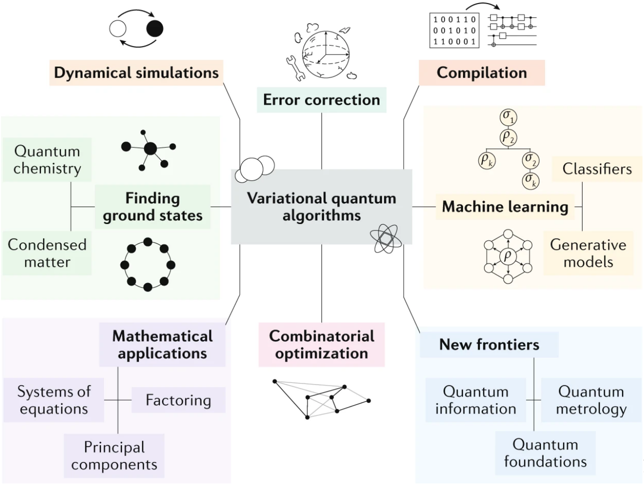
Variational quantum algorithms and Quantum Machine Learning

Cica Gustiani

21 June 2024

ICTP

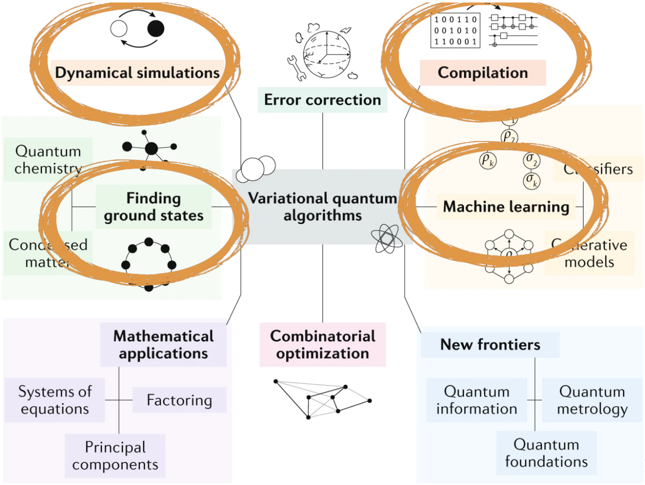
There are plenty of VQAs out there for various applications



[1]

[1] Cerezo, M., Arrasmith, A., Babbush, R. et al. Variational quantum algorithms. Nat Rev Phys 3, 625–644 (2021).

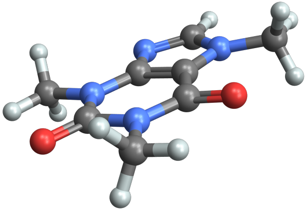
There are plenty of VQAs out there for various applications



[1]

[1] Cerezo, M., Arrasmith, A., Babbush, R. et al. Variational quantum algorithms. Nat Rev Phys 3, 625–644 (2021).

Simulating quantum systems on quantum computer



Static: ground-state

$$\hat{H} |\psi\rangle = E |\psi\rangle$$

Dynamics: the time-evolution operator

$$i\hbar \frac{d}{dt} |\psi(t)\rangle = \hat{H} |\psi(t)\rangle$$

$|\psi(t)\rangle = e^{-i\hat{H}t} |\psi(0)\rangle$, constant $\hat{H}$

Quantum Phase Estimation

Estimating the ground state energy (and excited states).

$H |\psi\rangle = E |\psi\rangle$

$$e^{-i\hat{H}\tau} |\psi\rangle = e^{-iE\tau} |\psi\rangle = e^{-i2\pi\varphi} |\psi\rangle$$

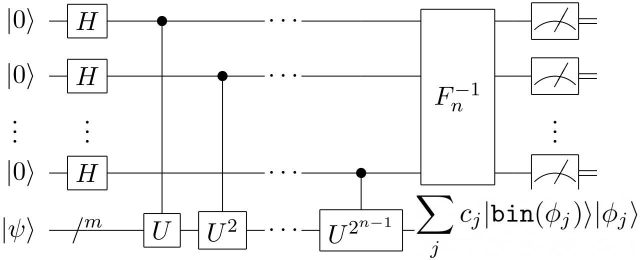
Quantum Phase Estimation

Estimating the ground state energy (and excited states).

$H|\psi\rangle = E|\psi\rangle$

$$e^{-i\hat{H}\tau}|\psi\rangle = e^{-iE\tau}|\psi\rangle = e^{-i2\pi\varphi}|\psi\rangle$$

$E = 2\pi\varphi/\tau$ and $U = e^{-iH\tau}$



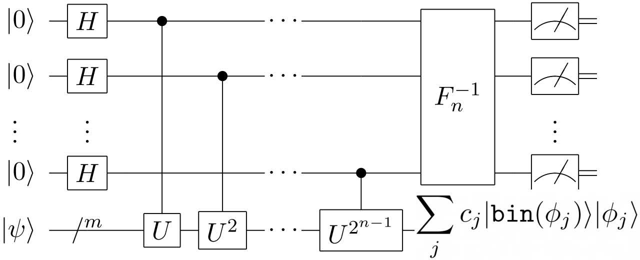
Quantum Phase Estimation

Estimating the ground state energy (and excited states).

$H|\psi\rangle = E|\psi\rangle$

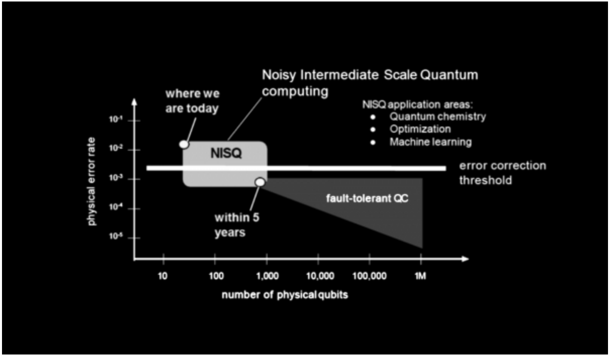
$$e^{-i\hat{H}\tau}|\psi\rangle = e^{-iE\tau}|\psi\rangle = e^{-i2\pi\varphi}|\psi\rangle$$

$E = 2\pi\varphi/\tau$ and $U = e^{-iH\tau}$



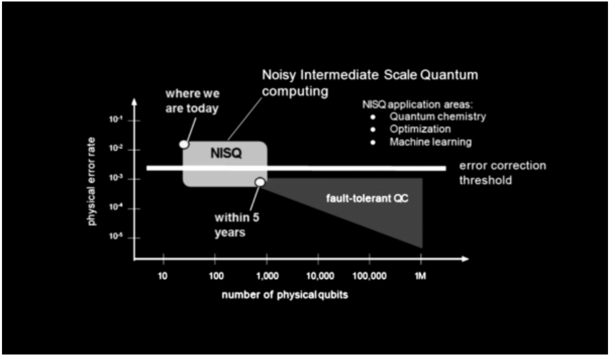
Grows by precision of φ and U.

Noisy Intermediate-Scale Quantum (NISQ) era



https://www.quandco.com/ja/blog/arxiv-1801-00862

Noisy Intermediate-Scale Quantum (NISQ) era



https://www.quandco.com/ja/blog/arxiv-1801-00862

Noise-resilient algorithms

Error mitigation

From variational method to VQE

The classic Rayleigh-Ritz method

$$E_0 = \min_{|\psi\rangle} \frac{\langle\psi| H |\psi\rangle}{\langle\psi|\psi\rangle}$$

test function $|\psi_t\rangle$

$$\frac{\langle\psi_t| H |\psi_t\rangle}{\langle\psi_t|\psi_t\rangle} \geq E_0, \quad \forall |\psi_t\rangle$$

From variational method to VQE

The classic Rayleigh-Ritz method

$$E_0 = \min_{|\psi\rangle} \frac{\langle\psi| H |\psi\rangle}{\langle\psi|\psi\rangle}$$

test function $|\psi_t\rangle$

$$\frac{\langle\psi_t| H |\psi_t\rangle}{\langle\psi_t|\psi_t\rangle} \geq E_0, \quad \forall |\psi_t\rangle$$

ansatz: $|\phi(\vec{\alpha})\rangle = \sum_i \alpha_i |\phi_i\rangle$

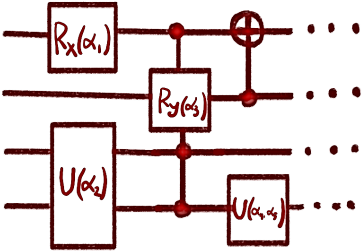
$$E_{est} = \min_{\vec{\alpha}} \frac{\langle\phi(\vec{\alpha})| H |\phi(\vec{\alpha})\rangle}{\langle\phi(\vec{\alpha})|\phi(\vec{\alpha})\rangle} = \min_{\vec{\alpha}} E(\vec{\alpha}) \geq E_0$$

classical optimisation over α via $\frac{\partial E(\vec{\alpha})}{\partial \alpha_j} = 0$

From variational method to VQE

The classic Rayleigh-Ritz method

test funct



$\frac{\psi\rangle}{\rangle}$

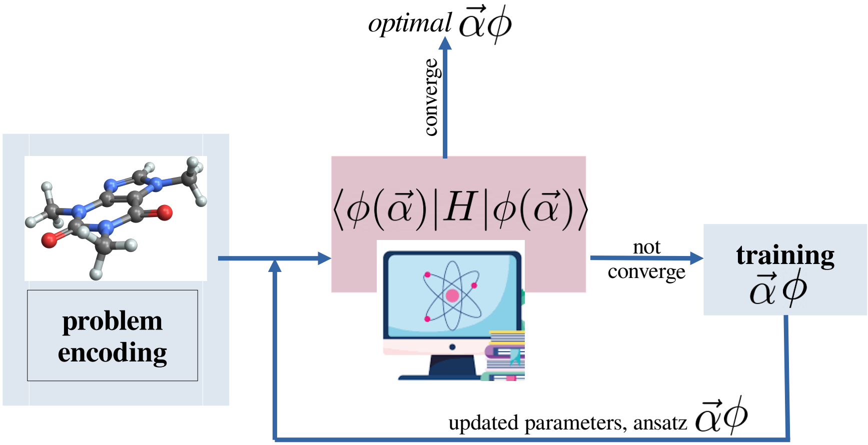
$\forall \, |\psi_t\rangle$

ansatz: $|\phi(\vec{\alpha})\rangle = \sum_i \alpha_i \, |\phi_i\rangle$

$$E_{est} = \min_{\vec{\alpha}} \frac{\langle\phi(\vec{\alpha})| \, H \, |\phi(\vec{\alpha})\rangle}{\langle\phi(\vec{\alpha})|\phi(\vec{\alpha})\rangle} = \min_{\vec{\alpha}} E(\vec{\alpha}) \geq E_0$$

classical optimisation over α via $\frac{\partial E(\vec{\alpha})}{\partial \alpha_j} = 0$

VQE framework



$$\min_{\vec{\alpha}} \langle\phi(\vec{\alpha})|\, H \,|\phi(\vec{\alpha})\rangle$$

Why VQE is attractive to NISQ-ers

Why VQE is attractive to NISQ-ers

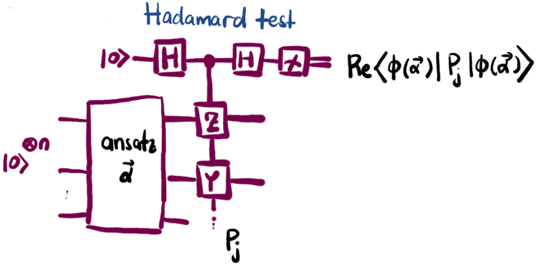
- Parameterised unitary (ansatz): $|\phi(\vec{\alpha})\rangle = \left(\prod_k g_k(\alpha_k)\right)|0\rangle^{\otimes n}$
- Hamiltonian as Pauli sums: $H = \sum_j c_j P_j, \quad$ Paulis P_j

Why VQE is attractive to NISQ-ers

- Parameterised unitary (ansatz): $|\phi(\vec{\alpha})\rangle = \left(\prod_k g_k(\alpha_k)\right)|0\rangle^{\otimes n}$
- Hamiltonian as Pauli sums: $H = \sum_j c_j P_j, \quad$ Paulis P_j
- By linearity: $\langle\phi(\vec{\alpha})| H |\phi(\vec{\alpha})\rangle = \sum_j c_j \langle\phi(\vec{\alpha})| P_j |\phi(\vec{\alpha})\rangle$parallel!

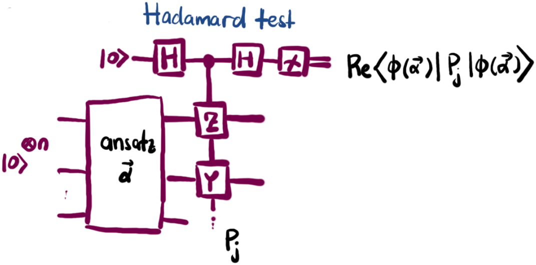
Why VQE is attractive to NISQ-ers

- Parameterised unitary (ansatz): $|\phi(\vec{\alpha})\rangle = \left(\prod_k g_k(\alpha_k)\right)|0\rangle^{\otimes n}$
- Hamiltonian as Pauli sums: $H = \sum_j c_j P_j$, Paulis P_j
- By linearity: $\langle\phi(\vec{\alpha})| H |\phi(\vec{\alpha})\rangle = \sum_j c_j \langle\phi(\vec{\alpha})| P_j |\phi(\vec{\alpha})\rangle$parallel!



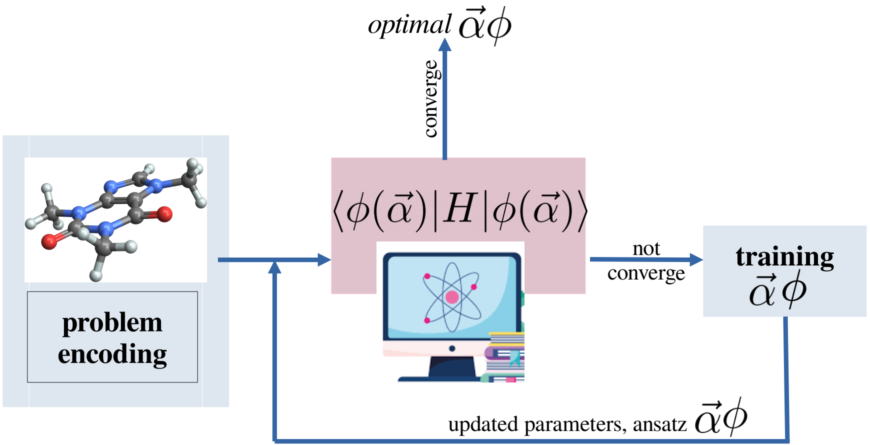
Why VQE is attractive to NISQ-ers

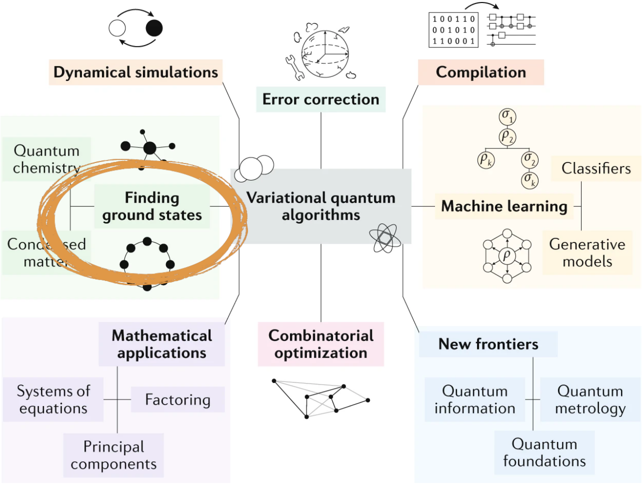
- Parameterised unitary (ansatz): $|\phi(\vec{\alpha})\rangle = (\prod_k g_k(\alpha_k))\,|0\rangle^{\otimes n}$
- Hamiltonian as Pauli sums: $H = \sum_j c_j P_j$, Paulis P_j
- By linearity: $\langle\phi(\vec{\alpha})|\,H\,|\phi(\vec{\alpha})\rangle = \sum_j c_j\,\langle\phi(\vec{\alpha})|\,P_j\,|\phi(\vec{\alpha})\rangle$ parallel!



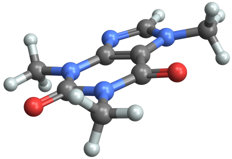
- Classical optimisation over $\vec{\alpha}$
- Some noise can be handled by automatic adjustment of $\vec{\alpha}$

VQE framework



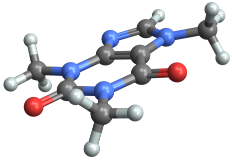

Dynamical simulations
Error correction
Compilation
Quantum chemistry
Finding ground states
Condensed matter
Variational quantum algorithms
Classifiers
Machine learning
Generative models
Mathematical applications
Systems of equations
Factoring
Principal components
Combinatorial optimization
New frontiers
Quantum information
Quantum metrology
Quantum foundations
100110
001010
110001

VQE: problem encoding



$$H_{\vec{R}} = -\sum_i \frac{\nabla^2_{R_i}}{2M_i} - \sum_i \frac{\nabla^2_{r_i}}{2} - \sum_{i,j} \frac{Z_i}{|R_i - r_j|} + \sum_{i,j>i} \frac{Z_i Z_j}{|R_i - R_j|} + \sum_{i,j>i} \frac{1}{|r_i - r_j|}$$

VQE: problem encoding



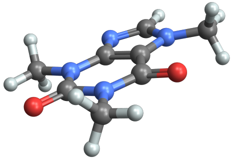
$$H_{\vec{R}} = -\sum_i \frac{\nabla^2_{R_i}}{2M_i} - \sum_i \frac{\nabla^2_{r_i}}{2} - \sum_{i,j} \frac{Z_i}{|R_i - r_j|} + \sum_{i,j>i} \frac{Z_i Z_j}{|R_i - R_j|} + \sum_{i,j>i} \frac{1}{|r_i - r_j|}$$

Jordan-Wigner mapping

$\Rightarrow$ **second quantisation**

Occupation $|01\ldots1\ldots\rangle \leftrightarrow$ qubit $|01\ldots1\ldots\rangle$

VQE: problem encoding



$$H_{\vec{R}} = -\sum_i \frac{\nabla^2_{R_i}}{2M_i} - \sum_i \frac{\nabla^2_{r_i}}{2} - \sum_{i,j} \frac{Z_i}{|R_i - r_j|} + \sum_{i,j>i} \frac{Z_i Z_j}{|R_i - R_j|} + \sum_{i,j>i} \frac{1}{|r_i - r_j|}$$

Jordan-Wigner mapping

$\Rightarrow$ **second quantisation**

Occupation $|01\ldots1\ldots\rangle \leftrightarrow$ qubit $|01\ldots1\ldots\rangle$

Hartree-Fock

1-electron orbitals: $\{\varphi_j(\vec{r}) = \sum_k c_k \phi_k(\vec{r})\} \rightarrow$ self-consistent field

Occupation state: Slater determinant of Spin orbitals $\{\varphi_j(\vec{r})_\uparrow, \varphi_j(\vec{r})_\downarrow\}$

$H_2 = \sum_{pq} h_{pq} a_p^\dagger a_q + \frac{1}{2} \sum_{pqrs} h_{pqrs} a_p^\dagger a_q^\dagger a_r a_s$

VQE: problem encoding

Occupation $|01\ldots1\ldots\rangle \mapsto$ qubit $|01\ldots1\ldots\rangle$

$H_2 = \sum_{pq} h_{pq} a_p^\dagger a_q + \frac{1}{2}\sum_{pqrs} h_{pqrs} a_p^\dagger a_q^\dagger a_r a_s$

$$a_j^\dagger \mapsto |1\rangle\langle 0|_j \otimes Z_{j-1} \otimes \cdots \otimes Z_1$$
$$a_j \mapsto |0\rangle\langle 1|_j \otimes Z_{j-1} \otimes \cdots \otimes Z_1$$

$|1\rangle\langle 0| = (X - iY)/2$ and $|0\rangle\langle 1| = (X + iY)/2$

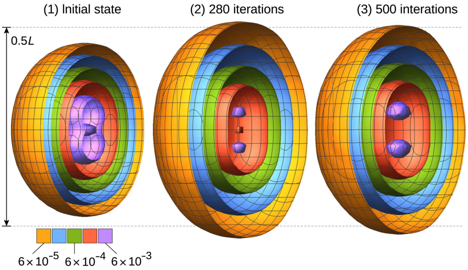
$$H' = \sum_k c_k P_k$$

Others: Bravyi-Kitaev, Parity

VQE: problem encoding

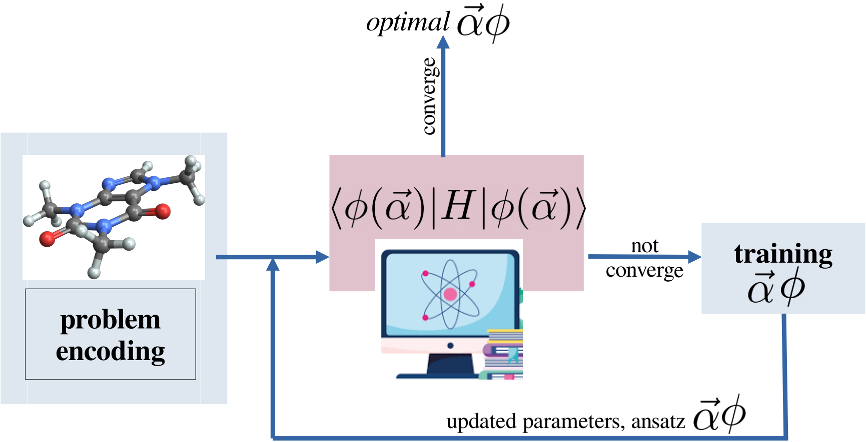
Or stick to the **first quantisation** (2105.12767v3)

Spatial electronic configuration of He_2 using grid method



www.science.org/doi/10.1126/sciadv.abo7484

VQE framework



Parameter training: $\vec{\alpha}$

Steepest gradient descent

$$G = \frac{\partial E(\vec{\alpha})}{\partial \alpha_j}$$

$$\vec{\alpha}' = \vec{\alpha} - \lambda G$$

[2] McArdle, S., Jones, T., Endo, S. et al. Variational ansatz-based quantum simulation of imaginary time evolution. npj Quantum Inf 5, 75 (2019). https://doi.org/10.1038/s41534-019-0187-2

Parameter training: $\vec{\alpha}$

Steepest gradient descent

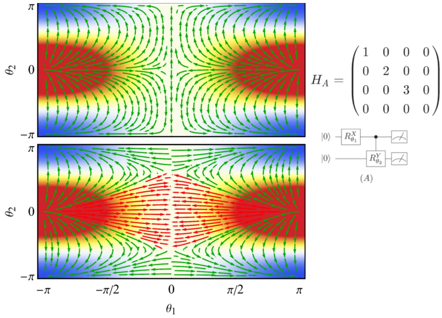
$$G = \frac{\partial E(\vec{\alpha})}{\partial \alpha_j}$$

$$\vec{\alpha}' = \vec{\alpha} - \lambda G$$

Imaginary-time evolution[2]/ Quantum natural gradient

$$\vec{\alpha}' = \vec{\alpha} - \lambda F^{-1} G$$

$$F_{ij} = 4\,\mathrm{Re}\left[\langle \partial_i \psi(\vec{\alpha}) | \partial_j \psi(\vec{\alpha}) \rangle - \langle \partial_i \psi(\vec{\alpha}) | \psi(\vec{\alpha}) \rangle \langle \psi(\vec{\alpha}) | \partial_i \psi(\vec{\alpha}) \rangle \right]$$

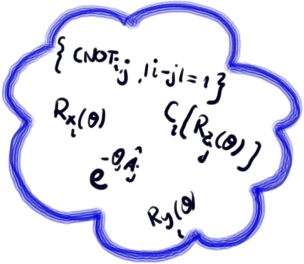


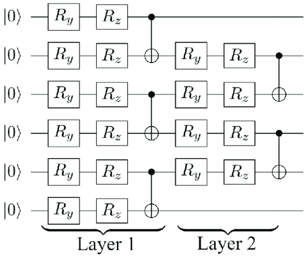
(a) H_A, $|\psi_A\rangle$

[2] McArdle, S., Jones, T., Endo, S. et al. Variational ansatz-based quantum simulation of imaginary time evolution. npj Quantum Inf 5, 75 (2019). https://doi.org/10.1038/s41534-019-0187-2

Training: ansatz ϕ

- Fixed ansatz
- Hardware-efficient ansatz
- Adaptive ansatz



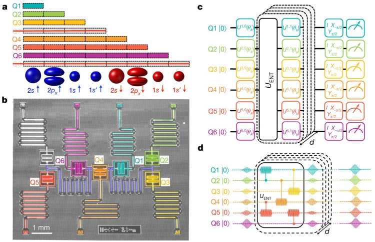


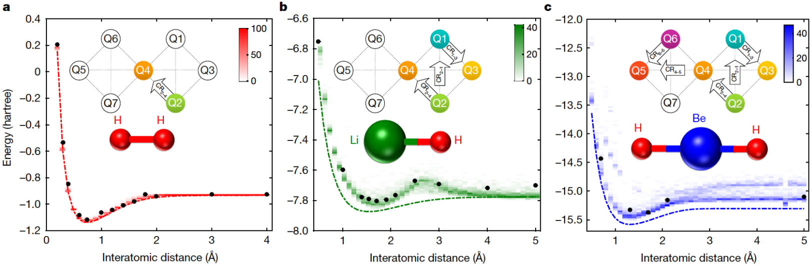
$$\{R\sigma_j(\alpha), C_i(R\sigma_j(\alpha))\}$$

$$\{e^{\hat{T}-\hat{T}^\dagger}\}$$

Bond-dissociation curve

Hardware-efficient ansatz



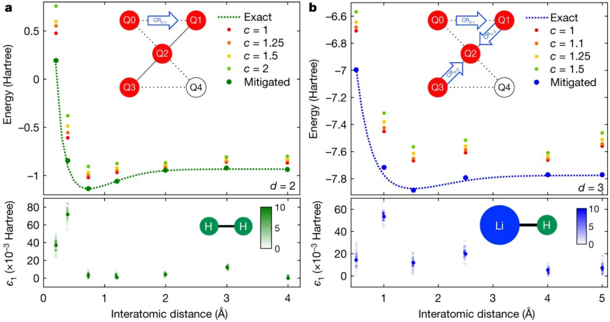


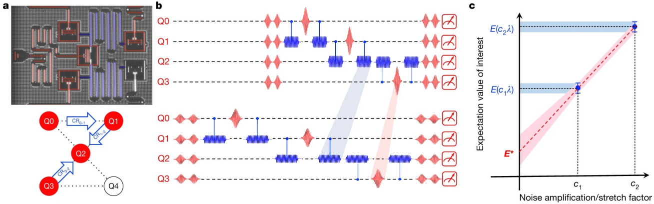
[3]

---

[3] https://www.nature.com/articles/nature23879

VQE + error mitigation

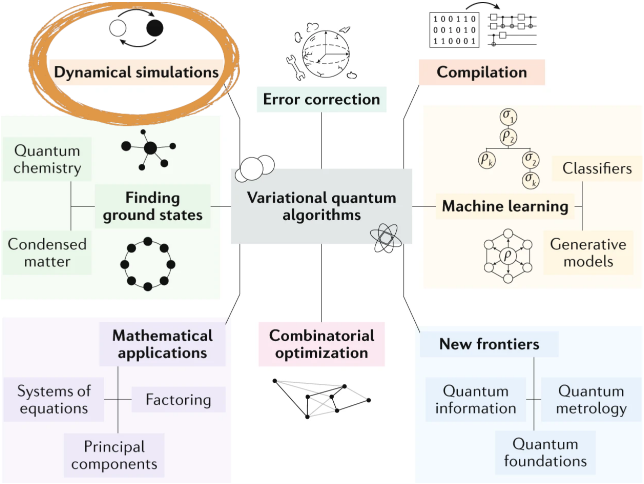




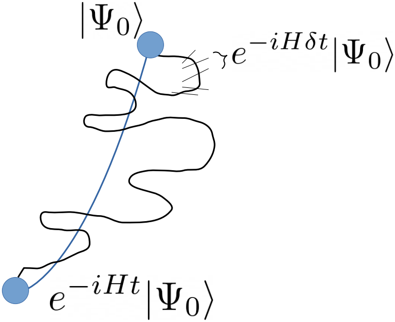
4

5

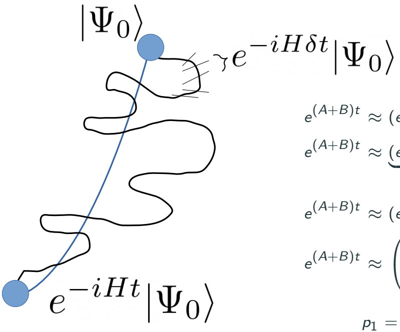
[5]https://www.nature.com/articles/s41586-019-1040-7


Dynamical simulations
Error correction
100110
001010
110001
Compilation
Quantum chemistry
Finding ground states
Condensed matter
Variational quantum algorithms
Classifiers
Machine learning
Generative models
Mathematical applications
Systems of equations
Factoring
Principal components
Combinatorial optimization
New frontiers
Quantum information
Quantum metrology
Quantum foundations

Dynamical simulation



Dynamical simulation



Trotterization

$$e^{(A+B)t} \approx (e^{At/n}e^{Bt/n})^n + O(\delta t^2)$$

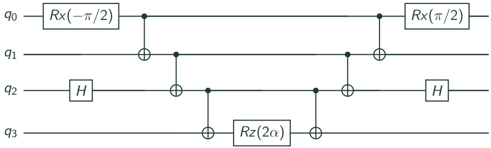
$$e^{(A+B)t} \approx \underbrace{(e^{At/2n}e^{Bt/n}e^{At/2n})(e^{Bt/2n}e^{At/n}e^{Bt/2n})(e^{At/2n}e^{Bt/n}e^{At/2n})\ldots}_{n\text{ terms}} + O(\delta t^3)$$

$$e^{(A+B)t} \approx (e^{\frac{7}{4}At/n}e^{\frac{2}{3}Bt/n}e^{\frac{3}{4}At/n}e^{\frac{-2}{3}Bt/n}e^{\frac{-1}{24}At/n}e^{Bt/n})^n + O(\delta t^4)$$

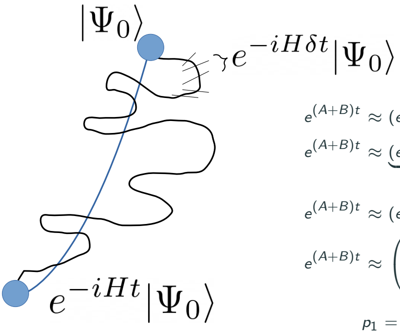
$$e^{(A+B)t} \approx \left(\prod_{i=1}^{5} e^{p_i At/2n}e^{p_i Bt/n}e^{p_i At/2n}\right)^n + O(\delta t^5),$$

$$p_1 = p_2 = p_4 = p_5 = \frac{1}{4 - 4^{1/3}}, \; p_3 = 1 - 4p_1.$$

$$e^{-i\alpha Y_0 Z_1 X_2 Z_3} =$$



Dynamical simulation



Trotterization

$$e^{(A+B)t} \approx (e^{At/n}e^{Bt/n})^n + O(\delta t^2)$$

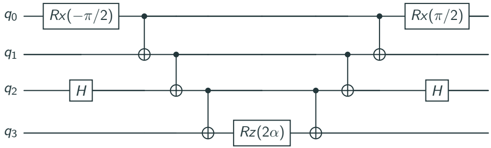
$$e^{(A+B)t} \approx \underbrace{(e^{At/2n}e^{Bt/n}e^{At/2n})(e^{Bt/2n}e^{At/n}e^{Bt/2n})(e^{At/2n}e^{Bt/n}e^{At/2n})\dots}_{n \text{ terms}} + O(\delta t^3)$$

$$e^{(A+B)t} \approx (e^{\frac{7}{4}At/n}e^{\frac{2}{3}Bt/n}e^{\frac{3}{4}At/n}e^{\frac{-2}{3}Bt/n}e^{\frac{-1}{24}At/n}e^{Bt/n})^n + O(\delta t^4)$$

$$e^{(A+B)t} \approx \left(\prod_{i=1}^{5} e^{p_i At/2n}e^{p_i Bt/n}e^{p_i At/2n}\right)^n + O(\delta t^5),$$

$$p_1 = p_2 = p_4 = p_5 = \frac{1}{4 - 4^{1/3}},\ p_3 = 1 - 4p_1.$$

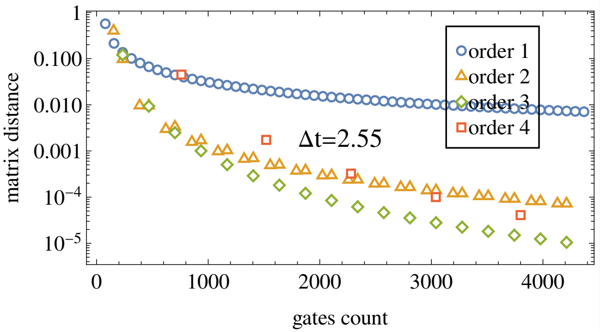
$$e^{-i\alpha Y_0 Z_1 X_2 Z_3} =$$



e.g., H_2 has 15 Pauli terms, H_2O has > 1500 Pauli terms

Trotterisation vs VQA

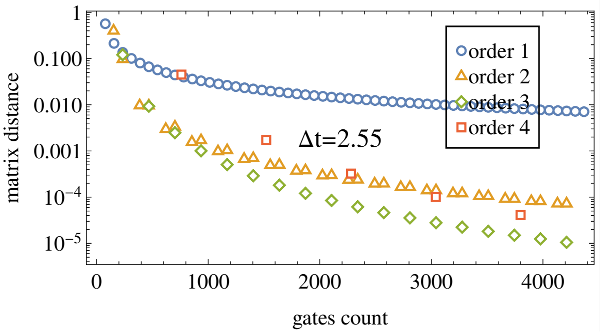
Trotterisation scales badly



H_2

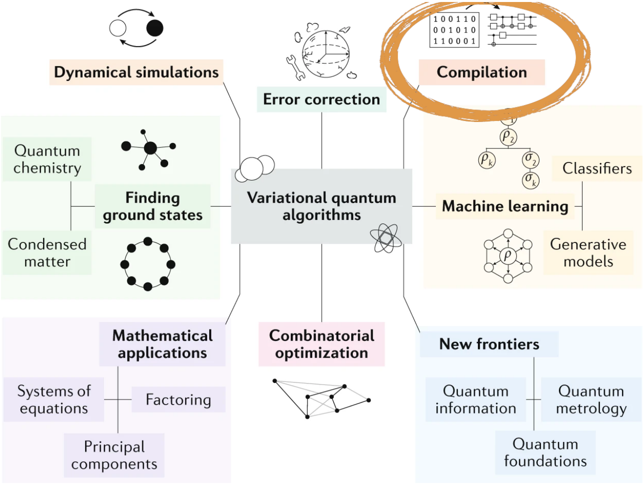
Trotterisation vs VQA

Trotterisation scales badly



H_2

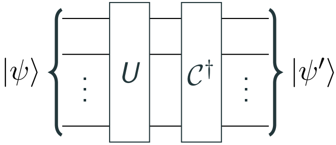
Dynamics simulation via VQA: compiling the propagators $e^{-i\alpha P_j}$



Dynamical simulations
Error correction
1 0 0 1 1 0
0 0 1 0 1 0
1 1 0 0 0 1
Compilation
Quantum chemistry
Finding ground states
Condensed matter
Variational quantum algorithms
Classifiers
Machine learning
Generative models
Mathematical applications
Systems of equations
Factoring
Principal components
Combinatorial optimization
New frontiers
Quantum information
Quantum metrology
Quantum foundations

Compilation

Goal: synthesise circuit $\mathcal{C}$ that approximates U

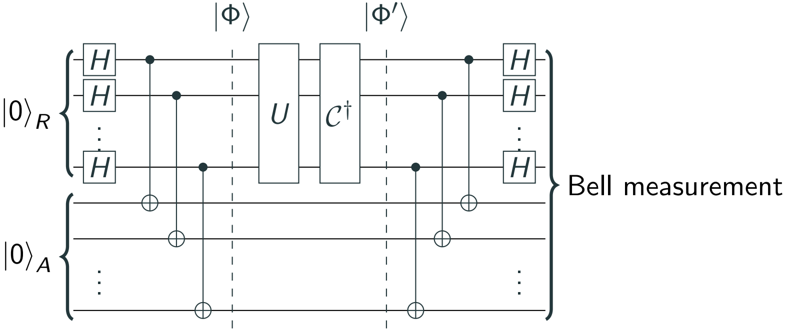


$$|\psi\rangle \sim |\psi'\rangle \iff \mathcal{C}^\dagger U = I$$

For all $|\psi\rangle$

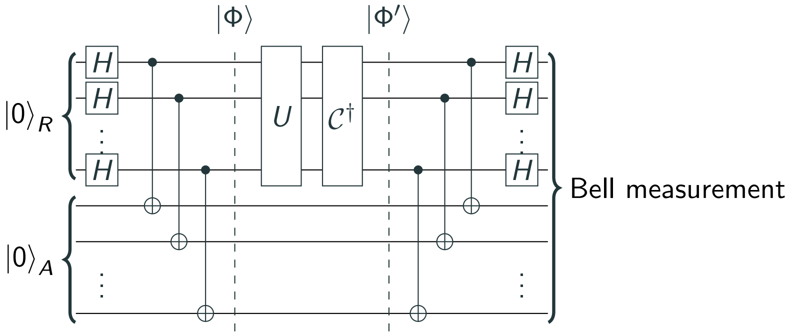
Compilation via Choi-Jamiołkowski state

Goal: synthesise circuit $\mathcal{C}$ that approximates U



Compilation via Choi-Jamiølkowski state

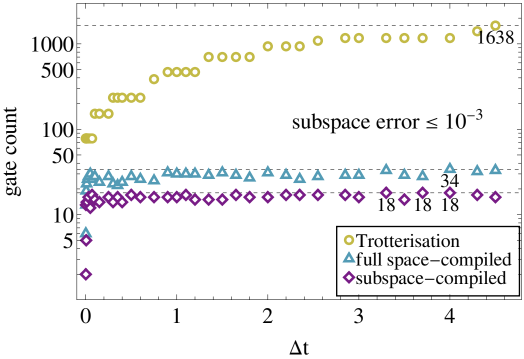
Goal: synthesise circuit $\mathcal{C}$ that approximates U



Cost: $1 - \left|\langle\Phi|\left(\mathcal{C}^\dagger U \otimes I\right)|\Phi\rangle\right|^2$

Cost$= 0 \iff \mathcal{C}^\dagger U = I$

Trotterisation vs VQA on H_2 dynamics

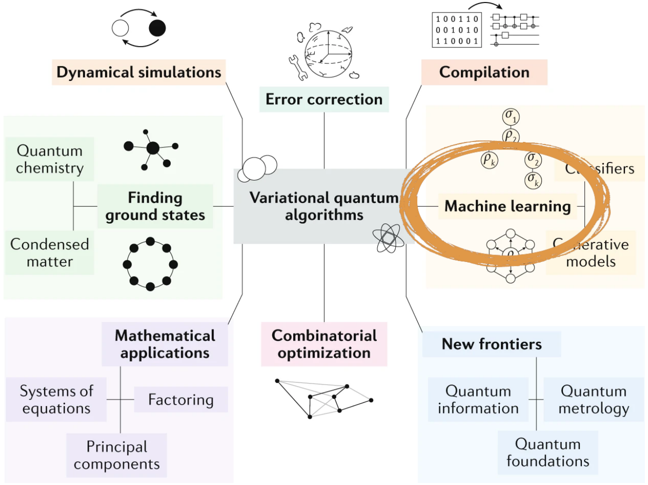


[6]

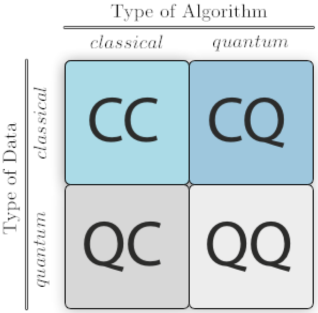
$Exp(-iH\Delta t)$ for H_2

Gates: $\{R\sigma(\alpha), C(R\sigma(\alpha))\}, \sigma$ are Paulis

6 https://arxiv.org/abs/2206.11246


Dynamical simulations
Error correction
Compilation
100110
001010
110001
Quantum chemistry
Finding ground states
Condensed matter
Variational quantum algorithms
Machine learning
models
Mathematical applications
Systems of equations
Factoring
Principal components
Combinatorial optimization
New frontiers
Quantum information
Quantum metrology
Quantum foundations

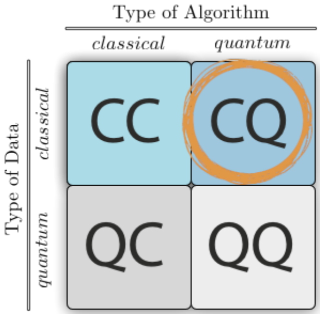
Approaches on Quantum Computing and Machine Learning



[7]

[7] Schuld, Maria, and Francesco Petruccione. Supervised learning with quantum computers. Vol. 17. Berlin: Springer, 2018.

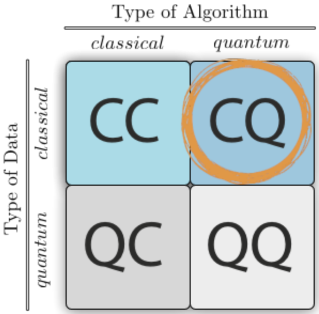
Approaches on Quantum Computing and Machine Learning



[7]

[7] Schuld, Maria, and Francesco Petruccione. Supervised learning with quantum computers. Vol. 17. Berlin: Springer, 2018.

Approaches on Quantum Computing and Machine Learning



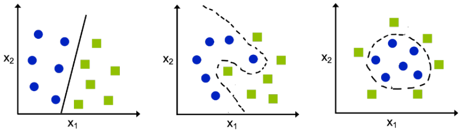
[7]

VQA on QML differs in the presence of data training

[7] Schuld, Maria, and Francesco Petruccione. Supervised learning with quantum computers. Vol. 17. Berlin: Springer, 2018.

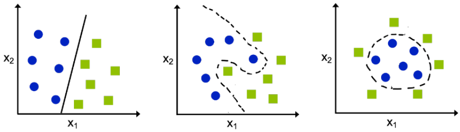
Machine learning: Kernel method

Classification problem: distinguishing two sets of data



Machine learning: Kernel method

Classification problem: distinguishing two sets of data



Kernel $\kappa(x^j, x)$, similarity measure between data points x^j and x

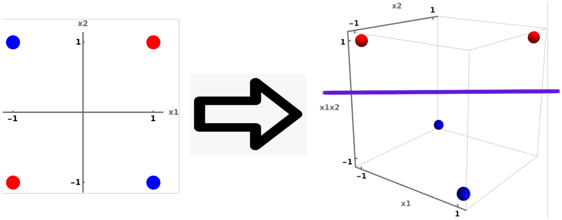
classifier using kernel κ:

$$y = \sigma(\sum_j y^j \kappa(x^j, x))$$

Machine learning: Feature map

Expanding the feature space: map the data into a larger space

Example: SVM (support vector machine)



$$\mathcal{F} : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} x_1 \\ x_2 \\ x_1 x_2 \end{pmatrix}$$

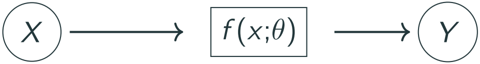
Classifier model is defined on the feature space

Machine learning: ingredients and basic

1. **Data** $(X, Y) \in \mathcal{X} \times \mathcal{Y}$

Machine learning: ingredients and basic

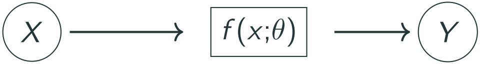
1. **Data** $(X, Y) \in \mathcal{X} \times \mathcal{Y}$


2. **Model** $f(x;\theta) \in \mathcal{F}$, where θ are free parameters


Machine learning: ingredients and basic

1. **Data** $(X, Y) \in \mathcal{X} \times \mathcal{Y}$

2. **Model** $f(x;\theta) \in \mathcal{F}$, where θ are free parameters

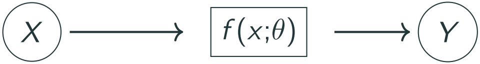


3. **Loss** $L(f(x;\theta), y) \in \mathbb{R}$: minimisation

Machine learning: ingredients and basic

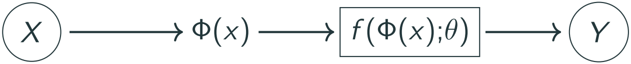
1. **Data** $(X, Y) \in \mathcal{X} \times \mathcal{Y}$

2. **Model** $f(x;\theta) \in \mathcal{F}$, where θ are free parameters

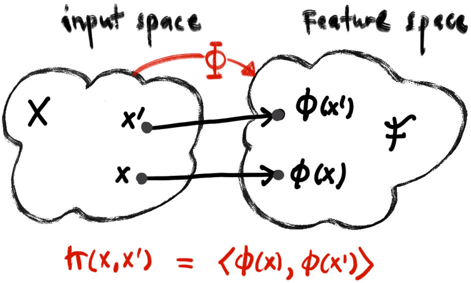


3. **Loss** $L(f(x;\theta), y) \in \mathbb{R}$: minimisation

Feature map Φ



Machine learning: inner product as kernel



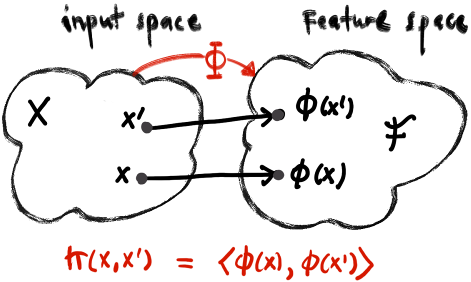
Theorem 1. *Let $\phi : \mathcal{X} \to \mathcal{F}$ be a feature map. The inner product of two inputs mapped to feature space defines a kernel via*

$$\kappa(x, x') := \langle \phi(x), \phi(x') \rangle_{\mathcal{F}}, \quad (1)$$

where $\langle \cdot, \cdot \rangle_{\mathcal{F}}$ is the inner product defined on $\mathcal{F}$. [8]

[8] Schuld, M. and Killoran, N., 2019. Quantum machine learning in feature hilbert spaces. Physical review letters, 122(4), p.040504.

Machine learning: inner product as kernel



Theorem 1. *Let $\phi : \mathcal{X} \rightarrow \mathcal{F}$ be a feature map. The inner product of two inputs mapped to feature space defines a kernel via*

$$\kappa(x, x') := \langle \phi(x), \phi(x') \rangle_{\mathcal{F}}, \qquad (1)$$

where $\langle \cdot, \cdot \rangle_{\mathcal{F}}$ is the inner product defined on $\mathcal{F}$. [8]

$\mathcal{F}$ has to be Hilbert space!

[8] Schuld, M. and Killoran, N., 2019. Quantum machine learning in feature hilbert spaces. Physical review letters, 122(4), p.040504.

Data encoding is the Quantum Feature Map

$$X \longrightarrow |\Phi(x)\rangle \longrightarrow \boxed{f(|\Phi(x)\rangle ;\theta)} \longrightarrow Y$$

Data encoding is the Quantum Feature Map

$$X \longrightarrow |\Phi(x)\rangle \longrightarrow \boxed{f(|\Phi(x)\rangle; \theta)} \longrightarrow Y$$

Examples:

- Binary encoding:

$$x = (1, 0, 1) \mapsto X \otimes I \otimes X |000\rangle$$

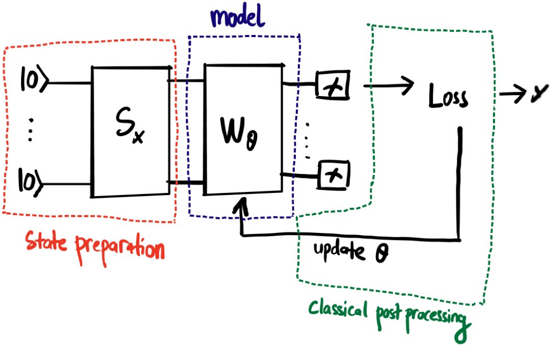
- Amplitude encoding is exponentially more compact but hard to prepare

$$x \mapsto \sum_j x_j |j\rangle$$

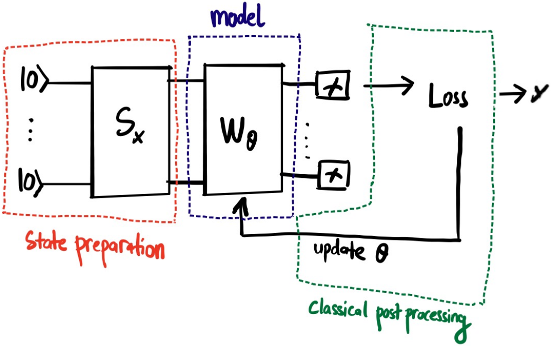
- Rotation angle encoding

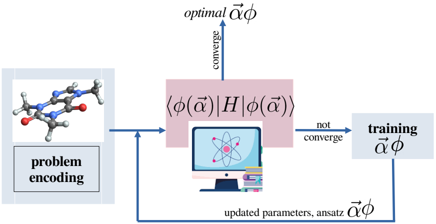
$$x \mapsto \bigotimes_j Rx(x_j) |0 \ldots 0\rangle$$

Variational circuits in QML model

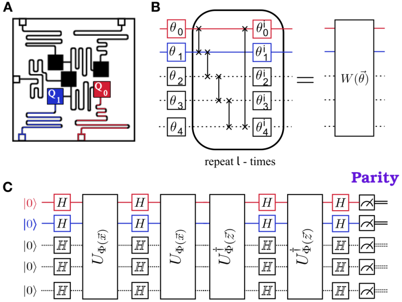


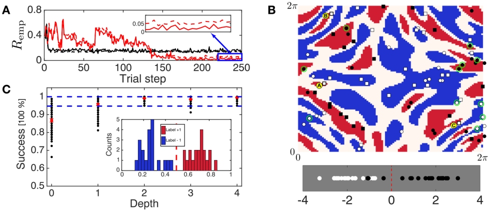
Variational circuits in QML model





Variational quantum classifier[9]





[8] Havlíček, Vojtěch, et al. "Supervised learning with quantum-enhanced feature spaces." Nature 567.7747 (2019): 209-212.

Thank you for your attention! ;)