scientific reports

OPEN

Meta-optimization of resources on quantum computers

Ijaz Ahamed Mohammad[1], Matej Pivoluska [1] & Martin Plesch [1,2,3,4✉]

The current state of quantum computing is commonly described as the Noisy Intermediate-Scale Quantum era. Available computers contain a few dozens of qubits and can perform a few dozens of operations before the inevitable noise erases all information encoded in the calculation. Even if the technology advances fast within the next years, any use of quantum computers will be limited to short and simple tasks, serving as subroutines of more complex classical procedures. Even for these applications the resource efficiency, measured in the number of quantum computer runs, will be a key parameter. Here we suggest a general meta-optimization procedure for hybrid quantum-classical algorithms that allows finding the optimal approach with limited quantum resources. This method optimizes the usage of resources of an existing method by testing its capabilities and setting the optimal resource utilization. We demonstrate this procedure on a specific example of variational quantum algorithm used to find the ground state energy of a hydrogen molecule.

Quantum computers, as a theoretical concept, have been suggested in the 1980's independently by Paul Benioff[1] and Yuri Manin[2]. Later they have been popularized by Richard Feynman in his seminal work on simulating quantum physics with a quantum mechanical computer[3]. In the following 30 years the potential of quantum computers to outperform classical computers in certain tasks was thoroughly studied[4] and many important quantum algorithms have been found, among them an algorithm to factorize large numbers in polynomial time[5] and an algorithm for a fast search in unstructured databases[6]. In recent years experimental quantum computing has achieved tremendous advances and thus the design of quantum algorithms has shifted from purely theoretical research towards more practical questions. One of the main new areas of research is the practical utilization of contemporary quantum processors, which are encumbered by noise that is reducing their reliability. These are called Noisy Intermediate-Scale Quantum (NISQ) computers[7] and in attempt to utilize them efficiently many so-called variational algorithms have been designed[8,9]. Variational algorithms use both quantum and classical computational resources and their main idea is to design a parameterized quantum circuit and a measurement method associated to a minimization problem at hand. The parameters of the quantum circuit minimizing the value of the target function are then searched for using classical function minimization subroutines. Algorithms of this type can be designed for optimization problems in a plethora of real world use cases, ranging from chemistry[10,11], through artificial intelligence[12,13] to financial market modelling[14].

Such a combination of classical and quantum approaches leads to many interesting challenges. The accuracy and reliability of the final output is inevitably limited by a combination of different factors. First, the classical optimization algorithm is not guaranteed to converge to the true minimum. This is because many optimization algorithms are stochastic by nature and they use randomness at least during initialization to choose the starting point. In many scenarios, randomness is required in each iteration. This randomness ensures that the optimization procedures can deal with a large family of different functions without getting stuck in a local minimum, but naturally also leads to stochasticity of the whole calculation, i.e. the algorithms are not guaranteed to succeed. Thus for obtaining a useful result with high probability, the procedure needs to be repeated several times.

Secondly, even in the noiseless quantum processor scenario the outcome of any useful quantum computation is stochastic. Typically, the outcome is a non-computational basis state (otherwise the computation would be classically efficiently simulable) and is characterized by frequencies of outcomes for different measurement settings. A single run of a quantum computer only provides a single snapshot of the state for one measurement setting. Any optimization procedure therefore inevitably comes with a trade-off between a more precise measurement on a single position in the parameter-space and less precise measurements of many positions. It is non-trivial to decide which of these two strategies leads to better results.

[1]Institute of Physics, Slovak Academy of Sciences, Dúbravská cesta 9, 841 04 Bratislava, Slovak Republic. [2]Faculty of Natural sciences and Informatics, Constantine the Philosopher University, Tr. A. Hlinku 1, 949 01 Nitra, Slovak Republic. [3]Institute of Physics of Materials, Czech Academy of Sciences, Žižkova 513/22, 616 00 Brno, Czech Republic. [4]Faculty of Natural Sciences, Matej Bel University, Tajovského 40, 974 01 Banská Bystrica, Slovakia. ✉email: martin.plesch@savba.sk

1

Last, but not least, one cannot rely on any kind of error correcting procedures for NISQ computers. The classical procedure itself will have to account for the fact that any result of the quantum computer is influenced, on top of the statistical fluctuations due to the intrinsic quantum randomness, also by experimental errors. All these three complications suggest that the classical procedure will have to go significantly further than just utilizing known optimization methods. For any near-future hybrid algorithms the main limiting factor will be the quantum part. Both from the time (queues) and financial (pay-per-shot) perspective, the efficiency of the algorithm will be measured by its ability to produce good results with as little utilization of quantum computers as possible. This generic statement can be simplified into a measure relating the reliability and precision of the result (the probability to get a result and how good the result is) to the number of utilizations of a quantum computer with a single measurement outcome (a shot). Simply speaking, the aim is to get the best possible result with a fixed budget counted in the number of shots (i.e., time or money).

Instead of suggesting a specific new technique on the classical level, we introduce a meta-optimization technique that can be utilized for a broad variety of classical-quantum scenarios. For any classical optimization procedure and connected quantum calculation the technique first samples the fluctuation of the results depending on the parameters, estimates this dependence and suggests the optimal setting. These will be expressed in the parameters of the classical and quantum part of the procedure and the optimal number of repetitions. We demonstrate this approach on a specific example on a Variational Quantum Eigensolver (VQE)[15] determining the ground energy of Hydrogen molecule in combination with the Simultaneous Perturbation Stochastic Approximation (SPSA)[16] classical minimization procedure. It turns out that for reasonable parameters, namely the number of available quantum shots counting in a few millions and the expected accuracy of the energy in small multiples of chemical precision, the optimal number of repetitions of the whole procedure is less than a dozen and the probability of obtaining the result within the expected precision ranges from 10 % to almost 90 %. This shows that one needs to be very careful by choosing the exact strategy during the optimization process.

The paper is organized as follows. In the rest of this chapter we introduce the VQE and SPSA. In the Results section, we describe the procedure in general terms, apply it to the specific Hydrogen molecule scenario and present the results. In Methods we provide details on the implementation of the quantum part of the algorithm on IBM quantum machines.

Variational Quantum Eigensolver

In this paper we study variational quantum eigensolver, introduced by Peruzzo et al.[15], and subsequently thoroughly studied by many authors (see review papers[8,9,17] on the topic of VQE and the references therein). VQE is used to calculate the ground state energy (E_0) of any Hamiltonian (H). It is based on quantum physics' variational principle, which asserts that any Hamiltonian's expectation with regard to any state ($|\psi\rangle$) is higher than the ground state energy:

$$\langle\psi|H|\psi\rangle \geq E_0. \tag{1}$$

Classical techniques become impractical to detect the ground energies of the Hamiltonian as the Hamiltonian grows in size. VQE is a hybrid method that circumvents this issue by combining quantum processors with classical optimizers. The working of a VQE algorithm is as follows:

- Preparation of a trial state ($|\psi_i\rangle$) from an initial state ($|\psi_0\rangle$) (typically a zero state) by introducing some parameters (θ_j), which can be updated to change the trial state in each iteration;
- Measurement of the state (possibly in several bases defined by the Hamiltonian), obtaining frequencies approximating probabilities. In this stage the trade-off is being made between a quick/cheap imprecise value and slower/more expensive precise value;
- Evaluation of the cost function (energy E_i) based on the probabilities approximated in the previous step;
- Using this evaluation as an input to the classical optimizer. This checks for convergence, and if the convergence criteria are not met, it modifies the parameters, resulting in a new iteration. This procedure is repeated until the convergence criteria are met.

The VQE algorithm's operation is depicted in Fig. 1.

Classical optimizers

Previous research in modifying the VQE algorithm to be more resource efficient approached the problem by tailoring the classical optimization part of the algorithm to economically distribute the shot budget among various required calculations. This is a very fruitful approach and many significant advances have been achieved, involving both gradient-based and non-gradient based algorithms. One example of such an algorithm is individual Coupled Adaptive Number of Shots (iCANS)[18]. The iCANS algorithm distributes the shot budget available for each iteration across all parameters in order to find partial derivatives in all directions. For each direction the amount of shots used to evaluate the gradient in each iteration i is inversely proportional to the square of the gradient in this direction in iteration $i − 1$. In such a way, low gradient directions in each iteration are evaluated with larger precision, which is necessary for successful convergence. Another approach is to distribute the shot budget for each iteration unevenly across different Hamiltonian terms. This has been explored in conjunction with several optimization algorithms, with the best performance observed with iCANS algorithm resulting in an algorithm called Random Operator Sampling for Adaptive Learning with Individual Number of shots (Rosalin)[19]. Last but not least, a more general discussion about evaluating only a randomly chosen subset of all
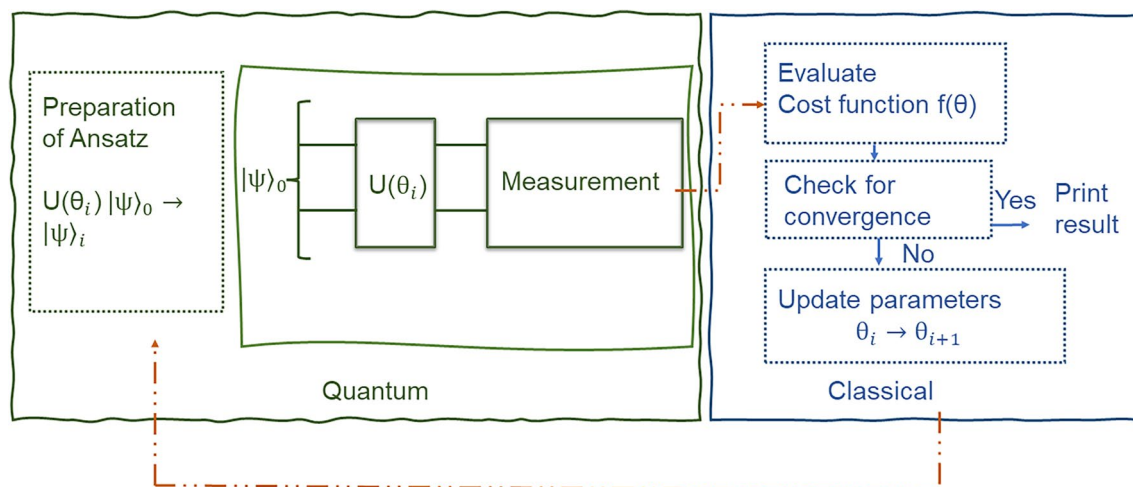
Figure 1. The VQE algorithm is divided into two parts. First, the ansatz ($|\psi_i\rangle$) is prepared from an initial state ($|\psi_0\rangle$) with the help of some parameterized circuit which can viewed as the action of a parameterized unitary operator ($U(\theta_i)$). The state is later measured and by using the counts of the outcomes of the measurement, the cost function is evaluated. If the convergence of the optimizer isn't achieved, this process is repeated until desired result with good convergence is obtained in the end.

possible Hamiltonian terms in each iteration, with conjunction with evaluating gradients only in a randomly chosen subset of directions (so-called stochastic gradient descent) can be found in[20].

In the commonly used stochastic gradient descent algorithms at each iteration, the partial derivative is evaluated by perturbing each parameter along and opposite to its direction requiring two function evaluations (alternatively in VQE one can use the parameter shift rule[21]). This can be computationally expensive for functions with many parameters, which suggests that we will need $2pt$ functions evaluations in total, where p is the length of the parameter vector θ and t is the total number of iterations used in optimization. In contrast, in Simultaneous Perturbation Stochastic Approximation (SPSA)[16], the gradient in each iteration is estimated by simultaneously perturbing all parameters θ randomly with only two function evaluations required for the whole gradient. Thereby resulting in $2t'$ function evaluations in total where t' is the number of iterations SPSA took to converge. The second advantage of SPSA is its stochasticity, which makes it noise-resistant. Because each parameter is already randomly perturbed, subsequent noise perturbations are less likely to disrupt the optimization process.

Costs of final estimation

While rather counter-intuitive, advanced optimization techniques are able to find the parameters of the optimum even if very few shots are used per each measurement. In other words, one can find the target even if almost blind, with one reservation – the resulting value of the cost function is very imprecise. In many cases this issue is not addressed at all and the final value is calculated by classical means[18,19,22]. This, however, is a simplification – classical recalculation of the energies becomes infeasible in instances where the quantum computer is really useful. Thus, if one wants to have a complete solution, it turns out that at the end of the calculation one needs to test the resulting state for the precise value of the cost function, investing a large number of shots – as we will show in the result section, this can be comparable to the whole optimization procedure budget. This further complicates the optimization – not only the number of repetitions of the procedure needs to be optimized, but also the distribution of the shots between the optimization procedure itself and the final estimation of the cost function value.

H_2 molecule with SPSA optimizer

We will apply the suggested method onto a specific example of Variational Quantum Eigensolver seeking the ground energy of an H_2 molecule. First, one needs to transform the physical Hamiltonian of the system in question expressed in creation and annihilation operators into a Hamiltonian suitable for quantum computers expressed in Pauli operators, called qubit Hamiltonian. Here we considered a qubit Hamiltonian of H_2 molecule with 2-qubits with the distance between the atoms set as 0.725 Å of the form

$$H = c_0(I \otimes I) + c_1(Z \otimes I) + c_1(I \otimes Z) + c_2(Z \otimes Z) + c_3(X \otimes X), \qquad (2)$$

where X, Z are the usual notations of the Pauli matrices and the coefficients c_0, c_1, c_2, c_3 are specified as -1.05016, 0.40421, 0.01135, and 0.18038 respectively[23]. Even for such a simple physical system and form of the Hamiltonian a significant cost is required to get the desired results from conventional use of the VQE algorithm. We will show by using SPSA method how our optimization can help to achieve better results.

Results

In the most general terms, we suggest a meta optimization procedure with following inputs:

- An optimization technique parameterized solely by the number of shots used per the whole run;
- Final result estimation method parameterized by its costs measured in the number of shots;
- Shot budget, i.e. the total number of shots available;
- Desired accuracy, i.e. how far away from the exact value of the cost function in optimum the result is considered as successful.

Regarding the last condition it is important to stress that the success of the optimization is evaluated in terms of the cost function and not the parameters. Finding a local minimum far away from the global one (e.g. an excited state), which has the value of cost function within the desired accuracy (e.g. if the accuracy is set as low as the first energy gap) is taken as a successful result.

The output of the meta optimization consists of

- Optimal number of repetitions of the optimization;
- Number of shots to be used per one optimization run;
- Number of shots to be used per final estimation;
- Probability of success.

Using these results, one has to take the steps as suggested, i.e., run the optimization method desired number of times, perform the final estimation of all results and select the best one based on this estimation (this might not be the best result according to the optimization method itself, due to a larger variance of energy calculations used during the run). The probability of success expresses the likelihood we get the result (at least) as good as we wanted. For large budget of shots, resulting into reliable final estimations, we will know how good the result is with high probability after obtaining it. This is not completely the case for small budgets, resulting into unreliable final estimation – here we might have a good result even if not estimated so, or overestimate the quality of a bad result.

Alternatively, one can re-formulate the procedure by exchanging the Success probability and Accuracy. If we state the expected probability to get a result, the meta-optimization procedure will suggest the parameters to obtain the best result (highest accuracy) with at least the given probability.

The cost of final estimation

While the input-output relation for optimization techniques are complex, the final estimation precision is in a simple mathematical relationship with the number of shots used. In VQAs, the objective function, which is typically an expectation of summations of different quantum operators with different weights, is calculated from the probability estimates from the quantum measurements. Such a quantum operator is of the form $F = \sum_i A_i$, with A_i denoting a collection of simultaneously-measurable quantum operators with $A_i = \sum_j c_j O_j$, where c_j and O_j denote coefficients and Pauli words, respectively. The cost of a final estimate (S) for a desired accuracy of ε is given by[24]

$$S := \frac{\left(\sum_i \sqrt{\mathrm{Var}(A_i)} \right)^2}{\varepsilon^2}. \tag{3}$$

The authors of the article[25] have shown that the variance can be computed as

$$\mathrm{Var}(A_i) = \sum_{jk} c_j c_k \mathrm{Cov}(O_j, O_k), \tag{4}$$

where $\mathrm{Cov}(\cdot)$ represents the covariance between the Pauli words. As discussed in the article[26], from the Cauchy-Schawrtz inequality the covariance between two Pauli words is upper bounded as

$$\mathrm{Cov}(O_j, O_k) \leq \left| \sqrt{\mathrm{Var}(O_j)\mathrm{Var}(O_k)} \right|. \tag{5}$$

As the magnitude of the covariance is upper bounded as shown in Eq. 5, one can consider the average from a random distribution of the covariance of cross terms to be zero. Using this and coupling it with the fact that the variance of a Pauli word is upper bounded by 1, the number of shots required to obtain the estimate of the objective function with accuracy ε is given by

$$S = \frac{\left(\sum_i \sqrt{\sum_j c_{ij}^2} \right)^2}{\varepsilon^2}. \tag{6}$$

Example of H_2 molecule

Measuring the eigenstate of our Hamiltonian's ground state energy allows us to examine the impact of the statistical disturbance brought on by measurements. Given that the Hamiltonian matrix in our example is small, we performed a classical calculation to identify the Hamiltonian's corresponding eigenvalues and eigenstates. We discovered that the energy of the ground state for our Hamiltonian was 1.8671 Ha. We then measured the quantum state with various numbers of shots, initialized it as the eigenstate corresponding to the ground state

energy in qiskit, and repeated the experiment 10, 000 times. We select various desirable ranges surrounding the true value, ranging from chemical precision (±0.0015 Ha) to 5 times the chemical precision range (±0.0075 Ha).

For chemical precision, the variation of energy with different final estimation is shown in Fig. 2. To better comprehend how the theoretical estimate mentioned in the subsection "The cost of final estimation" matches with the empirical results shown in Fig. 2, we sorted the energy data points, discarded extreme points based on the confidence level selected, and then compared them to the theoretical estimate as shown in Eq. (6), which corresponds to a confidence level of 68%. This is illustrated in Fig. 3.

It is worth to mention here that setting the number of shots in such a way that the final estimation will achieve the accuracy as desired might not be optimal in some scenarios. Naturally, the first and obvious one is when the total budget is comparable to the number of shots needed for estimation. Here one would significantly decrease the budget for the optimization itself and thus decrease the probability of obtaining a good result. In other
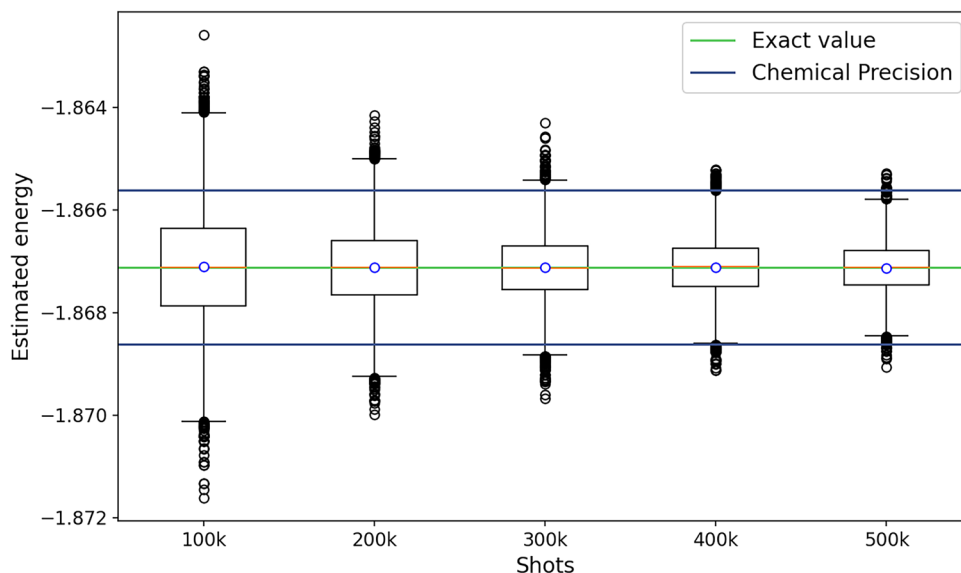

Figure 2. This boxplot represents 10000 energy values for different values of final estimation. The circle represents the mean of the distribution and the yellow line represents the median. It can be clearly seen here that even with the right eigenstate, one requires a large number of shots for the final estimation to get most of our results within chemical precision (±0.0015 Ha).
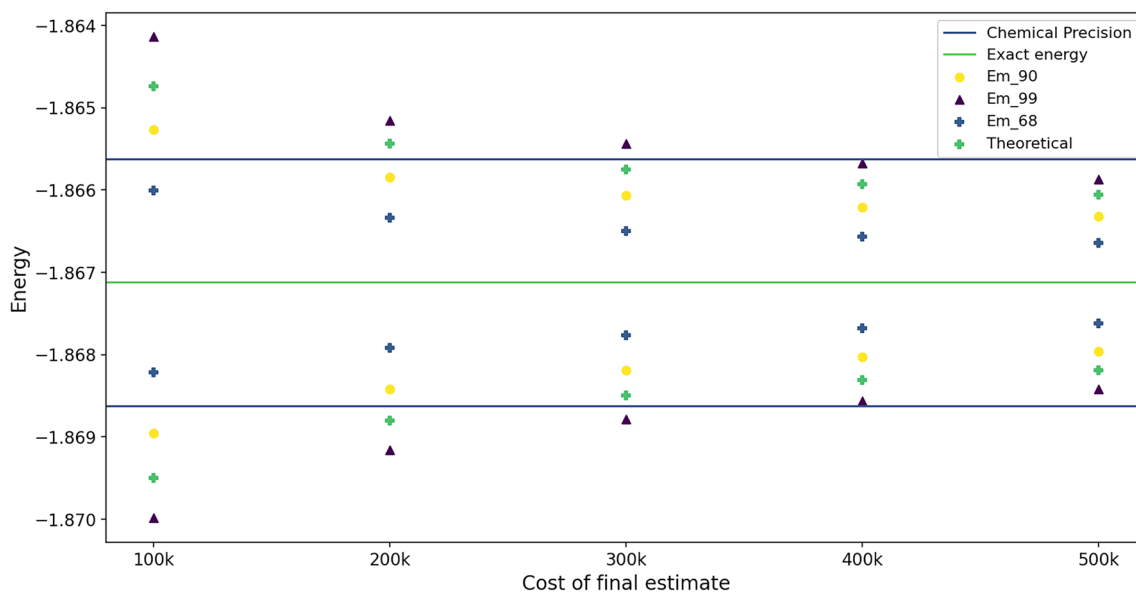

Figure 3. This plot shows a better comparison of theoretical and empirical estimates. Em_{90} represents the end point of energy distribution after discarding extreme points so that 90% of the energies are within this range. Similarly with Em_{68} and Em_{99}. The theoretical estimate is based on the upper bound of the variance and should correspond to Em_{68}, however one can see that the bound is rather conservative and close to about 95%.

words, we would know when we get the good result, but we never get it. In such a case it is better to search a bit more and risk a failure while increasing the probability of success. The other scenario is when the optimization procedure by itself is unreliable, thus resulting in a bad result in many cases even for a high-precision quantum subroutine. Here one will use a lot of repetitions and will have to reduce the final estimation costs accordingly.

Benchmarking of the optimization method

To be able to estimate the capabilities of the optimization method to deal with the optimization task depending on the available budget of shots, we will have to sample it. It is important to stress here that we will not set any internal parameters while sampling, like number of iterations, starting number of samples or the way the procedure changes the number of shots per measurement. The only criterion here will be the total number of shots used n and the quality of the result measured in the probability to be within the desired precision.

We model the method by an exponential function characterized by a set of three parameters

$$p_s = a(1 - e^{-bn}) + c, \tag{7}$$

where c is the "guessing" probability of success without any measurement (expected to be very close to zero for any realistic scenario), $a + c$ the probability of obtaining the desired result in the limit of infinite number of shots used (i.e. in the situation when the quantum subroutine would provide perfect outputs) and b is the parameter expressing how the enhancement of the result depends on the number of shots used.

To obtain such a parametrization, one needs to perform sampling of the procedure. This means running it repeatedly with different number of shots and evaluating the results. While the true minimum is not known, we expect the obtained results to have a mean around the physical value, i.e. there is no dominant local minima the optimization technique ends in. It is also important to stress that while the parameters a, b and c are dependent on the desired precision (the probability of success depends on the desired precision), all these values can be calculated from the same sampling data.

Showcase of fitting on H_2 molecule

To evaluate our method of meta-optimization technique, we used 10000 data points of energies to estimate probabilities of success for the energy to lie within the desired accuracy (expressed in multiples of the chemical precision (ChP)) by varying the shots per iteration, while fixing the `maxiter` (maximum number of iterations) to 100. Results are depicted for different values of desired accuracy in Fig. 4.

In Table 1 we tabulate the resulting fit parameters for different values of accuracy. As one can see, the c value is vanishing as expected, suggesting that random guessing of the state can hardly lead to a reasonable result. Value of b expresses the inversed typical number of shots needed to start obtaining reasonable results. This ranges from about 40000 for small accuracy to about 200000 for large accuracy, what is a very reasonable result.

The a value now expresses the limiting success probability of the optimization procedure. It ranges from about 30% to 60% depending on the desired accuracy, what might seem to be rather small. To validate this result independently, we did perform the same rounds of the optimization method, but using a state vector simulator instead of (simulation of) quantum computer, which virtually simulates infinite number of shots used. Results differ only by a few % compared to the fit parameters proving good extrapolation capabilities of the fit function.
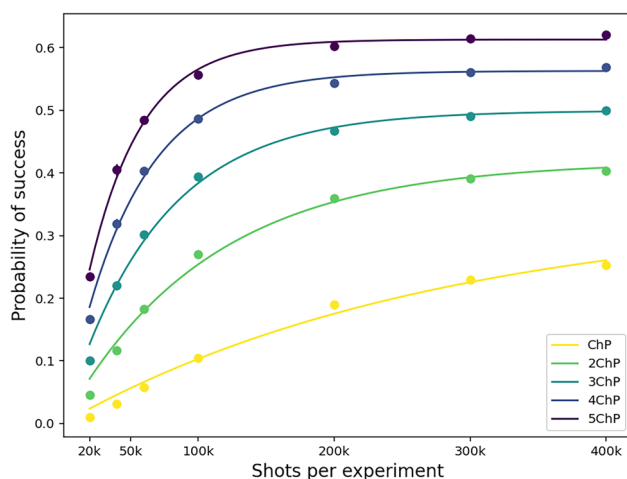

Figure 4. This is the pictorial representation of the probability of success with respect to shots per experiment which is shots per iteration $\times$ 100 (iterations). Differently colored lines represent different desired precision in multiples of chemical precission (ChP = 0.0015 Ha).

Precision (Hartree)	a	b	c	a_{sv}
ChP	0.3416	3.60e−06	9.56e−11	0.3119
2 × ChP	0.4185	9.35e−06	9.58e−18	0.4461
3 × ChP	0.4995	1.46e−05	3.71e−17	0.5313
4 × ChP	0.5629	2.01e−05	1.8e−26	0.5913
5 × ChP	0.613	2.56e−05	2.86e−17	0.6442

Table 1. This table shows the different values of fit parameters of the function $a(1 - e^{-bn}) + c$ corresponding to different accuracy of energies, measured in chemical precision (ChP = 0.0015 Ha). The last column $(a)_{sv}$ represents values obtained by using a state vector simulator instead of quantum measurements at each iteration in the optimization process. As one can see they correspond well to the obtain fitting parameters a, proving good extrapolation capabilities of the fit function.

Repetitions

Instead of investing the whole available budget into a single optimization that, as we have seen in the previous section, is likely to fail, it makes sense to repeat the procedure more times. On the other hand, very few shots have extremely low probability to succeed, so one might expect that there shall exist an optimum of number of repetitions of the whole procedure.

If we repeat the experiment r number of times, then the probability that at least one of the energies lies within desired range is given by

$$P = 1 - (1 - p_s)^r = 1 - (1 - p_s)^{B/n+m}, \tag{8}$$

where B represents the total shot budget and m represents the number of shots used for final estimation of the value of the observable. Note that p_s is also a function of n. To get the best possible results, we have to maximize this total probability function

$$P_{max} = \max_n (1 - (1 - p_s)^r). \tag{9}$$

Instead of maximizing the above expression, one can minimize the following expression to locate the point of maxima

$$\min_n (1 - p_s)^{B/n+m} \equiv \min_n (1 - p_s)^{1/n+m}. \tag{10}$$

Once the optimal n is found, one can calculate the number of repetitions (r) from B, m and n and repeat the experiment r number of times to get the best possible outcomes.

Reliability

Repeating the experiment r times, we get r different results. Then we have to select one of them that shall be presented as the final outcome of the meta-optimization technique. The only reasonable way to do that is to select the result that has the lowest energy stemming from the final estimation procedure. But while for large number of shots used for final estimation the outcoming result will be highly reliable, this is not the case for small budget anymore. In particular, if the precision of the final estimation will be comparable or lower than the desired precision defined for the meta-optimization technique, the single result provided as the outcome might not be the correct one. In other words, even if the optimization procedure in one of the r runs did find a "good" result, the insufficient final estimation was not able to identify it correctly.

To account for this aspect, we define a quantity called "Reliable probability". This is obtained by modifying the probability described in Eq. (8) to incorporate the reliability of the energy value by multiplying it with the probability that the final estimate is good enough. We first calculate the accuracy which can be obtained with a given budget for final estimation E from the Eq. (6)

$$\varepsilon = \sqrt{\frac{\left(\sum_i \sqrt{\sum_j c_{ij}^2}\right)^2}{m}}. \tag{11}$$

As this ε corresponds to standard error, it lies at one σ (standard deviation) away from the mean of our energy distribution. Thus we can then represent the desired precision (d) in terms of σ to get the Z-score (the number of standard deviations the value is above or below the mean value)

$$Z = \frac{d}{\sigma}, \tag{12}$$

and convert it into the corresponding confidence level of the value to lie within the desired range (γ). Finally the probability (P) defined in Eq. (8) is multiplied with this confidence level to get the reliable probability. In our case we define the final reliable probability as a multiplication of the maximal probability defined in Eq. (9) with the reliability factor γ

$$P_{reliable} = \gamma P_{max}. \tag{13}$$

Results for H_2 molecule

To showcase the working procedure of our technique, we provide a pictorial representation of reliable probabilities with varying repetitions and final estimation costs for different accuracy and budget (maximum number of shots that can be spent) in Fig. 5. Here it is clearly visible that for a given budget there exists a clear optimum both for the number of repetitions of the classical optimizer and for the value of final evaluation. This can be understood as follows.

If the final evaluation cost is too large, not enough shots remain for the optimization itself. Thus the final result is evaluated correctly, but is likely out of the desired precision. On the other hand, if the final evaluation costs are too low, one of the results might be good, but we will likely choose the wrong one. Similarly, if the number of repetitions is too low, we might fail to get a good results simply due to the imperfectness of the underlying procedure itself. On the other hand, if the number of repetitions is too high, the budget per repetition is small and does not allow for obtaining a good result with reasonable probability.

In the graph presented in Fig. 6 we show the final results of optimal reliable probability for a wide range of parameters. In essence, it aggregates the optima – red asterisks depicted in graphs Fig. 5 – for different input parameters. This graph showcases the capabilities of the meta optimization technique – for a given budget of shots and desired precision, it gives us the information on the maximum probability of obtaining the desired result. For each point in the graph, though not depicted, we also obtain the information on how to achieve this probability, namely the optimal number of repetitions to be performed and the optimal distribution of shots between the optimization itself and the final evaluation.

Discussion

In this paper we have presented a meta-optimization technique that addresses two major obstacles in using NISQ quantum computers for variational tasks. Our goal is an efficient use of quantum devices by optimizing the number of times they are used during the run of the algorithm.

Existing optimization techniques deal with two major issues. First, even if they are able to identify the optimal state relatively well, the result of the cost function is determined only with a very low accuracy. This is often circumvented by evaluating it in the end using classical means[18,19,22], which is not feasible in scenarios with moderately large inputs. We address this problem by introducing the final evaluation step and optimizing the ratio of quantum resources used for comparison of the obtained results in multiple runs to the cost of optimization itself. It turns out that it might be far from negligible, i.e. to get a reliable result, one has to invest a good portion of quantum resources for its final evaluation.

The other problem of existing techniques is that they are stochastic by their nature. This leads to the inevitable fact that in some cases it returns an incorrect result even if the quantum part works perfectly, i.e. with no noise and statistical deviations. Thus it makes sense to mitigate the risk of investing all quantum resources at hand
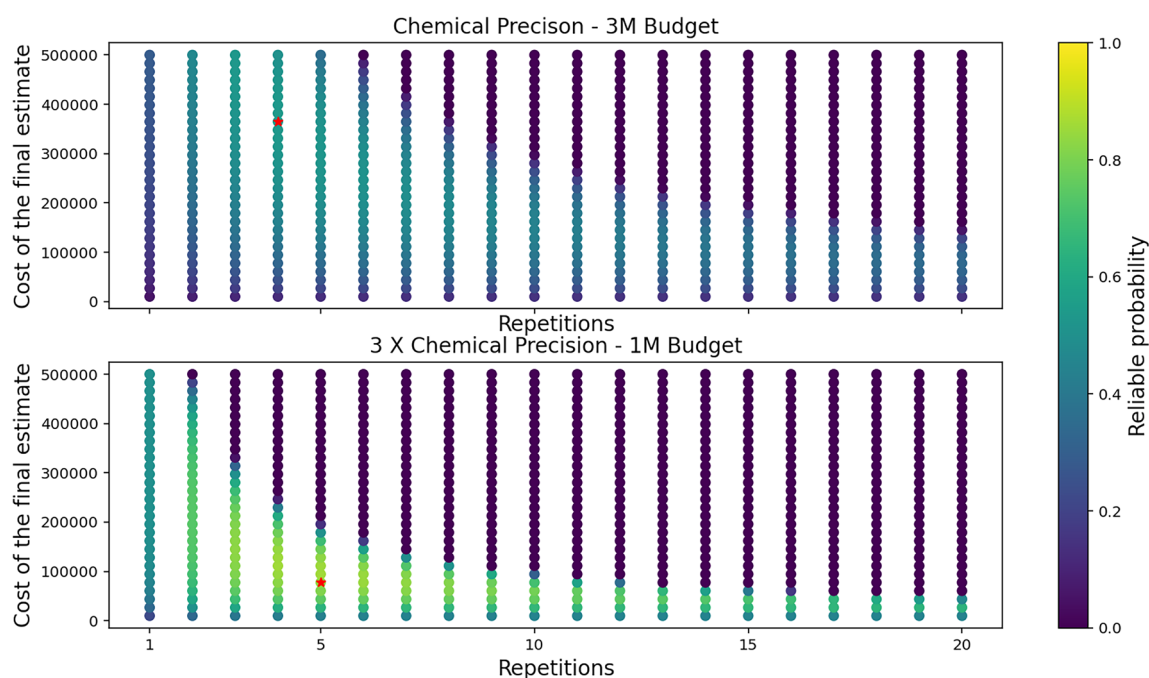

Figure 5. These plots represent the reliable probabilities to obtain a result with energy value within the desired range of accuracy for different values of final estimation and repetitions of the VQE for a fixed shot budget. The configuration which leads to maximum reliable probability is marked with a red asterisk.
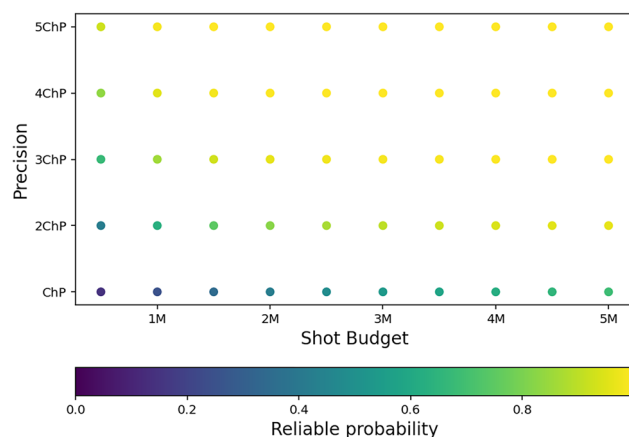
Figure 6. This figure illustrates the maximum reliable probabilities one can obtain for different shot budgets and accuracy. Every point in this figure is the red asterisk marked point in Fig. 5 for a different combination of accuracy and budget.
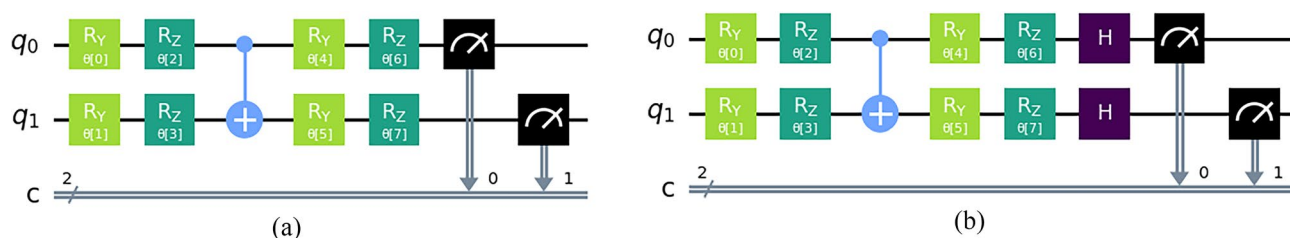

Figure 7. The qubits in the circuit (**a**) are measured in the computational $\{|0\rangle, |1\rangle\}$ basis. To measure the qubits in $\{|+\rangle, |-\rangle\}$ basis, in circuit (**b**) we applied Hadamard gates on both qubits in the end before the measurement.

into a single calculation. We presented a technique to benchmark the underlying optimization procedure and decide about the optimal number of repetitions. Interestingly, it turns out that this number is non-trivial (i.e. not one) for realistic scenarios.

While we deployed our technique on a specific example of SPSA optimization on a hydrogen molecule, it can be applied on a very broad variety of variational quantum algorithms that are currently of a large interest to the quantum computation community. We believe that our meta-optimization can provide an easy to implement step towards practical use of quantum computers. While the method was tested only on a rather simple problem, we believe that the nature of the approach makes the method suggested robust against changes of the system in question, i.e. it shall behave in a similar way even if applied to a larger system involving more qubits. This is due to the fact that the existence of an optimum between investing "too little" or "too much" into a single try is obvious and the nature of the fitting function mimics the expected dependency between precision and cost.

Yet, the exact efficiency and precision of determining this optimum using the method as suggested on large systems forms a standalone project for future research. In particular, there exist a risk that for more complicated systems the determination of optimization function will turn out to be to "costly" in comparison with the gain of it, or, as in any other scenario, unforeseen complications may arise.

Methods

For the quantum subroutine, we have used Ry-Rz ansatz to prepare the state from an initial zero state. This state was then measured in both $\{|0\rangle, |1\rangle\}$ and $\{|+\rangle, |-\rangle\}$ (the eigenstates of Pauli X operator) basis to calculate the expectation value of the Hamiltonian. The procedure was defined by eight parameters $\theta[0]$ to $\theta[7]$ and circuits are shown in Fig. 7.

Data availability

Data and programs used to derive the results presented in this paper are available from the corresponding author upon reasonable request.

References

1. Benioff, P. The computer as a physical system: A microscopic quantum mechanical hamiltonian model of computers as represented by turing machines. *J. Stat. Phys.* **22**, 563–591. https://doi.org/10.1007/BF01011339 (1980).

2. Manin, Y. *Computable and Uncomputable* (Sovetskoye Radio, 1980) ((**in Russian**)).
3. Feynman, R. P. Simulating physics with computers. *Int. J. Theor. Phys.* **21**, 467–488. https://doi.org/10.1007/BF02650179 (1982).
4. Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge University Press, 2000).
5. Shor, P. W. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comput.* **26**, 1484–1509. https://doi.org/10.1137/S0097539795293172 (1997).
6. Grover, L. K. A fast quantum mechanical algorithm for database search. In *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing, STOC '96* 212–219 (Association for Computing Machinery, 1996) https://doi.org/10.1145/237814.237866
7. Preskill, J. Quantum computing in the NISQ era and beyond. *Quantum* **2**, 79. https://doi.org/10.22331/q-2018-08-06-79 (2018).
8. Cerezo, M. *et al.* Variational quantum algorithms. *Nat. Rev. Phys.* **3**, 625–644. https://doi.org/10.1038/s42254-021-00348-9 (2021).
9. Bharti, K. *et al.* Noisy intermediate-scale quantum algorithms. *Rev. Mod. Phys.* **94**, 015004. https://doi.org/10.1103/RevModPhys.94.015004 (2022).
10. McCaskey, A. J. *et al.* Quantum chemistry as a benchmark for near-term quantum computers. *npj Quantum Inf.* **5**, 99 (2019).
11. McArdle, S., Endo, S., Aspuru-Guzik, A., Benjamin, S. C. & Yuan, X. Quantum computational chemistry. *Rev. Mod. Phys.* **92**, 015003 (2020).
12. Dunjko, V. & Briegel, H. J. Machine learning & artificial intelligence in the quantum domain: A review of recent progress. *Rep. Prog. Phys.* **81**, 074001 (2018).
13. Biamonte, J. *et al.* Quantum machine learning. *Nature* **549**, 195–202 (2017).
14. Orús, R., Mugel, S. & Lizaso, E. Quantum computing for finance: Overview and prospects. *Rev. Phys.* **4**, 100028 (2019).
15. Peruzzo, A. *et al.* A variational eigenvalue solver on a photonic quantum processor. *Nat. Commun.* **5**, 1–7. https://doi.org/10.1038/ncomms5213 (2014).
16. Spall, J. C. An overview of the simultaneous perturbation method for efficient optimization. *J. Hopkins APL Tech. Dig.* **19**, 482–492 (1998).
17. Tilly, J. *et al.* The variational quantum eigensolver: A review of methods and best practices, https://doi.org/10.48550/ARXIV.2111.05176 (2021).
18. Kübler, J. M., Arrasmith, A., Cincio, L. & Coles, P. J. An adaptive optimizer for measurement-frugal variational algorithms. *Quantum* **4**, 263. https://doi.org/10.22331/q-2020-05-11-263 (2020).
19. Arrasmith, A., Cincio, L., Somma, R. D. & Coles, P. J. Operator sampling for shot-frugal optimization in variational algorithms. https://doi.org/10.48550/ARXIV.2004.06252 (2020).
20. Sweke, R. *et al.* Stochastic gradient descent for hybrid quantum-classical optimization. *Quantum* **4**, 314. https://doi.org/10.22331/q-2020-08-31-314 (2020).
21. Schuld, M., Bergholm, V., Gogolin, C., Izaac, J. & Killoran, N. Evaluating analytic gradients on quantum hardware. *Phys. Rev. A* **99**, 032331 (2019).
22. Tamiya, S. & Yamasaki, H. Stochastic gradient line bayesian optimization for efficient noise-robust optimization of parameterized quantum circuits. *npj Quantum Inf.* **8**, 90 (2022).
23. Miháliková, I. *et al.* The cost of improving the precision of the variational quantum eigensolver for quantum chemistry. *Nanomaterials* **12**, 243 (2022).
24. Yen, T.-C. & Izmaylov, A. F. Cartan subalgebra approach to efficient measurements of quantum observables. *PRX Quantum* **2**, 040320 (2021).
25. Yen, T.-C., Ganeshram, A. & Izmaylov, A. F. *Deterministic improvements of quantum measurements with grouping of compatible operators, non-local transformations, and covariance estimates*. arXiv preprint arXiv:2201.01471 (2022).
26. Gonthier, J. F. *et al.* Measurements as a roadblock to near-term practical quantum advantage in chemistry: Resource analysis. *Phys. Rev. Res.* **4**, 033154 (2022).

Acknowledgements

Author contributions

All authors contributed to the idea, design and writing of the manuscript. IA was responsible for the numerical simulations and producing graphs and figures.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.