Learning to measure
Adaptive informationally complete POVMs for near-term quantum algorithms

Sabrina Maniscalco, University of Helsinki, Aalto University, Algorithmiq Oy - Finland
The team

**Finnish Team**
Guillermo Garcia-Perez, University of Turku, Algorithmiq
Matteo Rossi, University of Turku, Algorithmiq
Boris Sokolov, University of Helsinki, Algorithmiq

**IBM Quantum Team**
Francesco Tacchino, IBM Quantum, Zurich, Switzerland
Panagiotis Barkoutsos, IBM Quantum, Zurich, Switzerland
Guglielmo Mazzola, IBM Quantum, Zurich, Switzerland
Ivano Tavernelli, IBM Quantum, Zurich, Switzerland
where it all began
Learning to measure
Adaptive informationally complete POVMs for near-term quantum algorithms

Sabrina Maniscalco, University of Helsinki, Aalto University, Algorithmiq Oy - Finland
Quantum Computing

The Goal
Understanding if, and to which extent, it is possible to build computing machines able to perform tasks which are impossible to conventional (classical) computers.
Quantum Algorithms, exploiting quantum properties of the physical information carriers, outperform all currently existing classical algorithms. However, they run on ideal fault-tolerant universal quantum computers!
A quantum Turing machine or universal quantum computer is an abstract machine used to model the effects of a quantum computer. Any quantum algorithm can be expressed formally as a particular quantum Turing machine.
Information is physical

Physical realization of quantum computers

Algorithms as dynamical quantum processes!
Because of the extreme fragility of quantum information storage and processing in presence of environmental noise, error-correction techniques required to achieve fault-tolerance are still experimentally in their infancy.
It is unclear, at the moment, whether decoherence will eventually be a fundamental unavoidable limiting factor to the manufacturing of universal scalable quantum computers, or not.

The Grand Challenge
Learning to measure

Adaptive informationally complete POVMs for near-term quantum algorithms

Sabrina Maniscalco, University of Helsinki, Aalto University, Algorithmiq Oy - Finland
NISQ Devices

What are the useful problems which quantum computers can solve more efficiently than their classical counterparts?

Which subclass of such problems are experimentally less demanding, given the current state-of-the-art quantum hardware?
Feynman ’82 & Manin ’80

Simulating quantum systems is hard on classical computers. However, one can use quantum systems to simulate other quantum systems efficiently.
Abandoning the “exact”

Unlikely that quantum computers will be able to solve efficiently worst-case instances of NP-hard problems, like combinatorial optimisation problems.

However, it might be possible to build specific-purpose quantum devices able to find better approximate solutions or find such approximate solutions faster.
Variational Quantum Algorithms


Hybrid quantum-classical algorithm

Impact

Roadblocks

VQE

Quantum State preparation

New Iteration

Update \gamma Parameters

classical optimization

Quantum Hardware

Measurement Apparatus
Variational Quantum Eigensolver

minimizing the energy of a many-body correlated quantum system (chemistry examples)

\[ \lambda_{\text{min}} \leq \lambda_0 \equiv \langle \psi(\theta) | H | \psi(\theta) \rangle \]

explore the exponentially-large Hilbert space of the fermionic particles in order to find iteratively the ground state of the Hamiltonian, without solving the full diagonalisation problem.
Electronic Hamiltonian

\[ H_e = -\sum_i \frac{\nabla_i^2}{2} - \sum_{i,I} \frac{Z_I}{|r_i - R_I|} + \frac{1}{2} \sum_{i \neq j} \frac{1}{|r_i - r_j|} \]

Second quantization Hamiltonian

\[ H = \sum_{pq} h_{pq} a_p^+ a_q + \frac{1}{2} \sum_{pqrs} h_{pqrs} a_p^+ a_q^+ a_r a_s \]

The Hamiltonian contains up to \( N^4 \) terms

excitation of an electron into the single electron spin orbital \( p \)

kinetic energy terms of the electrons, and their Coulomb interaction with the nuclei

the electron-electron Coulomb repulsion

the knowledge of the ground state of a chemistry compound as a function of the bond length allows one to extract crucial information such as the equilibrium bond length, bond angle, and dissociation energy
Fermion to qubit mappings

**Jordan - Wigner mapping**
Pauli weight: $O(N)$

**Parity mapping**
Pauli weight: $O(N)$

**Bravyi-Kitaev mapping**
Pauli weight: $O(\log(N))$

**Jiang mapping**
Pauli weight: $O(\log(N))$

$Z. \ Jian \ et \ al., \ arXiv \ 1910.10746 \ (2020)$

\[ H = \sum_{pq} h_{pq} a_{p\dagger} a_q + \frac{1}{2} \sum_{pqrs} h_{pqrs} a_{p\dagger} a_{q\dagger} a_r a_s \]

\[ P_k = \bigotimes_{i=1}^{N} \sigma_{k_i}^{(i)} \]

$\sigma_0^{(i)} = I^{(i)}$, $\sigma_1^{(i)} = \sigma_x^{(i)}$

Pauli weight: number of non-identity single qubit Pauli operators in a Pauli string
<table>
<thead>
<tr>
<th>Fermion</th>
<th>Jordan-Wigner</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a</td>
<td>0001\rangle + b</td>
</tr>
<tr>
<td>$a_0$</td>
<td>$Q_0$</td>
</tr>
<tr>
<td>$a_1$</td>
<td>$Q_1 Z_0$</td>
</tr>
<tr>
<td>$a_2$</td>
<td>$Q_2 Z_1 Z_0$</td>
</tr>
<tr>
<td>$a_3$</td>
<td>$Q_3 Z_2 Z_1 Z_0$</td>
</tr>
<tr>
<td>$\hat{n}_i = a_i^\dagger a_i$</td>
<td>$</td>
</tr>
</tbody>
</table>

Jorden-Wigner transformation

stores the occupation number locally and the parity non-locally

Z are parity operators
### Parity Transformation

stores the parity locally, and the occupation number non-locally

<table>
<thead>
<tr>
<th>Fermion</th>
<th>Jordan-Wigner</th>
<th>Parity</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a</td>
<td>0001⟩ + b</td>
<td>0010⟩ + c</td>
</tr>
<tr>
<td>$a_0$</td>
<td>$Q_0$</td>
<td>$X_3 X_2 X_1 Q_0$</td>
</tr>
<tr>
<td>$a_1$</td>
<td>$Q_1 Z_0$</td>
<td>$X_3 \left( Q_1</td>
</tr>
<tr>
<td>$a_2$</td>
<td>$Q_2 Z_1 Z_0$</td>
<td>$X_3 \left( Q_2</td>
</tr>
<tr>
<td>$a_3$</td>
<td>$Q_3 Z_2 Z_1 Z_0$</td>
<td>$Q_3</td>
</tr>
<tr>
<td>$\hat{n}<em>i = a</em>\dagger_i a_i$</td>
<td>$</td>
<td>1⟩⟨1</td>
</tr>
<tr>
<td>Fermion</td>
<td>Jordan-Wigner</td>
<td>Parity</td>
</tr>
<tr>
<td>-----------------</td>
<td>---------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>$a</td>
<td>0001\rangle + b</td>
<td>0010\rangle$</td>
</tr>
<tr>
<td>$+c</td>
<td>0100\rangle + d</td>
<td>1000\rangle$</td>
</tr>
<tr>
<td>$a_0$</td>
<td>$Q_0$</td>
<td>$X_3X_2X_1Q_0$</td>
</tr>
<tr>
<td>$a_1$</td>
<td>$Q_1Z_0$</td>
<td>$X_3X_2 \left( Q_1</td>
</tr>
<tr>
<td>$a_2$</td>
<td>$Q_2Z_1Z_0$</td>
<td>$X_3 \left( Q_2</td>
</tr>
<tr>
<td>$a_3$</td>
<td>$Q_3Z_2Z_1Z_0$</td>
<td>$Q_3</td>
</tr>
<tr>
<td>$n_i = a_i^\dagger a_i$</td>
<td>$</td>
<td>1\rangle \langle 1</td>
</tr>
</tbody>
</table>

a midway point between the JW and parity methods: it compromises on the locality of occupation number and parity information.
Jiang Mapping

defined on ternary trees, maps any single Majorana operator of an n-mode fermionic system to a multi-qubit Pauli operator acting nontrivially on $O(\log(N))$ qubits
Variational ansatz

\[
\psi(\theta)
\]

State preparation

A state which has a sufficiently large overlap with the target eigenstate (ground state)

e.g., hardware efficient ansatz and chemically inspired ansatz
calculate the expectation value

\[ \langle H \rangle_{\psi(\theta)} = \sum_k c_k \langle P_k \rangle_{\psi(\theta)} \]

repeat to accumulate statistics
Impact

Solving classical intractable chemistry problems

- high-T superconductivity
- biochemical reactions
- transition metal catalysis

Designing new compounds

- drug design
- material science
Roadblocks

1 optimisation procedure can become difficult in practice because of the presence of many local minima
2 measurement problem

classical approaches → ansatz choice
For many interesting Hamiltonians, as we increase the size of the problem, the number of measurements needed, in a regime where the quantum core of VQE is advantageous, would be extremely large, so it may take years to reach the required precision (chemical accuracy).
Learning to measure
Adaptive informationally complete POVMs for near-term quantum algorithms

Sabrina Maniscalco, University of Helsinki, Aalto University, Algorithmiq Oy - Finland
Learning to Measure

A new perspective

A novel method which sensibly alleviates the demands on the number of needed measurements, paving the way to current experimental feasibility.
The measurement problem

Sub-optimal measurements scheme, as the variance of $O$ is the sum of the weighted variances of the individual Pauli strings.

\[ O = \sum_k c_k P_k \]

Error in the estimation

\[ \epsilon = \sqrt{\sum_k |c_k|^2 \text{Var}(P_k) / S} \]

\[ \text{Var}(P_k) \approx \langle P_k^2 \rangle - \langle P_k \rangle^2 \]

Grouped Pauli method aims to identify all the Pauli strings that can be measured simultaneously from the same data set.
\[ \Pi_m^{(i)} \quad m = 0, 1, 2, 3 \quad \text{effects} \]

Each effect is associated to one of the 4 possible outcomes of 2 qubit measurements.

\[ \text{Tr}[\rho \Pi_i] \]

Informationally complete positive operator-valued measures (IC-POVMs)

Can be used to estimate any expectation value of our choice!

Starting point
quantum-classical hybrid
Monte Carlo algorithm

measuring a linear combination of a large number of Pauli strings

Monte Carlo integral

\[
\mathcal{O} = \sum_{k} c_k \bigotimes_{i=1}^{N} \sigma^{(i)}_{k_i} = \sum_{k} c_k \bigotimes_{i=1}^{N} \left( \sum_{m_i} b^{(i)}_{k_i m_i} \Pi^{(i)}_{m_i} \right)
\]

\[
= \sum_{m} \left( \sum_{k} c_k \prod_{i=1}^{N} b^{(i)}_{k_i m_i} \right) \Pi_m = \sum_{m} w_m \Pi_m.
\]

\[
\langle \mathcal{O} \rangle = \text{Tr}[\rho \mathcal{O}] = \sum_{m} w_m \text{Tr}[\rho \Pi_m] = \sum_{m} w_m \rho_m
\]

\[
\langle \mathcal{O} \rangle = \langle \omega_m \rangle \{p_m\}
\]
We can estimate the weighted average of all the Pauli strings simultaneously, regardless of whether they commute or not, by exploiting IC data.

1. circumvents costly tomographic reconstruction of quantum states

2. naturally takes into account the covariance between all these parallel measurements

\[ \langle \mathcal{O} \rangle = \langle w_m \rangle \{ p_m \} \]

repeat the measurement \( S \) times using the local POVMs to sample from the probability distribution \( \{ p_m \} \)

estimator

\[ \bar{\mathcal{O}} = \frac{1}{S} \sum_{s=1}^{S} w_{m_s} \]

\[ w_{m_s} = \sum_k c_k \prod_{i=1}^{N} b^{(i)}_{k_i m_i} \]

can be calculated efficiently on a classical computer

estimation error \[ \sqrt{\text{Var}(w_m)/S} \]
POVM optimization
via classical gradient estimation

classical postprocessing routine
to navigate the space of POVMs
towards low-variance ones

one feeds the outcomes of the current POVM to
the classical routine, which uses them to
evaluate the variance of the new POVMs
	on the fly

calculate on
classical computer

gradient descent

\[
\text{Var}(w_r) = \sum_r w_r^2 \text{Tr}[\rho \Gamma_r] - \left( \sum_r w_r \text{Tr}[\rho \Gamma_r] \right)^2
= \langle \mathcal{O} \rangle^2
\]
does not depend on the POVMs

\[
\langle w_r^2 \rangle_{\{q_r\}} = \sum_r w_r^2 \text{Tr} \left[ \rho \bigotimes_{i=1}^N \left( \sum_{m_i} d_{r_i m_i}^{(i)} \Pi_{m_i}^{(i)} \right) \right]
= \sum_m p_m \sum_r \left( \prod_{i=1}^N d_{r_i m_i}^{(i)} \right) w_r^2.
\]
on the fly

we do not need to first optimise the POVM until it reaches a small-enough variance to then start estimating the expected value of the observable.

The intermediate POVMs used in the process are also IC, so they can be used for the estimation of the mean value of $O$ as well.

The strategy is to calculate a weighted average of the estimated means that minimises the resulting variance in the estimation.

This can be done iteratively as the optimisation algorithm progresses.
Key ideas

We use the IC data obtained with the POVM twice

1. to produce an estimation of the mean of the observable.

2. to find a better POVM in the next experiment

* the measurement learning procedure improves over the initial POVM with no additional measurement cost
Learning to measure

Adaptive informationally complete POVMs for near-term quantum algorithms

Sabrina Maniscalco, University of Helsinki, Aalto University, Algorithmiq Oy - Finland
estimated (dashed) and actual (solid) error in the estimation of the ground energy
8 qubits H2 Jiang

Learning to measure

Example 2

different colors = different qubits
about 100 repetitions
Number of shots $S_{\text{tar}}$ required to achieve a target error of 0.5 mHa for H chains as a function of the number of qubits $N$.

$S_{\text{tar}} = aN^b$

<table>
<thead>
<tr>
<th>Method</th>
<th>Parity</th>
<th>Mapping BK</th>
<th>Mapping JKMN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pauli</td>
<td>6.0 ± 0.4</td>
<td>6.8 ± 0.5</td>
<td>6.2 ± 0.2</td>
</tr>
<tr>
<td>Grouped Pauli</td>
<td>5.5 ± 0.4</td>
<td>6.4 ± 0.5</td>
<td>5.7 ± 0.4</td>
</tr>
<tr>
<td>SIC-POVM 1</td>
<td>5.8 ± 0.7</td>
<td>5.7 ± 0.6</td>
<td>4.9 ± 0.5</td>
</tr>
<tr>
<td>SIC-POVM 2</td>
<td>5.4 ± 0.5</td>
<td>4.4 ± 0.4</td>
<td>4.7 ± 0.2</td>
</tr>
<tr>
<td>Grad. POVM 1</td>
<td>4.3 ± 0.6</td>
<td>4.6 ± 0.5</td>
<td>3.2 ± 0.3</td>
</tr>
<tr>
<td>Grad. POVM 2</td>
<td>4.0 ± 0.6</td>
<td>4.4 ± 0.5</td>
<td>3.4 ± 0.3</td>
</tr>
</tbody>
</table>
Using the data obtained to calculate the mean energy for partial (2, 3 and 4 qubit) tomography.

Partial tomography
Learning to measure

Adaptive informationally complete POVMs for near-term quantum algorithms

Sabrina Maniscalco, University of Helsinki, Aalto University, Algorithmiq Oy - Finland
Conclusions

- New measurement strategy in which the optimal measurement of an operator average is learnt in an adaptive fashion with no measurement overhead
- Does not require exponentially-scaling classical computations
- Our algorithm is completely agnostic to the nature of the qubit Hamiltonian and it is exact.
- The measurement data can be reused to calculate other properties of the state, including its tomographic reconstruction.
We are one step closer!

Take home message

The road to useful quantum advantage requires new skillful ways to optimise every aspect of the near-term quantum algorithms.
Learning to measure

Adaptive informationally complete POVMs for near-term quantum algorithms

Sabrina Maniscalco, University of Helsinki, Aalto University, Algorithmiq Oy - Finland