Neuromorphic quantum computing

Christian Pehle [*]
Kirchhoff-Institute for Physics, Heidelberg University, Im Neuenheimer Feld 227, 69120 Heidelberg, Germany

Christof Wetterich [†]
Institute for Theoretical Physics, Heidelberg University, Philosophenweg 16, 69120 Heidelberg, Germany

(Received 7 May 2021; revised 28 February 2022; accepted 9 August 2022; published 31 October 2022)

Quantum computation builds on the use of correlations. Correlations could also play a central role for artificial intelligence, neuromorphic computing or "biological computing." As a step toward a systematic exploration of "correlated computing" we demonstrate that neuromorphic computing can perform quantum operations. Spiking neurons in the active or silent states are connected to the two states of Ising spins. A quantum density matrix is constructed from the expectation values and correlations of the Ising spins. We show for a two qubit system that quantum gates can be learned as a change of parameters for neural network dynamics. These changes respect restrictions which ensure the quantum correlations. Our proposal for probabilistic computing goes beyond Markov chains and is not based on transition probabilities. Constraints on classical probability distributions relate changes made in one part of the system to other parts, similar to entangled quantum systems.

DOI: 10.1103/PhysRevE.106.045311

I. INTRODUCTION

The particularity and strength of quantum computations relies on the use of correlations between different parts of a quantum system. For entangled quantum states the state of a given qubit is intrinsically related to the state of other qubits. Correlated systems act as a whole and cannot be separated into individual parts. This "global character" of the computation constitutes the potential of the novel algorithms for quantum computing. In a quantum system the correlations are encoded in the wave function, or more efficiently in the density matrix whose elements are related directly to certain correlation functions.

Correlations are not a unique feature of quantum systems. They are central ingredients for many aspects of nature. This raises the question if correlations are used effectively for other forms of "computing" by systems that do not obey the strict requirements of quantum systems, such as isolation from the environment, a nondissipative character of the preserved information or the avoidance of decoherence. It seems to us almost certain that "biological computation," as pattern recognition or memory in biological brains, employs correlations. We suggest that also forms of neuromorphic computing or artificial neural networks rely heavily on the use of correlations. The formal understanding of the use of correlations in computing is, however, not yet developed very far. The present note should be considered as a step in the direction outlined in Ref. [1], which proposes a general framework for computations in a stochastic environment which can be non-Markovian.

It has already been argued that several important tasks within quantum computation can be performed by neural networks [2–4] and frameworks exist which enable the implementation of several of these techniques (cf. [5] and references therein). Beyond this, the proposal of Refs. [1,6] suggests that a full quantum computation may be possible by use of stochastic information in a system of neurons. The present paper proposes a concrete implementation of quantum operations by spiking neurons. Our system learns the preservation of "quantum constraints" which guarantees the properties of quantum correlations. We propose that our system of spiking neurons can be realized as a physical system in a neuromorphic computer.

We demonstrate four important steps in the direction of correlated computing:

(1) We describe spiking neurons by a system of differential equations with synaptic weight parameters W. Expectation values and correlations of Ising spins can be extracted by following the evolution of the states of these neurons according to the dynamical system of interacting neurons. The correlations depend on the parameters W. We demonstrate that for every given two-qubit density matrix ρ the system of neurons can adapt or learn suitable parameters W such that expectation values and correlations of the Ising spins can be used to encode this density matrix.

(2) We show by a simulation of neuromorphic computing that the correlation map [1] from expectation values and correlations of classical Ising spins to a quantum density matrix is complete in the case of two qubits. This means that for every possible quantum density matrix ρ there exists a probability distribution p for the configurations of classical Ising spins such that ρ is realized by the correlation map. This completeness was not proven before.

[*]christian.pehle@kip.uni-heidelberg.de
[†]c.wetterich@thphys.uni-heidelberg.de

(3) We establish how unitary transformations of ρ corresponding to arbitrary quantum gates can be performed by a change in probability distributions p.

(4) We demonstrate that the minimal correlation map is not a complete bit-quantum map for three or more qubits. We propose an extended correlation map.

The quantum operations are performed by changes in the probability distributions and corresponding changes in the expectation values and correlations. These changes cannot be realized by Markov chains for which otherwise deterministic operations are performed with certain "transition probabilities." Probabilistic computing beyond Markov chains opens a rich area of new possibilities.

So far our approach centers on the use of expectation values and particular correlations of Ising spins for the construction of quantum density matrices. We concentrate first on two qubits and discuss a possible scaling to a higher number of qubits at the end. Correlated Ising systems are efficient models for many biological systems [7], and several aspects of the present paper can be taken over directly.

II. CLASSICAL AND QUANTUM CORRELATIONS

Also classical computing can use correlations but in a rather restricted way. A given algorithm may imply that two bits are always the same (maximal correlation) or opposite (maximal anticorrelation). It is, however, in general, not possible to extract from the values or correlations of a few given bits information on the state of many other bits. Quantum correlations, and presumably the correlations in artificial neural networks, neuromorphic computing or biological systems, are of a different more global nature. They link many parts of the system, and changes in one part affect directly many other parts.

Since quantum correlations are well understood formally, and they are well suited for a first investigation of this more global type of correlation. In this paper we demonstrate that neuromorphic computing can preserve quantum correlations for effective systems of two qubits by learning appropriate parameters which govern the dynamics of the firing of neurons. Whereas quantum systems are perhaps rather special systems for our more general interest in "correlated computing," a focus on the correlations in quantum systems offers two important advantages. First, performing arbitrary unitary transformations solves a well-understood task by the use of correlations. Second, there is control of the formal aspects by the quantum formalism. These two aspects contrast to other systems which almost certainly use correlations as artificial or biological intelligence and render first steps in an investigation of correlated computing easier. Our aim is not to propose a particularly efficient way of performing unitary operations. It remains open if an extension of the present method to many qubits is feasible with reasonable resources. We rather want to progress in the conceptional understanding of computing by use of correlations. The interesting perspectives are new or existing forms of probabilistic computing that use correlations, even though these correlations may not be pure quantum correlations. For example, biological systems are rather unlikely to employ precise quantum correlations, whereas it seems highly likely that some type of more global correlations are the key to efficiency for certain tasks.

Beyond the important aspect of a controlled formalism our investigation of a simple quantum system has also interesting conceptual consequences for quantum mechanics. We not only show that unitary transformations can be performed by classical statistical systems. This task can be performed much more efficiently by classical computers, which can be viewed as limiting cases of statistical systems as well. In our approach also the discrete nature of observables is directly realized, given by the discrete values of the Ising spins associated to active or quiet neurons or their correlations. Both the discrete spectrum of possible measurement values and the continuous expectation values are implemented directly. The possibility to measure the discrete observables is not given by a numerical solution of a discretized Schrödinger equation by a classical computer.

We first focus on six-system neurons in an dynamical environment of additional neurons. The correlation map constructs a quantum density matrix from expectation values and correlations of the six-system neurons. The probabilistic states associated with a particular density matrix typically correlate all six-system neurons. The system is, therefore, a good starting point for an investigation of "global correlations."

III. CORRELATION MAP FOR TWO QUBITS

The correlation map for a two-qubit quantum system involves six classical Ising spins [1] that are grouped in two pairs of three Ising spins $s_k^{(i)}$. Here $i = 1, 2$ corresponds to the two quantum spins, and $k = 1 \cdots 3$ is associated with the three Cartesian directions of a given quantum spin. One could, in principle, employ a higher number of Ising spins for a "classical computation" of a two-qubit quantum system. The six classical spins are sufficient for an identification of the Cartesian directions of quantum spins with the classical spins, both for the possible measurement values of observables (spectrum) and their expectation values. Also the classical correlation functions for the spins of different qubits coincide with the corresponding quantum correlations.

We want to employ the expectation values $\langle s_k^{(i)} \rangle$ and the correlations $\langle s_k^{(i)} s_l^{(j))} \rangle, i \neq j$, for the construction of the density matrix for the two-qubit system. For this purpose we can form a real 4×4 matrix χ of expectation values and correlations as

$$\chi_{00} = 1, \quad \chi_{0k} = \langle s_k^{(1)} \rangle, \quad \chi_{l0} = \langle s_l^{(2)} \rangle, \quad \chi_{kl} = \langle s_k^{(1)} s_l^{(2)} \rangle. \quad (1)$$

If we denote by $\tau_k, k = 1 \cdots 3$ the three Pauli matrices and by τ_0 the identity matrix, we can define the $U(4)$ generators,

$$L_{\mu\nu} = \tau_\mu \otimes \tau_\nu, \quad \mu = 0 \cdots 3, \quad \nu = 0 \cdots 3. \quad (2)$$

The bit quantum map organizes the expectation values and correlations (1) into a density matrix,

$$\rho = \tfrac{1}{4} \chi_{\mu\nu} L_{\mu\nu}. \quad (3)$$

(Summation over double indices is implied.) We denote this map by

$$f: \mathbf{R}^{4\times4} \to \mathbf{C}^{4\times4}, \quad \chi \mapsto \rho = \tfrac{1}{4} \chi_{\mu\nu} L_{\mu\nu}. \quad (4)$$

The density matrix ρ is a complex Hermitian 4×4 matrix with $\mathrm{tr}(\rho) = 1$. In terms of the real and imaginary parts C_R and C_I of a complex matrix C we define the real 8×8-matrix A as

$$A = 1_2 \otimes C_R + I_2 \otimes C_I = \begin{pmatrix} C_R & -C_I \\ C_I & C_R \end{pmatrix}. \tag{5}$$

This construction is compatible with matrix multiplication and can be used to define a real representation of the density matrix ρ and of unitary transformations acting on it.

A quantum density matrix has to be positive—all eigenvalues λ of ρ have to obey $\lambda \geqslant 0$. This imposes restrictions on the classical probability distribution p or the expectation values (1) that can realize a quantum density matrix. These restrictions are called the quantum constraints. The quantum constraints from the positivity of the density matrix entail important relations between expectation values and correlations for the two qubits. These are crucial for many quantum aspects as the uncertainty relation for spins of a given qubit in different directions. The quantum constraints are preserved by unitary transformations. As long as the quantum constraints of a positive density matrix are obeyed, the classical expectation values and correlations (1) coincide precisely with their counterparts in the quantum system.

In particular, all relations between the classical correlations and the expectation values (1) follow precisely the laws of quantum mechanics. For example, $s_1^{(1)}$ and $s_2^{(1)}$ cannot have simultaneously sharp values. On the other hand, for a density matrix (3) realizing a pure quantum state with $\langle s_k^{(i)} \rangle = 0$ we know that some of the correlations $\langle s_k^{(i)} s_l^{(j)} \rangle$ have to differ from zero. The quantum constraints enforce directly that changes in one part of the system entail changes in other parts. An important part of the present paper demonstrates that a neuromorphic computer can learn transformations of the expectation values and correlations (1) that preserve the quantum constraints.

The states $\tau = 0 \cdots 63$ of the classical spin system correspond to the $2^6 = 64$ configurations of Ising spins $\{s_1^{(1)}, s_2^{(1)}, s_3^{(1)}, s_1^{(2)}, s_2^{(2)}, s_3^{(2)}\}$ that can take the values of $s = \pm 1$. The classical probabilistic system is defined by associating with each τ a probability $p_\tau \geqslant 0, \quad \sum_\tau p_\tau = 1$. The expectation values (1) are formed in the standard way, multiplying the values ± 1 of $(s_\gamma)_\tau$ or $(s_\gamma)_\tau (s_\delta)_\tau$ in a given state τ with p_τ and summing over τ. This results in a weighted sum of probabilities,

$$\chi_{k0} = \sigma_\tau^{(k0)} p_\tau, \quad \chi_{0k} = \sigma_\tau^{(0k)} p_\tau, \quad \chi_{kl} = \sigma_\tau^{(kl)} p_\tau, \tag{6}$$

with signs $\sigma_\tau^{(a)} = \pm 1$ given by

$$\sigma_\tau^{(k0)} = (-1)^{1+\mathrm{bin}(\tau)[k]}, \tag{7}$$

$$\sigma_\tau^{(0k)} = (-1)^{1+\mathrm{bin}(\tau)[3+k]}, \tag{8}$$

$$\sigma_\tau^{(kl)} = (-1)^{1+\mathrm{bin}(\tau)[k]} (-1)^{1+\mathrm{bin}(\tau)[3+l]}. \tag{9}$$

Here $\mathrm{bin}(\tau)$ denotes the six-bit binary representation of a number $0 \cdots 63$ and $\mathrm{bin}(\tau)[k]$ the entry at index k of that vector. For the example $\tau = (1, -1, -1, 1, 1, 1)$ one has $\mathrm{bin}[\tau] = (1, 0, 0, 1, 1, 1)$, $\mathrm{bin}(\tau)[2] = 0$ reads out the value at the second place, and $\sigma_\tau^{(20)} = -1$ is the value of the spin

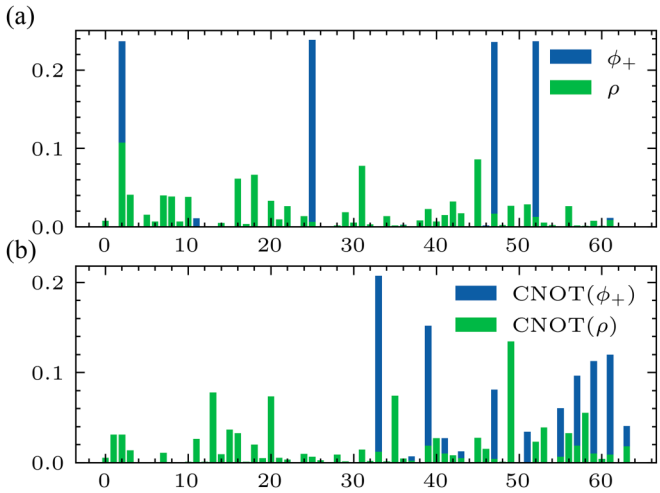

FIG. 1. Classical probabilities for quantum states. In (a) we indicate the probabilities corresponding to the density matrices of the maximally entangled pure state $\psi_+ = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ (blue) and a randomly generated density matrix ρ (green). We also show their transformation under a (b) CNOT gate. The labels $0 \cdots 63$ are best thought of as bit vectors and label the state of six classical Ising spins $s_k^{(i)}$, $i = 1, 2$, $k = 1 \cdots 3$. For example, the label 3 corresponds to the spin state $[-1, -1, -1, -1, 1, 1]$. This figure demonstrates that entangled quantum states can be realized by classical probability distributions, and quantum gates by changes in these probability distributions.

$s_2^{(1)}$ in state τ. We denote this map by

$$g: \mathbf{R}^{64} \to \mathbf{R}^{4\times 4}, \quad p \mapsto \chi. \tag{10}$$

The bit quantum map b can be seen as a map from the classical probability distribution p to the quantum density matrix,

$$b = f \circ g: \mathbf{R}^{64} \to \mathbf{C}^{4\times 4}, \quad p \mapsto \rho. \tag{11}$$

It has the property that the particular classical expectation values and correlations (6) coincide with the expectation values of quantum spins in the Cartesian directions and the corresponding quantum correlations. The density matrices can be constructed (or reconstructed) by use of the particular classical or quantum correlations. This is analogous to the reconstruction of the density matrix for photons by appropriate correlations [8–10]. Examples for probability distributions p realizing particular quantum density matrices ρ are shown in Fig. 1. In Fig. 2 we indicate the corresponding spin expectation values and correlations.

IV. COMPLETENESS OF THE CORRELATION MAP FOR $Q = 2$

Our first task is a demonstration that the bit quantum map b is complete, in the sense that for every positive Hermitian normalized quantum density matrix there exists, at least, one classical probability distribution p which realizes ρ by use of Eq. (11). For this task it will be convenient to work with the "classical wave function" q which is a type of probability

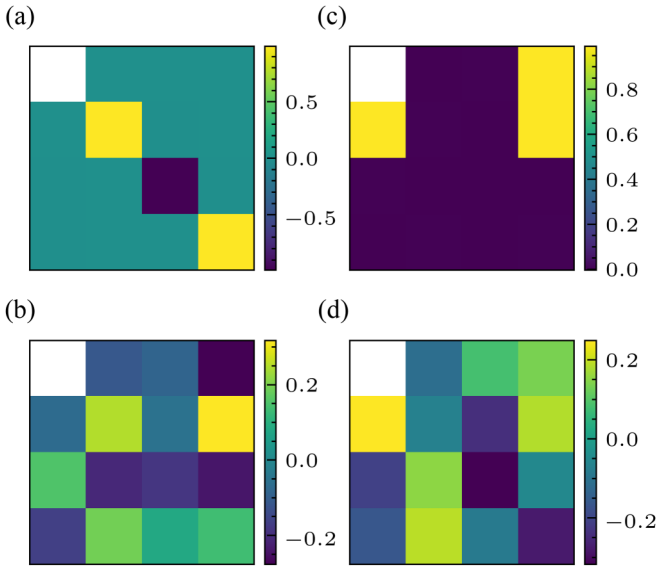

FIG. 2. We show the $15 = 2 \times 3 + 9$ spin expectation values $\langle s_k^{(1)} \rangle$ and correlations $\langle s_k^{(1)} s_l^{(2)} \rangle$ corresponding to (a) ψ_+ and (b) ρ. The first row contains the three expectation values $\langle s_k^{(1)} \rangle$, the first column the three expectation values $\langle s_k^{(2)} \rangle$. The remaining 3×3 entries are made up of the correlations $\langle s_k^{(1)} s_l^{(2)} \rangle$. The transformed spin expectation values and correlations related to the density matrices CNOT(ψ_+) and CNOT(ρ) are shown in (c) and (d) respectively. Note the different color scale between (a) and (b) and (c) and (d).

amplitude. Its components q_τ are related to p_τ by

$$p_\tau = \frac{q_\tau^2}{\sum_\tau q_\tau^2}. \tag{12}$$

This defines one further map,

$$h: \mathbf{R}^{64} \to \mathbf{R}^{64}, \quad q \mapsto p = \frac{q^2}{|q|_2^2}. \tag{13}$$

In short, a classical wave function defines the classical probabilities and, therefore, the expectation values and correlations (5). In turn, these define the complex or real representations of the density matrix by Eq. (4). This associates a density matrix to every classical wave function. We will see that the map from the classical wave function to the density matrix is not invertible.

Changes in the probability distribution correspond to rotations of the vector q with $p_\tau > 0$ and $\sum_\tau p_\tau = 1$ guaranteed by the construction (12). We need to find for each given density 4×4 matrix $\rho \in \mathbf{C}^{4\times4}$ a vector $q \in \mathbf{R}^{64}$ such that it maps to the density matrix ρ under the composition $b \circ h$. For most density matrices q is not unique as can be verified by simple examples [1].

In order to find one such q we minimize

$$l_\rho(q) = |(b \circ h)(q) - \rho|_2^2, \tag{14}$$

by gradient descent on q. To verify numerically that this approach works we performed the following test: Starting with a randomly generated density matrix ρ_0, we iteratively apply the three unitary transformations CNOT, Hadamard $H \otimes I$ and rotation $R_{1/8} \otimes I$ to obtain density matrices ρ_i. The quantum

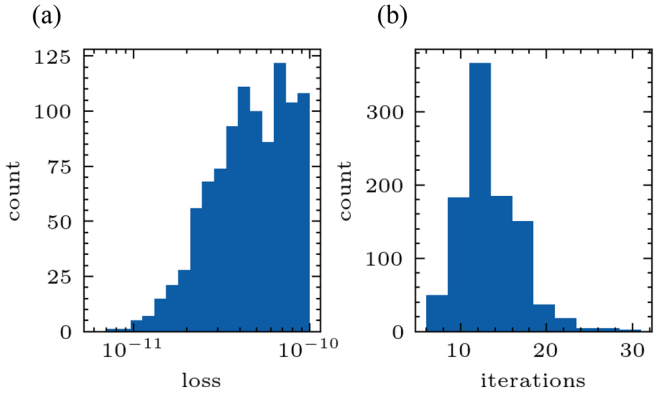

FIG. 3. Optimization results for finding vectors $q_i \in \mathbf{R}^{64}$, such that the loss $l_{\rho_i}(q_i)$ (14) is minimized. We generate 10^3 density matrices ρ_i by starting from a randomly generated density matrix ρ_0 and iteratively apply three unitary transformations CNOT, Hadamard $H \otimes I$ and rotation $R_{1/8} \otimes I$. The optimization is stopped once the loss falls below 10^{-10}. (a) is a histogram of the resulting final losses, and (b) is a histogram of the required iterations.

gates preserve the positivity of ρ. Applying them many times leads to a dense covering to the space of quantum density matrices, which becomes a full covering if the number of steps goes to infinity. We further explore the space of ρ by investigating different ρ_0. For each ρ_i we solve the optimization problem with loss $l_{\rho_i}(q)$ to obtain corresponding vectors q_i. The results of this optimization are shown in Fig. 3. For all ρ_i's we find vectors q realizing the bit quantum map $b \circ h$ with high precision already after rather few iterations. This strongly suggests that for two qubits the bit quantum map is complete.

Considering one of the vectors q found in this way as a representation of ρ, our second step asks how a given unitary transformation of ρ can be represented as a transformation of q. Considering q as a classical probabilistic object, we ask how the classical system can learn a quantum operation. The goal is to find for any vector $q \in \mathbf{R}^{64}$ a matrix $M \in \mathbf{R}^{64\times64}$, such that the transformed vector Mq yields the density matrix $U\rho U^\dagger$ related to ρ by a given unitary transformation U. Since unitary transformations preserve the positivity of the density matrix, a probabilistic system that learns the transformation Mq also learns how the preserve the quantum constraint for the correlations of the system. Even further, since unitary transformations preserve a pure quantum state, the associated classical probabilistic system learns all the relations between expectation values and correlations that characterize a pure quantum state.

A simple learning or optimization goal consists of minimizing the Frobenius norm between the two density matrices $U\rho U^\dagger = U(b \circ h)(q)U^\dagger \equiv U[(b \circ h)(q)]$ and $(b \circ h)(Mq)$,

$$l_U(M) = |U[(b \circ h)(q)] - (b \circ h)(Mq)|_2^2. \tag{15}$$

Here we use the notation $U(\rho) = U\rho U^\dagger$. This minimization can again be implemented using gradient descent. In Figs. 1 and 2 we show the resulting probabilities,

$$p = h(q), \quad p' = h(Mq), \tag{16}$$

as well as the corresponding matrices of expectation values and correlations,

$$\chi = (g \circ h)(q), \quad \chi' = (g \circ h)(Mq), \tag{17}$$

when this minimization procedure is applied to two example initial density matrices and their associated q vectors. The unitary transformation U is taken to be CNOT. We have investigated many different ρ's and always found a satisfactory M with this procedure. Thus, the classical system learns the matrix M necessary for a unitary quantum gate. We emphasize that M depends on q such that $Mq = M(q)q$ is a nonlinear map [1].

V. QUANTUM COMPUTING WITH SPIKING NEURONS

We finally turn to our third task of implementation by neuromorphic computing. So far the Ising spins s entered only indirectly through their state probabilities p and associated correlations and expectation values χ. We next want to study classical systems for which correlations and expectation values can be determined from temporal averages. In the context of neuroscience there is a long history of applying spin-glass models to the study of biological neural networks [7]. The general idea is to consider each biological neuron to have two states *active* and *silent*. A neuron is considered active (or refractory) if it has produced a *spike* within a certain sampling time window and silent otherwise. Within this framework, experimental work has been carried out to study pairwise and higher correlations of biological neurons (e.g., Refs. [11,12]).

One immediate way of obtaining the state probabilities p is from *neural sampling* [13,14]. Since we need only the expectation values and correlations (1) we directly focus on these quantities and do not aim to resolve the probability distributions completely. The six spin variables $s_k^{(i)}(t)$ are associated with six particular neurons in a larger network with $s = 1$ if the neuron is active (refractory) and $s = -1$ if it is silent. Expectation values can be formed by measuring the duration of active and silent states,

$$\langle s_k^{(i)} \rangle = \frac{1}{T} \int_0^T s_k^{(i)}(t) dt, \tag{18}$$

$$\langle s_k^{(1)} s_l^{(2)} \rangle = \frac{1}{T} \int_0^T s_k^{(1)}(t) s_l^{(2)}(t) dt. \tag{19}$$

Instead of spin variables s which take values $\{1, -1\}$, we want to consider variables z with values in $\{1, 0\}$, $s = 2z - 1$. A given selected neuron has the value of $z = 1$ during the refractory period (active state), and $z = 0$ otherwise (silent state). We choose a simple so-called *leaky-integrate and fire* (LIF) neuron model. The model is specified by $3n$ state variables $v_i, I_i, r_i, i = 1 \cdots n$, called the membrane voltages, synaptic currents, and refractory states. The dynamics has phases of continuous evolution and jumps or spikes. The continuous evolution obeys the differential equations,

$$\tau_{\text{mem}} \dot{v} = [1 - \Theta(r)](v_l - v + RI) \tag{20}$$

$$\tau_{\text{syn}} \dot{I} = -I + I_{\text{in}} \tag{21}$$

$$\dot{r} = -\frac{1}{t_{\text{refrac}}} \Theta(r). \tag{22}$$

Here Θ denotes the Heaviside function and multiplication in (20) is pointwise for every i separately. Characteristic for a spiking neuron model are the jumps or discrete changes in the state variables v, I, r at particular times. Whenever one of the membrane voltages v_i reaches a threshold $(v_{\text{th}})_i$ during the continuous evolution,

$$v_i - (v_{\text{th}})_i = 0, \tag{23}$$

the state variables undergo discontinuous jumps. The transition equations,

$$v^+ = v^- + \xi(v_r - v_{\text{th}}), \tag{24}$$

$$I^+ = I^- + W_{\text{rec}} \xi, \tag{25}$$

$$r^+ = r^- + \xi, \tag{26}$$

specify the state before and after the transition, $v^\pm = v(t^\pm)$, $I^\pm = I(t^\pm)$, and $r^\pm = r(t^\pm)$. Here $\xi \in \mathbf{R}^n$ is a binary vector, $\xi_i = 1$ for the value of i for which the threshold voltage is reached and (23) obeyed, and $\xi_j = 0$ for $v_j \neq (v_{\text{th}})_j$. In Eq. (24) the multiplication is pointwise such that only the voltage of the spiking neuron changes. In (23)–(26) v_{th}, v_r, $v_l \in \mathbf{R}^n$ are the *threshold*, *reset*, and *leak* potential. The variables τ_{mem} and τ_{syn} are the membrane and synaptic time constants and R is a resistance chosen to be 1. Finally I_{in} is an input current and W_{rec} is a matrix in $\mathbf{R}^{n \times n}$ which parametrizes the response of the currents I_j of all neurons to the "firing" of the spiking neuron i.

Equation (22) specifies a linear decrease in r_i until $r_i = 0$ is reached. As long as $r_i > 0$ the neuron is considered to be active. The neuron remains active for a refractory period t_{refrac} after its firing at t_k. For this period its voltage does not change,

$$\dot{v}_i([t_k, t_k + t_{\text{refrac}}]) = 0, \tag{27}$$

as implied by Eq. (20).

Ising spins $s(t)$ or the associated occupation numbers $z(t)$ are defined by

$$z(t) = \Theta[r(t)], \tag{28}$$

which is one during the refractory period and zero otherwise. Six selected neurons define the Ising spins $s_k^{(i)} = 2z_k^{(i)} - 1$ for which the expectation values and correlations (18) and (19) define the quantum density matrix ρ by equations (1) and (3).

The last quantity to be specified is the input current to the n LIF neurons. We model spike input from m additional input spike sources. At times t_l the source neuron q_l fires. The response of the n LIF neurons is given by

$$(I_{\text{in}})_j = \sum_l (W_{\text{in}})_{j,q_l} \delta(t - t_l), \quad q_l \in 1, \ldots, m. \tag{29}$$

Here $\delta(t - t_l)$ is the Dirac-δ distribution such that we model an immediate response of all n LIF neurons to every input spike at t_l. The $n \times m$ matrix $W_{\text{in}} \in \mathbf{R}^{n \times m}$ parametrizes the height of the jump in the currents I_j upon arrival of a spike at input source q_l. In our experiments the arrival times t_l of the input spikes are the result of m-independent Poisson point processes with rates λ_q, $q = 1 \cdots m$.

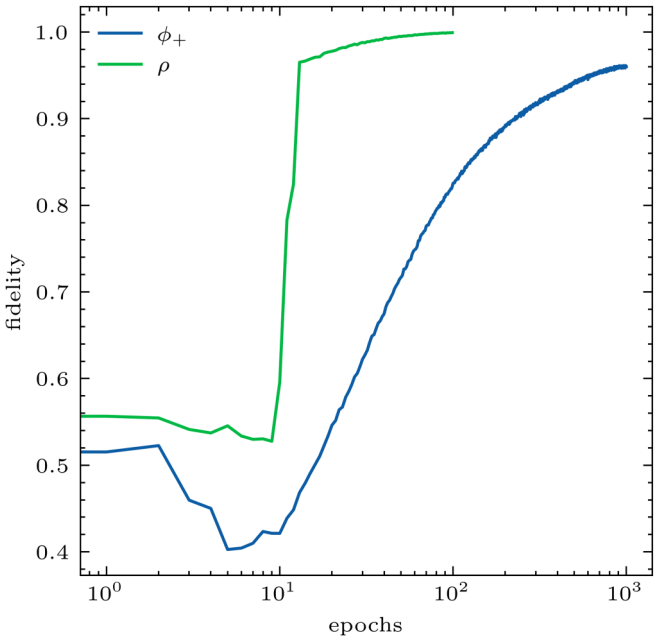

FIG. 4. Fidelity of density matrices obtained by correlations of spiking neurons as a function of training epochs. Shown are approximation results for the density matrix of a maximally entangled pure state $\phi_+ = \frac{1}{\sqrt{2}}(|11\rangle + |00\rangle)$ and a random density matrix ρ.

By solving the differential equations with jumps for given input spikes we can compute $z_j(t)$ and, therefore,

$$s_k^{(i)}(t) = 2z_k^{(i)}(t) - 1. \tag{30}$$

In the experiments we carried out we took $n > 6$. The six spin variables $s_k^{(i)}$ are recovered by projecting to the first six components,

$$\pi : \mathbf{R}^n \to \mathbf{R}^6, \quad (s_1, \ldots, s_n) \mapsto (s_1, \ldots, s_6). \tag{31}$$

Given a density matrix ρ, we can now formulate an optimization problem for W_{in} and W_{rec}. One obstacle in doing so is that the jumps introduce discontinuities, which need to be taken into account. We choose here to solve this issue by introducing a smooth approximation to the Heaviside function Θ_ϵ. More specifically we use a fast sigmoid as in Ref. [15] with parameter $1/\epsilon = 100$. This is a common way in which spiking neuron models can be made amendable to gradient descent optimization [15–18]. The loss function is then,

$$l_\rho(W_{\text{rec}}, W_{\text{in}}) = |\rho - f\{\chi[\pi(s)]\}|_2^2, \tag{32}$$

where χ is defined as before (1) by the spin expectation values and correlations in Eqs. (18) and (19).

In Figs. 4 and 5 we show the result of the optimization process for the spin expectation and correlation matrices χ. We also indicate in Fig. 6 a view of the resulting recurrent weight matrix W_{rec} restricted to the spins $s_k^{(i)}$. The membrane threshold is set to one, $v_{\text{th}} = 1$, and the leak and reset potentials to zero, $v_l = v_r = 0$. The synaptic and membrane time constants are $\tau_{\text{syn}} = 5$ and $\tau_{\text{mem}} = 10\,\text{ms}$, and we integrate for $T = 10^5$ time steps. As a typical example for units in present neuromorphic computing we may associate the integration step to $dt = 1\,\text{ms}$. We choose an input dimension

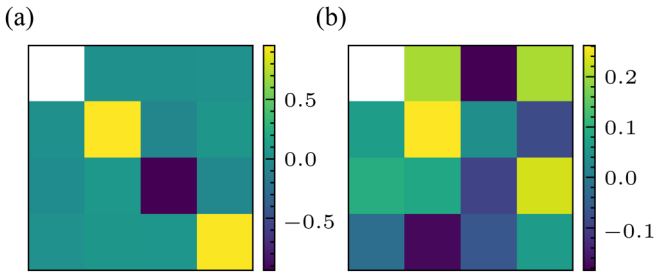

FIG. 5. Spin expectation and correlation matrices χ corresponding to the pure state (a) ψ_+ and (b) ρ, respectively.

of $m = 128$ and consider $n = 64$ recurrently connected LIF neurons. We set $t_{\text{refrac}} = dt$ (this eliminates the need to take the equations for the refractory state into account) and draw the input spikes with Poisson frequency $\lambda = 700\,\text{Hz}$. The learning rate of the gradient descent starts at $\eta = 10$ and is exponentially decreased with a decay constant of $1/100$. The numerical implementation was performed in JAX [19] using simple forward Euler integration.

For the result of the optimization we plot the fidelity, which is a common measure to judge how well a given quantum state is approximated. The fidelity is defined by $F(\rho, \sigma) = (\text{tr}\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}})^2$. Instead of optimizing for the fidelity directly we minimize the square of the Frobenius norm $|\rho - \sigma|_2^2$. By the Fuchs–van de Graaf inequalities we know that $1 - \sqrt{F(\rho,\sigma)} \leqslant \frac{1}{2}|\rho - \sigma|_1 \leqslant \sqrt{1 - F(\rho,\sigma)}$. Since the trace norm is, in turn, bounded by the Frobenius norm the fidelity approaches one as the square of the Frobenius norm goes to zero. Computing the fidelity directly is computationally more expensive and has the added disadvantage that it is not real valued for arbitrary complex matrices ρ, σ.

Whereas the learning for the entangled pure state takes somewhat longer than for the randomly chosen state, it is clear that after a reasonable learning time the neuronal dynamics has adapted to represent the quantum density matrix with acceptable fidelity. Combined with the learning of the unitary quantum gates above one concludes that this type of neural network can learn unitary transformations for two-qubit quantum systems. Details of an optimal learning algorithm for spiking neural networks remain to be worked out.

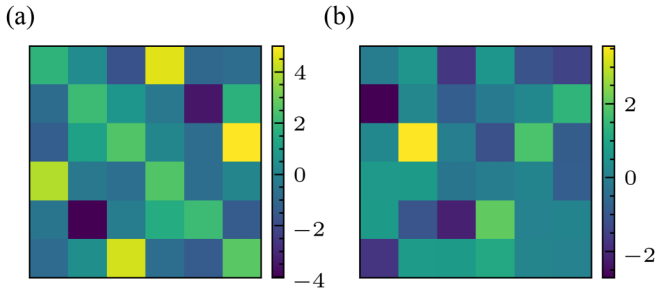

FIG. 6. Final recurrent weight matrices W_{rec} of the six recurrently connected neurons, whose refractory state corresponds to the spins $s_k^{(i)}$ of the pure state (a) ψ_+ and (b) ρ, respectively.

VI. GENERALIZATIONS TO MANY QUBITS

So far we focused on the case of two qubits. The correlation map can be extended to n qubits. Using generators,

$$L_{\mu_1\cdots\mu_n} = \bigotimes_{i=1}^{n} \tau_{\mu_i}, \quad \mu_i = 0\cdots 3, \tag{33}$$

we write the density matrix as

$$\rho = \frac{1}{2^n}\chi_{\mu_1\cdots\mu_n} L_{\mu_1\cdots\mu_n}. \tag{34}$$

For the minimal correlation map the coefficients $\chi_{\mu_1\cdots\mu_n}$ are determined by correlating $3n$ spins as follows: Write $s_\mu^{(i)}(t) = [1, s_k^{(i)}(t)]$ with $i = 1\cdots n$, $k = 1\cdots 3$, and $\mu = 0\cdots 3$. Define

$$\sigma_{\mu_1\cdots\mu_n}(t) = s_{\mu_1}^{(1)}(t)\cdots s_{\mu_n}^{(n)}(t), \tag{35}$$

and

$$\chi_{\mu_1\cdots\mu_n} = \frac{1}{T}\int_0^T \sigma_{\mu_1\cdots\mu_n}(t)dt. \tag{36}$$

This procedure involves up to n-point correlations. One may ask if the minimal correlation map remains a complete bit-quantum map for three qubits. We can employ the same methodology for $Q = 3$ as previously for $Q = 2$. Expectation values and correlations are evaluated by time averaging for nine Ising spins and corresponding selected nine neurons.

Time averaging of Ising spins representing some quantity above a threshold ($s_j = 1$) or below a threshold ($s_j = -1$) is comparatively economical in this respect. It is sufficient to measure for all Ising spins s_j the times above or below threshold. With $t_j^- = T - t_j^+$ one can use identities of the type,

$$\langle s_j\rangle = \frac{t_j^+ - t_j^-}{T} = \frac{2t_j^+}{T} - 1, \tag{37}$$

$$\langle s_j s_i\rangle = \frac{2(t_{ji}^{++} + t_{ji}^{--})}{T} - 1, \tag{38}$$

$$\langle s_k s_j s_i\rangle = \frac{2(t_{jik}^{+++} + t_{jik}^{+--} + t_{jik}^{-+-} + t_{jik}^{--+})}{T} - 1, \tag{39}$$

where t_{ji}^{++} is the time when both s_j and s_i are above threshold and so on.

We find that for three qubits an obstruction prevents the minimal correlation map to be complete. There are valid quantum density matrices for which no classical correlation functions can be realized that obey Eqs. (34), (35), and (36). For this purpose we concentrate on the gigahertz (GHZ) states,

$$\psi = \frac{1}{\sqrt{2}}(|+++\rangle + \varepsilon|---\rangle), \quad |\varepsilon| = 1. \tag{40}$$

The only nonvanishing elements of the corresponding pure state density matrix ρ_{GHZ} are

$$\rho_{+++,+++} = \rho_{---,---} = \frac{1}{2},$$
$$\rho_{+++,---} = \frac{\varepsilon^*}{2}, \quad \rho_{---,+++} = \frac{\varepsilon}{2}. \tag{41}$$

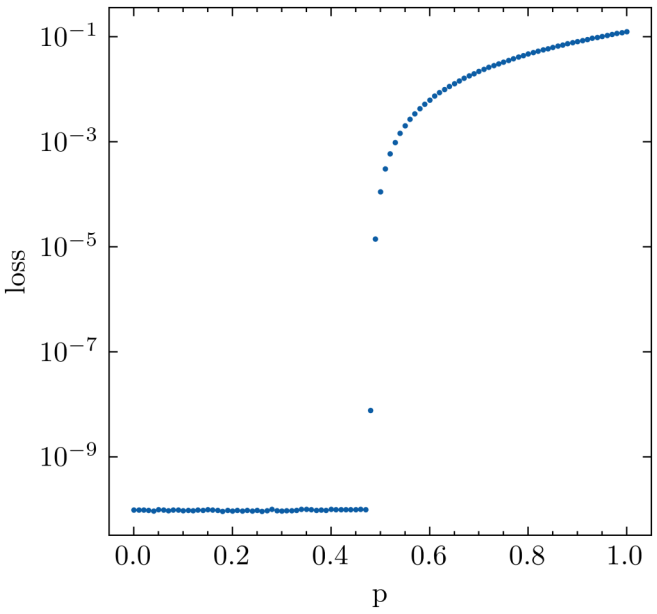

FIG. 7. Final loss after training up to 10^4 epochs to approximate $\rho(p) = p\rho_{\mathrm{GHZ}} + (1-p)\bar{\rho}$, where ρ is a randomly chosen density matrix and ρ_{GHZ} is the density matrix of the GHZ state in dimension 3.

The elements $\sigma_{\mu_1\mu_2\mu_3}$ are given by

$$\begin{aligned}\rho &= \tfrac{1}{2}\sigma_{\mu_1\mu_2\mu_3}\tau_{\mu_1}\otimes\tau_{\mu_2}\otimes\tau_{\mu_3}\\ &= \tfrac{1}{16}\{(1+\tau_3)\otimes(1+\tau_3)\otimes(1+\tau_3)\\ &\quad +(1-\tau_3)\otimes(1-\tau_3)\otimes(1-\tau_3)\\ &\quad +\varepsilon(\tau_1 - i\tau_2)\otimes(\tau_1 - i\tau_2)\otimes(\tau_1 - i\tau_2)\\ &\quad +\varepsilon^*(\tau_1 + i\tau_2)\otimes(\tau_1 + i\tau_2)\otimes(\tau_1 + i\tau_2)\},\end{aligned} \tag{42}$$

which results in nonzero values for

$$\sigma_{000} = \sigma_{330} = \sigma_{303} = \sigma_{033} = 1, \tag{43}$$

$$\sigma_{111} = -\sigma_{221} = -\sigma_{122} = -\sigma_{212} = \frac{1}{2}(\varepsilon + \varepsilon^*), \tag{44}$$

$$\sigma_{112} = \sigma_{211} = \sigma_{121} = -\sigma_{222} = -\frac{i}{2}(\varepsilon - \varepsilon^*). \tag{45}$$

For randomly chosen density matrices one finds for most cases suitable probability distributions that realize them by Eqs. (34), (35), and (36). If we look instead at a special class of density matrices,

$$\rho = p\rho_{\mathrm{GHZ}} + (1-p)\bar{\rho}, \tag{46}$$

where $\bar{\rho}$ is chosen randomly and $0 \leqslant p \leqslant 1$, for large enough p no such probability distribution can be found anymore. This is demonstrated in Fig. 7, where we plot the loss after training for up to 10^4 epochs against p.

This is reflected in the fidelity and relative entropy for a comparison of the finally optimized matrix and the true GHZ-density matrix, shown in Fig. 8. Again we observe a clear insufficiency of the minimal correlation map for a reproduction of the density matrices close to the GHZ state. This insufficiency sets in rather sharply at a certain value of p.

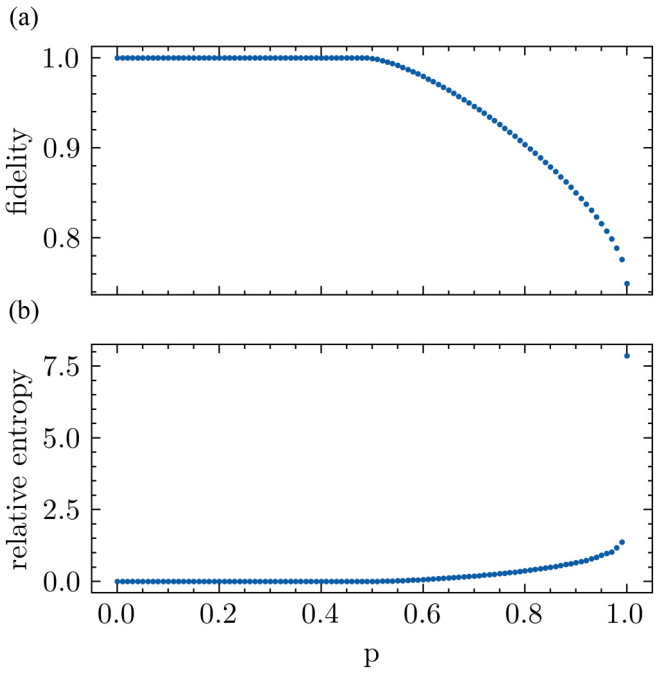

FIG. 8. (a) Final fidelity and (b) relative entropy after training up to 10^4 epochs to approximate $\rho(p) = p\rho_{\mathrm{GHZ}} + (1-p)\bar{\rho}$, where ρ is a randomly chosen density matrix and ρ_{GHZ} is the density matrix of the GHZ state in dimension 3.

One can understand the obstruction analytically. For this reason we consider the GHZ-density matrix with $\varepsilon = 1$. The expectation values (43), (44), and (45) can be realized by a factorizing probability distribution,

$$p = p_1\big[s_3^{(i)}\big]\, p_2\big[s_1^{(i)}, s_2^{(i)}\big]. \tag{47}$$

The first factor p_1 has to realize the correlations,

$$\begin{aligned} \big\langle s_3^{(1)} s_3^{(2)}\big\rangle = \big\langle s_3^{(1)} s_3^{(3)}\big\rangle = \big\langle s_3^{(2)} s_3^{(3)}\big\rangle = 1, \\ \big\langle s_3^{(i)}\big\rangle = 0, \quad \big\langle s_3^{(1)} s_3^{(2)} s_3^{(3)}\big\rangle = 0. \end{aligned} \tag{48}$$

This is achieved by the probabilities,

$$p_{+++} = p_{---} = \tfrac{1}{2}. \tag{49}$$

The second factor p_2 depends on the Ising spins $s_1^{(i)}$ and $s_2^{(i)}$ and has to realize the correlations,

$$\big\langle s_1^{(1)} s_2^{(2)} s_2^{(3)}\big\rangle = \big\langle s_2^{(1)} s_1^{(2)} s_2^{(3)}\big\rangle = \big\langle s_2^{(1)} s_2^{(2)} s_1^{(3)}\big\rangle = -1, \tag{50}$$

and

$$\big\langle s_1^{(1)} s_1^{(2)} s_1^{(3)}\big\rangle = 1. \tag{51}$$

This is not possible for a classical statistical setting. Equation (51) requires that for the three spins $(s_1^{(1)} s_1^{(2)} s_1^{(3)})$ only the configurations $(+++)$, $(+--)$, $(-+-)$, $(--+)$ can have a nonzero probability. For the case of the configuration $(+++)$ for $s_1^{(i)}$ Eq. (50) requires that the only configurations for $s_2^{(i)}$ that can have nonvanishing probabilities must obey $s_2^{(2)} s_2^{(3)} = -1$, $s_2^{(1)} s_2^{(2)} = -1$, $s_2^{(1)} s_2^{(3)} = -1$. Otherwise one of the three three-point functions (50) would be larger than -1. If $s_2^{(1)}$ and $s_2^{(2)}$ have opposite signs, and $s_2^{(1)}$ and $s_2^{(3)}$ have opposite signs, one infers that $s_2^{(2)}$ and $s_2^{(3)}$ have the same sign in contradiction to $s_2^{(2)} s_2^{(3)} = -1$. One concludes that the probability for $(s_1^{(1)} s_1^{(2)} s_1^{(3)}) = (+++)$ must be zero. Similar chains of arguments show that the three other configurations for $s_1^{(i)}$, namely, $(+--)$, $(-+-)$, $(--+)$, cannot have a nonzero probability either. In consequence, Eq. (51) cannot be obeyed. There is no probability distribution p_2 that can generate the set of correlations (51) and (50). This argument generalizes to probability distributions that are not of the direct product form (47). We conclude that for three qubits the minimal correlation map is not complete.

One may envisage an extended correlation map with additonal 27 spins $(s_{k_1k_2}^{(12)}, s_{k_1k_3}^{(13)}, s_{k_2k_3}^{(23)})$, $k_i = 1\cdots 3$. The spins $\sigma_{\mu_1\mu_2\mu_3}$ with precisely one index zero are given by

$$\sigma_{k_1k_20} = s_{k_1k_2}^{(12)}, \quad \sigma_{k_10k_2} = s_{k_1k_3}^{(13)}, \quad \sigma_{0k_2k_3} = s_{k_2k_3}^{(23)}, \tag{52}$$

instead of Eq. (35) for the minimal correlation map. Equation (36) continues to hold for all $\chi_{\mu_1\mu_2\mu_3}$ with one, two, or three indices equal to zero. The coefficients of the density matrix with only nonzero indices are correlations of "pair spins" $s_{kl}^{(ij)}$ and "single spins" $s_k^{(i)}$,

$$\chi_{k_1k_2k_3} = \big\langle s_{k_1k_2}^{(12)} s_{k_3}^{(3)}\big\rangle = \big\langle s_{k_1k_3}^{(13)} s_{k_2}^{(2)}\big\rangle = \big\langle s_{k_2k_3}^{(23)} s_{k_1}^{(1)}\big\rangle. \tag{53}$$

Equation (53) requires new quantum constraints on the classical probability distribution since all three correlations have to be the same. In the presence of this constraint knowledge of one set of correlations, say $\langle s_{k_1k_2}^{(12)} s_{k_3}^{(3)}\rangle$, gives access to $\langle s_{k_1k_3}^{(13)} s_{k_2}^{(2)}\rangle$ and $\langle s_{k_2k_3}^{(23)} s_{k_1}^{(1)}\rangle$. This feature is typical for quantum systems.

The GHZ state with $\varepsilon = 1$ can be realized by the extended correlation map. For an explicit construction of a classical probability distribution realizing this state we choose probabilities which only differ from zero if

$$s_{11}^{(23)} = -s_{22}^{(23)} = s_1^{(1)}, \quad s_{12}^{(23)} = s_{21}^{(23)} = -s_2^{(1)}, \tag{54}$$

$$s_{11}^{(13)} = -s_{22}^{(13)} = s_1^{(2)}, \quad s_{12}^{(13)} = s_{21}^{(13)} = -s_2^{(2)}, \tag{55}$$

$$s_{11}^{(12)} = -s_{22}^{(12)} = s_1^{(3)}, \quad s_{12}^{(12)} = s_{21}^{(12)} = -s_2^{(3)}. \tag{56}$$

This guarantees the relations (44),

$$\sigma_{111} = -\sigma_{221} = -\sigma_{122} = -\sigma_{212} = 1, \tag{57}$$

together with the constraint (53) for the four quantities in (57). The correlations in Eq. (43) are obeyed if nonzero probabilities occur only for

$$s_{33}^{(12)} = s_{33}^{(13)} = s_{33}^{(23)} = 1. \tag{58}$$

With Eqs. (57) and (58) the configurations with nonzero probabilities can be characterized by the values of the 21 spins $\bar{s}_a = (s_k^{(i)}, s_{13}^{(ij)}, s_{23}^{(ij)}, s_{31}^{(ij)}, s_{32}^{(ij)})$. With the conditions,

$$\langle \bar{s}_a\rangle = 0, \quad \langle \bar{s}_a \bar{s}_b\rangle = 0, \tag{59}$$

all other coefficients except σ_{111}, σ_{221}, σ_{122}, σ_{212}, σ_{330}, σ_{303}, σ_{033}, and σ_{000} vanish. The relation (59) can be realized by equipartition for the configurations of spins $\bar{s}_a$. This probability distribution realizes the GHZ state. So far we did not find other three-qubit density matrices that cannot be realized by the extended correlation map, but we do not have sufficient

material for a judgment if the latter should be considered as complete.

One can generalize the extended correlation map to a higher number of qubits $Q > 3$. One chooses independent spins for all $\sigma_{\mu_1 \cdots \mu_Q}$'s with one or two-"space"-indices k, l taking values of $k, l = 1 \cdots 3$ and all other μ_i's equal zero. These are $\frac{9}{2}Q^2 - \frac{3}{2}Q$ Ising spins, such that the number of classical spins grows quadratically with the number of qubits. Coefficients $\chi_{\mu_1 \cdots \mu_Q}$ with one or two space indices are given by the expectation values of the corresponding $\sigma_{\mu_1 \cdots \mu_Q}$. Coefficients with three-space indices are correlations of one pair spin and one simple spin with quantum constraint (53) extended correspondingly. For four-space indices of $\chi_{\mu_1 \cdots \mu_Q}$ one takes correlations of two pair spins with corresponding quantum constraints. Five spin indices are realized by three-point functions of two pair spins and one single spin, and so forth. There is never more than one single spin in the correlations. Again it is not known if this extended correlation map is a complete bit-quantum map, or if new obstructions arise for a certain number of qubits.

VII. INCOMPLETE PROBABILISTIC INFORMATION

The number 2^{2Q} of elements of the density matrix grows very rapidly with an increasing number Q of qubits. This issue is common to all approaches which use at every step of the computation the full information about the quantum state. We find it unlikely that computations by real quantum systems or artificial neurons need the full information contained in the density matrix ρ. It becomes then an important task to find out which part of the probabilistic information is involved for practical questions.

As an example we consider the search for the ground-state energy of the one-dimensional quantum Ising model with Hamiltonian,

$$H = -J \sum_{i=1,\ldots,n} \tau_3^{(i)} \tau_3^{(i+1)} - h \sum_i \tau_1^{(i)}, \tag{60}$$

one of the benchmark tasks in Ref. [2]. The goal is to find a density matrix ρ such that $\mathrm{tr}(H\rho)$ is minimized. This requires only a small subset of the $\chi_{\mu_1 \cdots \mu_n}$, namely, those n two-point functions for which $\mu_i = \mu_{i+1} = 3$ and all other $\mu_k = 0$, as well as n expectation values for which $\mu_i = 1$ and all other $\mu_k = 0$'s. For the minimal correlation map the first set of χ is given by two point correlation functions. These two-point correlation functions cannot take arbitrary values, however. The positivity of the density matrix imposes quantum constraints [1,20] that these correlation functions have to obey. Due to these constraints the minimal value of $\langle H \rangle$ is higher than for unconstrained correlation functions. For the extended correlation map also the first set of χ is given by expectation values. Again, quantum constraints restrict the possible values.

The explicit use of all quantum constraints seems a difficult task for a high number of qubits. It may, however, be sufficient to impose only part of the quantum constraints for obtaining already a reasonable approximation to the minimal value of $\langle H \rangle$. These partial constraints could then be associated with the information that is relevant for a given problem, whereas additional information concerning the full set of quantum constraints may be discarded. A probabilistic view on expectation values and correlations may help to focus on the relevant information needed for a given quantum problem.

VIII. CONCLUSIONS

Neuromorphic computers can be trained to learn and transform the correlations necessary for certain computational tasks. We have explicitly demonstrated this for the quantum correlations of a two-qubit quantum system. Already for this very simple system the learned correlations are of a rather global type, connecting typically all six classical Ising spins, which represent the quantum system. In view of the large amount of probabilistic information that can be stored in correlations it seems likely to us that correlations involving many neurons at once could be used for an efficient performance of tasks in neuromorphic computing. The important question for future investigations amounts to finding out which is the precise form of such correlations.

In our model neuromorphic computing is probabilistic computing. The dynamical equations realize expectation values and correlations for neurons. Whereas every neuron has only two possible states at any given time, the sharp states with given fixed values of all neurons play no role in the computation. Only time-averaged probabilistic expectation values and correlations matter. This probabilistic computing is of a particular type, however. It is not of the type where updating steps between sharp neuron or bit configurations are performed with certain probabilities. Such probabilistic updatings are Markov chains for which the initial information gets diluted or lost after a certain number of updating steps. Since our sample system can perform unitary quantum operations no information in the subsystem described by the density matrix is lost. This demonstrates that "classical probabilistic" computing is possible without loss of the relevant initial information.

Other approaches to sampling with spiking neurons [14] make use of an approximate correspondence between Poisson-stimulated recurrently connected leaky integrate and fire neurons in the "high-conductance" state and restricted Boltzmann machines. This allows, in particular, to translate between the weight parameters of the restricted Boltzmann machine and the weight parameters of the network of spiking neurons. Here in contrast we train the spiking neural network directly using backpropagation through time [21] without knowing or assuming a certain equilibrium distribution. In particular, the synaptic connections of the spiking network are initialized nonsymmetrically and any structure in the optimized weight matrix arises from the training procedure. Independent of the application reported here, this is a demonstration of this approach to stochastic computation with spiking neurons. We give a brief explanation of the method in a purely classical context and a simple application in Appendix B.

For the realization of tasks of quantum computing by neuromorphic computers it seems hard to perform such operations precisely for systems of many qubits. The total information in the density matrix is simply too high to be stored, even worse to be learned. What seems more promising is the learning of the relevant correlations needed to perform a given "quantum task" with an acceptable approximation.

Finding out the relevant probabilistic information for a given task seems to us a crucial question. It is conceivable that for a given task for which quantum computers are very promising it is actually not necessary that the correlations obey the particular quantum constraints. Other approaches employing effectively nonquantum, but still "global" correlations could be efficient as well. From the point of view of correlated probabilistic computing the fields of quantum computing, artificial intelligence, neuromorphic computing, or brain-inspired computing may appear as different facets of a more general common setting based on correlations.

In our approach the computation of the expectation values $\chi_{\mu_1\cdots\mu_n}$, and the associated quantum density matrix involves purely classical probabilistic settings as neurons firing in biological or artificial systems. No low temperature or a high degree of isolation of small subsystems is needed for such a realization of quantum features. Our findings are an example how quantum evolution can be realized by probabilistic classical systems [20,22,23]. This concerns the full evolution of a quantum subsystem with observables and the quantum law for expectation values realized in the same way as for usual quantum systems. This goes far beyond the solution of a Schrödinger equation, or its adaptation to discrete time steps, by a classical computer. The latter can realize the computation of expectation values as real numbers, but there is no possible measurement process with discrete values of observables and a probabilistic distribution of outcomes associated such a classical computer. In our proposal any measurement of observables yields a discrete value—a neuron is active or not. Expectation values and correlations of suitable observables can be obtained by repeated measurements as in real quantum systems.

A physical quantum system involves two ingredients. The first are observables, represented by operators whose eigenvalues (spectrum) correspond to the possible measurement values of the observable. Quantum mechanics makes probabilistic statements about expectation values and correlations for the observables. The second ingredient is the time evolution as described by the Schrödinger or von Neumann equation or its counterpart for discrete time steps. A numerical solution of the Schrödinger equation or the corresponding sequence of unitary transformations can realize the second ingredient effectively up to a number of ca. 50 qubits. The first ingredient of observables that can be measured is absent, however. Our proposal for quantum operations by a neuromorphic computer comes much closer to a physical quantum system with possible measurements of observables.

For our proposal the neurons are classical two-level observables that are represented in the quantum subsystem by operators whose discrete spectrum reflects the possible measurement values of the observables. All expectation values and correlations for those classical observables that can be computed from the density matrix have precisely the same time evolution as for the associated quantum system. For a real neuromorphic computer operating according to our prescription the values for the classical observables can be measured at a given time. Expectation values and correlations can be extracted by many identical measurements, in complete analogy to quantum systems. The subset of classical bits employed for the density matrix behaves as a physical quantumlike system.

Not all quantum observables are directly realized in this way, however. The quantum spin observables in directions different from the Cartesian directions have no direct analog as states of neurons that can be directly measured. The number of quantum observables that have a classical correspondence can be increased by using additional neurons for the implementation of the density matrix for the quantum subsystem together with additional quantum constraints [1]. With a finite number of neurons the number of quantum observables that have a direct classical correspondence remains finite as well. In this sense a neuromorphic computer can be partially a physical quantum system without realizing a complete physical quantum system for all the possible quantum observables. This is another facet of the difference between more general correlated probabilistic computing and precise quantum computing.

The implementation uses the publicly available open source software PYTORCH [24], JAX [19] and NORSE [25]. Code and data to reproduce the experiments will be made available here: [26] .

ACKNOWLEDGMENTS

This work was supported by the EU Horizon 2020 Framework Program under Grant Agreements No. 720270, No. 785907, and No. 945539 (HBP), the Heidelberg Graduate School for Fundamental Physics (HGSFP), the DFG Collaborative Research Centre SFB 1225 (ISOQUANT), and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy Grant No. EXC 2181/1-390900948 (the Heidelberg STRUCTURES Excellence Cluster).

APPENDIX A: SURROGATE GRADIENT LEARNING

In this Appendix we give more details on the method to achieve the results in Sec. V. The overall strategy is to implement the differential equations with jumps in some simple numerical integration scheme and then use backpropagation through time together with surrogate gradients to implement end-to-end learning of the given objective function.

More concretely, we use simple forward Euler time-discretization to implement the differential equations (20) and (21),

$$v_{n+1} = v_n + \frac{dt}{\tau_{\text{mem}}}(v_{\text{l}} - v_n + RI_n), \tag{A1}$$

$$I_{n+1} = I_n - \frac{dt}{\tau_{\text{syn}}} I_n, \tag{A2}$$

since we choose the absolute refractory time to be identical with the integration time-step dt, it does not have to be tracked explicitly.

The jump condition is computed as

$$z_{n+1} = \Theta(v_{n+1} - v_{\text{th}}), \tag{A3}$$

and jumps are computed as

$$v_{n+1} = (1 - z_{n+1})v_{n+1} + z_{n+1}v_{\text{reset}}, \tag{A4}$$

$$I_{n+1} = I_{n+1} + W_{\text{rec}}z_{n+1} + W_{\text{in}}z^{\text{in}}_{n+1}. \tag{A5}$$

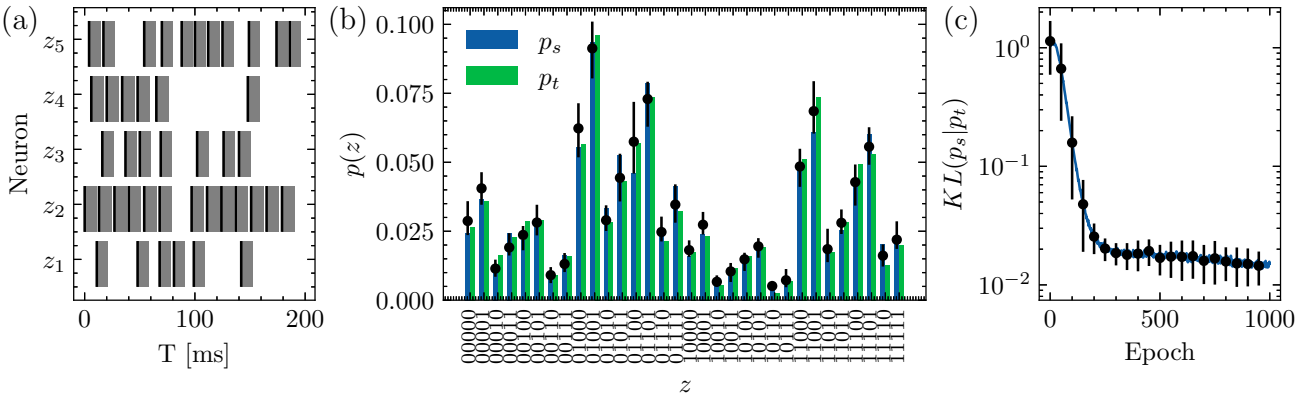

FIG. 9. Approximation of a categorical probability distribution by sampling from a recurrently connected spiking neural network of leaky-integrate and fire neurons. The refractory state is treated as a binary random variable z_i for each of the neurons. In (a) we indicate the state of these binary random variables for one sampling run in a 200-ms interval with the spike times marked by vertical black lines and the resulting absolute refractory period marked by gray rectangles. The resulting sampled probability distribution p_s (blue) is compared to a given target probability distribution p_t (green) in (b), error bars indicate the minimal and maximal values over 16 sampling runs. The average loss and standard deviation over five instances of this optimization problem is shown in (c). We use the ADAM optimizer with default parameters and batch size $B = 16$. The final loss after 10^3 optimization steps is similar to the one report in Ref. [14] for $T = 10^4$ sampling time steps.

Overall this defines a function,

$$[z_{n+1}, (v_{n+1}, I_{n+1})] = f[(v_n, I_n), z^{\text{in}}_{n+1}], \tag{A6}$$

which is not differentiable due to the occurrence of the Heaviside function. As already mentioned in the main article, this can be resolved by using a scaled derivative of a (piecewise) smooth approximation of the Heaviside function to compute the derivative of f instead [16,18]. Here we use the approximation used in Ref. [15], whose derivative is

$$\Theta'_\alpha(x) = \frac{1}{(\alpha|x| + 1)^2}. \tag{A7}$$

It is then possible to regard the repeated application of the function f as one differentiable function for which parameter gradients can be computed using backpropagation through time [21]. Concretely the repeated application is implemented in JAX using the "scan" primitive.

Using an analogous approach it is also possible to treat the case that the refractory time is a multiple of the time-step dt, whereas this was not use to obtain the results in the Sec. V, the following section demonstrates this extension.

APPENDIX B: STOCHASTIC COMPUTATION WITH SPIKING NEURONS

Whereas in Sec. V we used surrogate-gradient based training of a recurrently connected population of neurons subject to Poissonian input to perform end-to-end optimization approximating a given quantum density matrix, the method itself can also be applied in a purely classical context. In this Appendix we show that we can approximate small categorical probability distributions using a similar end-to-end learning approach. "Neural sampling" was previously demonstrated in either abstract models [13] or in LIF neuron models in the high-conductance state utilizing an approximate connection to restricted Boltzmann machines [14]. Here we instead train a network leaky-integrate and fire neurons subject to Poissonian input noise without reference to some equilibrium model. Approximating categorical probability distributions in this way could be used to obtain the results in Secs. III and IV. Moreover, it also connects our approach to other ways of approximating quantum density matrices using spiking neural networks.

To illustrate this, consider a concrete example: A recurrently connected population of $N = 5$ leaky-integrate and fire neurons with $v_{\text{leak}} = 0.9 v_{\text{th}}$ and an absolute refractory period of $\tau_{\text{refrac}} = 10\,\text{ms}$ receive Poissonian input with frequency 300 Hz from $K = 256$ input sources. Just like in Sec. V, the refractory state can be considered as a binary random variable as defined in Eq. (28). Here the refractory time is a multiple of the integration step so that the refractory state variable has to be treated explicitly. We use the same surrogate gradient function as before but use a publicly available implementation of LIF neurons with refractory state [25].

Based on integration over some sampling time T, we can then compute the sample probability of a given binary state vector, for example,

$$p_s(1, 0, 0, 1, 0) = \frac{1}{T}\int_0^T z_1 \bar{z}_2 \bar{z}_3 z_4 \bar{z}_5 dt, \tag{B1}$$

where $\bar{z}_i = (1 - z_i)$. This is in contrast to Sec. V where we instead interpreted the refractory state as a spin variable using the transformation $s = 2z - 1$.

We use the Kullback-Leibler divergence between the sampled probability distribution p_s and a fixed target distribution p_t as a loss function for gradient-based optimization. In the particular case of categorical probability distributions it is simply given by

$$\text{KL}(p_s|p_t) = -\sum_z p_s(z; w) \log\, p_t(z) + \sum_z p_s(z; w) \log\, p_s(z; w). \tag{B2}$$

Overall this can then be used to perform end-to-end gradient-based optimization of the synaptic weight

parameters w, where the derivative of the Heaviside function is again replaced by the derivative of an approximation as in Eq. (A7).

In Fig. 9, we show one example of such an optimization problem as well as optimization results sampled over five random initializations. The target distribution in all cases was generated from sampling from a LIF network with randomly initialized weights subject to Poissonian noise of the same input frequency 300 Hz.

Clearly this method can be generalized to other neuron models for which no correspondence to an equilibrium model, such as the one considered in Ref. [14] is known. Moreover, it also directly allows to use recent advances in surrogate gradient-based training on neuromorphic hardware [27].

[1] C. Wetterich, Quantum computing with classical bits, Nucl. Phys. B **948**, 114776 (2019).

[2] G. Carleoand M. Troyer, Solving the quantum many-body problem with artificial neural networks, Science **355**, 602 (2017).

[3] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, Neural-network quantum state tomography, Nat. Phys. **14**, 447 (2018).

[4] O. Sharir, Y. Levine, N. Wies, G. Carleo, and A. Shashua, Deep Autoregressive Models for the Efficient Variational Simulation of Many-Body Quantum Systems, Phys. Rev. Lett. **124**, 020503 (2020).

[5] M. Broughton, G. Verdon, T. McCourt, A. J. Martinez, J. H. Yoo, S. V. Isakov, P. Massey, M. Y. Niu, R. Halavati, E. Peters, M. Leib, A. Skolik, M. Streif, D. V. Dollen, J. R. McClean, S. Boixo, D. Bacon, A. K. Ho, H. Neven, and M. Mohseni, Tensorflow quantum: A software framework for quantum machine learning, arXiv:2003.02989.

[6] C. Pehle, K. Meier, M. Oberthaler, and C. Wetterich, Emulating quantum computation with artificial neural networks, arXiv:1810.10335.

[7] J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, Proc. Natl. Acad. Sci. USA **79**, 2554 (1982).

[8] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, Measurement of qubits, Phys. Rev. A **64**, 052312 (2001).

[9] M. Parisand J. Rehacek, *Quantum State Estimation* (Springer, Berlin, 2004), Vol. 649.

[10] N. Kiesel, Experiments on multiphoton entanglement, Ph.D. thesis, Ludwig Maximilian University of Munich, 2007.

[11] E. Schneidman II, M. J. Berry, R. Segev, and W. Bialek, Weak pairwise correlations imply strongly correlated network states in a neural population, Nature (London) **440**, 1007 (2006).

[12] I. E. Ohiorhenuan, F. Mechler, K. P. Purpura, A. M. Schmid, Q. Hu, and J. D. Victor, Sparse coding and high-order correlations in fine-scale cortical networks, Nature (London) **466**, 617 (2010).

[13] L. Buesing, J. Bill, B. Nessler, and W. Maass, Neural dynamics as sampling: A model for stochastic computation in recurrent networks of spiking neurons, PLoS Comput. Biol. **7**, 1 (2011).

[14] M. A. Petrovici, J. Bill, I. Bytschok, J. Schemmel, and K. Meier, Stochastic inference with spiking neurons in the high-conductance state, Phys. Rev. E **94**, 042312 (2016).

[15] F. Zenkeand S. Ganguli, Superspike: Supervised learning in multilayer spiking neural networks, Neural Comput. **30**, 1514 (2018).

[16] S. K. Esser, P. A. Merolla, J. V. Arthur, A. S. Cassidy, R. Appuswamy, A. Andreopoulos, D. J. Berg, J. L. McKinstry, T. Melano, D. R. Barch, C. di Nolfo, P. Datta, A. Amir, B. Taba, M. D. Flickner, and D. S. Modha, Convolutional networks for fast, energy-efficient neuromorphic computing, Proc. Natl. Acad. Sci. USA **113**, 11441 (2016).

[17] G. Bellec, D. Salaj, A. Subramoney, R. Legenstein, and W. Maass, Long short-term memory and learning-to-learn in networks of spiking neurons, in *Advances in Neural Information Processing Systems* (2018), pp. 787–797.

[18] E. O. Neftci, H. Mostafa, and F. Zenke, Surrogate gradient learning in spiking neural networks: Bringing the power of gradient-based optimization to spiking neural networks, IEEE Signal Processing Magazine **36**, 51 (2019).

[19] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, and S. Wanderman-Milne, JAX: composable transformations of Python+NumPy programs (2018).

[20] C. Wetterich, The probabilistic world, arXiv:2011.02867.

[21] P. Werbos, Backpropagation through time: what it does and how to do it, Proc. IEEE **78**, 1550 (1990).

[22] C. Wetterich, Information transport in classical statistical systems, Nucl. Phys. B **927**, 35 (2018).

[23] C. Wetterich, Quantum formalism for classical statistics, Ann. Phys. (NY) **393**, 1 (2018).

[24] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai *et al.*, Pytorch: An imperative style, high-performance deep learning library, in *Advances in Neural Information Processing Systems 32*, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Curran, Red Hook, NY, 2019), pp. 8024–8035.

[25] C. Pehle and J. E. Pedersen, Norse - A deep learning library for spiking neural networks (2021), https://github.com/norse/norse.

[26] https://github.com/cpehle/neuromorphic-quantum-computing.

[27] B. Cramer, S. Billaudelle, S. Kanya, A. Leibfried, A. Grübl, V. Karasenko, C. Pehle, K. Schreiber, Y. Stradmann, J. Weis *et al.*, Surrogate gradients for analog neuromorphic computing, Proc. Natl. Acad. Sci. USA **119**, e2109194119 (2022).