Variational Quantum Machine Learning

Episode IX, Warsaw Quantum Computing Group
Talk given at: Google Poland
host: Pawel Gora

Jacob D. Biamonte

Deep Quantum Laboratory
Skolkovo Institute of Science and Technology
3 Nobel Street
Moscow Russia 143026

J.Biamonte@Skoltech.ru
http://Quantum.Skoltech.ru

Review Articles

1. **Quantum Machine Learning**
   Jacob Biamonte and others
   Nature 549, 195-202 (2017)
2. **Tensor Networks in a Nutshell**
   Jacob Biamonte and Ville Bergholm
   in review (2019) arXiv:1708.00006
3. **Complex Networks from Classical to Quantum**
   Jacob Biamonte, Mauro Faccin and Manlio De Dominico
   Communications Physics 2, 53 (2019)
4. **Pushing Tensor Networks to the Limit**
   Anastasiia A. Pervishko and Jacob Biamonte
   Physics 12, 59 (2019)
5. **Charged String Tensor Networks**
   Jacob Biamonte
   Proceedings of the National Academy of Sciences 114, 2447 (2017)

Why ‘simulators’?

1. **Concept:** Time-scales of physical processes can be much shorter than time to simulate those processes with conventional computer algorithms
2. **Past:** Wind tunnels simulate fluid dynamics—now nearly replaced with supercomputers
3. **Future:** Quantum systems are difficult to simulate with supercomputers

$\rightarrow$ replace computer algorithms with actual physics

Experiments towards new computers

1. **Idea:** build a physical system and measure it $\rightarrow$ 'simulator'
2. **Goal:** increased control $\rightarrow$ programmable simulator $\rightarrow$ universal computer...
3. Manufacturing technologies now enable precise control at the small scale
4. The world is racing to utilize natural physical processes to compute
5. **Example:** replace simulated annealing with physical annealing!

Today's talk

Part I. Breaking Things: when do simulators fail?

1. When do stochastic annealers fail?
2. When do variational quantum approximate optimization algorithms fail (QAOA)?

Part II. Optimistic Predictions: how far can simulators be pushed?

3. Merging quantum simulation with machine learning?
4. Does variational quantum computation admit a universal model of quantum computation?

Part I. When do simulators fail?

1. When do stochastic annealers fail?

Computational Phase Transition Signature in Gibbs Sampling
H. Philathong, V. Akshay, I. Zacharov, J. Biamonte
in review (2019) `arXiv:1906.10705`

H. Philathong

V. Akshay

I. Zacharov

Satisfiability Instances

1. The k-satisfiability (k-SAT) problem is a decision problem determining satisfiability of Boolean formula.
2. We let a k-SAT instance consist of M clauses over N Boolean variables.
3. The clause density of a random instance is defined by the simple faction $\alpha = M/N$.

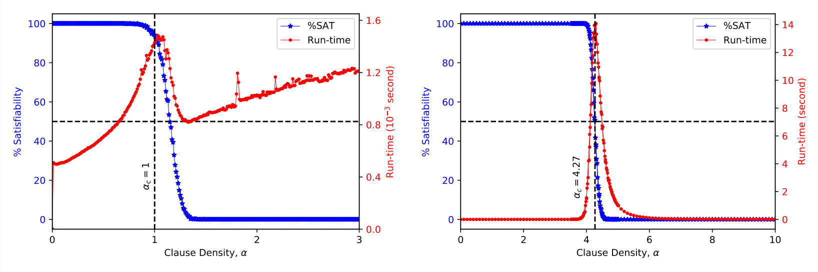


Figure 1: Percent of satisfiabity and run-time versus clause density. For every clause density α, we generated $1,000$ 2-SAT instances (left) and 3-SAT instances (right). $\alpha_c = 1$ and $\alpha_c = 4.27$ for 2-SAT and 3-SAT respectively; Simple resolution SAT for $\alpha < \alpha_c$ and UNSAT for $\alpha > \alpha_c$; Hard resolution for $\alpha \sim \alpha_c$.

Ground State Occupancy of Thermal States

1. We encode 2 (3)-SAT instances into Hamiltonians $\rightarrow$ ground state minimizes instance.
2. **2-SAT maps directly to two-body Hamiltonians; requires long-range $\rightarrow$ all-to-all connectivity**
3. Thermal equilibrium ideally described by Gibbs state,

$$\rho_\beta = \frac{e^{-\beta\mathcal{H}}}{\mathcal{Z}}, \qquad \mathcal{Z} = \mathrm{tr}\{e^{-\beta\mathcal{H}}\}, \tag{1}$$

   $\mathcal{H}$ is the 2 (3)-SAT Hamiltonian, β is inverse temperature.
4. Occupancy in the (possibly degenerate d) lowest-energy, λ_{min}, subspace becomes

$$p\left(\lambda_{\text{min}}, \beta\right) = \frac{1}{\mathcal{Z}} \sum_{i=1}^{d} \langle i| \, e^{-\beta\mathcal{H}} \, |i\rangle = \frac{d}{\mathcal{Z}} e^{-\beta\lambda_{\text{min}}}. \tag{2}$$

5. The quantity $p\left(\lambda_{\text{min}}, \beta\right)$ accesses the difficulty of sampling the solution.

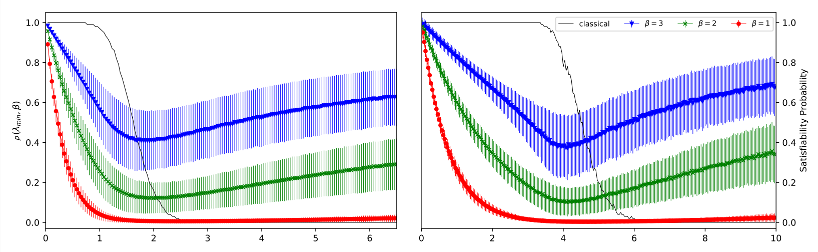


Figure 2: Occupancy of the thermal ground state corresponding to Hamiltonians embedding 2-SAT instances (left) and 3-SAT instances (right) across the algorithmic phase transition. (26 spins; $\beta = 1, 2, 3$; vertical bars, standard deviation).

Empirical. For fixed β, there exists problem instances requiring *significant* sampling time to recover the ground-state.

What average inverse temperature is required to ensure that the occupancy of the ground thermal state, $p(\lambda_{\text{min}}, \beta)$, is greater than 0.9?

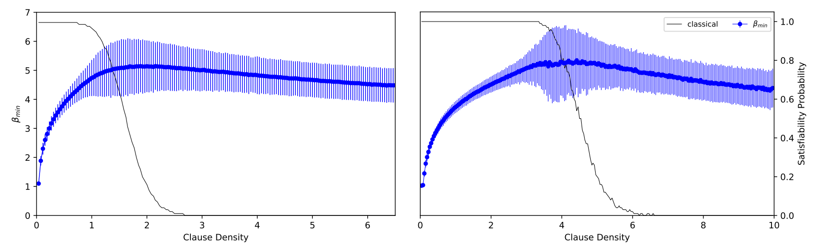


Figure 3: Minimum β for 2-SAT (left) and 3-SAT (right) such that the ground state occupancy, $p(\lambda_{\text{min}}, \beta)$, is greater than 0.9 across the algorithmic phase transition (26 spins).

Towards variational quantum circuits as machine learning models

1. A hybrid quantum-classical algorithm to minimize the energy of a spin Hamiltonian [Nat. Comm. 5, 4213 (2014)]
2. Frarhi and others `arXiv:1411.4028`, 2014
3. **A quantum algorithm to train neural networks using low-depth circuits**
   Guillaume Verdon, Michael Broughton, Jacob Biamonte
   in review (2019) arXiv:1712.05304
4. **Machine Learning Phase Transitions with a Quantum Processor**
   Alexey Uvarov, Andrey Kardashin, Jacob Biamonte
   in review (2019) arXiv:1906.10155

1. Select a parametrized quantum circuit (an ansatz)

$$|\psi(\boldsymbol{\theta})\rangle = U_k(\theta_k)\dots U_1(\theta_1)\,|\mathbf{0}\rangle$$

2. Minimize some objective function
3. (**Example.**) $\langle\psi(\boldsymbol{\theta})|\,H\,|\psi(\boldsymbol{\theta})\rangle$ over $\boldsymbol{\theta}$ by estimating each term $\langle\psi(\boldsymbol{\theta})|\,H_{\boldsymbol{\alpha}}\,|\psi(\boldsymbol{\theta})\rangle$ separately

Definition (Variational Statespace)

The variational statespace of a l-parameterized n-qubit state preparation process is the complex linear extension of $|\psi(\boldsymbol{\theta})\rangle$ over all possible assignments of real numbers $\boldsymbol{\theta}$.

$$\mathrm{Span}\{|\psi(\boldsymbol{\theta})\rangle\,|\boldsymbol{\theta} \subset \mathbb{R}^{\times l}\} \subseteq \mathbb{C}_2^{\otimes n} \tag{3}$$

arXiv:1903.04500

Definition (Bounded Objective Function)

We call a family of objective functions efficiently computable when uniformly generated by calculating the expected value of an operator with bounded linear extension over

$$\Omega \subset \{\mathbf{1}, X, Y, Z\}^{\otimes n}. \tag{4}$$

Variational Statespace Examples

Examples include preparing $|\psi(\boldsymbol{\theta})\rangle$ by either

1. a quantum circuit with $\boldsymbol{\theta} \in (0, 2\pi]^{\times l}$ tunable parameters as

$$|\psi(\boldsymbol{\theta})\rangle = \Pi_l U_l \, |0\rangle^{\otimes n} \tag{5}$$

   where U_l is adjusted by θ_l

2. by tuning accessible time-dependent and appropriately bounded parameters ($\theta_k(t)$ corresponding to Hermitian $A^{(k)}$) as

$$|\psi\rangle = \mathcal{T}\{e^{-\imath \sum \theta_k(t) A^{(k)}}\} \, |0\rangle^{\otimes n} \tag{6}$$

   where $\mathcal{T}$ time orders the sequence and superscript k indexes the kth operator $A^{(k)}$

An interesting connection between variational algorithms and quantum control is only now being explored.

Definition (Poly-Computable Objective Function)

An objective function

$$f : |\phi\rangle^{\times O(\text{poly}(n))} \to \mathbb{R}_+ \tag{7}$$

is called poly-computable provided poly(n) independent physical copies of $|\phi\rangle$ can be efficiently prepared to evaluate poly(n) expected values taken from a subset of the Pauli algebra on n qubits.

We say an objective function **accepts** $|\phi\rangle$ iff

$$f(|\phi\rangle^{\times O(\text{poly}(n))}) = f(|\phi\rangle, |\phi\rangle, \ldots, |\phi\rangle) < \Delta \tag{8}$$

evaluates strictly less than a chosen parameter Δ.

Objective Function Examples

Efficiently computable objective function examples include:

1. Hamiltonian's in the Pauli basis known to be non-vanishing for at most some poly(n) terms.
2. Calculating the expected value

$$\langle\phi|\,\mathcal{H}\,|\phi\rangle = \sum \mathcal{J}^{ab\ldots c}_{\alpha\beta\ldots\gamma}\ \langle\phi|\,\sigma^a_\alpha\sigma^b_\beta\cdots\sigma^c_\gamma\,|\phi\rangle \tag{9}$$

In variational quantum computing, (9) is evaluated term-wise (i.e. each $\langle\phi|\,\sigma^a_\alpha\sigma^b_\beta\cdots\sigma^c_\gamma\,|\phi\rangle$ is evaluated on the quantum processor, scaled and then the entire sum is evaluated classically). The goal is to vary over $|\phi\rangle$ and minimize (9).

Boolean Satisfiability

1. SAT-instances are embedded into Hamiltonians,

$$\mathcal{H}_{\mathsf{SAT}} = \sum_{l} \mathcal{P}(l), \tag{10}$$

   where l indexes each clause in the instance.

2. Unsatisfiable assignments are penalized with at-least 1 unit of energy
3. QAOA with $\mathcal{H}_x = \sum_i \sigma_x^{(i)}$ and $\mathcal{V} = \mathcal{H}_{\mathsf{SAT}}$ calculates the energy approximation,

$$E_g^{\mathrm{QAOA}} = \min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \langle \psi(\boldsymbol{\alpha},\boldsymbol{\beta}) | \, \mathcal{H}_{\mathsf{SAT}} \, | \psi(\boldsymbol{\alpha},\boldsymbol{\beta}) \rangle \,. \tag{11}$$

Self-driving eigensolver

1. We can consider expected values of the RIG

$$\{H, \cdot, +, \mathbb{R}\}^{O(\ln n)} \tag{12}$$

2. The divergence $\mathbb{E}^2 = \langle H^2 \rangle - \langle H \rangle^2$ vanishes if and only if the prepared state is an eigenstate of the Hamiltonian.
3. Here $\langle H \rangle^2$ is calculated by first calculating $\langle H \rangle$ and $\langle H^2 \rangle$ is calculated by expanding $\left(\sum \mathcal{J}^{ab\ldots c}_{\alpha\beta\ldots\gamma} \ \sigma^a_\alpha \sigma^b_\beta \cdots \sigma^c_\gamma \right)^2$ and evaluating not more than $\sim poly(n)^2$ expected values.

2. When do variational quantum approximate optimization algorithms fail?

Reachability Deficits in Quantum Approximate Optimization
V. Akshay, H. Philathong, M.E.S. Morales, J. Biamonte
in review (2019) `arXiv:1906.11259`

V. Akshay

H. Philathong

M.E.S. Morales

Quantum Approximate Optimization

1. Create ansatz states selecting $2p$-tunable parameters,

$$|\psi(\boldsymbol{\alpha}, \boldsymbol{\beta})\rangle = \prod_{i=1}^{p} \mathcal{U}(\alpha_i, \beta_i) \left|+\right\rangle^{\otimes n}, \tag{13}$$

   with

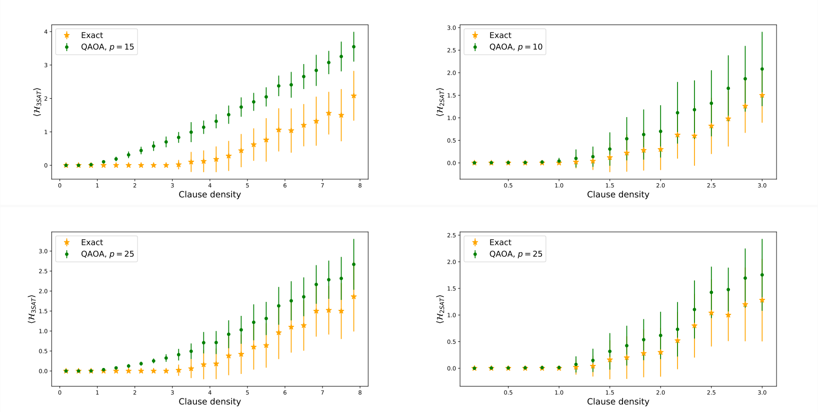
$$\mathcal{U}(\alpha_k, \beta_k) = \exp\{-i\beta_k \mathcal{H}_x\}. \exp\{-i\alpha_k \mathcal{V}\}. \tag{14}$$

   .

2. Measurement returns a bit sting $\to$ evaluated by classical objective function (to be minimized)
3. Classical optimization assigns $2p$-parameters; process is repeated

For QAOA, the objective function to be minimized is the optimization problem. (encoded in the Hamiltonian $\mathcal{V}$)

$$F_{\mathcal{V}} : \{0, 1\}^n \to \mathbb{R}_+ \tag{15}$$



Energy versus clause density for 3-SAT (Left) and 2-SAT (Right). Dots, averaged energies obtained from QAOA; stars, exact values averaged on 50 randomly generated SAT instances for $n = 6$. Observed convergence to exact values for increasing depth.

Definition (Reachability Deficits)

Let $|\psi\rangle$, be the ansatz states generated from a p–depth QAOA circuit as shown in (13). Then

$$\Delta = \min_{\psi \subsetneq \mathcal{H}} \langle\psi| \mathcal{V} |\psi\rangle - \min_{\phi \in \mathcal{H}} \langle\phi| \mathcal{V} |\phi\rangle , \tag{16}$$

characterises the performance of QAOA.

$\Delta = f(p, \alpha, N)$ and for $p \in \mathbb{N}$ and fixed problem size, $\exists$ density, $\alpha > \alpha_c$ such that $\Delta \neq 0$.

Conclusion. The performance of QAOA exhibits strong dependence on the *density* of the problem instance.

Implication. Provides the first known heuristic and the first general limiting feature of the algorithm.

Reachability Deficits in Quantum Approximate Optimization
V. Akshay, H. Philathong, M.E.S. Morales, J. Biamonte
in review (2019) `arXiv:1906.11259`

3. NISQ application: quantum enhanced machine learning plus quantum simulation

Machine Learning Phase Transitions with a Quantum Processor

A. Uvarov, A. Kardashin, J. Biamonte

in review (2019) `arXiv:1906.10155`

A. Uvarov

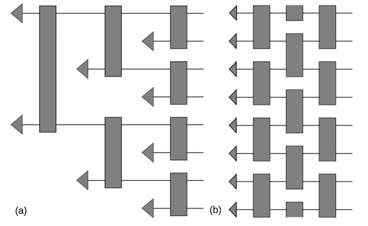
A. Kardashin

Quantum Machine Learning Phase Transitions

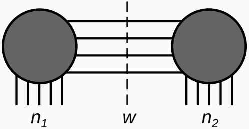
1. **Entanglement Scaling in Quantum Supremacy Benchmarks**
   Mauro E. S. Morales, Dax Enshan Koh, Jacob D. Biamonte
   in review (2019) arXiv:1808.00460

Tensor network states

A quantum circuit can be treated with tensor network theory:

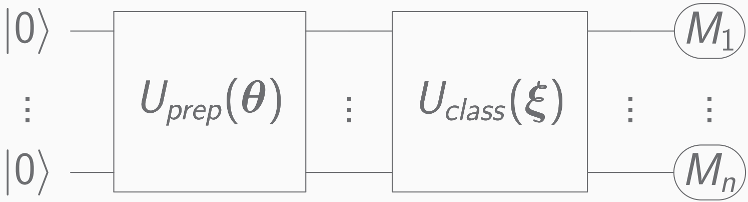


Max possible sustainable ebits generated between two regions connected by a quantum circuit is limited by the number of gates needed to cut to isolate a contiguous region, and the minimum qubits among two regions: $\min\{n_1, n_2, w\}$.



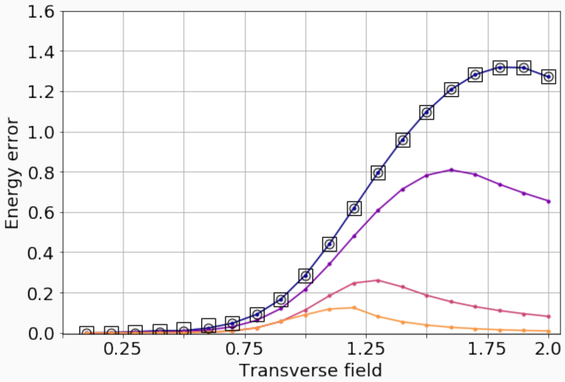
Quantum classifier

A parametrized circuit can work as a classifier. Prepare the input state, then feed it in the classifier circuit, then measure all qubits (in any basis):



Treat the average output as an assigned label and minimize log loss to fit the labels of the data. Quantum data feed to quantum classifier.

VQE approximation quality



Energy error of VQE solutions for the transverse field Ising model. Hollow squares: rank-1 ansatz, hollow circles: tree tensor network, filled circles: checkerboard states (darkest: 1 layer, brightest: 4 layers).

More layers of checkerboard $\rightarrow$ better approximation. Saturates entanglement scaling upperbound.

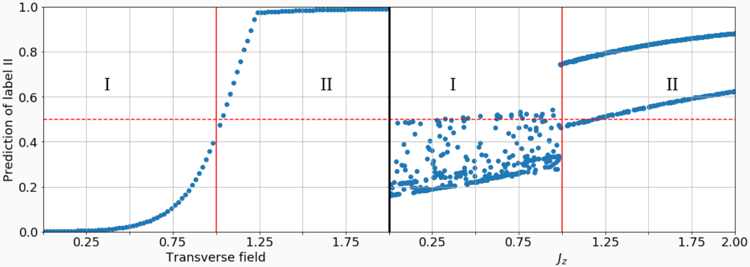


Figure 4: Output classification labels. Left: transverse field Ising model. Right: XXZ model

99% accuracy for transverse field Ising model (4 layers of checkerboard)

94% accuracy for Heisenberg XXZ model (6 layers of checkerboard)

Part II. How far can simulators be pushed?

4. Universality of variaitonal quantum computation?

Universal Variational Quantum Computation

Jacob Biamonte

in review (2019) `arXiv:1903.04500`

Proof Strategy

1. Stability theorem (relates expected energy with state-overlap)
2. Telescopes (the backbone construction)

2.a Clifford invariance of penalty functions

2.b Existence of an accepting sequence (telescopes)

3. Modified Feynman-Kitaev construction

3.a Set input with a telescope

3.b Prove gap

3.c Prove log qubit clock

3.d Prove boosting lemma $\rightarrow$ show existence of an accepting sequence

Hence or otherwise, prove that variational quantum computation is universal.

Lemma (Telescoping Lemma)

Let $\Pi_l U_l |0\rangle^{\otimes n}$ be an l-gate quantum circuit preparing state $|\psi\rangle$ on n-qubits and containing not more than $O(poly(\ln n))$ non-Clifford gates.

Then there exists a non-degenerate Hamiltonian $\mathcal{H} \geq 0$ on n-qubits with $poly(l, n)$ terms, gap Δ and ground eigenvector $|\psi\rangle \propto \Pi_l U_l |0\rangle^{\otimes n}$. In particular, if

$$0 \leq \langle\phi| \mathcal{H} |\phi\rangle < \Delta \tag{17}$$

then stability follows as

$$1 - \frac{\langle\phi| \mathcal{H} |\phi\rangle}{\Delta} \leq |\langle\phi|\psi\rangle|^2 \leq 1 - \frac{\langle\phi| \mathcal{H} |\phi\rangle}{Tr\{H\}}. \tag{18}$$

Clocking Gates in Variational Quantum Computation

Theorem

Given a quantum circuit of L gates on n-qubits producing state $\Pi_l U_l \left|0\right\rangle^{\otimes n}$*, there exists an objective function (Hamiltonian,* $\mathcal{H}$*) with* $O(L^2)$ *terms, non-degenerate ground state and spectral gap* $\Delta \geq O(L^{-2})$ *acting on* $n + O(\ln_2(L))$ *qubits such that* $\arg\min\{\mathcal{H}\}$ *with some* $O(L^{-1})$ *trials produces* $\Pi_l U_l \left|0\right\rangle^{\otimes n}$*.*

A new model of quantum computation

arXiv:1903.04500

1. Variational quantum computation admits a universal model
2. Scalars in the Hamiltonian implemented in classical evaluation of objective function
3. Unlike adiabatic quantum computation, k-body Hamiltonian terms are implemented as a simplistic measurement: support of penalty functions in the Pauli basis invariant under Clifford operations
4. Iterative process. Consider some $q < L$ gates. Preform q gates, causing the penalty to accept. Variational reduce sequence length to implement q gates. Increase $q + 1$

Papers we didn't have time to discuss

1. **Experimental neural network enhanced quantum tomography**
   Adriano Macarone Palmieri, Egor Kovlakov, Federico Bianchi, Dmitry Yudin, Stanislav Straupe, Jacob Biamonte, Sergei Kulik
   in review (2019) arXiv:1904.05902
2. **Learning Tensor Network States on a Quantum Computer**
   Jacob Biamonte
   in review (2019) arXiv:1804.02398
3. **Deep Learning Super-Diffusion in Multiplex Networks**
   Vito M. Leli, Saeed Osat, Timur Tlyachev, Jacob D. Biamonte
   in review (2019) arXiv:1811.04104

Further Discussion

Jacob D. Biamonte
Deep Quantum Laboratory
Skolkovo Institute of Science and Technology
3 Nobel Street
Moscow Russia 143026

J.Biamonte@Skoltech.ru
http://Quantum.Skoltech.ru

Thank you for your attention!