The Discrete Langevin Machine: Bridging the Gap Between Thermodynamic and Neuromorphic Systems

Lukas Kades\textsuperscript{1} and Jan M. Pawlowski\textsuperscript{1}
\textsuperscript{1}Institut für Theoretische Physik, Universität Heidelberg, Philosophenweg 16, 69120 Heidelberg, Germany

A formulation of Langevin dynamics for discrete systems is derived as a new class of generic stochastic processes. The dynamics simplify for a two-state system and suggest a novel network architecture which is implemented by the Langevin machine. The Langevin machine represents a promising approach to compute successfully quantitative exact results of Boltzmann distributed systems by LIF neurons. Besides a detailed introduction of the new dynamics, different simplified models of a neuromorphic hardware system are studied with respect to a control of emerging sources of errors.

Keywords: Langevin dynamics, Discrete systems, Boltzmann machine, Neuromorphic systems, LIF neurons, BrainScaleS

I. INTRODUCTION

The rapidly increasing progress on neuromorphic computing and the ongoing research of spiking systems such as the third generation of neural networks calls for a better understanding of the fundamental processes of neuromorphic hardware systems [1–6], for a recent review on neuromorphic computing see [7]. As a parallel computing platform, these systems may be used in the long run to accurately simulate and compute large systems, in particular given their low energy consumption.

Possible applications range from an effective implementation of artificial neural networks and further machine learning methods [8–13], a better understanding of biological processes in our brains [14, 15] to the computation of physical and stochastic interesting systems [16–18]. Many artificial neural networks and physical systems are described by Boltzmann distributed systems. For a quantitatively accurate computation of such systems, it is necessary to deduce an exact representation on neuromorphic hardware systems [11, 19–21], in particular for systematic error estimates.

Our work is motivated by the similarity of the Langevin dynamics and leaky integrate-and-fire (LIF) neurons for performing stochastic inference [22]. Indeed, the fundamental dynamics of LIF neurons is governed by Langevin dynamics. Apart from its obvious relevance for the description of stochastical processes, the Langevin equation [23] can also be used for simulating quantum field theories with stochastic quantization [24–26]. In this approach the Euclidean path integral measure is obtained as the stationary distribution of a stochastic process. This paves the way to the heuristic approach of using complex Langevin dynamics as a potential method for accessing real time dynamics and sign problems. The latter problem is e.g. prominent in QCD at finite chemical potential [27–29]. A further interesting application of the Langevin equation can be found in [30]. There, Langevin dynamics is combined with a stochastic gradient descent algorithm to perform Bayesian learning which enables an uncertainty estimation of resulting parameters.

Many of the above mentioned systems are discrete ones, the simplest one being a two-state system. This suggests the formulation of a discrete analogue of the continuous Langevin dynamics for the accurate description of discrete systems. In the present work we show that a formulation of Langevin dynamics for discrete systems leads to a new class of a generic stochastic process, namely the Langevin equation for discrete systems,

\[ \phi' = \phi + (\nu - \phi) \Theta \left( -1 - \frac{\epsilon}{2\lambda} \Delta S(\nu, \phi) + \sqrt{\epsilon} \eta \right) \]

where \( \phi \) is the current state, \( \nu \) is a proposal state and \( \phi' \) is the updated state. A more detailed derivation of (1) including a discussion of its properties can be found in Section II.

The present work concentrates on the potential of the novel process for a more accurate implementation of Boltzmann distributed systems on the neuromorphic hardware. This leads to a new architecture of neurons based on a self-interacting contribution. The self-interacting term changes manifestly the dynamics of the neural network. This results in an activation function which is much closer to a logistic distribution, the activation function of a Boltzmann machine, than existing approaches. The new architecture can be applied to both, discrete two state systems and neuromorphic hardware systems with a continuous membrane potential and a spiking character. In this work spiking character refers to an effective mapping of a continuous potential to two discrete neuron states in an interacting system. The dynamics differ in their kind of noise that is uncorrelated in the former case and autocorrelated in the latter case. Figure 1 compares the different network structures and gives an overview over existing and contributed new dynamics of this work.

An exact representation of the activation function of the Boltzmann machine is necessary to obtain correct statistics in coupled systems on neuromorphic hardware systems. In the present work we show in a detailed numerical analysis that small deviations in the activation function propagate if a rectangular refractory mechanism or interactions between neurons are taken into account. These small deviations have a large impact on the re-
sulting correlation functions and observables. The numerical results demonstrate that a reliable estimation, understanding and control of different sources of errors are essential for a correct computation of Boltzmann distributed systems in the future.

The paper is organised as follows. The Langevin equation for discrete systems is derived in Section II. In Section III, the so called sign-dependent discrete Langevin machine is introduced as a special case of the Langevin equation for discrete systems. The mappings of the different dynamics for discrete systems onto an Ornstein-Uhlenbeck process with spiking character is discussed in Section IV. Section V recapitulates relations between the discrete Langevin machine and the neuromorphic hardware system. In Section VI, numerical results of the introduced and of existing dynamics are presented, possible sources of errors are extracted and the propagation of errors for different abstractions of a neuromorphic hardware system is analysed. The conclusion and outlook can be found in Section VII.

II. DISCRETE LANGEVIN DYNAMICS

The Langevin equation for discrete systems is derived inspired by a comparison of the Metropolis algorithm and the Langevin dynamics (for a detailed comparison see Appendix A). The general formulation of a Langevin equation for discrete systems presented in this Section is similar to a Monte Carlo algorithm and is driven by a Gaussian noise contribution. The transition probability to a proposed new state is regulated by the introduction of truncating Gaussian noise. It is shown that the accuracy of the process strongly depends on an intrinsic parameter \( \epsilon \) and the scale of the energy contribution.

Certain necessary properties of a possible Langevin equation for discrete systems can be stated beforehand based on the comparison of the Langevin dynamics and the Metropolis algorithm in Appendix A. First of all, an infinitesimal change of the microscopic state/field is not possible. Therefore, one has to switch from a parallel to a random sequential update mechanism. Further, the computer time directly is the time scale of the approach. The proposal field has to be chosen from a discrete distribution. One may select the proposal field according to some distribution around the current field. However, since a parallelisation is not possible, the uniform selection probability of a Metropolis algorithm can be adopted.

Assuming the same acceptance probability as in the continuous case, a starting point is the following proportionality,

\[
W(\phi \rightarrow \phi') \propto \exp \left( -\frac{S(\phi') - S(\phi)}{2} \right). \tag{2}
\]

With the help of a relation between the cumulative Gaussian distribution and the exponential function, given by (B1), and with \( \Delta S(\phi', \phi) = S(\phi') - S(\phi) \), this can be rewritten in the following way,

\[
W(\phi \rightarrow \phi') \propto \Phi \left( -\frac{1}{\sqrt{\epsilon}} - \frac{\sqrt{\epsilon}}{2\lambda} \Delta S(\phi', \phi) \right) = P \left( \hat{\eta} < -\frac{1}{\sqrt{\epsilon}} - \frac{\sqrt{\epsilon}}{2\lambda} \Delta S(\phi', \phi) \right), \tag{3}
\]
for $\epsilon \to 0$. An analytical expression for the additional scaling factor $\lambda$ is given in equation (B2).

The Gaussian noise contribution $\tilde{\eta}$ is uncorrelated and has variance 1,

$$\langle \tilde{\eta}_i, \tilde{\eta}_j \rangle_{\tilde{\eta}} = \delta(j - i)\delta(t' - t), \quad \langle \tilde{\eta}_i \rangle_{\tilde{\eta}} = 0. \quad (4)$$

Taking the current state $\phi$ into account, one can transform the sampling from the cumulative normal distribution into a general stochastic update rule with Gaussian noise and a proposal state $\nu$. This leads us to (1), already presented in the introduction,

$$\phi' = \phi + (\nu - \phi)\Theta \left[ -1 - \frac{\epsilon}{2\lambda\epsilon} \Delta S(\nu, \phi) + \sqrt{\epsilon} \tilde{\eta} \right], \quad (5)$$

where $\epsilon$ needs to be chosen sufficiently small. $\Theta(x)$ represents the Heaviside function.

The update formalism corresponds to a single spin flip Monte Carlo algorithm with a random sequential update mechanism, driven by Gaussian noise. It can be immediately seen within the present form that a flip to a proposed field gets the more unlikely the smaller $\epsilon$. Adaptations of the Gaussian noise term to truncated Gaussian noise can help to improve the dynamics, i.e., to increase the probability of a spin flip. In principle, this corresponds to a rescaling of the transition probability term similar to a maximization of the spin flip probability in the Metropolis algorithm.

The truncated Gaussian noise term can be expressed by the following parametrization,

$$\tilde{\eta}^T \in \left[ \frac{1}{\sqrt{\epsilon}} + \alpha, \infty \right], \quad (6)$$

where $\alpha$ is in the range of

$$-\infty \leq \alpha \leq -\frac{\sqrt{\epsilon}}{2\lambda\epsilon} \Delta_{\text{max}}, \quad \text{with} \quad \Delta_{\text{max}} = |\Delta S(\nu, \phi)|. \quad (7)$$

The improved update rule is

$$\phi' = \phi + (\nu - \phi)\Theta \left[ -1 - \frac{\epsilon}{2\lambda\epsilon} \Delta S(\nu, \phi) + \sqrt{\epsilon} \tilde{\eta}^T \right]. \quad (8)$$

For $\alpha \to -\infty$ this reduces to the update formalism (5) and for $\alpha = -\frac{\sqrt{\epsilon}}{2\lambda\epsilon} \Delta_{\text{max}}$ one obtains spin flip probabilities up to 1. This can be seen under consideration of the explicit transition probability of the update rule (8),

$$W(\phi \to \nu) = \frac{\Phi \left( -\frac{1}{\sqrt{\epsilon}} - \frac{\sqrt{\epsilon}}{2\lambda\epsilon} \Delta S(\nu, \phi) \right)}{\Phi \left( -\frac{1}{\sqrt{\epsilon}} - \alpha \right)}. \quad (9)$$

Transition probabilities of further standard Monte Carlo algorithms can be emulated by other choices of $\alpha$. Note that for a uniform random number $r \in [0, 1]$ and a proposal field $\nu$, an equivalent formulation to (8) can be stated for the transition probability (2),

$$\phi' = \phi + (\nu - \phi)\Theta \left[ \exp \left( -\frac{S(\nu) - S(\phi) + \Delta_{\text{max}}}{2} - r \right) \right]. \quad (10)$$

Processes with a different value of $\alpha$, i.e., a different rescaling of the transition probability, can always be mapped onto each other by a respective rescaling of the time. Given a transition probability $W(\phi \to \mu)$ and a scaling factor $a$, the following relation holds,

$$W(\phi \to \mu) \to aW(\phi \to \mu) \iff t \to \frac{t}{a}. \quad (11)$$

Most of the existing single spin flip algorithms can be reformulated into a Langevin equation for discrete systems with the same derivation, as presented in this Section. However, it can be shown that for the particular choice of the transition probability according to equation (2), the resulting order of accuracy in the detailed balance equation is the best one, with $O(\epsilon \Delta S(\nu, \phi)^3)$.

The update formalism (8) represents a Langevin like equivalent for discrete systems to the Langevin dynamics of continuous systems. As for continuous systems, the dynamics depends on Gaussian noise and is based on a rather simple expression. The algorithms can also be applied to continuous systems due to the equivalence to standard Monte Carlo algorithms in the limit $\epsilon \to 0$.

III. SIGN-DEPENDENT DISCRETE LANGEVIN MACHINE

The Langevin equation for discrete systems (8) turns into a rather simple expression for a two-state system. The resulting dynamics is introduced in the following as sign-dependent discrete Langevin machine (LM$^2$). The LM$^2$ represents a new architecture for interacting neurons with the particularity of a self-interacting contribution. The derived network structure results in a new basic dynamics with different weights and biases compared to the Boltzmann machine. It has the unique property that the equilibrium distribution converges in the limit $\epsilon \to 0$, despite a different underlying dynamics, to a logistic distribution, the activation function of the Boltzmann machine.

We define the energy of the Boltzmann machine in the common way by

$$E = -\sum_{i<j} W_{ij} z_i z_j - \sum_i b_i z_i, \quad (12)$$

where $W_{ij}$ are symmetric weights between the neurons $i$ and $j$ and $b_i$ is some additional bias. The domain of definition of the states $z_i$ at each neuron is given by $z_i \in \{0, 1\}$.

For applying the generalised update rule (8) we need the following identifications: $S \rightarrow E$ and $\phi_i \rightarrow z_i$. As discussed in Appendix D, the following simplified update rule can be derived for the LM$^2$,

$$z'_i = \Theta \left[ W_{ii} z_i + \sum_j W_{ij} z_j + b'_i + \tilde{\eta}^T \right], \quad (13)$$
where the transformed parameters are defined as follows:

\[ W'_{ii} = \frac{2}{\sqrt{\epsilon}} \]  
\[ W'_{ij} = \frac{\sqrt{\pi}}{2\sqrt{\epsilon}} W_{ij} \]  
\[ b'_i = \left( \frac{\sqrt{\pi}}{2\sqrt{\epsilon}} b_i - \frac{1}{\sqrt{\epsilon}} \right) . \]

Figure 2 illustrates a comparison between the structure of the Boltzmann machine and the new update dynamics. The activation function of the LM\(^2\) is given in the limit of \(\epsilon \to 0\) by a logistic distribution,

\[ \lim_{\epsilon \to 0} P_{\text{LM}^2}(z_i = 1) = \frac{1}{1 + \exp \left( -\sum_j W_{ij} z_j - b_i \right)} . \]  

The term *Langevin machine* is chosen because of the similarity of the network to the Boltzmann machine and to Langevin dynamics. The adjective *discrete* is added to avoid confusion with the Langevin machine presented in [31]. The noise term in the dynamics can be chosen according to equation (6), i.e., it can be a Gaussian noise or a truncated Gaussian noise. The self-interaction term \( W'_{ii} \in \{0, 2/\sqrt{\epsilon}\} \) and the contribution \(-1/\sqrt{\epsilon}\) of the bias \(b'_i\) lead in dependency of the state of the neuron for small values of \(\epsilon\) to a strong shift of the mean value into a positive or negative direction. Respectively, the neuron stays very long in an active regime or in an inactive regime in the case of Gaussian noise. The process fluctuates between two different fundamental descriptions. The addition *sign-dependent* is used to emphasize this property and to point out that the so far presented dynamics is a particular realisation of the *discrete Langevin machine*, a larger class of network implementations with a Gaussian noise distribution. This is discussed in more detail in Section V. The exponent ‘2’ in the abbreviation signifies the fluctuation between the two regimes. The implicit dynamics (13) allows different interpretations and implementations.

Absorbing the Gaussian noise term into the bias, the resulting network has a rectangular decision function and can be interpreted as a neural network with a noisy bias. The simplicity of the update rule might be especially for neuromorphic systems very helpful for a computation of physical statistical systems, which are Boltzmann dis-tributed. The implementation of an exponential function in the system is much more challenging than generating Gaussian noise. The rectangular decision function further coincides with the threshold function of spiking neurons. Accordingly, a possible adaptation of the dynamics on neuromorphic systems is obtained by the introduction of an additional time scale and staggered Gaussian noise peaks.

In the present work we pursue an alternative approach which is discussed in the next Section. Instead of performing an implicit update, it is also possible to explicitly compute the probability for an activation of the neuron in the next step. This probability is given by

\[ W_{\text{LM}^2}(z_i \to 1) = \Phi \left( W'_{ii} z_i + \sum_j W'_{ij} z_j + b'_i \right) . \]  

In contrast to the Boltzmann machine, the transition probability is not the same probability as the activation function (14).

Finally, the LM\(^2\) exhibits a totally different dynamics than the Boltzmann machine. The dynamics is characterised by a Gaussian noise term as stochastic input, a self-interacting term, its simplicity and multiple possible implementations. Transition probabilities and correlation times can be easily controlled by usage of truncated Gaussian noise. Finite values of \(\epsilon\) lead to a greater or lesser extent to deviating observables, depending on the structure of the Boltzmann machine. The source of error is given by the error term of order \(O(\epsilon m^2)\) in the Taylor expansion of the detailed balance equation. Here, \(m_i\) corresponds to the total input for a neuron \(i\), according to Appendix D. Exact results of a Boltzmann machine can be obtained by an extrapolation to the limit \(\epsilon\).

### IV. NEUROMORPHIC HARDWARE SYSTEM

In this Section, we discuss different approaches for a projection and an accurate computation of the Boltzmann machine on a neuromorphic hardware system. This is one of the possible application of the sign-dependent Langevin machine. Several steps are necessary for a successful projection, as indicated in Figure 3. Each step describes a different level of abstraction of a neuromorphic system. A separate consideration of different aspects of such a system enables a clear distinction and identification of different sources of errors. Note that the diagram in Figure 3 is not the only possible approach to such a projection.

In [22], an analytic expression for the neural activation function of leaky integrate-and-fire (LIF) neurons has been derived for the hardware of the BrainScaleS project in Heidelberg. It was demonstrated how the neuromorphic hardware system can be used to perform stochastic inference with spiking neurons in the high-conductance state. The microscopic dynamics of the membrane potential of a neuron can be approximated in this state with...
Poisson-driven LIF neurons by an Ornstein-Uhlenbeck process. The spiking character of the system is obtained by a threshold function which maps the system onto an effective two-state system.

Particular properties of LIF sampling are:

(1) a description of the microscopic state of a neuron by a continuous membrane potential,
(2) an autocorrelated noise contribution to the membrane potential,
(3) a spiking character with an asymmetric refractory mechanism and
(4) non-trivial and non-constant interaction kernels between neurons.

In the present Section we study simplified dynamics of LIF sampling. This allows us to analyse the impact of particular hardware related properties and sources of resulting errors on different levels of abstraction. After a short introduction to the principles of LIF sampling, a mapping of the \( \text{LM}^2 \) is presented with respect to several particularities of the hardware system. Possible sources of errors of the mapping are discussed. We also relate our novel approach to the standard approach which relies on a fit of the activation function of the hardware system to the activation function of the Boltzmann machine.

The Section ends with an analysis of the impact of a refractory mechanism on the dynamics as a further step towards LIF sampling.

A. LIF Sampling

The spikey neuromorphic system of the BrainScaleS project emulates spiking neural networks with physical models of neurons and synapses implemented in mixed-signal microelectronics [22, 32]. With the help of Poisson-driven leaky integrate-and-fire (LIF) neurons, it is possible to obtain stochastic inference with deterministic spiking neurons. The dynamics of the free membrane potential \( u_{\text{eff}}(t) \) of a neuron can be approximated in the high-conductance state by an Ornstein-Uhlenbeck process,

\[
\frac{du_{\text{eff}}(t)}{dt} = \theta [\mu - u_{\text{eff}}(t)] + \sigma \eta(t).
\]

In (16), \( \theta \) determines the strength of the attractive force towards the mean value \( \mu = \mu^{\text{leak}} + \mu^{\text{average noise}} \). The mean value consists of some leak potential and an additional averaged noise contribution. \( \sigma \) depends on the contribution from the Poisson background.

Inspired by a biological neuron [33], the neuron emits a spike when the membrane potential exceeds a certain threshold \( \vartheta \). It is active and is reset to \( \varrho \) for a refractory time \( \tau_{\text{ref}} \) afterwards, where the neuron is considered as inactive. This is also sketched in Figure 4. One has to distinguish between the effective membrane potential \( u_{\text{eff}}(t) \) (red curve), which is unaffected by the spiking dynamics, and the real membrane potential \( u(t) \) (blue curve). As in [22], it is assumed that the convergence of \( u(t) \) from \( \varrho \) to \( u_{\text{eff}}(t) \) takes place in a negligible time after the finite refractory time has elapsed.

1. Activation Function

One can calculate distributions for the so called burst lengths and the mean first passage times of the membrane potential with the help of transition probabilities \( p(u_{i+1} | u_i) \). These are given by a corresponding Fokker-Planck equation of the Ornstein-Uhlenbeck process in the high conductance state. The burst length \( n \) is the number of consecutive spikes. The mean first passage time corresponds to the mean duration which it takes for the membrane potential to pass the threshold \( \vartheta \) from a lower
function is a cumulative Gaussian distribution, of the projected neuron states instead of their actual spatial is above a certain threshold, and as inactive otherwise. By: 

$$P(z = 1) = \frac{\sum P_n^{n\tau_{ref}}}{\sum P_n^{n\tau_{ref}} + \sum_{k=1}^{n-1} \frac{1}{\tau_k} + T_n},$$ (17)

where $\tau_k$ corresponds to the mean drift time from the resting potential $\vartheta$ to $\vartheta$. The distribution over burst lengths is represented by $P_n$ and the distribution over the times between burst regimes is given by $T_n$.

2. A Simplified Model

If the refractory time $\tau_{ref}$ is neglected, the neuron can be interpreted as active if the effective membrane potential is above a certain threshold, and as inactive otherwise. The resulting dynamics corresponds to the level of abstraction (c) in Figure 3. The neuron state is given by: $z(t) := \Theta[u_{eff}(t) - \vartheta]$. The process is referred to as an Ornstein-Uhlenbeck process with spiking character. Interacting contributions are implemented on the basis of the projected neuron states instead of their actual effective continuous membrane potential. The activation function is a cumulative Gaussian distribution,

$$P_{\text{eq}}(z = 1) = \int_{\vartheta}^{\infty} P_{\text{eq}}(u_{\text{eff}})du_{\text{eff}} = \Phi\left(\frac{\sqrt{2\mu\vartheta}}{\sigma}(\mu - \vartheta)\right),$$ (18)

with the equilibrium distribution $P_{\text{eq}}$ of the Ornstein-Uhlenbeck process (16),

$$P_{\text{eq}}(u_{\text{eff}}) = \sqrt{\frac{\theta}{\pi\sigma^2}} \exp\left(-\frac{\theta(u_{\text{eff}} - \mu)^2}{2\sigma^2}\right).$$ (19)

For convenience, the threshold potential $\vartheta$ is set to zero in further considerations.

The more accurate expression of the activation function, given by equation (17), takes the finite refractory time into account. The actual activation function is somewhere between the logistic distribution and the cumulative Gaussian distribution.

In the following, we neglect the finiteness of the refractory time. Therefore, we consider mostly simplified theoretical models of the hardware system of the level of abstraction (c). The models can be used to analyse Gaussian noise as stochastic input and the impact of autocorrelated noise in a system with a microscopic real time evolution of the membrane potential. Given by the spiking character, the considered models correspond effectively to interacting two-state systems. A discussion with respect to a refractory mechanism, as a process of the level of abstraction (d), is given in Section IV D.

B. Boltzmann Machine

In this Section we discuss a mapping of the Boltzmann machine (BM) onto an Ornstein-Uhlenbeck process with spiking character as a simplified model of a neuromorphic hardware system. A successful mapping demands a logistic distribution as activation function and a correct handling of interactions between neurons.

A possible approach is a fit of the activation function with a scaling parameter $r$ and a shift parameter $\mu_0$ to the desired logistic distribution according to $P_{\text{eq}}(z = 1) = \Phi\left(\frac{\sqrt{2\mu\vartheta}}{\sigma}(\mu - \vartheta)\right) \approx \sigma(\mu)$, with $\sigma(x) = [1 + \exp(-x)]^{-1}$ [12, 20, 22]. Interactions can be taken into account by absorbing their contributions into the mean value $\mu$ of the Ornstein-Uhlenbeck process according to: $\mu \rightarrow \mu + \mu_{\text{interaction}}$. It is assumed that the time to equilibrium is negligible after a change of an interacting neuron.

A mapping of the Boltzmann machine to a process of the level of abstraction (c) can be achieved by setting the average noise contribution to zero, adjusting the leak potential to $b_i$, and by taking the interacting contributions into account,

$$\mu_{\text{average noise}} = 0,$$

$$\mu_{\text{leak}} = b_i,$$

$$\mu_{\text{interaction}} = \sum_{\text{syn}} W_{ij} z_j(t).$$ (20)

Then, from the dynamics of equation (16) with a correct scaling of the interaction strength (see Appendix E) the following Ornstein-Uhlenbeck process with spiking char-
acter is obtained,

\[ \frac{du_{i,\text{eff}}(t)}{dt} = \theta \frac{1}{\sqrt{2}} \left[ \sum_{\text{syn}} W_{ij} z_j(t) + b_i - \mu_i^0 - u_{i,\text{eff}}(t) \right] \]

\[ + \sigma \eta(t), \]

with \( z_j(t) = \Theta[u_{j,\text{eff}}(t) - \vartheta] \) and where \( W_{ij} = \frac{A_{ij}}{\alpha} \) and \( A_{ij} \to \frac{A_{ij}}{\alpha} \). With \( \sigma = \sqrt{2} \), \( \theta = 1 \) we obtain the following activation function,

\[ P_{\text{OU}^1}(z_i = 1) = \Phi \left( \frac{\sum_{\text{syn}} W_{ij} z_j + b_i - \mu_i^0}{r} \right). \]

The process is abbreviated in the following by OU\(^1\). The '1' in the exponent is chosen in compliance with the LM\(^1\) (13) and indicates that the process takes place in one regime, i.e. the process does not fluctuate between two fundamental dynamics. The fitting of the activation function to the logistic distribution is indicated by the additional 'F'. The process without any fitting parameters \( (r = 1, \mu_i^0 = 0) \) is denoted as OU\(^1\).

We can also formulate an update rule with the same resulting activation function for a discrete two-state system, i.e., a system without a membrane potential. The resulting system is build upon an immediate representation of the neuron state, as it is the case for the BM and the LM\(^2\). The resulting process corresponds to a transition from the level of abstraction (c) to the level (b) and is driven by uncorrelated Gaussian noise.

The related update rule of level of abstraction (b) is derived in a similar manner as for the Langevin equation for discrete systems. It is given by,

\[ z'_i = \Theta \left[ \sum_{\text{syn}} W_{ij} z_j + b_i + \tilde{\eta} \right], \] (23)

where the updates take place in computer time. The corresponding transition probability reads,

\[ W_{\text{OU}^1F}(z_i \to 1) = \Phi \left( \frac{\sum_{\text{syn}} W_{ij} z_j + b_i - \mu_i^0}{r} \right). \] (24)

The dynamics has an additive Gaussian noise term and the Heaviside function as a projection onto the domain of definition of \( z_i \). Therefore, it is very similar to the sign-dependent discrete Langevin machine (13). The update rule is studied in [19] in more detail and introduced in [10] as an approximation of the so called \textit{Synaptic Sampling Machine}. In compliance with the LM\(^2\) and the OU\(^1\), we use the abbreviation LM\(^1\) for the process. When the activation function is fitted to the logistic distribution, LM\(^1\) is the corresponding acronym. In the latter case, sources of errors are resulting deviations due to an imperfect fit and finite times to equilibrium if an interacting neuron changes its macroscopic state [19].

The dynamics can be interpreted as another realisation of the discrete Langevin machine, as will be discussed in Section V. Properties and similarities of the two presented processes, i.e., the LM\(^1\) (23) and the OU\(^1\) (16), (21), are numerically investigated in Section VI.

C. Sign-dependent Langevin Machine

The sign-dependent discrete Langevin machine and LIF neurons exhibit similar underlying dynamics. This motivates a mapping of the LM\(^2\) onto an Ornstein-Uhlenbeck process with spiking character in the same manner as in the previous Section, i.e. from the level of abstraction (b) to (c). The resulting process represents a continuous counterpart to the sign-dependent discrete Langevin machine and is referred to as \textit{sign-dependent Ornstein-Uhlenbeck process} (OU\(^2\)). The activation function of the OU\(^2\) process converges in the limit \( \epsilon \to 0 \) also to a logistic function. As illustrated in Figure 1, the two processes differ in their microscopic representation and their time scales. The LM\(^2\) corresponds to a process with two discrete states and the computer time as time scale. The OU\(^2\) process describes the temporal evolution of a membrane potential in real time, whereas the interactions between neurons are based on a projection of the potential onto two states.

The total input \( \sum_j W_{ij} z_j + b_i \) of the dynamics of the previous Section is exchanged for a mapping onto an Ornstein-Uhlenbeck process by the redefined membrane potential of the sign-dependent discrete Langevin machine: \( W_{ii} z_i + \sum_j W'_{ij} z_j + b'_i \). This leads to the following dynamics of the sign-dependent Ornstein-Uhlenbeck process,

\[ \frac{du_{i,\text{eff}}(t)}{dt} = \theta \left[ W'_{ii} z_i(t) + \sum_{\text{syn}} W'_{ij} z_j(t) + b'_i - u_{i,\text{eff}}(t) \right] \]

\[ + \sigma \tilde{\eta}(t), \]

with \( z_i(t) = \Theta[u_{i,\text{eff}}(t) - \vartheta] \). The additional scaling factor of \( \lambda_i \) is omitted, i.e., it holds: \( W'_{ii} = \frac{\sqrt{2}}{\sqrt{\epsilon}} \), \( W'_{ij} = \sqrt{\frac{2}{\epsilon}} W_{ij} \) and \( b'_i = \left( \sqrt{\frac{2}{\epsilon}} b_i - \frac{1}{\alpha} \right) \). The term \textit{sign-dependent} reflects again the property of the neuron to stay very long in an active regime \((+1)\), or in an inactive regime \((-1)\).

Intuitive arguments for a convergence of the activation function of the process to a logistic distribution can be given by considering the regimes separately. The equilibrium distribution \( P_{eq}(u_{i,\text{eff}}) \) for a dynamics, which is only in the active regime, is given by equation (18) with
mu_i(t) = -\frac{1}{\sqrt{\sigma}} - \frac{\sqrt{\epsilon}}{2} \left( \sum_{j} W_{ij} z_j(t) + b_i \right). The equilibrium distribution P_{eq}^{\pm}(u_{i,eff}) of the inactive regime is implemented by mu_i(t) = -\frac{1}{\sqrt{\sigma}} + \frac{\sqrt{\epsilon}}{2} \left( \sum_{j} W_{ij} z_j(t) + b_i \right). The tails of these two distributions overlap in the region around \theta, as illustrated in Figure 5a. The separately considered equilibrium distributions are a good approximation for small values of \epsilon after reweighting them with the corresponding stationary probability distributions. If the membrane potential randomly crosses the threshold \theta, it perceives a strong drift towards the other regime, due to the changing mean value, i.e., due the switch z_i = 0 \rightarrow z_i = 1, and the large gap between the mean value of the two regimes. An immediate return to its initial regime is less likely. According to this argument, the transition probabilities of the sign-dependent Ornstein-Uhlenbeck process to become active or inactive can be approximated in the limit \epsilon \rightarrow 0 by the probability that the membrane potential reaches the threshold potential \theta = 0. This corresponds to the identifications

W_{OU^1}(0 \rightarrow 1) \simeq P_{eq}^{+}(u_{i,eff} = 0),

W_{OU^2}(0 \rightarrow 1) \simeq P_{eq}^{+}(u_{i,eff} = 0),

The resulting activation function of the sign-dependent Ornstein-Uhlenbeck process can be obtained by computing

P_{OU^2}(z_i = 1) = \frac{W_{OU^2}(0 \rightarrow 1)}{W_{OU^2}(0 \rightarrow 1) + W_{OU^2}(1 \rightarrow 0)}. \quad (27)

The distribution converges in the limit of \epsilon \rightarrow 0 to the logistic distribution

\lim_{\epsilon \rightarrow 0} P_{OU^2}(z_i = 1) = \frac{1}{1 + \exp \left[ -\sum_{j} W_{ij} z_j - b_i \right]}, \quad (28)

the activation function of the Boltzmann machine.

One resulting source of error is the worse approximation of the transition probabilities of the OU^2 process with P_{eq}^{+}(u_{i,eff}) and P_{eq}^{+}(u_{i,eff}) for larger values of \epsilon. The deviations can be seen in Figure 5a, when one compares the tail distributions of the fitted weighted equilibrium distributions of u_i of the two regimes with the actual measured distribution. The distributions of the two regimes are closer together and crossings between the active and the inactive state take place more often. This can be compensated to some extent by the introduction of an additional rescaling factor \tau as it has been done in the previous Section for the fitting of the OU^{1\tau} process. An advantage is that this can also be done after running the
A comparison of the dynamics in (25) with the mapping of the Boltzmann machine in equation (21) shows that the self-interaction and the dynamics of the discrete Langevin machine are key ingredients for a successful mapping onto the spiking system. The properties of the OU² process are dominated essentially by the self-interacting term. Therefore, the process is not just an Ornstein-Uhlenbeck process, but represents a new kind of dynamics with a different resulting equilibrium distribution and, up to now, non-investigated properties. Similar dynamics which contain projected values of interacting potentials might serve as a starting point for an entire new class of dynamics.

D. Refractory Mechanism

A possible further step towards LIF sampling is to take into account a refractory mechanism. This step is given in Figure 3 by the level of abstraction (d). The refractory mechanism can be also considered for a discrete system, for example for the Boltzmann machine. This approach represents a different ordering of the different abstractions of Figure 3.

In a simplified model, it can be assumed that a neuron stays active for the refractory time τref, after it got activated. An imbalance between the active and the inactive state is caused by this property. This asymmetry can be compensated by reducing the transition probability to become active by a factor of 1/τref, as discussed in [21]. The factor can be absorbed into the membrane potential by a shift of the activation function by log(τref), i.e., by $b_i \rightarrow b_i - \log(\tau_{ref})$. Note that the sign-dependent processes lead to a reformulation of the neuron computability condition of [21] due to the inherent dependency of the dynamics on the neuron state itself.

For the cumulative Gaussian distribution, an absorption of the factor of 1/τref is not possible anymore. The resulting activation function with a finite refractory time is deformed. The deformation gets larger for larger refractory times, as can be seen in Figure 7. We conclude that the errors of the activation function to the logistic distribution without a refractory mechanism propagate and increase for dynamics with finite refractory times τref. The resulting deformation of the activation function can be identified as a further source of error.

Within the last level of abstraction of Figure 3, interactions between neurons or with the neuron itself are in general not constant. The so called post-synaptic potential (PSP) corresponds to the received input potential of an interacting neuron [22]. In Appendix E, the relation between a correct implementation of the weights based on the interaction kernel is discussed in more detail. In this work, only rectangular PSP shapes are considered. An investigation of exponential PSP shapes is postponed to future work.

It is important to distinguish between the refractory mechanism as a property of each neuron itself and the post-synaptic potential. The latter needs only to be taken into account if interactions between neurons are considered. In particular, this means that the PSP shape affects only the activation function of the sign-dependent processes due to their self-interacting contribution.

V. DISCRETE LANGEVIN MACHINE

The Ornstein-Uhlenbeck process with a spiking dynamics and correlated noise offers the possibility to simulate a discrete two-state system by an underlying continuous dynamics. The discrete Langevin machine can be interpreted as a discrete counterpart to those spiking systems with uncorrelated noise, as indicated in Figure 1. It has been shown, that it is possible to map different realisations of simplified theoretical models of the neuromorphic hardware system onto a two-state system (see Figure 3 as well as the dynamics (21) ↔ (23) and (25) ↔ (13)). Denoting the hardware as $HW(p)$, with parameters $p = \{W_{ij}, b_i, z_i, r\}$, and the discrete Langevin machine as $LM(h)$, with $h := h(p)$, we can state the important relation,

$$\{LM(h)\} = \{HW(p)\}, \tag{29}$$

i.e., there exists a discrete two-state system for each set of parameters $p$ of the hardware which emulates the dy-
The numerical and the analytic scaling factor \( \lambda_c = \lambda(\sqrt{\epsilon}) \) for the sign-dependent Langevin machine.

Figure 7. Illustration of the non-symmetric deformation of the LM and the OU process for larger refractory times. The activation functions are shifted to comply: \( P(z_i = 1) \) for all realisations of the brane potential and spiking character and processes in terms of processes in real time with a continuous membrane potential and spiking character and processes in computer time with discrete states. For an exact mapping \( h(p) \), the magnitudes of the sources of error have to be matched. Table II gives an overview of the presented theoretical models and their properties regarding different levels of abstraction of a neuromorphic system.

VI. APPLICATIONS

Numerical results are discussed for the Langevin equation for discrete systems of Section II, for the newly introduced sign-dependent Ornstein-Uhlenbeck process as well as for existing approaches. We start with an analysis of the Clock model in Subsection VIA. Dynamics and equilibrium distributions of the free membrane potential are compared in Subsections VIB and VIC for the discrete Langevin machine and for abstractions of the neuromorphic hardware system, according to Figure 3 and Table II. The focus is on a correct implementation of the logistic distribution of the Boltzmann machine and on a detailed analysis of the impact of different sources of errors. The systems are considered with and without an asymmetric refractory mechanism with a rectangular post-synaptic shape. The Section ends with a computation of the Ising model by a projection of the model on the Boltzmann machine and with a numerical investigation of a Boltzmann machine with three neurons in subsection VID. Both models serve as a benchmark for Boltzmann distributed systems with interacting neurons.

A. \( q \)-state Clock Model

The \( q \)-state clock model [34, 35] describes spins \( \theta_i = 2\pi n \) with \( q \) different states which are parametrised by \( n \in \{1, 2, \ldots, q\} \). It is used to verify numerically the Langevin equation for discrete systems, as a first example. The
Table II. Comparison of the different analysed dynamics. An extrapolation to the exact solution and, therefore, a control of sources of errors is possible for the LM$^2$ and the OU$^2$ for both with and without a refractory mechanism.

Figure 8. Comparison of observables obtained by a standard Monte Carlo algorithm and with the LM$^2$ for the 4-state clock model on a $16 \times 16$ lattice. The results of the LM$^2$ converge for $\epsilon \to 0$ to the results of the standard Monte Carlo algorithm. Relative deviations of the inverse critical temperatures in dependence of $\epsilon$ are illustrated in Figure 9.

model has the following Hamiltonian,

\[ H_c = -J_c \sum_{\langle i,j \rangle} \cos (\theta_i - \theta_j) . \]  

(32)

The sum runs over all nearest neighbour spin pairs $\langle i,j \rangle$. In a complex plane one can interpret the spin states as equally distributed states on an unit circle. The common Potts model \[36\] is derived from this initial model. For $q = 2$ the model corresponds to the Ising model and in the limit of $q \to \infty$, it describes the continuous XY-model. For $q = 4$ the system emulates two independent Ising models.

The clock model exhibits for $q \leq 4$ a second order phase transition. It exists an exact solution for the inverse critical temperature for $q = 4$, which is as follows \[37\],

\[ J_c \beta_c^{q=4} = 2J_c \beta_c^{q=2}, \]  

(33)
where the Boltzmann constant $k_b$ has been set to 1. An appropriate order parameter for the system is the average magnetization per spin, which can be defined as

$$m = \frac{1}{N} \left| \sum_{k} e^{\frac{i 2 \pi \cdot n_k}{q}} \right|,$$

(34)

where the sum runs over all spins $n_k$ of a lattice with $N$ sites. The specific heat capacity $c$ per spin is considered as a further observable and is defined as

$$c = \frac{\beta^2}{N} \left( \langle E^2 \rangle - \langle E \rangle^2 \right),$$

(35)

where $\langle \cdot \rangle$ denotes the expectation value [38].

Numerical results for the magnetization and the specific heat are illustrated in Figure 8a and Figure 8b in dependency of the inverse temperature $\beta$ and of different values of $\epsilon$. Results of the Metropolis algorithm as a standard Monte Carlo algorithm (MC) serve as benchmark. The inverse critical temperature can be read off from the maximum of the specific heat. In Figure 9, the relative deviations of the inverse critical temperature to the inverse critical temperature of the Metropolis algorithm are plotted against $\epsilon$.

The resulting deviations for finite values of $\epsilon$ can be explained by a detailed error analysis of the transition probabilities of the Langevin equation for discrete systems. For this purpose, one has to check the compliance of the detailed balance equation,

$$\frac{W_{LM^2}(\theta \to \theta')}{W_{LM^2}(\theta' \to \theta)} + O(\epsilon \Delta H_e(\theta', \theta)^3) = \frac{P_{MC}(\theta')}{P_{MC}(\theta)}.$$

(36)

In Figure 16a, it can be seen that the absolute error of the cumulative Gaussian distribution is asymmetric around $x = 0$. This imbalance leads to a shift of the effective fraction of transition probabilities and therefore to a change of the equilibrium distribution of the spin states. The strength of this shift grows for larger values of $x$, which corresponds to larger values of $\beta \Delta H_e$, and larger values of $\epsilon$. The effect can be nicely observed in the change of the specific heat in Figure 8b with growing $\beta$. In general, it holds: the larger $|\beta \Delta H_e|$, the worse is the compliance of the detailed balance equation and the larger is the resulting shift of the equilibrium distribution.

**B. Neuromorphic Hardware vs. Langevin Machine**

We analyse numerically the mapping between dynamics of the discrete Langevin machine and the continuous dynamics according to relation (29) by an explicit consideration of transition probabilities. It is discussed the impact of deviations in the transition probabilities as well as a mapping of the temporal evolution of the different processes onto each other with respect to resulting activation functions for the free membrane potential.

Differences of the two dynamics which are given by construction are illustrated in Figure 1. The processes correspond to the levels of abstraction (b) and (c) of a neuromorphic hardware system in Figure 3. The essential differences are the source of noise which is for the Langevin machine uncorrelated and for the Ornstein-Uhlenbeck process correlated as well as the representation of a microscopic state. The dynamics is described in the former case by two discrete states in computer time and in the latter one by the evolution of a continuous membrane potential with spiking character in real time. We evaluate the impact of the different sources of errors on the deviation to the expected logistic distribution for the sign-dependent and for the fitted processes: LM$^2$, OU$^2$, LM$^{1F}$ and OU$^{1F}$.

**1. Activation Function**

Figures 7 and 10a illustrate the activation functions of the free membrane potential in dependency of the bias $b$ in the network for the different presented dynamics. The results of the LM$^1$ and the OU$^1$ process coincide exactly and their deviation to the cumulative Gaussian distribution emerges from numerical errors. In concordance to these observations, the fitted LM$^{1F}$ and OU$^{1F}$ process have the same deviations to the logistic distribution in Figure 10a. In the case of the LM$^2$ and the OU$^2$ process, the observed deviations mirror the theoretical errors for finite values of $\epsilon$. As depicted in Figure 10c, both activation functions converge in the limit of $\epsilon \rightarrow 0$. The rate of convergence of the OU$^2$ process is much smaller than the one of the LM$^2$ for equal values of $\epsilon$. This can be reasoned by the different sources of errors for the two processes, as discussed in detail at the ends of Sections III and IV C. The results of the OU$^2$ process motivate the research for an analytic or numerical expression of the dependency $r(\epsilon)$. In contrast to the LM$^{1F}$ and OU$^{1F}$, deviations to the logistic distribution are limited only due to larger correlation times for smaller values of $\epsilon$ for the OU$^2$ process.
2. Dynamics - Time Evolution

It has been found numerically that the computer time and the real time coincide for the LM\(^1\) and the Ornstein-Uhlenbeck process. All simulations in real time are performed with finite time steps of 0.02. All processes in computer time are computed with a random sequential update formalism and in real time by a parallel update scheme. The time scale in all figures is chosen in units of the computer time.

Figure 11 compares trajectories of the different discussed processes with respect to a uniform time scale. It can be observed in the evolution of the membrane potential for all processes that there occur fast changes if the membrane potential is close to the threshold value \( \vartheta = 0 \). These perturbations seem to have no influence on the time evolution and the equilibrium distribution.

As discussed in Section II, a scaling factor \( a \) can be found for a correct mapping of the temporal evolution of two processes A and B if both processes exhibit the same equilibrium distribution. The scaling factor is given under usage of equation (11) and by a computation of the transition probabilities by

\[
    a = \frac{W_A(0 \rightarrow 1)}{W_B(0 \rightarrow 1)}. \tag{37}
\]

Analytic expressions for the transition probabilities of the considered processes are given in Table I. The given
transition probabilities have been validated numerically. For that purpose we have mapped the temporal ensemble evolution of the different dynamics onto the evolution of the Boltzmann machine with respect to the computed scaling factors. A scaling factor $a \neq 1$ reflects the increase/decrease of the correlation time for processes with different transition probabilities. In Figure 12, the dependency of the scaling factor $a$ on $\epsilon$ is plotted for the sign-dependent processes.

The considerations of the time evolution reinforce that the relation of equation (29) corresponds to an exact mapping of the dynamics of a discrete system with uncorrelated noise to a continuous system with correlated noise. This property is not self-evident. However, the dependency $h(p)$ is in some cases non-trivial, due to different occurring sources of errors of the considered models.

C. Refractory Mechanism

In this Section we investigate the impact of an asymmetric refractory mechanism of a neuromorphic system with a rectangular PSP shape. This has been introduced in Section IV D as the level of abstraction $(d)$ with regard to Figure 3.

1. Control of the Refractory Mechanism

We concentrate on a correct representation of the logistic function with a refractory mechanism. The imbalance between the inactive and the active state can in general not be compensated entirely by a trivial shift of the bias by $\log(\tau_{\text{ref}})$ for the cumulative Gaussian distribution. The activation functions of the sign-dependent processes and the fitted dynamics are close to a logistic distribution. Therefore, a shift can be used to approximate the logistic distribution with a refractory mechanism. However, for large refractory times this approximation gets worse, as indicated in Figure 7.

In this work, the shift of the activation function is determined by the constraint that $p(z_i = 1)_{k=0} = 0.5$. The actual shifts of the bias are slightly different for the LM$^{1F}$ and the OU$^{1F}$ as a consequence of resulting deformations of the cumulative Gaussian distribution for larger refractory times (see Figure 7). We also introduce a further time constant $\tau'_{\text{ref}}$. This allows us to distinguish clearly between the refractory time $\tau_{\text{ref}}$ of a neuron and the resulting optimal shift $\log(\tau'_{\text{ref}})$ for a correct fixing of the activation function. Ideally, one should derive a dependency $\tau'_{\text{ref}}(\tau_{\text{ref}})$ to preserve a consistent fixing of the activation function.

For the LM$^2$ process it holds true that $\tau_{\text{ref}} \approx \tau'_{\text{ref}}$, since the deviation of the activation function to the logistic distribution is nearly symmetric around $b = 0$. Nevertheless, a shift by $\log(\tau'_{\text{ref}})$ leads to a worse approximation of the transition probability, since the Taylor expansion of the cumulative Gaussian distribution of the LM$^2$ dynamics is performed around 0. This is a further error source.

For the OU$^2$ process, the necessary shift of the bias by $\log(\tau'_{\text{ref}})$ is much larger than $\log(\tau_{\text{ref}})$. Dependencies of $\tau'_{\text{ref}}(\tau_{\text{ref}})$ for fixed values of epsilon and $\tau_{\text{ref}}(\epsilon)$ for $\tau_{\text{ref}} = 8$ are illustrated in Figure 13. The large differences in $\tau_{\text{ref}}$ and in $\tau'_{\text{ref}}$ can be traced back to the different microscopic dynamics of the processes and to the different origin of a correct implementation of the activation function for the processes without a refractory mechanism.

In contrast to the OU$^3$ process, the dynamics of the OU$^2$ process fluctuates between the active and the inactive regime which are distinguished and driven by the self-interacting contribution. Integrating a refractory mechanism, a change from the active to the inactive regime by the self-interacting term is suppressed as long...
as the neuron is captured in its refractory mode. This can be seen in the lower plot of Figure 5b. Consequently, the distributions \( P^{\text{eq}}(u_{i,\text{eff}}) \) and \( P^{\text{eq}}'(u_{i,\text{eff}}) \) have a dissimilar impact on the resulting distribution of \( P(u_{i,\text{eff}}) \). The lower tail distribution of \( P^{\text{eq}}(u_{i}) \), i.e., the distribution for \( u_{i,\text{eff}} < \vartheta \), biases the distribution \( P(u_{i,\text{eff}}) \) around the threshold \( \vartheta = 0 \) within the refractory time. In contrast, the upper tail distribution of \( P^{\text{eq}}'(u_{i}) \) does not affect \( P(u_{i,\text{eff}}) \), since the dynamics changes for \( u_{i,\text{eff}} > \vartheta \) from the inactive to the active regime. The local minimum of \( P(u_{i,\text{eff}}) \) around \( u_{i,\text{eff}} = \vartheta \) as well as the entire distribution \( P(z_{i}) \) are shifted to smaller values, as a result of this asymmetry, as illustrated in Figure 5b. Further, the absolute value of the minimum is larger than the one for the process without a refractory mechanism. The imbalance between \( P^{\text{eq}}(u_{i,\text{eff}}) \) and \( P^{\text{eq}}'(u_{i,\text{eff}}) \) results in larger deviations of the activation function for the OU\(^2\) process with a refractory mechanism. This asymmetry corresponds to a further source of error.

The equilibrium distribution needs to undergo a larger shift by \( \log(\tau_{\text{ref}}') \) than the OU\(^{1}\)F process for a compensation of the impact of the refractory mechanism. This is a consequence of a partially suppression of the change of the dynamics to the inactive regime. For the OU\(^{1}\)F process, the underlying dynamics is not affected by the refractory mode due to the absence of a self-interacting term. Therefore, the only purpose of the shift by \( \log(\tau_{\text{ref}}') \) is to fix the transition probabilities to correctly compensate the emerging asymmetry of the refractory mechanism. Respectively, the resulting transition probabilities are expressed in dependence of \( \log(\tau_{\text{ref}}) \). Analytic expressions are given in Table 1.

2. Activation Function and Time Evolution

Figure 10b compares the impact of a refractory mechanism on the different dynamics regarding their deviations to the logistic distribution. The deviations of the LM\(^2\) and the OU\(^2\) process have increased, as expected by the introduced asymmetry of the refractory mechanism. Nevertheless, the error vanishes for \( \epsilon \to 0 \), as illustrated in Figure 10c. Further, Figure 10d shows that a further increase of the refractory time has a very low impact on the deviations which ensures an applicability for large refractory times, in practice. As discussed in Section IV D, the cumulative Gaussian distribution is non-symmetrically deformed by the shift by \( \log(\tau') \). This leads to deviations in the activation function of the LM\(^1\) and the OU\(^1\) process that can be compensated to a certain extent by an adaptation of the variance, i.e., of the scaling parameter \( r \).

We conclude that a refractory mechanism with rectangular PSP shape has no impact on a possible control of the sources of errors for the sign-dependent processes.

D. Interacting Systems

We consider the Ising model [39] and the Boltzmann machine [40] to investigate the presented abstractions of a neuromorphic hardware system with interactions between neurons. The Ising model can be easily mapped onto the Boltzmann machine. A numerical analysis can be understood as a proof of concept that the presented processes also work in a more complex network set-up. As a second model, we study a Boltzmann machine with three neurons. We compare the results for all presented models with and without a refractory mechanism with a rectangular PSP shape.

The Ising model describes a two-state spin system. The spin states are \( s_{i} \in \{-1, +1\} \), which are likewise also referred to as spin up and spin down \( s_{i} \in \{\downarrow, \uparrow\} \). The Hamiltonian is defined as

\[
H = -J \sum_{\langle i,j \rangle} s_{i}s_{j} - h \sum_{i} s_{i} .
\]

The external magnetic field \( h \) is set to zero in the following numerical analysis and \( J = 1 \) is some coupling constant. For this particular case, we can consider the averaged absolute value of the magnetization per spin as an order parameter. This is then given by

\[
m = \frac{1}{N} \left| \sum_{i} s_{i} \right| ,
\]

where the sum runs again over all spins of the lattice for a given configuration. From theoretical considerations, an exact expression for the inverse critical temperature of the model with a vanishing external field can be obtained [41],

\[
J\beta_{c} = \frac{\ln(1 + \sqrt{2})}{2} .
\]

For a computation with the presented algorithms, we need a mapping between the Boltzmann machine and the Ising model onto the correct domain of definition. The mapping of \( s_{i} = -1 \to z_{i} = 0 \) and \( s_{i} = 1 \to z_{i} = 1 \) can be obtained by the following identifications between \( J \) and \( h \) and \( W_{ij} \) and \( b_{i} \) [33, 42],

\[
W_{ij} = 4J ,
\]
\[ b_i = 2h - 2Jd, \]  

(41)

where \( d \) corresponds to the dimension of the system. The spin state can be computed by \( s_i = 2z_i - 1 \).

The Boltzmann machine can have an arbitrarily complex network structure. Particular implementations like the restricted Boltzmann machine turn the Boltzmann machine to an interesting class of networks, which has many applications in different areas of research, see e.g. [16–18, 43]. To study the impact of systems with a higher possible variability, we consider a Boltzmann machine with three neurons and different weights an biases around zero. The Kullback-Leibler divergence [44] serves as a measure to numerically classify the quality of the presented processes. We compute the Kullback-Leibler divergence based on the history of a process, starting from a random initial state according to

\[ D_{KL}(P_{BM}||P_{AM}) = - \sum_{c \in \Omega} P_{BM}(c) \log \frac{P_{AM}(c)}{P_{BM}(c)}. \]  

(42)

BM indicates the exact probability distribution of the
Boltzmann machine and AM corresponds to the approximated probability distribution of some other model. The sum runs over all possible neuron configurations \( c \). The probabilities are approximated by the corresponding histograms of the history of the trajectory in the configuration space.

Figures 14a and 14b show the absolute value of the magnetization for the Ising model with a vanishing external magnetic field for the dynamics without and with a refractory mechanism. The observables are computed for the LM\(^2\) and the OU\(^2\) process for different values of \( \epsilon \) and for the LM\(^{1F}\) and the OU\(^{1F}\) process. In Figure 9, the deviation of the derived inverse critical temperatures is plotted in dependency of \( \epsilon \) for the processes without a refractory mechanism. Figure 15 illustrates the convergence of the considered processes for vanishing \( \epsilon \). The resulting deviations reflect the magnitude of errors in the representation of the activation function. This reinforces, together with the results of Figure 14, the argument that already small changes in the activation function can lead to large deviations in the resulting observables. This argument also explains the partially worse performance of the processes for the dynamics with a refractory mechanism.

The observables exhibit for the LM\(^2\) and the OU\(^2\) process the same tendency as the results for the 4-state clock model, despite their different sources of errors. The equilibrium distributions are shifted to smaller values of \( \beta \) as described in the discussion of Section VI.A. As before, the shift grows with larger values of \( \beta \) and of \( \epsilon \). The similar behaviour of the OU\(^2\) process can be justified by the similar trend in the deviation of the activation functions of the two processes. The higher rate of convergence of the LM\(^2\) process is a result of the different source of errors.

The comparison of the Kullback-Leibler divergence of the different processes in Figures 14c and 14d reinforces the better representation of the logistic distribution by the sign-dependent processes and illustrates again the dependency on \( \epsilon \).

VII. CONCLUSIONS AND OUTLOOK

In the present work we have introduced the discrete Langevin machine with discrete Langevin dynamics (1), (5), see in particular the Sections II, III, V.

The newly introduced dynamics pave the way for possible new applications and the discovery of new physics. This includes, for example, a formulation of Langevin dynamics for discrete systems with respect to a possible computation of Hamiltonians with a complex contribution, similar to complex Langevin dynamics [27, 28]. A further interesting task is to investigate the novel network architecture of the sign-dependent Ornstein-Uhlenbeck process (OU\(^2\)), given by the dynamics (25) in Section IV, with its particularity of a self-interacting term and resulting fluctuating dynamics, on a neuromorphic hardware system.

The numerical analysis of different abstractions of a LIF network in Section VI demonstrates that the novel network architecture of the sign-dependent discrete Langevin machine (LM\(^2\)) and the OU\(^2\) process is suitable for an exact computation of correlation functions of Boltzmann distributed systems. This applies to both, a discrete two-state system with uncorrelated noise and a continuous system with autocorrelated noise. The numerical results show that an exact implementation of the logistic distribution or at least a correct estimation of errors is necessary to obtain quantitative exact observables.

It remains to be seen whether this statement is also sufficient and valid for non-trivial PSP shapes, as a last step towards LIF sampling. In particular, one has to analyse the impact of marginal deviations to the activation function on observables of larger and more complex systems than the one considered in this work. Moreover, one may ask whether an exact representation of an activation function with a self-interacting term is sufficient to also obtain reliable and accurate results for interacting neurons independent of the interaction kernel, i.e. the post-synaptic potential, respectively. In other words: Is it possible to extend findings for a single self-interacting neuron to a general complex interacting system? These questions are postponed to future work. Either way, we expect that the existence of a self-interacting contribution in the OU\(^2\) process helps to better understand arising non-linearities of the neuromorphic hardware.

In summary, the potentially more accurate implementation of Boltzmann machines by the dynamics (13) and (25) represents a further step towards an integration of deep learning and neuroscience [5, 9, 12]. We believe that the present work offers a tool for a better
comparison of classical artificial networks and neuromorphic networks.

ACKNOWLEDGEMENTS

We thank M. A. Petrovici, A. Baumbach and K. Meier for discussions and collaboration on related subjects. This work is supported by Deutsche Forschungsgemeinschaft (DFG) under Germany’s Excellence Strategy EXC-2181/1 - 390900948 (the Heidelberg Excellence Cluster STRUCTURES).

Appendix A: Metropolis Algorithm vs. Langevin Dynamics

The Section reviews shortly properties of the Metropolis algorithm [45] and Langevin dynamics for continuous systems to get a first intuition on how a possible Langevin equation for a discrete system could look like. The considerations are inspired by the work in [46–48].

We adopt the common formulation of the Langevin equation within the study of Euclidean quantum field theories [24, 25],

\[ \partial_{\tau} \phi_x (\tau) = -\frac{\delta S_E}{\delta \phi_x (\tau)} + \eta_x (\tau), \]

where \( S_E \) corresponds to the Euclidean action, which depends on fields \( \phi_x (\tau) \) on a \( (3+1) \) dimensional hypercubic lattice in Euclidean space. The Langevin equation describes the evolution of the quantum fields \( \phi_x (\tau) \) in an additional fictitious time dimension, the Langevin time \( \tau \). Quantum fluctuations are emulated by the additional white Gaussian noise term with the properties to be uncorrelated,

\[ \langle \eta_i, \eta_j \rangle = \delta (j - i) \delta (t' - t), \quad \langle \eta_i \rangle = 0. \]

It can be shown that in equilibrium the resulting distribution of the fields coincides with the Boltzmann distribution: \( \lim_{\tau \to \infty} P(\phi, \tau) = P_{eq}(\phi) = \frac{1}{Z} \exp(-S_E). \) A common approach to prove this is to derive the equivalence of the Langevin equation and the Fokker-Planck equation in a first step and to compute the static solution in a second step [24]. This property renders Langevin dynamics a powerful tool in QCD [27, 28].

1. Transition Probability of the Langevin Equation

The transition probabilities of the discrete Langevin equation are computed, in the following. These are used in a second step for an interpretation of the dynamics as a standard Monte Carlo algorithm. Starting from the discrete Langevin equation

\[ \phi' = \phi - \epsilon \frac{\delta S}{\delta \phi_x} + \sqrt{\epsilon} \eta, \]

with: \( \phi := \phi (\tau) \) and \( \phi' := \phi (\tau + \epsilon) \), it is straightforward to compute the transition probabilities of an infinitesimal change, \n
\[ W(\phi \rightarrow \phi') = \frac{1}{\sqrt{2\epsilon}} \varphi \left( \frac{\phi' - \phi}{\sqrt{2\epsilon}} + \sqrt{\epsilon} \frac{\delta S}{\delta \phi} \right). \quad (A4) \]

Inserting the standard normal distribution \( \varphi (x) = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{x^2}{2} \right) \) and computing the square in the exponent one obtains

\[ W(\phi \rightarrow \phi') = \frac{1}{\sqrt{4\pi \epsilon}} \exp \left[ -\frac{1}{2} \left( \frac{\phi' - \phi}{\sqrt{2\epsilon}} + \frac{\sqrt{\epsilon} \delta S}{\delta \phi} \right)^2 \right] = \varphi \left( \frac{\phi' - \phi}{\sqrt{2\epsilon}} \right) \exp \left[ -\frac{\phi' - \phi}{\sqrt{2\epsilon}} \frac{\delta S}{\delta \phi} + O(\epsilon) \right]. \quad (A5) \]

With the identifications \( \delta \phi \simeq \phi' - \phi \) and \( \delta S \simeq S(\phi') - S(\phi) \), this can be further simplified to

\[ W(\phi \rightarrow \phi') = \frac{1}{\sqrt{2\epsilon}} \varphi \left( \frac{\phi' - \phi}{\sqrt{2\epsilon}} \right) \exp \left[ -\frac{S(\phi') - S(\phi)}{2} + O(\epsilon) \right]. \quad (A6) \]

Apparently, the transition probability satisfies the detailed balance equation since the first factor is symmetric to an exchange of \( \phi' \) and \( \phi \),

\[ \frac{W(\phi \rightarrow \phi')}{W(\phi' \rightarrow \phi)} = \exp \left[ -(S(\phi') - S(\phi)) \right]. \quad (A7) \]

2. Equivalence to a Monte Carlo Algorithm

As shown previously, the discrete Langevin equation corresponds to an infinitesimal update step of a field \( \phi := \phi (\tau) \) to a new field \( \phi' := \phi (\tau + \epsilon) \) with transition probability

\[ W(\phi \rightarrow \phi') = \frac{1}{\sqrt{2\epsilon}} \varphi \left( \frac{\phi' - \phi}{\sqrt{2\epsilon}} \right) \exp \left[ -\frac{S(\phi') - S(\phi)}{2} \right]. \quad (A8) \]

The first factor can be interpreted as the selection probability of a new field and the second term as the acceptance probability. There are two adaptations to the Metropolis algorithm. Firstly, the acceptance probability has the additional factor of \( \epsilon \). This factor is necessary to satisfy the detailed balance equation. Secondly, the proposal state is chosen to be an infinitesimal change, \( \phi \rightarrow \phi' \), as can be seen by replacing the selection probability by the delta distribution,

\[ \int_{-\infty}^{\infty} d\phi W(\phi \rightarrow \phi') = 1, \]

\[ \int_{-\infty}^{\infty} d\phi W(\phi \rightarrow \phi') = \frac{1}{\sqrt{2\epsilon}} \varphi \left( \frac{\phi' - \phi}{\sqrt{2\epsilon}} \right) \exp \left[ -\frac{S(\phi') - S(\phi)}{2} \right]. \]

\[ (A9) \]
\[ \int_{-\infty}^{\infty} d\phi' \delta(\phi' - \phi) \exp \left[ - \frac{S(\phi') - S(\phi)}{2} \right] = 1. \]  

(A9)

The update mechanism of the Langevin dynamics can be parallelised because of these two properties.

Accordingly, the Langevin equation (A1) can be interpreted as a standard Monte Carlo algorithm with a Gaussian distribution as selection probability. The proposal state is chosen implicitly by an absorption of the acceptance probability into the selection probability and by a corresponding sampling with Gaussian noise. Since the nearest neighbour sites can be assumed to be nearly constant in one Monte Carlo step, it is possible to switch from a random sequential update formalism to a parallel update of the entire lattice. The Langevin time is introduced as a temporal measure for a lattice update. In principle, the delta distribution can be exchanged by any other positive representation. In some cases one can derive Langevin dynamics with another noise source by an equivalent approximation (for example, Cauchy noise).

Having the knowledge of this Section, a Langevin equation for discrete systems can be constructed by considering a standard Monte Carlo algorithm.

**Appendix B: Relation between the Cumulative Normal Distribution and the Exponential Function**

Turning to discrete states, the normal distribution as a density probability distribution transforms to differences of cumulative normal distributions.

To be able to define an update formalism with a Gaussian noise term, we need a relation between the exponential function of cumulative normal distributions. Such a relation exists and is given by

\[ \lim_{\epsilon \to 0} n_{\epsilon,0}(x) = \lim_{\epsilon \to 0} \frac{\Phi(-\frac{1}{\sqrt{\epsilon}} + \sqrt{\epsilon} x)}{\Phi(-\frac{1}{\sqrt{\epsilon}})} = \exp(x) + \mathcal{O}(\epsilon x^2), \]

(B1)

with a scaling factor

\[ \lambda_{\epsilon} = \frac{\sqrt{\epsilon} \varphi(-\frac{1}{\sqrt{\epsilon}})}{\Phi(-\frac{1}{\sqrt{\epsilon}})}, \]

(B2)

and the cumulative normal distribution \( \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp \left[ - \frac{1}{2} t^2 \right] dt \).

As shown in Appendix C, the denominator in the limit corresponds to a scaling of the zero order term of the Taylor expansion of the cumulative normal distribution. The rescaling of \( x \) in the argument corrects the first order term. The resulting order of accuracy is therefore of second order in \( \sqrt{\epsilon} \).

The relation can be extended to the \( m \)-th derivative of the cumulative distribution with \( m > 0 \), according to

\[ \lim_{\epsilon \to 0} n_{\epsilon,m}(x) = \lim_{\epsilon \to 0} \frac{\partial^m \Phi \left( -\frac{1}{\sqrt{\epsilon}} + \sqrt{\epsilon} t \right)}{\partial \epsilon^m} \bigg|_{t=\frac{x}{\sigma_{m,\epsilon}}} = \exp(x) + \mathcal{O}(\epsilon x^2), \]

(B3)

where the scaling factor \( \sigma_{m,\epsilon} \) is defined as

\[ \sigma_{m,\epsilon} = \frac{\sqrt{\epsilon} \text{He}_m \left( -\frac{1}{\sqrt{\epsilon}} \right)}{\text{He}_{m-1} \left( -\frac{1}{\sqrt{\epsilon}} \right)}, \]

(B4)

and where \( \text{He}_m(x) \) denote the \( m \)-th probabilists’ Hermite polynomials.

For \( m = 1 \), this corresponds to a similar computation as in Appendix A1 for the Langevin equation of continuous systems. The scaling factor of \( x \) becomes 1 and the resulting identity simplifies to

\[ \lim_{\epsilon \to 0} n_{\epsilon,1}(x) = \lim_{\epsilon \to 0} \frac{\varphi \left( -\frac{1}{\sqrt{\epsilon}} + \sqrt{\epsilon} x \right)}{\varphi \left( -\frac{1}{\sqrt{\epsilon}} \right)} = \exp(x) + \mathcal{O}(\epsilon x^2). \]

(B5)

**Appendix C: Derivation of the Relation between the Cumulative Normal Distribution and the Exponential Function**

The relations (B1) and (B3) are derived. For reasons of readability, \( \sqrt{\epsilon} \) is abbreviated by \( \varepsilon \) and the shorthand notation \( \frac{\partial^m}{\partial \varepsilon^m} = \partial^m \) is used in the following.

We start with a Taylor series around \( x = 0 \) of the \( m \)-th derivate of the cumulative Gaussian contribution,

\[ \partial^m \Phi \left( -\frac{1}{\varepsilon} + \varepsilon x \right) = \sum_{n=0}^{\infty} \frac{n!}{\varepsilon^{n+1}} \partial^{n+m} \Phi \left( -\frac{1}{\varepsilon} + \varepsilon x \right) \bigg| \varepsilon = 0^+ x^n. \]

(C1)

The following important identity between the cumulative Gaussian distribution \( \Phi(x) \) and the probabilists’ Hermite polynomials \( \text{He}_n(x) \) is useful for an evaluation of the Taylor expansion,

\[ \partial^{n+m} \Phi \left( -\frac{1}{\varepsilon} + \varepsilon x \right) = (-\varepsilon)^{n+m-1} \text{He}_{n+m-1} \left( -\frac{1}{\varepsilon} + \varepsilon x \right) \partial \Phi \left( -\frac{1}{\varepsilon} + \varepsilon x \right), \]

(C2)

for \( n > 0 \). Since this relation holds only for \( n > 0 \), the cases for \( m = 0 \) and \( m > 0 \) have to be treated separately, which coincides with the relations (B1) and (B3).
This gives

\[ \Phi \left( \frac{1}{\varepsilon} + \varepsilon x \right) = \Phi \left( \frac{1}{\varepsilon} \right) + \sum_{n=1}^{\infty} \frac{1}{n!} (-\varepsilon)^{n-1} x^n \]

\times \text{He}_{n-1} \left( \frac{1}{\varepsilon} + \varepsilon x \right) \bigg|_{x=0} \partial \Phi \left( \frac{1}{\varepsilon} + \varepsilon x \right) \bigg|_{x=0} x^n = \]

\[ = \Phi \left( \frac{1}{\varepsilon} \right) + \sum_{n=1}^{\infty} \frac{1}{n!} (-1)^{n-1} \varepsilon^n \text{He}_{n-1} \left( \frac{1}{\varepsilon} \right) \varphi \left( \frac{1}{\varepsilon} \right) x^n. \quad (C3) \]

A comparison with the Taylor expansion of \( \exp(x) = 1 + x + \mathcal{O}(x^2) \) shows that the first two terms in the above expression can be fixed by a division of the entire equation by \( \Phi \left( \frac{1}{\varepsilon} \right) \) and an additional rescaling of \( x \) by

\[ \lambda(\varepsilon) = \frac{\varepsilon \varphi \left( \frac{1}{\varepsilon} \right)}{\Phi \left( \frac{1}{\varepsilon} \right)}. \quad (C4) \]

This gives

\[ \Phi \left( \frac{1}{\varepsilon} + \varepsilon \frac{x}{\lambda(\varepsilon)} \right) = \]

\[ = 1 + x + \sum_{n=2}^{\infty} \frac{1}{n!} (-1)^{n-1} \varepsilon^n \text{He}_{n-1} \left( \frac{1}{\varepsilon} \right) \lambda(\varepsilon) x^n. \quad (C5) \]

It remains to show that the fractional factor converges to 1 for \( \varepsilon \to 0 \) and for arbitrary values of \( n \). This is done in two steps. Firstly, we argue that \( \lim_{\varepsilon \to 0} \lambda(\varepsilon) = 1 + \mathcal{O}(\varepsilon^2) \) and secondly, a limit is derived for the fractional factor.

The limit of \( \lim_{\varepsilon \to 0} \lambda(\varepsilon) \) can be derived by showing the identity that \( \lim_{\varepsilon \to 0} \Phi \left( \frac{1}{\varepsilon} \right) = \lim_{\varepsilon \to 0} \varphi \left( \frac{1}{\varepsilon} \right) \). By the substitution \( u := \frac{1}{x} \) and a subsequent partial integration, one finds that a second order term in \( x = \frac{1}{\varepsilon} \) vanishes and arrives directly at the identity which entails that

\[ \lim_{\varepsilon \to 0} \lambda(\varepsilon) = 1 + \mathcal{O}(\varepsilon^2). \quad (C6) \]

Since the highest order term of the \( n \)-th probabilists’ Hermite polynomial equals \( x^n \), it can be directly concluded that

\[ \lim_{\varepsilon \to 0} \varepsilon^n \text{He}_n \left( \frac{1}{\varepsilon} \right) = (-1)^n + \mathcal{O}(\varepsilon^n). \quad (C7) \]

Using the Taylor expansion

\[ \frac{1}{(1 + x)^{n+1}} = 1 - (n - 1) x + \mathcal{O}(x^2), \quad (C8) \]

and inserting the two limits (C6) and (C7), one arrives at the following limit for the fractional factor,

\[ \lim_{\varepsilon \to 0} \frac{(-1)^{n-1} \varepsilon^{n-1} \text{He}_{n-1} \left( \frac{1}{\varepsilon} \right)}{\lambda(\varepsilon)^{n-1}} = 1 + \mathcal{O}(\varepsilon^2). \quad (C9) \]

The final limit between the cumulative normal distribution and the exponential function can be stated with the corresponding order of accuracy,

\[ n_{\varepsilon,0}(x) = \frac{\Phi \left( \frac{1}{\varepsilon} + \varepsilon \frac{x}{\lambda(\varepsilon)} \right)}{\Phi \left( \frac{1}{\varepsilon} \right)} = \exp(x) + \mathcal{O}(\varepsilon^2 x^2). \quad (C10) \]

The existence of this limit can also be proven by applying L’Hospital’s rule to relation (B5) as shown in \cite{49}.

2. Evaluation for \( m > 0 \)

Proceeding similarly as for the case of \( m = 0 \), the Taylor expansion can be written as

\[ \partial^m \Phi \left( \frac{1}{\varepsilon} + \varepsilon x \right) = \partial^m \Phi \left( \frac{1}{\varepsilon} + \varepsilon x \right) \bigg|_{x=0} + \sum_{n=1}^{\infty} \frac{1}{n!} \times \]

Figure 16. Illustrations regarding the limit \( \lim_{\varepsilon \to 0} n_{\varepsilon,m}(x) \) of relations (B1) and (B3) for \( m = 0 \) and \( m = 1 \). The vertical lines in Figures 16b and 16c indicate the respective fixed value of \( \epsilon \) and \( x \). In general, the limit with the cumulative Gaussian distribution (\( m = 0 \)) has a lower deviation for equal values of \( \epsilon \) then the limit with the Gaussian distribution (\( m = 1 \)).
\[ \times (-1)^{n+m-1} \varepsilon^{n+m} H_{n+m-1} \left( -\frac{1}{\varepsilon} \right) \varphi \left( -\frac{1}{\varepsilon} \right) x^n. \]  

(C11)

The fixing of the first two order terms of the exponential function leads to

\[ \frac{\partial^n \varphi \left( -\frac{1}{\varepsilon} \right)}{\partial x^n} \bigg|_{x=0} = 1 + x + \sum_{n=2}^{\infty} \frac{1}{n!} \varepsilon^n H_{n-1} \left( -\frac{1}{\varepsilon} \right) \varphi \left( -\frac{1}{\varepsilon} \right) \sigma_m (\varepsilon)^n x^n, \]  

(C12)

where we have evaluated the expression \( \partial^n \varphi \left( -\frac{1}{\varepsilon} \right) \bigg|_{x=0} \) on the right hand side and where the scaling factor \( \sigma_m (\varepsilon) \) is given by

\[ \sigma_m (\varepsilon) = -\frac{\varepsilon H_m \left( -\frac{1}{\varepsilon} \right)}{H_{m-1} \left( -\frac{1}{\varepsilon} \right)}. \]  

(C13)

Again, the asymptotic behaviour of the fractional factor has to be computed for \( \varepsilon \to 0 \). From the highest order term of the probabilists’ Hermite polynomial, it can be deduced that

\[ \lim_{\varepsilon \to 0} \frac{\varepsilon^n H_{n+m-1} \left( -\frac{1}{\varepsilon} \right)}{H_{m-1} \left( -\frac{1}{\varepsilon} \right)} = (-1)^n + O(\varepsilon^n). \]  

(C14)

For \( n = 1 \), this corresponds to \((-1)\) times the scaling factor \( \sigma(\varepsilon) \), therefore,

\[ \lim_{\varepsilon \to 0} \sigma(\varepsilon) = 1 + O(\varepsilon). \]  

(C15)

It can be derived with the same arguments as for the case of \( m = 0 \), that the fractional factor converges to 1 and that the limit with its order of accuracy is given by

\[ n \varepsilon^2 = \frac{\partial^n \varphi \left( -\frac{1}{\varepsilon} \right)}{\partial x^n} \bigg|_{x=0} = \exp(x) + O(\varepsilon^2 x^2). \]  

(C16)

Appendix D: Derivation of the Dynamics of the Langevin Machine

Starting from the energy of the Boltzmann machine

\[ E = -\sum_{i<j} W_{ij} z_i z_j - \sum_i b_i z_i, \]  

(D1)

the total input for a neuron \( i \) can be defined as

\[ m_i := -\sum_j W_{ij} z_j - b_i. \]  

(D2)

The possible resulting energy differences for a change of the state of the neuron are given by

\[ \Delta E(z'_i = 1, z_i = 0) = m_i, \]

\[ \Delta E(z'_i = 0, z_i = 1) = -m_i. \]  

(D3)

The proposed state for a two-state system always corresponds to the other state. This results in the following two update rules for a transition from \( z_i = 0 \to 1 \) and \( z_i = 1 \to 0 \),

\[ z'_i = \Theta \left[ 1 - \frac{\varepsilon}{2 \lambda_c} m_i + \varepsilon \tilde{\eta}_i \right], \]  

(D4)

Taking the current state into account, the relations can be merged into a common update rule,

\[ z'_i = \Theta \left[ (2 z_i - 1) - \frac{\varepsilon}{2 \lambda_c} m_i + \varepsilon \tilde{\eta}_i \right]. \]  

(D5)

After a division by \( \sqrt{\varepsilon} \) and a rearranging of the summands, one arrives at the following update rule for the Langevin machine,

\[ z'_i = \Theta \left[ 2 \varepsilon z_i + \sum_j \sqrt{\frac{\varepsilon}{2 \lambda_c}} W_{ij} z_j + \sqrt{\frac{\varepsilon}{2 \lambda_c}} b_i \right] \]  

(D6)

By the identifications

\[ W'_{ii} = \frac{2}{\sqrt{\varepsilon}}, \quad W'_{ij} = \frac{\sqrt{\varepsilon}}{2 \lambda_c} W_{ij}, \quad b'_i = \left( \frac{\sqrt{\varepsilon}}{2 \lambda_c} b_i - \frac{1}{\sqrt{\varepsilon}} \right), \]  

(D7)

the update rule can be written as

\[ z'_i = \Theta \left[ W'_{ii} z_i + \sum_j W'_{ij} z_j + b'_i + \tilde{\eta}_i \right]. \]  

(D8)

Appendix E: Interactions on the Neuromorphic Hardware System

The interaction between neurons is non-trivial. The post-synaptic potential (PSP) corresponds to the input potential of a connected neuron in case of firing. As shown in [22], the interaction term can be approximated by

\[ \mu_i(t)_{\text{interaction}} = \sum_{\text{synaptic spikes}} A_{ij} \kappa(t, t_{s,j}), \]  

(E1)

with \( A_{ij} = \frac{w_{ij} (k_{ij} - \langle w_{ij} \rangle)}{\langle y_{tot} \rangle} \) and where \( t_{s,j} \) is the time of the last spike. \( \kappa(t, t_{s,j}) \) describes the PSP shape and
depends in general on the time constants: $\tau_{\text{ref}}$, $\tau_{\text{syn}}$ and $\tau_{\text{eff}}$.

The actual PSP shape has the following form \cite{22},

$$\kappa(t, t_{s,j}) = \frac{\exp \left[ -\frac{t - t_{s,j}}{\tau_{\text{ref}}} \right] - \exp \left[ -\frac{t - t_{s,j}}{\tau_{\text{syn}}} \right]}{\tau_{\text{eff}} - \tau_{\text{syn}}}.$$  \hspace{1cm} (E2)

Weights $W_{ij}$ can be translated onto the neuromorphic system by the assumption that the area under a PSP shape is equal to $W_{ij}\tau_{\text{ref}}\alpha$, where $\alpha$ is some scaling factor,

$$W_{ij}\tau_{\text{ref}}\alpha = \int_0^{\tau_{\text{ref}}} A_{ij} \kappa(t, t_{s,j}) \, dt.$$  \hspace{1cm} (E3)

Ideally, the PSP shape would have a rectangular form,

$$\kappa(t, t_{s,j})_{\text{rect}} = \Theta \left[ t - t_{s,j} \right] - \Theta \left[ t - t_{s,j} - \tau_{\text{ref}} \right].$$  \hspace{1cm} (E4)

For $\tau_{\text{ref}} \to 0$, the neuron $j$ is in the firing mode as long as $u_j(t) > \vartheta_j$ and the interaction term turns to

$$\mu_{i}(t)_{\text{interaction}} = \sum_{\text{synj}} A_{ij} \Theta \left[ u_j(t) - \vartheta_j \right].$$  \hspace{1cm} (E5)

---

[1] W. Maass, Neural Networks 10, 1659 (1997).
[2] Y. Huang and R. P. Rao, in Advances in Neural Information Processing Systems 27, edited by Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger (Curran Associates, Inc., 2014) pp. 1943–1951.
[3] D. Jimenez Rezende and W. Gerstner, Frontiers in Computational Neuroscience 8, 38 (2014).
[4] V. Korniijcuk, H. Lim, J. Y. Seok, G. Kim, S. K. Kim, I. Kim, B. J. Choi, and D. S. Jeong, Frontiers in Neuroscience 8, 10651 (2018), arXiv:1709.08166 [cs.NE].
[5] C. M. Lapilli, P. Pfeifer, and C. Wexler, Phys. Rev. Lett. 116, 10651 (2016).
[6] W. Gerstner, H. Sprekeler, and G. Deco, Science 38, 60 (2012).
[7] C. D. Schuman, T. E. Potok, R. M. Patton, J. D. Birdwell, M. E. Dean, G. S. Rose, and J. S. Plank, ArXiv e-prints (2017), arXiv:1705.06963.
[8] W. Maass, Proceedings of the IEEE 102, 860 (2014).
[9] E. Neftci, S. Das, B. Pedroni, K. Kreutz-Delgado, and G. Cauwenberghs, Frontiers in Neuroscience 7, 272 (2014).
[10] E. O. Neftci, B. U. Pedroni, S. Joshi, M. Al-Shedivat, and G. Cauwenberghs, Frontiers in Neuroscience 10, 241 (2016).
[11] B. U. Pedroni, S. Das, E. Neftci, K. Kreutz-Delgado, and G. Cauwenberghs, in The 2013 International Joint Conference on Neural Networks (IJCNN) (2013) pp. 1–6.
[12] L. Leng, R. Martel, O. Breitwieser, I. Bytschok, W. Senn, L. Buesing, K. Meier, and M. A. Petrovici, Scientific Reports 8, 10651 (2018), arXiv:1709.08166 [cs.NE].
[13] G. E. Hinton, Neural Comput. 14, 1771 (2002).
[14] D. Pccevski, L. Buesing, and W. Maass, PLOS Computational Biology 7, 1 (2011).
[15] B. Nessler, M. Pfeiffer, L. Buesing, and W. Maass, PLOS Computational Biology 7, 1 (2011).
[16] M. A. Petrovici, J. Bill, I. Bytschok, J. Schemmel, and K. Meier, Phys. Rev. E 94, 042312 (2016).
[17] D. S. Lemons and A. Gythiel, American Journal of Physics 65, 1079 (1997).
[18] P. H. Danggaard and H. Huffel, Phys. Rept. 152, 227 (1987).
[19] G. G. Batrouni, G. R. Katz, A. S. Kronfeld, G. P. Lepage, B. Svetitsky, and K. G. Wilson, Phys. Rev. D 32, 2736 (1985).
[20] G. Parisi and Y.-s. Wu, Sci. Sin. 24, 483 (1981).
[21] G. Aarts and I.-O. Stamatescu, Journal of High Energy Physics 2008, 018 (2008).
[22] G. Aarts, F. A. James, J. M. Pawlowski, E. Seiler, D. Sexty, and I.-O. Stamatescu, JHEP 03, 073 (2013), arXiv:1212.5231 [hep-lat].
[23] G. Aarts, Proceedings, 30th International Symposium on Lattice Field Theory (Lattice 2012): Cairns, Australia, June 24–29, 2012, PoS LATTICE2012, 017 (2012), arXiv:1302.3028 [hep-lat].
[24] M. Welling and Y. W. Teh, in Proceedings of the 28th International Conference on International Conference on Machine Learning, ICML’11 (Omnipress, USA, 2011) pp. 681–688.
[25] P. S. Neelakanta, R. Sudhakar, and D. Degroff, Biol. Cybern. 65, 331 (1991).
[26] J. Schemmel, D. Breiderle, A. Giibi, M. Hock, K. Meier, and S. Millner, in Proceedings of 2010 IEEE International Symposium on Circuits and Systems (2010) pp. 1947–1950.
[27] M. A. Petrovici, Form Versus Function: Theory and Models for Neuronal Substrates, Ph.D. thesis (2016).
[28] R. B. R. Potts, “The mathematical investigation of some cooperative phenomena,” (1951), typescript.
[29] R. B. Potts, Proc. Cambridge Phil. Soc. 48, 106 (1952).
[30] F. Y. Wu, Rev. Mod. Phys. 54, 235 (1982).
[31] G. G. Batrouni, G. R. Katz, A. S. Kronfeld, G. P. Lepage, B. Svetitsky, and K. G. Wilson, Phys. Rev. D 32, 2736 (1985).
[32] G. Aarts, F. A. James, J. M. Pawlowski, E. Seiler, D. Sexty, and I.-O. Stamatescu, JHEP 03, 073 (2013), arXiv:1212.5231 [hep-lat].
[33] M. Welling and Y. W. Teh, in Proceedings of the 28th International Conference on International Conference on Machine Learning, ICML’11 (Omnipress, USA, 2011) pp. 681–688.
[34] P. S. Neelakanta, R. Sudhakar, and D. Degroff, Biol. Cybern. 65, 331 (1991).
[35] J. Schemmel, D. Breiderle, A. Giibi, M. Hock, K. Meier, and S. Millner, in Proceedings of 2010 IEEE International Symposium on Circuits and Systems (2010) pp. 1947–1950.
[36] M. A. Petrovici, Form Versus Function: Theory and Models for Neuronal Substrates, Ph.D. thesis (2016).
[37] R. B. R. Potts, “The mathematical investigation of some cooperative phenomena,” (1951), typescript.
[38] R. B. Potts, Proc. Cambridge Phil. Soc. 48, 106 (1952).
[39] F. Y. Wu, Rev. Mod. Phys. 54, 235 (1982).
[40] G. G. Batrouni, G. R. Katz, A. S. Kronfeld, G. P. Lepage, B. Svetitsky, and K. G. Wilson, Phys. Rev. D 32, 2736 (1985).
[41] G. Aarts, F. A. James, J. M. Pawlowski, E. Seiler, D. Sexty, and I.-O. Stamatescu, JHEP 03, 073 (2013), arXiv:1212.5231 [hep-lat].
[42] M. Welling and Y. W. Teh, in Proceedings of the 28th International Conference on International Conference on Machine Learning, ICML’11 (Omnipress, USA, 2011) pp. 681–688.
[39] E. Ising, Zeitschrift für Physik 31, 253 (1925).
[40] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, Cognitive Science 9, 147 (1985).
[41] L. Onsager, Phys. Rev. 65, 117 (1944).
[42] I. Bytschok, D. Dold, J. Schemmel, K. Meier, and M. A. Petrovici, arXiv e-prints, arXiv:1707.01746 (2017), arXiv:1707.01746 [q-bio.NC].
[43] G. Montúfar, CoRR abs/1806.07066 (2018), arXiv:1806.07066.
[44] S. Kullback and R. A. Leibler, Ann. Math. Statist. 22, 79 (1951).
[45] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. 21, 1087 (1953).
[46] C. F. Baillie and D. A. Johnston, Phys. Rev. D 39, 1246 (1989).
[47] P. Meakin, H. Metiu, R. G. Petschek, and D. J. Scalapino, J. Chem. Phys. 79, 1948 (1983).
[48] R. Ettelaie and M. A. Moore, Journal of Physics A: Mathematical and General 17, 3505 (1984).
[49] C. Yi, Quantitative Finance 10, 957 (2010).