

# Q-MAML: Quantum Model-Agnostic Meta-Learning for Variational Quantum Algorithms

Junyong Lee<sup>1\*</sup>, Jeihee Cho<sup>2\*</sup>, Shiho Kim<sup>2</sup>

<sup>1</sup> BK21 Graduate Program in Intelligent Semiconductor Technology, Yonsei University, Korea

<sup>2</sup> Yonsei University, Korea

jjunilee@yonsei.ac.kr, jeiheec@gmail.com, shiho@yonsei.ac.kr

## Abstract

In the Noisy Intermediate-Scale Quantum (NISQ) era, using variational quantum algorithms (VQAs) to solve optimization problems has become a key application. However, these algorithms face significant challenges, such as choosing an effective initial set of parameters and the limited quantum processing time that restricts the number of optimization iterations. In this study, we introduce a new framework for optimizing parameterized quantum circuits (PQCs) that employs a classical optimizer, inspired by Model-Agnostic Meta-Learning (MAML) technique. This approach aims to achieve better parameter initialization that ensures fast convergence. Our framework features a classical neural network, called *Learner*, which interacts with a PQC using the output of *Learner* as an initial parameter. During the pre-training phase, *Learner* is trained with a meta-objective based on the quantum circuit cost function. In the adaptation phase, the framework requires only a few PQC updates to converge to a more accurate value, while the learner remains unchanged. This method is highly adaptable and is effectively extended to various Hamiltonian optimization problems. We validate our approach through experiments, including distribution function mapping and optimization of the Heisenberg XYZ Hamiltonian. The result implies that the *Learner* successfully estimates initial parameters that generalize across the problem space, enabling fast adaptation.

## Introduction

In the past decade, quantum computing has seen remarkable advancements, leading to the development of variational quantum algorithms (VQAs) that utilize parameterized quantum circuits (PQC), where gradient-based optimization methods are applied to gain optimal parameters (Cerezo et al. 2021). The popularity of this paradigm is due to its flexibility, widening the area of quantum applications and approaches (Havlíček et al. 2019; Huang et al. 2021). However, these methods face significant challenges due to the unfavorable characteristics of the optimization landscape.

One example is the Barren Plateau (BP), which occurs when the gradient of the cost function becomes exponentially small, caused by a flat optimization landscape (Mc-

\*These authors contributed equally.

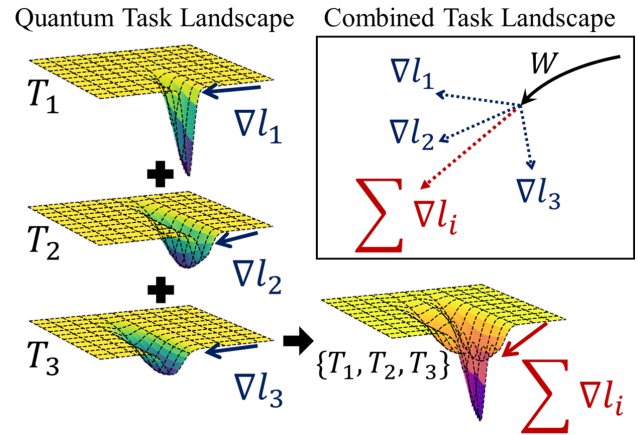


Figure 1: Illustration of the multiple individual optimization landscapes ( $\nabla l_1, \nabla l_2, \nabla l_3$ ) combine to form a single landscape ( $\sum \nabla l_i$ ). The model trained using the combined landscape can capture the general properties of various tasks, which results in the improvement of convergence in the optimization as in the MAML.

Clean et al. 2018; Wang et al. 2021; Holmes et al. 2022). Lack of meaningful gradient value leads to failure of the learning, resulting in slow or non-convergence of the optimization process. BP is especially problematic as it scales unfavorably with the size of the quantum system, making it a crucial problem to solve for future advancement and usage of quantum computing (Ragone et al. 2023).

Even with techniques designed to overcome the challenges posed by the optimization landscape, the iterative nature of quantum algorithms demands significant quantum computing resources. This makes the optimization process both time-consuming and costly, which can also be a burden for near-term quantum hardware. Therefore, achieving rapid convergence becomes essential to minimize computational expenses. A key strategy to enhance the performance and efficiency of PQC is identifying good initial parameters (Friedrich and Maziero 2022). Properly selected initial parameters can help mitigate the BP, facilitate more effective optimization, and ultimately improve the algorithms' ability to find meaningful solutions.

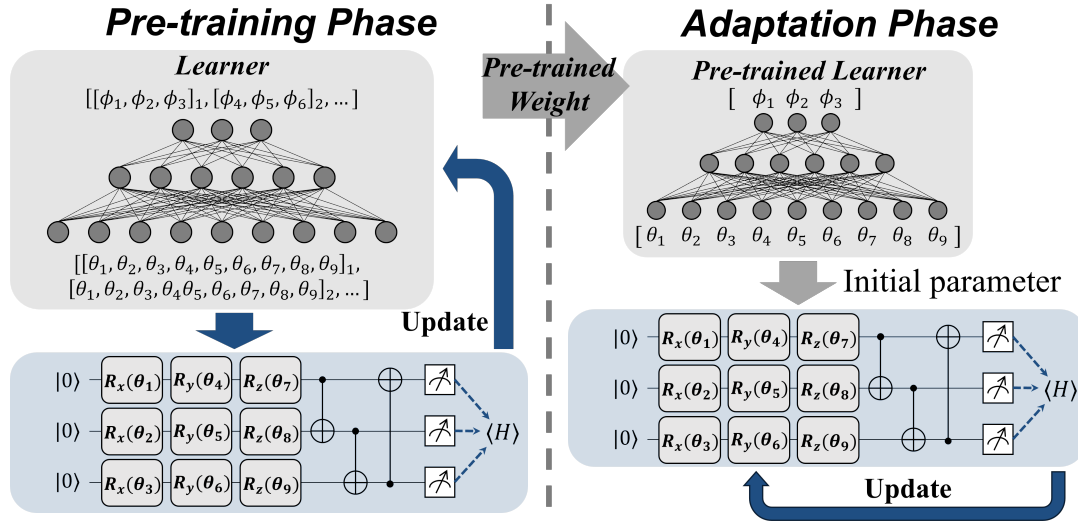


Figure 2: The overall learning path of Q-MAML. The pre-training phase includes an update of the *Learner* while the adaptation phase trains PQC on specific tasks with parameter initialized using pre-trained *Learner*.

Zhang et al. (2022) suggest a Gaussian distribution-based parameter sampling method that can increase the expectation of gradient norm to reduce the likelihood of encountering BP. Similarly, the reduced-domain parameter initialization technique shows promise in enhancing trainability (Wang et al. 2023). However, these approaches limit the circuit design and therefore can not completely resolve the problem in more complex circuits or diverse VQA tasks. Additionally, they often overlook the specific characteristics of the Hamiltonian governing the quantum system, which can result in suboptimal performance in certain scenarios.

To tackle the challenges, we propose Q-MAML, a quantum-classical hybrid framework that uses the classical neural network, *Learner*, to provide initial parameters for a PQC on optimization of various Hamiltonian. By leveraging Model-Agnostic Meta-Learning (MAML), *Learner* can predict helpful initial parameters on various tasks, and therefore enable fast convergence of the PQC. We empirically confirm that our approach not only addresses the challenges posed by barren plateaus but also paves the way for more efficient and scalable quantum optimization solutions.

## Related Works

**Variational Quantum Algorithms (VQA)** VQAs have emerged as a pivotal approach in leveraging near-term quantum devices to solve complex optimization problems (Cerezo et al. 2021). These algorithms combine PQCs with classical optimization techniques, making them versatile for a wide range of applications, including quantum chemistry, machine learning, and combinatorial optimization (Lubasch et al. 2020; Beckey et al. 2022; Amaro et al. 2022; Jaksch et al. 2023). VQAs are particularly appealing in that they are the leading candidate for achieving a quantum advantage in the NISQ era.

One of the most promising VQAs is Variational Quantum

Eigensolvers (VQEs), which are designed for solving quantum chemistry and material science problems (Peruzzo et al. 2014). The goal of VQE is to find the ground state energy of a Hamiltonian  $H$  by minimizing the cost function, defined as follows:

$$E(\theta) = \langle \psi(\theta) | H | \psi(\theta) \rangle \quad (1)$$

where  $\psi$  is quantum state and  $\theta$  is parameter of PQC.

Despite their potential, VQAs face significant challenges, particularly in terms of trainability and scalability. One of the primary issues is the BP phenomenon (McClean et al. 2018), where the gradient of the cost function becomes exponentially small, hindering the optimization process. These limitations underscore the importance of developing more robust initialization techniques that incorporate Hamiltonian-specific information, as our research proposes. By aligning initial parameters with the intrinsic properties of the Hamiltonian, our approach aims to enhance the trainability and convergence of VQAs, offering a more effective solution to the challenges faced by current methods.

There are several approaches to enhance trainability with proper initialization of circuits (Zhang et al. 2022; Cervera-Lierta, Kottmann, and Aspuru-Guzik 2021), however, they require specific circuit architecture or lack of explanation of generalization.

**Model-Agnostic Meta-Learning (MAML)** MAML is a highly influential approach in meta-learning, designed to allow models to adapt to new tasks with minimal data. Introduced by Finn, Abbeel, and Levine (2017), MAML operates by optimizing a model’s parameters such that only a few gradient updates are required for fast adaptation to new tasks. This makes MAML particularly effective for few-shot learning, where the goal is to generalize from limited examples.

MAML’s versatility lies in its model-agnostic characteristic, meaning it can be applied across a wide range of tasks, including supervised learning, reinforcement learning, and

---

**Algorithm 1: MAML for Pre-training Phase**

---

**Parameter**  $\alpha$ : learning rate  
**Require:**  $p(\mathcal{T})$ : distribution of tasks  
**Require:** classical optimizer  $\mathcal{O}$ , cost function  $l_T(\theta)$

- 1: randomly initialize  $W$
- 2: **while** not done **do**
- 3:   Sample batch of task  $T_i \sim p(\mathcal{T})$
- 4:   **for all**  $T_i$  **do**
- 5:     Get  $\theta$  from *Learner* :  $\theta = h_W(\phi_i)$
- 6:     Prepare quantum state  $|\psi(\theta)\rangle = g(\theta) |0\rangle$
- 7:     Evaluate cost function  $l_{T_i}(\theta) = \langle \psi(\theta_k) | \hat{H} | \psi(\theta_k) \rangle$
- 8:     Calculate the gradient  $\nabla_W l_{T_i}(g(h_W(\phi_i)))$
- 9:   **end for**
- 10:   Update parameter  
     $W \leftarrow W + \alpha \nabla_W \sum_{T_i} l_{T_i}(g(h_W(\phi)))$
- 11: **end while**

---

regression. By focusing on creating a universal initialization that performs well across different tasks, MAML has become a cornerstone in meta-learning research, influencing a broad range of subsequent studies and applications (Wang et al. 2022).

Inspired by the advantages of MAML, we aim to provide a universal initialization point for various VQE tasks. In VQE, the Hamiltonian is considered a single task, and the task space is defined as the space of optimization in the D-dimensional parameter space, where D is the total number of parameters being optimized. We apply the MAML method to train the *Learner* by constructing this task space with different Hamiltonians. With Q-MAML, the *Learner* gains the ability to guess good initial parameters for optimizing various Hamiltonians.

## Method

In this section, we illustrate detailed procedures of Q-MAML that consist of two key phases: the pre-training phase and the adaptation phase.

In the pre-training phase, *Learner* explores the combined task space (see Figure 1) to learn a robust initialization point of a quantum circuit that can quickly adapt to various quantum tasks. Similar to MAML, the *Learner* is generalized across different Hamiltonian tasks, capturing essential features and common patterns within this domain. In the adaptation phase, a specific Hamiltonian is selected and the PQC is fine-tuned to minimize the cost/loss function with a classical optimizer. During this stage, the PQC refines its weights to optimize cost on the target task as in VQE.

This dual-phase approach allows faster convergence and better optimization under the constraints of limited quantum processing time. The overview of Q-MAML is provided in Figure 2.

### Pre-training Phase

The pre-training phase focuses on determining an effective initial set of parameters for the PQC, denoted as  $\theta$ . During this phase, a classical neural network (*Learner*) is employed

---

**Algorithm 2: Adaptation to Single Task**

---

**Input:**  $T^*$ : Target Task (Hamiltonian)  
**Parameter:**  $\alpha$ : learning rate  
**Require:** Parameterized Quantum Circuit  $g(\theta)$   
**Require:** classical optimizer  $\mathcal{O}$ , cost function  $l_{T^*}(\theta)$   
**Require:**  $h_{W^*}$ : pre-trained *Learner*

- 1: Get initial parameters  $\theta = h_{W^*}(\psi)$
- 2: **while** not done **do**
- 3:   Prepare quantum state  $|\psi(\theta)\rangle = g(\theta) |0\rangle$
- 4:   Evaluate cost function  $l_{T^*}(\theta_k) = \langle \psi(\theta_k) | \hat{H} | \psi(\theta_k) \rangle$
- 5:   Calculate the gradient  $\nabla_{\theta} l_{T^*}(\theta_k)$
- 6:   Update parameter  
     $\theta \leftarrow \theta + \alpha \nabla_{\theta} l_{T^*}(\theta_k)$
- 7: **end while**

**Output** Final optimization  $\theta^*$ , minimal cost  $l_{T^*}(\theta^*)$

---

to approximate the initial parameters by minimizing a meta-objective function based on the quantum circuit’s cost function (Equation 2). The mathematical background of Hamiltonian task space is provided in the Appendix. The detailed algorithm is provided in Algorithm 1.

The *Learner*’s input is the Hamiltonian task vector ( $\phi$ ), and the *Learner*’s output is assigned as a parameter of the PQC ( $\theta$ ), and the observable of the PQC ( $\langle \mathcal{O} \rangle$ ) is used for the cost calculation. In this structure, the Meta-Learning Objective is as follows:

$$\underset{W}{\operatorname{argmin}} \sum_{T_i \sim p(\mathcal{T})} l_{T_i}(g(h_W(\phi_i))) \quad (2)$$

where  $l_{T_i}$  is the cost function for task  $T_i$ . Given the known fact that the combination of cost functions is feasible in PQC training (Nemkov, Kiktenko, and Fedorov 2024), we have taken advantage of this by differentiating the *Learner*’s weight concerning the combined cost function, instead of computing and directly updating  $\theta$  using its Hessian as in MAML. This approach allows for more efficient and scalable optimization, aligning the parameter updates with the underlying quantum task characteristics. The *Learner*’s weight is trained via backpropagation of the gradients, where the chain rule is still applicable. With generated observable  $\langle \mathcal{O} \rangle$ , the gradient to update  $W$  can be calculated as follows:

$$\frac{dl_T}{dW} = \frac{dl_T}{d\langle \mathcal{O} \rangle} \frac{d\langle \mathcal{O} \rangle}{d\theta} \frac{d\theta}{dW} \quad (3)$$

where  $\frac{dl_T}{d\langle \mathcal{O} \rangle}$  is differentiable as cost functions are typically defined to be differentiable. Additionally,  $\frac{d\langle \mathcal{O} \rangle}{d\theta}$  is differentiable as is confirmed through training in VQA and Nemkov, Kiktenko, and Fedorov (2024), and  $\frac{d\theta}{dW}$  represents standard differentiation in neural network training. The *Learner* is trained using a variety of Hamiltonians to ensure that the initial parameters it generates are generalizable and serve as a good starting point for different quantum tasks. This phase aims to reduce the burden on the quantum processor by providing a well-informed initialization, which reduces the number of iterations required during the adaptation phase.

## Adaptation Phase

Following the pre-training phase, the adaptation phase involves fine-tuning the PQC using the initial parameters  $\theta$  estimated by the *Learner*. Unlike traditional VQAs, where parameter optimization is performed from scratch, our method utilizes the pre-learned parameters as a starting point. Other components such as cost and classical optimizer are the same as traditional VQAs. In this phase, only the PQC is updated, and no further adjustments are made to the *Learner*. This focused adaptation ensures that the quantum circuit can adapt to the specific problem at hand with minimal quantum resource usage. Detailed algorithm is provided in Algorithm 2.

## Experiments and Results

We conduct experiments on two different applications: Heisenberg XYZ Hamiltonian, and Molecule Hamiltonian. To show the expandability of the framework, we perform additional application, and distribution function embedding and provide results in the Appendix.

Throughout the experiments, the *Learner* is trained for 30 epochs using Adam optimizer with a learning rate of 0.001. For the adaptation phase, PQC is trained for 2000 iterations by Adam optimizer with 0.001 as a learning rate. We use four different initialization methods that are commonly used: Zero initialization,  $\pi$  initialization, reduced-domain Uniform initialization (Wang et al. 2023), and Gaussian initialization (Zhang et al. 2022). Zero and  $\pi$  initialization method use zero and  $\pi$  for all parameters. Parameters of reduced-domain Uniform initialization is sampled from  $\mathcal{U}[-\alpha\pi, \alpha\pi]$  where  $\alpha$  is set as 0.05 to ensure sufficiently small values are sampled, and Gaussian initialization from  $\mathcal{N}(0, \gamma^2)$  and set  $\gamma^2 = \frac{1}{4S(L+2)}$  where  $S$  is the number of non-identity Pauli operators in observable and  $L$  is the number of layers in the circuit that follows the setting of Zhang et al. (2022).

### Heisenberg XYZ Hamiltonian

The Heisenberg XYZ Hamiltonian is a quantum mechanical model that describes interactions between spins on a lattice. It is an extension of the Heisenberg model, which is used to understand the magnetic properties of materials at the quantum level. The XYZ Hamiltonian specifically accounts for anisotropic interactions in three different spatial directions (X, Y, and Z), meaning that the interaction strengths along these directions can differ. Heisenberg XYZ Hamiltonian is defined via three parameters,  $J_x$ ,  $J_y$ , and  $J_z$  and can be obtained by calculating following equation:

$$\begin{aligned} \hat{H}_{XYZ} &= - \sum_{n=1}^{N-1} (J_x \sigma_n^1 \sigma_{n+1}^1 + J_y \sigma_n^2 \sigma_{n+1}^2 + J_z \sigma_n^3 \sigma_{n+1}^3 + h \sigma_n^3) \end{aligned} \quad (4)$$

where  $\sigma^1$ ,  $\sigma^2$ , and  $\sigma^3$  represent the Pauli X, Y, and Z matrices, respectively, and  $h$  is the effect of the external magnetic field where we set as 0.

**Task Space Definition** Heisenberg XYZ Task Space consists of vectors representing each Hamiltonian. As Hamiltonian varies by  $J_i$  (Equation 4), each task (Hamiltonian) can be defined by  $J^i = [J_x^i, J_y^i, J_z^i]$  where  $i$  indicates task number. We randomly sample three values within the range  $[-3, 3]$  with an interval of 0.1 to construct  $J^i$ .

**Experiment Setting** The *Learner* is constructed using one input layer with a size of 3, two hidden layers with 256 neurons, and an output layer with a size of the number of parameters in PQC. We vary the number of qubits by 12 and 20 to check the performance of Q-MAML on a small and large number of qubits. Experiments using 10, 14, 16, and 18 qubits are provided in the Appendix along with the design of the ansatz. The ansatz of PQC consists of blocks that have IsingXX, IsingYY, and IsingZZ gates. We randomly select 16 samples to check the adaptation performance of Q-MAML.

### Molecule Hamiltonian

The Molecule Hamiltonian is the crucial point for solving the Schrödinger equation for a molecule, which yields important information about the molecule’s electronic structure, bond lengths, bond angles, and other physical and chemical properties. For  $M$  spin molecular orbital (MO), the basis is denoted as  $|n_0, n_1, \dots, n_{M-1}\rangle$ . Each basis can take 0 or 1 which indicates the electron occupancy of the corresponding orbital. It can be written as:

$$\hat{H} = \sum_{p,q} h_{p,q} c_p^\dagger c_q + \frac{1}{2} \sum_{p,q,r,s} c_p^\dagger c_q^\dagger c_r c_s, \quad (5)$$

$c^\dagger$  and  $c$  are the electron creation and annihilation operators, respectively, while the coefficients  $h_{pq}$  and  $h_{pqrs}$  correspond to the one and two-electron Coulomb integrals, estimated by the Hartree-Fock orbitals. Using  $M$  qubits to encode the  $M$  numbers of basis. Through the Jordan-Wigner transformation, fermionic operators are mapped onto qubit operators, expressed as a tensor product of Pauli matrices  $I, X, Y, Z$  action on the qubits.

$$\hat{H} = \sum_j C_j \otimes_i \sigma_i^{(j)} \quad (6)$$

where  $\sigma \in I, X, Y, Z$ .

**Task Space Definition** Molecule Hamiltonian Task Space is constructed with the vector  $C$ , which is the coefficient of the molecule Hamiltonian. It should be noted that each task can have a different length.

**Experiment Setting** We apply the same *Learner* architecture as in Heisenberg XYZ Hamiltonian. The only difference is the input size of the *Learner*, where the max length of  $C$  is used in this experiment. If  $C$  is shorter than the max length, the remaining values are zero-padded. The PQC consists of 7 layers of StronglyEntanglingLayer (Schuld et al. 2020), with an example of the circuit provided in the Appendix. We randomly select 6 samples (which is 10% of the total dataset) to evaluate the performance of adaptation.

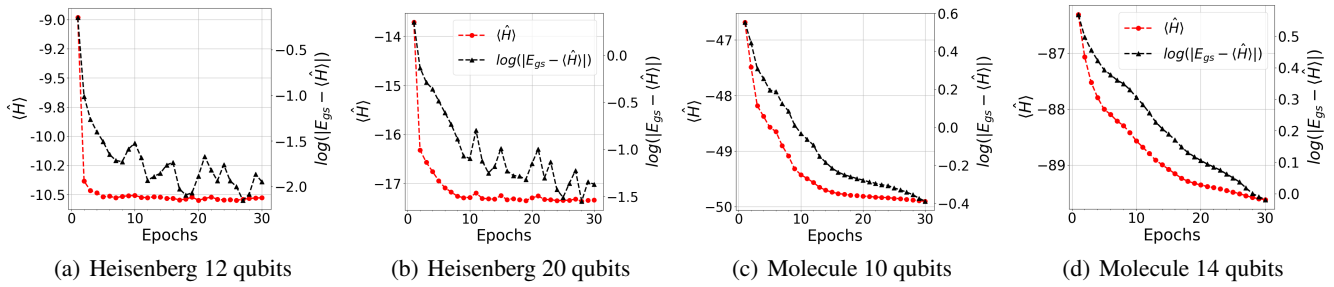


Figure 3: Training trajectory of *Learner* with different Hamiltonian dataset.

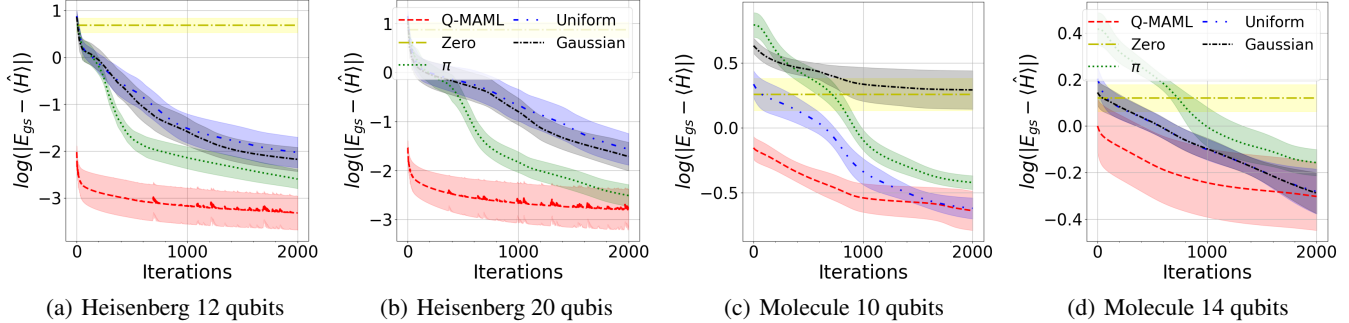


Figure 4: Training trajectory of PQC in the adaptation phase. The average gap value is plotted using 16 and 6 different Hamiltonian for Heisenberg and Molecule, respectively. The shaded regions indicate one standard deviation ( $[\mu - \sigma/2, \mu + \sigma/2]$ ).

**Dataset Generation** We use PennyLane function `qml.molecular_hamiltonian` to generate a Hamiltonian of molecules with specific conditions. We give 12 distinct bond lengths to generate different Hamiltonian as many as possible. As the number of qubits required for Hamiltonian is related to the number of active orbitals, we set it as 5 and 7 to generate 10 and 14 qubits Hamiltonian. Detailed parameters to generate the data are provided in the Appendix.

**Results** We first validate the learning of the *Learner* on the task space. In Figure 3, we can tell from the result that the *Learner* successfully learned the pattern of the task space and achieved proper estimation resulting in a small gap between the ground state energy concerning the dataset and the number of qubits.

With pre-trained *Learner*, the initial parameter for various tasks is generated for the adaptation phase. In Figure 4, we visualize the learning curve of PQC to assess its convergence speed and performance. The gap between ground state energy calculated using exact diagonalization and expectation of PQC is measured. It should be noted that, unlike training of the *Learner*, separate PQC is used and trained per task. The  $\gamma^2$  for Gaussian initialization is set as  $\frac{1}{176}$  where  $S = 2$  and  $L = 20$  as we use at most 20 layers for experiments using Heisenberg dataset, and  $\frac{1}{360}, \frac{1}{504}$  respectively for 10 and 14 qubit Molecule experiments. The result implies that when initialized with Q-MAML, PQC can converge faster and therefore requires lower quantum resources compared to other methods.

## Discussion

In this section, we delve into the key findings. Focusing on three primary of interest: the learning behavior of the *Learner*, its impact on gradient norm characteristic, and the barren plateau phenomenon. First, we examine what the *Learner* is actually learning during the pre-training phase, and how this knowledge translates into effective parameter initialization for the quantum circuit. Next, we analyze the influence of the *Learner*'s initialization on the gradient norms throughout the optimization process, shedding light on its role in improving convergence. Finally, we address the ongoing challenge of barren plateaus, discussing how our approach mitigates this issue and comparing our results with existing strategies in the field.

### What the *Learner* Discovers During Training?

First, we define the Task Space  $\mathcal{T}$  and hypothesize the relationship between this space and the quantum circuit parameter  $\theta$ . Q-MAML approach is designed to learn this relationship during the pre-training phase, aiming to produce better initial parameters for a given task. The key insight extracted from the following experiment is that the *Learner* learns to capture the distinctive patterns across the task space, and therefore effectively tailor the initialization for each Hamiltonian.

Figure 5 shows the statistical properties of the parameters initialized using Q-MAML, compared to Uniform and Gaussian initialization methods, across different Task Space and qubits configurations. In both figures, Q-MAML shows

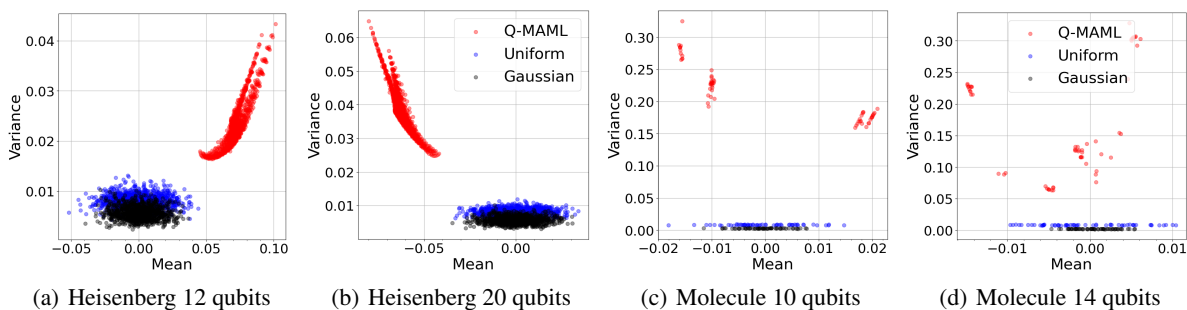


Figure 5: Statistics of the parameter initialized using Q-MAML, Uniform, and Gaussian.

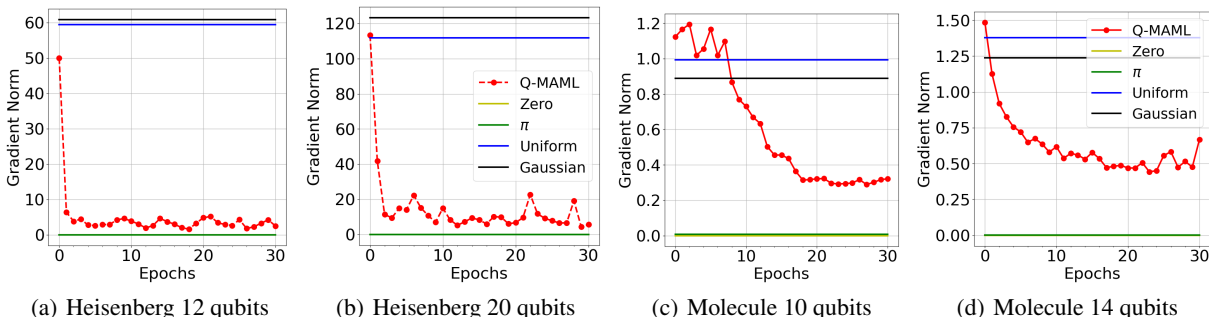


Figure 6: Trend of the PQC's gradient norm during pre-training phase

a broader distribution in terms of variance and mean. This indicates that the *Learner* is identifying and leveraging specific patterns within each Task Space to generate diverse and task-specific initializations. These patterns are not present in the Uniform or Gaussian methods, which result in more constrained parameter distributions.

By adapting to the specific characteristics of Task Space, the *Learner* not only provides task-specific initialization but also significantly enhances the optimization process. This is demonstrated in the pre-training and adaption trajectories (Figure 3, 4), where Q-MAML consistently exhibits superior starting points compared to other methods. As Figure 4 illustrates, Q-MAML's initialized parameters display a broader and more effective distribution, which directly contributes to more efficient convergence. The *Learner's* ability to identify and incorporate distinct patterns across different Task Spaces results in these advantageous starting positions, highlighting the effectiveness of our approach and its potential to address the challenges posed by barren plateaus in VQAs.

### How the *Learner* affect Gradient Norms?

The norm of the gradient is a widely recognized tool for assessing the convergence behavior of VQAs. In many studies, tracking the gradient norm provides valuable insights into the optimization process and helps identify the factors that influence convergence rates. In this section, we explore the fast adaptation observed with Q-MAML by analyzing how the gradient norm evolves during the pre-training process. By examining the train trends in the gradient norm in differ-

ent qubit configurations, as illustrated in Figure 6, we aim to understand the mechanisms that contribute to the enhanced convergence performance of our approach.

In classical neural networks, the norm of the gradient is a critical indicator of learning stability and convergence. A well-balanced gradient norm ensures stable learning, preventing issues such as "exploding gradients," where overly large gradients cause instability, or "vanishing gradients," where excessively small gradients hinder progress (Hochreiter et al. 2001; Glorot and Bengio 2010; Pascanu, Mikolov, and Bengio 2013). By maintaining a moderate gradient norm, the learning process can make consistent and meaningful updates to the model parameters, leading to efficient and reliable convergence. This concept underscores the importance of monitoring gradient norms to achieve optimal performance during the gradient-based learning process, which applies to both neural network and PQC training.

To better predict where our initial point will be and how the model will converge, we measure the average gradient norm at the starting point of each epoch across the entire dataset. This approach allows us to track how the initial gradients evolve and gives us insight into the expected trajectory of the optimization process. In addition, it is one of the important evidence to predict the overall stability of the VQA optimization.

In Figure 6, we depict the trends of the gradient norm in different configurations. Overall, the gradient norm of Q-MAML converges to a specific value, which is *moderate* compared to other approaches. Although this observation needs further discussion, we can assume that the *Learner*

provides better initial  $\theta$  for adaptation relative to other approaches with larger gradient norm. This implies that Q-MAML effectively mitigates issues such as exploding gradients, resulting in more efficient and reliable convergence against traditional initialization methods. An additional discussion on this topic is concluded in the following section, *A New Perspective on Barren Plateaus*.

### A New Perspective on Barren Plateaus

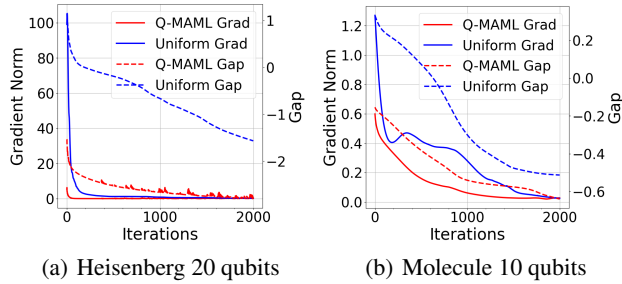


Figure 7: Fluctuation of gradient norm of PQC while training in the adaptation phase.

To further investigate the challenge of barren plateaus, we focus on the adaptation process and use gradient norms as a diagnostic tool. Examining the gradient norms throughout the adaptation process allows us to identify the effective regions of the learning curve where optimization is most effective. In addition, we analyze the statistical properties of these regions to assess the suitability of the Q-MAML initialization and adaptation strategy. This approach provides a deeper understanding of how our method navigates these landscapes, ultimately enhancing the efficiency and effectiveness of the optimization process.

During the adaptation process, the relative positions of each initial parameter can be empirically compared using the gradient norm and the gap between the expected value and ground state energy. Based on Figure 6 and 7, the parameters initialized by Q-MAML exhibit a moderate gradient norm and are positioned close to the optimal solution, allowing for quicker convergence. Additionally, the mean and variance of these parameters differ from those produced by traditional methods.

Building on these observations, we provide a conceptual figure of optimization landscape in Figure 8. Through observation, we can imply that the initialization point provided by Q-MAML is located in space with a moderate gradient norm, which is closer to the optimal value. On the contrary, slower convergence, and a larger gradient norm of Uniform initialization can be understood as located far from the optimal value.

Similar to the context of classical neural networks training, excessively large gradients can cause unstable updates to the parameters which leads to instability and slow convergence. Our findings suggest that the initialization points identified by Q-MAML reside in regions of the parameter space where gradient norms remain moderate, effectively mitigating the problem of gradients.

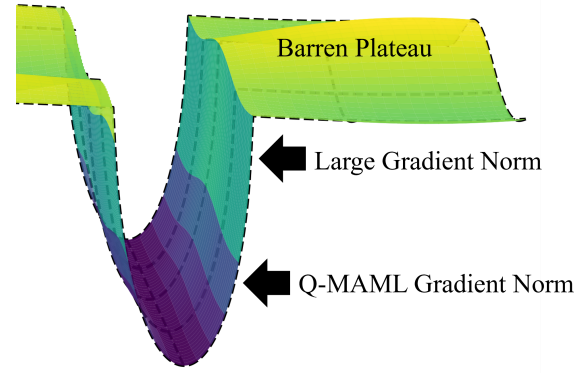


Figure 8: Visualization of optimization landscape in Q-MAML. The "Barren Plateau" region, is depicted in yellow. The green and blue regions indicate areas with "Large Gradient Norm" and "Q-MAML Gradient Norm".

### Conclusion

We introduced a quantum-classical hybrid framework for optimizing PQCs utilizing a MAML-inspired approach. Our methodology aims to improve the initial parameter of PQC to enhance convergence speed in the adaptation phase, which is critical for effective quantum computation in the NISQ era. Through rigorous experimentation, our results demonstrate that Q-MAML significantly improves the training dynamics of PQCs across various quantum task spaces, including Heisenberg XYZ Hamiltonian and Molecule Hamiltonian.

Our analyses reveal that Q-MAML effectively circumvents the common challenges posed by barren plateaus through superior initialization strategies, which are tailored to specific quantum tasks by learning from a broad task landscape. This approach not only facilitates a deeper understanding of the gradient norms but also allows us to pinpoint optimal learning regions, thus contributing to more efficient and precise quantum optimization.

Furthermore, the comparative study against traditional initialization methods like Uniform and Gaussian showed that Q-MAML's initialization leads to better positioning of the learning trajectory, enhancing the model's adaptability and overall performance. The implementation of this framework holds great promise for advancing the capabilities of quantum computing algorithms, particularly in solving complex optimization problems that are pivotal in the field of quantum mechanics.

Building on our findings, we raise the issue of "optimal gradients" in VQAs, a challenge similar to that faced in classical neural networks. Our future work will focus on refining the Q-MAML framework to incorporate more diverse quantum tasks and explore further applications in other quantum computing paradigms. We believe that the continued development of meta-learning strategies in quantum computing will significantly contribute to overcoming the current limitations of PQCs and pave the way for more robust, scalable quantum technologies.

## Acknowledgements

This research was supported by Quantum Computing based on Quantum Advantage challenge research(RS-2023-00257561) through the National Research Foundation of Korea(NRF) funded by the Korean government (Ministry of Science and ICT(MSIT)).

## References

- Amaro, D.; Rosenkranz, M.; Fitzpatrick, N.; Hirano, K.; and Fiorentini, M. 2022. A case study of variational quantum algorithms for a job shop scheduling problem. *EPJ Quantum Technology*, 9(1): 5.
- Beckey, J. L.; Cerezo, M.; Sone, A.; and Coles, P. J. 2022. Variational quantum algorithm for estimating the quantum fisher information. *Physical Review Research*, 4(1): 013083.
- Cerezo, M.; Arrasmith, A.; Babbush, R.; Benjamin, S. C.; Endo, S.; Fujii, K.; McClean, J. R.; Mitarai, K.; Yuan, X.; Cincio, L.; et al. 2021. Variational quantum algorithms. *Nature Reviews Physics*, 3(9): 625–644.
- Cervera-Lierta, A.; Kottmann, J. S.; and Aspuru-Guzik, A. 2021. Meta-Variational Quantum Eigensolver: Learning Energy Profiles of Parameterized Hamiltonians for Quantum Simulation. *PRX Quantum*, 2: 020329.
- Finn, C.; Abbeel, P.; and Levine, S. 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning*, 1126–1135. PMLR.
- Friedrich, L.; and Maziero, J. 2022. Avoiding barren plateaus with classical deep neural networks. *Physical Review A*, 106(4): 042433.
- Glorot, X.; and Bengio, Y. 2010. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 249–256. JMLR Workshop and Conference Proceedings.
- Havlíček, V.; Córcoles, A. D.; Temme, K.; Harrow, A. W.; Kandala, A.; Chow, J. M.; and Gambetta, J. M. 2019. Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747): 209–212.
- Hochreiter, S.; Bengio, Y.; Frasconi, P.; Schmidhuber, J.; et al. 2001. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies.
- Holmes, Z.; Sharma, K.; Cerezo, M.; and Coles, P. J. 2022. Connecting ansatz expressibility to gradient magnitudes and barren plateaus. *PRX Quantum*, 3(1): 010313.
- Huang, H.-L.; Du, Y.; Gong, M.; Zhao, Y.; Wu, Y.; Wang, C.; Li, S.; Liang, F.; Lin, J.; Xu, Y.; et al. 2021. Experimental quantum generative adversarial networks for image generation. *Physical Review Applied*, 16(2): 024051.
- Jaksch, D.; Givi, P.; Daley, A. J.; and Rung, T. 2023. Variational quantum algorithms for computational fluid dynamics. *AIAA journal*, 61(5): 1885–1894.
- Lubasch, M.; Joo, J.; Moinier, P.; Kiffner, M.; and Jaksch, D. 2020. Variational quantum algorithms for nonlinear problems. *Physical Review A*, 101(1): 010301.
- McClean, J. R.; Boixo, S.; Smelyanskiy, V. N.; Babbush, R.; and Neven, H. 2018. Barren plateaus in quantum neural network training landscapes. *Nature communications*, 9(1): 4812.
- Nemkov, N. A.; Kiktenko, E. O.; and Fedorov, A. K. 2024. Barren plateaus are swamped with traps. *arXiv preprint arXiv:2405.05332*.
- Pascanu, R.; Mikolov, T.; and Bengio, Y. 2013. On the difficulty of training recurrent neural networks. In *International conference on machine learning*, 1310–1318. Pmlr.
- Peruzzo, A.; McClean, J.; Shadbolt, P.; Yung, M.-H.; Zhou, X.-Q.; Love, P. J.; Aspuru-Guzik, A.; and O’Brien, J. L. 2014. A variational eigenvalue solver on a photonic quantum processor. *Nature communications*, 5(1): 4213.
- Ragone, M.; Bakalov, B. N.; Sauvage, F.; Kemper, A. F.; Marrero, C. O.; Larocca, M.; and Cerezo, M. 2023. A unified theory of barren plateaus for deep parametrized quantum circuits. *arXiv preprint arXiv:2309.09342*.
- Schuld, M.; Bocharov, A.; Svore, K. M.; and Wiebe, N. 2020. Circuit-centric quantum classifiers. *Physical Review A*, 101(3): 032308.
- Wang, H.; Wang, Y.; Sun, R.; and Li, B. 2022. Global convergence of maml and theory-inspired neural architecture search for few-shot learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9797–9808.
- Wang, S.; Fontana, E.; Cerezo, M.; Sharma, K.; Sone, A.; Cincio, L.; and Coles, P. J. 2021. Noise-induced barren plateaus in variational quantum algorithms. *Nature communications*, 12(1): 6961.
- Wang, Y.; Qi, B.; Ferrie, C.; and Dong, D. 2023. Trainability enhancement of parameterized quantum circuits via reduced-domain parameter initialization. *arXiv preprint arXiv:2302.06858*.
- Zhang, K.; Liu, L.; Hsieh, M.-H.; and Tao, D. 2022. Escaping from the barren plateau via gaussian initializations in deep variational quantum circuits. *Advances in Neural Information Processing Systems*, 35: 18612–18627.