Measuring ansatz capability and benchmarking their performances

Arunava Majumder,[1,*] Bibhas Adhikari,[2] and Leong Chuan Kwek[3,4,5,6]

[1] *Department of Physics, Indianinstituteoftechnology, Kharagpur, India*, 721302
[2] *Department of Mathematics, IndianInstituteofTechnology, Kharagpur, India*, 721302
[3] *Centre for Quantum Technologies, NationalUniversityofSingapore, Singapore*117543, *Singapore*
[4] *MajuLab, CNRS − UNS − NUS − NTUInternationalJointResearchUnit, SingaporeUMI*3654, *Singapore*
[5] *National Institute of Education, NanyangTechnologicalUniversity, Singapore*637616, *Singapore*
[6] *Quantum Science and Engineering Center, NanyangTechnologicalUniversity, Singapore*637616, *Singapore*

Currently available Noisy Intermediate Scale Quantum Devices (NISQ) are small scale devices and error free large scale devices are likely to be available in the next 10-12 years. In recent days the most successful class of quantum algorithms for use on near-term hardware is the so-called variational quantum algorithm (VQA), which employ a classical optimizer to train a parametrized quantum circuit. These algorithms can be used to optimized various spin systems which can be really difficult in classical computer. One such spin system is the *Spin Glass problem*. Many NP-hard problems can be seen as the task of finding a ground state of a disordered highly connected Ising spin glass. They constitute a universality class that is significantly different from the classical, thermal phase transitions. In this article we tried to optimize different models of spin glass system with different number of qubits. The complexity in such systems arises due to the random choices of interaction coefficients which are gaussian in nature. We also tried to observe the variation of ground state with variance and mean of the interaction coefficients.

CONTENTS

* arunavaangshuman@iitkgp.ac.in

I. INTRODUCTION

Quantum computing has emerged as the leading candidate after the realization that the classical simulation of quantum systems is an extremely difficult task as the computational resources required to simulate such systems grows exponentially with the size of the system. The progress in science and engineering as well as considerable investments have increased our capabilities to precisely control quantum systems, leading to the development of quantum devices that are capable of impressive feats. Applications such as simulating large quantum systems or solving large-scale linear algebra problems are immensely challenging for classical computers due their extremely high computational cost. Quantum computing promises to deal with such issues. Not only in the domain of physics but others like finance, drugs, artificial intelligence and few more fields have also realized the advantage of using quantum computing over the classical one.

Nevertheless, the true promise of quantum computers, speedup for practical applications, which is often called quantum advantage, has yet to be realized. Moreover, the availability of fault-tolerant quantum computers appears to still be many years, or even decades, away. The key technological question is therefore how to make the best use of today's Noisy Intermediate Scale Quantum (NISQ) devices to achieve quantum advantage. Any such strategy must account for: Limited numbers of qubits, Limited connectivity of qubits, and Coherent and incoherent errors that limit quantum circuit depth.

Variational Quantum Algorithms (VQAs) have emerged as the leading strategy to obtain quantum advantage on NISQ devices. Canonical examples of VQAs are the variational quantum eigensolver (VQE) and the quantum approximate optimization algorithm (QAOA). To harness the potential of parameterized quantum states, it is crucial to understand how a parameterized quantum state changes as the underlying parameters θ are varied. Moreover, VQAs leverage the toolbox of classical optimization, as VQAs employ parametrized quantum circuits to be run on the quantum computer, and then outsource the parameter optimization to a classical optimizer. This has the added advantage of keeping the quantum circuit depth shallow and hence mitigating noise, in contrast toquantum algorithms developed for the fault-tolerant era.

The report is organized as follows. Here, Sec. II provides the basic knowledge of Quantum Spin glass with two well known hamiltonian model Sherrington–Kirkpatrick and Edwards–Anderson model. Sec.III tells the basic concept of Variational Quantum algorithm. In section IV we briefly introduce the three best possible *problem inspired ansatz*. Section V and VI show the formation of cost function following the given problem and the gradient of the cost, necessary during optimization, respectively. Next in sec.VII we consider two specific models for Quantum Spin Glass and show the optimization of each model with the help of different ansatz from sec.IV. In sect.VIII we describe the overpaarametrization phenomenon that is responsible during the optimization of various models. We also show the possibility of Barren Plateaus in our optimization in the next sec.IX. The measure of efficiency of the ansatz is provided in sec.X with the help of entanglement spectrum. While performing any VQA's the computational complexity is a key factor to consider because of the restricted device size. We show the complexity of the problem ansatz in sec.XI.

II. THE SPIN GLASS PROBLEM

In this work we focus on spin glass problems that have features in common with real life hard optimizations problems and do not admit analytic solutions. A quantum spin glass is a magnetic system that can be described by a quantum mechanical Hamiltonian with spin-glass like features (randomness and frustration). Finding the lowest energy levels of a spin glass system is one of the hardest task for classical computer due to intrinsic random disorder.

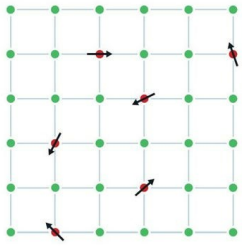

FIG. 1: A spin glass is a metal alloy where the iron atoms (**red**), for example, are randomly mixed into a grid of copper atoms (**green**). Each iron atom behaves like a small magnet, or spin, which is affected by other magnets around it. However, in a spin glass they are frustrated and have difficulty choosing which direction to point.

The behaviour of the spin glass model can be described using various Hamiltonian models.

A. Sherrington–Kirkpatrick spin glass

The Sherrington–Kirkpatrick (SK) spin glass Hamiltonian H_{SK} (Sherrington and Kirkpatrick 1975) is defined on n spins as

$$H = -\sum_{i<j} J_{ij}\sigma_i\sigma_j \tag{1}$$

where σ_j=$(\sigma^x, \sigma^y, \sigma^z)$ are the quantum pauli spin matrices and the couplings J_{ij} are drawn independently from the normal distribution $N(\mu, \sigma_{sk}^2)$ with mean μ and variance σ_{sk}^2.

The hardness of finding the ground state follows directly from the hardness of the field strength $h_j = 0$. This is not true in general for different distributions of field strength h_j. There are known examples in which fields can destroy spin glass behaviour(see, e.g. Young and Katzgraber 2004 , Feng et al 2014). In particular, if the field strengths are much larger than the coupling strengths ($|h_j| >> |J_{ij}|$ for all i, j), then the energy is minimized trivially when all the spins each minimize the energy with respect to their individual fields. While the distribution of field strengths could be used to tune the problem hardness, we do not use it in this way here, and only consider cases where the field and coupling strengths are drawn from the same distribution.

SK models are in fact NP hard [12] for classical computers to optimize when the system size grows but quantum computers can efficiently reduce the amount of resources required to minimize the energy. So SK model with larger systems can be efficiently tackled using VQE algorithm on an actual hardware. Later we will provide few examples where we showed how we can minimize the structure.

B. Edwards–Anderson model

In this model, we have spins arranged on a d-dimensional lattice with only nearest neighbor interactions similar to the Ising model. This model can be solved exactly for the critical temperatures and a glassy phase is observed to exist at low temperatures. The Hamiltonian for this spin system is given by:

$$H = -\sum_{<ij>} J_{ij}\sigma_i\sigma_j \tag{2}$$

where σ_i refers to the Pauli spin matrix for the spin-half particle at lattice point i. A negative value of J_{ij} denotes an antiferromagnetic type interaction between spins at points i and j. The sum runs over all nearest neighbor positions on a lattice, of any dimension. The variables J_{ij} representing the magnetic nature of the spin-spin interactions are called bond or link variables and again the J_{ij}'s are taken from gaussian distribution.

III. BASIC CONCEPTS AND TOOLS OF VQA

One of the main advantages of Variational Quantum Algorithms (VQAs) is that they provide a general framework that can be used to solve a wide range of problems. The first step is to consider a task one wishes to solve with a total

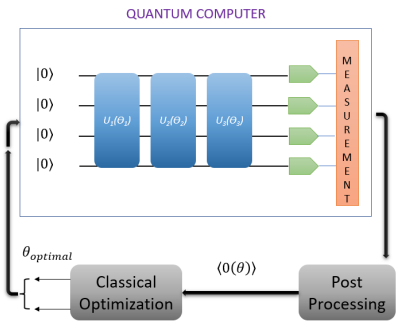

FIG. 2: **Schematic diagram of a Variational Quantum Algorithm (VQA).** The inputs to a VQA are: a cost function $C(\theta)$ which encodes the solution to the problem, an ansatz whose parameters are trained to minimize the cost, and (possibly) a set of training data used during the optimization.

description of the problem, and also possibly a set of training data. As shown in Fig. 2, the first step to developing a VQA is to define a cost (or loss) function $C(\theta)$ which encodes the solution to the problem. One then proposes an ansatz, i.e., a quantum operation depending on a set of continuous or discrete parameters θ that can be optimized (see below for a more in-depth discussion of ansatze). This ansatz is then trained (with data from the training set) in a hybrid quantum-classical loop to solve the optimization task. As in Fig. 2 VQAs employ a quantum computer to estimate the cost function $C(\theta)$ (or its gradient) while leveraging the power of classical optimizers to train the parameters θ.

IV. ANSATZ

Perhaps the most paradigmatic application of VQAs is estimating low-lying eigenstates and corresponding eigenvalues of a given Hamiltonian[VQE] by minimizing the efficient cost obtained with the help of a suitable ansatz. Thus the ansatz $U(\theta)$ is an important ingredient of VQA.

The specific structure of an ansatz will generally depend on the task at hand, as in many cases one can employ information about the problem to tailor an ansatz. These are the so-called *problem-inspired ansatze*. On the other hand, some ansatz architectures are generic and problem-agnostic, meaning that they can be used even when no relevant information is readily available.

In general an ansatz can be expressed as as the product of L sequentially applied unitaries [2] i.e. the circuit has a total L repetitions of the same initial layer.

$$\begin{aligned} U(\theta) &= \prod_{l=1}^{L} U_L(\theta_l) W_l \\ &= \prod_{l=1}^{L} e^{-i\theta_l V_l} W_l \end{aligned} \tag{3}$$

where V_l are the local Hermitian operators and W_l are the unparameterized gates in L^{th} layer of the circuit [2,11].

We used three different kinds of ansatz to perform the variational optimization on different spin glass models. Below only one layer of each ansatz is presented and in general according to the given problem there can be any L number of layers in a circuit

1. Type 1 Ansatz

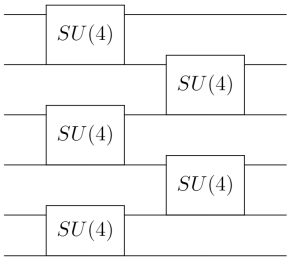

FIG. 3: This type of ansatz is called universal layered ansatz (ULA) with a depth = 1 here.

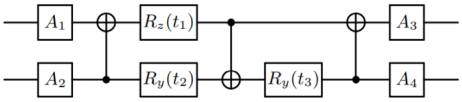

FIG. 4: This is the decomposition of $SU(4)$ [13]. Here A_1,A_2,A_3 and A_4 are the general $SU(2)$ gates.

2. Type 2 Ansatz

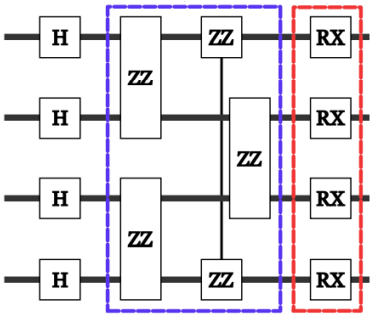

FIG. 5: Here ZZ are the quantum gate corresponding to the Z-Z interaction i.e. $e^{-iZ\otimes Z\theta}$ [6] and so on. This is also called Hamiltonian Variational Ansatz (HVA) and here too the circuit depth is set to be 1.

3. Type 3 Ansatz

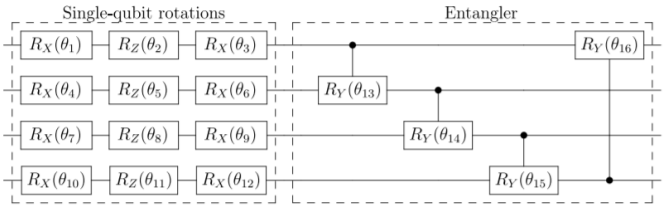

FIG. 6: This is another kind of ansatz where in one part of the layer we have single qubit gates and the other contains entangler gates and the depth of the circuit is chosen to be 1. [14]

V. COST FUNCTION

A crucial aspect of a VQA is encoding the problem into a cost function. Similar to classical machine learning, the cost function maps values of the trainable parameters θ to real numbers. More abstractly, the cost defines a hyper-surface usually called the cost landscape (see Fig. 1) such that the task of the optimizer is to navigate through the landscape and find the global minima. Without loss of generality, we consider that the cost can be expressed as

$$\begin{aligned} C(\theta) = E(\theta) &= \langle 0|U(\theta)^\dagger H U(\theta)|0\rangle \\ &= \langle \psi(\theta)|H|\psi(\theta)\rangle \end{aligned} \tag{4}$$

where $|\psi(\theta)\rangle$ is the called the parameterized state, $U(\theta)$ is the parametrized unitary and H is the problem Hamiltonian. According to the Rayleigh-Ritz variational principle, the cost is meaningful and faithful as $C(\theta) \geq E_G$ with equality holding if $|\psi(\theta)\rangle$ is the ground state $|\psi_G\rangle$ of H. Here E_G will be the lowest state energy and θ being the array of variational parameters.

In general if we consider Quantum Neural Network(QNN) then we have a large training set of various measurement operators and input states then the cost will be changed accordingly

$$C(\theta) = \sum_k f_k(Tr[O_k U^\dagger(\theta)\rho_k U(\theta)]) \tag{5}$$

where ρ_k are input states from a training set, O_k are a set of observables, and f_k are functions which encode the task at hand.

Let us now discuss desirable criteria that the cost function should meet. The cost must encode the solution of the problem. Second, one must be able to efficiently estimate $C(\theta)$ by performing measurements on a quantum computer and possibly performing classical post-processing. Also, the cost must be trainable, which means that it must be possible to efficiently optimize the parameters θ to reach the solution of the given problem.

For a given VQA problem to be implementable in NISQ hardware, the quantum circuits that estimate $C(\theta)$ must keep the circuit depth and ancilla requirements low enough. This is due to the fact that NISQ devices are vulnerable to gate errors, have limited qubit counts, and that these qubits have low coherence time. Hence the construction of efficient cost evaluation circuits is an important aspect of VQA research.

VI. GRADIENTS

After having the cost function and ansatz, the next step is to train the parameters θ and solve the optimization problem of Eq.4. It is known that for many optimization tasks using information in the cost function gradient (or in higher-order derivatives) can help in speeding up and guaranteeing the convergence of the optimizer. One of the main advantages of many VQAs is that, as discussed below, one can analytically evaluate the cost function gradient. And calculating the gradient is the most important part while updating the parameters during the optimization.

When the random parameterized quantum circuits(RPQCs) are parameterized in the way showed in eq.4, the gradient of the objective function takes a simple form:

$$\partial_k C(\theta) = \frac{\partial_k C(\theta)}{\partial \theta_k} = i\langle 0|U_-^\dagger [V_k, U_+^\dagger H U_+] U_-|0\rangle \tag{6}$$

where we define $U_- \equiv \prod_{l=0}^{k-1} U_L(\theta_l) W_l$ and $U_+ \equiv \prod_{l=k}^{L} U_L(\theta_l) W_l$.

The gradient can be easily obtained from a hardware-friendly protocol to evaluate the partial derivative of $C(\theta)$ with respect to θ_k often referred to as the *parameter-shift rule*.

VII. APPLICATION

1. SK-model

Firstly we will deal with SK model described in eq.1 and try to optimize two different structures Fig.7 and 11. We will show different variaition of the model based on different ansatz showed in Fig.3, 5 and 6. We will run the optimization for 10 different runs for each variance point and finally average over all the

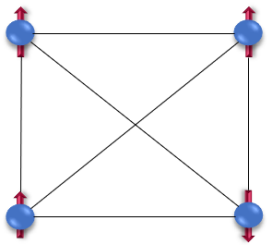

FIG. 7: In this figure we have a 4 qubit complete graph, the nodes are representing spin qubits and the edges are representing the interactions among them.

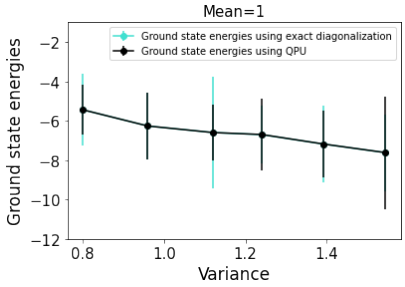

FIG. 8: Here we are optimizing the system shown in Fig.7 using the ULA ansatz, where we are varying the lowest energy with the varience σ^2_{SK} from where the interaction coefficients J_{ij}'s are drawn and the mean of the gaussian distribution is set to be 1 in this case.

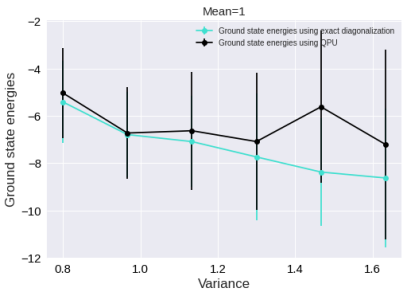

FIG. 9: Here we use the ansatz shown in Fig.5 to optimize the 4 qubit SK model shown in Fig.7

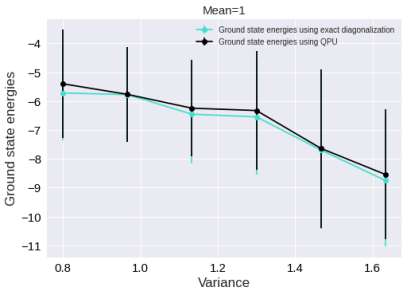

FIG. 10: Here we use the ansatz shown in Fig.6 to optimize the 4 qubit SK model shown in Fig.7 i.e. the type 3 ansatz.

2. *Edwards–Anderson model*

Here we will take a look at few simple structures that can be drawn from the ED model shown in eq.2. We use two specific structures of systems an open and a closed chain of 6 qubits.

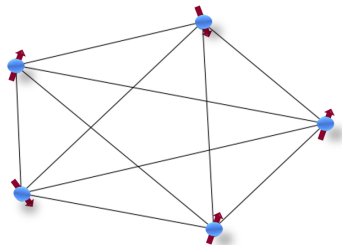

FIG. 11: In this case the SK model is having 5 qubits so it would be a 5 qubit complete graph.

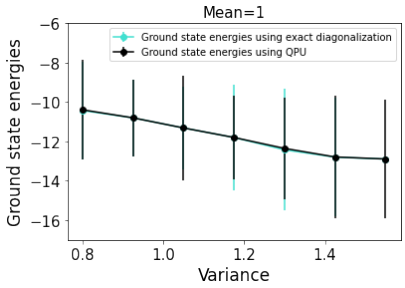

FIG. 12: Here we are optimizing the system shown in Fig.11 using the ULA ansatz. Once again the mean of the gaussian distribution is set to be 1 in this case.

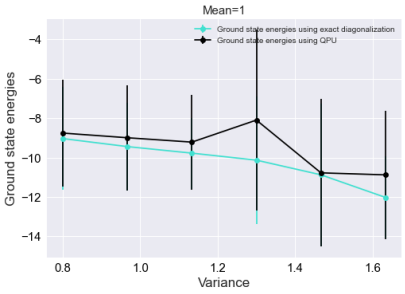

FIG. 13: Here we are using the HVA ansatz shown in 5

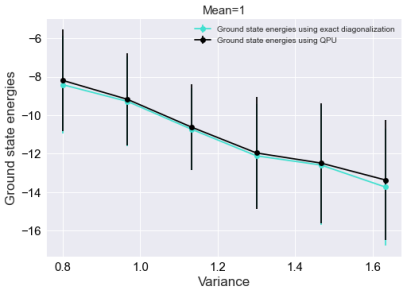

FIG. 14: Using the ansatz shown in Fig.6

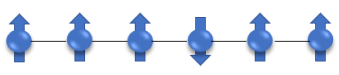

FIG. 15: Nodes represents qubits and edges represents their in between interactions.

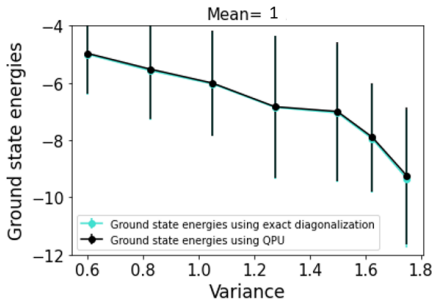

FIG. 16: To perform this optimization we used ULA in Fig. 3.

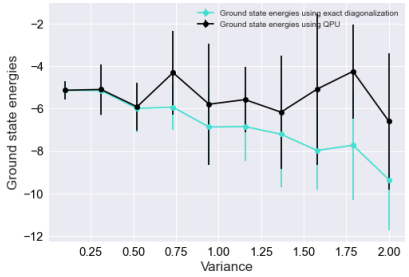

FIG. 17: To perform this optimization we used HVA in Fig. 5.

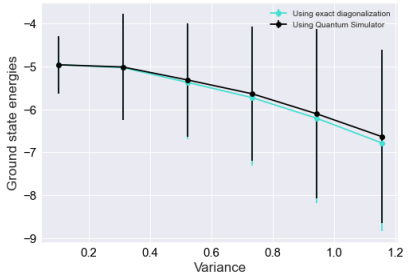

FIG. 18: To perform this optimization we used hardware efficient ansatz in fig. 6.

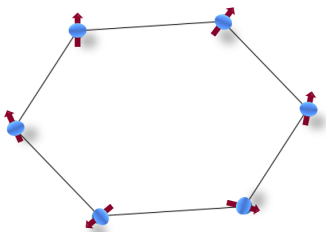

FIG. 19: To perform this optimization we used hardware efficient ansatz in fig. 6.

VIII. OVERPARAMETERIZATION AND COST LANDSCAPE

The prospect of achieving quantum advantage using VQA is exciting. It can be good to know how the properties (e.g., the total number of parameters, say, M) affect the loss landscape and it is crucial to solve large scale problems. In this section, we rigorously show the overparametrization [5] phenomenon in Parametrized Quantum Circuits(PQC). We define overparametrization as the regime where the quantum circuit has more than a critical number of parameters, say, M_c that allows it to explore all relevant directions in state space and also the cost can effectively produced a landscape where the global minimum can contain the solution of the problem. We show that with the gradient descent always converges to an accurate solution when the total number of parameters is M_c or greater and that a

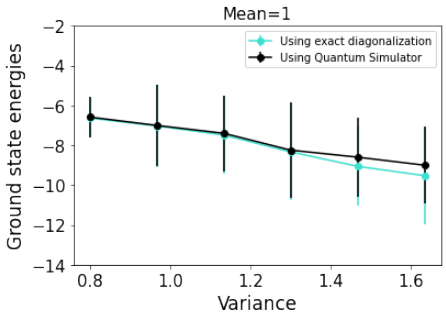

FIG. 20: To perform this optimization we used ULA ansatz in Fig. 3

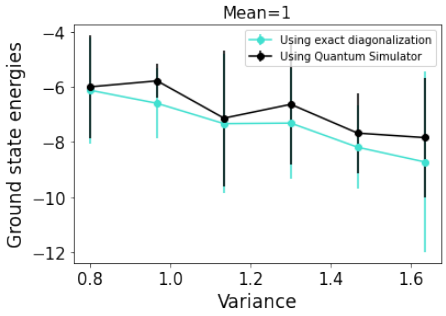

FIG. 21: To perform this optimization we used HVA in fig. 5.

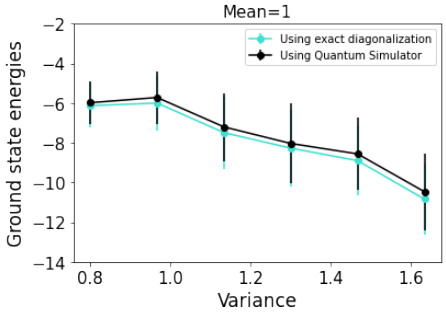

FIG. 22: To perform this optimization we used hardware efficient ansatz in fig. 6.

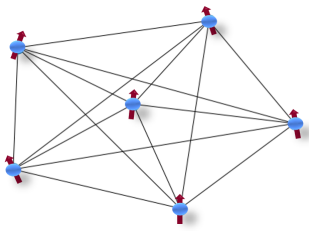

FIG. 23: Considering 6 qubit complete graph corresponding to the SK model.

"computational phase transition" is observed between an underparametrization ($< M_c$) and an overparametrization ($> M_c$).

Although for VQE algorithms it is clear that we have $minE(\theta)_{L+1} \leq minE(\theta)_L$, it is not clear if this minimum can be found consistently due to the nonconvexity of the energy landscape.

In Fig. 24, 25, 26, 27 and 28 we try to explain how overparametrization can effect the cost landscape.

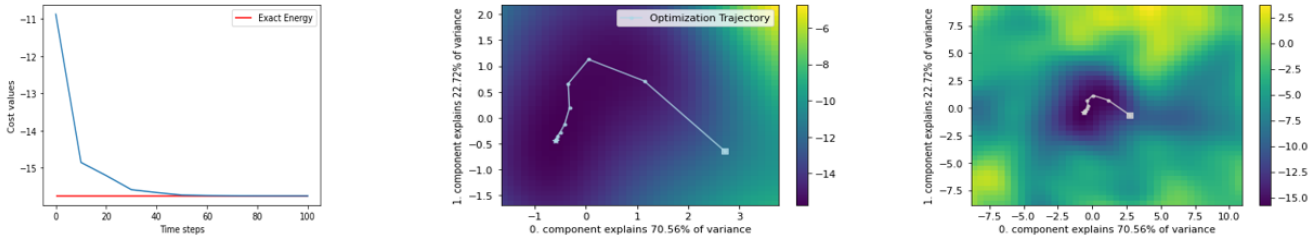

FIG. 24: Here we are optimizing the 6 qubit SK model showed in Fig.23 using the ULA in Fig.3. The ULA already has a lot of parameters so wihthin **depth = 1** we can have *overparametrization* and the cost landscape can efficiently contain the minimum energy. The path in the 2^{nd} figure shows the evolution of the parameters during optimizarion and the 3rd one is zoom out view of the higher dimensional cost landscape.

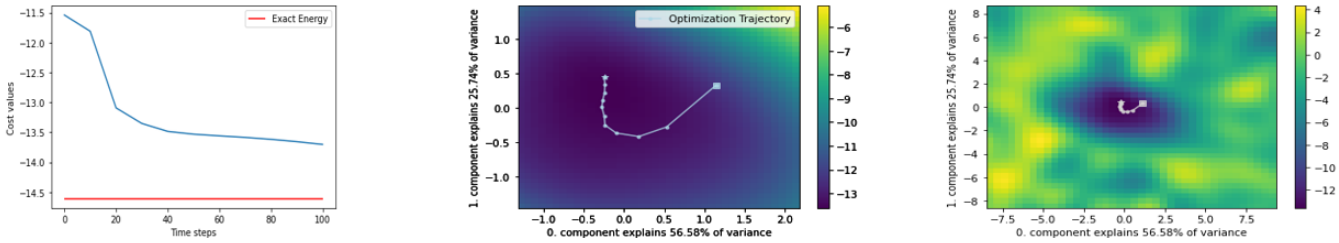

FIG. 25: Here we are using HVA the in Fig.5 to optimize the system with **depth = 1**. Now in this case due to the phenomenon called *underparametrization* the cost landscape doesn't produce the global minimum that can contain the solution of the problem (in zoom out view).

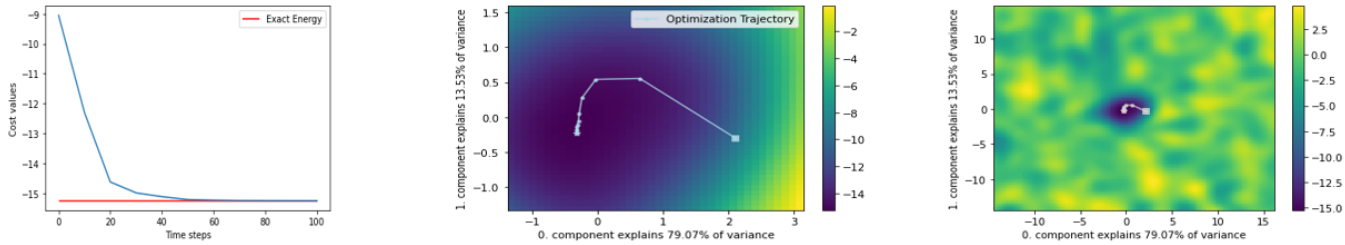

FIG. 26: Here we are using the same HVA the to optimize the system, now, with **depth = 2**. In this case we successfully overcome *underparametrization* and due to *overparametrization* the cost can efficiently reach the actual minimum energy.

IX. BARREN PLATEAUS IN VQE

Barren-plateau (BP) is a phenomenon that has been observed for the VQE on random quantum circuits, where all it's gradients are exponentially close to zero with overwhelmingly high probability, making local optimization within the ansatz space extremely challenging. The barren plateau phenomenon is due to the fact that RQCs consisting of single- and two-qubit gates form a 2-design, which means that the gradients of the energy objective function will obey the same concentration of measure properties as if the circuits are Haar random unitaries [2,10,11]. And due to Haar randomness the circuits are highly expressive. Expressibility [10] of quantum circuits helps them to reach any point on Hilbert space and that leads to a suitable cost landscape which ultimately helps the optimizer to reach the global minimum with true ground state energy.

BP was first noticed in deep unstructured parametrized quantum circuits, exhibit BPs when randomly initialized. This can be avoided by initializing with *problem inspired ansatz* i.e. one needs to select a suitable ansatz according

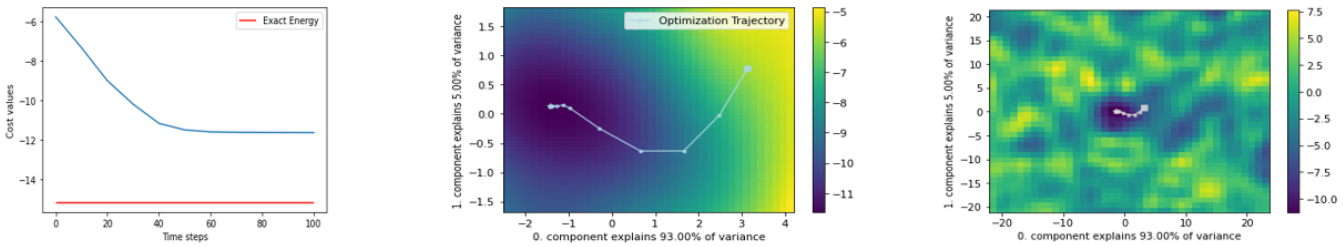

FIG. 27: Using the ansatz in Fig.6 we now try to optimize the system with **depth = 1**. Here too due to the *underparametrization* the cost is not expressive enough to contain the solution of the problem. From the zoom out view in 3^{rd} figure we can also realise that the landscape contain multiple local minimum.

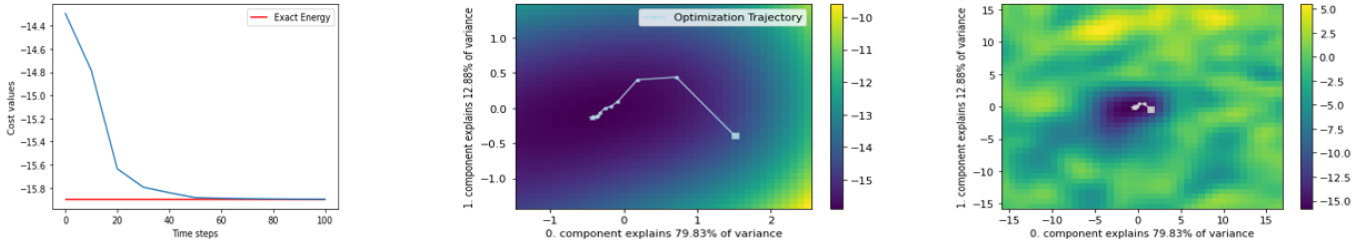

FIG. 28: Now with the same ansatz in Fig.6 with **depth = 2** again we successfully overcome *underparametrization* and the *overparametrization* helps the cost to be more expressive. From the 3^{rd} figure it is also clear that the cost dosen't contain that much of local minimum which is a huge barrier during optimization.

to the given problem that lowers the probability of getting stucked in the BPs. Randomly initializing the parameters of an ansatz can lead to the algorithm starting far from the solution, near a local minima, or even in a region with barren plateaus. Hence, optimally choosing the seed for θ at the beginning of the optimization is an important task.

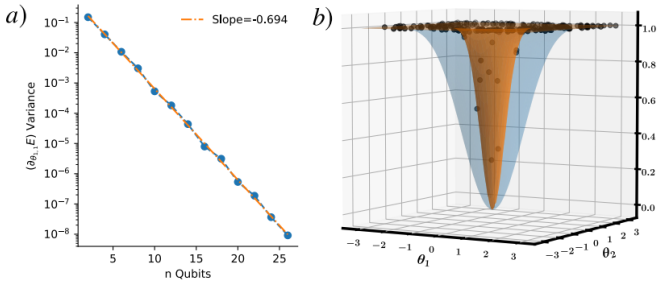

FIG. 29: Barren Plateau (BP) phenomenon. a) Variance of the cost function partial derivative versus number of qubits. Results were obtained from a VQE implementation with a deep unstructured ansatz. The y-axis is on a *log* scale. As the number of qubits increases the variance vanish exponentially with the system size. b) Visualization of the landscape of a global cost function which exhibits a BP for the quantum compilation implementation, taken from. The orange (blue) landscape was obtained for $n = 24(n = 4)$ qubits. As the number of qubits increases, the landscape becomes flatter. Thus the cost values shrinks exponentially with n

Barren Plateau (BP) is an important concept to consider in VQA problems and avoiding this phenomenon is a crucial task to perform during optimization. The conclusion from Fig.30 would be that our all three *problem inspired ansatz* are suitable interms of BPs but the two ansatz in Fig.5 and 6 have high probability that they are not expressive enough as the ULA (Fig.3) with a *depth* $= 1$ because they don't have a certain threshold number of parameters to produce a suitable cost landscape. But with a *depth* $= 2$ or more, they become *overparametrized* and the expressibility of those ansatz become more probable. But again more expressiveness leads to greater probability of having BPs so one has to be careful while increasing the number of layers in a Random PQC (RPQC).

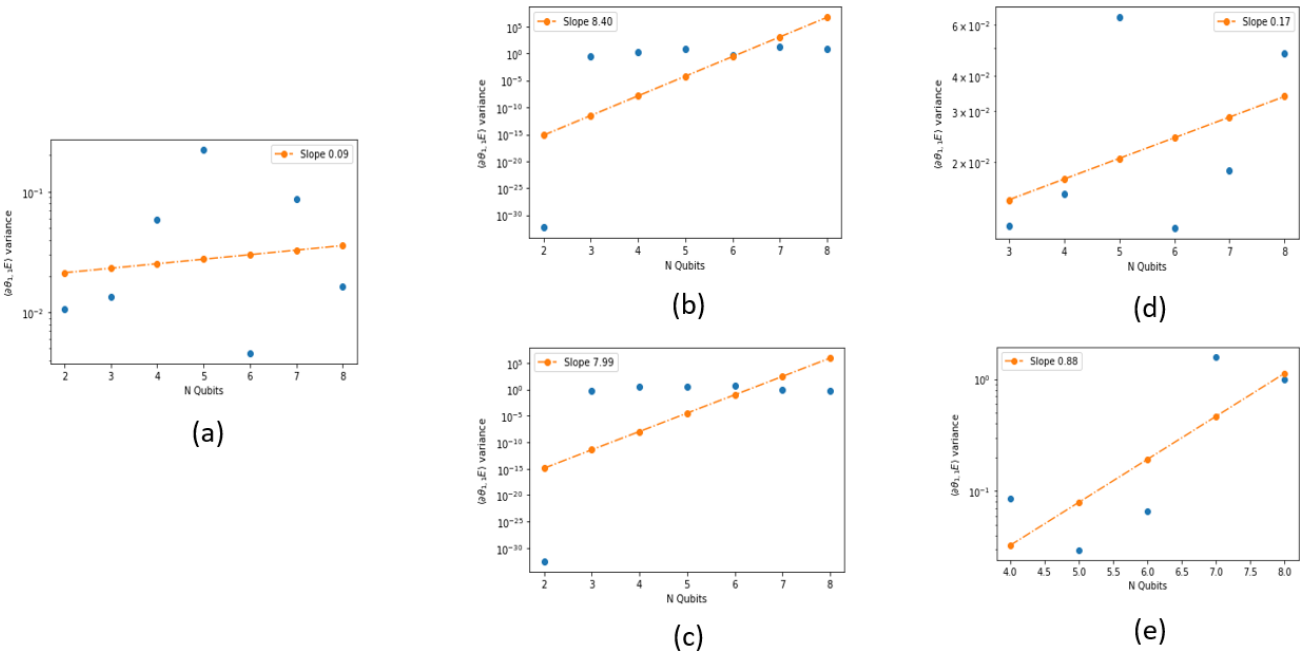

FIG. 30: Here, for the cost function $C(\theta)$ we consider the SK (Eq.1) hamiltonian (complete graph) for different N number of qubits (x-axis). The y-axis is on a *log* scale shows the variance of gradient of the cost $Var[\partial_k C(\theta)]$ $w.r.t$ the last parameter but in general can be done for any parameter θ_k. (a) Shows the $Var[\partial_k C(\theta)]$, while considering ULA in Fig.3 as the problem inspired ansatz, almost remain constant for all $N = 2$ to $N = 8$. (b) Shows the varience remains constant too with HVA of Fig.5 with $depth = 1$. (c) Here again we take the HVA but now with $depth = 2$ and again the varience remains almost constant upto $N = 8$. (d) Considering the ansatz in Fig.6 with $depth = 1$ and we are getting a decent variance for all system size. (e) Lastly, using the same ansatz as (d) but with $depth = 2$ the variance remains non-zero.

X. ENTENGLEMENT SPECTRUM

In classical world, one uses entropy to quantify the lack of knowledge of the state of the system due to thermal fluctuations. However, for a quantum system at zero temperature, the entropy of a subsystem has a different origin which is entanglement. To quantify it, we use the bipartite entanglement entropy, which is defined as the von Neumann entropy of the reduced density matrix ρ_A. To obtain this reduced density matrix, we divide the system into two subsystems A and B and trace out subsystem B just as shown in Fig.31

$$\rho_A(|\psi\rangle) = Tr_B(|\psi\rangle_{AB}\langle\psi|) \tag{7}$$

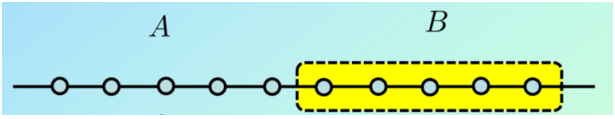

FIG. 31: Example of a 10 qubit closed chain. The division of the full system into two subsystems A (blue) and B (red) on a 1D chain.

where $|\psi\rangle$ is a pure state. For example, for an eight-spin qubit model on a cycle graph, a typical bipartition is given in Fig.31. In recent years, the importance of entanglement in condensed-matter physics has been elucidated in several systems through the study of the scaling behavior of the entanglement entropy, which has enabled the identification and characterization of exotic phases of matter such as topological quantum states and quantum spin liquids. Full characterization of the entanglement properties of a system cannot be done by looking solely at the entanglement entropy. The so-called entanglement spectrum has a much richer structure and has been used to study many-body

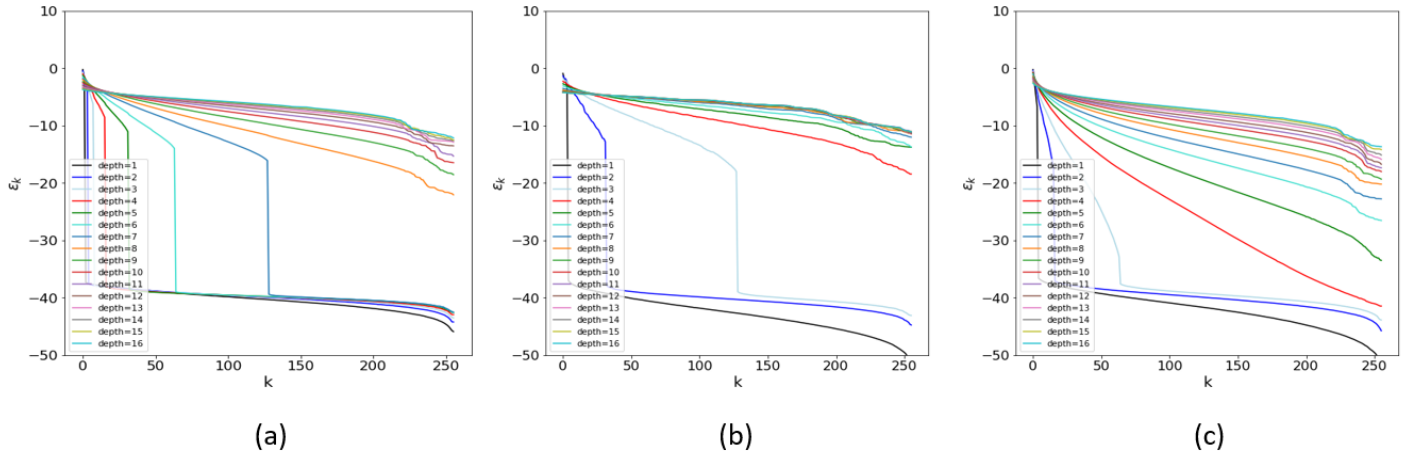

FIG. 32: Here we take a 16 qubit ansatz where we fixed the input state $|\psi_0\rangle$ to be the same for all three ansatz. We then traced out the half of the structure and obtain the entanglement(ENT) spectrum for the reduced density matrix ρ_A explained in Sec.X. (a) shows the ENT spectrum for ULA in Fig. 3. It tells us that it requires more than depth = 8 to reach the MP distribution. (b) It shows the ENT spectrum for HVA in Fig.5 and requires depth =4 to reach MP distribution. (c) Shows the ENT spectrum for ansatz in Fig.6 and it also takes depth around 8 to reach the MP distribution.

localization. In addition, the entanglement spectrum has been used to study the properties of variational methods such as the restricted Boltzmann machine. The entanglement spectrum is defined as the logarithm of the eigenvalue spectrum of the reduced density matrix.

$$\{\xi_k\} = \{log(p_k)\} \tag{8}$$

Where $\{p_k\}$ denotes the eigenvalues of the reduced density matrix ρ_A [8].

A. Application of entanglement entropy

The effectiveness of a VQE optimization is determined by two factors. First, one requires an expressive enough ansatz space that contains the ground state. The ansatz space of a specific unitary with depth-L refers to the set of all possible quantum states that can be reached by applying a depth-L RPQC corresponding to a fixed initial state $|\psi_0\rangle$. Second, the non-convex cost landscape induced by the variational energy, showed in sec. VIII must be favorable, in the sense that the optimization does not get stuck in local minima and can reliably produce as well as reach the true ground state.

Here, we investigate the properties of the ansatz space by examining the entanglement spectra of the quantum states generated from all three ansatz with random initialized parameters sampled uniformly in the range $[0, 2\pi]$. For each ansatz, we sample 200 sets of random parameters and calculate the average entanglement spectrum of the resulting state.

If the average spectrum of the sampled states follows a distribution close to the Marchenko-Pastur (MP) distribution [6,8], a random HVA state has an entanglement spectrum that resembles that of a Haar random state [11]. On the contrary, a distribution far away from the MP distribution indicates a restricted manifold of states that has a nonrandom structure.

The entangles spectrum for each the three ansatz in Fig.3, 5 and 6 is shown in Fig.32.

XI. COMPUTATIONAL COMPLEXITY OF THE PROBLEMS

While working with three suitable ansatz with a required depth to reach the minima we also need to ensure the computational cost of these ansatz and it must be limited to run on a real noisy hardware.

A. Type 1

The ansatz showed in Fig.3 i.e. the ULA is a depth one circuit. It consists of a bunch of $SU(4)$ gates placed between every pair of qubits. Now each $SU(4)$ has 15 variational parameters for 15 single qubit parametrized gates and 3 unparametrized CNOT gates. The sixe of ULA solely depends on the number of qubits. Thus, for n qubits there would be a total of $n-1$ number of $SU(4)$ gates. Then gate complexity would be

$$18 \times (n-1) = O(n) \tag{9}$$

and the total number of parameters for any n qubit ULA would be of the order of

$$15 \times (n-1) = O(n) \tag{10}$$

B. Type 2

The ansatz showed in Fig.5 i.e. the HVA is a depth = 1 or 2 circuit depending on the problem. Now this type of ansatz is highly dependent on the structure of the Hamiltonian i.e. which type of graph it represents. In Fig.5, ignoring the 1^{st} layer of unparametrized gates, the ansatz would contain only XX, YY and ZZ two qubit gates for each of the edges for both SK (eq.1) and ED (eq.2) model Hamiltonians and thus each edge would contain a total of $(7+7+3) = 17$ gates.

For the worst possible case, say for *SK model*, the corresponding graph will be a complete graph and thus having a total of $\frac{n(n-1)}{2}$ number of edges. So the gate as well as parameter complexity becomes

$$17 \times \frac{n(n-1)}{2} = O(n^2) \tag{11}$$

And for other cases apart from *complete dense graph* the complexity remains $O(n)$.

C. Type 3

The third ansatz doesn't depend on the topology of the system. Thus it only varies with the number of qubits. The gate complexity remains

$$3 \times n + 2 \times n = O(n) \tag{12}$$

$3n$ for single qubit gates and $2n$ for two-qubit controlled rotation.

XII. EXPRESSIBILITY OF VARIOUS ANSATZES

XIII. SENSITIVITY ANALYSIS OF PARAMETRIZED QUANTUM CIRCUITS

In this section we shall focus on analysing the sensitivity of Parametrized Quantum Circuits(PQC). Such an analysis is necessary to understand how much of a change in parameter in the circuit affects the change in quantum state i.e. the output probability distribution. For such analysis, we shall perturb the parametrized quantum circuit slightly with a magnitude t along a unit vector $\overline{\delta}$. To measure how much the output quantum state changes w.r.t perturbation, we shall use the notion of *Fidelity* instead of standard Euclidean distance measure because all parameter components generally do not contribute equally to the output quantum state. So a better measure is to find the angular deviation of the original output state to the perturbed one. In pursuit of such measurement, fidelity is extremely useful. Suppose, we have two pure quantum states viz. $|\psi(\overline{\theta})\rangle$ and $|\psi(\overline{\theta'})\rangle$. Then, The Fidelity $\mathcal{F}_{\overline{\theta},\overline{\theta'}}$ between $|\psi(\overline{\theta})\rangle$ and $|\psi(\overline{\theta'})\rangle$ is equal to $|\langle\psi(\overline{\theta'})|\,|\psi(\overline{\theta})\rangle\,|^2$. We shall derive two lower bounds for the Fidelity generalized PQCs for slight perturbations. However, we shall also provide a universal bound which is greater than 0 for heavily perturbed PQCs as well. Such a lower bound has been non-existent in literature yet and this provides a better lower bound than the previously existing lower bound 0, which is an obvious lower bound for Fidelity between any two quantum states. Our results are as follows:

Theorem XIII.1 *Let $|\psi(\overline{\theta})\rangle$represent a pure quantum state where $|\psi(\overline{\theta})\rangle = U(\overline{\theta})\,|\psi\rangle = \Pi_{k=1}^{K} e^{-\iota\theta_k H_k}\,|\psi\rangle$ where H_k's are known Hermitian matrices and $|\psi\rangle$ is an unit vector . Provided a fixed direction unit vector $\overline{\delta}$ and a perturbation t, the Fidelity between the quantum states*

$$|\langle\psi(\overline{\theta}+t\overline{\delta})|\,|\psi(\overline{\theta})\rangle\,|^2 = \begin{cases} \geq |1-|\mathbf{H}^2/2||^2 \ \textit{if}\ \mathbf{H}\leq 1 \\ \geq |e^{-\mathbf{H}}/\mathbf{H}|^2 \ \textit{if}\ \mathbf{H}\leq 1 \end{cases}$$

where $\mathbf{H} = \sum_{k=1}^{K}|t\delta_k|||H_k||$ and $||.||$ is a matrix norm.

Proof: Let $|\psi\rangle$ be an initial state vector and $U(\overline{\theta})$ be a unitary matrix which is represents a parametrized quantum circuit having a mathematical representation $\Pi_{k=1}^{K} e^{-\iota\theta_k H_k}\,|\psi\rangle$. Then we have $|\psi(\overline{\theta})\rangle = U(\overline{\theta})\,|\psi\rangle = \Pi_{k=1}^{K} e^{-\iota\theta_k H_k}\,|\psi\rangle$. Now let us perturb the parametrized state vector by t units in the direction of the unit vector $\overline{\delta} = (\delta_j)_{1\times K}$. Then the perturbed vector $|\psi(\overline{\theta}+t\overline{\delta})\rangle$ can be written as

$$|\psi(\overline{\theta}+t\overline{\delta})\rangle = |\psi(\overline{\theta})\rangle + t\frac{d\,|\psi(\overline{\theta})\rangle}{dt}|_{t=0} + \frac{t^2}{2!}\frac{d^2\,|\psi(\overline{\theta})\rangle}{dt^2}|_{t=0} + \dots \tag{13}$$

Thus , the fidelity

$$\mathcal{F} = |\langle\psi(\overline{\theta})|\,|\psi(\overline{\theta}+t\overline{\delta})\rangle\,|^2 = |1 + t\,\langle\psi(\overline{\theta})|\,\frac{d\,|\psi(\overline{\theta})\rangle}{dt}|_{t=0} + \frac{t^2}{2!}\,\langle\psi(\overline{\theta})|\,\frac{d^2\,|\psi(\overline{\theta})\rangle}{dt^2}|_{t=0} + \dots|^2 \tag{14}$$

since $\psi(\overline{\theta})$ is an unit vector. The expressions of each derivative gets cumbersome as the order of derivative becomes higher. However, we shall explicitly write the expressions of the first and second order derivatives left multiplied by the inverse of the parametrized circuit because we need to calculate the Fidelity which is $|\langle\psi(\overline{\theta}+t\overline{\delta})|\,|\psi(\theta)\rangle\,|^2$. We see that $\frac{d|\psi(\overline{\theta})\rangle}{dt}|_{t=0} = -\iota(\sum_{j=1}^{k}\delta_j\Pi_{i=1}^{j-1}e^{-\iota\theta_i H_i}H_j\pi_{p=j}^{k}e^{-\iota\theta_p H_p})\,|\psi\rangle$. From that we can derive

$$\begin{aligned} \frac{d^2\,|\psi(\overline{\theta})\rangle}{dt^2}|_{t=0} &= -(\sum_{j=1}^{K}\delta_j^2(\Pi_{m=1}^{j-1}e^{-\iota\theta_m H_m})H_j^2(\Pi_{p=j}^{K}e^{-\iota\theta_p H_p}) + \\ (\sum_{1\leq i\leq j\leq K} 2\delta_i\delta_j(\Pi_{m=1}^{i-1}e^{-\iota\theta_m H_m})H_i(\Pi_{n=i}^{j-1}e^{-\iota\theta_n H_n})H_j(\Pi_{p=j}^{K}e^{-\iota\theta_n H_n}))\,|\psi\rangle & \end{aligned} \tag{15}$$

From the above equations we can obtain $\langle\psi(\overline{\theta})|\,\frac{d|\psi(\overline{\theta})\rangle}{dt}|_{t=0}$ and $\langle\psi|\,(U(\overline{\theta}))^*\,\frac{d^2|\psi(\overline{\theta})\rangle}{dt^2}|_{t=0}$. Similar expressions are obtained for higher order derivatives. However, we omit the explicit calculation as it is extremely cumbersome and lengthy. It is a matter of trivial calculation to see using standard matrix norm inequality that $|t\,\langle\psi(\overline{\theta})|\,\frac{d|\psi(\overline{\theta})\rangle}{dt}|_{t=0}| \leq \sum_{j=1}^{K}|t||\delta_j|||H_j||$. Let us denote the expression on the right i.e.$\sum_{j=1}^{K}|t||\delta_j|||H_j||$ as $\mathbf{H}$ Similarly, $|\frac{t^2}{2!}\,\langle\psi(\overline{\theta})|\,\frac{d^2|\psi(\overline{\theta})\rangle}{dt^2}|_{t=0}| \leq \frac{1}{2!}\mathbf{H}^2$. Using induction, one can check that $|\frac{t^r}{r!}\,\langle\psi(\overline{\theta})|\,\frac{d^r|\psi(\overline{\theta})\rangle}{dt^r}|_{t=0}| \leq \frac{1}{r!}\mathbf{H}^r \forall r \in \mathbb{N}$. If $\mathbf{H}\leq 1$. Then clearly we get that

$$\begin{aligned} \mathcal{F} \approx |1 + t\,\langle\psi(\overline{\theta})|\,\frac{d\,|\psi(\overline{\theta})\rangle}{dt}|_{t=0} + \frac{t^2}{2!}\,\langle\psi(\overline{\theta})|\,\frac{d^2\,|\psi(\overline{\theta})\rangle}{dt^2}|_{t=0}|^2 \geq ||1 + \langle\psi|\,(\sum_{j=1}^{K}\delta_j(\Pi_{m=j}^{K}e^{-\iota\theta_m^2 H_m})\,|\psi\rangle\,t\frac{d\,|\psi(\overline{\theta})\rangle}{dt}|_{t=0}| - \\ |\langle\psi|\,(\sum_{j=1}^{K}\delta_j(\Pi_{m=j}^{K}e^{-\iota\theta_m^2 H_m})\,|\psi\rangle\,\frac{t^2}{2!}\frac{d\,|\psi(\overline{\theta})\rangle}{dt^2}|_{t=0}|| \geq ||1 + t\,\langle\psi|\,(\sum_{j=1}^{K}\delta_j(\Pi_{m=j}^{K}e^{-\iota\theta_m^2 H_m})\,|\psi\rangle\,\frac{d\,|\psi(\overline{\theta})\rangle}{dt}|_{t=0}| - \frac{\mathbf{H}^2}{2!}| \end{aligned} \tag{16}$$

Now, $\langle\psi|\,(U(\overline{\theta}))^* t\frac{d|\psi(\overline{\theta})\rangle}{dt}|_{t=0} = -\iota\,\langle\psi|\,(\sum_{j=1}^{K}\delta_j(\Pi_{m=j}^{K}e^{-\iota\theta_m^2 H_m})\,|\psi\rangle$ is purely imaginary since the expression $\langle\psi|\,(\sum_{j=1}^{K}\delta_j(\Pi_{m=j}^{K}e^{-\iota\theta_m^2 H_m})\,|\psi\rangle$ is real. Hence we get $|1 + t\frac{d|\psi(\overline{\theta})\rangle}{dt}|_{t=0}| \geq |1 - |\frac{\mathbf{H}^2}{2}||^2$.

For the second case i.e. if $\mathbf{H} > 1$, then we get an universal bound such that we notice that $\mathbf{H}\geq 1$ Thus $|\mathbf{H}^r| \geq |\mathbf{H}^{r-1}|$. $|\langle\psi(\overline{\theta})|\,|\psi(\overline{\theta}+t\overline{\delta})\rangle\,|^2 = |\sum_{r=0}^{\infty}(-\iota)^r\frac{t^r}{r!}\,\langle\psi(\overline{\theta})|\,\frac{d^r|\psi(\overline{\theta})\rangle}{dt^r}|_{t=0}|$. It is trivial to see from the above derived inequalities that the expression is greater then $(e^{-\mathbf{H}}/\mathbf{H})^2$ and hence the result follows.

Theorem XIII.2 *Let $|\psi(\overline{\theta})\rangle$represent a pure quantum state where $|\psi(\overline{\theta})\rangle = U(\overline{\theta})\,|\psi\rangle = \Pi_{j=1}^{k} e^{-\iota H_k}\,|\psi\rangle$ where $H_k = \sum_{j=1}^{r} c_{kj}H_{kj}$ are known Hermitian matrices and $|\psi\rangle$ is an unit vector . Provided a fixed direction vector $\overline{\delta}$ and a small perturbation t, the Fidelity between the quantum states*

$$|\langle\psi(\overline{\theta}+t\overline{\delta})|\,|\psi(\overline{\theta})\rangle\,|^2=\begin{cases}\geq |1-|\mathbf{H}|-|\mathbf{H}|^2/2|^2 \ \textit{if}\ \mathbf{H}<1\\ \geq |e^{-\mathbf{H}}/\mathbf{H}|^2 \ \textit{if}\ \mathbf{H}\geq 1\end{cases}$$

where $\mathbf{H}=\sum_{i=1}^{k}(\sum_{j=1}^{r}|t\delta_{ij}|||H_{ij}||)e^{2||H_i||}$ *and* $||.||$ *is a matrix norm.*

Proof: The proof follows similar to the previous theorem. However, since the H_i's are not separated and are written as sum of local Hamiltonians, the previous derivative of matrix exponential doesn't work in this case. For this case we use the genealized derivative of matrix exponential which says that

$$\frac{d}{dt}e^{X(t)}=\int_0^1 e^{\alpha X(t)}(\frac{d}{dt}X(t))e^{(1-\alpha)X(t)}d\alpha$$

Since expressions of each derivative gets even more cumbersome as the order of derivative becomes higher, we shall explicitly write the expressions of the first order derivatives left multiplied by the inverse of the parametrized circuit because we need to calculate the Fidelity which is $|\langle\psi(\overline{\theta}+t\overline{\delta})|\,|\psi(\theta)\rangle\,|^2$.

Using the generalized derivative and perturbing the system, we see that

$$\frac{d\,|\psi(\overline{\theta})\rangle}{dt}|_{t=0}=(\sum_{j=1}^{k}(\Pi_{i=1}^{p-1}(\sum_{n=0}^{\infty}\frac{1}{n!}(\sum_{m=0}^{n}\binom{n}{m}\beta(m+1,n-m+1)(-\iota)^{n+1}H_p^m\triangle_p H_p^{n-m}))\Pi_{l=p+1}^{k})\,|\psi\rangle \tag{17}$$

where β is the beta function of first kind and $\triangle_p$ is $\frac{d}{dt}H_p(\overline{c}+t\overline{\delta})|_{t=0}$i.e. $\delta_p=\sum_{j=1}^{r}\delta_{pj}H_{pj}$. The second order derivative follows similarly, however, the expression becomes too complicated. However, from the above equations we can obtain

$\langle\psi(\overline{\theta})|\,\frac{d|\psi(\overline{\theta})\rangle}{dt}|_{t=0}$ and Similar expressions are obtained for higher order derivatives. However, we omit the explicit calculation as it is extremely cumbersome and lengthy. It is a matter of trivial calculation to see using standard matrix norm inequality that

$|t\,\langle\psi(\overline{\theta})|\,\frac{d|\psi(\overline{\theta})\rangle}{dt}|_{t=0}|\leq\sum_{i=1}^{k}(\sum_{j=1}^{r}|t\delta_{ij}|||H_{ij}||)e^{2||H_i||}$. Let us denote the expression on the right i.e.$\sum_{i=1}^{k}(\sum_{j=1}^{r}|t\delta_{ij}|||H_{ij}||)e^{2||H_i||}$ as $\mathbf{H}$ Similarly, $|\frac{t^2}{2!}\,\langle\psi(\overline{\theta})|\,\frac{d^2|\psi(\overline{\theta})\rangle}{dt^2}|_{t=0}|\leq\frac{1}{2!}\mathbf{H}^2$. Using induction, one can check that $|\frac{t^r}{r!}\,\langle\psi(\overline{\theta})|\,\frac{d^r|\psi(\overline{\theta})\rangle}{dt^r}|_{t=0}|\leq\frac{1}{r!}\mathbf{H}^r\forall r\in\mathbb{N}$. If $\mathbf{H}\leq 1$. Then clearly we get that

$$\begin{aligned}\mathcal{F}\approx|1+t\,\langle\psi(\overline{\theta})|\,\frac{d\,|\psi(\overline{\theta})\rangle}{dt}|_{t=0}+\frac{t^2}{2!}\,\langle\psi(\overline{\theta})|\,\frac{d^2\,|\psi(\overline{\theta})\rangle}{dt^2}|_{t=0}|^2&\geq||1+\langle\psi|\,(\sum_{j=1}^{K}\delta_j(\Pi_{m=j}^{K}e^{-\iota\theta_m^2H_m})\,|\psi\rangle\,t\frac{d\,|\psi(\overline{\theta})\rangle}{dt}|_{t=0}|-\\ |\langle\psi|\,(\sum_{j=1}^{K}\delta_j(\Pi_{m=j}^{K}e^{-\iota\theta_m^2H_m})\,|\psi\rangle\,\frac{t^2}{2!}\frac{d\,|\psi(\overline{\theta})\rangle}{dt^2}|_{t=0}||&\geq||1+t\,\langle\psi|\,(\sum_{j=1}^{K}\delta_j(\Pi_{m=j}^{K}e^{-\iota\theta_m^2H_m})\,|\psi\rangle\,\frac{d\,|\psi(\overline{\theta})\rangle}{dt}|_{t=0}|-\frac{\mathbf{H}^2}{2!}|\\ &\geq||1-|t\,\langle\psi|\,(\sum_{j=1}^{K}\delta_j(\Pi_{m=j}^{K}e^{-\iota\theta_m^2H_m})\,|\psi\rangle\,|-\frac{1}{2!}\mathbf{H}^2||\\ &\geq||1-|\mathbf{H}||-\frac{1}{2!}\mathbf{H}^2||\end{aligned} \tag{18}$$

Rest of the cases and arguments follows exactly similar to the previous theorem.

XIV. CONCLUSION

Here in this article we explore the Quantum Spin glass problem with the help of Variational Quantum Eigensolver. Further we observe the properties of various problem inspired ansatz. We show that even the best structured ansatz are not capable of reaching the true ground state of the spin glass Hamiltonians of different system size. As a reason we show the cost landscape of each of the ansatz with various *depth*. The expressibility of those ansatz is the main reason behind this as none of them has vanishing gradient as the system size grows even with random initialization.

XV. FUTURE PROSPECTS

In future as a continuation of this work we would love to show, analytically, why the expressibility matters so much in some cases to have a more deeper understanding of such optimization problems where expressiveness play a crucial role. We would likely explore the concept of Quantum Fisher Information (QFI) in this context as QFI, in recent days, has emerged as a new frontier to explore the power of VQE.

Try to plot BP for other parameters instead of only the last one and see if any other case shows BP or not. Also try to plot with more depth like in [6] and also try to take initial parameters in such a way that overall operation becomes identity just as showed in [6].

XVI. ACKNOWLEDGEMENTS

I am extremely grateful to Prof. Bibhas Adhikari from the dept. of Mathematics, IIT Kharagpur, for giving me the opportunity to do my Master's Thesis on Quantum computing under his guidance. I would also like to thanks Prof. Leong Chuan Kwek from Centre for Quantum Technologies(CQT), NUS, Singapore, for his constant support towards my thesis throughout the semester. This work would have been incomplete without their invaluable insights and support.

XVII. REFERENCES

[**1**]. A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.- Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, "A variational eigenvalue solver on a photonic quantum processor," Nature Communications 5, 4213 (2014).

[**2**].Cerezo, M., Arrasmith, A., Babbush, R. et al. Variational quantum algorithms. Nat Rev Phys 3, 625–644 (2021). https://doi.org/10.1038/s42254-021-00348-9

[**3**].Alexandre Choquette, Agustin Di Paolo *et al.*Choquette, A., Di Paolo, A., Barkoutsos, P. K., Sénéchal, D., Tavernelli, I., & Blais, A. (2021). Quantum-optimal-control-inspired ansatz for variational quantum algorithms. Physical Review Research, 3(2), 023092.

[**4**].Venturelli, D., Mandrà, S., Knysh, S., O'Gorman, B., Biswas, R., & Smelyanskiy, V. (2015). Quantum optimization of fully connected spin glasses. Physical Review X, 5(3), 031040.

[**5**].Larocca, M., Ju, N., García-Martín, D., Coles, P. J., & Cerezo, M. (2021). Theory of overparametrization in quantum neural networks. arXiv preprint arXiv:2109.11676.

[**6**].Wiersema, R., Zhou, C., de Sereville, Y., Carrasquilla, J. F., Kim, Y. B., & Yuen, H. (2020). Exploring entanglement and optimization within the hamiltonian variational ansatz. PRX Quantum, 1(2), 020319.

[**7**].Liu, J. G., Zhang, Y. H., Wan, Y., & Wang, L. (2019). Variational quantum eigensolver with fewer qubits. Physical Review Research, 1(2), 023025.

[**8**].Yang, Z. C., Chamon, C., Hamma, A., & Mucciolo, E. R. (2015). Two-component structure in the entanglement spectrum of highly excited states. Physical review letters, 115(26), 267206.

[**9**].Haug, T., Bharti, K., & Kim, M. S. (2021). Capacity and quantum geometry of parametrized quantum circuits. arXiv preprint arXiv:2102.01659.

[**10**].Sim, S., Johnson, P. D., & Aspuru-Guzik, A. (2019). Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms. Advanced Quantum Technologies, 2(12), 1900070.

[**11**].McClean, J. R., Boixo, S., Smelyanskiy, V. N., Babbush, R., & Neven, H. (2018). Barren plateaus in quantum neural network training landscapes. Nature communications, 9(1), 1-6.

[**12**].Venturelli, D., Mandrà, S., Knysh, S., O'Gorman, B., Biswas, R., & Smelyanskiy, V. (2015). Quantum optimization of fully connected spin glasses. Physical Review X, 5(3), 031040.

[**13**].Vatan, F., & Williams, C. (2004). Optimal quantum circuits for general two-qubit gates. Physical Review A, 69(3), 032315.

[**14**].Kardashin, A., Pervishko, A., Biamonte, J., & Yudin, D. (2021). Numerical hardware-efficient variational quantum simulation of a soliton solution. Physical Review A, 104(2), L020402.