Approximate supervised learning of quantum gates via ancillary qubits

Document Version:
Peer reviewed version

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

APPROXIMATE SUPERVISED LEARNING OF QUANTUM GATES VIA ANCILLARY QUBITS

LUCA INNOCENTI

Centre for Theoretical Atomic, Molecular, and Optical Physics, School of Mathematics and Physics, Queen's University Belfast, BT7 1NN
Belfast, United Kingdom
linnocenti01@qub.ac.uk

LEONARDO BANCHI

QOLS, Blackett Laboratory, Imperial College London, London SW7 2AZ
London, United Kingdom

SOUGATO BOSE

Department of Physics and Astronomy, University College London, Gower Street, WC1E 6BT
London, United Kingdom

ALESSANDRO FERRARO

Centre for Theoretical Atomic, Molecular, and Optical Physics, School of Mathematics and Physics, Queen's University Belfast, BT7 1NN
Belfast, United Kingdom

MAURO PATERNOSTRO

Centre for Theoretical Atomic, Molecular, and Optical Physics, School of Mathematics and Physics, Queen's University Belfast, BT7 1NN
Belfast, United Kingdom

We present strategies for the training of a qubit network aimed at the ancilla-assisted synthesis of multi-qubit gates based on a set of restricted resources. By assuming the availability of only time-independent single and two-qubit interactions, we introduce and describe a supervised learning strategy implemented through momentum-stochastic gradient descent with automatic differentiation methods. We demonstrate the effectiveness of the scheme by discussing the implementation of non-trivial three qubit operations, including a Quantum Fourier Transform (QFT) and a half-adder gate.

1. Introduction

Machine learning techniques find fruitful application in several branches of quantum information theory.[1–6] Supervised learning, in particular, provides powerful tools

2

to build algorithms able to pick up patterns from sets of pre-labelled data. In the context of quantum information science, machine learning techniques have been showcased as a flexible tool to solve complex optimisation tasks in different areas.[7–17] In particular, supervised learning techniques were recently demonstrated to solve *gate design* problems.[15, 17]

Here, by *gate design problem* we mean the task of identifying a time-independent Hamiltonian generating a target evolution, under a series of restrictions imposed on the allowed Hamiltonian terms. This problem is well suited for machine learning techniques because, while mathematically well-defined, it presents a complex and strongly context-depend phenomenology. At the same time, the ability to implement non-trivial gates using time-independent dynamics and experimentally-feasible interactions is potentially of great interest for practical implementations of quantum algorithms. It was recently shown in Ref.[17] that, by exploiting supervised learning techniques and ideas, it is possible to flexibly tackle arbitrary gate design problems. A mathematical framework to further improve the efficiency of the numerics by providing improved ansatz for the optimisations can be formalised and used. In this paper, we illustrate the methodology sketched in Ref.[17] and give further examples of time-independent Hamiltonians found via supervised learning optimisation, exploring cases in which ancillary qubits are used to catalyse the evolution, but are traced out at the end of the dynamics in such a way that the effective evolution over the system qubits is the desired one.

The remainder of this paper is organized as follows. In Sec. 2 we introduce the context of the problem that we address, stating the mathematical conditions that qualify a successful gate design (or synthesis). Sec. 3 is dedicated to the illustration of the supervised learning approach that we use to tackle our problem. In Sec. 4 we discuss in details the stochastic momentum gradient descent approach that we choose to ensure the fulfilment of the mathematical conditions for gate design. Such methods are used in Sec. 5 to address quantitatively the training of a qubit network for the synthesis of three non-trivial three-qubit gates, including a half-adder and a quantum Fourier transform. We demonstrate that extremely high-quality gates can be synthesised using only evolutions coupling at most two qubits per time. Moreover, such restricted-resource designs are relatively stable against variations in the controlled parameters of the synthesis, which makes our approach robust, and potentially very interesting for implementation in a number of context, from superconducting quantum systems to trapped ions. Finally, in Sec. 6 we draw our conclusions.

2. Optimisation of qubit networks

Consider a set of qubits whose dynamics is described by the time-independent Hamiltonian $\mathcal{H}_{\boldsymbol{\lambda}} = \sum_k \lambda_k A_k$, which is specified by the set of real parameters $\boldsymbol{\lambda} = \{\lambda_1, \lambda_2, \ldots, \lambda_k, \ldots\}$ and has been decomposed over the (Hermitian) operatorial basis $\{A_k\}$. We consider some of the qubits of the set, say the first M ones, as

composing the *system* $\mathcal{S}$ under investigation. The others embody a set $\mathcal{A}$ of *ancillary qubits*, whose collective initial state is set to be $|\phi\rangle_{\mathcal{A}}$. We call $|\psi\rangle_{\mathcal{S}}$ the initial state of the system qubits, which we assume to be uncorrelated to the set of ancillae and will thus evolve as

$$\rho_{\text{out}}^{\mathcal{S}}(t, \boldsymbol{\lambda}) = \text{Tr}_{\mathcal{A}}\left[\exp(it\mathcal{H}_{\boldsymbol{\lambda}})(\psi \otimes \phi)\right], \tag{1}$$

where $\text{Tr}_{\mathcal{A}}$ denotes the partial trace of the ancillary qubits and we have used the shortcut notation

$$\exp(it\mathcal{H}_{\boldsymbol{\lambda}})(\psi \otimes \phi) = e^{it\mathcal{H}_{\boldsymbol{\lambda}}}(|\psi\rangle\langle\psi|_{\mathcal{S}} \otimes |\phi\rangle\langle\phi|_{\mathcal{A}})e^{-it\mathcal{H}_{\boldsymbol{\lambda}}}. \tag{2}$$

Without loss of generality, in the following we can assume units such that $t = 1$, so that the evolution time can be incorporated into a suitable rescaling of the the Hamiltonian parameters $\boldsymbol{\lambda}$. The evolved system state will henceforth be written as $\rho_{\text{out}}^{\mathcal{S}}(t, \boldsymbol{\lambda}) \equiv \rho_{\text{out}}^{\mathcal{S}}(\boldsymbol{\lambda})$. Our goal is to find a set of parameters $\boldsymbol{\lambda}_0$ such that, for a given target evolution $\mathcal{U}_{target}$, we have the successful gate synthesis $\mathcal{U}_{target} = \exp(i\mathcal{H}_{\boldsymbol{\lambda}_0})$. This is equivalent to having

$$\rho_{\text{out}}^{\mathcal{S}}(\boldsymbol{\lambda}_0) = \mathcal{U}_{target}(|\psi\rangle\langle\psi|_{\mathcal{S}})\mathcal{U}_{target}^{\dagger}, \qquad \forall |\psi\rangle_{\mathcal{S}}. \tag{3}$$

A natural way to quantify how much a given $\boldsymbol{\lambda}$ is close to the needed set $\boldsymbol{\lambda}_0$ is to use the state fidelity

$$\mathcal{F}_{\boldsymbol{\lambda}}(\psi) = {}_{\mathcal{S}}\langle\psi|\mathcal{U}_{target}^{\dagger}\rho_{\text{out}}^{\mathcal{S}}(\boldsymbol{\lambda})\mathcal{U}_{target}|\psi\rangle_{\mathcal{S}}. \tag{4}$$

It then follows that

$$\mathcal{F}_{\boldsymbol{\lambda}_0}(\psi) = 1 \quad \forall |\psi\rangle \Leftrightarrow \exp(i\mathcal{H}_{\boldsymbol{\lambda}_0}) = \mathcal{U}_{target}. \tag{5}$$

Restricting ourselves to Hamiltonians having at most pairwise interactions, which are easier to implement in many experimental architectures, we can consider the following decomposition over the Pauli basis

$$\mathcal{H}_{\boldsymbol{\lambda}} = h_0\mathbb{1} + \sum h_i^{\alpha}\sigma_i^{\alpha} + \sum J_{ij}^{\alpha\beta}\sigma_i^{\alpha}\sigma_j^{\beta}. \tag{6}$$

Here, $i, j = 1, ..., N$ are indices for the qubits of the overall set (system and ancillae), $\alpha, \beta \in \{1, 2, 3\}$ identify the element of the Pauli vector $\boldsymbol{\sigma} = (\sigma^1, \sigma^2, \sigma^3)$ being considered (we use the correspondence $\sigma^1 \to \sigma^x, \sigma^2 \to \sigma^y, \sigma^3 \to \sigma^z$), $J_{ij}^{\alpha\beta}$ is a coupling strength, and $\boldsymbol{\lambda} = \{h_0, h_i^{\alpha}, J_{ij}^{\alpha\beta}\}$.

Note that, for a given choice of $\mathcal{U}_{target}$ and $\mathcal{H}_{\boldsymbol{\lambda}}$, it is not known whether a solution to the question posed by Eq. (5) even exists. Indeed, a simple parameter-counting argument shows it does not in general: not all possible evolutions produced using higher-order interactions can be generated when only one- and two-qubit couplings are available. A concrete example of the gate-synthesis problem discussed here is given by the implementation of a Toffoli gate[19] over a three-qubit network with a restricted set of interactions. Yet, Ref.[15] has shown that the use of a single ancillary qubit catalyses the synthesis of three-qubit gates with the restricted resources mentioned above. Moreover, it was recently shown[17] that the same task can be

4

achieved *without* ancillary qubits and under the stronger restriction of having available only *diagonal* pairwise interactions. Whether a given gate can be realized with a specific time-independent set of operations remains an open question.

3. Supervised learning of the Hamiltonian parameters

A standard approach to the design of Hamiltonians producing a target gate is the direct optimization of the average fidelity $\overline{\mathcal{F}}_{\boldsymbol{\lambda}} \equiv \int \mathcal{F}_\psi \, d\psi$, where the integral is performed over the set of all pure states $|\psi\rangle$. Explicit closed expressions for $\overline{\mathcal{F}}_{\boldsymbol{\lambda}}$ are known,[20–22] so that global optimisation algorithms can be applied directly. However, the resulting complex parameter landscape makes such optimization inefficient. In order to avoid the issues associated with global optimisation techniques, we thus resort to a different approach.

Instead of directly optimising $\overline{\mathcal{F}}_{\boldsymbol{\lambda}}$, we focus on single input states and optimise $\mathcal{F}_{\boldsymbol{\lambda}}(\psi)$ for many different $|\psi\rangle$. Formulated in this manner, the problem can be phrased in terms of supervised machine learning tasks, akin to the training of a neural network model. Indeed, qubit networks can be trained to implement target evolutions just as neural network are trained to implement desired functional relationships.

One of the most used techniques to train neural network models is the so-called Stochastic Gradient Descent (SGD),[23, 24] a workhorse of many state-of-the-art machine learning algorithms. Given a parametrised functional relation $f_{\boldsymbol{w}}(\boldsymbol{x})$, SGD looks for sets of parameters $\boldsymbol{w}_0$ such that $f_{\boldsymbol{w}_0}(\boldsymbol{x})$ is minimised for all $\boldsymbol{x}$. This is done via the following iterative procedure: 1) choose a random input $\boldsymbol{x}_0$, 2) perform a few gradient descent steps over $\boldsymbol{w}$, and 3) go back to the first step, until a good result is obtained. In this way some of the difficulties associated with local optimisation are avoided, because for every input $\boldsymbol{x}$ one has a different parameter landscape $\boldsymbol{w} \mapsto f_{\boldsymbol{w}}(\boldsymbol{x})$ over which the gradient descent is performed. This new parameter landscape does not in general have the same local minima as the previous ones, while on the other hand the global minimum is bound to be a minimum for all $\boldsymbol{x}$. In the qubit network scenario that we are tackling, this translates into the fidelity being unitary for all $|\psi\rangle_{\mathcal{S}}$ only for the sets $\boldsymbol{\lambda}$ such that $\exp(i\mathcal{H}_{\boldsymbol{\lambda}}) = \mathcal{U}_{target}$.

Following the large body of research that the area of machine learning has seen in recent years,[25, 26] a number of well-honed tools to efficiently tackle these kinds of tasks has been developed and made freely available. In particular, software frameworks such as Theano,[27] TensorFlow[28] and PyTorch[29] allow to efficiently train arbitrarily built models with state-of-the-art SGD algorithms. One particular aspect of these frameworks which allows to further speed-up the training procedure is that they allow to perform *automatic differentiation*[23, 30, 31] of arbitrary computations. Automatic differentiation is a technique to efficiently compute, numerically, the gradient of an arbitrary multivalued function $f(\boldsymbol{x})$, without using numerical approximation methods nor requiring the analytical expression of the gradient. Such technique is especially valuable to speed-up SGD optimisations, as it significantly decreases the cost of computing $\boldsymbol{\nabla}_{\boldsymbol{w}} f$.

4. Implementation details

We implement supervised learning of arbitrary qubit networks using Theano,[27] a Python library for machine learning of widespread use. More specifically, we use Theano to perform momentum-SGD[24] over the network parameters $\boldsymbol{\lambda}$, exploiting the automatic differentiation capabilities provided by Theano to speed-up the optimisation. The implemented protocol can be summarized in the following steps

(1) Choose an initial set of parameters $\boldsymbol{\lambda}$.
(2) Generate a random set of input states $|\psi_k\rangle$, $k = 1, ..., N_b$, with N_b the size of the mini-batches chosen beforehand.
(3) For each k, compute $\boldsymbol{\nabla_\lambda}\mathcal{F}_{\boldsymbol{\lambda}}(\psi_k)$ using the automatic differentiation capabilities offered by Theano.
(4) Change the coupling strengths $\boldsymbol{\lambda}$ according to the chosen updating rule. A standard choice, adopted in this work, is momentum-SGD, which is characterised by the following updating rule:

$$
\begin{aligned}
\boldsymbol{v} &\to \gamma\boldsymbol{v} + \eta\boldsymbol{\nabla_\lambda}\mathcal{F}_{\boldsymbol{\lambda}}(\psi_k), \\
\boldsymbol{\lambda} &\to \boldsymbol{\lambda} + \boldsymbol{v},
\end{aligned}
\tag{7}
$$

where the *learning rate* η and the *momentum term* γ are *hyperparameters* that define the optimisation protocol. The value of the learning rate is also chosen to be decreasing with the iteration number.
(5) Return to point (2), until a satisfactory value of the fidelity is obtained.

For the cost function $\mathcal{F}_{\boldsymbol{\lambda}}$ we use the expression given in Eq. (11), with a general Hamiltonian model with at most pairwise interactions, as given in Eq. (6). This means that, for example, to train the four-qubit networks implementing Half-adder and TofFredkin gates, we start with a general Hamiltonian like in Eq. (6) with 9 parameters h_i^α for the single-qubit fields, plus 3×9 parameters $J_{ij}^{\alpha\beta}$ for the two-qubit interactions, amounting to a total of 36 parameters to be trained. The h_0 parameter can be left out of the training as it only amounts to an unobservable global phase. A similar calculation tells us that for the $3 + 5$ network used to train the QFT gate, a total amount of 276 parameters are trained.

All the codes used to generate the results reported in this paper are freely available from Ref.[32]

5. Results

We trained qubit networks of various sizes to generate the Quantum Fourier Transform (QFT) over three qubits, a special doubly-controlled gate that shares characteristics with both Toffoli and Fredkin gate, to which we will refer to as *TofFredkin* in the following, and the half-adder gate.[33]

6

The three-qubit QFT transformation can be written as

$$\text{QFT} \, |a\rangle_1 \, |b\rangle_2 \, |c\rangle_3 = \frac{1}{\sqrt{8}} \left(|0\rangle + e^{i\pi c} \, |1\rangle \right)_1 \left(|0\rangle + e^{i\pi(b+\frac{c}{2})} \, |1\rangle \right)_2 \left(|0\rangle + e^{i\pi(a+\frac{b}{2}+\frac{c}{4})} \, |1\rangle \right)_3 \tag{8}$$

with $|k\rangle_j$ the states of the system qubit $j = 1, 2, 3$ ($k = a, b, c = 0, 1$).

We define the TofFredkin gate as the three-qubit transformation

$$\mathcal{U}_{\text{TF}} = |0\rangle \langle 0|_1 \otimes \text{CNOT}_{23} + |1\rangle \langle 1|_1 \otimes \text{SWAP}_{23}, \tag{9}$$

which thus performs a CNOT (SWAP) gate on qubits 2 and 3, akin to a quantum Toffoli (Fredkin) gate, when the state of the control qubit 1 is $|0\rangle_1$ ($|1\rangle_1$).

Finally, the half-adder acts on a register of three qubits as

$$\mathcal{U}_{HA} = \text{CNOT}_{12}\text{CCNOT}_{123}, \tag{10}$$

where CCNOT_{123} stands for the quantum Toffoli gate. When applied to a logical state $|a\rangle_1 \, |b\rangle_2 \, |c\rangle_3$, the half adder returns the state $|a\rangle_1 \, |a \oplus b\rangle_2 \, |c \oplus \text{carry}\rangle_3$ with carry=0,1 the carry over of $a \oplus b$.

To assess the quality of the obtained gates we use the averaged gate fidelity $\overline{\mathcal{F}}(\mathcal{E}, \mathcal{U})$, defined as

$$\overline{\mathcal{F}}(\mathcal{E}, \mathcal{U}) = \int d\psi \, \langle \psi | \mathcal{U}^\dagger \mathcal{E}(|\psi\rangle\langle\psi|) \mathcal{U} | \psi \rangle \,. \tag{11}$$

This quantity is used to characterize how close the action of a map $\mathcal{E}$ is to the action of a unitary $\mathcal{U}$. The map is in our case the operation corresponding to the action of a unitary in the enlarged system+ancillae space, followed by tracing the ancillary qubits. In our case, if $\tilde{\mathcal{U}}$ is the unitary acting on the full system+ancillae space, obtained from the learning procedure, then the map is defined as $\mathcal{E}(\rho) = \text{Tr}_A[\tilde{\mathcal{U}}(\rho \otimes |0\rangle\langle0|)\tilde{\mathcal{U}}^\dagger]$, where $|0\rangle\langle0|_A$ is the initial state of the ancillary system. Equation (11) can be explicitly computed as

$$\overline{\mathcal{F}}(\mathcal{E}, \mathcal{U}) = \frac{1}{d+1} \left[1 + \frac{1}{d} \sum_{ij} \langle i | \mathcal{U}^\dagger \mathcal{E}(|i\rangle\langle j|) \mathcal{U} | j \rangle \right]. \tag{12}$$

In what follows, we use of Eq. (12) to assess the quality of the reported results.

Figure 1 shows the sets of Hamiltonian parameters that were obtained through the training procedure, for each of these operations. The TofFredkin gate was found with unit fidelity, up to numerical precision, using a single ancillary qubit. In Fig. 2 is shown how the fidelity varies with the Hamiltonian parameters. Fig. 3 reports the final matrix over the full four-qubit network.

In the case of the QFT, we performed a series of training procedures with different sizes for the ancillary system, from the case of no ancilla to a maximum of five ancillary qubits being used. We observed the best results when using 0, 3, and 5 ancillae, the average fidelities being 0.987, 0.991, and 0.988 respectively. It is important to note that due to the heuristic nature of the optimization method we employed, better results could be possible by using different initial conditions

or optimization parameters. In Figs. 1 and 4 we show the final parameters, and variation of final fidelities against the parameters, for the case of 5 ancillary qubits.[34]

Finally, the half-adder gate was realized with average fidelity $\overline{\mathcal{F}} \simeq 0.999997$ with a single ancilla. Again, in Fig. 5 is reported the behaviour of $\mathcal{F}_{\boldsymbol{\lambda}}$ upon variations of $\boldsymbol{\lambda}$, and in Fig. 6 we show the final unitary gate implemented over the full four-qubit network.

Figures 2, 4 and 5 show the relative stability of the gates with respect to changes of the Hamiltonian parameters. In particular, for TofFredkin and QFT, the fidelity remains above 95% upon a 25% variation of the evolution time. The half-adder appears to be less stable, but this is only consequence of the larger values of its interactions, as shown in Fig. 1(c). Indeed, the hardness of tuning the Hamiltonian parameters with sufficient precision will vary strongly between different gates, as well as between different implementation of the same gate, and between different parameters in any given implementation.

These results provide further evidence in support of the power and flexibility of the supervised learning approach presented in,[15,17] which clearly applies to the cases where ancillary degrees of freedom are exploited during the evolution.

6. Conclusions

We have reported on strategies for supervised learning-assisted synthesis of multi-qubit quantum gates. The general scheme of our gate design is based on the use of a limited set of resources (such as two-qubit gates) that, in general, would not be sufficient for arbitrary gate design. However, the catalyst effect brought about by the use of machine learning – specifically, stochastic momentum gradient descent – is sufficient to overcome the limitation of our resources and deliver high-quality complex quantum gates. We have demonstrated the performance of our scheme by addressing the significant cases of a half-adder and a QFT circuit. Moreover, we have shown the flexibility of our approach by addressing a novel type of gate that puts together the paradigmatic quantum Toffoli and Fredkin gates.

We believe that the proposed methodology will be able to significantly help in the task of designing complex evolutions suitable for the simulation of the dynamics of networks of information carriers, lowering the complexity associated with gate decomposition, and simply adopting well-known techniques of classical machine learning.
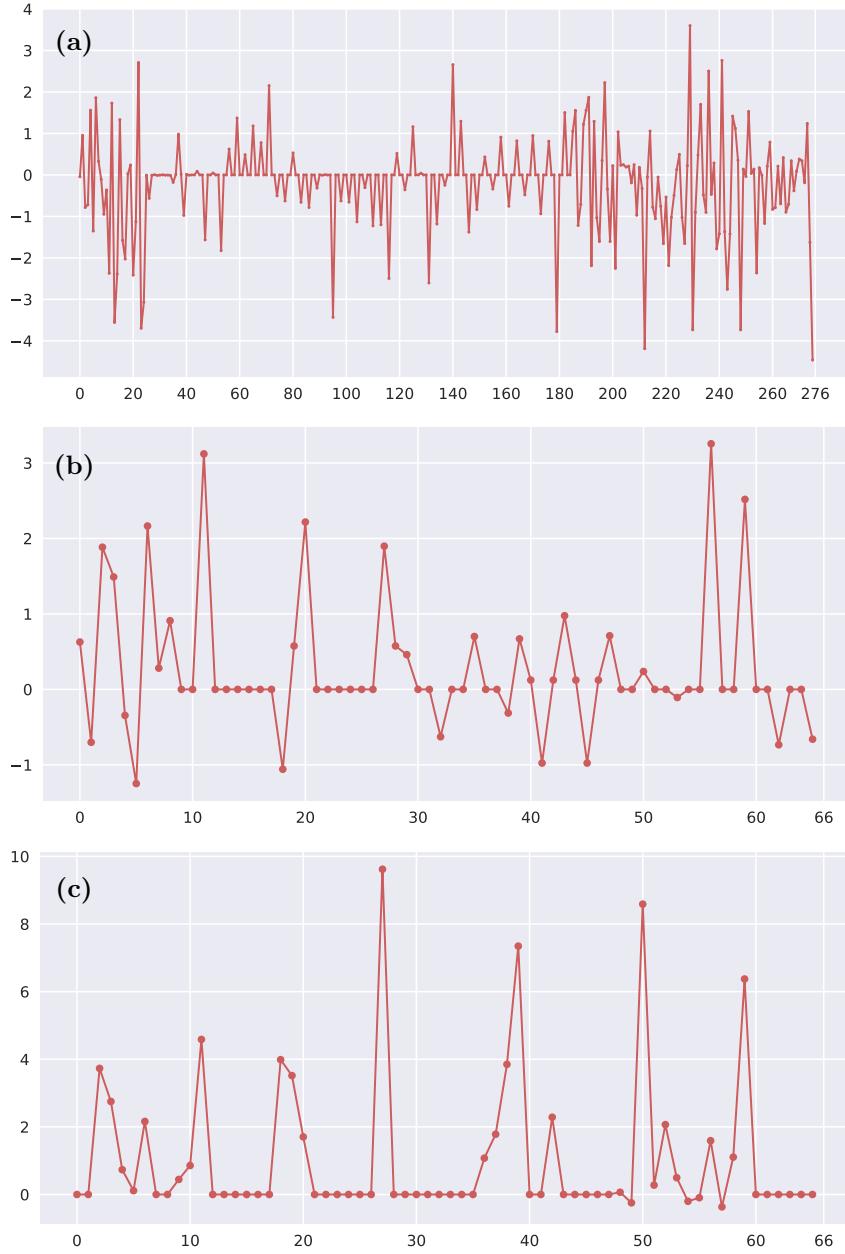
7. Acknowledgments

8

Fig. 1. Trained parameters for the three-qubit QFT, TofFredkin and half-adder networks. We plot the values taken by the elements of the full set of parameters entering Eq. (6) and ordered as $\boldsymbol{\lambda} = \{h_0, h_1^1, \ldots, h_N^3, J_{11}^{11}, \ldots, J_{NN}^{33}\}$ with N the number of elements of the qubit network.
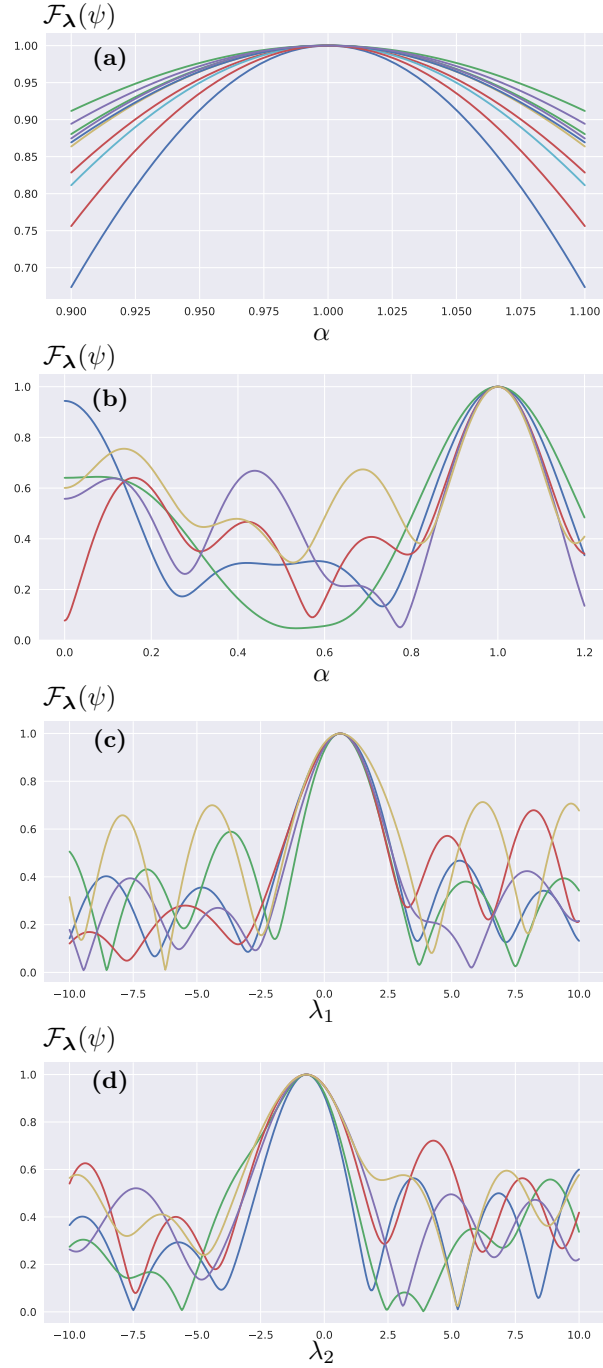
Fig. 2. Fidelity $\mathcal{F}_{\boldsymbol{\lambda}}(\psi)$ vs variations of $\boldsymbol{\lambda}$, for different test states, for the **TofFredkin** gate. The five test states ψ are sampled randomly. **(a)** Global relative variations of $\boldsymbol{\lambda}$, that is, plotting the fidelity against $\alpha\boldsymbol{\lambda}$ for $0.9 \leq \alpha \leq 1.1$. Note that this is equivalent to studying how the fidelity changes with respect to uncertainties in the evolution time, that is, how much does $\exp(i\mathcal{H}t')$ differ from $\exp(i\mathcal{H}t)$. **(b)** Same as *(a)* but with $0 \leq \alpha \leq 1.2$. **(c)** Plot of $\mathcal{F}_{\boldsymbol{\lambda}}(\psi)$ against *absolute* variations of a single element of $\boldsymbol{\lambda}$, in this case the first one, i.e. we take $\lambda_1 \in [-10, 10]$. **(d)** Like *(c)* but for λ_2.
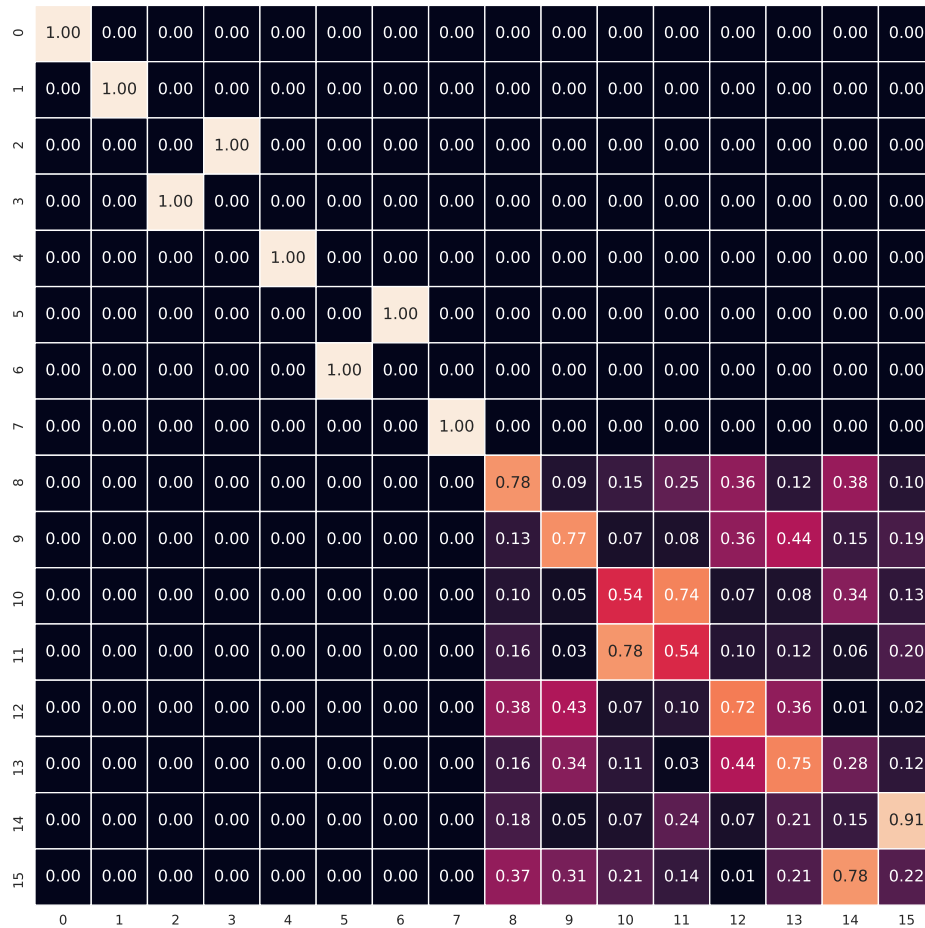
10

Fig. 3. Final TofFredkin gate found from the supervised learning training. In this representation, the ancillary qubit is the first one, so that the top-left 8×8 submatrix describes the evolution of states when the ancillary qubit starts as $|0\rangle$. Notably, it is clear from the matrix that the gate acts diagonally on the ancillary qubits, which therefore effectively acts as a control qubit. When this control ancillary qubit is $|0\rangle$, the other three qubits evolve according to the TofFredkin gate.
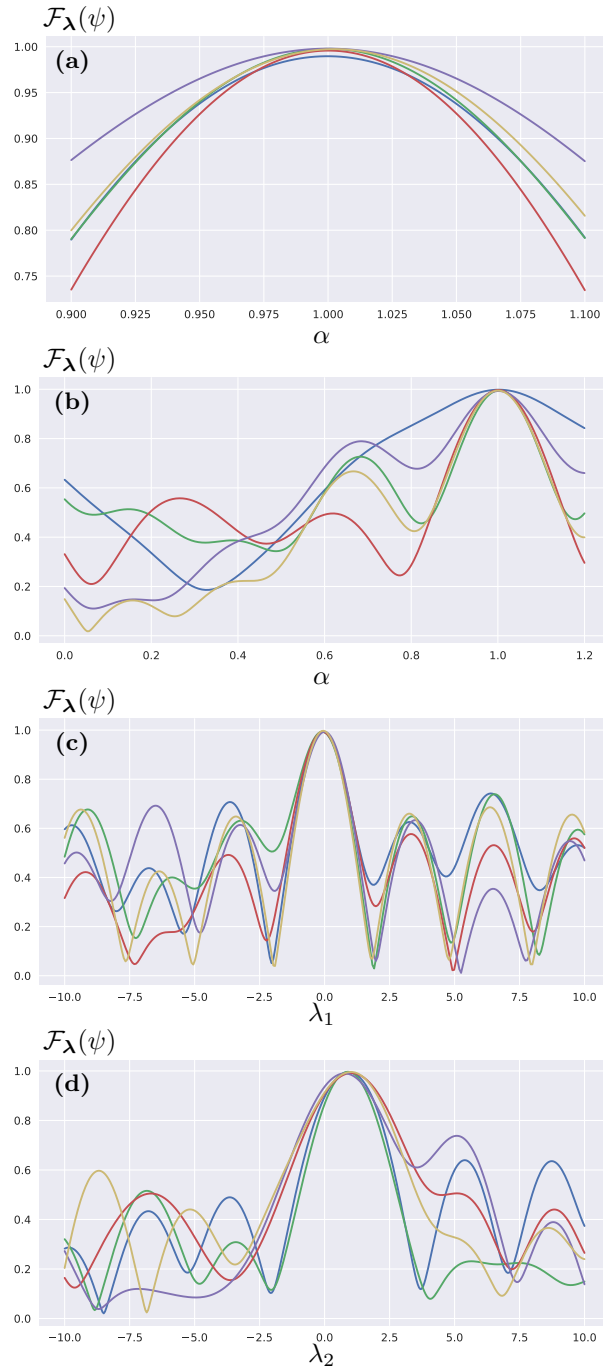
Fig. 4. Fidelity $\mathcal{F}_{\boldsymbol{\lambda}}(\psi)$ vs variations of $\boldsymbol{\lambda}$, for different test states, for the **three-qubit QFT** gate. The five test states ψ are sampled randomly. **(a)** Global relative variations of $\boldsymbol{\lambda}$, that is, plotting the fidelity against $\alpha\boldsymbol{\lambda}$ for $0.9 \leq \alpha \leq 1.1$. Note that this is equivalent to studying how the fidelity changes with respect to uncertainties in the evolution time, that is, how much does $\exp(i\mathcal{H}t')$ differ from $\exp(i\mathcal{H}t)$. **(b)** Same as *(a)* but with $0 \leq \alpha \leq 1.2$. **(c)** Plot of $\mathcal{F}_{\boldsymbol{\lambda}}(\psi)$ against *absolute* variations of a single element of $\boldsymbol{\lambda}$, in this case the first one, i.e. we take $\lambda_1 \in [-10, 10]$. **(d)** Like *(c)* but for λ_2.
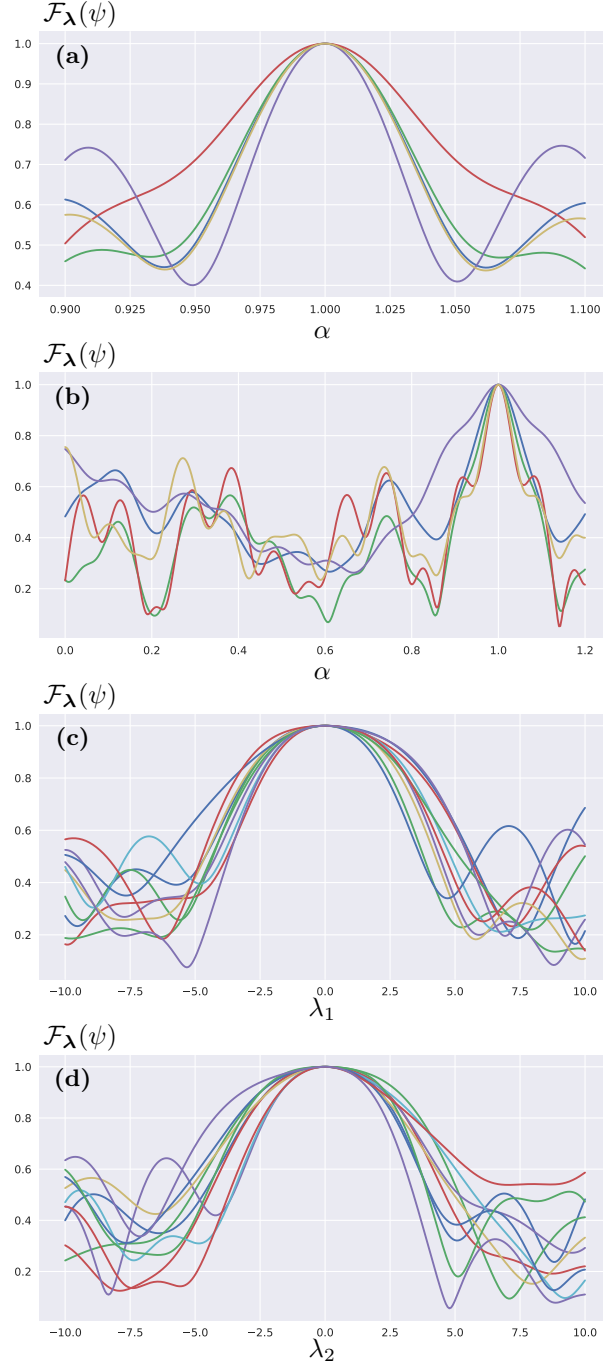
12

Fig. 5. Fidelity $\mathcal{F}_{\boldsymbol{\lambda}}(\psi)$ vs variations of $\boldsymbol{\lambda}$, for different test states, for the **half-adder** gate. The five test states ψ are sampled randomly. **(a)** Global relative variations of $\boldsymbol{\lambda}$, that is, plotting the fidelity against $\alpha\boldsymbol{\lambda}$ for $0.9 \leq \alpha \leq 1.1$. Note that this is equivalent to studying how the fidelity changes with respect to uncertainties in the evolution time, that is, how much does $\exp(i\mathcal{H}t')$ differ from $\exp(i\mathcal{H}t)$. **(b)** Same as *(a)* but with $0 \leq \alpha \leq 1.2$. **(c)** Plot of $\mathcal{F}_{\boldsymbol{\lambda}}(\psi)$ against *absolute* variations of a single element of $\boldsymbol{\lambda}$, in this case the first one, i.e. we take $\lambda_1 \in [-10, 10]$. **(d)** Like *(c)* but for λ_2.

Fig. 6. Final half-adder gate found from the supervised learning training. In this representation, the ancillary qubit is the first one, so that the top-left 8×8 submatrix describes the evolution of states when the ancillary qubit starts as $|0\rangle$. Notably, it is clear from the matrix that the gate acts diagonally on the ancillary qubits, which therefore effectively acts as a control qubit, which when $|0\rangle$ induces the other three qubits to evolve according to the half-adder gate.

14

References

1. Mehta, P., Bukov, M., Wang, C. H., Day, A. G., Richardson, C., Fisher, C. K., and Schwab, D. J. *arXiv:1803.08823* (2018).
2. Spears, B. K. *arXiv:1712.08523* (2017).
3. J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, *Nature* **549** (2017) 195.
4. M. Schuld, I. Sinayskiy, and F. Petruccione, *Contemporary Physics* **56** (2015) 172.
5. Ciliberto, C., Herbster, M., Ialongo, A. D., Pontil, M., Rocchetto, A., Severini, S., and Wossnig, L. *Proc. R. Soc. A* **474(2209)** (2018) 20170551.
6. Perdomo-Ortiz, A., Benedetti, M., Realpe-Gómez, J., and Biswas, R. *arXiv:1708.09757* (2017).
7. L. Zdeborová, *Nature Physics* **13** (2017) 420.
8. J. Carrasquilla and R. G. Melko, *Nature Physics* **13** (2017) 431.
9. G. Carleo and M. Troyer, *Science* **355** (2017) 602.
10. E. P. van Nieuwenburg, Y.-H. Liu and S. D. Huber, *Nature Physics* **13** (2017) 435.
11. S. S. Schoenholz, E. D. Cubuk, D. M. Sussman, E. Kaxiras and A. J. Liu, *Nature Physics* **12** (2016) 469.
12. G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko and G. Carleo, *Nature Physics* **14** (2018) 447.
13. A. Rocchetto, S. Aaronson, S. Severini, G. Carvacho, D. Poderini, I. Agresti, M. Bentivegna and F. Sciarrino, *arXiv:1712.00127* (2017)
14. A. A. Melnikov, H. P. Nautrup, M. Krenn, V. Dunjko, M. Tiersch, A. Zeilinger and H. J. Briegel, *Proceedings of the National Academy of Sciences* (2018) p. 201714936.
15. L. Banchi, N. Pancotti and S. Bose, *NPJ Quantum Information* **2** (2016) 16019.
16. H. Fujita, Y. O. Nakagawa, S. Sugiura and M. Oshikawa, *Physical Review B* **97** (2018) 075114.
17. L. Innocenti, L. Banchi, A. Ferraro, S. Bose and M. Paternostro, *arXiv:1803.07119* (2018).
18. E. Zahedinejad, J. Ghosh and B. C. Sanders, *Physical Review Letters* **114** (2015) 200502.
19. Y. Shi, *arXiv:0205115* (2002).
20. L. Banchi, A. Bayat, P. Verrucchi and S. Bose, *Physical Review Letters* **106** (2011) 140501.
21. E. Magesan, R. Blume-Kohout and J. Emerson, *Physical Review A* **84** (2011) 012309.
22. L. H. Pedersen, N. M. Møller and K. Mølmer, *Physics Letters A* **367** (2007) 47.
23. R. E. Wengert, *Communications of the ACM* **7** (1964) 463.
24. S. Ruder, *arXiv:1609.04747* (2016).
25. Y. LeCun, Y. Bengio, and G. Hinton, *Nature* **521** (2015) 436.
26. D. Silver, et al., *Nature* **529** (2016) 7587.
27. T. T. D. Team, *arXiv:1605.02688* (2016).
28. M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu and X. Zheng, TensorFlow: Large-scale machine learning on heterogeneous systems (2015). Software available from tensorflow.org.
29. A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga and A. Lerer, Automatic differentiation in pytorch, in *NIPS-W* (20170; available from https://openreview.net/forum?id=BJJsrmfCZ.

15

30. M. Bartholomew-Biggs, S. Brown, B. Christianson and L. Dixon, *Journal of Computational and Applied Mathematics* **124** (2000) 171.
31. A. G. Baydin, B. A. Pearlmutter, A. A. Radul and J. M. Siskind, *arXiv:1502.05767* (2015).
32. github.com/lucainnocenti/quantum-gate-learning-1803.07119.
33. G. A. Barbosa, *Physical Review A* **73** (2006) 052321.
34. More   details,   as well as all of the data , are available at github.com/lucainnocenti/quantum-gate-learning-1803.07119/notebooks/QFT/qft_over_three_qubits.ipynb.
35. G.   Goh,   "*Why   Momentum   Really   Works*",   *Distill*,   2017. http://doi.org/10.23915/distill.00006.