widetilde{\mathbb{V}}_d^{\mathrm{opt}} = \{V' \in \mathbb{V}_d \colon W = (V')^{\dagger}U \in \mathbb{S}'\}. \tag{F18}$$

While any matrix in $\mathbb{V}_d^{\mathrm{opt}}$ saturates the inequality in (F14), only a subset will also saturate (F17). Hence, $\widetilde{\mathbb{V}}_d^{\mathrm{opt}} \subseteq \mathbb{V}_d^{\mathrm{opt}}$, and C_{LLET} exhibits weak-OPR to Noise Model 3 according to definition 6.   □

Appendix G. Proof of corollaries 1–8

Corollary 1. *The cost functions C_{HST} and C_{LHST} exhibit strong-OPR to a noise model that includes the following: (1) all noise processes in Noise Model 1, as well as (2) a noise process during the implementation of*

$\mathcal{W} = \mathcal{W}_k \circ \cdots \circ \mathcal{W}_1 = \mathcal{V}^\dagger \circ \mathcal{U}$ (i.e. in the time interval between τ_1 and τ_2) in which global Pauli channels $\{\mathcal{P}_1^A, \ldots,$ $\mathcal{P}_k^A\}$ act on system A, such that the overall channel on A is $\mathcal{P}_k^A \circ \mathcal{W}_k \cdots \circ \mathcal{P}_1^A \circ \mathcal{W}_1$, provided that the following condition is satisfied:

$$(\mathcal{P}_k^A \circ \mathcal{W}_k \cdots \circ \mathcal{P}_1^A \circ \mathcal{W}_1)(\cdot) = (\mathcal{W}_k \circ \mathcal{W}_{k-1} \cdots \circ \mathcal{W}_1 \circ \widehat{\mathcal{P}}^A)(\cdot). \tag{G1}$$

Here $\widehat{\mathcal{P}}^A$ is also a Pauli channel, and the channels $\mathcal{U}$, $\mathcal{V}^\dagger$, and $\mathcal{W}$ correspond to conjugating the state by the unitaries U, $V^\dagger$, and W, respectively.

Proof. This follows from the fact that the overall noisy channel acting during the implementation of $\mathcal{W}$ is mathematically equivalent to a Pauli channel followed by the unitary $\mathcal{W}$, as described in the condition (G1) and by invoking theorem 1, which allows for Pauli channel noise at time τ_1.□

Corollary 2. *Let the* $W = V^\dagger U$ *gate sequence have the form* $W = W_2^A W_1^A$ *with* W_1^A *be composed only of Clifford gates. Then the cost functions* C_{HST} *and* C_{LHST} *exhibit strong-OPR to a noise model that includes the following: (1) all noise processes in Noise Model 1, as well as (2) a noise process during the implementation of* $\mathcal{W}_1^A = \mathcal{W}_{1,k} \circ \cdots \circ \mathcal{W}_{1,1}$, *in which global Pauli channels* $\{\mathcal{P}_1^A, \ldots, \mathcal{P}_k^A\}$ *act on system A, such that the overall channel on A is* $\mathcal{P}_k^A \circ \mathcal{W}_{1,k} \cdots \circ \mathcal{P}_1^A \circ \mathcal{W}_{1,1}$.

Proof. From lemma 1 it follows that Clifford unitaries satisfy the condition in (G1). Therefore, corollary 2 is a special case of corollary 1. □

Corollary 3. *Let the* $W = V^\dagger U$ *gate sequence have the form* $W = W_2^A W_1^A$ *with* $W_1^A = W_1^{A'} \otimes W_1^{A''}$ *being a tensor product, i.e.,* W *is a tensor product up to a particular time. Then the cost functions* C_{HST} *and* C_{LHST} *exhibit strong-OPR to a noise model that includes the following: (1) all noise processes in Noise Model 1, as well as (2) a noise process during the implementations of* $\mathcal{W}_1^{A'} = \mathcal{W}_{1,k}^{A'} \circ \cdots \circ \mathcal{W}_{1,1}^{A'}$ *and* $\mathcal{W}_1^{A''} = \mathcal{W}_{1,l}^{A''} \circ \cdots \circ \mathcal{W}_{1,1}^{A''}$ *in which local depolarizing channels* $\{\mathcal{D}_{1,1}^{A'}, \ldots, \mathcal{D}_{1,k}^{A'}\}$ *and* $\{\mathcal{D}_{1,1}^{A''}, \ldots, \mathcal{D}_{1,l}^{A''}\}$ *act on subsystems* A' *and* A'', *respectively, such that the overall channel on* $A'A''$ *is* $(\mathcal{D}_{1,k}^{A'} \circ \mathcal{W}_{1,k}^{A'} \ldots \mathcal{D}_{1,1}^{A'} \circ \mathcal{W}_{1,1}^{A'}) \otimes (\mathcal{D}_{1,l}^{A''} \circ \mathcal{W}_{1,l}^{A''} \ldots \mathcal{D}_{1,1}^{A''} \circ \mathcal{W}_{1,1}^{A''})$.

Proof. Let ρ denote a quantum state. Consider the following chain of equalities:

$$(\mathcal{D}_p^{A'} \otimes \mathcal{D}_q^{A''})(\mathcal{W}^{A'} \otimes \mathcal{W}^{A''})(\rho) = (\mathcal{I}^{A'} \otimes \mathcal{D}_q^{A''})(p(\mathcal{W}^{A'} \otimes \mathcal{W}^{A''}(\rho)) + (1-p)\pi^{A'} \mathrm{Tr}_{A'}((\mathcal{W}^{A'} \otimes \mathcal{W}^{A''})(\rho)) \tag{G2}$$

$$= (\mathcal{I}^{A'} \otimes \mathcal{D}_q^{A''})(p(\mathcal{W}^{A'} \otimes \mathcal{W}^{A''}(\rho)) + (1-p)\pi^{A'} \mathrm{Tr}_{A'}((\mathcal{I}^{A'} \otimes \mathcal{W}^{A''})(\rho)) \tag{G3}$$

$$= (\mathcal{I}^{A'} \otimes \mathcal{D}_q^{A''})(\mathcal{W}^{A'} \otimes \mathcal{W}^{A''})(p\rho + (1-p)\pi^{A'} \mathrm{Tr}_{A'}(\rho)) \tag{G4}$$

$$= (\mathcal{I}^{A'} \otimes \mathcal{D}_q^{A''})(\mathcal{W}^{A'} \otimes \mathcal{W}^{A''})(\mathcal{D}_p^{A'}(\rho)) \tag{G5}$$

$$= (\mathcal{W}^{A'} \otimes \mathcal{W}^{A''})(\mathcal{D}_p^{A'} \otimes \mathcal{D}_q^{A''})(\rho), \tag{G6}$$

where $\pi^{A'}$ is a maximally mixed state on system A'. Therefore, the result follows by applying (G6) several times and invoking corollary 1. □

Corollary 4. *The cost functions* C_{HST} *and* C_{LHST} *exhibit strong-OPR to the following noise model: (1) all noise processes in Noise Model 2, as well as (2) a noise process during the implementation of* $\mathcal{W} = \mathcal{W}_k \circ \cdots \circ \mathcal{W}_1 = \mathcal{V}^\dagger \circ \mathcal{U}$ *(i.e. in the time interval between* τ_1 *and* τ_2*) in which global non-unital Pauli channels* $\{\mathcal{P}_{\mathrm{NU},1}^A, \ldots, \mathcal{P}_{\mathrm{NU},k}^A\}$ *act on system A such that the overall channel on A is* $\mathcal{P}_{\mathrm{NU},k}^A \circ \mathcal{W}_k \cdots \circ \mathcal{P}_{\mathrm{NU},1}^A \circ \mathcal{W}_1$, *provided that the following condition is satisfied:*

$$(\mathcal{P}_{\mathrm{NU},k}^A \circ \mathcal{W}_k \cdots \circ \mathcal{P}_{\mathrm{NU},1}^A \circ \mathcal{W}_1)(\cdot) = (\mathcal{W}_k \circ \mathcal{W}_{k-1} \cdots \mathcal{W}_1 \circ \widehat{\mathcal{P}}_{\mathrm{NU}}^A)(\cdot), \tag{G7}$$

where $\widehat{\mathcal{P}}_{\mathrm{NU}}^A$ *is also a Pauli channel.*

Proof. This follows from the fact that the overall noisy channel acting during the implementation of $\mathcal{W}$ is mathematically equivalent to a non-unital Pauli channel followed by the unitary $\mathcal{W}$, as described in the condition (G7) and by invoking theorem 2, which allows for non-unital Pauli noise at time τ_1.□

Corollary 5. *The cost functions* C_{HST} *exhibits strong-OPR to the following noise model: (1) global depolarizing noise acting continuously throughout the circuit, (2) global non-unital Pauli noise on system* A *at a fixed time in between* τ_1 *and* τ_2.

Proof. Let us decompose $\mathcal{W}$ as $\mathcal{W} = \mathcal{W}_2 \circ \mathcal{W}_1$ such that the non-unital Pauli channel $\mathcal{P}_{NU}^A$ acts at time τ' between $\mathcal{W}_1$ and $\mathcal{W}_2$, with the overall channel between τ_1 and τ_2 given by $\mathcal{W}_2 \circ \mathcal{P}_{NU}^A \circ \mathcal{W}_1$. The state at time τ_1 is

$$\rho^{(1)} = p^{(1)}|\Phi^+\rangle\langle\Phi^+| + (1 - p^{(1)})\mathbb{1}/d, \tag{G8}$$

where $p^{(1)} = p^{(k,1)} \cdots p^{(1,1)}$ corresponds to the continuous depolarizing channel as discussed in appendix D. We break up the time interval in between τ' and τ_1 into l steps. The state at time τ' is given by

$$\widetilde{\rho}^{(2)} = \mathcal{P}_{NU}^A \circ \mathcal{D}_{q^{(2,l)}}^{AB} \circ \mathcal{W}_1^l \cdots \circ \mathcal{D}_{q^{(2,l)}}^{AB} \circ \mathcal{W}_1^1(\rho^{(1)}) \tag{G9}$$

$$= \mathcal{P}_{NU}^A(p^{(1)}q^{(2)}\mathcal{W}_1(|\Phi^+\rangle\langle\Phi^+|) + (1 - p^{(1)}q^{(2)})\mathbb{1}/d) \tag{G10}$$

$$= p^{(1)}q^{(2)}\mathcal{P}_{NU}^A(\mathcal{W}_1(|\Phi^+\rangle\langle\Phi^+|)) + (1 - p^{(1)}q^{(2)})\mathbb{1}/d + (1 - p^{(1)}q^{(2)})\frac{1}{d}\sum_{(g,h)\neq(0,0)} d_{g,h}X_A^g Z_A^h \otimes \mathbb{1}_B, \tag{G11}$$

where $q^{(2)} = q^{(2,k)} \cdots q^{(2,1)}$ and $\mathcal{W}_1 = \mathcal{W}_1^l \cdots \mathcal{W}_1^1$. Similarly, we break up the the time interval between τ_2 and τ' into m steps. The term that depends on $\mathcal{W}$ at time τ_2 is given by

$$\widetilde{\sigma}^{(2)} = p^{(1)}q^{(2)}r^{(2)}\mathcal{W}_2 \circ \mathcal{P}_{NU}^A \circ \mathcal{W}_1(|\Phi^+\rangle\langle\Phi^+|) + r^{(2)}(1 - p^{(1)}q^{(2)})\frac{1}{d}\sum_{(g,h)\neq(0,0)} d_{g,h}W_2 X_A^g Z_A^h W_2^\dagger \otimes \mathbb{1}_B. \tag{G12}$$

Let

$$\widetilde{F}_{HST}(V) \propto f(V) := \mathrm{Tr}[|\Phi^+\rangle\langle\Phi^+|\widetilde{\sigma}^{(2)}]. \tag{G13}$$

Moreover, for simplicity we denote

$$f_1(V) := \mathrm{Tr}[|\Phi^+\rangle\langle\Phi^+|(\mathcal{W}_2 \circ \mathcal{P}_{NU}^A \circ \mathcal{W}_1)(|\Phi^+\rangle\langle\Phi^+|)], \tag{G14}$$

$$f_2(V) := \mathrm{Tr}[|\Phi^+\rangle\langle\Phi^+|(W_2 X_A^g Z_A^h W_2^\dagger \otimes \mathbb{1}_B)]. \tag{G15}$$

Consider the followings:

$$f_1(V) = \mathrm{Tr}[|\Phi^+\rangle\langle\Phi^+|(\mathcal{W}_2^A \circ \mathcal{P}_{NU}^A)((\mathcal{I}_A \otimes (\mathcal{W}_1^T)^B)(|\Phi^+\rangle\langle\Phi^+|_{AB}))] \tag{G16}$$

$$= \mathrm{Tr}[(\mathcal{I}_A \otimes (\mathcal{W}_1^*)^B)(|\Phi^+\rangle\langle\Phi^+|)(\mathcal{W}_2^A \circ \mathcal{P}_{NU}^A)(|\Phi^+\rangle\langle\Phi^+|)] \tag{G17}$$

$$= \mathrm{Tr}[((\mathcal{W}_1^\dagger)^A \otimes \mathcal{I}_B)(|\Phi^+\rangle\langle\Phi^+|)(\mathcal{W}_2^A \circ \mathcal{P}_{NU}^A)(|\Phi^+\rangle\langle\Phi^+|)] \tag{G18}$$

$$= \mathrm{Tr}[|\Phi^+\rangle\langle\Phi^+|(\mathcal{W}_1^A \circ \mathcal{W}_2^A \circ \mathcal{P}_{NU}^A)(|\Phi^+\rangle\langle\Phi^+|)] \tag{G19}$$

$$\leqslant f_1(V'), \tag{G20}$$

where $V' \in \mathbb{V}_d^{opt}$, and where $\mathbb{V}_d^{opt}$ denote the sets of unitaries that optimize $F_{HST}(V)$ (and hence $C_{HST}(V)$) as defined in (D18). The first and third equalities follow from the ricochet property. The last equality corresponds to the case when there is non-unital Pauli noise at time τ_1 and no other noise in the HST circuit, which is a special case of theorem 2. Therefore, the inequality follows from theorem 2. Moreover, by using the arguments similar to (E10)–(E12), we find that $f_2(V)$ is independent of W. This completes the proof. $\qquad\square$

Corollary 6. *The cost functions C_{LET} and C_{LLET} exhibit weak-OPR to a noise model that includes the following: (1) all noise processes in Noise Model 3, as well as (2) a noise process during the implementation of $\mathcal{W} = \mathcal{W}_k \circ \cdots \circ \mathcal{W}_1 = \mathcal{V}^\dagger \circ \mathcal{U}$ in which global Pauli channels $\{\mathcal{P}_1, \ldots, \mathcal{P}_k\}$ act, such that the overall channel is $\mathcal{P}_k \circ \mathcal{W}_k \cdots \circ \mathcal{P}_1 \circ \mathcal{W}_1$, provided that the following condition is satisfied:*

$$(\mathcal{P}_k \circ \mathcal{W}_k \cdots \circ \mathcal{P}_1 \circ \mathcal{W}_1)(\cdot) = (\mathcal{W}_k \circ \mathcal{W}_{k-1} \cdots \circ \mathcal{W}_1 \circ \widehat{\mathcal{P}})(\cdot). \tag{G21}$$

where $\widehat{\mathcal{P}}$ is also a Pauli channel.

Proof. This follows from arguments similar to corollary 1 and by invoking theorem 3. $\qquad\square$

Corollary 7. *Let the $W = V^\dagger U$ gate sequence have the form $W = W_2^A W_1^A$ with W_1^A be composed only of Clifford gates. Then the cost functions C_{LET} and C_{LLET} exhibit weak-OPR to a noise model that includes the following: (1) all noise processes in Noise Model 3, as well as (2) a noise process during the implementation of $\mathcal{W}_1^A = \mathcal{W}_{1,k} \circ \cdots \circ \mathcal{W}_{1,1}$, in which global Pauli channels $\{\mathcal{P}_1^A, \ldots, \mathcal{P}_k^A\}$ act on system A, such that the overall channel on A is $\mathcal{P}_k^A \circ \mathcal{W}_{1,k} \cdots \circ \mathcal{P}_1^A \circ \mathcal{W}_{1,1}$.*

Proof. This corollary is a special case of corollary 6, since lemma 1 implies that Clifford unitaries satisfy (G21). $\qquad\square$

Corollary 8. *Let the $W = V^\dagger U$ gate sequence have the form $W = W_2^A W_1^A$ with $W_1^A = W_1^{A'} \otimes W_1^{A''}$ being a tensor product,* i.e. *W is a tensor product up to a particular time. Then the cost functions C_{LET} and C_{LLET} exhibit weak-OPR to a noise model that includes the following: (1) all noise processes in Noise Model 3, as well as (2) a noise process during the implementations of $\mathcal{W}_1^{A'} = \mathcal{W}_{1,k}^{A'} \circ \cdots \circ \mathcal{W}_{1,1}^{A'}$ and $\mathcal{W}_1^{A''} = \mathcal{W}_{1,l}^{A''} \circ \cdots \circ \mathcal{W}_{1,1}^{A''}$ in which local depolarizing channels $\{\mathcal{D}_{1,1}^{A'}, \ldots, \mathcal{D}_{1,k}^{A'}\}$ and $\{\mathcal{D}_{1,1}^{A''}, \ldots, \mathcal{D}_{1,l}^{A''}\}$ act on subsystems A' and A'', respectively, such that the overall channel on $A'A''$ is $(\mathcal{D}_{1,k}^{A'} \circ \mathcal{W}_{1,k}^{A'} \ldots \mathcal{D}_{1,1}^{A'} \circ \mathcal{W}_{1,1}^{A'}) \otimes (\mathcal{D}_{1,l}^{A''} \circ \mathcal{W}_{1,l}^{A''} \ldots \mathcal{D}_{1,1}^{A''} \circ \mathcal{W}_{1,1}^{A''})$.*

Proof. This follows from arguments similar to the proof of corollary 3 and by invoking corollary 6. □

ORCID iDs

Kunal Sharma ⓘ https://orcid.org/0000-0003-3132-1088
M Cerezo ⓘ https://orcid.org/0000-0002-2757-3170

References

[1] Preskill J 2018 Quantum computing in the NISQ era and beyond *Quantum* **2** 79
[2] Temme K, Bravyi S and Gambetta J M 2017 Error mitigation for short-depth quantum circuits *Phys. Rev. Lett.* **119** 180509
[3] Linke N M, Johri S, Figgatt C, Landsman K A, Matsuura A Y and Monroe C 2018 Measuring the renyi entropy of a two-site fermi-hubbard model on a trapped ion quantum computer *Phys. Rev.* A **98** 052334
[4] Subaşı Y, Cincio L and Coles P J 2019 Entanglement spectroscopy with a depth-two quantum circuit *J. Phys. A: Math. Theor.* **52** 044001
[5] Murali P, Baker J M, Javadi-Abhari A, Chong F T and Martonosi M 2019 Noise-adaptive compiler mappings for noisy intermediate-scale quantum computers *Proc. Twenty-Fourth Int. Conf. on Architectural Support for Programming Languages and Operating Systems* (New York: ACM) pp 1015–29
[6] Cincio L, Subaşı Y, Sornborger A T and Coles P J 2018 Learning the quantum algorithm for state overlap *New J. Phys.* **20** 113022
[7] McClean J R, Romero J, Babbush R and Aspuru-Guzik A 2016 The theory of variational hybrid quantum–classical algorithms *New J. Phys.* **18** 023023
[8] Peruzzo A, McClean J, Shadbolt P, Yung M-H, Zhou X-Q, Love P J, Aspuru-Guzik A and O'Brien J L 2014 A variational eigenvalue solver on a photonic quantum processor *Nat. Commun.* **5** 4213
[9] Farhi E, Goldstone J and Gutmann S 2014 A quantum approximate optimization algorithm arXiv:1411.4028
[10] Johnson P D, Romero J, Olson J, Cao Y and Aspuru-Guzik A 2017 QVECTOR: an algorithm for device-tailored quantum error correction arXiv:1711.02249
[11] Romero J, Olson J P and Aspuru-Guzik A 2017 Quantum autoencoders for efficient compression of quantum data *Quantum Sci. Technol.* **2** 045001
[12] LaRose R, Tikku A, O'Neel-Judy É, Cincio L and Coles P J 2019 Variational quantum state diagonalization *npj Quantum Inf.* **5** 8
[13] Arrasmith A, Cincio L, Sornborger A T, Zurek W H and Coles P J 2019 Variational consistent histories as a hybrid algorithm for quantum foundations *Nat. Commun.* **10** 3438
[14] Cerezo M, Poremba A, Cincio L and Coles P J 2019 Variational quantum fidelity estimation arXiv:1906.09253
[15] Jones T, Endo S, McArdle S, Yuan X and Benjamin S C 2019 Variational quantum algorithms for discovering hamiltonian spectra *Phys. Rev.* A **99** 062304
[16] Yuan X, Endo S, Zhao Q, Benjamin S and Li Y 2019 *Quantum* **3** 191
[17] Li Y and Benjamin S C 2017 Efficient variational quantum simulator incorporating active error minimization *Phys. Rev.* X **7** 021050
[18] Kokail C *et al* 2019 Self-verifying variational quantum simulation of lattice models *Nature* **569** 355
[19] Khatri S, LaRose R, Poremba A, Cincio L, Sornborger A T and Coles P J 2019 Quantum-assisted quantum compiling *Quantum* **3** 140
[20] Jones T and Benjamin S C 2018 Quantum compilation and circuit optimisation via energy dissipation arXiv:1811.03147
[21] Heya K, Suzuki Y, Nakamura Y and Fujii K 2018 Variational quantum gate optimization arXiv:1810.12745
[22] Carolan J *et al* 2020 Variational quantum unsampling on a quantum photonic processor *Nat. Phys.* **16** 322–7
[23] Devitt S J, Munro W J and Nemoto K 2013 Quantum error correction for beginners *Rep. Prog. Phys.* **76** 076001
[24] Fowler A G, Mariantoni M, Martinis J M and Cleland A N 2012 Surface codes: towards practical large-scale quantum computation *Phys. Rev.* A **86** 032324
[25] Chong F T, Franklin D and Martonosi M 2017 Programming languages and compiler design for realistic quantum hardware *Nature* **549** 180
[26] Häner T, Steiger D S, Svore K and Troyer M 2018 A software methodology for compiling quantum programs *Quantum Sci. Technol.* **3** 020501
[27] Venturelli D, Do M, Rieffel E and Frank J 2018 Compiling quantum circuits to realistic hardware architectures using temporal planners *Quantum Sci. Technol.* **3** 025004
[28] Cross A W, Bishop L S, Smolin J A and Gambetta J M 2017 Open quantum assembly language arXiv:1707.03429
[29] Horodecki M, Horodecki P and Horodecki R 1999 General teleportation channel, singlet fraction, and quasidistillation *Phys. Rev.* A **60** 1888
[30] Nielsen M A 2002 A simple formula for the average gate fidelity of a quantum dynamical operation *Phys. Lett.* A **303** 249–52
[31] Cerezo M, Sone A, Volkoff T, Cincio L and Coles P J 2020 Cost-function-dependent barren plateaus in shallow quantum neural networks arXiv:2001.00550
[32] Goussev A, Jalabert R A, Pastawski H M and Wisniacki D A 2012 Loschmidt echo *Scholarpedia* **7** 11687
[33] Nielsen M A and Chuang I L 2010 *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press)
[34] Wilde M M 2017 *Quantum Information Theory* 2nd edn (Cambridge: Cambridge University Press)
[35] Shi Y 2003 Both toffoli and controlled-not need little help to do universal quantum computing *Quantum Inf. Comput.* **3** 84–92
[36] Shor P 1997 Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer *SIAM J. Comput.* **26** 1484–509

[37] Wang Z, Rubin N C, Dominy J M and Rieffel E G 2020 *Phys. Rev.* A **101** 012320

[38] Shende V V and Markov I L 2009 On the CNOT-cost of toffoli gates *Quantum Inf. Comput.* **9** 0461–86

[39] Bärtschi A and Eidenbenz S 2019 Deterministic preparation of Dicke states *Fundamentals of Computation Theory* vol 11651 (Berlin: Springer) (https://doi.org/10.1007/978-3-030-25027-0_9)

[40] Cruz D *et al* 2019 Efficient quantum algorithms for ghz and w states, and implementation on the ibm quantum computer *Adv. Quantum Technol.* **2** 1900015

[41] Aleksandrowicz G *et al* 2019 Qiskit: An Open-source Framework for Quantum Computing

[42] Qiskit 2019 Qiskit/qiskit-tutorials

[43] Mitarai K, Negoro M, Kitagawa M and Fujii K 2018 Quantum circuit learning *Phys. Rev.* A **98** 032309

[44] Kübler J M, Arrasmith A, Cincio L and Coles P J 2019 An adaptive optimizer for measurement-frugal variational algorithms arXiv:1909.09083

[45] Gentini L, Cuccoli A, Pirandola S, Verrucchi P and Banchi L 2019 Noise-assisted variational hybrid quantum-classical optimization arXiv:1912.06744

[46] Khodjasteh K and Viola L 2009 Dynamically error-corrected gates for universal quantum computation *Phys. Rev. Lett.* **102** 080501

[47] Bravo-Prieto C, LaRose R, Cerezo M, Subaşı Y, Cincio L and Coles P J 2019 arXiv:1909.05820Variational quantum linear solver: a hybrid algorithm for linear systems

[48] Hardy G H and Littlewood J E 1952 Karreman mathematics research collection *Inequalities, Cambridge Mathematical Library* ed G Pólya *et al* (Cambridge: Cambridge University Press)