Error mitigation in variational quantum eigensolvers using probabilistic machine learning

John Rogers,1 Gargee Bhattacharyya,2 Marius S. Frank,2 Tao Jiang,3 Ove Christiansen,4 Yong-Xin Yao,3,5† and Nicola Lanata†2,6

1Department of Physics and Astronomy, Texas A&M University, College Station, Texas 77845, USA
2Department of Physics and Astronomy, Aarhus University, 8000, Aarhus C, Denmark
3Ames Laboratory, Ames, Iowa 50011, USA
4Department of Chemistry, Aarhus University, 8000, Aarhus C, Denmark
5Department of Physics and Astronomy, Iowa State University, Ames, Iowa 50011, USA
6Nordita, KTH Royal Institute of Technology and Stockholm University, Roslagstullsbacken 23, 10691 Stockholm, Sweden

(Dated: November 18, 2021)

Quantum-classical hybrid schemes based on variational quantum eigensolvers (VQEs) may transform our ability of simulating materials and molecules already within the next few years. However, one of the main obstacles to overcome in order to achieve practical near-term quantum advantage is to improve our ability of mitigating the “noise effects”, characteristic of the current generation of quantum processing units (QPUs). To this end, here we design a method based on probabilistic machine learning, which allows us to mitigate the noise by imbuing within the computation prior (data independent) information about the variational landscape. We perform benchmark calculations of a 4-qubit impurity model using the IBM open-source framework for quantum computing Qiskit, showing that our method improves dramatically the accuracy of the VQE outputs. Finally, we show that applying our method makes quantum-embedding simulations of the Hubbard model with a VQE impurity solver considerably more reliable.

I. INTRODUCTION

Quantum computers have the potential of boosting dramatically our ability of simulating quantum-mechanical systems. Computational chemistry is believed to be one of the most promising fields that will benefit significantly from the noisy intermediate-scale quantum technologies within the next few years [1–6]. As the number of qubits required for representing the active degrees of freedom of molecules of limited size is relatively small. Furthermore, the quantum embedding (QE) methods [7, 8] may allow us to benefit from quantum devices also for simulating large systems (both molecules and periodic materials), by treating with QPUs only the most active fraction of degrees of freedom, while treating the other modes (at the mean-field level) on classical devices [9–12]. A cardinal obstacle to realizing this ambitious program is that, even if several effective error-mitigation methodologies have been already developed and implemented [13, 15], the data produced by quantum devices are not yet sufficiently reliable. In this work we propose a method to tackle this problem, which is complementary to the existing approaches and may improve substantially the accuracy obtained in applications of VQE frameworks [16,19].

All error-mitigation techniques proposed so far focus on improving the computation of expectation values with respect to a parametrized quantum circuit, at a single point in parameter space [13, 15, 20, 27]. For example, the zero-noise extrapolation technique measures an observable at a single parameter point for a set of equivalent circuits with different noise strengths, followed by a fitting with analytical functions such as polynomials and extrapolates expectation value to zero-noise limit [13, 14, 20, 22, 23]. Probabilistic error cancellation and randomized compiling approaches convert the expectation value with respect to a parametrized circuit at a single point to a sum of estimations with equivalent random circuits, which effectively tailor the coherent noise to stochastic error [15, 20, 21, 24]. Rather than mitigating the error of the VQE measurements for each individual variational state, as in the methods mentioned above, here we propose a complementary approach, based on the idea of mitigating the error for the whole variational landscape simultaneously (exploiting the underlying correlations between the VQEs outputs for different variational parameters). Specifically, using a probabilistic machine learning method based on “Gaussian process regression (GPR),” rooted in Bayesian statistics, we imbue within the computation: (i) our prior knowledge about general (data independent) features of the variational landscape and (ii) the expectation value of the desired observable for variational parameters corresponding to limiting cases, where the calculations can be done exactly on a classical computer, at a negligible computational cost. This approach can be applied in combination with any of the error-mitigation techniques mentioned above, improving considerably the accuracy of VQE calculations.

The manuscript is structured as follows. In Secs. I1 and I4 we present our error-mitigation formalism from a general perspective; in Sec. I5 we show benchmark calculations of a Fermionic impurity model using our method the IBM quantum-computing framework...
II. GENERAL FEATURES OF THE VARIATIONAL LANDSCAPE

Let us consider a generic Hamiltonian $\hat{H}$ and a variational space $\{|\Psi(\theta)\rangle\}$, where the components of $\theta = (\theta^1, \ldots, \theta^L)$ are real numbers parametrizing the trial quantum states. Our goal is to determine:

$$\hat{\theta} = \arg\min_{\theta} E(\theta), \quad (1)$$

$$E(\theta) = \langle \phi(\theta) | \hat{H} | \phi(\theta) \rangle, \quad (2)$$

Within the VQE frameworks, the state $|\phi(\theta)\rangle$ is prepared utilizing parametrized quantum gates, and $E(\theta)$ is estimated from a series of quantum measurements for each $\theta$. However, the outputs of real available intermediate-scale quantum devices are affected by spurious effects such as decoherence and hardware imperfections, resulting in both random and systematic noise.

The key idea at the basis of our method for mitigating the noise in VQE frameworks is to exploit the fact that the variational landscape satisfies exact properties that are known beforehand. A trivial example is that $E(\theta)$ is generally a smooth function. Furthermore, depending on the specific variational ansatz, a wealth of additional information is available. Here we illustrate this point focusing on a generic ansatz represented as follows:

$$|\phi(\theta)\rangle = \hat{U}(\theta)|\phi_0\rangle, \quad (3)$$

where $|\phi_0\rangle$ is a single-particle state (e.g., the Hartree-Fock solution of $\hat{H}$) and $\hat{U}(\theta)$ is a unitary transformation represented as follows:

$$\hat{U}(\theta) = \prod_{i=1}^{d} \prod_{l=1}^{M_i} e^{\hat{G}_{ilm} \theta^l}, \quad (4)$$

where $\theta^l$ are variational parameters, the generators $\hat{G}_{ilm}$ are Pauli strings expressed in the Hartree-Fock basis, $d$ is the number of variational parameters and $M_i$ is the number of Pauli strings for each variational parameter. Note that a Pauli string is defined as a generic tensor product of Pauli operators $\hat{P} = \otimes_i \hat{P}_i$, where $\hat{P}_i \in \{I, X^i, Y^i, Z^i\}$, $i$ is a generic site label, $X^i$, $Y^i$ and $Z^i$ are the corresponding local Pauli operators and $I$ is the identity.

A first observation is that, for the variational ansatz above, Eq. (3) is known exactly when $\theta^l = 0 \forall l$, as in this case the variational state reduces to the Hartree-Fock solution. Furthermore, Eq. (4) can be evaluated exactly (at a negligible cost) for all transformations $\hat{U}(\theta)$ involving only single Pauli rotation gates. This can be readily understood by noticing that, in qubit representation, such states are represented as: $|\phi(\theta)\rangle = \otimes_i \hat{U}_i(\theta_i)|\phi_0\rangle$, where $|\phi_0\rangle = \otimes_i |s_i\rangle$, $|s_i\rangle$ are eigenstates of $Z_i$ with eigenvalues $s_i \in \{0, 1\}$ and $\hat{U}_i(\theta_i)$ are single-qubit unitary transformations. Therefore, by expressing $\hat{H}$ as a weighted sum of Pauli strings $\hat{P}_h = \otimes_i \hat{P}_{hl}$:

$$\hat{H} = \sum_h w_h \hat{P}_h \quad (5)$$

the expectation value of the Hamiltonian can be calculated as follows:

$$\langle \phi(\theta) | \hat{H} | \phi(\theta) \rangle = \sum_h w_h \langle \phi(\theta) | \hat{P}_h | \phi(\theta) \rangle, \quad (6)$$

where:

$$\langle \phi(\theta) | \hat{P}_h | \phi(\theta) \rangle = \prod_i \langle s_i | \hat{U}_i(\theta_i) \hat{P}_{hi} \hat{U}_i(\theta_i) | s_i \rangle. \quad (7)$$

Another important observation is that, as pointed out in Ref. [33], since $G_{lm}^2 = I \forall l, m$ we have:

$$e^{i\hat{G}_{lm} \theta^l} = \cos(\theta^l) I + i \sin(\theta^l) \hat{G}_{lm}. \quad (8)$$

Therefore, the variational energy [Eq. (2)] can be expanded in the following form:

$$E(\theta) = \sum_{s=1}^{S} \xi_s T_s(\theta), \quad (9)$$

where $S = \prod_{l=1}^{d} (2M_l + 1)$ and $T_s(\theta)$ are known analytical functions, while the coefficients $\xi_s$ depend on the specific operator $\hat{H}$. Since $S$ grows exponentially as a function of $d$, the number of terms in Eq. (9) can be managed more effectively when the number of variational parameters is smaller. Nevertheless, this equation poses valuable non-trivial constraints on the form of the variational landscape, for all values of $d$ and $M_i$.

III. GPR-BASED ERROR MITIGATION

Our approach consists in using the GPR framework for imbuing within the solution our prior knowledge about global properties of the variational landscape $E(\theta)$ (such as the information mentioned above). As we are going to show, since such information correlates the data with each other, it can mitigate substantially the error arising from individual QPU measurements.

A. The GPR framework

We formulate the problem assuming that our prior information about $\tilde{E}(\theta)$ is not necessarily deterministic, but probabilistic. In particular, we assume that, based on our prior expectations (before to make any measurement), $E(\theta)$ can be any function sampled from a probability distribution $P[E]$ (called prior). Following the GPR
framework, we will also assume that \( P[\mathcal{E}] \) is Gaussian-distributed. Our specific choice of \( P[\mathcal{E}] \) will be explained below in Sec. 3.B.

Let us assume to have evaluated the variational energy on QPUs for a finite set of variational parameters \( \{ \theta_\alpha \} \), obtaining the following measurements:

\[
D = \{ (\theta_\alpha, \mathcal{E}_\alpha) \mid \alpha = 1, \ldots, n \} .
\] (10)

Within our context of application, it will be essential to take into account that the evaluation of the data is non-deterministic, as this is intrinsically true also for quantum-mechanical measurements performed on fault-free machines. Here we are going to model such uncertainty in terms of the following probability distribution:

\[
P[D|\mathcal{E}] \propto \exp \{ -U_D[\mathcal{E}] \} \tag{11}
\]

\[
U_D[\mathcal{E}] = \sum_{\alpha=1}^{n} \frac{1}{2\sigma^2(\theta_\alpha)} (\mathcal{E}(\theta_\alpha) - \mathcal{E}_\alpha)^2 ,
\] (12)

representing a Gaussian noise for \( \mathcal{E}_\alpha \), with respect to the (unknown) underlying respective values \( \mathcal{E}(\theta_\alpha) \). Note that we have allowed the different data \( \mathcal{E}_\alpha \) to have different uncertainties \( \sigma(\theta_\alpha) \) (which, in the GPR related literature, is referred to as “heteroscedastic GPR”). This will be particularly important for us, as some of our data will be generated with the QPU, while other data — corresponding to limiting cases, where the variational energy is known exactly — will be determined exactly.

From the Bayes rule, given the data \( D \) (see Eq. (10)) and a prior probability \( P[\mathcal{E}] \), the “posterior” conditional probability for the function \( \mathcal{E} \) is the following:

\[
P[\mathcal{E}|D] \propto P[\mathcal{E}] \exp \{ -U_D[\mathcal{E}] \} ,
\] (13)

which represents the probability distribution for the function \( \mathcal{E} \) to learn, given a data set \( D \). Remarkably, since both \( P[\mathcal{E}] \) and \( P[D|\mathcal{E}] \), are Gaussian, the following quantities can be calculated exactly:

\[
\langle \mathcal{E}(\theta) \rangle = \int \mathcal{D}[\mathcal{E}] P[\mathcal{E}|D] \mathcal{E}(\theta) \tag{14}
\]

\[
\Sigma^2(\theta) = \int \mathcal{D}[\mathcal{E}] P[\mathcal{E}|D] \left( \mathcal{E}^2(\theta) - \langle \mathcal{E}(\theta) \rangle^2 \right) , \tag{15}
\]

where \( \mathcal{D}[\mathcal{E}] \) is the functional-integral measure, Eq. (14) represents our prediction for \( \mathcal{E} \) at \( \theta \) and Eq. (15) allows us to estimate the expected uncertainty.

Below we are going to derive explicitly Eqs. (14) and (15) for our specific application of GPR as an error-mitigation method.

### B. Parametric prior

In the present work we will focus on the case discussed in Sec. 1, where it is known that \( \mathcal{E}(\theta) \) can be expressed as in Eq. (9). We formalize this information by introducing the following prior:

\[
P[\mathcal{E}] \propto \prod_r d\xi_r \exp \left\{ \frac{1}{2p} \int d\theta \left( \xi(\theta) - \sum_s \xi_s T_s(\theta) \right)^2 \right\} \times \exp \left\{ -\frac{t}{2} \int d\theta \mathcal{E}^2(\theta) \right\} , \tag{16}
\]

where:

\[
d\theta = \prod_{t=1}^d d\theta_t ,
\] (17)

\( d \) is the number of variational parameters and \( \eta \to 0 \). The parameter \( t \) encodes our prior information concerning the range of the variational energy, as the second factor in Eq. (16) vanishes exponentially for \( \mathcal{E}(\theta) \gg t^{1/2} \).

### C. Parametric GPR formalism

Our goal is to calculate the “moments” of the energy function with respect to the posterior probability, i.e.:

\[
\langle \mathcal{E}^p(\theta) \rangle = \int \mathcal{D}[\mathcal{E}] P[\mathcal{E}|D] \mathcal{E}^p(\theta) , \tag{18}
\]

for an arbitrary integer power \( p \), where \( \mathcal{D}[\mathcal{E}] \) is a “functional-integral measure”, whose precise mathematical meaning is the following. We start by considering:

\[
\mathcal{D}[\mathcal{E}] = \prod_{\theta \in M_\ell} d\mathcal{E}(\theta) , \tag{19}
\]

where the parameters \( \bar{\theta} \) are distributed on a uniform mesh \( M_\ell \) within the domain of \( \mathcal{E} \), with spacing \( \epsilon \). Furthermore, the integrals appearing in the exponents of Eq. (16) have to be interpreted as discrete summations, i.e.:

\[
P[\mathcal{E}] \propto \prod_r d\xi_r \prod_{\theta \in M_\ell} \delta(\mathcal{E}(\theta) - \sum_s \xi_s T_s(\theta)) \times \exp \left\{ -\frac{t}{2} \sum_{\theta \in M_\ell} \epsilon^d \mathcal{E}^2(\theta) \right\} , \tag{20}
\]

where we have already taken the limit for \( \eta \to 0 \). Subsequently, we will evaluate the resulting integral [Eq. (18)] in the “continuum limit,” i.e., for \( \epsilon \to 0 \).

Let us partition the mesh \( M_\ell \) as follows:

\[
M_\ell = \{ \theta \} \cup \{ \theta_\alpha \mid \alpha = 1, \ldots, n \} \cup M_\ell' , \tag{21}
\]

where \( \theta \) is the parameter appearing in Eq. (18) (which is generally called “test point”), \( \{ \theta_\alpha \mid \alpha = 1, \ldots, n \} \) are the so-called “training data points”, see Eq. (22), while \( \{ \theta' \in M_\ell' \} \) are the other points of the mesh. With these definitions, Eq. (18) can be expressed as follows:

\[
\langle \mathcal{E}^p(\theta) \rangle \propto \prod_r d\xi_r \left( \sum_s \xi_s T_s(\theta) \right)^p
\]

where 

\[
\frac{1}{2p} \int d\theta \left( \xi(\theta) - \sum_s \xi_s T_s(\theta) \right)^2
\]

and 

\[
\exp \left\{ -\frac{t}{2} \int d\theta \mathcal{E}^2(\theta) \right\}
\]

are the two factors in the integrand.
\[ \times \exp \left\{ -\frac{t}{2} \sum_{\theta_a} \xi_t T_s(\theta)^2 \right\} \]
\[ \times \exp \left\{ -\frac{t}{2} \sum_{\theta_a} \epsilon_t \sum_{s} \xi_t T_s(\theta_a)^2 \right\} \]
\[ \times \exp \left\{ -\frac{n}{2} \sum_{\alpha}^{1} \frac{1}{2\sigma^2(\theta_\alpha)} \left( \sum_{s} \xi_t T_s(\theta_\alpha) - \xi_{\alpha} \right)^2 \right\} \]
\[ \times \exp \left\{ -\frac{t}{2} \sum_{\theta' \in M_E} \epsilon_t \sum_{s} \xi_t T_s(\theta')^2 \right\} . \] (22)

Since the training data set is finite, the summations at the exponents of the second, third and fifth factors of Eq. (22) simplify as follows for \( \epsilon \to 0 \):

\[ \langle E^p(\theta) \rangle = \int \prod_r d\xi_r P[\xi|D] \left( \sum_s \xi_s T_s(\theta) \right)^p , \] (23)

where:

\[ P[\xi|D] \propto \exp \left\{ -\frac{n}{2} \sum_{\alpha}^{1} \frac{1}{2\sigma^2(\theta_\alpha)} \left( \sum_{s} \xi_t T_s(\theta_\alpha) - \xi_{\alpha} \right)^2 \right\} \]
\[ \times \exp \left\{ -\frac{t}{2} \int d\theta \left( \sum_{s} \xi_t T_s(\theta) \right)^2 \right\} . \] (24)

In fact, the discrete summation of the fifth factor of Eq. (22) contains all points of the mesh \( M_E \), except for a finite subset (the training data points \{ \theta_\alpha \} and the test point \( \theta \)), whose contribution becomes irrelevant in the continuum limit.

The probability distribution [Eq. (24)] is a Gaussian function of \( \xi \), which can be represented as follows:

\[ P[\xi|D] \propto \exp \left\{ -\sum_{s}^{1} \frac{1}{2} A_{ss'} \xi_s \xi_{s'} + \sum_s J_s \xi_s \right\} , \] (25)

where the coefficients \( A_{ss'} \) and \( J_s \) depend on the training data, the “hyper parameter” \( t \) and the following integrals:

\[ M_{ss'} = \int d\theta T_s(\theta) T_{s'}(\theta) \] (26)

(which can generally be evaluated analytically).

The corresponding expectation value for \( E(\theta) \), see Eq. (14), is given by the following equation:

\[ \langle E(\theta) \rangle = \sum_s \xi_s T_s(\theta) , \] (27)

where:

\[ \xi_s = \int \prod_r d\xi_r P[\xi|D] \xi_s = \sum_{s'} A_{ss'}^{-1} J_{s'} . \] (28)

Instead, the corresponding variance, see Eq. (15), is given by:

\[ \Sigma^2(\theta) = \sum_{ss'} \left( A_{ss'}^{-1} - \xi_s \xi_{s'} \right) T_s(\theta) T_{s'}(\theta) , \] (29)

where we used that:

\[ \int \prod_r d\xi_r P[\xi|D] \xi_s \xi_{s'} = A_{ss'}^{-1} . \] (30)

D. Simplified formalism without energy cutoff

Imbuing our knowledge about the parameter \( t \) within the prior would presumably contribute to reducing the number of necessary training data (i.e., the number of necessary QPU evaluations). However, for simplicity, in the present work we will not exploit such information, i.e., we will take the limit for \( t \to 0^+ \) (as if no bound to \( E(\theta) \) was available a priori).

Within this working hypotheses, the coefficients \( A_{ss'} \) and \( J_s \) are the following:

\[ A_{ss'} = \sum_{\alpha}^{1} \frac{1}{\sigma^2(\theta_\alpha)} T_s(\theta_\alpha) T_{s'}(\theta_\alpha) \] (31)
\[ J_s = \sum_{\alpha}^{1} \frac{1}{\sigma^2(\theta_\alpha)} f_s T_s(\theta_\alpha) . \] (32)

Note that, Eqs. (31) and (32), which were derived here from a probabilistic perspective (for the special case \( t \to 0^+ \)), coincide with the result obtained based on the so-called “generalized least-square method” for \( \xi_s \) (see Eq. (28)).

It shall be noted that in this section we assumed that the data \( D \) are Gaussian-distributed around the underlying exact function, see Eq. (12). While this is rigorously justified for fault-free devices (because of the central limit theorem), this is not the case for real QPUs. On the other hand, since Eq. (6) is a non-trivial condition correlating \( E(\theta) \) over its entire domain, we shall expect that enforcing it can contribute to mitigate the spurious sources of systematic noise present in real quantum devices—as we are going to show explicitly in Sec. IV.

IV. ENERGY MINIMIZATION AND EVALUATION OF OBSERVABLES

In this work, the problems of evaluating the variational landscape and calculating its minimum will be decoupled as follows:

- Use the available prior information and the data produced by QPUs for evaluating the expected variational landscape, see Eq. (27).
- Perform the minimization of \( \langle E(\theta) \rangle \), whose explicit analytical form is provided by Eq. (27), on a classical computer.
This allows us to evaluate the optimal variational parameters $\theta$ and the corresponding energy $E(\theta)$.

We point out that the method outlined above, which is based on the idea of learning the entire variational landscape before to perform the energy minimization, is not the only possibility. For example, in another context [34], molecular structures (whose atomic positions are local minima of the electronic energy) have been obtained using minimization of GPR representations updated on the fly. This suggests that Eqs. (27) and (29) could be applied to our problem in a similar fashion. Alternative approaches such as this, or classic Bayesian-optimization methods, will be investigated and tested in future work.

Let us now consider the problem of evaluating the expectation value of a generic observable $\hat{O}$ with respect to our approximation $|\phi(\theta)\rangle$ to the ground state of $\hat{H}$. Rather than performing repeated measurements of $\hat{O}$ with respect to $|\phi(\theta)\rangle$, we will perform measurements on a series of different variational parameters:

$$D_{\theta} = \{ |\theta_{\alpha}, O_{\alpha}\rangle | \alpha = 1, \ldots, n \},$$

and apply the same procedure used above, for estimating the function:

$$O(\theta) = \langle \phi(\theta) | \hat{O} | \phi(\theta) \rangle.$$  

In fact, all steps described above for $\hat{H}$ remain applicable for any observable $\hat{O}$. Once the expansion coefficient characterizing $O(\theta)$ are calculated with this procedure, the ground-state expectation value of $\hat{O}$ is estimated by evaluating Eq. (34) at $\theta$.

In the next section we will benchmark the methodology proposed above on a 4-qubit impurity-model Hamiltonian.

V. BENCHMARK CALCULATIONS OF A 4-QUBIT IMPURITY MODEL

Let us consider a 4-qubit Fermionic Hamiltonian represented as follows:

$$\hat{H}_{\text{emb}} = \frac{U}{2} (\hat{n}_c - 1)^2 + D \sum_{\sigma = \uparrow, \downarrow} (c_{\sigma}^\dagger d_{\sigma} + d_{\sigma}^\dagger c_{\sigma})$$

$$+ \lambda \sum_{\sigma = \uparrow, \downarrow} d_{\sigma}d_{\sigma}^\dagger,$$  

(35)

where $c_{\sigma}^\dagger$ and $c_{\sigma}$ are the creation and annihilation operators of the so-called “impurity degrees of freedom”, $d_{\sigma}^\dagger$ and $d_{\sigma}$ are the creation and annihilation operators of the so-called “bath degrees of freedom”, $\sigma \in \{ \uparrow, \downarrow \}$ is the spin label, $\hat{n}_c = \sum_{\sigma} c_{\sigma}^\dagger c_{\sigma}$, $U$ is the Hubbard-repulsion parameter for the impurity degrees of freedom, while $\lambda$ and $D$ are the coupling constants characterizing the bath of the impurity model and its coupling to the impurity, respectively. From now on we will refer to Eq. (35) as “embedding Hamiltonian” (EH). This terminology will be clarified in Sec. (VI), where we will explain that a broad class of QE methods require to calculate recursively the ground-state of Eq. (35), for different EH parameters $\lambda$ and $D$.

In this section we will benchmark our error-mitigation method for estimating the ground state $|\Phi(D, \lambda, \theta)\rangle$ of $\hat{H}_{\text{emb}}$ within the subspace generated by states with 2 electrons (i.e., half of the maximum possible occupation within the EH space). The QPU data will be generated using the IBM open-source framework for quantum computing Qiskit [28], which provides methods for manipulating quantum programs on real quantum computers, as well as on classical QPU simulators. Real QPU calculations were made using the IBM Quantum Experience [35] which serves as a interface for submitting and running jobs on IBM QPUs.

A. Qubit representation of the EH

As in Ref. [11], we transform the EH (see Eq. (35)) into the so-called “parity mapping” [36], leading to the following qubit representation:

$$|\phi(\theta)\rangle = e^{\frac{i}{2} \theta_1 Y_2} e^{\frac{i}{2} \theta_2 Y_1} e^{\frac{i}{2} \theta_3 Y_1} e^{\frac{i}{2} \theta_4 Y_2} |\phi_0\rangle,$$  

(37)

where $\theta = (\theta_1, \theta_2)$, with each angle in the range of $[-\pi, \pi]$ and $|\phi_0\rangle$ is the spin-restricted Hartree-Fock ground-state solution of $\hat{H}_{\text{emb}}$. The angles for the single-qubit rotations are set to be equal because of spin rotational symmetry.

Using the procedure described in Sec. [11] it can be readily shown that the variational energy $E(\theta)$ can be expressed as a linear combination of 25 trigonometric functions:

$$E(\theta) = \sum_{i=0}^{4} \sum_{j=0}^{4} \xi_{ij} \cos \left( \frac{\theta_1}{2} \right)^i \sin \left( \frac{\theta_1}{2} \right)^{4-i} \times \cos \left( \frac{\theta_2}{2} \right)^j \sin \left( \frac{\theta_2}{2} \right)^{4-j}.$$  

(38)

Note that here we use for convenience a composite index $s = (i, j)$ to label the coefficients $\xi_s$. 

B. The VQE ansatz

To calculate the spin-singlet ground state of the EH we use the unitary coupled cluster ansatz with single and double excitations (UCCSD) [37–39]:

$$|\phi(\theta)\rangle = e^{\frac{i}{2} \theta_1 Y_2} e^{\frac{i}{2} \theta_2 Y_1} e^{\frac{i}{2} \theta_3 Y_1} e^{\frac{i}{2} \theta_4 Y_2} |\phi_0\rangle,$$  

(37)

where $\theta = (\theta_1, \theta_2)$, with each angle in the range of $[-\pi, \pi]$ and $|\phi_0\rangle$ is the spin-restricted Hartree-Fock ground-state solution of $\hat{H}_{\text{emb}}$. The angles for the single-qubit rotations are set to be equal because of spin rotational symmetry.

Note that Eq. (37) is a special case of Eqs. (3) and (4). Using the procedure described in Sec. [11] it can be readily shown that the variational energy $E(\theta)$ can be expressed as a linear combination of 25 trigonometric functions:
Figure 1. Benchmark calculations of the EH performed on the IBM 5 qubit QPU ibmq_quito. Total energy $E$, impurity double-occupancy $\langle \hat{n}_c \hat{n}_c \rangle$ and components of the single-particle density matrix $\sum_\sigma \langle c_{\sigma \uparrow}^\dagger c_{\sigma \uparrow} \rangle$, $\sum_\sigma \langle d_{\sigma \uparrow}^\dagger d_{\sigma \uparrow} \rangle$. The exact values are compared with the VQE outputs produced using 3 different methods: (i) the COBYLA optimizer, (ii) the 1D optimizer and (iii) our parametric-GPR optimizer.

C. Parametric GPR Optimization

As explained in Sec. IV in the present work we decouple the problems of learning the energy landscape and calculating the optimal variational parameters. To determine the coefficients $\bar{\xi}_s$ we generate data on the following mesh of variational parameters:

$$T = \left\{ \left( \frac{\pi}{5} a, \frac{\pi}{5} b \right) \mid a, b \in \{ -5, -4, \cdots , 4 \} \right\} .$$

Note that the number of points $\theta_\alpha \in T$ (Eq. (39)) is 100, which exceeds the number of trigonometric functions necessary for expanding the variational landscape (that is 25, as explained in Sec. V B). This allows us to learn the whole variational landscape from data at once.

For the 10 data points in $T$ such that $b = 0$, the variational energy can be determined on a classical computer at a negligible computational cost, with uncertainty $\sigma_\alpha = 0$ (see Eqs. (11) and (12)). In fact, as explained in Sec. III $\theta = 0$ corresponds to the Hartree-Fock solution, while the other 9 points in $T$ with $b = 0$ correspond to single-qubit rotations.

The other 90 training data points were generated by averaging the energy over $N = 4096$ QPU measurements (shots). On fault-free machines the corresponding uncertainty could be assumed to be given by the following equation:

$$\sigma^2_\alpha = \frac{1}{N} \sum_{\nu} \left( E_{\alpha \nu} - E_\alpha \right)^2 .$$

Instead, on real QPUs, we consider $\sigma_\alpha$ as a machine-dependent parameter, which depends mainly on details such as hardware imperfections.
Figure 2. Benchmark calculations of the EH performed with the Qiskit QASM simulator with noise model and coupling map extracted from the IBM machine ibmq_qquito. Total energy $E$, impurity double-occupancy $\langle \hat{n}_c^\uparrow \hat{n}_c^\downarrow \rangle$ and components of the single-particle density matrix $\sum_\sigma \langle c^{\dagger}_\sigma d_\sigma \rangle, \sum_\sigma \langle d^{\dagger}_\sigma d_\sigma \rangle$. The exact values are compared with the VQE outputs produced using 3 different methods: (i) the COBYLA optimizer, (ii) the 1D optimizer and (iii) our parametric-GPR optimizer.

As explained in Sec. [IV], once the coefficients of the variational landscape are estimated from the data, using Eq. (27), the corresponding analytical approximation to $E(\theta)$ can be minimized on a classical computer, and all of the desired observables can be evaluated.

We point out that, besides enforcing that the variational landscape satisfies exactly the mathematical form discussed in Sec. [V B], the training data points with $\theta_2 = 0$ act as boundary conditions, enforcing a-priori that the GPR variational landscape learned from QPU measurements is exact for these variational parameters. The resulting error mitigation will be assessed below with numerical experiments, both on simulators and actual quantum devices.

D. Numerical tests on real machines and simulators

In this subsection we show ground-state calculations of the EH for the total energy $E$, the impurity double-occupancy $\langle \hat{n}_c^\uparrow \hat{n}_c^\downarrow \rangle$ and the components of the single-particle density matrix $\sum_\sigma \langle c^{\dagger}_\sigma d_\sigma \rangle, \sum_\sigma \langle d^{\dagger}_\sigma d_\sigma \rangle$. In all of our calculations we have set $2D$ as energy unit and assumed that $\sigma_0/D = 0.2$.

The calculations shown in Fig. 1 were obtained using data produced by the IBM 5 qubit QPU ibmq_quito, while the calculations performed in Fig. 2 were generated using the Qiskit QASM simulator with noise model and coupling map extracted from the IBM machine ibmq_quito. For both the simulator and real device calculations, complete measurement calibration to mitigate readout errors...
was applied on all individual measurements, as implemented in Qiskit [28].

To benchmark the performance of our parametric GPR method, we compare the results obtained with the procedure described above in Sec. VI C with the exact solution, as well as with: (i) the standard COBYLA optimizer [40] and (ii) the sequential 1-dimensional optimizer, recently proposed in Ref. [53]. Both sequential optimizers underwent 20 optimization iterations. Note that each iteration in COBYLA requires a single energy evaluation. Instead, each iteration for the 1-dimensional optimizer requires to perform 2 1-dimensional fits (one for each variational parameter), which costs 8 energy evaluations for our variational ansatz. Therefore, in total, our calculations with COBYLA used 20 energy evaluations for each EH, while our calculations with the 1-dimensional optimizer used 160 energy evaluations. As explained above, our parametric GPR calculations were performed using 90 energy evaluations for learning the variational landscape of each EH, while the variational optimization was performed subsequently, on a classical computer.

Both for the simulations performed on the real quantum device (Fig. 1) and for the simulations performed with the simulator (Fig. 2), the results obtained using the parametric GPR method are considerably closer to the exact solution (Fig. 2), the results obtained using the parametric GPR method are considerably closer to the exact solution, as expected, since as at $U = 0$ the Hartree-Fock solution, which is imbued exactly within our variational landscape, is exact.

Additional numerical experiments are provided in the Appendix, confirming that our error-mitigation method is effective also for different values of the parameter $\lambda$, as well as using different quantum devices.

VI. APPLICATION AS A RISB IMPURITY SOLVER

A very promising route for achieving near-term quantum advantage is to use QE theoretical frameworks [7,8], treating with a VQE only the most active fraction of degrees of freedom, while treating the other modes at the mean-field level [9,11], on classical devices.

Well-known examples of QE schemes are dynamical mean field theory (DMFT) [11,12], the multi-orbital Gutzwiller approximation (GA) [13,46], the recently-developed ghost Gutzwiller approximation (g-GA) [47], the rotationally invariant slave boson theory (RISB) [29,31] and density matrix embedding theory [48].

Here we focus on a QE method called rotationally invariant slave boson theory (RISB) [29,31], which is equivalent to the Gutzwiller approximation [47,48] at the mean-field level. As shown in Ref. [46], the algorithmic structure of this framework consists in solving self-consistency equations involving the ground-state calculation of relatively-small auxiliary systems called “embedding Hamiltonians” (EH) —which is a task where a quantum computer with tens of qubits could potentially surpass the classical algorithms currently available [11,12]. In particular, this opens the possibility of implementing quantum-classical hybrid approaches, exploiting VQEs for solving the EH [11] and classical computers for solving the quantum-embedding self-consistency equations.

A. GA theory of the single-band Hubbard model

Here we utilize the general formulation of the GA derived in Refs. [31,40]. For completeness, here we summarize the GA equations for the special case of the periodic single-band Hubbard model, that we will consider in Sec. VI C for our benchmark calculations:

$$H = \sum_{i,j=1}^{N} t_{ij} c_{i\sigma}^\dagger c_{j\sigma} + \sum_{i} \left( \frac{U}{2}(\hat{n}_i - 1)^2 - \mu \hat{n}_i \right) ,$$

(41)

where $\hat{n}_i = c_{i\sigma}^\dagger c_{i\sigma}$, $\hat{n}_i = \hat{n}_{\uparrow} + \hat{n}_{\downarrow}$, $\mu$ is the chemical potential and $N$ is the total number of lattice sites. Furthermore, we will focus exclusively on the “normal phase,” i.e., we will assume that no symmetry is spontaneously broken. The GA solution of Eq. (41) is obtained by calculating the saddle point of the following Lagrange function [31,46]:

$$L_n(R, \lambda, D, \lambda^*, \mu) = 2 \int d\rho(\epsilon)(R^2 \epsilon + \lambda - \mu) + 2n\epsilon n + \langle \Phi(D, \lambda^*, U) | \hat{H}_{emb}(D, \lambda^*, U) | \Phi(D, \lambda^*, U) \rangle + 2(\lambda + \lambda^*)n + 4D \mathbb{R}[n(1-n)]^2,$$

(42)

where:

$$\rho(\epsilon) = N^{-1} \sum_{k=1}^{2n} \delta(\epsilon - \epsilon_k),$$

(43)

$\epsilon_k$ are the eigenvalues of $\hat{t}$, $2n$ is the total number of electrons per site, $\hat{H}_{emb}$ is the EH defined in Eq. (45) and $| \Phi(D, \lambda^*, U) \rangle$ is the corresponding ground state, within the subspace generated by states with 2 electrons (i.e., half of the maximum possible occupation within the EH space).

Below we use the half-bandwidth of $\rho(\epsilon)$ as energy unit and assume a semicircular density of states (corresponding to a Bethe lattice with infinite coordination number):

$$\rho(\epsilon) = \frac{2}{\pi} \sqrt{1 - \epsilon^2}.$$

(44)

The physical observables can be expressed in terms of the saddle-point GA variational parameters. In particular, at the saddle point, $L_n$ reduces to the system’s ground-state total energy per site, while the parameters $R, \lambda$ determine the GA approximation to the self-energy:

$$\Sigma(\omega) = -\omega \frac{1 - Z}{Z} + \frac{\lambda}{Z},$$

(45)

and $Z = R^2$ is the so-called quasi-particle weight.
Figure 3. Self-consistent GA benchmark calculations of the for the single-band Hubbard model: quasi-particle weight (left panels) and the double occupancy (right panels). The solution obtained using exact diagonalization as impurity solver is compared with the Qiskit QASM simulator with noise model and coupling map extracted from the IBM machine ibmq_bogota, using 2 different methods: the 1D optimizer (bottom panels) and our parametric-GPR optimizer (top panels).

B. Algorithmic structure

The saddle-point equations for Eq. \ref{eq:42} are the following:

\begin{align*}
\int d\epsilon \rho(\epsilon) f(\epsilon + \lambda - \mu) &= n \quad \text{(46)} \\
\mathcal{D}[n(1-n)]^{\frac{1}{2}} &= \mathcal{R} \int d\epsilon \rho(\epsilon) \epsilon f(\epsilon + \lambda - \mu) \quad \text{(47)} \\
2\mathcal{D}\mathcal{R}[n(1-n)]^{\frac{1}{2}} \left( \frac{1}{2} - n \right) + \lambda + \lambda^c &= 0 \quad \text{(48)} \\
\mathcal{F}^{(1)} &= f_1(\mathcal{D}, \lambda^c, U) - \mathcal{R}[n(1-n)]^{\frac{1}{2}} = 0 \quad \text{(49)} \\
\mathcal{F}^{(2)} &= f_2(\mathcal{D}, \lambda^c, U) - n = 0 \quad \text{(50)}
\end{align*}

where:

\begin{align*}
f_1 &= \langle \Phi(\mathcal{D}, \lambda^c, U) | \frac{1}{2} \sum \sigma c_{\sigma}^d c_{\sigma} | \Phi(\mathcal{D}, \lambda^c, U) \rangle \quad \text{(51)} \\
f_2 &= \langle \Phi(\mathcal{D}, \lambda^c, U) | \frac{1}{2} \sum \sigma d_{\sigma}^d d_{\sigma} | \Phi(\mathcal{D}, \lambda^c, U) \rangle \quad \text{(52)}
\end{align*}

Using Eqs. \ref{eq:46}--\ref{eq:48} one can readily compute \( \mathcal{D} \) and \( \lambda^c \) as a function of \( \mathcal{R} \) and \( \lambda \). Therefore, the saddle-point equations can be tackled by solving Eqs. \ref{eq:49}--\ref{eq:50}, which is ultimately a root problem for \( \mathcal{F} = (\mathcal{F}^{(1)}, \mathcal{F}^{(2)}) \), as a function of \( X = (\mathcal{R}, \lambda) \).

In summary, the GA algorithmic structure consists in calculating recursively the ground state of the EH and evaluating the corresponding components of the so-called single-particle density matrix, i.e., Eqs. \ref{eq:51} and \ref{eq:52}. 
Below we present numerical simulations of the Hubbard model performed using QPU simulators for this task.

C. Numerical tests of VQE as EH solver

In this subsection we benchmark our method as an impurity solver, for performing GA calculations of the single-orbital Hubbard model, within the algorithmic structure outlined above (Sec. VI A). For this purpose, due to limited QPU resources, we use only the Qiskit QASM simulator, with noise model and coupling map extracted from the IBM machine ibmq_bogota. Based on the QPU calculations performed in Sec. VI D we expect that using real machines would yield similar results.

In Fig. 3 we calculate the quasi-particle weight $Z$ and the double occupancy $\langle n_{i\uparrow} n_{i\downarrow} \rangle$, comparing our parametric GPR with the sequential 1-dimensional optimizer. The COBYLA optimizer could not be used, as the numerical stability of this approach proved to be insufficient for achieving numerical convergence of the GA self-consistency conditions.

Our results confirm that the parametric GPR method mitigates substantially the noise. Consistent with the benchmark calculations of Figs. 1 and 2, the improvement in accuracy is systematic, and it is particularly evident for small values of $U$. As pointed out in Sec. VI D this is thanks to the fact that the $U = 0$ the Hartree-Fock solution is imbeded exactly within the GPR framework, and our prior information concerning the parametric form of the variational landscape correlates with each other the GPR evaluations at different variational parameters.

Alternative ways of exploiting our error-mitigation method within QE frameworks, such as generating training data within the machine-learning approach recently proposed in Ref. [49], will be tested in future work.

VII. CONCLUSIONS

We derived a method (rooted on probabilistic machine learning) for mitigating the error in variational quantum eigensolvers, which we call “parametric GPR”. Rather than mitigating the error of the VQE measurements for each individual variational state, as in the methods mentioned in the introduction, our approach is based on the idea of exploiting the underlying correlations between the VQEs outputs for different variational parameters. We performed benchmark calculations both on real IBM devices and using noise-model simulators, showing that our method mitigates substantially the noise. In particular, we performed GA calculations of the single-band Hubbard model, using a variational quantum eigensolver as “impurity solver”. We found that our methodology allows us to solve the GA self-consistency conditions more effectively than other approaches, providing more reliable results. Our method will be applicable to all QE framework with an algorithmic structure similar to the GA, such as the recently-developed “ghost Gutzwiller approximation” and density matrix embedding theory. Our approach may contribute to speed up the advent of quantum advantage for simulations of large molecules and periodic materials, capitalizing more effectively on the potential of QPU already in the near future. In particular, we believe that our method will prove useful for unlocking access to cluster calculations with very large impurities, allowing us to describe dynamical-correlation effects beyond the reach of classical impurity solvers. Particularly relevant here is that, interestingly, there has been significant recent efforts into developing quantum GPR algorithms. This further supports that probabilistic machine learning and QPUs together in the near-term may allow quantum computations to enter previously uncharted territory, paving the way to simulations of large molecules and materials with near-term quantum devices.

ACKNOWLEDGEMENTS

We thank Gunnar Schmitz and Gabriel Kotliar for useful discussions. We gratefully acknowledge funding from VILLUM FONDEN through the Villum Experiment project 00028019 and the Centre of Excellence for Dirac Materials (Grant. No. 11744). We thank support from the Novo Nordisk Foundation through the Exploratory Interdisciplinary Synergy Programme project NNF19OC0057790. Part of the quantum computing method development and implementation by Y. Yao, and part of the calculations and analyses by Y. Yao and T. Jiang are supported by the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Co-design Center for Quantum Advantage (C2QA) under contract number DE-SC0012704. We acknowledge use of the IBM Quantum Experience, through the IBM Quantum Researchers Program. The views expressed are those of the authors, and do not reflect the official policy or position of IBM or the IBM Quantum team.

[1] John Preskill, “Quantum computing in the NISQ era and beyond,” Quantum 2, 79 (2018).
[2] Alán Aspuru-Guzik, Anthony D Dutoi, Peter J Love, and Martin Head-Gordon, “Simulated quantum computation of molecular energies,” Science 309, 1704–1707 (2005).
[3] James D. Whitfield, Jacob Biamonte, and Alán Aspuru-Guzik, “Simulation of electronic structure hamiltonians using quantum computers,” Molecular Physics 109, 735–750 (2011).
[4] Dave Wecker, Bela Bauer, Bryan K. Clark, Matthew B. Hastings, and Matthias Troyer, “Gate-count estimates for performing quantum chemistry on small quantum computers,” Phys. Rev. A 90, 022305 (2014).

[5] Ryan Babbush, Jarrod McClean, Dave Wecker, Alán Aspuru-Guzik, and Nathan Wiebe, “Chemical basis of trotter-suzuki errors in quantum chemistry simulation,” Phys. Rev. A 91, 022311 (2015).

[6] Sam McArdle, Suguru Endo, Alán Aspuru-Guzik, Simon C. Benjamin, and Xiao Yuan, “Quantum computational chemistry,” Rev. Mod. Phys. 92, 015003 (2020).

[7] Paul R. C. Kent and Gabriel Kotliar, “Toward a predictive theory of correlated materials,” Science 361, 348–354 (2018).

[8] Sun Qiming and Chan Garnet Kin-Lic, “Quantum embedding theories,” Acc. Chem. Res. 49, 2705 (2016).

[9] He Ma, Marco Govoni, and Giulia Galli, “Quantum simulations of materials on near-term quantum computers,” npj Computational Materials 6, 85 (2020).

[10] Bela Bauer, Dave Wecker, Andrew J. Millis, Matthew B. Hastings, and Matthias Troyer, “Hybrid quantum-classical approach to correlated materials,” Phys. Rev. X 6, 031045 (2016).

[11] Yongxin Yao, Feng Zhang, Cai-Zhuang Wang, Kai-Ming Ho, and Peter P. Orth, “Gutzwiller hybrid quantum-classical computational approach for correlated materials,” Phys. Rev. Research 3, 013184 (2021).

[12] Ben Jaderberg, Abhishek Agarwal, Karsten Leonhardt, Martin Kiffner, and Dieter Jaksch, “Minimum hardware requirements for hybrid quantum-classical dmft,” Quantum Science and Technology 5, 034015 (2020).

[13] Ying Li and Simon C. Benjamin, “Efficient variational quantum simulator incorporating active error minimization,” Phys. Rev. X 7, 021050 (2017).

[14] Abhinav Kandala, Kristan Temme, Antonio D. Córcoles, Antonio Mezzacapo, Jerry M. Chow, and Jay M. Gambetta, “Error mitigation extends the computational reach of a noisy quantum processor,” Nature 567, 491–495 (2019).

[15] Suguru Endo, Simon C. Benjamin, and Ying Li, “Practical quantum error mitigation for near-future applications,” Phys. Rev. X 8, 031027 (2018).

[16] Jarrod McClean, Jonathan Romero, Ryan Babbush, and Alán Aspuru-Guzik, “The theory of variational hybrid quantum-classical algorithms,” New Journal of Physics 18, 023023 (2016).

[17] Peter JJ O’Malley, Ryan Babbush, Ian D Kivlichan, Jonathan Romero, Jarrod McClean, Rami Barends, Julian Kelly, Pedram Roushan, Andrew Tranter, Nan Ding, et al., “Scalable quantum simulation of molecular energies,” Physical Review X 6, 031007 (2016).

[18] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M. Chow, and Jay M Gambetta, “Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets,” Nature 549, 242–246 (2017).

[19] Jonathan Romero, Ryan Babbush, Jarrod McClean, Cornelius Hempel, Peter J Love, and Alán Aspuru-Guzik, “Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz,” Quantum Science and Technology 4, 014008 (2018).

[20] Kristan Temme, Sergey Bravyi, and Jay M. Gambetta, “Error mitigation for short-depth quantum circuits,” Phys. Rev. Lett. 119, 180509 (2017).

[21] Joel J Wallman and Joseph Emerson, “Noise tailoring for scalable quantum computation via randomized compiling,” Physical Review A 94, 052325 (2016).

[22] Ryan LaRose, Andrea Mari, Sarah Kaiser, Peter J Karalekas, Andre A Alves, Piotr Czarnik, Mohamed El Mandouh, Max H Gordon, Yousef Hindiy, Aaron Robertson, et al., “Mitig: A software package for error mitigation on noisy quantum computers,” arXiv preprint arXiv:2009.04147 (2020).

[23] Youngseok Kim, Christopher J Wood, Theodore J Yoder, Seth T Merkel, Jay M Gambetta, Kristan Temme, and Abhinav Kandala, “Scalable error mitigation for noisy quantum circuits produces competitive expectation values,” arXiv preprint arXiv:2108.09197 (2021).

[24] Shuaining Zhang, Yao Lu, Kuan Zhang, Wentao Chen, Ying Li, Jing-Ning Zhang, and Kihwan Kim, “Error-mitigated quantum gates exceeding physical fidelities in a trapped-ion system,” Nature communications 11, 1–8 (2020).

[25] Xavi Bonet-Monroig, Raúl Sagastibaral, M Singh, and TE O’Brien, “Low-cost error mitigation by symmetry verification,” Physical Review A 98, 062339 (2018).

[26] Sam McArdle, Xiao Yuan, and Simon Benjamin, “Error-mitigated digital quantum simulation,” Physical review letters 122, 180501 (2019).

[27] Jarrod R McClean, Zhang Jiang, Nicholas C Rubin, Ryan Babbush, and Hartmut Neven, “Decoding quantum errors with subspace expansions,” Nature communications 11, 1–9 (2020).

[28] Héctor Abraham, Adouffe, Rochisha Agarwal, Ismail Yunus Akhawayn, Gadi Aleksandrowicz, Thomas Alexander, Matthew Amy, Eli Arbel, Arijit02, Abhik Banerjee, Aman Bansal, Panagiotis Barkoutsos, Ashish Barnawal, George Barron, George S. Barron, Luciano Bello, Yael Ben-Haim, Daniel Bevenius, Arjun Bhobe, Lev S. Bishoff, Chris V. Bosman, Soeren Boes, Brandon, Sergey Bravyi, Bryce-Fuller, David Bucher, Artemiy Burov, Fran Cabrera, Padraic Calpin, Lauren Capelluto, Jorge Carballo, Gines Carrales, Adrian Chen, Chun-Fu Chen, Edward Chen, Jielun (Chris) Chen, Richard Chen, Jerry M. Chow, Spencer Churchill, Christian Claus, Christian Claus, Romilly Cocking, Filipe Correa, Abigail J. Cross, Andrew W. Cross, Simon Cross, Juan Cruz-Benito, Chris Culver, Antonio D. Córcoles-Gonzales, Sean Dague, Tareq El Dandachi, Marcus Daniels, Matthieu Darbailly, DavideFrr, Abdón Rodríguez Davila, Anton Dekusar, Delton Ding, Jun Doi, Eric Drechsler, Drew, Eugene Dumitrescu, Karel Dumon, Ivan Duran, Karim El-Safty, Eric Eastman, Pieter Eendebak, Daniel Egger, Mark Everitt, Paco Martín Fernández, Axel Hernández Ferrera, Romain Fouilland, FranckChevallier, Albert Frisch, Andreas Fuhrer, Bryce Fuller, MELVIN GEORGE, Julien Gacon, Borja Godoy Gago, Claudio Gambella, Jay M. Gambetta, Adhisha Gammapilana, Luis Garcia, Tanya Garg, Shelly Garion, Austin Gilliam, Aditya Giridharan, Juan Gomez-Mosquera, Gonzalo, Salvador de la Puente González, Jesse Gorzinski, Ian Gould, Donny Greenberg, Dmitry Grinko, Wen Guan, John A. Gunnels, Mikael Haglund, Isabel Haide, Ikko Hamamura, Omar Costa Hamido, Frank Harksins, Vojtech Havlivcek, Joe Hellmers, Lukasz Herok, Stefan Hillmich, Hiroshi Horii, Connor Howing-
ton, Shaohan Hu, Wei Hu, Junye Huang, Rolf Huisman, Haruki Imai, Takashi Imamura, Kazuaki Ishizaki, Raman Iten, Toshinari Itoko, James Seaward, Ali Javadi, Ali Javadi-Abhari, Wahaj Javed, Jessica, Madhav Jivrajani, Kiran Johns, Scott Johnstone, Jonathan Shomaker, Vismai K, Tal Kachmann, Akshay Kale, Naoki Kanazawa, Kang-Bae, Anton Karazeev, Paul Kassebaum, Josh Kelso, Spencer King, Knabberjoe, Yuri Kobyayashi, Arseny Kovyrshin, Rajiv Krishnakumar, Vivek Krishnan, Kevin Krsulich, Prasad Kumkar, Gawan Kus, Ryan LaRose, Enrique Laca, Raphael Lambert, John Lapeyre, Joe Latone, Scott Lawrence, Christina Lee, Gushu Li, Dennis Liu, Peng Liu, Yunho Maeng, Kahan Majmudar, Aleksei Malyshev, Joshua Manela, Jakub Marecek, Manoel Marques, Dmitri Maslov, Dolph Mathews, Atsushi Matsuo, Douglas T. McClure, Cameron McGarry, David McKay, Dan McPherson, Srujan Meesala, Thomas Metcalfe, Martin Mevissen, Andrew Meyer, Antionio Mezzacapo, Rohit Midha, Zlatko Minev, Abby Mitchell, Nikolaj Moll, Jhon Montenez, Gabriel Montiero, Michael Duane Mooreing, Renier Morales, Niall Moran, Mario Motta, MrF, Prakash Murali, Jan Miéggenburg, David Nadlinger, Ken Nakashima, Giovanni Nannincini, Paul Nation, Edwin Navarro, Yehuda Nevhe, Scott Wyman Neage, Patrick Neuweiler, Johan Nicander, Pradeep Niroula, Hassan Nolten, NuoWenLei, Lee James O’Riordan, Ohwotabo Ogumbayo, Pauline Olitrault, Raul Otoleao, Steven Oud, Dan Padilla, Hanhee Paik, Soham Pal, Yuchen Pang, Vincent R. Pascuzzi, Simone Perriello, Anna Phan, Francesco Piro, Marco Pistoia, Christophe Piveto, Pierre Pocrea, Alejandro Pozas-Kerstens, Milos Prokop, Viktor Prutyanov, Daniel Puzzuoli, Jesús Pérez, Quintii, Rafay Iqbal Rahman, Arun Raja, Nipun Ramagiri, Anirudh Rao, Rudy Raymond, Rafael Martín-Cuevas Redondo, Max Reuter, Julia Rice, Matt Riedemann, Marcello La Rocca, Diego M. Rodriguez, RohitKur, Max Rossmannek, Mingyi Ruyi, Tharrmashastha SAPV, SamFerracin, Martin Sandberg, Hirmay Sandesara, Rtvik Sapna, Hayk Sarayyan, Aniruddha Sarkar, Inad Sathaye, Bruno Schmitt, Chris Schnabel, Zachary Schoenfeld, Travis L. Scholten, Enrico Schoute, Joel Sethna, Israel Schwaab, Christian Schollwöck, Kanav Setia, Hassan Sethy, Ivan Setyan, Andrei Silva, Andrea Simonetto, Nick Singstock, Yuval Sisim, Yogeeta Sitdikov, Irosmis, Misha Slinkin, Sergey Slinkin,_xlabel, ylabel, year}

[39] Jonathan Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Jefrey, E. Lucero, A. Megrant, J. Y. Mutus, M. Neeley, C. Neill, C. Quintana, D. Sank, A. Vainsencher, J. Wettiger, T. C. White, P. V. Coveney, P. J. Love, H. Neven, A. Aspuru-Guzik, and J. M. Martinis, “Scalable quantum simulation of molecular energies,” Phys. Rev. X 6, 031007 (2016).

[40] Michael JD Powell, “A direct search optimization method that models the objective and constraint functions by linear interpolation,” in Advances in optimization and numerical analysis (Springer, 1994) pp. 51–67.
[41] A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, “Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions,” Rev. Mod. Phys. 68, 13 (1996).

[42] V. Anisimov and Y. Izumov, Electronic Structure of Strongly Correlated Materials (Springer, 2010).

[43] M. C. Gutzwiller, “Correlation of Electrons in a Narrow Band,” Phys. Rev. 137, A1726 (1965).

[44] J. Büinemann, W. Weber, and F. Gebhard, “Multiband Gutzwiller wave functions for general on-site interactions,” Phys. Rev. B 57, 6896 (1998).

[45] X.-Y. Deng, L. Wang, X. Dai, and Z. Fang, “Local density approximation combined with Gutzwiller method for correlated electron systems: Formalism and applications,” Phys. Rev. B 79, 075114 (2009).

[46] Nicola Lanatà, Yong Xin Yao, Cai-Zhuang Wang, Kai-Ming Ho, and Gabriel Kotliar, “Phase diagram and electronic structure of praseodymium and plutonium,” Phys. Rev. X 5, 011008 (2015).

[47] Nicola Lanatà, Tsung-Han Lee, Yong-Xin Yao, and Vladimir Dobrosavljević, “Emergent bloch excitations in mott matter,” Phys. Rev. B 96, 195126 (2017).

[48] G. Knizia and G. K.-L. Chan, “Density matrix embedding: A simple alternative to dynamical mean-field theory,” Phys. Rev. Lett. 109, 186404 (2012).

[49] John Rogers, Tsung-Han Lee, Sahar Pakdel, Wenhu Xu, Vladimir Dobrosavljević, Yong-Xin Yao, Ove Christiansen, and Nicola Lanatà, “Bypassing the computational bottleneck of quantum-embedding theories for strong electron correlations with machine learning,” Phys. Rev. Research 3, 013101 (2021).

[50] Marius S. Frank, Tsung-Han Lee, Gargee Bhattacharyya, Pak Ki Henry Tsang, Victor L. Quito, Vladimir Dobrosavljević, Ove Christiansen, and Nicola Lanatà, “Quantum embedding description of the anderson lattice model with the ghost gutzwiller approximation,” Phys. Rev. B 104, L081103 (2021).

[51] Daniele Guerci, Massimo Capone, and Michele Fabrizio, “Exciton mott transition revisited,” Phys. Rev. Materials 3, 054605 (2019).

[52] T. Maier, M. Jarrell, T. Pruschke, and M. H. Hettler, “Quantum cluster theories,” Rev. Mod. Phys. 77, 1027 (2005).

[53] A. I. Lichtenstein and M. I. Katsnelson, “Antiferromagnetism and d-wave superconductivity in cuprates: A cluster dynamical mean-field theory,” Phys. Rev. B 62, R9283 (2000).

[54] M. Potthoff, M. Aichhorn, and C. Dahnken, “Variational cluster approach to correlated electron systems in low dimensions,” Phys. Rev. Lett. 91, 206402 (2003).

[55] Zhikuan Zhao, Jack K. Fitzsimons, and Joseph F. Fitzsimons, “Quantum-assisted gaussian process regression,” Phys. Rev. A 99, 052331 (2019).

[56] Zhikuan Zhao, Jack K. Fitzsimons, Michael A. Osborne, Stephen J. Roberts, and Joseph F. Fitzsimons, “Quantum algorithms for training gaussian processes,” Phys. Rev. A 100, 012304 (2019).

[57] M Otten, (2020), arXiv:cond-mat/2004.11280.
Appendix: Additional benchmark calculations of the EH for different values of $\lambda^c$

Figure 4. Benchmark calculations of the EH performed on the IBM 5 qubit QPU *ibmq_quito*. Total energy $\mathcal{E}$, impurity double-occupancy $\langle \hat{n}_c^\uparrow \hat{n}_c^\downarrow \rangle$ and components of the single-particle density matrix $\sum_\sigma \langle c_\sigma^\dagger d_\sigma \rangle$, $\sum_\sigma \langle d_\sigma^\dagger d_\sigma \rangle$. The exact values are compared with the VQE outputs produced using 3 different methods: (i) the COBYLA optimizer, (ii) the 1D optimizer and (iii) our parametric-GPR optimizer.
Figure 5. Benchmark calculations of the EH performed with the Qiskit QASM simulator with noise model and coupling map extracted from the IBM machine `ibmq_quito`. Total energy $E$, impurity double-occupancy $\langle \hat{n}_{c\uparrow}\hat{n}_{c\downarrow} \rangle$ and components of the single-particle density matrix $\sum_\sigma \langle c^\dagger_\sigma d_\sigma \rangle$, $\sum_\sigma \langle d^\dagger_\sigma d_\sigma \rangle$. The exact values are compared with the VQE outputs produced using 3 different methods: (i) the COBYLA optimizer, (ii) the 1D optimizer and (iii) our parametric-GPR optimizer.
Figure 6. Benchmark calculations of the EH performed on the IBM 5 qubit QPU *ibmq_bogota*. Total energy $E$, impurity double-occupancy $\langle \hat{n}_{c\uparrow}\hat{n}_{c\downarrow} \rangle$ and components of the single-particle density matrix $\sum_{\sigma} \langle c_{\sigma}\dagger c_{\sigma} d_{\sigma} \rangle$, $\sum_{\sigma} \langle d_{\sigma}\dagger d_{\sigma} \rangle$. The exact values are compared with the VQE outputs produced using 3 different methods: (i) the COBYLA optimizer, (ii) the 1D optimizer and (iii) our parametric-GPR optimizer.
Figure 7. Benchmark calculations of the EH performed with the Qiskit QASM simulator with noise model and coupling map extracted from the IBM machine *ibmq_bogota*. Total energy $E$, impurity double-occupancy $\langle \hat{n}_{c\uparrow} \hat{n}_{c\downarrow} \rangle$ and components of the single-particle density matrix $\sum_{\sigma} \langle c_{\sigma}^\dagger c_{\sigma} \rangle$, $\sum_{\sigma} \langle d_{\sigma}^\dagger d_{\sigma} \rangle$. The exact values are compared with the VQE outputs produced using 3 different methods: (i) the COBYLA optimizer, (ii) the 1D optimizer and (iii) our parametric-GPR optimizer.
Figure 8. Benchmark calculations of the EH performed on the IBM 5 qubit QPU ibmq_lima. Total energy $\mathcal{E}$, impurity double-occupancy $\langle \hat{n}_{c_{\uparrow}}\hat{n}_{c_{\downarrow}} \rangle$ and components of the single-particle density matrix $\sum_{\sigma} \langle c_{\sigma}^\dagger d_{\sigma} \rangle$, $\sum_{\sigma} \langle d_{\sigma}^\dagger d_{\sigma} \rangle$. The exact values are compared with the VQE outputs produced using 3 different methods: (i) the COBYLA optimizer, (ii) the 1D optimizer and (iii) our parametric-GPR optimizer.
Figure 9. Benchmark calculations of the EH performed with the Qiskit QASM simulator with noise model and coupling map extracted from the IBM machine *ibmq_lima*. Total energy $\mathcal{E}$, impurity double-occupancy $\langle \hat{n}_{c\uparrow}\hat{n}_{c\downarrow}\rangle$ and components of the single-particle density matrix $\sum_{\sigma} \langle \hat{c}^{\dagger}_{\sigma}\hat{d}_{\sigma}\rangle$, $\sum_{\sigma} \langle \hat{d}^{\dagger}_{\sigma}\hat{d}_{\sigma}\rangle$. The exact values are compared with the VQE outputs produced using 3 different methods: (i) the COBYLA optimizer, (ii) the 1D optimizer and (iii) our parametric-GPR optimizer.
Figure 10. Benchmark calculations of the EH performed on the IBM 5 qubit QPU `ibmq_belem`. Total energy $\mathcal{E}$, impurity double-occupancy $\langle \hat{n}_{c\uparrow} \hat{n}_{c\downarrow} \rangle$ and components of the single-particle density matrix $\sum_\sigma \langle c_{\sigma}^\dagger d_{\sigma} \rangle$, $\sum_\sigma \langle d_{\sigma}^\dagger d_{\sigma} \rangle$. The exact values are compared with the VQE outputs produced using 3 different methods: (i) the COBYLA optimizer, (ii) the 1D optimizer and (iii) our parametric-GPR optimizer.
Figure 11. Benchmark calculations of the EH performed with the Qiskit QASM simulator with noise model and coupling map extracted from the IBM machine *ibmq_belem*. Total energy $E$, impurity double-occupancy $\langle \hat{n}_c \hat{n}_c \rangle$ and components of the single-particle density matrix $\sum_\sigma \langle c^\dagger_\sigma d_\sigma \rangle, \sum_\sigma \langle d^\dagger_\sigma d_\sigma \rangle$. The exact values are compared with the VQE outputs produced using 3 different methods: (i) the COBYLA optimizer, (ii) the 1D optimizer and (iii) our parametric-GPR optimizer.