Improving the speed of variational quantum algorithms for quantum error correction

Fabio Zoratti, Giacomo De Palma, Bobak Kiani, Quynh T. Nguyen, Milad Marvian, Seth Lloyd, and Vittorio Giovannetti

Improving the speed of variational quantum algorithms
for quantum error correction

Fabio Zoratti
Scuola Normale Superiore, I-56126 Pisa, Italy

Giacomo De Palma
Department of Mathematics, University of Bologna, 40126 Bologna, Italy

Bobak Kiani and Quynh T. Nguyen
Department of Electrical Engineering and Computer Science,
MIT, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

Milad Marvian
Department of Electrical and Computer Engineering and Center for
Quantum Information and Control, University of New Mexico, USA

Seth Lloyd
Department of Mechanical Engineering, MIT 77 Massachusetts Avenue, Cambridge, MA 02139, USA
Turing Inc., Cambridge, MA 02139, USA

Vittorio Giovannetti
NEST, Scuola Normale Superiore and Istituto Nanoscienze-CNR, I-56126 Pisa, Italy

We consider the problem of devising a suitable Quantum Error Correction (QEC) procedures for a generic quantum noise acting on a quantum circuit. In general, there is no analytic universal procedure to obtain the encoding and correction unitary gates, and the problem is even harder if the noise is unknown and has to be reconstructed. The existing procedures rely on Variational Quantum Algorithms (VQAs) and are very difficult to train since the size of the gradient of the cost function decays exponentially with the number of qubits. We address this problem using a cost function based on the Quantum Wasserstein distance of order 1 (QW_1). At variance with other quantum distances typically adopted in quantum information processing, QW_1 lacks the unitary invariance property which makes it a suitable tool to avoid to get trapped in local minima. Focusing on a simple noise model for which an exact QEC solution is known and can be used as a theoretical benchmark, we run a series of numerical tests that show how, guiding the VQA search through the QW_1, can indeed significantly increase both the probability of a successful training and the fidelity of the recovered state, with respect to the results one obtains when using conventional approaches.

I. INTRODUCTION

Performing reliable computations on physically imperfect hardware is something that has become usual nowadays, given the current state of classical computers, which can produce perfect results without any software-side mitigation of the imperfections of the physical media where the computation happens. Error correction is based on the fact that these machines automatically perform, on the hardware side, procedures that allow errors to happen and to be fixed without any intervention from the end user. This kind of setting is even more crucial in a quantum scenario where the current noisy intermediate-scale quantum computers (NISQ) have a much larger error rate than their classical counterparts [1]. Performing reliable computations with a trustworthy error correction procedure has direct implications not only in quantum computation [2, 3], but potentially also in all the other sectors of quantum technology which indirectly relay on it (e.g. quantum communication or quantum key distribution [4–6]).

In the typical Quantum Error Correction (QEC) scheme, the quantum information that has to be protected is stored in a subspace of a larger Hilbert space, using an *encoding* procedure. Stabilizer codes [7, 8], which are within the best analytical results in this field, are not universal because they are tailored for a generic noise acting on a small but unknown subset of qubits. Several attempts have already been made to create a numerical optimization procedure to find an error correction code for specific noise models [9–12], but these studies are not universal because they rely heavily on the type of noise on the specific quantum circuit and this is a problem because real quantum devices are not characterized by a single kind of quantum noise. Some attempts have been made to characterize the noise of the current and near-term devices [13, 14], but these methods will become very difficult to implement soon because classical computers are not able to simulate efficiently quantum circuits when the number of qubits increases. Near-term devices with approximately 50 qubits may already be intractable to simulate for supercomputers [15].

If we define a figure of merit of the quality of the state after the action of the noise and its corresponding correction, the obvious choice for the kind of maximization algorithm is a Variational Quantum Algorithm (VQA) [16]. These are hybrid algorithms that couple a quantum computer with a classical one. In this setting, usually, a parametric quantum circuit is applied to some reference state, some measurements are performed on the system, and the outcomes are given to the classical computer to perform a minimization procedure of a given cost function (from this point of view the optimization procedure in a VQA can be seen as the training phase in machine learning). Some examples of this class of algorithms are the variational quantum eigensolver [17] and the Quantum Approximate Optimization Algorithm [18]. Proposals to use VQAs to address QEC problems are already present in literature [19]. Unfortunately, VQAs usually suffer from the phenomenon of barren plateaus [20, 21], namely the gradient of the cost function decays exponentially with respect to the number of qubits of the system, leading to an untrainable model. The fundamental theoretical reason for such behavior has been associated with the presence of *barren plateaus* which originate when the cost function of the problem is global, i.e. mediated by a highly non-local operator [21]. To avoid these effects we propose here to guide the VQA search using cost functions inspired to Quantum Wasserstein distance of order 1 (or QW_1 in brief) introduced in Ref. [22] as a quantum generalization of the Hamming distance [23] on the set of bit strings. As will detail in the following, at variance with more conventional quantum distances typically adopted in quantum information, QW_1 is lacking a fundamental symmetry (unitary invariance) which makes it a suitable candidate to avoid the barren plateau problem. The rationale behind this is that for unitarily invariant distances as the trace distance or the distances derived from the fidelity, all the states of the computational basis are equally orthogonal and thus have all maximum distance one with respect to the other. The QW_1 functional instead measures how many qubits are different between the two states allowing the VQA gradient to be less flat in the regions that are not already very close to a local minimum. While this special property of QW_1 has been already observed in other contexts, such as the study of quantum Generative Adversarial Networks presented in [24–29], here we test its effectiveness in the identifying effective QEC procedures. For this purpose, we run a series of numerical tests which compare the performances of a VQA that adopts a conventional (i.e. unitary invariant) cost function, with that of a VQA which instead refers to QW_1-like distances. Our findings confirm that in the second case the effectiveness of the numerical optimization significantly increases both in terms of the probability of a successful training and in the fidelity of the recovered state.

The manuscript is organized as follows: in Section II we present a concise, yet rather complete review on the QW_1 distance for qubits; in Sec. III we present some basic notions on conventional QEC procedures which allow us to set the notation and the theoretical background; in Sec. IV we introduce our VQA discussing the different choices of cost functions that can be used in order to guide it; in Sec. V we present our numerical results where comparing the performances of the VQA implemented with different types of cost functions. Conclusions are given in Sec. VI.

II. THE QUANTUM WASSERSTEIN DISTANCE OF ORDER 1 FOR QUBITS

The theory of optimal mass transport [30–32] considers probability distributions on a metric space as distributions of a unit amount of mass. The key element of such theory is the Monge–Kantorovich distance between probability distributions, which is the minimum cost that is required to transport one distribution onto the other, assuming that moving a unit of mass for a unit distance has cost one [33–35]. Such distance is also called earth mover's distance or Wasserstein distance of order 1, often shortened to W_1 distance. The exploration of the theory of optimal mass transport has led to the creation of an extremely fruitful field in mathematical analysis, with applications ranging from differential geometry and partial differential equations to machine learning [31, 35, 36].

The most natural distance on the set of the strings of n bits is the Hamming distance [23], which counts the number of different bits. The resulting W_1 distance on the set of the probability distributions on strings of n bits is called Ornstein's $\bar{d}$-distance [37]. Ref. [22] proposed a generalization of the W_1 distance to the space of the quantum states of a finite set of qubits, called quantum W_1 distance (or QW_1 in brief). The generalization is based on the notion of neighboring quantum states. Two quantum states of a finite set of qubits are neighboring if they coincide after discarding one qubit. The quantum W_1 distance of Ref. [22] is the distance induced by the maximum norm that assigns distance at most 1 to any couple of neighboring states; in the case of quantum states diagonal in the computational basis it recovers Ornstein's $\bar{d}$-distance and inherits most of its properties.

The QW_1 quantity can be computed with a semidefinite program, whose formulation requires to define a notion of Lipschitz constant for quantum observables. The Lipschitz constant of the observable $\hat{H}$ acting on the Hilbert space of n qubits is [22]

$$\|\hat{H}\|_L = 2 \max_{i=1,\,\dots,\,n} \min_{\hat{H}_{i^c}} \left\| \hat{H} - \hat{\mathbb{1}}_i \otimes \hat{H}_{i^c} \right\|_\infty , \qquad (1)$$

where the minimization is performed over all the observables $\hat{H}_{i^c}$ that *do not* act on the i-th qubit. The quantum W_1 distance between the quantum states $\hat{\rho}$ and $\hat{\sigma}$ can then be expressed as [22]

$$\|\hat{\rho} - \hat{\sigma}\|_{W_1} = \max_{\|\hat{H}\|_L \leq 1} \mathrm{Tr}\left[(\hat{\rho} - \hat{\sigma}) \hat{H} \right] . \qquad (2)$$

The present paper is based on the following lower bound to the quantum W_1 distance. Let

$$\hat{H}^{(\text{wass})} = \sum_{i=1}^{n} |1\rangle_i \langle 1| \otimes \hat{\mathbb{1}}_{i^c} , \tag{3}$$

be the quantum observable that counts the number of ones in the computational basis. We have $\left\| \hat{H}^{(\text{wass})} \right\|_L = 1$ [22], therefore for any quantum state $\hat{\rho}$ we have

$$\left\| \hat{\rho} - |0\rangle\langle 0|^{\otimes n} \right\|_{W_1} \geq \text{Tr}\left[\hat{\rho}\, \hat{H}^{(\text{wass})} \right] . \tag{4}$$

QW_1 has found several applications in quantum information theory and many-body quantum physics, among which we mention a proof of the equivalence between the microcanonical and the canonical ensembles of quantum statistical mechanics [38] and a proof of limitations of VQA [39, 40]. Furthermore, QW_1 has been extended to quantum spin systems on infinite lattices [41]. In the context of quantum state tomography, the quantum W_1 distance has been employed as a quantifier of the quality of the learned quantum state and has led to efficient algorithms to learn Gibbs states of local quantum Hamiltonians [42–44]. In the context of quantum machine learning, the quantum W_1 distance has been employed as a cost function of the quantum version of generative adversarial networks [24, 26–28].

A.   Related approaches

Several quantum generalizations of optimal transport distances have been proposed. One line of research by Carlen, Maas, Datta and Rouzé [45–51] defines a quantum Wasserstein distance of order 2 from a Riemannian metric on the space of quantum states based on a quantum analog of a differential structure. Exploiting their quantum differential structure, Refs. [47, 48, 52] also define a quantum generalization of the Lipschitz constant and of the Wasserstein distance of order 1. Alternative definitions of quantum Wasserstein distances of order 1 based on a quantum differential structure are proposed in Refs. [53–56]. Refs. [57–59] propose quantum Wasserstein distances of order 1 based on a distance between the vectors of the canonical basis.

Another line of research by Golse, Mouhot, Paul and Caglioti [60–70] arose in the context of the study of the semiclassical limit of quantum mechanics and defines a family of quantum Wasserstein distances of order 2. Ref. [71] proposes another quantum Wasserstein distance of order 2 where the optimal transport is implemented with quantum channels.

The quantum Wasserstein distance between two quantum states can be defined as the classical Wasserstein distance between the probability distributions of the outcomes of an informationally complete measurement performed on the states, which is a measurement whose

probability distribution completely determines the state. This definition has been explored for Gaussian quantum systems with the heterodyne measurement in Refs. [72–74].

III.   PRELIMINARIES ON QEC

Let Q be a quantum register we wish to protect (at least in part) from the action of some external noise source. In a typical QEC scenario [75] this problem is addressed through the following three-step procedure:

i) Before the action of the noise, a unitary encoding gate $\hat{V}_{QA}$ is used to distribute the information originally contained in Q on the larger system QA. Here A is an auxiliary quantum register that is assumed to be initialized in a fiduciary quantum state, and that is affected by the same noise that tampers with Q;

ii) After the action of the noise, a measurement on QA is performed to reveal the nature of the latter and, based on the associated outcome, a unitary recovery operation is applied to the system. Equivalently this step can be described by introducing yet an extra quantum register B (also initialized on a fiduciary state but *not* affected by the noise) that is coupled with QA trough a recovering unitary transformation $\hat{W}_{QAB}$ which effectively mimics the measurement and the recovery operation;

iii) The inverse of the gate $\hat{V}_{QA}$ is finally used on QA to refocus the recovered information in Q.

Denoting with $|\psi\rangle_Q$ the input state of Q, the corresponding output state of QA that emerges from the process at the end of the step *iii)* can be expressed as the density matrix

$$\hat{\rho}_{QA}^{(V,W)}(\psi) := \text{tr}_B\Big\{ \mathcal{V}_{QA}^\dagger \circ \mathcal{W}_{QAB} \circ \Phi_{QA} \tag{5}$$
$$\circ \mathcal{V}_{QA}\Big(|\psi\rangle_Q\langle\psi| \otimes |\emptyset\rangle_A\langle\emptyset| \otimes |\emptyset\rangle_B\langle\emptyset| \Big)\Big\}$$
$$:= \mathcal{V}_{QA}^\dagger \circ \Phi_{QA}^{(R)} \circ \Phi_{QA} \circ \mathcal{V}_{QA}\Big(|\psi\rangle_Q\langle\psi| \otimes |\emptyset\rangle_A\langle\emptyset| \Big)$$

where $|\emptyset\rangle_X$ represents the fiduciary state of the X register, $\text{tr}_B\{\cdots\}$ is the partial trace over B, and given a unitary $\hat{U}_X$ on X we adopted the symbol $\mathcal{U}_X(\cdots) := \hat{U}_X \cdots \hat{U}_X^\dagger$ to denote its action as super-operator. In the above expressions Φ_{QA} is the LCPT quantum channel [75] describing the noise on Q and A, while $\Phi_{QA}^{(R)}(\cdots) := \text{tr}_B\{\mathcal{W}_{QAB}(\cdots \otimes |\emptyset\rangle_B\langle\emptyset|)\}$ is the LCPT (recovery) quantum channel on QA originating from the interaction with B, that attempts to undo the action of Φ_{QA}.

An ideal QEC procedure able to completely remove the noise from the system will make sure that $\hat{\rho}_{QA}^{(V,W)}(\psi)$ corresponds to $|\psi\rangle_Q|\emptyset\rangle_A$, irrespectively from the specific

choice of $|\psi\rangle_Q$. A bona-fide figure of merit to characterize the effectiveness of a generic QEC scheme is hence provided by the average input-output fidelity

$$\overline{F}(V,W) := \int d\mu_\psi \ _Q\langle\psi|_A\langle\varnothing|\hat{\rho}_{QA}^{(V,W)}(\psi)|\psi\rangle_Q|\varnothing\rangle_A \ , \ (6)$$

where $d\mu_\psi$ is the uniform measure on the set of the input states of Q originated from the Haar measure on the associated unitary group [76] or from an exact or approximate unitary 2-design $\mathcal{S}$ [75, 77] that simulates the latter[1]. Notice that by expressing $|\psi\rangle_Q = \hat{U}_Q|\varnothing\rangle_Q$, Eq. (6) can equivalently be casted in the more compact form

$$\overline{F}(V,W) = \ _{QA}\langle\varnothing|\hat{\rho}_{QA}^{(V,W)}|\varnothing\rangle_{QA} \ , \qquad (7)$$

with $|\varnothing\rangle_{QA} := |\varnothing\rangle_Q \otimes |\varnothing\rangle_A$ and where the state

$$\hat{\rho}_{QA}^{(V,W)} := \frac{1}{|\mathcal{S}|}\sum_{\hat{U}_Q\in\mathcal{S}} \mathcal{U}_Q^\dagger \circ \mathcal{V}_{QA}^\dagger \circ \Phi_{QA}^{(R)} \circ \Phi_{QA}$$

$$\circ \ \mathcal{V}_{QA} \circ \mathcal{U}_Q\Big(|\varnothing\rangle_{QA}\langle\varnothing|\Big) \ , \qquad (8)$$

now includes the average over all possible inputs. An ideal QEC procedure will enable one to get $\overline{F}(V,W) = 1$. A natural benchmark for lowest admissible $\overline{F}(V,W)$ is represented instead by the value one would get if one decides not to perform corrections on the register that we compute by setting $\hat{V}_{QA}$ and $\hat{W}_{QAB}$ equal to the identity operators i.e.[2]

$$\overline{F}_0 := \ _{QA}\langle\varnothing|\hat{\rho}_{QA}^{(\mathbb{1},\mathbb{1})}|\varnothing\rangle_{QA} \ . \qquad (9)$$

IV. VARIATIONAL QUANTUM ALGORITHM

While enormous progress has been made in the study of QEC procedures, identifying efficient choices for the operations that lead to (non trivial) high values of $\overline{F}(V,W)$ for a specific noise model, is still a challenging open problem. A possible solution, in this case, is to employ variational

---

[1] We remind that a unitary 2-design is a probability distribution over the set of unitary operators which can duplicate properties of the probability distribution over the Haar measure for polynomials of degree 2 or less. When Q is a single qubit, a 2-design can be realized by a uniform sampling over a set $\mathcal{S}$ composed by only 6 elements $\hat{\mathbb{1}}$, $\hat{\sigma}_1$, $e^{\pm i\pi/4\hat{\sigma}_1}$, $e^{\pm i\pi/4\hat{\sigma}_2}$ that maps its logical state $|0\rangle_Q$ into the vectors $|0\rangle_Q$, $|1\rangle_Q$, $(|0\rangle_Q \pm i|1\rangle_Q)/\sqrt{2}$, $(|0\rangle_Q \mp |1\rangle_Q)/\sqrt{2}$.

[2] Equation (9) accounts for the noise effects both on Q and A. A more conservative estimation of $\overline{F}_0$ can be obtained by focusing directly on the noise on Q alone, i.e. tracing out the A component of $\hat{\rho}_{QA}^{(\mathbb{1},\mathbb{1})}$ and studying its fidelity with $|\varnothing\rangle_Q$, i.e. $\overline{F}_0^{(\text{strong})} := \ _Q\langle\varnothing|\hat{\rho}_Q^{(\mathbb{1},\mathbb{1})}|\varnothing\rangle_Q \geq \overline{F}_0$, with $\hat{\rho}_Q^{(\mathbb{1},\mathbb{1})} := \text{tr}_A\,\hat{\rho}_{QA}^{(\mathbb{1},\mathbb{1})}$. Notice that for the noise model of Sec. IV C the two are directly connected via the identity $\overline{F}_0 = \overline{F}_0^{(\text{strong})} - \frac{n-1}{n}p(1 - |\langle 0|\hat{\sigma}|0\rangle|^2)$.
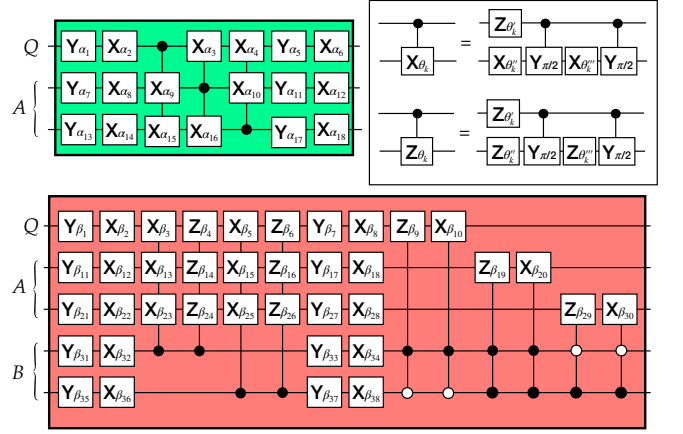

Figure 1. Parametric gates $\hat{V}_{QA}(\vec{\alpha})$ (green element) and $\hat{W}_{QAB}(\vec{\beta})$ (red element) used for case of quantum registers Q, A, and B with $k = 1$, $n - k = 2$, and $r = 2$ qubits respectively. Indicating with $\hat{\sigma}_1$, $\hat{\sigma}_2$, and $\hat{\sigma}_3$ the Pauli operators, the X_θ, Y_θ, and Z_θ elements of the figure represent single qubit rotations $e^{-i\theta\hat{\sigma}_1}$, $e^{-i\theta\hat{\sigma}_2}$, and $e^{-i\theta\hat{\sigma}_3}$ with the angles θ determined to the components of the vectors $\vec{\alpha}$, $\vec{\beta}$, respectively. Vertical lines indicate instead quantum control operations which are activated when the control qubits (indicated by the full or empty circles) are in the logical state $|1\rangle$ (full circle) or in $|0\rangle$ (empty circle). As shown on the inset, each one of those gates depend parametrically upon elements of the control vectors $\vec{\alpha}$ and $\vec{\beta}$ through single qubit operations.

quantum algorithms to run numerical searches. Our approach follows a training strategy inspired by the work of Johnson et al. [19]. Assuming hence Q, A, and B to be formed by collections of independent qubits (k for Q, $n - k$ for A, and r for B), we introduce a manifold of transformations $\hat{V}_{QA}(\vec{\alpha})$, $\hat{W}_{QAB}(\vec{\beta})$ parametrized by classical controls vectors $\vec{\alpha}$, $\vec{\beta}$ (see Fig. 1), and construct the quantum circuit of Fig. 2. The method then proceeds along the following stages:

1. Having selected the values of $\vec{\alpha}$ and $\vec{\beta}$, the register Q is prepared into a collection of known quantum state $\{|\psi(1)\rangle_Q, \cdots, |\psi(m)\rangle_Q\}$ operating on the vector $|\varnothing\rangle_Q = |0\rangle^{\otimes k}$ through action of the control gates $\hat{U}_Q(1), \cdots, \hat{U}_Q(m)$ (first cyan element of the figure) which define the 2-design $\mathcal{S}$ entering in Eq. (8). Each of such inputs is hence evolved via a circuit (pale-orange area of the figure) that emulates both the effect of the noise (patterned square of the figure, see Section IV C and Fig. 4), and the transformations $\hat{V}_{QA}(\vec{\alpha})$, $\hat{W}_{QAB}(\vec{\beta})$, and $\hat{V}_{QA}^\dagger(\vec{\alpha})$ that are meant to implement the steps $ii)$ and $iii)$ of the QEC procedure (green and red elements of the figure). Notice that in the ideal case (i.e. if $\hat{V}_{QA}(\vec{\alpha})$ and $\hat{W}_{QAB}(\vec{\beta})$ manage to completely suppress the noise) then in correspondence with the input $|\psi(j)\rangle_Q$ the registers QA should emerge in the state $|\psi(j)\rangle_Q \otimes |\varnothing\rangle_A := |\psi(j)\rangle_Q \otimes |0\rangle^{\otimes n-k}$,
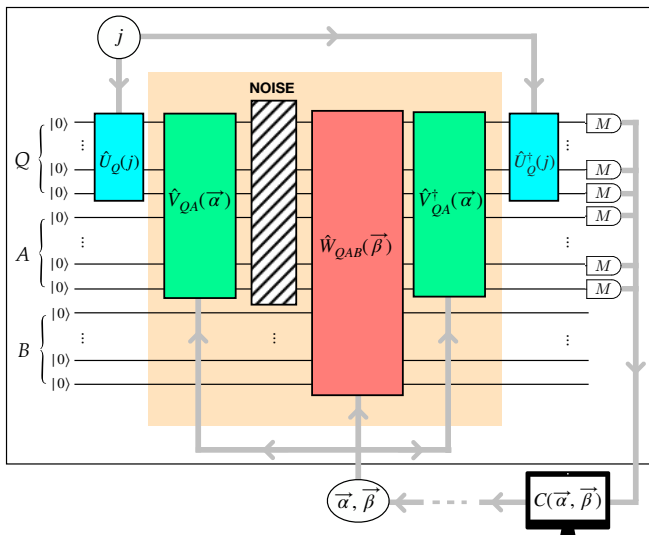
Figure 2. (Color online) Sketch of the variational quantum algorithm: Q, A and B are quantum registers formed respectively by k, $n-k$ and r qubits. The initial information we wish to protect is written in Q by the unitary gate $\hat{U}_Q(j)$ extracted from a 2-design set $\mathcal{S}$; A and B are two auxiliary elements (containing respectively $n-k$ and r qubits) that are used to implement the QEC procedure described by the parametric gates $\hat{V}_{QA}(\vec{\alpha})$, $\hat{W}_{QAB}(\vec{\beta})$, and $\hat{V}_{QA}^{\dagger}(\vec{\alpha})$ of Fig. 1. The patterned element in the central part of the scheme represents the noise on Q and A (no noise is assumed to be active on B). Lastly, the D-shaped measurements at the end of the circuit represent local measurements on QA whose outcomes over the entire collection of the possible inputs generated by the entire set $\mathcal{S}$, are processed by a classical computer which, evaluating the cost function $C(\vec{\alpha}, \vec{\beta})$ defined in Section IV B, decides how to update the values of the parameters $\vec{\alpha}$ and $\vec{\beta}$. Thick grey lines in the figure represent classical control lines.

which will be hence mapped into the final configuration $|\varnothing\rangle_{QA} := |0\rangle^{\otimes n}$ by the inverse $\hat{U}_Q^{\dagger}(j)$ of the state preparation gate (second cyan element of the figure).

2. For each choice of the index $j \in \{1, \cdots, m\}$ a measurement on the system is performed at the end of the transformations described in stage 1 and the resulting m collected outcomes used to compute a cost function $C(\vec{\alpha}, \vec{\beta})$ which evaluates the effectiveness of the adopted QEC strategy in leading large values of the average input-output fidelity. The specific choice of the cost function is very important and is discussed in Section IV A.

3. A classical computer decides, given the results of the measurement, how to change the value of the parameters $\vec{\alpha}$ and $\vec{\beta}$ to be used in the subsequent run in order to minimize the cost function $C(\vec{\alpha}, \vec{\beta})$. This is discussed in detail in Section IV B.

A. Cost function

The natural choice for the cost function at the stage 2 of our algorithm is provided by the expectation value of the self-adjoint operator

$$\hat{H}_{QA}^{(\text{fid})} := \hat{\mathbb{1}}_{QA} - |\varnothing\rangle_{QA} \langle\varnothing| , \qquad (10)$$

computed on the mean state of system QA which emerges at the output of the quantum circuit of Fig. 2, i.e. the quantity

$$C^{(\text{fid})}(\vec{\alpha}, \vec{\beta}) := \text{tr}\{\hat{\rho}_{QA}^{(V(\vec{\alpha}), W(\vec{\beta}))} \hat{H}_{QA}^{(\text{fid})}\} , \qquad (11)$$

where $\hat{\rho}_{QA}^{(V(\vec{\alpha}), W(\vec{\beta}))}$ is the density matrix (8) evaluated for $\hat{V}_{QA} = \hat{V}_{QA}(\vec{\alpha})$ and $\hat{W}_{QAB} = \hat{W}_{QAB}(\vec{\beta})$. This choice has two main advantages. First of all, the expectation value $C^{(\text{fid})}(\vec{\alpha}, \vec{\beta})$ can be evaluated by performing (simple) local measurement on the qubits of Q and A (indeed it can be computed by simply checking whether of not each one of them is in the logical state $|0\rangle$). Most importantly, since by explicit evaluation one has that $C^{(\text{fid})}(\vec{\alpha}, \vec{\beta}) = 1 - \overline{F}(V(\vec{\alpha}), W(\vec{\beta}))$, it is clear that by using (11) the algorithm will be forced to look for values of $\vec{\alpha}$, $\vec{\beta}$ that yield higher average input-output fidelities. Despite all this, the use of $C^{(\text{fid})}(\vec{\alpha}, \vec{\beta})$ as a cost function, has a major drawback associated with the fact that the spectrum of the Hamiltonian $\hat{H}_{QA}^{(\text{fid})}$ exhibits maximum degeneracy with respect to space orthogonal to the target state $|\varnothing\rangle_{QA}$ (see Fig. 3). Due to this fact a numerical search based on a training procedure that simply target the minimization of $C^{(\text{fid})}(\vec{\alpha}, \vec{\beta})$, has non trivial chances to get stuck somewhere in the large flat plateau associated with the eigenvalue 1 of $\hat{H}_{QA}^{(\text{fid})}$ without finding any good direction. in the large flat plateau A possible way to avoid this problem is to introduce new cost-functions Hamiltonians which, while maintaining the target vector $|\varnothing\rangle_{QA}$ as a unique ground state and still being easy to compute, manage to remove the huge degeneracy of the excited part of the spectra of $\hat{H}_{QA}^{(\text{fid})}$. Our choice is based on the quantum Wasserstein distance of order 1 (W_1) introduced Ref. [22] which, even though it lacks some interesting properties that the fidelity has, is less likely to be affected by the barren plateaus phenomena [21]. As mentioned in Sec. II good estimation of the W_1 distance that separate $\hat{\rho}_{QA}^{(V(\vec{\alpha}), W(\vec{\beta}))}$ from the target state, is provided by the following quantity

$$C^{(\text{wass})}(\vec{\alpha}, \vec{\beta}) := \text{tr}\{\hat{\rho}_{QA}^{(V(\vec{\alpha}), W(\vec{\beta}))} \hat{H}_{QA}^{(\text{wass})}\} , \qquad (12)$$

$$\hat{H}_{QA}^{(\text{wass})} := \sum_{j=1}^{n} j \, \hat{\Pi}_{QA}^{(j)} , \qquad (13)$$

where $\hat{H}_{QA}^{(\text{wass})}$ is the Hamiltonian (3) which we express here in terms of the projectors $\hat{\Pi}_{QA}^{(j)}$ on the sub-space of the register QA in which we have j qubits in $|1\rangle$ and the
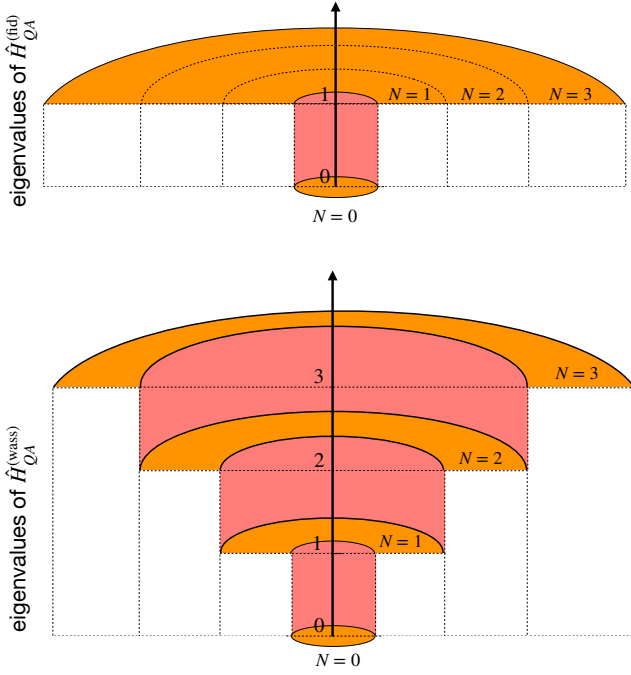
Figure 3. Pictorial rendering of the spectra of the Hamiltonians $\hat{H}_{QA}^{(\text{fid})}$ (top panel) and $\hat{H}_{QA}^{(\text{wass})}$ (lower panel). While $\hat{H}_{QA}^{(\text{fid})}$ is characterized by a unique, flat plateau that includes all the excited state, $\hat{H}_{QA}^{(\text{wass})}$ partially removes the associated degeneracy assigning higher energy to subspaces that have higher number of qubits in the logical state $|1\rangle$.

remaining one in $|0\rangle$. Observe that, as already anticipated, $\hat{H}_{QA}^{(\text{wass})}$ is the sum of the number operators acting on the individual qubits of the register QA as in (3): accordingly, as $C^{(\text{fid})}(\vec{\alpha}, \vec{\beta})$, $C^{(\text{wass})}(\vec{\alpha}, \vec{\beta})$ can be computed from local measurement. What $C^{(\text{wass})}(\vec{\alpha}, \vec{\beta})$ does is to count the total number of logical ones present in the system. To understand why using (12) could in principle lead to a more efficient numerical search than the one obtained by using (11), notice that Eq. (10) can be equivalently written as $\hat{H}_{QA}^{(\text{fid})} = \sum_{j=1}^{n} \hat{\Pi}_{QA}^{(j)}$. A comparison with (13) reveals hence that indeed while both $\hat{H}_{QA}^{(\text{fid})}$ and $\hat{H}_{QA}^{(\text{wass})}$ admit $|\varnothing\rangle_{QA}$ as a unique ground state, the Wasserstein Hamiltonian removes large part of the degeneracy of the high energy spectrum of the fidelity Hamiltonian. Accordingly, it is reasonable to expect that a numerical search that uses $\hat{H}_{QA}^{(\text{wass})}$, has fewer chances to get trapped into regions of constant energy (barren plateau) than a search based on $\hat{H}_{QA}^{(\text{fid})}$ [3].

---

[3] It goes without mentioning that alternative choices for the cost function Hamiltonians are also available. For instance, one can use operators that also remove the residual degeneracies that

B.   Descent algorithm

The algorithm that we used for this work is a gradient descent algorithm with momentum [78]. To overcome the numerical difficulties of using finite differences to estimate the gradients of the cost function $C(\vec{\alpha}, \vec{\beta})$, we exploit a variation of the parameter-shift rule introduced in [79] which reduces the problem to compute linear combinations of the function itself evaluated in different points that are not infinitesimally close. Specifically, we observe that, irrespectively from the choice of the operator $\hat{H}_{QA}$, the functional dependence of $C(\vec{\alpha}, \vec{\beta})$ upon the j-th component of the vector $\vec{\beta}$ is of the form

$$C(\vec{\alpha}, \vec{\beta}) = f(\beta_j) := \sum_k \text{tr} \left\{ \hat{\Omega}_1^{(k)} e^{i\beta_j \hat{\sigma}} \hat{\Omega}_2^{(k)} e^{-i\beta_j \hat{\sigma}} \right\}, \quad (14)$$

with $\hat{\Omega}_{1,2}^{(k)}$ being multi-qubits operators which do not depend upon β_j, and with $e^{-i\beta_j \hat{\sigma}}$ a single qubit rotation generated by an element $\hat{\sigma}$ of the Pauli set. Therefore its gradient can be written as

$$\frac{\partial C(\vec{\alpha}, \vec{\beta})}{\partial \beta_j} = i \sum_k \text{tr} \left\{ \hat{\Omega}_1^{(k)} e^{i\beta_j \hat{\sigma}} [\hat{\sigma}, \hat{\Omega}_2^{(k)}] e^{-i\beta_j \hat{\sigma}} \right\}$$
$$= f(\beta_j + \tfrac{\pi}{4}) - f(\beta_j - \tfrac{\pi}{4}), \quad (15)$$

where in the last passage we used the identity

$$i[\hat{\sigma}, \hat{\Omega}_2^{(k)}] = e^{i\frac{\pi}{4}\hat{\sigma}} \hat{\Omega}_2^{(k)} e^{-i\frac{\pi}{4}\hat{\sigma}} - e^{-i\frac{\pi}{4}\hat{\sigma}} \hat{\Omega}_2^{(k)} e^{i\frac{\pi}{4}\hat{\sigma}}. \quad (16)$$

The gradient with respect the vector $\vec{\alpha}$ can be computed similarly. In this case however we observe that, due to the fact that $\hat{\rho}_{QA}^{(V(\vec{\alpha}), W(\vec{\beta}))}(\psi)$ depends upon the parameters $\vec{\alpha}$ via $\hat{V}_{QA}(\vec{\alpha})$ and through its adjoint $\hat{V}_{QA}^{\dagger}(\vec{\alpha})$, the dependence of $C(\vec{\alpha}, \vec{\beta})$ upon the j-th component of $\vec{\alpha}$ is slightly more complex. Indeed in this case we have

$$C(\vec{\alpha}, \vec{\beta}) = g(\alpha_j, \alpha_j), \quad (17)$$

where $g(\alpha_j^{(1)}, \alpha_j^{(2)})$ is the function

$$g(\alpha_j^{(1)}, \alpha_j^{(2)}) := \sum_k \text{tr} \left\{ \hat{\Omega}_1^{(k)} e^{i\alpha_j^{(1)} \hat{\sigma}} \hat{\Omega}_2^{(k)} e^{-i\alpha_j^{(1)} \hat{\sigma}} \right. \quad (18)$$
$$\left. \times \hat{\Omega}_3^{(k)} e^{i\alpha_j^{(2)} \hat{\sigma}} \hat{\Omega}_4^{(k)} e^{-i\alpha_j^{(2)} \hat{\sigma}} \right\},$$

with $\hat{\Omega}_{1,2,3,4}^{(k)}$ representing multi-qubits operators which do not depend neither upon $\alpha_j^{(1)}$ nor $\alpha_j^{(2)}$. It is important

---

affect $\hat{H}_{QA}^{(\text{wass})}$ – e.g. using the operator $\hat{H}_{QA}^{(\text{full})} = \sum_{\ell=1}^{n} w_\ell \hat{\pi}_\ell$ with ω_ℓ positive weights selected so that different allocation of $|1\rangle$ states inside the eigenspaces of $\hat{H}_{QA}^{(\text{wass})}$ get an assigned ordering. Our numerical analysis however seems to indicate that these refinements do not contribute significantly in improving numerical search of the algorithm.

to stress that $g(\alpha_j^{(1)}, \alpha_j^{(2)})$ can be computed using the same circuit of Fig. 2, by simply replacing the phases α_j of $\hat{V}_{QA}(\vec{\alpha})$ and $\hat{V}_{QA}^\dagger(\vec{\alpha})$ with $\alpha_j^{(1)}$ and $\alpha_j^{(2)}$ respectively. Notice finally that exploiting the identity Eq. (16) we can write

$$\frac{\partial C(\vec{\alpha}, \vec{\beta})}{\partial \alpha_j} = \left.\frac{\partial g(\alpha_j^{(1)}, \alpha_j)}{\partial \alpha_j^{(1)}}\right|_{\alpha_j^{(1)}=\alpha_j} + \left.\frac{\partial g(\alpha_j, \alpha_j^{(2)})}{\partial \alpha_j^{(2)}}\right|_{\alpha_j^{(2)}=\alpha_j} \quad (19)$$

$$= g(\alpha_j + \tfrac{\pi}{4}, \alpha_j) - g(\alpha_j - \tfrac{\pi}{4}, \alpha_j)$$
$$+ g(\alpha_j, \alpha_j + \tfrac{\pi}{4}) - g(\alpha_j, \alpha_j - \tfrac{\pi}{4}) ,$$

which shows that computing the gradient of $C(\vec{\alpha}, \vec{\beta})$ with respect to α_j simply accounts to evaluate the circuit that express $g(\alpha_j^{(1)}, \alpha_j^{(2)})$ for four distinct values of the parameters.

C. Noise model

The scheme presented so far can in principle be applied to arbitrary classes of noises. In our research however we focused on a specific model that has been extensively studied in the literature producing explicit examples of efficient QEC solutions which can be used as a theoretical benchmark for our variational search. Specifically we assume Q and A to be respectively a single qubit register ($k = 1$) and a two qubit register ($n = 3$), globally affected by a given species of single-qubit noise [80, 81]. These transformations can be represented in terms of a LCPT map of the form

$$\Phi_{QA}(\cdots) = \sum_{\ell=0}^{n} \hat{K}_{QA}^{(\ell)} \cdots \hat{K}_{QA}^{(\ell)\dagger} , \quad (20)$$

with Kraus operators [75]

$$\hat{K}_{QA}^{(0)} := \sqrt{1-p}\, \mathbb{1}_{QA} , \qquad \hat{K}_{QA}^{(\ell)} := \sqrt{\tfrac{p}{n}}\, \hat{\sigma}^{(\ell)} , \quad (21)$$

where for $\ell \in \{1, \cdots, n\}$, $\hat{\sigma}^{(\ell)}$ is the Pauli operator acting on the ℓ-th qubit of QA which defines the noise species we have selected. For instance, in the case we choose to describe phase-flip noise then $\hat{\sigma}^{(\ell)} = \hat{\sigma}_3^{(\ell)}$, while for describing bit-flip we have $\hat{\sigma}^{(\ell)} = \hat{\sigma}_1^{(\ell)}$. Explicit examples of $\hat{V}_{QA}$, $\hat{W}_{QAB}$ which allow for exact suppression of the noise ($\overline{F}(V, W) = 1$) are shown in Fig. 5. Notice that by construction the circuit parametrization of $\hat{V}_{QA}(\vec{\alpha})$, $\hat{W}_{QAB}(\vec{\beta})$ given in Fig. 2 include such gates as special solution: accordingly if properly guided by an efficient cost function, our numerical VQA search has a chance to find the solution of Fig. 5.

V. RESULTS

In this section we study the impact of the cost function on the efficiency of the optimization algorithm of
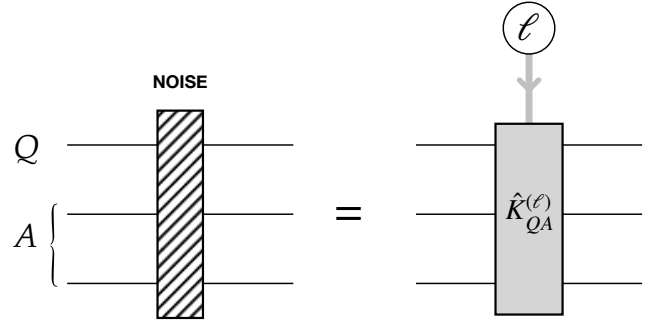

Figure 4. Circuital implementation of the noise element of Fig. 2: here $\hat{K}_{QA}^{(\ell)}$ are weighted unitaries of Eq. (21).

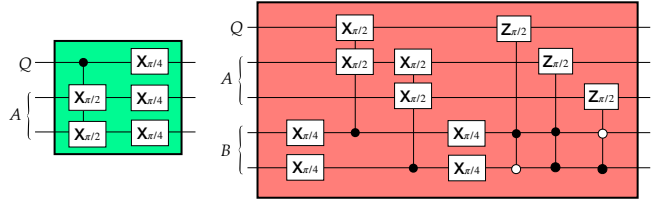Figure 5. Circuital implementations of the ideal transformations $\hat{V}_{QA}(\vec{\alpha})$ (left) and $\hat{W}_{QAB}(\vec{\beta})$ (right) which allow for exact noise suppression of a single-qubit bit-flip noise model [i.e. (20) with $\hat{\sigma}^{(\ell)} = \hat{\sigma}_1^{(\ell)}$] using a quantum register B with $r = 2$ qubit. gates.

Sec. IV. Assuming the single-qubit noise model detailed in Sec. IV C and taking B to be a $r = 2$ qubit register, we run two distinct numerical searches: the first obtained by identifying $C(\vec{\alpha}, \vec{\beta})$ with $C^{(\text{fid})}(\vec{\alpha}, \vec{\beta})$ and the second choosing instead $C^{(\text{wass})}(\vec{\alpha}, \vec{\beta})$. Results are reported in Figs. 6 and 7 for two different choices of the noise models (20), i.e. phase-flip and bit-flip. For both, we compare the input-output average fidelity (7) at the end of the procedure obtained with the two different cost functions, and the number of iterations M needed for convergence. Regarding this last quantity we set a maximum value M_{max} equal to 2000 before convergence and we chose this limit mainly with practical choices like the maximum time for the simulation, enforcing that a single run does not require more than a few hours of computational time: in case the algorithm fails to reach the convergency we simply stop the numerical search (this is the reason for the peak at the end of the upper orange plot in Fig. 7). The plots report only the simulations that manage to achieve an average fidelity that is greater or equal than no-correction threshold bound $\overline{F}_0$.

The first thing to observe is that for both noise models, $C^{(\text{fid})}(\vec{\alpha}, \vec{\beta})$ has problem in reaching the do-nothing threshold $\overline{F}_0$: the probability of success being 2.6% for the phase-flip case of Fig. 7 and only 0.2% for the bit-flip case of Fig. 6 (for both noise models the total number of simulations analyzed was 500). Observe also that in this
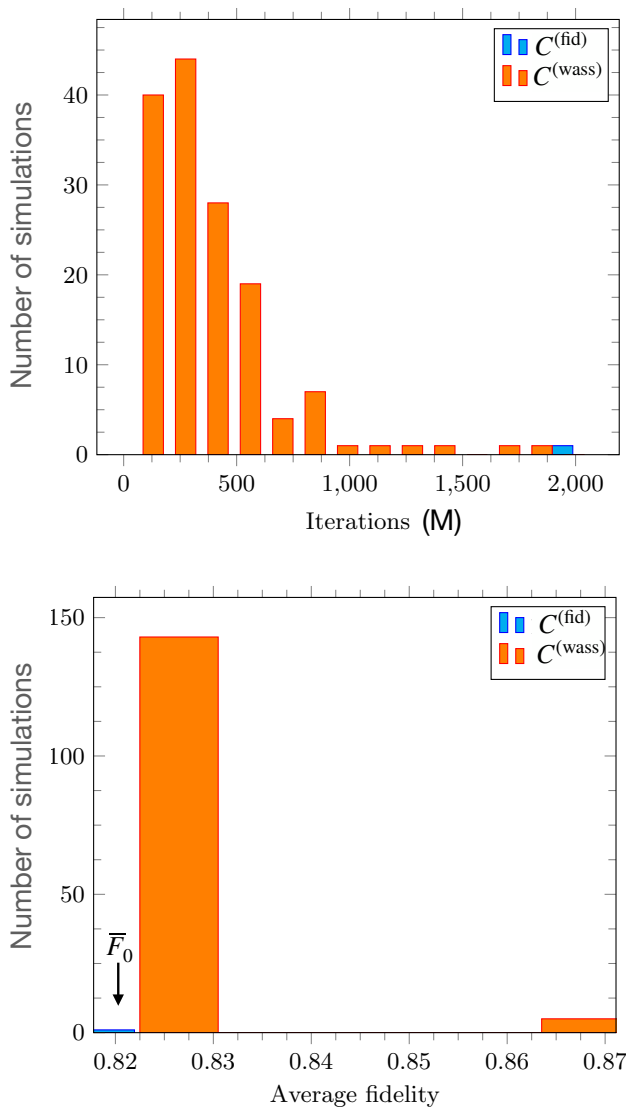
Figure 6. Comparison of the the input-output average fidelity (7) attainable by running our optimization algorithm using the cost function $C^{(\mathrm{fid})}(\vec{\alpha}, \vec{\beta})$ (blue data) and $C^{(\mathrm{wass})}(\vec{\alpha}, \vec{\beta})$ (orange data). Here the error model is a single-qubit bit-flip noise ($\hat{\sigma} = \hat{\sigma}_1$ in (20)) with $p = 0.8$. The no error correction threshold (9) of this scheme is $\overline{F}_0 \approx 0.822$ – orange peak in the fidelity plot, up to numerical precision. Only the runs that produced a fidelity of at least $\overline{F}_0$ have been included: for $C^{(\mathrm{fid})}(\vec{\alpha}, \vec{\beta})$ this happens in the 0.2% of the runs, while for $C^{(\mathrm{wass})}(\vec{\alpha}, \vec{\beta})$ for the 29.6%.

_____

[4] It has to be said that in few cases the figure of merit is worse after the second optimization – see the negative bar in right panel of Fig. 8. This is due to the fact that when using $C^{(\mathrm{wass})}(\vec{\alpha}, \vec{\beta})$ we are not maximizing the fidelity but minimizing a function whose stationary point corresponds to the maximum of the latter: accordingly the final point of convergence for $C^{(\mathrm{wass})}(\vec{\alpha}, \vec{\beta})$ can be slightly off mark in terms of fidelity. This is not a problem because these two functions do not have a constant ratio, and we checked that the inequalities between them are still satisfied.
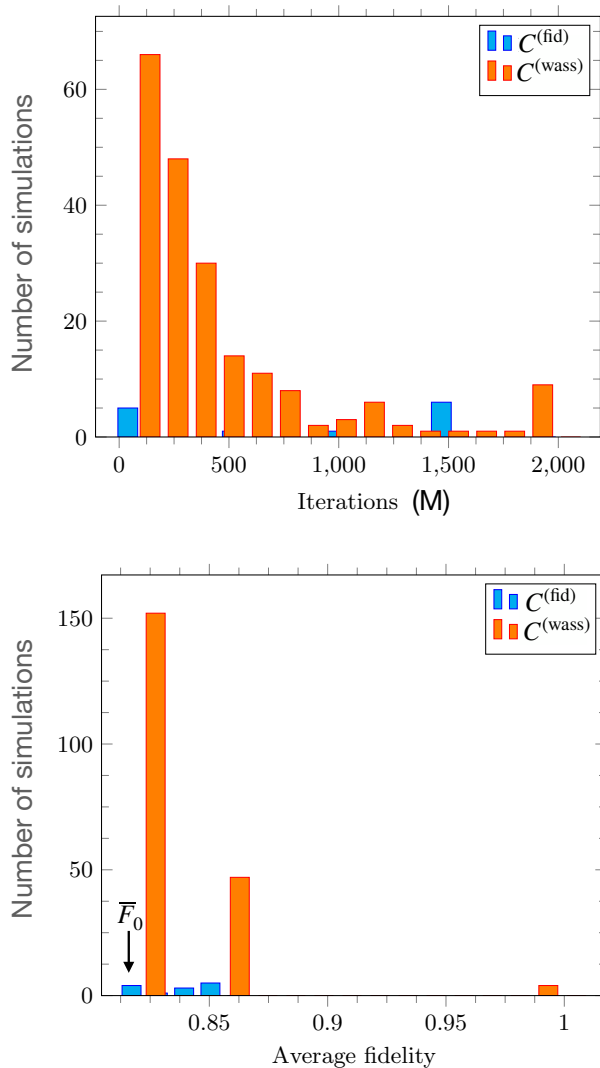
last case the algorithm never yields average input-output fidelity values strictly larger than $\overline{F}_0$ and that, even in those cases, it requires a number M of iterations which saturate the maximum allowed value $M_{\max}$ (blue peak in the upper plot of Fig. 7). $C^{(\mathrm{was})}(\vec{\alpha}, \vec{\beta})$ performs definitely better: to begin with it succeeds in overcoming the threshold $\overline{F}_0$ in one third of the simulations (specifically 40.6% for the phase-flip noise model and 29.6% for the bit-flip noise model). Furthermore, the algorithm reaches convergency with a number of iterations that are typically smaller than those required by $C^{(\mathrm{fid})}(\vec{\alpha}, \vec{\beta})$.

To better enlighten the differences between the two cost functions, we proceeded with further simulations, whose results are summarized in Fig. 8. The idea here is to run a two-step optimization process composed by two sequences of runs: in the first run we start the optimization procedure from a random point in the parameter space $(\vec{\alpha}, \vec{\beta})$ with one of the two cost functions (say $C^{(\mathrm{fid})}(\vec{\alpha}, \vec{\beta})$), up to convergence; after that we start a second optimization run using the other cost function (say $C^{(\mathrm{wass})}(\vec{\alpha}, \vec{\beta})$) but assuming as initial condition for the parameters the final point reached by the first run. The plots report the difference in fidelity between the second and the first run: when we start using the $C^{(\mathrm{wass})}(\vec{\alpha}, \vec{\beta})$ in the first run, the fidelity cannot further improve the result that is already found, and this is represented by the fact that the best improvement is of the order of 10^{-5}; on the contrary, if we started employing $C^{(\mathrm{fid})}(\vec{\alpha}, \vec{\beta})$ in the first run, the use of $C^{(\mathrm{wass})}(\vec{\alpha}, \vec{\beta})$ in the second run typically yields substantial improvements of the performance[4]. Moreover, we sampled some single descent processes and plotted the cost in function of the iteration. When we move from fidelity to W_1, the descent part after the change of cost function is qualitatively indistinguishable from starting from a random point.

VI. CONCLUSIONS

To summarize, we have shown a variational quantum algorithm that allows finding the most suitable error correction procedure for a specific noise on quantum hardware. We compared the performance of two different versions of this algorithm using two different cost functions, the fidelity and an approximation of the quantum Wasserstein distance of order one. We compared the difference

Figure 7. Comparison of the the input-output average fidelity (7) attainable by running our optimization algorithm using the cost function $C^{(\mathrm{fid})}(\vec{\alpha},\vec{\beta})$ (blue data) and $C^{(\mathrm{wass})}(\vec{\alpha},\vec{\beta})$ (orange data). Here the error model is a single-qubit phase-flip noise ($\hat{\sigma} = \hat{\sigma}_3$ in (20) with $p = 0.8$. The no error correction threshold (9) of this scheme is $\overline{F}_0 \approx 0.822$ – orange peak in the fidelity plot, up to numerical precision. Only the runs that produced a fidelity of at least $\overline{F}_0$ have been included: for $C^{(\mathrm{fid})}(\vec{\alpha},\vec{\beta})$ this corresponds to the 2.6% of the runs, while for $C^{(\mathrm{wass})}(\vec{\alpha},\vec{\beta})$ the success probability is 40.6%.

in speed and the ability to obtain a useful solution between the two algorithms, finding really different trends between the two optimization procedures. The optimization process based on the fidelity suffers greatly from the phenomenon of the barren plateaus, leading to very slow convergence or no convergence at all, while the algorithm based on the quantum W_1 distance allows us to find the configurations that correct the errors in the examples that we explored. The obtained results show a clear improvement and allow us to explore further improvements of these methods, as using different algorithms for the

minimization process, e.g. stochastic gradient descent or higher-order algorithms like Newton or pseudo-Newton algorithms.

Given that the gradient can be expressed only with the cost function evaluated in a small number of circuits that differ only in the parameter choice, the gradient of the cost function can be computed on the same hardware that will be used for the correction procedure. Moreover, simulating this circuit may be difficult because of the exponential scaling of the dimension of the Hilbert space of a set of qubits, but this problem does not apply when all the circuit is built on hardware, gaining a quantum advantage. For the same reason, the same procedure can be iterated to compute the exact Hessian of the cost function and then apply a second-order method like the Newton method as a descent algorithm. However, this has not been done because the circuits that we marked as useful have a relatively big number of parameters, and computing the hessian scales quadratically with this number, leading to intractable computations.

With this work, we have shown a clear advantage in the use of this QW_1 distance approximation in a gradient-based optimization algorithm. In future work, it may be interesting to study the effect of this cost function also on gradient-free optimization algorithms, that do also suffer from the barren-plateaus phenomenon, as shown in Ref. [82].

VII. BIBLIOGRAPHY

[1] J. W. Z. Lau, K. H. Lim, H. Shrotriya, and L. C. Kwek, Nisq computing: where are we and where do we go?, AAPPS Bulletin **32**, 27 (2022).

[2] J. Preskill, Quantum computing and the entanglement frontier (2012).

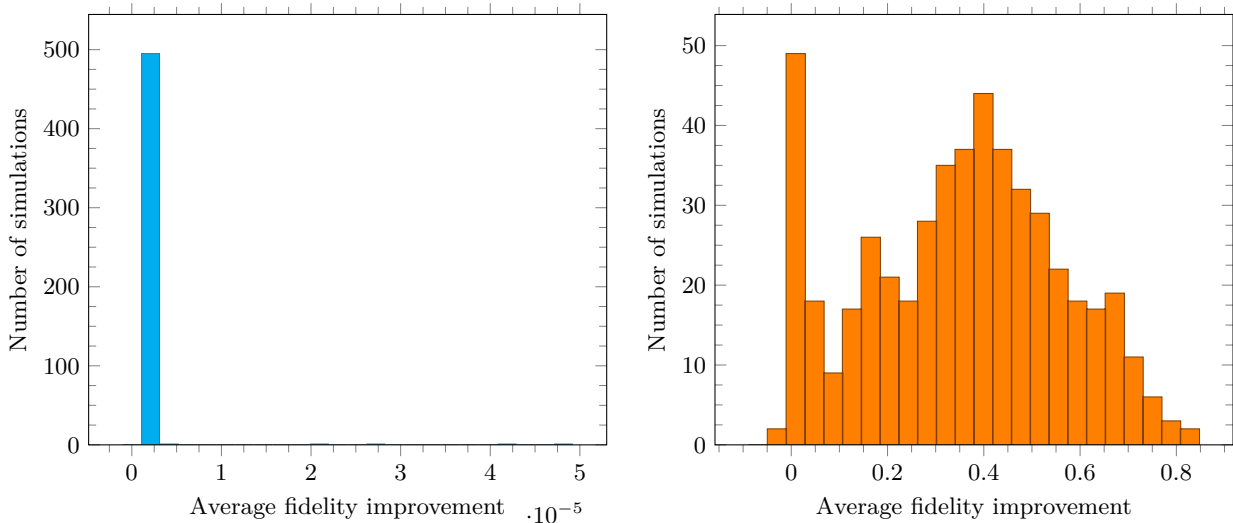[3] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

Figure 8. Improvement of simulations when changing the cost function in a two run optimization process that uses different cost functions to drive the descent algorithm. In the left plot, we started the descent on a random initial point, ran the optimization using $C^{(\text{wass})}(\vec{\alpha}, \vec{\beta})$ as cost function until convergence, and then we started the descent algorithm again but using $C^{(\text{fid})}(\vec{\alpha}, \vec{\beta})$ as cost function, starting from the final point of the previous descent. In the right part, the roles of the two cost functions are inverted (we start using $C^{(\text{fid})}(\vec{\alpha}, \vec{\beta})$ and then we use $C^{(\text{wass})}(\vec{\alpha}, \vec{\beta})$). The histograms represent the difference in average input-output fidelity (6) after the change of cost function, namely the difference between the fidelity achieved after the second descent and the fidelity after the first descent (positive values correspond to improved performances).Please notice the scale difference on the x-axis between the left and right plot.

[4] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, Quantum cryptography, Reviews of Modern Physics **74**, 145 (2002).

[5] H.-K. Lo, M. Curty, and K. Tamaki, Secure quantum key distribution, Nature Photonics **8**, 595 (2014).

[6] S. Pirandola, U. L. Andersen, L. Banchi, M. Berta, D. Bunandar, R. Colbeck, D. Englund, T. Gehring, C. Lupo, C. Ottaviani, J. L. Pereira, M. Razavi, J. S. Shaari, M. Tomamichel, V. C. Usenko, G. Vallone, P. Villoresi, and P. Wallden, Advances in quantum cryptography, Adv. Opt. Photon. **12**, 1012 (2020).

[7] D. Gottesman, Stabilizer codes and quantum error correction (1997), arXiv:quant-ph/9705052 [quant-ph].

[8] E. Knill, R. Laflamme, R. Martinez, and C. Negrevergne, Implementation of the five qubit error correction benchmark, arXiv preprint quant-ph/0101034 (2001).

[9] A. S. Fletcher, P. W. Shor, and M. Z. Win, Channel-adapted quantum error correction for the amplitude damping channel, IEEE Transactions on Information Theory **54**, 5705 (2008).

[10] R. L. Kosut, A. Shabani, and D. A. Lidar, Robust quantum error correction via convex optimization, Phys. Rev. Lett. **100**, 020502 (2008).

[11] S. Taghavi, R. L. Kosut, and D. A. Lidar, Channel-optimized quantum error correction, IEEE Transactions on Information Theory **56**, 1461 (2010).

[12] M. Chiani and L. Valentini, Short Codes for Quantum Channels With One Prevalent Pauli Error Type, IEEE Journal on Selected Areas in Information Theory **1**, 480 (2020).

[13] J. Koch, T. M. Yu, J. Gambetta, A. A. Houck, D. I. Schuster, J. Majer, A. Blais, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, Charge-insensitive qubit design derived from the cooper pair box, Physical Review A **76**, 042319 (2007).

[14] M. J. Peterer, S. J. Bader, X. Jin, F. Yan, A. Kamal, T. J. Gudmundsen, P. J. Leek, T. P. Orlando, W. D. Oliver, and S. Gustavsson, Coherence and decay of higher energy levels of a superconducting transmon qubit, Phys. Rev. Lett. **114**, 010501 (2015).

[15] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, Characterizing quantum supremacy in near-term devices, Nature Physics 10.1038/s41567-018-0124-x (2018).

[16] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and et al., Variational quantum algorithms, Nature Reviews Physics **3**, 625–644 (2021).

[17] J. Tilly, H. Chen, S. Cao, D. Picozzi, K. Setia, Y. Li, E. Grant, L. Wossnig, I. Rungger, G. H. Booth, and J. Tennyson, The variational quantum eigensolver: a review of methods and best practices (2021).

[18] S. Hadfield, Z. Wang, B. O'Gorman, E. Rieffel, D. Venturelli, and R. Biswas, From the quantum approximate optimization algorithm to a quantum alternating operator ansatz, Algorithms **12**, 34 (2019).

[19] P. D. Johnson, J. Romero, J. Olson, Y. Cao, and A. Aspuru-Guzik, Qvector: an algorithm for device-tailored quantum error correction (2017), arXiv:1711.02249 [quant-ph].

[20] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, Nature Communications 10.1038/s41467-018-07090-4 (2018).

[21] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Cost function dependent barren plateaus in shallow parametrized quantum circuits, Nature Communications 10.1038/s41467-021-21728-w (2021).

[22] G. De Palma, M. Marvian, D. Trevisan, and S. Lloyd, The quantum wasserstein distance of order 1, IEEE Transactions on Information Theory **67**, 6627 (2021).

[23] R. W. Hamming, Error detecting and error correcting codes, The Bell System Technical Journal **29**, 147 (1950).

[24] B. T. Kiani, G. D. Palma, M. Marvian, Z.-W. Liu, and S. Lloyd, Learning quantum data with the quantum earth mover's distance, Quantum Science and Technology **7**, 045002 (2022).

[25] L. Kim, S. Lloyd, and M. Marvian, Hamiltonian Quantum Generative Adversarial Networks, arXiv:2211.02584 10.48550/ARXIV.2211.02584 (2022).

[26] D. Herr, B. Obert, and M. Rosenkranz, Anomaly detection with variational quantum generative adversarial networks, Quantum Science and Technology **6**, 045004 (2021).

[27] E. R. Anschuetz and B. T. Kiani, Quantum variational algorithms are swamped with traps, Nature Communications **13**, 7760 (2022).

[28] B. Coyle, *Machine learning applications for noisy intermediate-scale quantum computers*, Ph.D. thesis, University of Edinburgh (2022).

[29] S. Chakrabarti, H. Yiming, T. Li, S. Feizi, and X. Wu, Quantum wasserstein generative adversarial networks, in *Advances in Neural Information Processing Systems* (2019) pp. 6781–6792.

[30] C. Villani, *Optimal transport: old and new*, Vol. 338 (Springer Science & Business Media, 2008).

[31] L. Ambrosio, Gradient flows in metric spaces and in the spaces of probability measures, and applications to fokker-planck equations with respect to log-concave measures, Bollettino dell'Unione Matematica Italiana **1**, 223 (2008).

[32] L. Ambrosio, A. Bressan, D. Helbing, A. Klar, E. Zuazua, L. Ambrosio, and N. Gigli, A user's guide to optimal transport, Modelling and Optimisation of Flows on Networks: Cetraro, Italy 2009, Editors: Benedetto Piccoli, Michel Rascle , 1 (2013).

[33] G. Monge, *Mémoire sur la théorie des déblais et des remblais* (Meémoires de l'Acadeémie royale des sciences de Paris vol 1781, 1781) p. 625–704.

[34] L. V. Kantorovich, On the translocation of masses, Journal of Mathematical Sciences **133**, 1381 (2006).

[35] A. M. Vershik, Long history of the monge-kantorovich transportation problem, The Mathematical Intelligencer **35**, 1 (2013).

[36] G. Peyré and M. Cuturi, Computational Optimal Transport: With Applications to Data Science, Foundations and Trends® in Machine Learning **11**, 355 (2019).

[37] D. S. Ornstein, An application of ergodic theory to probability theory, The Annals of Probability **1**, 43 (1973).

[38] G. De Palma and C. Rouzé, Quantum Concentration Inequalities, Annales Henri Poincaré **23**, 3391 (2022).

[39] G. De Palma, M. Marvian, C. Rouzé, and D. S. França, Limitations of variational quantum algorithms: a quantum optimal transport approach, PRX Quantum **4**, 010309 (2023).

[40] C.-N. Chou, P. J. Love, J. S. Sandhu, and J. Shi, Limitations of Local Quantum Algorithms on Random MAX-k-XOR and Beyond, in *49th International Colloquium on Automata, Languages, and Programming (ICALP 2022)*, Leibniz International Proceedings in Informatics (LIPIcs),

Vol. 229, edited by M. Bojańczyk, E. Merelli, and D. P. Woodruff (Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2022) pp. 41:1–41:20.

[41] G. De Palma and D. Trevisan, The wasserstein distance of order 1 for quantum spin systems on infinite lattices, arXiv:2210.11446 (2022).

[42] C. Rouzé and D. S. França, Learning quantum many-body systems from a few copies, arXiv:2107.03333 10.48550/ARXIV.2107.03333 (2021).

[43] F. B. Maciejewski, Z. Puchała, and M. Oszmaniec, Exploring quantum average-case distances: proofs, properties, and examples, arXiv:2112.14284 10.48550/ARXIV.2112.14284 (2021).

[44] E. Onorati, C. Rouzé, D. S. França, and J. D. Watson, Efficient learning of ground & thermal states within phases of matter, arXiv:2301.12946 10.48550/ARXIV.2301.12946 (2023).

[45] E. A. Carlen and J. Maas, An analog of the 2-Wasserstein metric in non-commutative probability under which the Fermionic Fokker–Planck equation is gradient flow for the entropy, Communications in Mathematical Physics **331**, 887 (2014).

[46] E. A. Carlen and J. Maas, Gradient flow and entropy inequalities for quantum Markov semigroups with detailed balance, Journal of Functional Analysis **273**, 1810 (2017).

[47] E. A. Carlen and J. Maas, Non-commutative calculus, optimal transport and functional inequalities in dissipative quantum systems, Journal of Statistical Physics **178**, 319 (2020).

[48] C. Rouzé and N. Datta, Concentration of quantum states from quantum functional and transportation cost inequalities, Journal of Mathematical Physics **60**, 012202 (2019).

[49] N. Datta and C. Rouzé, Relating relative entropy, optimal transport and Fisher information: A quantum HWI inequality, Annales Henri Poincaré **21**, 2115 (2020).

[50] T. Van Vu and Y. Hasegawa, Geometrical Bounds of the Irreversibility in Markovian Systems, Phys. Rev. Lett. **126**, 010601 (2021).

[51] M. Wirth, A dual formula for the noncommutative transport distance, Journal of Statistical Physics **187**, 1 (2022).

[52] L. Gao, M. Junge, and N. LaRacuente, Fisher information and logarithmic sobolev inequality for matrix-valued functions, Annales Henri Poincaré **21**, 3409 (2020).

[53] Y. Chen, T. T. Georgiou, L. Ning, and A. Tannenbaum, Matricial Wasserstein-1 distance, IEEE control systems letters **1**, 14 (2017).

[54] E. K. Ryu, Y. Chen, W. Li, and S. Osher, Vector and matrix optimal mass transport: theory, algorithm, and applications, SIAM Journal on Scientific Computing **40**, A3675 (2018).

[55] Y. Chen, T. T. Georgiou, and A. Tannenbaum, Matrix optimal mass transport: a quantum mechanical approach, IEEE Transactions on Automatic Control **63**, 2612 (2018).

[56] Y. Chen, T. T. Georgiou, and A. Tannenbaum, Wasserstein geometry of quantum states and optimal transport of matrix-valued measures, in *Emerging Applications of Control and Systems Theory* (Springer, 2018) pp. 139–150.

[57] J. Agredo, A Wasserstein-type distance to measure deviation from equilibrium of quantum Markov semigroups, Open Systems & Information Dynamics **20**, 1350009 (2013).

[58] J. Agredo, On exponential convergence of generic quantum Markov semigroups in a Wasserstein-type distance, International Journal of Pure and Applied Mathematics

107, 909 (2016).

[59] K. Ikeda, Foundation of quantum optimal transport and applications, Quantum Information Processing **19**, 25 (2020).

[60] F. Golse, C. Mouhot, and T. Paul, On the mean field and classical limits of quantum mechanics, Communications in Mathematical Physics **343**, 165 (2016).

[61] E. Caglioti, F. Golse, and T. Paul, Towards Optimal Transport for Quantum Densities, arXiv:2101.03256 10.48550/ARXIV.2101.03256 (2021).

[62] F. Golse, The quantum N-body problem in the mean-field and semiclassical regime, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **376**, 20170229 (2018).

[63] F. Golse and T. Paul, The Schrödinger equation in the mean-field and semiclassical regime, Archive for Rational Mechanics and Analysis **223**, 57 (2017).

[64] F. Golse and T. Paul, Wave packets and the quadratic Monge–Kantorovich distance in quantum mechanics, Comptes Rendus Mathematique **356**, 177 (2018).

[65] E. Caglioti, F. Golse, and T. Paul, Quantum optimal transport is cheaper, Journal of Statistical Physics **181**, 149 (2020).

[66] S. Friedland, M. Eckstein, S. Cole, and K. Życzkowski, Quantum Monge-Kantorovich Problem and Transport Distance between Density Matrices, Phys. Rev. Lett. **129**, 110402 (2022).

[67] S. Cole, M. Eckstein, S. Friedland, and K. Życzkowski, Quantum Optimal Transport, arXiv:2105.06922 10.48550/ARXIV.2105.06922 (2021).

[68] R. Duvenhage, Optimal quantum channels, Phys. Rev. A **104**, 032604 (2021).

[69] R. Bistroń, M. Eckstein, and K. Życzkowski, Monotonicity of the quantum 2-Wasserstein distance, arXiv:2204.07405 10.48550/ARXIV.2204.07405 (2022).

[70] T. Van Vu and K. Saito, Thermodynamic Unification of Optimal Transport: Thermodynamic Uncertainty Relation, Minimum Dissipation, and Thermodynamic Speed Limits, arXiv preprint arXiv:2206.02684 (2022).

[71] G. De Palma and D. Trevisan, Quantum optimal transport with quantum channels, Annales Henri Poincaré **22**, 3199 (2021).

[72] K. Życzkowski and W. Slomczynski, The Monge distance between quantum states, Journal of Physics A: Mathematical and General **31**, 9095 (1998).

[73] K. Życzkowski and W. Slomczynski, The Monge metric on the sphere and geometry of quantum states, Journal of Physics A: Mathematical and General **34**, 6689 (2001).

[74] I. Bengtsson and K. Życzkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement* (Cambridge University Press, 2017).

[75] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2000).

[76] E. B. Vinberg, *Linear representations of groups* (Boston: Birkhauser Verlag, 1989).

[77] C. Dankert, R. Cleve, J. Emerson, and E. Livine, Exact and approximate unitary 2-designs and their application to fidelity estimation, Physical Review A **80**, 012304 (2009).

[78] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. (Springer, New York, NY, USA, 2006).

[79] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, Evaluating analytic gradients on quantum hardware, Physical Review A **99**, 032331 (2019).

[80] D. Gottesman, An introduction to quantum error correction and fault-tolerant quantum computation (2009).

[81] E. Knill, R. Laflamme, R. Martinez, and C. Negrevergne, Benchmarking quantum computers: The five-qubit error correcting code, Physical Review Letters **86**, 5811 (2001).

[82] A. Arrasmith, M. Cerezo, P. Czarnik, L. Cincio, and P. J. Coles, Effect of barren plateaus on gradient-free optimization, Quantum 5, 558 (2021).