

Experimental pairwise entanglement estimation for an N -qubit system

A machine learning approach for programming quantum hardware

Nathan L. Thompson · N.H. Nguyen ·
E. C. Behrman · James E. Steck

Received: date / Accepted: date

Abstract Designing and implementing algorithms for medium and large scale quantum computers is not easy. In previous work we have suggested, and developed, the idea of using machine learning techniques to train a quantum system such that the desired process is “learned,” thus obviating the algorithm design difficulty. This works quite well for small systems. But the goal is macroscopic physical computation. Here, we implement our learned pairwise entanglement witness on Microsoft’s Q#, one of the commercially available gate model quantum computer simulators; we perform statistical analysis to determine reliability and reproducibility; and we show that after training the system in stages for an incrementing number of qubits (2, 3, 4, ...) we can infer the pattern for mesoscopic N from simulation results for three-, four-, five-, six-, and seven-qubit systems. Our results suggest a fruitful pathway for general quantum computer algorithm design and for practical computation on noisy intermediate scale quantum devices.

Keywords quantum machine learning · entanglement · quantum gates · quantum simulator

1 Introduction

For several decades now the prospect of macroscopic quantum computers, able to solve large classes of difficult problems, has been “ten years away.” We do

N.L. Thompson · N.H. Nguyen · E.C. Behrman
Department of Mathematics, Statistics and Physics, Wichita State University, Wichita, KS 67260-0033
Tel.: +13169783953
E-mail: behrman@math.wichita.edu

J.E. Steck
Department of Aerospace Engineering, Wichita State University, Wichita, KS 67260-0044

have thousand-qubit size “quantum annealing” machines [1], to solve optimization problems through adiabatic evolution to the ground state of a designed Hamiltonian, but programmable quantum computers remain small and their applicability limited. One major obstacle is the construction of algorithms that take advantage of the fundamental quantum nature of reality. There are still only a very few. Most fall into one of three categories: those using a quantum Fourier transform, like Shor’s algorithm [2]; those using amplitude amplification, like Grover’s algorithm [3]; and those using quantum walks [4]. Speedup varies: Shor’s, and some quantum walk algorithms, provide an exponential advantage over the best known classical algorithm in each case, but the speedup with Grover is only quadratic. We do not yet know whether there exists any quantum advantage for broad classes of problems [5,6], much less, what it will be in each case. Nor do we have a general process to factor an arbitrary N -qubit unitary efficiently to generate the quantum machine language necessary, in the case of the gate model; or to design a Hamiltonian whose ground state will be the answer to an optimization problem, in the case of quantum annealing.

For some time now our research group has been investigating the advantages of a marriage of machine learning and quantum computing to answer this need [7,8,9]. The basic idea is that a quantum system can itself act as a neural network: The state of the system at the initial time is the “input”; a measurement on the system at the final time is the “output”. If we know enough about the computation desired to be able to construct a comprehensive set of input-output pairs from which the net can generalize, then, we can use techniques of machine learning to bypass the algorithm-construction problem. Moreover, this approach is scalable [10] as we are able to train iteratively in larger and larger stages and use knowledge of a smaller system to make systematic inferences about a larger one. In addition, our method promises to be generally robust to both noise and to decoherence [11,12]. Machine learning may also be helpful in the factorization problem [13], and in the Hamiltonian design problem [14].

Entanglement estimation is a good example of an intrinsically quantum calculation for which we have no general algorithm. Indeed, it has been shown that the quantum separability problem (determination of entanglement) is NP-hard [15]. In previous work we succeeded in mapping a function of a measurement at the final time to a witness of the entanglement of a two-qubit system in its initial state [9]. The “output” (result of the measurement of the witness at the final time) will change depending on the time evolution of the system, which is of course controlled by the Hamiltonian: by the tunneling amplitudes $\{K\}$, the qubit biases $\{\varepsilon\}$, and the qubit-qubit coupling ζ . Thus we can consider these functions $\{K_A, K_B, \varepsilon_A, \varepsilon_B, \zeta\}$ to be the “weights” to be trained. We then use a quantum version [9] of backpropagation [16] to find optimal functions such that our desired mapping is achieved. Full details are provided in [8,9]. From a training set of only four pure states, our quantum neural network successfully generalized the witness to large classes of states, mixed as well as pure [10]. Qualitatively, what we are doing is using machine

learning techniques to find a “best” hyperplane to divide separable states from entangled ones, in the Hilbert space.

Now, this method finds a time dependent Hamiltonian that solves the given problem, a procedure more reminiscent of a quantum annealing approach [1] than the gate approach. But of course the unitary operator of time development can be represented as a product of simple gates; indeed, it is a theorem [17] that any quantum computation can be performed as a succession of simple operators belonging to any universal set. Thus, a universal quantum computer need only be able to execute each of the members of that set [18]. There is now a large number of quantum simulators available online [19], including Microsoft Quantum Development Kit [20] and IBM’s Quantum Experience [21], which implement a universal set of quantum operators (gates) plus many more that are useful in encoding quantum computations, such as the Pauli spin matrices, the Hadamard gate, the CNOT gate, and others. The difficulty arises in determining exactly how to represent a particular calculation: first, in terms of finding the unitary for that problem; and second, in terms of these gates so that algorithms may be eventually implemented on real quantum hardware.

Once we do have a unitary, there are several approaches [22,23] available for decomposing an arbitrary $2^N \times 2^N$ unitary matrix, representing a quantum computation on an N -qubit system, into “simple” gates: single qubit and the two-qubit CNOT operations implemented in the languages associated with one of the online systems. However, none of these techniques is straightforward, and often the result is a large sequence of gates to represent the desired unitary. This inherent difficulty is another reason machine learning techniques are enticing [24]: If we can determine methods where the machines themselves develop and refine the algorithms they are using, we circumvent part of this intrinsic challenge of quantum computing. In this paper, because we want to demonstrate the advantages of our stage training paradigm, we will solve the two-qubit problem by hand, then use these results to generalize to larger systems.

2 Two-qubit Quantum Neural Network

We begin with a simple two-qubit system.

2.1 Reverse Engineering of Entanglement Witness

The system evolves in time according to the Hamiltonian

$$H = K_A X_A + K_B X_B + \varepsilon_A Z_A + \varepsilon_B Z_B + \zeta Z_A Z_B \quad (1)$$

where X and Z are the Pauli operators corresponding to qubits A and B , K_A and K_B are the tunneling amplitudes, ε_A and ε_B are the biases, and ζ is the

qubit-qubit coupling. The state of the system as a function of time can then be written, for a pure state, as

$$|\psi(t)\rangle = \exp\left(\frac{-iHt}{\hbar}\right) |\psi(0)\rangle. \quad (2)$$

It is convenient to consider the Hamiltonian H as a sum of single qubit and two-qubit operations

$$H = \underbrace{K_A X_A + \varepsilon_A Z_A}_{H_A} + \underbrace{K_B X_B + \varepsilon_B Z_B}_{H_B} + \underbrace{\zeta Z_A Z_B}_{H_{AB}}. \quad (3)$$

We now consider the evolution to be broken into several ‘‘time chunks’’ where the parameters $\{K_A, K_B, \varepsilon_A, \varepsilon_B, \zeta\}$ are held constant on each interval. For most of the paper, we will use four time chunks or intervals. We can approximate the operator as the product of several operators as follows:

$$\exp\left(\frac{-iHt}{\hbar}\right) = \prod_{k=0}^3 \exp\left(\frac{-iH_k t}{4\hbar}\right), \quad (4)$$

where on each time chunk k the operator is approximated using the first-order Trotter-Suzuki formula [25]

$$\exp\left(\frac{-iH_k t}{4\hbar}\right) = \exp\left(\frac{-i(H_A + H_B + H_{AB})_k t}{4\hbar}\right) \quad (5)$$

$$\approx \exp\left(\frac{-iH_{A,k} t}{4\hbar}\right) \exp\left(\frac{-iH_{B,k} t}{4\hbar}\right) \exp\left(\frac{-iH_{AB,k} t}{4\hbar}\right). \quad (6)$$

This last equation is only approximate, because while H_A and H_B commute, neither commutes with H_{AB} . The Trotter-Suzuki formula introduces an error of the type

$$\exp(d(P + Q)) = \exp(dP) \exp(dQ) + O(d^2) \quad (7)$$

for a scalar parameter d and non-commuting ($[P, Q] \neq 0$) operators P and Q . In our application, this d corresponds to the time evolution variable, which is measured on the order of nanoseconds and should not introduce much error. During each interval or time chunk the functions $\{K_A, K_B, \varepsilon_A, \varepsilon_B, \zeta\}$ are constant, so, we may, for a given time interval $\Delta t = t/4$, rewrite the operator given by (6) as a product of physically realizable quantum gates, such as those implemented in the Q# or Qiskit languages. Comparisons between the final density matrices made in the training process and the full gate based decomposition as well the effect of a finer discretization are discussed in Sections 2.2 and 4.2, respectively.

We start with the single qubit part of the operator for a single time interval, $\exp(-iH_A \Delta t/\hbar)$, and use the well-known identity for the exponential of Pauli matrices

$$e^{-i\alpha(\hat{\mathbf{n}} \cdot \boldsymbol{\sigma})} = I \cos(\alpha) - i(\hat{\mathbf{n}} \cdot \boldsymbol{\sigma}) \sin(\alpha) \quad (8)$$

where I is the 2×2 identity matrix, α is an angle of rotation about axis $\hat{\mathbf{n}}$ (a unit vector) on the Bloch sphere, and $\boldsymbol{\sigma}$ is a vector of the Pauli matrices $\{X, Y, Z\}$. Looking at the definition of H_A (or H_B) in equation (3), we see that it is easy to express the exponent in the form of representation (8):

$$\begin{aligned} \frac{\Delta t}{\hbar} H_A &= \frac{\Delta t}{\hbar} (K_A X_A + 0 Y_A + \varepsilon_A Z_A) \\ &= \underbrace{\frac{\Delta t}{\hbar} \sqrt{K_A^2 + \varepsilon_A^2}}_{\alpha} \underbrace{\left(\frac{K_A}{\sqrt{K_A^2 + \varepsilon_A^2}} X_A + 0 Y_A + \frac{\varepsilon_A}{\sqrt{K_A^2 + \varepsilon_A^2}} Z_A \right)}_{\hat{\mathbf{n}} \cdot \boldsymbol{\sigma}}. \end{aligned} \quad (9)$$

Interpreting the operator as a rotation on the Bloch sphere, we have a formula for a rotation α about an axis $\hat{\mathbf{n}}$ [17]

$$R_{\hat{\mathbf{n}}}(\alpha) = R_z(\gamma) R_y(\beta) R_z(\alpha) R_y(-\beta) R_z(-\gamma), \quad (10)$$

where the rotations $R_x(\theta)$, $R_y(\theta)$, and $R_z(\theta)$ are defined as

$$\begin{aligned} R_x(\theta) &= e^{-i\frac{\theta}{2}X} = \begin{bmatrix} \cos \frac{\theta}{2} & -i \sin \frac{\theta}{2} \\ -i \sin \frac{\theta}{2} & \cos \frac{\theta}{2} \end{bmatrix} \\ R_y(\theta) &= e^{-i\frac{\theta}{2}Y} = \begin{bmatrix} \cos \frac{\theta}{2} & -\sin \frac{\theta}{2} \\ \sin \frac{\theta}{2} & \cos \frac{\theta}{2} \end{bmatrix} \\ R_z(\theta) &= e^{-i\frac{\theta}{2}Z} = \begin{bmatrix} e^{-i\frac{\theta}{2}} & 0 \\ 0 & e^{i\frac{\theta}{2}} \end{bmatrix} \end{aligned}$$

The Q# and Qiskit environments [20,21] have access to a function which computes the rotation of a state about the x , y , or z axis of the Bloch sphere by a specified angle, so this expression will suffice supposing that we can find the appropriate values for α , β , and γ in (10). To do this, we use some analogues to this expression in terms of Pauli matrices and spherical coordinates:

$$\begin{aligned} R_{\hat{\mathbf{n}}}(\alpha) &= I \cos\left(\frac{\alpha}{2}\right) - i(\hat{\mathbf{n}} \cdot \boldsymbol{\sigma}) \sin\left(\frac{\alpha}{2}\right) \\ &= I \cos\left(\frac{\alpha}{2}\right) - i(\sin \beta \cos \gamma X + \sin \beta \sin \gamma Y + \cos \beta Z) \sin\left(\frac{\alpha}{2}\right). \end{aligned} \quad (11)$$

Our expression (9) matches (11) perfectly, and now we need only solve the following system of three equations with three unknowns:

$$\sin \beta \cos \gamma = \frac{K_A}{\sqrt{K_A^2 + \varepsilon_A^2}}, \quad \sin \beta \sin \gamma = 0, \quad \cos \beta = \frac{\varepsilon_A}{\sqrt{K_A^2 + \varepsilon_A^2}}. \quad (12)$$

We notice immediately that $\sin \gamma = 0$ (since $\sin \beta$ cannot be zero due the first equation), and so $\gamma = c\pi$ for some integer c . This forces $\cos \gamma$ to be ± 1 . Last,

$\beta = \sin^{-1} \left(\pm K_A / \sqrt{K_A^2 + \varepsilon_A^2} \right)$. We see that the relative sizes of K_A and ε_A are constrained by the sine and cosine relationship between

$$\sin \beta = \pm \frac{K_A}{\sqrt{K_A^2 + \varepsilon_A^2}} \quad \cos \beta = \frac{\varepsilon_A}{\sqrt{K_A^2 + \varepsilon_A^2}}.$$

A change of indices gives us the operator H_B in a similar way. The only remaining step is to write H_{AB} of (3) in a form using practical quantum gates.

The two-qubit part of the Hamiltonian is $H_{AB} = \zeta Z_A Z_B$. The matrix form of this operator is generated by taking the Kronecker product, $Z_A Z_B = Z \otimes Z$. After setting $w_0 = \zeta \Delta t / \hbar$ and taking the exponential of the operator, we have

$$\exp \left(\frac{-i H_{AB} \Delta t}{\hbar} \right) = \begin{bmatrix} e^{-i w_0} & 0 & 0 & 0 \\ 0 & e^{i w_0} & 0 & 0 \\ 0 & 0 & e^{i w_0} & 0 \\ 0 & 0 & 0 & e^{-i w_0} \end{bmatrix}. \quad (13)$$

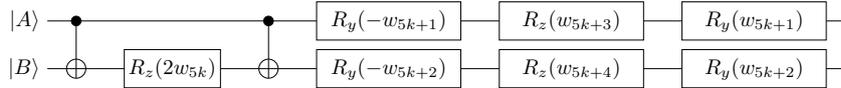
Since this is a two-qubit operator, it is necessary to represent it using a two-qubit gate. The primary tool for this is the CNOT (controlled NOT) gate,

$$\text{CNOT} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \quad (14)$$

In general the CNOT operator in addition to single-qubit phase gates forms a universal set with which to build an arbitrary (N -qubit) operator. For our purposes, we may represent matrix (13) using the following expression:

$$\exp \left(\frac{-i H_{AB} \Delta t}{\hbar} \right) = \text{CNOT} \begin{bmatrix} e^{-i w_0} & 0 & 0 & 0 \\ 0 & e^{i w_0} & 0 & 0 \\ 0 & 0 & e^{-i w_0} & 0 \\ 0 & 0 & 0 & e^{i w_0} \end{bmatrix} \text{CNOT}. \quad (15)$$

The interior matrix is $I \otimes R_z(2w_0)$, which is a rotation on only the B qubit. With the above decompositions for H_A , H_B , and H_{AB} , we can now express each time chunk of our quantum operator in terms of a quantum circuit



where we have relabeled variables as $w_{5k} = \frac{\zeta_k \Delta t}{\hbar}$,

$$w_{5k+1} = \sin^{-1} \left(\frac{K_{A,k}}{\sqrt{K_{A,k}^2 + \varepsilon_{A,k}^2}} \right), \quad w_{5k+2} = \sin^{-1} \left(\frac{K_{B,k}}{\sqrt{K_{B,k}^2 + \varepsilon_{B,k}^2}} \right),$$

$$w_{5k+3} = \frac{\Delta t}{\hbar} \sqrt{K_{A,k}^2 + \varepsilon_{A,k}^2}, \quad \text{and} \quad w_{5k+4} = \frac{\Delta t}{\hbar} \sqrt{K_{B,k}^2 + \varepsilon_{B,k}^2}.$$

Collected formulaically, the gate decomposition of operator (4) is

$$\exp\left(\frac{-iHt}{\hbar}\right) = \prod_{k=0}^3 U_{A,k} U_{B,k} U_{AB,k} \quad (16)$$

where

$$U_{A,k} = [R_y(w_{5k+1})R_z(w_{5k+3})R_y(-w_{5k+1})] \otimes I \quad (17)$$

$$U_{B,k} = I \otimes [R_y(w_{5k+2})R_z(w_{5k+4})R_y(-w_{5k+2})] \quad (18)$$

$$U_{AB,k} = \text{CNOT} [I \otimes R_z(2w_{5k})] \text{CNOT}. \quad (19)$$

2.2 Numerical computation

In our original work [8,9,10] on the entanglement witness, we used piecewise constant functions for $\{K_A, K_B, \varepsilon_A, \varepsilon_B, \zeta\}$; in subsequent work [26] we used continuum functions, for which we found training was much more rapid and complete. Because current technology does not allow for continuous-time control of gate functions, we return to our original piecewise formulation; however, we have retrained using our more recent codes to improve our earlier results.

Physically, we imagine that the system would be allowed to evolve for a specified time under a Hamiltonian whose parameter functions we could control. At the end of that time we perform a measurement whose average value would represent the entanglement witness. The training of the net is a process whereby we find an optimal mapping of the desired physical property (here, the entanglement) to that chosen measurement. We chose, as that measurement, the (square of the) qubit-qubit correlation function at that final time, $\langle Z_A(t_f)Z_B(t_f) \rangle^2$.

To perform the retraining, we used our (newer) continuum codes implemented in MATLAB [27], following the general approach of [9] with some modifications. Each parameter $\{K_A, K_B, \varepsilon_A, \varepsilon_B, \zeta\}$ is represented as a Fourier series and allowed to evolve for a specified time under the prescribed Hamiltonian. After each epoch (single pass through the whole training set) of training, the parameter functions are then approximated as piecewise constant functions obtained by averaging over each of the four time chunks. Next, these piecewise constant functions are used as the initial parameters for the following training epoch. Each time, the parameters functions are allowed to evolve as Fourier series, but are “chunked” and averaged back into a piecewise constant form. After a sufficient amount of training, we use the piecewise constant parameters for time evolution to calculate the expectation value of the final time correlation function, and, therefore, the error. Training data are shown in Table 1. The “Desired” column is the goal of the training for the final time correlation function, showing that we seek a value of one for a fully entangled state and zero for a product state. Because we are trying to optimize our entanglement witness, we find a value intermediate between zero and one for the target value for the partially entangled state $|P\rangle$; this (optimized) value is 0.443.

The column labelled “Trained” shows the asymptotic value for that final time correlation function after the training of the network. Training was of course less efficient than with the greater flexibility offered by the continuous-time functions; nonetheless, RMS error for the training set was only 0.05% after 200 epochs. The piecewise constant values found for the parameter functions are shown in Table 2.

Table 1 QNN entanglement witness trained for 200 epochs using piecewise constant parameter functions, and compared with calculated results using first chunked time propagators, then with the sequence of gates, and finally on the Q# simulator [20]. The training set of four [9] includes one completely entangled state $|\text{Bell}\rangle = \frac{1}{\sqrt{2}}[|00\rangle + |11\rangle]$, one unentangled state $|\text{Flat}\rangle = \frac{1}{2}[|00\rangle + |01\rangle + |10\rangle + |11\rangle]$, one classically correlated but unentangled state $|\text{C}\rangle = \frac{1}{\sqrt{5}}[2|00\rangle + |01\rangle]$, and one partially entangled state $|\text{P}\rangle = \frac{1}{\sqrt{3}}[|01\rangle + |10\rangle + |11\rangle]$. Errors for each method are shown in the final line.

| Input state | Desired | Trained | Chunked | Gates | Q# |
|-----------------------|---------|-----------------------|-----------------------|-----------------------|-----------------------|
| $ \text{Bell}\rangle$ | 1.0 | 0.999 | 0.999 | 0.999 | 0.999 |
| $ \text{Flat}\rangle$ | 0.0 | 7.99×10^{-5} | 5.99×10^{-7} | 5.99×10^{-7} | 6.14×10^{-5} |
| $ \text{C}\rangle$ | 0.0 | 1.08×10^{-4} | 1.87×10^{-5} | 1.87×10^{-5} | 8.17×10^{-5} |
| $ \text{P}\rangle$ | 0.443 | 0.440 | 0.446 | 0.446 | 0.446 |
| Total RMS error | | 5.0×10^{-4} | 1.4×10^{-3} | 1.4×10^{-3} | 1.7×10^{-3} |

Table 2 Trained parameter functions for the entanglement witness for the two-qubit system, in MHz. Total time of evolution for the two time propagation methods was 1.58 ns.

| Parameter | Interval 1 | Interval 2 | Interval 3 | Interval 4 |
|---------------------------------|------------|------------|------------|------------|
| $K_A = K_B$ | 2.49 | 2.47 | 2.48 | 2.49 |
| ζ | 0.0382 | 0.128 | 0.117 | 0.0382 |
| $\varepsilon_A = \varepsilon_B$ | 0.0930 | 0.116 | 0.0954 | 0.0833 |

We now use these trained values for the piecewise constant parameter functions in the equations derived for the sequence of operators in the previous section. Note that there are two separate sources for the error: the approximation in Equation (6), which assumes that the matrices commute; and the approximation of the substitution of the products of the gate operators for the time-propagation operator. We can separate these two sources by calculating the density matrix for the final time, using “chunked time.” That is, instead of calculating the time propagation correctly as in the QNN training, we separate the Hamiltonian into H_A , H_B , and H_{AB} for each of the four time intervals. This assumes that the pieces commute, which of course is an approximation. The column in Table 1 labelled “Chunked,” shows the calculation of the entanglement witness using this approximation. The “Gates” column re-

peats these calculations using the matrix decomposition outlined in Section 2.1. The final column, labelled “Q#”, shows the entanglement witness values of the sequence of gates as measured on Microsoft’s quantum simulator [20]. (Calculations performed using IBM’s Quantum Experience simulator [21] produced almost identical results [28].) Note that the calculated numbers for the entanglement witness in the two last columns are extremely close, as are, of course, the RMS errors for each method. The Frobenius norm of the difference between the density matrix as trained by the QNN technique and the (non-commuting) chunked time propagation matrices is in each case 1 to 2%; while the norm of the difference between the density matrix calculated by the chunked time propagator and by the sequence of applied gates in each case is on the order of 10^{-15} , i.e., within round-off error. Clearly all of the error comes from the non-commutation. This validates our replacement of the time evolution operator by the product of gates.

3 Statistical Evaluation of Entanglement Witness in Q#

With the entanglement witness properly reverse engineered to run on the hardware simulators, we now need to understand how to utilize it in applied situations. Both the Q# and Qiskit systems implement measurements of the qubit along a standard axis x , y , or z in the Bloch sphere. Each individual measurement only returns an eigenvalue of ± 1 . To generate a useable expected value, several thousand measurements must be done to average these eigenvalues to get a valid approximation of the expectation value $\langle Z_A(t_f)Z_B(t_f) \rangle^2$. Using the Q# built-in simulator, we did 100 iterations at several different “shot counts” (number of individual experiments and measurements) to gauge how many times a particular experiment must be run to generate a high confidence value for the entanglement witness. Our code is available at [29].

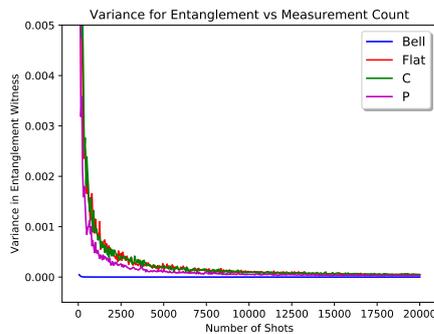


Fig. 1 Variance in entanglement witness for 100 iterations of each state measured at shot counts ranging from 50 to 20,000 in 50 shot increments. As the shot count increases, we see that the measurement variance quickly goes to zero.

Figure 1 shows the variance of the expectation value $\langle Z_A(t_f)Z_B(t_f) \rangle^2$ as a function of numbers of shot counts. We can see plainly that the law of large numbers is in effect for determining the entanglement witness. High confidence values for the witness are achieved near 15,000 iterations of the experiment. This is easier to see in Figure 2, which shows a 95% confidence interval surrounding the computed square of the qubit-qubit correlation for the witness on the $|\text{Bell}\rangle$ and $|\text{P}\rangle$ states, where the width of the interval shrinks to 0.0015. Results for the $|\text{Flat}\rangle$ and $|\text{C}\rangle$ states are similar.

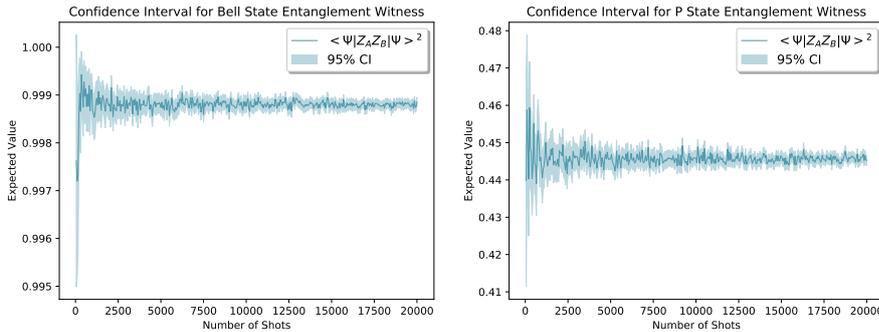


Fig. 2 Q# entanglement witness values for the $|\text{Bell}\rangle$ and $|\text{P}\rangle$ states with a 95% confidence interval as a function of the shot count. The confidence interval (CI) width reaches its minimum of ~ 0.0015 after approximately 15,000 shots.

4 Iterative Staging

We have constructed a sequence of hardware gates that mimics our trained two-qubit entanglement witness quite well. While this is interesting it is perhaps of somewhat limited use, as it pertains only to a two-qubit system. We now extend our results to an N -qubit system.

4.1 Searching for an Asymptotic Limit

As demonstrated in our previous work [8,9,10,26], it is both possible and very beneficial to use knowledge of smaller scale systems to make systematic inferences about larger ones. Hence, we train our network in iterative stages, using the trained parameter functions for the two-qubit system $\{K_A = K_B, \varepsilon_A = \varepsilon_B, \zeta\}$ as an initial guess for those parameters in the three-qubit case. We then train the three-qubit system from that point to minimize the error. Once the three-qubit trained functions are found, we start from those to train the four-qubit system, and so on. The benefit is that while there are large changes in the tunneling, bias, and coupling parameters as the system size increases initially, those percentage changes diminish as the system size N increases,

due to the increased connectivity. Hence, training five-, six-, and seven-qubit systems require fewer and fewer additional epochs. Because of the symmetry of the problem, all the parameter functions can be taken to be the same (that is, $K_A(t) = K_B(t) = K_C(t)$, $\varepsilon_A(t) = \varepsilon_B(t) = \varepsilon_C(t)$, and so on); imposing this as a constraint also reduces the training time. We now look for an asymptotic limit as N increases. Figures 3 and 4 show the results of the training.

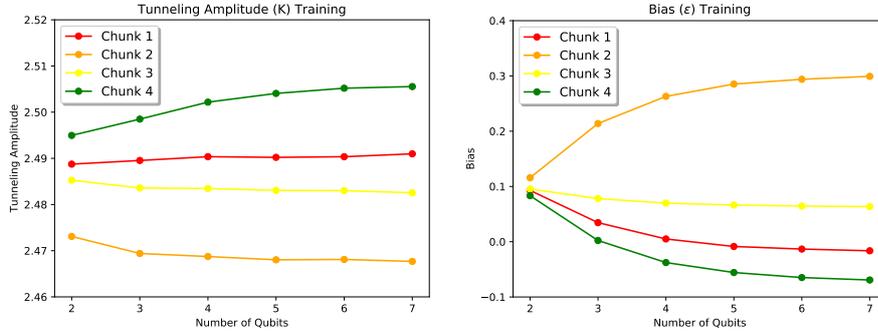


Fig. 3 Trained values for the tunneling amplitude K and for the bias ε , for each time chunk, as the number of qubits in the system is increased. Both demonstrate clear asymptotic behavior.

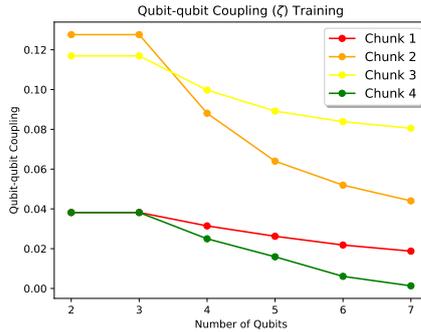


Fig. 4 Trained values for the qubit-qubit coupling ζ , for each time chunk, as the number of qubits in the system is increased. The values show a clear trend, but do not become asymptotic as quickly as with the other parameters.

All parameters show an asymptotic trend, with the tunneling amplitudes K and biases ε showing swift convergence to a limiting value. The qubit-qubit coupling ζ also has a trend emerging at the number of qubits increases, indicating that an N -qubit limit is likely. We infer that the parameters for the seven-qubit system are a reasonable approximation for the entanglement witness of an N -qubit system based on the limiting behavior observed in K and

ε . This is important, because once quantum computers become only a very little larger we will no longer be able to simulate them on classical computers (the point of so-called “quantum supremacy” [30].) Table 3 contains the parameters for the fully symmetric seven-qubit system.

Table 3 Trained parameter values at each time interval, for the pairwise entanglement witness for the seven-qubit system, in MHz. By symmetry, each of the tunneling functions K , each of the biases ε , and each of the pairwise couplings ζ is the same. We take these values to be an approximation to the asymptotic limit of the parameters for an N -qubit quantum system.

| Parameter | Interval 1 | Interval 2 | Interval 3 | Interval 4 |
|---------------|------------|------------|------------|------------|
| K | 2.49 | 2.47 | 2.48 | 2.51 |
| ζ | 0.0188 | 0.0440 | 0.0805 | 0.00132 |
| ε | -0.0164 | 0.299 | 0.0636 | -0.0693 |

Training for these parameters was relatively efficient for an N -qubit system, taking only 100 additional epochs past the previously trained $(N - 1)$ -qubit system to train the pairwise entanglement witness. While the number of training pairs, $4\binom{N}{2}$, does increase with the number of qubits, the increased connectivity meant that the system needed less additional training each time. The total RMS value for the training of the two-qubit parameters is 6.0×10^{-4} , and only increased slightly as qubits were added, with six-qubits having an RMS of 1.6×10^{-3} (at 60 training pairs) and seven-qubits 1.8×10^{-3} (84 training pairs). Mesoscopic systems will still require some training to decrease initial errors, but this amount should be very small or negligible since we already see the parameters nearing asymptotic values, and we anticipate that this (small) additional training can be done online and need not be simulated.

4.2 Comparing the Discrete and Continuum Cases

Having established the scalability of our results in terms of growing system size, we now show how the results for the chunked system compare to our more sophisticated model for the entanglement witness studied in [26]. In that work, the tunneling, bias, and coupling parameters were all modeled using continuous functions of time. Allowing continuous parameters added a great deal more flexibility and assisted training immensely. This approach to quantum machine learning resulted in smaller errors and faster training than the piecewise constant “chunked” model. However, with the current gate based model of quantum computing, we have no expectation of being able to implement or train a continuum parameter solution on developing or proposed hardware. Therefore, we examine the relationship between entanglement witnesses built and trained using the chunked and continuous versions of the K , ε , and ζ parameters.

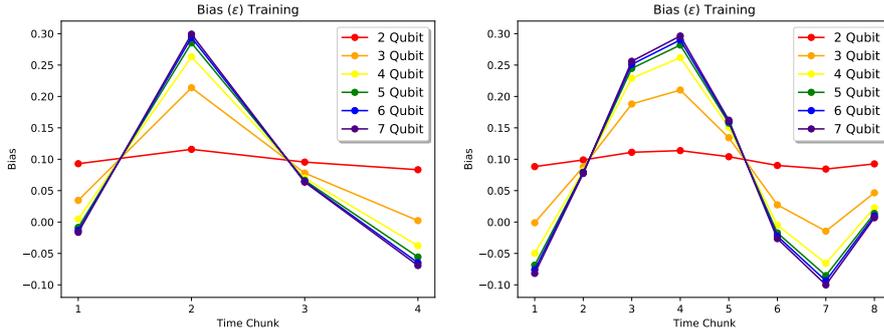


Fig. 5 Trained bias, for 4 and 8 time chunks, as functions of time, for systems of increasing numbers of qubits N .

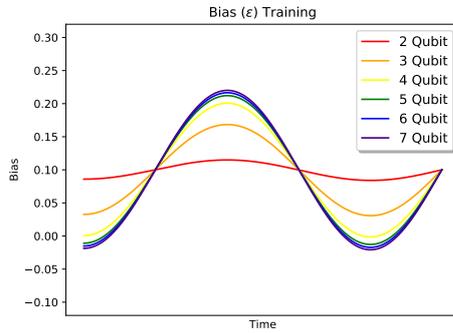


Fig. 6 Trained bias functions for the continuum model of the entanglement witness, for systems of increasing numbers of qubits N . Note how the graphs in Figure 5 are close approximations of the shapes and values of the function for each number of qubits.

Figure 5 shows the bias as a function of time in the 4 chunk model, for two- through seven-qubit systems, compared with a similar 8 chunk model, where the time chunks or intervals were halved. In each case we see that as the number of qubits increases a similar curve takes shape as the bias function seems to reach asymptotic values. Figure 6 is the continuum case [26], for the same two- through seven-qubit systems. Each system was trained using the techniques outlined in [10, 26] with the imposed symmetry constraints as discussed in the previous section. We observe that the 4 and 8 chunk cases show strong qualitative resemblance to the shape of the parameter function in the continuum case. Quantitatively, the exact values of the discretized functions and the continuum model do not match, but the disagreement is small and the overshooting can be attributed to fitting error. In previous work we have shown [11, 12] both that the calculation is relatively insensitive to the exact values of the time dependent functions and that as the system size N is increased that robustness increases; thus the disagreement is probably irrelevant and in any

case becomes more irrelevant with increasing N . Total RMS error for each of these simulations is shown in Table 4.

| Qubits | RMS error | | |
|--------|----------------------|----------------------|----------------------|
| | 4 Chunk | 8 Chunk | Continuous |
| 2 | 6.0×10^{-4} | 6.4×10^{-4} | 5.2×10^{-4} |
| 3 | 6.0×10^{-4} | 8.8×10^{-4} | 1.2×10^{-3} |
| 4 | 4.2×10^{-4} | 1.2×10^{-3} | 4.3×10^{-4} |
| 5 | 3.6×10^{-4} | 1.0×10^{-3} | 4.1×10^{-4} |
| 6 | 1.6×10^{-3} | 1.2×10^{-3} | 6.4×10^{-4} |
| 7 | 1.8×10^{-3} | 1.4×10^{-3} | 6.6×10^{-4} |

Table 4 Total RMS error for the training set in the 4, 8, and continuum models of the entanglement witness for system sizes ranging from two- to seven-qubits. Training followed the methods of [10,26] with the additional condition that all parameters are fully symmetric. The continuum model shows the best accuracy, but the discretized versions also trained well and are viable approximations of the continuum model (not realizable in the current hardware.)

5 Conclusions

As an example of using machine learning techniques to train quantum systems to do computations for which no algorithm is known, we have trained a system of qubits to return a witness to its initial pairwise entanglement, by manipulating parameters in a time-dependent Hamiltonian. The training process utilized classical computation performed in MATLAB which were later verified using the Q# simulator to determine that the computation was in fact implementable in the gate model decomposition, returned results consistent with error bounds projected by our training, and generated usable results without requiring an extremely large number of experiments (shots) within the hardware simulation.

This procedure is reminiscent of a physical setup like the quantum annealing processors [1], which have a time-dependent Hamiltonian (though the parameter flexibility is still severely limited.) But the approach outlined in this paper is a kind of bridge between the annealing and gate approaches to quantum computing: with systematic Hamiltonian design, QA computers could be used as programmable machines as well [14]. The entanglement witness was well approximated by a series of implementable gates. A thorough statistical analysis was done, and a good confidence interval of about 0.0015 is reached after 15,000 shots. The discretized parameter setup models the entanglement witness accurately with respect to two kinds of scaling: increasing the number of qubits and increasing the number of time chunks in the piecewise-constant parameter functions. Agreement was excellent, and the calculation generalizes well and easily as the number of qubits N increases to seven.

Two of the parameter functions so learned seem already to have reached an asymptote, which would mean that the witness could probably be implemented

with only small error for much larger values of N , or, at minimum, could be trained online with little effort [31]. The qubit-qubit coupling could not definitively be said to have reached its asymptote, but it is at least plausible that a large fraction of the training necessary has already been accomplished, and, again, has reduced the amount of further training necessary. This potential training reduction is significant because in the near term we can expect that quantum computers will both be noisy and have to operate with a severely limited number of ancillary qubits. Both noise and decoherence continue to be problematic, and with only between 50 to a few hundred qubits total, practical computations cannot afford even the minimum five ancilla per correction [35]. Our research suggests another approach for these noisy, intermediate-scale quantum (NISQ) [32] devices: To perform offline (simulation) training followed by online (physical device) training using reinforcement learning [31] or automatic differentiation [33] to fine tune the system parameters. Our work here strongly suggests that the amount of online training necessary should not be large; moreover, that our quantum machine learning technique should be robust to both noise and to decoherence [11, 12], which will be a great advantage. We are currently working on determining the most promising approach [34]. And, while we have done this work using the calculation of a pairwise entanglement witness as exemplar, there is no reason to think that our results are unique to that particular calculation: in all probability the technique presented here could be used for a general scaleup paradigm.

Physical implementation still poses some problems. There are major limitations in both connectivity and decoherence with the target hardware. For viable hardware implementation, the main consideration is the computational fidelity. Fidelity is lost to both time and inefficient computations. On the available IBM hardware, coherence times are approximately $60 \mu\text{s}$ for both depolarization and spin dephasing [36]. The time required to apply a single-qubit gate is about $0.130 \mu\text{s}$ and two-qubit gates are between $0.250 \mu\text{s}$ and $0.450 \mu\text{s}$. Any state preparation and quantum circuit operations must be completed within the $60 \mu\text{s}$ interval. Our implementation of the chunked pairwise entanglement witness uses 28 single-qubit and 8 two-qubit gates, which yields a smaller than $8 \mu\text{s}$ total time (plus up to $2 \mu\text{s}$ to prepare a state); despite this, reproducibility on IBM hardware was not good [28]. Gate fidelity also affects computations. Single qubit readouts are accurate 96% of the time, and single- and two-qubits maintain fidelity at a rate of 99.7% and 96.5%, respectively [36]. Available hardware and circuit implementation techniques will of course improve. Developers are working on two important avenues to combat decoherence: higher fidelity physical implementation of quantum gates [37, 38], and the reduction of the so-called T -depth for circuits [23, 39]. The value of fidelity increases are obvious, and reducing the physical time required to perform the operations of a circuit will improve computational accuracy. It should be noted that as the coherence times of the hardware improve, our training paradigm increases in value as we can give better models with finer discretizations of the continuum training for our entanglement witness, and, of

course, other desired calculations. Moreover, machine learning solutions may also have robustness advantages [11, 12].

Optimization of the discretization of universal circuits for operators involving very small numbers of qubits at a time is a major advance towards universal quantum computation. But it is not the whole answer. For one thing, many times we do not know the unitary operator that will perform the computation, since we do not have an algorithm. For another, we still do not have optimal ways of reducing an N qubit unitary to building blocks involving only one or two qubits. Machine learning holds a great deal of promise for both tasks. Our work here seems to show that with iterative staging we can fairly easily extend small simulational results to larger systems. And, even when a unitary is known that performs the desired calculation, a clever neural network approach may find one with more efficiency or better speedup [40].

Acknowledgements We thank Patrick Coles (LANL) and William Ingle (WSU), for helpful discussions, and Henry Elliott (WSU) for the comparative Qiskit [21] calculations and hardware implementation.

References

1. K. Karimi, N.G. Dickson, F. Hamze, M.H.S. Amin, M. Drew-Brook, F.A. Chudak, P.I. Bunyk, W.G. Macready, and G. Rose, *Investigating the performance of an adiabatic quantum optimization processor*, Quantum Inf. Process. **11**, pp. 77-88 (2012).
2. P.W. Shor, *Algorithms for quantum computation: discrete logarithms and factoring*, Proceedings of the 35th Annual Symposium on Foundations of Computer Science (IEEE) (1994).
3. L.K. Grover, *A fast quantum mechanical algorithm for data base search*, Proceedings of the 28th Annual ACM Symposium on the Theory of Computing 212, (1996).
4. A.M. Childs, R. Cleve, E. Deotto, E. Farhi, S. Gutmann, and D.A. Spielman, *Exponential algorithmic speedup by quantum walk*, Proceedings of the 35th Symposium on Theory of Computing, 59-68 (2003).
5. S. Bravyi, D. Gosset, and R. Konig, *Quantum advantage with shallow circuits*, arXiv:1704.00690v1 (2017).
6. T.F. Ronnow, Z. Wang, J. Job, S. Boixo, S.V. Isakov, D. Wecker, J.M. Martinis, D.A. Lidar, and M. Troyer, *Defining and detecting quantum speedup*, Science **345**, 420 (2014).
7. E.C. Behrman, L.R. Nash, J.E. Steck, V. Chandrashekar, and S.R. Skinner, *Simulations of quantum neural networks*, Inf. Sci. **128**, 257 (2000).
8. E.C. Behrman, V Chandrashekar, Z. Wang, C.K. Belur, J.E. Steck, and S.R. Skinner, *A quantum neural network computes entanglement*, arXiv: quant-ph/0202131 (2002).
9. E.C. Behrman, J.E. Steck, P. Kumar, and K.A. Walsh, *Quantum algorithm design using dynamic learning*, Quantum Inf. Comput. **8** pp. 12-29 (2008).
10. E. C. Behrman and J. E. Steck, *Multiqubit entanglement of a general input state*. Quantum Inf. Comput. **13**, pp. 36-53 (2013).
11. E.C. Behrman, N.H. Nguyen, J.E. Steck, M. McCann, *Quantum neural computation of entanglement is robust to noise and decoherence*, in Quantum Inspired Computational Intelligence: Research and Applications, S. Bhattacharyya, ed. (Morgan Kauffman, Elsevier) pp.3-33 (2016).
12. N.H. Nguyen, E.C. Behrman, and J.E. Steck, *Quantum learning with noise and decoherence: a robust quantum neural network*, Quantum Machine intelligence (to appear); arXiv: 1612.07593 (2019).
13. M. Swaddle, L. Noakes, L. Salter, H. Smallbone, and J. Wang, *Generating 3 qubit quantum circuits with neural networks*, Phys. Lett. A **381**, 3391 (2017).

14. E.C. Behrman, J.E. Steck, and M.A. Moustafa, *Learning quantum annealing*, Quantum Inf. Comput. **17**, 0469-0487 (2017).
15. L. Gurvitz, *Classical deterministic complexity of Edmonds problem and quantum entanglement*, in *Proc. 35th ACM Symp. on the Theory of Comput.*, pp. 10-19 (ACM Press, New York, 2003).
16. P. Werbos, in *Handbook of Intelligent Control*, Van Nostrand Reinhold, pp. 79-80 and 339-344 (1992); Yann le Cun, *A theoretical framework for back-propagation* in *Proc. 1998 Connectionist Models Summer School*, D. Touretzky, G. Hinton, and T. Sejnowski, eds., Morgan Kaufmann, (San Mateo), pp. 21-28 (1988).
17. M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press (Cambridge, England) (2001).
18. S. Lloyd, *Universal quantum simulators*. Science **273**, 1073 (1996).
19. R. LaRose, *Overview and comparison of gate level quantum software plaltforms*, arXiv: 1807.02500 (2018).
20. Microsoft Quantum Development Kit. <https://docs.microsoft.com/en-us/quantum/?view=qsharp-preview>, (2018). accessed September 2018
21. The IBM Quantum Experience. <https://quantumexperience.ng.bluemix.net/qx>, (2018). accessed October 2018
22. Y.G. Chen and J.B. Wang, *QCompiler: quantum compilation with CSF method*, arXiv:quant-ph/1208.0194v2 (2012).
23. V. Kliuchnikov, *Synthesis of unitaries with Clifford+ T circuits*, arXiv:1306.3200 (2013); V. Kliuchnikov, D. Maslov, M. Mosca, *Asmptotically optimal approximation of single qubit unitaries by Clifford and T circuits using a constant number of ancillary qubits*. Phys. Rev. Lett. **110**, 190502 (2013).
24. L. Cincio, Y. Subaşı, A.T. Sornborger, and P.J. Coles, *Learning the quantum algorithm for state overlap*, New J. Phys. **20**, 113022 (2018).
25. Hatano N., Suzuki M. *Finding exponential product formulas of higher orders*, In: Das A., K. Chakrabarti B. (eds) Quantum Annealing and Other Optimization Methods. Lecture Notes in Physics, vol 679. Springer, Berlin, Heidelberg (2005).
26. E.C. Behrman, R.E.F. Bonde, J.E. Steck, and J.F. Behrman, *On the correction of anomalous phase oscillation in entanglement witnesses using quantum neural networks*, IEEE Trans. on Neural Networks and Learning Systems **25**, pp 1696-1703 (2014).
27. MATLAB and Statistics Toolbox Release 2018a, The MathWorks, Inc., Natick, Massachusetts, United States
28. J.E. Steck, E.C. Behrman, and N.L. Thompson, *Machine learning applied to programming quantum computers*, AIAA Scitech 2019 Forum, (January 2019). DOI: 10.2514/6.2019-0956
29. Available at <https://github.com/williamingle/WichitaStateQNN>
30. J. Preskill, *Quantum computing and the entanglement frontier*, arXiv:1203.5813v3 (2013).
31. M.Y. Niu, S. Boixo, V. Smelyanskiy, and H. Neven, *Universal quantum control through deep reinforcement learning*, arXiv:1803.01857v2 (2018).
32. J. Preskill, *Quantum Computing in the NISQ era and beyond*, Quantum **2**, 79, arXiv:1801.00862v3 (2018).
33. M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, *Evaluating analytic gradients on quantum hardware*, arXiv:1811.11184 (2018).
34. N.L. Thompson, N.H. Nguyen, W. Ingle, E.C. Behrman, and J.E. Steck, *Optimized quantum machine learning*, in progress (2020).
35. E. Knill, R. Laflamme *Theory of error-correcting codes*, Phys. Rev. A **55**, 900. (1997)
36. N. M. Linke, D. Maslov, M. Rotteler, S. Debnath, C. Figgatt, K. A. Landsman, K. Wright, and C. R. Monroe, *Experimental comparison of two quantum computing architectures*, arXiv:1702.01852 (2017).
37. X. Wang, E. Barnes, S.D. Sarma, *Improving the gate fidelity of capacitively coupled spin qubits*, NPJ Quantum Information **1**, 15003, 10.1038/npjqi.2015.3 (2015).
38. B. Pokharel, N. Anand, B. Fortman, D.A. Lidar, *Demonstration of fidelity improvement using dynamical decoupling with superconducting qubits*, Phys. Rev. Lett., **121**, 220502, DOI: 10.1103/PhysRevLett.121.220502 (2016).
39. N.J. Ross and P. Selinger, *Optimal ancilla-free Clifford+T approximation of z-rotations*, arXiv: quant-ph/1403.2975 (2014).

-
40. M.J. Rethinam, A.K. Javali, A.E. Hart, E.C. Behrman, and J.E. Steck, *A genetic algorithm for finding pulse sequences for nmr quantum computing*, Paritantra – Journal of Systems Science and Engineering **20**, 32-42. arXiv:quant-ph/0404170 (2011).