Direct implementation of a perceptron in superconducting circuit quantum hardware

Marek Pechal,[1,2] Federico Roy [1,3,4] Samuel A. Wilkinson [5] Gian Salis [1] Max Werninghaus,[1,4,6]
Michael J. Hartmann [5,7] and Stefan Filipp [1,4,6,8]

[1]*IBM Quantum, IBM Research Europe-Zurich, Säumerstrasse 4, 8803 Rüschlikon, Switzerland*
[2]*ETH Zurich, Department of Physics, 8093 Zürich, Switzerland*
[3]*Theoretical Physics, Saarland University, 66123 Saarbrücken, Germany*
[4]*Walther-Meißner-Institut, Bayerische Akademie der Wissenschaften, 85748 Garching, Germany*
[5]*Department of Physics, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Staudtstrasse 7, 91058 Erlangen, Germany*
[6]*Physik-Department, Technische Universität München, 85748 Garching, Germany*
[7]*Max Planck Institute for the Science of Light, 91058 Erlangen, Germany*
[8]*Munich Center for Quantum Science and Technology (MCQST), Schellingstraße 4, 80799 München, Germany*

The utility of classical neural networks as universal approximators suggests that their quantum analogues could play an important role in quantum generalizations of machine-learning methods. Inspired by the proposal in Torrontegui and García-Ripoll [Europhys. Lett. **125**, 30004 (2019)], we demonstrate a superconducting qubit implementation of a controlled gate, which generalizes the action of a classical perceptron as the basic building block of a quantum neural network. In a two-qubit setup we show full control over the steepness of the perceptron activation function, the input weight and the bias by tuning the gate length, the coupling between the qubits, and the frequency of the applied drive, respectively. In its general form, the gate realizes a multiqubit entangling operation in a single step, whose decomposition into single- and two-qubit gates would require a number of gates that is exponential in the number of qubits. Its demonstrated direct implementation as perceptron in quantum hardware may therefore lead to more powerful quantum neural networks when combined with suitable additional standard gates.

I. INTRODUCTION

Artificial neural networks and engineered quantum systems are both quickly developing technologies with far-reaching potential applications. The promise that quantum computing can solve certain problems exponentially faster than classical computing technology and the ever growing thirst for computational power of machine learning applications has triggered substantial interest in the development of quantum machine learning methodology [1–8]. Whereas some work explored the acceleration of specific computational tasks in machine learning with quantum encodings [9], a major part of the research considers quantum neural networks as counterparts to artificial neural networks in classical software. Quantum neural networks are largely implemented as variational quantum circuits [10–13], that are composed of parametrized gates and where finding optimal parameters corresponds to the training of the network. Moreover, it has been shown that quantum speedup is possible in supervised machine learning [14], providing a perspective for hardware-efficient machine learning realizations.

An important question for quantum neural networks is their expressivity, i.e., which mappings between input and output states can be realized by the network. In the classical domain, expressivity is guaranteed by the universal approximation theorem, which requires a nonlinear activation function for the perceptrons in the network. In the quantum setting, there is currently no unique accepted way to extend the concept of classical neural networks to quantum systems [15,16]. In one approach, the inputs are encoded as classical parameters of applied gates [17], and the network approximates functions mapping these parameters to quantum states. Another option is to encode the neuron inputs as quantum states and to design a quantum perceptron as a unitary whose action on a set of basis states matches a classical perceptron [18]. Such a building block can serve as a tool to enable experimental studies and development of quantum neural networks. Since such an approach includes the functionality of classical neural networks in limiting cases, it has the expressive power guaranteed by the universal approximation theorem in those cases and thus forms a natural starting point for exploring quantum generalizations. Another advantage of this approach is that the perceptron may be directly realized at the hardware level, rather than encoded in software: this neuromorphic approach is more hardware efficient and may thus offer significant advantages
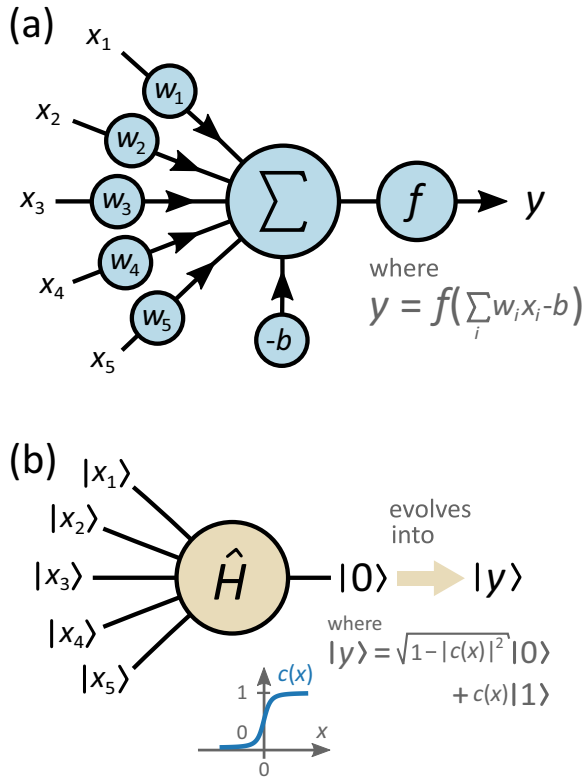
FIG. 1. (a) Schematic representation of a classical perceptron with $n = 5$ inputs. The perceptron output y is given by a nonlinear activation function $f(\cdot)$ applied to the sum of input values $x_1, ..., x_n$ individually weighted by $w_1, ..., w_n$ and biased by a term b. A quantum analog (b) can be constructed, where the inputs are encoded in quantum states $|x_1\rangle, ..., |x_n\rangle$ of n qubits and an output qubit undergoes a rotation by an angle of $\theta = 2\arcsin[c(x)]$ determined by the inputs values, leading to an excited population of $|c(x)|^2$ when starting the output qubit in the ground state.

for scaling the concept to larger and more data intensive applications.

The construction of the quantum counterpart to a perceptron is easily illustrated with the example of a binary classifier perceptron. This simple classical perceptron takes a number of binary inputs $x_i \in \{0, 1\}$, each with an associated weight w_i, and outputs another binary value y depending on whether the linear combination $x_{\text{in}} = w_1 x_1 + \ldots$ exceeds a given threshold or not:

$$y = \begin{cases} 0 \text{ if } x_{\text{in}} + b < 0, \\ 1 \text{ if } x_{\text{in}} + b \geqslant 0, \end{cases}$$

or, written compactly in terms of the Heaviside step function, $y = \Theta(x_{\text{in}} + b)$ with an additional bias b. More generally, as illustrated in Fig. 1(a), the step function may be replaced by a continuous activation function f and the inputs and outputs promoted from binary values to continuous ones. The key advantage of such continuous activation functions is that they have finite gradients which can be used for training the network in gradient descent approaches.

In a quantum generalization of the classifier perceptron, the binary variables can be naturally represented by qubit states, which we will label as $|0\rangle$ and $|1\rangle$. The action of the perceptron

can then be represented by a gate U whose effect on the output qubit depends on the states of the input qubits:

$$|x_1 x_2 \ldots\rangle_{\text{in}} \otimes |0\rangle_{\text{out}} \xrightarrow{U} \begin{cases} |x_1 x_2 \ldots\rangle_{\text{in}} \otimes |0\rangle_{\text{out}} \\ \quad \text{if } x_{\text{in}} + b \equiv \sum_j w_j x_j + b < 0, \\ \\ |x_1 x_2 \ldots\rangle_{\text{in}} \otimes |1\rangle_{\text{out}} \\ \quad \text{if } x_{\text{in}} + b \equiv \sum_j w_j x_j + b \geqslant 0. \end{cases}$$

When the input is a product state $|x_1 x_2 \ldots\rangle_{\text{in}}$, then the unitary is simply a rotation of the output qubit by 0 or π depending on the sign of $x_{\text{in}} + b$ [19]. This concept of a quantum perceptron can be further generalized, similarly to the classical case, by letting the rotation angle have a different dependence on x than a simple Heaviside function. The action of the perceptron, shown schematically in Fig. 1(b), is then described as

$$|x_1 x_2 \ldots\rangle_{\text{in}} \otimes |0\rangle_{\text{out}} \xrightarrow{U} |x_1 x_2 \ldots\rangle_{\text{in}} \otimes |y\rangle_{\text{out}},$$

$$\text{where } |y\rangle_{\text{out}} = \sqrt{1 - |c(x)|^2}|0\rangle + c(x)|1\rangle. \quad (1)$$

Here the excitation amplitude $c(x)$ is a continuous, steplike function of $x \equiv x_{\text{in}} + b$ such as depicted in Fig. 1. The previously described binary perceptron is the special case where $c(x)$ is the Heaviside step function $\Theta(x)$.

Since the perceptron gate is a unitary operation, it can in principle be implemented via any universal set of quantum gates, such as single-qubit rotations and controlled-NOT (CNOT) gates. However, the depth of such a decomposition in general grows exponentially in the number of input qubits. Therefore, we instead realize the perceptron gate directly, making use of an adiabatic protocol. This approach allows us to implement the gate with a single pulse whose duration does not scale with the number of input qubits: From Eq. (1) it can be seen that the perceptron acts like a multiqubit controlled gate. However, instead of triggering an operation on the target only when the control register is in the $|1 \ldots 1\rangle$ state (like a Toffoli gate [20], which applies an X gate to the target only when the control qubits are in the state $|11\rangle$), the perceptron gate applies a different operation to the output qubit for each possible input basis state $|x_1 x_2 \ldots\rangle_{\text{in}}$. As the number of input basis states grows exponentially in the number of qubits, a standard decomposition into elementary gates leads to an expontial growth of the circuit depth (see Sec. III), which is in contrast to the implementation discussed in the following.

II. THE PERCEPTRON GATE

A. Concept

We construct the perceptron unitary with an approach that is a slight modification of a theoretical proposal by Torrontegui and García-Ripoll [18], as described below. An implementation of a similar operation was also proposed [19] based on repeated measurements and feedback, while a gate-based approach to construction of activation functions was demonstrated in Ref. [21].

Our protocol, which we implement in a device with two fixed-frequency superconducting transmon qubits [22] interacting via a tunable coupler [23] [schematically depicted in Fig. 2(a)] is based on a chirped pulse applied to the output qubit. The resulting unitary U_{ad} is designed in such a way that
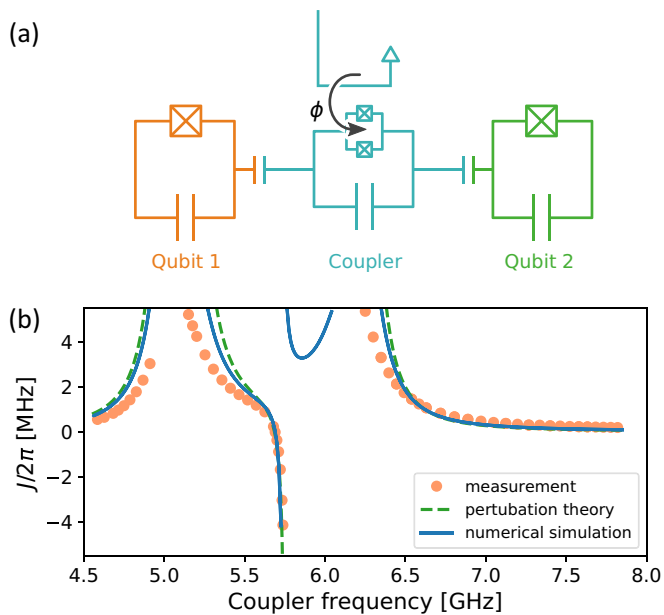
(a)

(b)

FIG. 2. (a) Diagram of the two-qubit experimental setup: two fixed frequency transmons are capacitively coupled via a flux-tunable coupler whose frequency determines the ZZ coupling between the qubits. (b) Measured value of the ZZ coupling as a function of the tunable coupler's frequency (orange dots). Here we plot the coupling strength J conventionally defined via $H_{\mathrm{int}} = J|11\rangle\langle 11|$. In the perceptron context, the weight w from Eq. (3) is given by $w = -J$. The solid line is the result of numerically diagonalizing the Hamiltonian of the system and the dashed line is the result of the lowest-order perturbative calculation (Appendix B).

the final excited state amplitude depends on the detuning Δ of a microwave drive frequency from the output qubit frequency via the steplike response function $c(\Delta)$:

$$|0\rangle_{\mathrm{out}} \xrightarrow{U_{\mathrm{ad}}} \sqrt{1 - |c(\Delta)|^2}|0\rangle_{\mathrm{out}} + c(\Delta)|1\rangle_{\mathrm{out}}. \quad (2)$$

Note that this becomes exactly the mapping in Eq. (1) if the detuning Δ is equal to the linear combination of inputs $\sum_j w_j x_j + b$. We achieve this linear dependence of Δ on the input states by subjecting the system to a ZZ interaction between each of the input qubits and the perceptron qubit:

$$H_{\mathrm{int}} = -\sum_j w_j |1\rangle\langle 1|_j \otimes |1\rangle\langle 1|_{\mathrm{out}}. \quad (3)$$

Here j enumerates the input qubits and "out" labels the output qubit. If the input qubits are in a product state $|x_1 \ldots\rangle$, this is equivalent to shifting the frequency of the output by $-\sum_j w_j x_j$, or $-w_1 x_1$ in the case of a single input.

B. Implementation

The ZZ interactions need to be configurable in-situ to allow training of the network. This can be achieved by using tunable couplers [23] mediating interactions between the output qubit and each input qubit. When the couplers are sufficiently far detuned from the qubits, the dispersive approximation is valid and the interactions effectively introduce ZZ coupling terms, giving rise to the Hamiltonian in Eq. (3). The individual interaction strengths representing the network

weights can be tuned by changing the frequencies of the couplers [24]. Using this scheme we can control the weight of each input by tuning of the respective coupler frequency and the bias point by changing the frequency of the microwave drive.

In the two-qubit device used for our experimental demonstration, qubit 1 at a frequency $\omega_1/2\pi = 6.189\,\mathrm{GHz}$ serves as the output while qubit 2 at a frequency $\omega_2/2\pi = 5.089\,\mathrm{GHz}$ is the single-valued input register. The qubits have anharmonicities $\alpha_1/2\pi = -286\,\mathrm{MHz}$ and $\alpha_2/2\pi = -310\,\mathrm{MHz}$. The coupler has a frequency tunable from its maximum of approximately $\omega_{\mathrm{c,max}}/2\pi = 7.8\,\mathrm{GHz}$ to well below the frequencies of both qubits. The qubit-coupler interaction strengths are approximately $g_1/2\pi = 142\,\mathrm{MHz}$ and $g_2/2\pi = 116\,\mathrm{MHz}$.

Using Ramsey-type measurements with the input qubit either in $|0\rangle$ or in $|1\rangle$, we characterize the dependence of the ZZ coupling on the frequency of the coupler and observe that it can be tuned over a range of a few MHz (see Fig. 2), covering both positive and negative values. The main features of the measured dependence are well reproduced by an analytical expression obtained from fourth-order perturbation theory [see Eq. (B2)], and similarly by the numerical diagonalization of the Hamiltonian [see Eq. (B1)].

The single-qubit operation described by Eq. (2), which underlies the quantum perceptron protocol is realized by applying a chirped pulse to qubit 1 [25,26]. The pulse's initial frequency ω_{i} is far below the output qubit's frequency ω_1 (where ω_1 is defined in the absence of the interaction H_{int}) while its final frequency ω_{f} is detuned from it by $b \equiv \omega_{\mathrm{f}} - \omega_1$. This means the initial detuning of the drive from the qubit is negative while the final detuning is $\Delta = \sum_j w_j x_j + b$. Note that the bias b, being equal to the final chirp detuning from the unshifted qubit frequency, can be tuned arbitrarily by changing ω_{f}.

Under perfect adiabatic conditions, as illustrated in Figs. 3(a) and 3(b), if the initial and the final detuning have the same sign, then the two basis states $|0\rangle$ and $|1\rangle$ remain unchanged by the pulse (up to accumulated phases). If they have opposite signs, then the states are flipped. The resulting dependence of the final excited state population on the qubit frequency is illustrated in Fig. 3(c). Since we choose the initial frequency of the chirped pulse ω_{i} significantly lower than the lowest possible frequency of the output qubit ($\omega_1 - \sum_{j, w_j > 0} w_j$), we neglect the rising edge of the function. The adiabatic operation is then exactly described by Eq. (1) with $|c(\Delta)|^2 = \Theta(\Delta)$ and $\Delta = \sum_j w_j x_j + b$. Smooth response functions $|c(\Delta)|^2$ arise naturally from imperfect adiabaticity of the chirp pulse: With a pulse of finite duration T, the process becomes nonadiabatic when the detuning Δ is on the order of $1/T$ or smaller. This leads to a smoothening of the step in the response function and a finite width of roughly $1/T$. In our experiment, the chirped pulse has a time-dependent frequency $\omega_{\mathrm{p}}(t) = \omega_{\mathrm{i}} + (\omega_{\mathrm{f}} - \omega_{\mathrm{i}})\sin^2(\pi t/2T)$ and amplitude $\Omega(t) = \Omega_0 \sin(\pi t/T)$, (Fig. 3(d)). As a time-transformed version of a hyperbolic secant pulse [27,28], for which analytical solutions of the population transfer exist, it leads to hyperbolic shaped activation functions with few fitting parameters (see Appendix C).
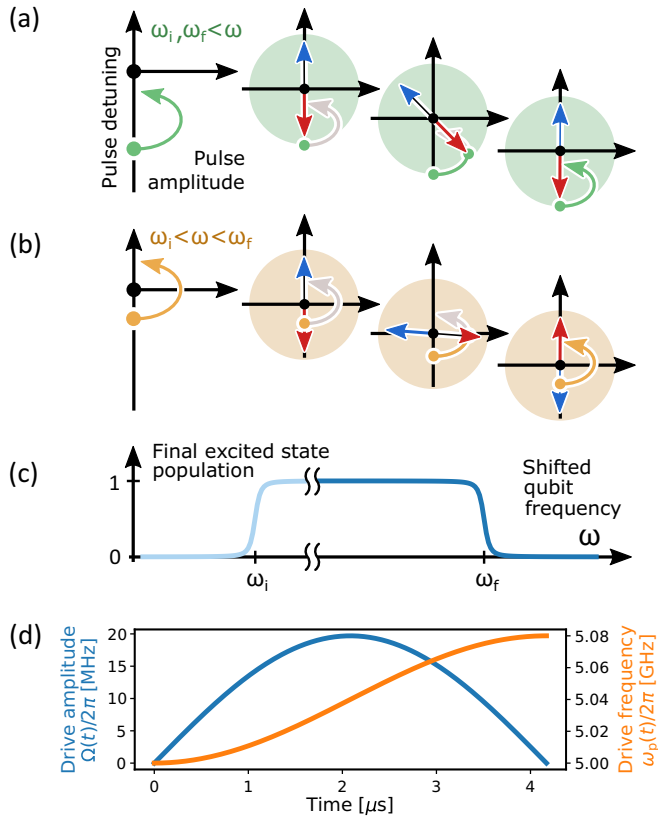
FIG. 3. (a, b) Illustration of the effect of a chirped pulse with initial and final frequencies ω_i and ω_f on a qubit with a frequency ω. (a) If the initial and final detunings have the same sign, then the basis states $|0\rangle$ and $|1\rangle$ are unchanged (up to a phase). (b) If the signs are opposite, then the states get swapped. This results in a dependence of the final state on the shifted qubit frequency shown in panel (c). The perceptron activation function is realized by this dependence in the vicinity of $\omega \approx \omega_f$ (darker line). (d) Time-dependence of the amplitude $\Omega(t)$ and frequency $\omega_p(t)$ of an exemplary chirped pulse realizing the output-qubit rotation.

Figure 4(a) shows the measured effect of the pulse on the qubit state as a function of the bias $b = \omega_f - \omega_1$ for several durations, T, of the pulse. Since the control qubit is left in its ground state, i.e., $x_1 = 0$, the curves are independent of the weight w_1. In these measurements, performed with a pulse amplitude $\Omega_0/2\pi = 19.7$ MHz (as calibrated by a comparison with the π-pulse) and initial frequency $\omega_i = \omega_f - 80$ MHz, the qubit is initialized in its ground state and its excited state population p_e after the pulse is observed to follow a steplike curve, where the slope of the transition increases with the pulse length. The broadening of the response curve due to nonadiabatic effects is well reproduced by unitary simulations, where we simply evolve the driven two-level-system Hamiltonian in the rotating frame, taking the rotating wave approximation. We choose $T = 1.67\,\mu s$ to be the default gate time for the following results, as it provides a compromise between a narrow activation function, maximizing the range that can be sampled with the achievable weights, and fast evolution, minimizing the amount of decoherence introduced by the gate.
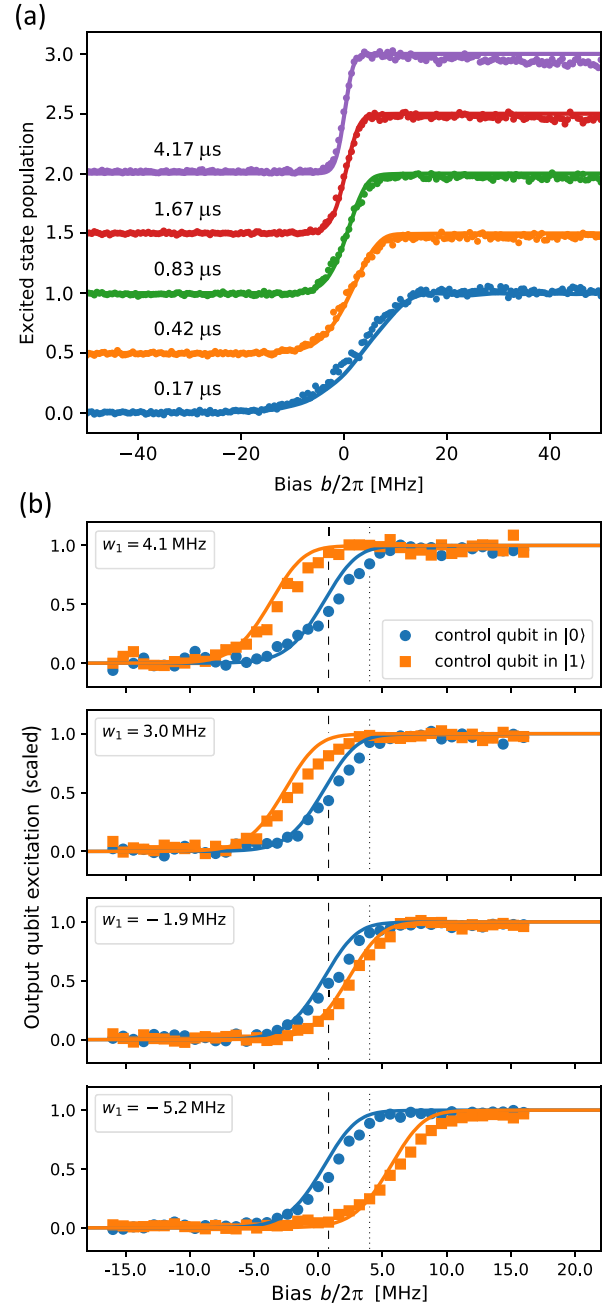

FIG. 4. (a) The quantum perceptron activation functions: The excited state population p_e of the output qubit as a function of its final detuning Δ after a chirped pulse is applied on the ground state of two-qubit system. Results for different pulse lengths T are offset for visualization. The characteristic width of the activation function is inversely proportional to the pulse lengths T. The solid lines result from a simulation of unitary evolution, with no fit parameters in this model. (b) Frequency shift of the activation function dependent on the state of the input qubit, for a pulse length of $T = 1.67\,\mu s$ Both the magnitude and sign of this shift can be varied by tuning the frequency of the coupler. In these plots, the output qubit excitations are rescaled to correct for relaxation and readout imperfections, to highlight the agreement of the curve's shape with the simulation results. The frequency spacing between the simulated curves is not obtained by fitting but rather predicted from measured J values [see Fig. 2(b)]. The vertical lines indicate values of $b/2\pi = 0.8$ MHz (dashed) and $b/2\pi = 4.0$ MHz (dotted) which are used in Fig. 5(a).
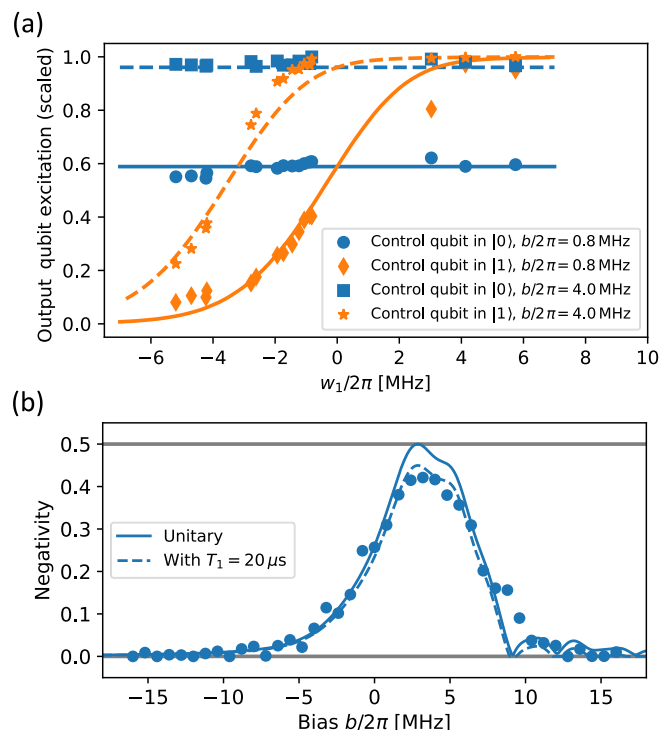
FIG. 5. (a) Final population of the output qubit as a function of the weight $w_1 = -J$ for two different bias values $b/2\pi$ of 0.8 MHz and 4.0 MHz. As expected, when the input qubit is in the $|0\rangle$ state, the final output state does not depend on the weight, while for control state $|1\rangle$ it follows the nonlinear activation function. The solid and dashed lines are derived from numerical simulations with no fit parameters. (b) Negativity of the final two-qubits state after preparing the input qubit in an equal superposition state and the output qubit in its ground state and applying the perceptron gate with a weight $w_1/2\pi = -5.2$ MHz. Negativity of a two-qubit state can range between 0 for a separable state to 1/2 for a maximally entangled state. Lines show the simulated negativity for perfect unitary dynamics (solid) and when including relaxation of the qubits with decay times of $T_1 = 20 \,\mu$s (dashed).

The effect of the ZZ coupling between the output and the input qubit is a shift of the S-shaped activation function dependent on the input qubit being in the excited state, as demonstrated in Fig. 4(b). The curves corresponding to the input qubit being in the ground state (blue) are unshifted, while the ones for an excited input qubit (orange) show a shift which is adjustable by changing the coupler frequency. This shift equals the ZZ coupling strength $J = -w_1$ (where the sign flip is included in the definition of the weight such that the activation function is increasing, as per convention) in the single connection between the output and the input qubit in the quantum perceptron gate. Depending on the relative frequency of the coupler with respect to the qubit frequencies, the weight can be made positive or negative (see Fig. 2).

As the next step toward trainability of the network, we control the weight w_1 at a fixed bias [see Fig. 5(a)]. For the input qubit in the ground state, the final state of the output does not depend on w_1, and is simply given by $f(b)$ since $x_1 = 0$. For the control in the excited state, the dependence on

w_1 is simply the activation function itself, shifted by choosing different values for the bias b.

We further characterize the perceptron gate with quantum process tomography [29] and verify several important aspects of the implemented process for the representative case of $w_1/2\pi = -5.2$ MHz. First, for each bias value, we calculate the average fidelity [30] of the process as compared to the controlled gate

$$U = |0\rangle \langle 0| \otimes R_z(\alpha_0)R_x(\theta_0)R_z(\beta_0)$$
$$+ |1\rangle \langle 1| \otimes R_z(\alpha_1)R_x(\theta_1)R_z(\beta_1)e^{i\gamma}, \quad (4)$$

where the R_i are rotations about the i axis. Since only the angles θ_i determine the populations in the activation function, we set these from the simulation, and optimize over α_i, β_i and γ to obtain a fidelity of approximately 0.86 independent of the bias. Next, to show its entangling property, we prepare the input qubit in an equal superposition state and the output qubit in its ground state. We then apply the gate, extract the density matrix of the final two-qubit state, and calculate its negativity [31], which reaches nearly the maximum value of 1/2 for bias values at which the activation functions for the two computational states of the input qubit are well separated [Fig. 5(b)]. In the limit of large weight, the ideal operation at the mid-point between the activation function slopes would become equivalent (up to local phases) to a controlled NOT gate, preparing a Bell state and reaching maximum negativity. Indeed, the gate is expected to act as a controlled gate—that is, if the input qubit is prepared in one of the basis states $|0\rangle$, $|1\rangle$, then the gate should leave it in this state, independently of the output qubit's initial state. We calculate a fidelity measure capturing how well the process $\mathcal{M}$ satisfies this condition, i.e.,

$$\overline{\mathcal{F}} = \frac{1}{2} \sum_{i \in \{0,1\}} \langle \mathrm{Tr}(|i\rangle\langle i| \otimes \mathbb{1}) \mathcal{M}[|i\rangle\langle i| \otimes |\varphi\rangle\langle\varphi|]\rangle_\varphi, \quad (5)$$

where $\langle \cdot \rangle_\varphi$ denotes averaging over the initial state $|\varphi\rangle$ of the output qubit. This average is calculated from the process matrix of $\mathcal{M}$ by directly using Eq. (5) and the identity $\langle |\varphi\rangle\langle\varphi| \rangle_\varphi = \mathbb{1}/2$. The obtained fidelity values are independent of the bias and lie between 0.95 and 0.97, indicating, in conjunction with the results above, that the implemented operation is to a good approximation a controlled gate. Finally, to properly capture decoherence effects we also evaluate the purity of the final state averaged over the initially prepared state Ψ, i.e.,

$$\overline{\mathcal{P}} = \langle \mathrm{Tr}((\mathcal{M}[|\Psi\rangle\langle\Psi|])^2) \rangle_\Psi, \quad (6)$$

and obtain a value around 0.78, independent of the qubit-drive detuning. This confirms that the gate is mostly a unitary process, limited slightly by qubit decoherence. The average purity value is consistent with typical observed qubit dephasing times which vary in our experiment between 10 μs and 20 μs.

III. EQUIVALENT CIRCUIT AND SCALING COMPLEXITY

For a general, multiqubit input register the perceptron's action in Eq. (1) can be considered as a generalized multiqubit conditional gate: A separate operation $V(x)$ is applied for each of the basis states $|x\rangle \equiv |x_1 x_2 \ldots x_N\rangle$ of the N input qubits.
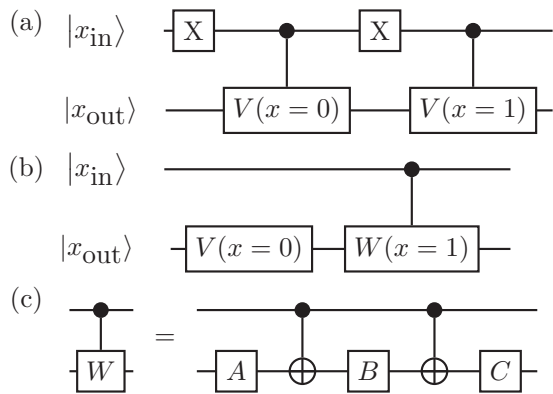
FIG. 6. (a) A circuit that implements the equivalent of the perceptron gate with a single input and a single output qubit. A controlled-V operation needs to be applied for each possible string of the inputs, in this case $x_{\text{in}} \in \{0, 1\}$. (b) The circuit can be simplified by applying one of the multiqubit controlled operations as a single-qubit gate and adjusting the following operations. Here, $W(x = 1) = V(x = 1) \cdot V(x = 0)^{\dagger}$. (c) The controlled-$W$ gate is further decomposed into at most two CNOT gates and three single qubit gates A, B, and C.

This is in contrast to a single multiqubit controlled operation, that applies the operation V to a output qubit only if all input qubits are in the state $|1\rangle$. Thus, to implement the perceptron gate requires a multiqubit controlled gate for each input basis state, in total 2^N operations. This can be reduced to $2^N - 1$ by implementing the $V(x)$ of one of the input strings as a single qubit gate and adjusting all multiqubit controlled operations accordingly. For small $N \lesssim 10$, a multiqubit controlled gate can be decomposed into $2^{N+1} - 2$ two-qubit gates [32], which gives us a total gate count for the equivalent circuit of $N_g = (2^N - 1)(2^{N+1} - 2)$, approaching $\sim 2^{2N+1}$ for large number of input qubits.

More specifically, the single-input perceptron implemented in this paper can be decomposed into single qubit gates and two CNOT gates (Fig. 6). The fidelity and total duration of a sequence is typically limited by the number of two-qubit gates in transmon-based architectures. Hence, when ignoring single qubit gates, the sequence equivalent to the perceptron gate would be implemented in a gate time of ~ 120 ns and with an estimated fidelity of $\sim 99.4\%$. Here, we have assumed state of the art CNOT/CZ gates with fidelities of up to 99.7% [33,34] and gate times of 60 ns. The advantage of our protocol, with approximately constant time, over the equivalent unitary circuit becomes apparent as the number of input qubits is increased, as discussed in Appendix A. For example, when decomposing the perceptron gate with three input qubits the best scenario estimate suggests a fidelity of $\sim 74.5\%$ with a gate time of ~ 5.9 μs.

IV. DISCUSSION AND OUTLOOK

By realizing the perceptron dynamics that activates the output qubit depending on the state of connected input qubits directly in hardware, we have demonstrated the basic building block of a quantum feed-forward neural network. In this codesign approach, we show that by changing the length of the perceptron pulse the shape of the activation function can be modified, whereas the ZZ shift mediated by a tunable coupler and the drive final frequency can be tuned to modify the weights and bias, which allows for training of the neural network.

As an extension of the demonstrated single-digit input scenario, an efficient implementation of a multiinput quantum perceptron can be realized by coupling multiple input qubits, each with its own tunable coupler, to a single output qubit. Its cumulative frequency shift would be of the form shown in Eq. (3) [35]. It is noteworthy that the gate presented here leads to complex multiqubit operations with two-body interactions only and with a gate time that is independent of the number of inputs. In contrast, the time to run the equivalent circuit scales exponentially in the number of qubits, with significant advantage of the codesigned perceptron expected already for the three-qubit input case. While the quantum perceptron we codesign exponentially reduces the number of required conventional gates, all-to-all connectivity between the qubits of adjacent layers is desired to fully benefit from this advantage. Limitations imposed by the number of couplers that can be physically attached to a single qubit or compromises in the connectivity between layers may reduce this advantage. Therefore, multiqubit couplers [36–39] may become increasingly useful, as will be architectures that combine the presented implementation of a perceptron with standard digital gate sequences. Moreover, a recent study shows that the presence of many-body terms can be beneficial to the expressivity of quantum neural networks [40].

To ensure proper trainability of the network, the range of achievable weigths and biases should allow to probe the activation function at both limits $f(x) \approx 1$ and $f(x) \approx 0$. For our implementation, the possible range of ZZ shifts should be comparable to the charateristic width of the activation curve, that is in turn limited by the ability to follow the eigenstates adiabatically given by the speed of the gate. Therefore, to reduce the perceptron gate time one could envisage either nonadiabatic versions of the perceptron [40], or an extended range of achievable ZZ coupling by AC [41,42] or DC [43] pulses on the coupler.

While we have demonstrated the entangling power of the perceptron gate, evaluating quantum advantage in larger networks is subject to forthcoming investigations. For instance, it will be important to understand how decoherence will affect the result and what role the entanglement between the input and output layer plays, in particular if the input layer is in a superposition state. Partially tracing out input and hidden layers and recycling qubits might allow for larger networks and protect the results from decoherence of earlier layers. Moreover, an investigation of the effects of higher connectivity on the expressivity of the nextwork will be required.

Finally, neural networks based on our perceptron gates may also have applications not directly related to machine learning. As an example, it is plausible that their quantum counterparts could serve well to parametrize unitaries in variational quantum algorithms [10,11,13], similar to the efficient approximation capability found in classical neural networks.

APPENDIX A: EQUIVALENT UNITARY

The equivalent unitary matrix for the proposed perceptron can be expressed in terms of standard gates such a two-qubit controlled NOT and single qubit gates. However, the depth of the equivalent circuit grows exponentially with the number of input qubits.

The action of the perceptron in Eq. (1) can be thought of as a generalized multiconditional gate. Whereas a "conventional" N-controlled-V operation applies the gate V to a output qubit only if the control qubit (or qubits, for a multiqubit controlled gate) is in the $|1\rangle$ state, in the perceptron gate for every computational basis state $|x\rangle \equiv |x_1 x_2 \ldots x_N\rangle$ with $x_i \in \{0, 1\}$ of the control qubits a different gate $V(x)$ is applied to the output qubit. The matrix representation for such a controlled-gate has a block-diagonal structure

$$
U = \begin{pmatrix} V(x=0\ldots0) & & & \\ & V(x=0\ldots1) & & \\ & & \ddots & \\ & & & V(x=1\ldots1) \end{pmatrix},
$$

with each block $V(x)$ corresponding to a particular input string x. Thus, the perceptron gate can be expressed in terms of multiqubit controlled unitaries as illustrated for a single input qubit in Fig. 6(a) and for two input qubits in Fig. 7(a). However, the number of multiqubit controlled unitaries can be decreased by one by applying one the $V(x)$ gates as a single qubit gates, e.g., the unitary $V(x=0\ldots0)$ conditioned on all inputs being in the ground state, and adjusting all other multiqubit controlled gates to account for the extra operation, in this example $V(x) \rightarrow W(x) = V(x)V(0\ldots0)^\dagger$. See Figs. 6(b) and 7(b) for one- and two-input qubits examples.

When acting on an output qubit initialized in the $|0\rangle$ state, the perceptron gate corresponds to the map

$$
|x\rangle \otimes |0\rangle \rightarrow |x\rangle \otimes (\sqrt{1 - |c(x)|^2}\,|0\rangle + c(x)\,|1\rangle). \quad \text{(A1)}
$$

By imposing unitarity, we can infer the action of the perceptron gate on a output qubit in the $|1\rangle$ state, and the controlled-$V(x)$ gate with

$$
V(x) = \begin{pmatrix} \sqrt{1 - |c(x)|^2} & c^*(x) \\ -c(x) & \sqrt{1 - |c(x)|^2}. \end{pmatrix}. \quad \text{(A2)}
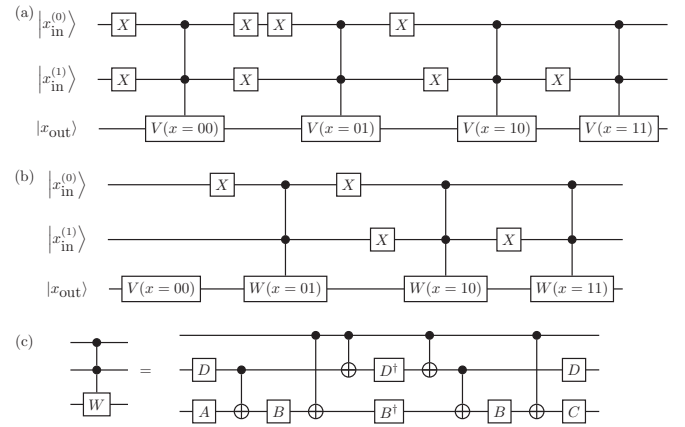$$

FIG. 7. (a) Decomposition of the perceptron gate with two inputs and a single output qubit into controlled-controlled-V operations and single-qubit operations X. (b) Improved circuit decomposition: one of the controlled-controlled operations, $V(00)$, is applied as a single-qubit gate with a simple adjustment $V(x) \rightarrow W(x) = V(x)V(00)^\dagger$ for the other controlled-controlled operations. (c) Decomposition of a controlled-controlled-W gate in terms of 6 CNOT gates and eight single-qubit gates.

For real $c(x)$, we can write $c(x) = \sin\theta_x/2$ and $\sqrt{1 - c^2(x)} = \cos\theta_x/2$ and express $V(x)$ as

$$
V(x) = \begin{pmatrix} \cos\theta_x/2 & \sin\theta_x/2 \\ -\sin\theta_x/2 & \cos\theta_x/2 \end{pmatrix} = R_y(\theta_x). \quad \text{(A3)}
$$

That is to say, the perceptron reduces to a sequence of rotations around the X, Y, and Z axes. The number of multiqubit controlled rotations in this decomposition is equal to the number of possible input bitstrings, which scales exponentially with the number of input qubits. As discussed in the main text, this may lead to a potential advantage of this protocol over more standard paramaterized gates.

Moreover, near-term quantum computers cannot generally implement multiqubit controlled gates at once, but rather must decompose them into a series of one- and two-qubit gates as shown in Figs. 6(c) and 7(c) by using CNOT gates as the primitive two-qubit operation.

In the main text, we use gate times and fidelities quoted for CZ instead of CNOT gates, being $F = 99.7\%$ and $\tau = 60$ ns as these are currently better for state-of-the-art transmon qubit architectures and since a CZ can easily be transformed to a CNOT via single qubit gates.

With these decompositions, and assuming that the two-qubit gates are the dominant source of error, we obtain estimates for the equivalent fidelities and gate times quoted in

TABLE I. Number of CNOT gates N_g required for a circuit equivalent to the adiabatic protocol with N input qubits, along with estimates for the total circuit time t and the circuit fidelity $\mathcal{F}$.

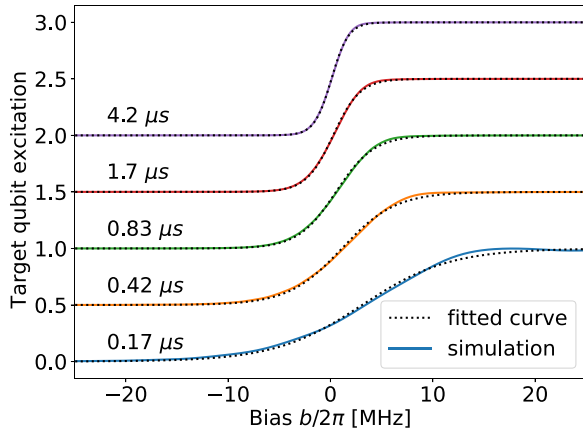| N	1	2	3	4
N_g	2	18	98	450
t (μs)	0.12	1.08	5.88	27
$\mathcal{F}$	0.994	0.947	0.745	0.259

FIG. 8. Comparison for different gate times of the simulated populations, solid line, and fits to the analytical [Eq. (C2)] from Ref. [27]. Results for different pulse lengths T are offset for visualization. The analytical form well describes the simulated results, with small deviations for short gate times.

the main text: for the two-qubit (single-input, single-output) gate, 2 CNOTs are needed, leading to a fidelity of ~99.4% and a gate time of ~120 ns; for the three-qubit (two inputs,

one output) gate 18 CNOTs are needed, leading to a fidelity of ~94.7% and a gate time of ~1.08 μs; for the four-qubit (three inputs, one output) gate 98 CNOTs are needed, leading to a fidelity of ~74.5% and a gate time of ~5.88 μs. Gate time and fidelity estimates for more inputs are summarised in Table I.

APPENDIX B: HAMILTONIAN AND PERTURBATION THEORY

The Hamiltonian describing the system is

$$H/\hbar = \sum_{i=1,2,c} \omega_i a_i^\dagger a_i + \frac{\alpha_i}{2} a_i^\dagger a_i^\dagger a_i a_i$$
$$+ \sum_{i=1,2} g_{ic}(a_i - a_i^\dagger)(a_c - a_c^\dagger), \quad (B1)$$

where $a_i^\dagger$ (a_i) are creation (annihilation) operators, ω_i are the frequencies and α_i are the anharmonicities of qubits ($i = 1, 2$) and coupler ($i = c$), and g_{ic} are their respective couplings. The anharmonicities α_i are defined as the difference $\omega_{12} - \omega_{01}$ between the $0 \leftrightarrow 1$ and $1 \leftrightarrow 2$ transition frequencies; values for the measured parameters can be found in Ref. [42].

Using fourth-order perturbation theory, we can approximate the *ZZ* coupling strength *J* as

$$J = 2g_{1c}^2 g_{2c}^2 \frac{\alpha_1\alpha_2(\Delta_1 + \Delta_2)^2 + \alpha_2\alpha_c\Delta_1^2 + \alpha_1\Delta_1^2(\Delta_1 + \Delta_2) + \alpha_2\Delta_2^2(\Delta_1 + \Delta_2) + \alpha_c(\Delta_1 + \Delta_2)(\Delta_1 - \Delta_2)^2}{\Delta_1^2\Delta_2^2(\Delta_1 + \Delta_2 + \alpha_c)(\Delta_2 - \Delta_1 - \alpha_2)(\Delta_1 - \Delta_2 - \alpha_1)}, \quad (B2)$$

>where $\Delta_{1,2} = \omega_c - \omega_{1,2}$ are the detunings of the coupler from the two qubits.

APPENDIX C: ANALYTIC SOLUTION FOR CHIRPED PULSES

The driven two-level system, although ubiquitous, has proven to notoriously difficult to solve analytically. The most well known analytical solution is the Landau-Zener linear ramp [44]. Soon after, another exact solution was found for a hyperbolic secant pulse [45] and was later extensively used [46], analyzed, and extended [27]. The hyperbolic secant pulse is chirped to have a detuning $\omega_p(t)$ and modulated in amplitude $\Omega(t)$ according to

$$\omega_p(t) = \omega_i + (\omega_f - \omega_i)\tanh^2(\pi t/T),$$
$$\Omega(t) = \Omega_0\mathrm{sech}(\pi t/T). \quad (C1)$$

Assuming that the pulse begins at time $t = -\infty$ and ends at time $t = +\infty$ one can find an analytical expression for the population transfer at the end of the pulse. This is given as

$$P_1(+\infty) = \mathrm{sech}[(\omega_i + \Delta_f)T/2]\mathrm{sech}[(\omega_i - \Delta_f)T/2]\left[\sin^2(\sqrt{\Omega_0^2 + \Delta_f^2}T/2) + \sinh^2(\Delta_f T/2)\right]. \quad (C2)$$

The chirped pulses in our experiment can be obtained from the ones described above by applying the time transformation

$$t \to \mathrm{artanh}[-\cos(\pi * t'/T)] * T$$

and indeed define the same trajectory in the detuning versus amplitude plane. When we compare the analytical solution described above with the simulation results we find agreement in the curve behavior (Fig. 8). Differences in steepness of the curves are due to the time transformation between the two cases. Nonetheless the underlying sigmoid behavior remains, and the curves can be precisely matched by allowing for the fitting parameter T and a further detuning of the whole curve δ, given by $\omega_i \to \omega_i + \delta$.

[1] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, Nature (London) **549**, 195 (2017).

[2] A. Rocchetto, E. Grant, S. Strelchuk, G. Carleo, and S. Severini, Learning hard quantum distributions with variational autoencoders, npj Quantum Inf. **4**, 28 (2018).

[3] S. Das Sarma, D.-L. Deng, and L.-M. Duan, Machine learning meets quantum physics, Phys. Today **72**, 48 (2019).

[4] C. Zoufal, A. Lucchi, and S. Woerner, Quantum generative adversarial networks for learning and loading random distributions, npj Quantum Inf. **5**, 103 (2019).

[5] R. G. Melko, G. Carleo, J. Carrasquilla, and J. I. Cirac, Restricted Boltzmann machines in quantum physics, Nat. Phys. **15**, 887 (2019).

[6] A. Abbas, D. Sutter, C. Zoufal, A. Lucchi, A. Figalli, and S. Woerner, The power of quantum neural networks, Nat. Comput. Sci. **1**, 403 (2021).

[7] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, Supervised learning with quantum-enhanced feature spaces, Nature (London) **567**, 209 (2019).

[8] J. R. Glick, T. P. Gujarati, A. D. Corcoles, Y. Kim, A. Kandala, J. M. Gambetta, and K. Temme, Covariant quantum kernels for data with group structure, arXiv:2105.03406.

[9] A. W. Harrow, A. Hassidim, and S. Lloyd, Quantum Algorithm for Linear Systems of Equations, Phys. Rev. Lett. **103**, 150502 (2009).

[10] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm, arXiv:1411.4028 (2014).

[11] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, A variational eigenvalue solver on a photonic quantum processor, Nat. Commun. **5**, 4213 (2014).

[12] N. Moll, P. Barkoutsos, L. S. Bishop, J. M. Chow, A. Cross, D. J. Egger, S. Filipp, A. Fuhrer, J. M. Gambetta, M. Ganzhorn *et al.*, Quantum optimization using variational algorithms on near-term quantum devices, Quant. Sci. Technol. **3**, 030503 (2018).

[13] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, Nature (London) **549**, 242 (2017).

[14] Y. Liu, S. Arunachalam, and K. Temme, A rigorous and robust quantum speedup in supervised machine learning, Nat. Phys. **17**, 1013 (2021).

[15] E. Behrman, L. Nash, J. Steck, V. Chandrashekar, and S. Skinner, Simulations of quantum neural networks, Inf. Sci. **128**, 257 (2000).

[16] M. Schuld, I. Sinayskiy, and F. Petruccione, The quest for a quantum neural network, Quant. Inf. Process. **13**, 2567 (2014).

[17] A. Pérez-Salinas, D. López-Núñez, A. García-Sáez, P. Forn-Díaz, and J. I. Latorre, One qubit as a universal approximant, Phys. Rev. A **104**, 012405 (2021).

[18] E. Torrontegui and J. J. García-Ripoll, Unitary quantum perceptron as efficient universal approximator, Europhys. Lett. **125**, 30004 (2019).

[19] Y. Cao, G. G. Guerreschi, and A. Aspuru-Guzik, Quantum neuron: An elementary building block for machine learning on quantum computers, arXiv:1711.11240 (2017).

[20] E. Fredkin and T. Toffoli, Conservative logic, Int. J. Theor. Phys. **21**, 219 (1982).

[21] S. Yan, H. Qi, and W. Cui, Nonlinear quantum neuron: A fundamental building block for quantum neural networks, Phys. Rev. A **102**, 052421 (2020).

[22] J. Koch, T. M. Yu, J. Gambetta, A. A. Houck, D. I. Schuster, J. Majer, A. Blais, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, Charge-insensitive qubit design derived from the Cooper pair box, Phys. Rev. A **76**, 042319 (2007).

[23] D. C. McKay, S. Filipp, A. Mezzacapo, E. Magesan, J. M. Chow, and J. M. Gambetta, Universal gate for fixed-frequency qubits via a tunable bus, Phys. Rev. Appl. **6**, 064007 (2016).

[24] X. Li, T. Cai, H. Yan, Z. Wang, X. Pan, Y. Ma, W. Cai, J. Han, Z. Hua, X. Han, Y. Wu, H. Zhang, H. Wang, Y. Song, L. Duan, and L. Sun, Tunable coupler for realizing a controlled-phase gate with dynamically decoupled regime in a superconducting circuit, Phys. Rev. Appl. **14**, 024070 (2020).

[25] M. Roth, N. Moll, G. Salis, M. Ganzhorn, D. J. Egger, S. Filipp, and S. Schmidt, Adiabatic quantum simulations with driven superconducting qubits, Phys. Rev. A **99**, 022323 (2019).

[26] G. Salis, N. Moll, M. Roth, M. Ganzhorn, and S. Filipp, Time-resolved tomography of a driven adiabatic quantum simulation, Phys. Rev. A **102**, 062611 (2020).

[27] F. T. Hioe, Solution of bloch equations involving amplitude and frequency modulations, Phys. Rev. A **30**, 2100 (1984).

[28] J. Baum, R. Tycko, and A. Pines, Broadband and adiabatic inversion of a two-level system by phase-modulated pulses, Phys. Rev. A **32**, 3435 (1985).

[29] M. Mohseni, A. T. Rezakhani, and D. A. Lidar, Quantum-process tomography: Resource analysis of different strategies, Phys. Rev. A **77**, 032322 (2008).

[30] E. Magesan, R. Blume-Kohout, and J. Emerson, Gate fidelity fluctuations and quantum process invariants, Phys. Rev. A **84**, 012309 (2011).

[31] G. Vidal and R. F. Werner, Computable measure of entanglement, Phys. Rev. A **65**, 032314 (2002).

[32] A. Barenco, C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. A. Smolin, and H. Weinfurter, Elementary gates for quantum computation, Phys. Rev. A **52**, 3457 (1995).

[33] M. Kjaergaard, M. E. Schwartz, A. Greene, G. O. Samach, A. Bengtsson, M. O'Keeffe, C. M. McNally, J. Braumüller, D. K. Kim, P. Krantz *et al.*, Programming a quantum computer with quantum instructions, arXiv:2001.08838 (2020).

[34] M. Kjaergaard, M. E. Schwartz, J. Braumüller, P. Krantz, J. I.-J. Wang, S. Gustavsson, and W. D. Oliver, Superconducting qubits: Current state of play, Annu. Rev. Condens. Matter Phys. **11**, 369 (2020).

[35] S. E. Nigg and S. M. Girvin, Stabilizer Quantum Error Correction Toolbox for Superconducting Qubits, Phys. Rev. Lett. **110**, 243604 (2013).

[36] A. Mezzacapo, U. Las Heras, J. Pedernales, L. DiCarlo, E. Solano, and L. Lamata, Digital quantum Rabi and Dicke models in superconducting circuits, Sci. Rep. **4**, 1 (2014).

[37] N. Chancellor, A. Kissinger, J. Roffe, S. Zohren, and D. Horsman, Graphical structures for design and verification of quantum error correction, arXiv:1611.08012.

[38] H. Paik, A. Mezzacapo, M. Sandberg, D. T. McClure, B. Abdo, A. D. Córcoles, O. Dial, D. F. Bogorin, B. L. T. Plourde, M. Steffen, A. W. Cross, J. M. Gambetta, and J. M. Chow, Experimental Demonstration of a Resonator-Induced Phase Gate in a Multiqubit Circuit-QED System, Phys. Rev. Lett. **117**, 250502 (2016).

[39] M. Sameti, A. Potočnik, D. E. Browne, A. Wallraff, and M. J. Hartmann, Superconducting quantum simulator for topological order and the toric code, Phys. Rev. A **95**, 042330 (2017).

[40] Y. Ban, X. Chen, E. Torrontegui, E. Solano, and J. Casanova, Speeding up quantum perceptron via shortcuts to adiabaticity, Sci. Rep. **11**, 5783 (2021).

[41] A. Noguchi, A. Osada, S. Masuda, S. Kono, K. Heya, S. P. Wolski, H. Takahashi, T. Sugiyama, D. Lachance-Quirion, and Y. Nakamura, Fast parametric two-qubit gates with suppressed residual interaction using the second-order nonlinearity of a cubic transmon, Phys. Rev. A **102**, 062408 (2020).

[42] M. Ganzhorn, D. J. Egger, S. Filipp, G. R. von Salis, and N. Moll, Cross-talk compensation in quantum processing devices (2019), U.S. Patent 10,452,991.

[43] H. Xu, W. Liu, Z. Li, J. Han, J. Zhang, K. Linghu, Y. Li, M. Chen, Z. Yang, J. Wang *et al.*, Realization of adiabatic and diabatic CZ gates in superconducting qubits coupled with a tunable coupler, Chin. Phys. B **30**, 044212 (2021).

[44] C. Zener, Non-adiabatic crossing of energy levels, Proc. R. Soc. London, Ser. A **137**, 696 (1932).

[45] N. Rosen and C. Zener, Double Stern-Gerlach experiment and related collision phenomena, Phys. Rev. **40**, 502 (1932).

[46] S. L. McCall and E. L. Hahn, Self-induced transparency, Phys. Rev. **183**, 457 (1969).