Spiking neural networks as superior generative and discriminative models

Luziwei Leng*, Mihai A. Petrovici*, Roman Martel, Ilja Bytschok,
Oliver Breitwieser, Johannes Bill, Johannes Schemmel, Karlheinz Meier

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386 · Kirchhoff-Institute for Physics · Electronic Vision(s) · BrainScaleS ScaleS · The Manfred Stärk Foundation
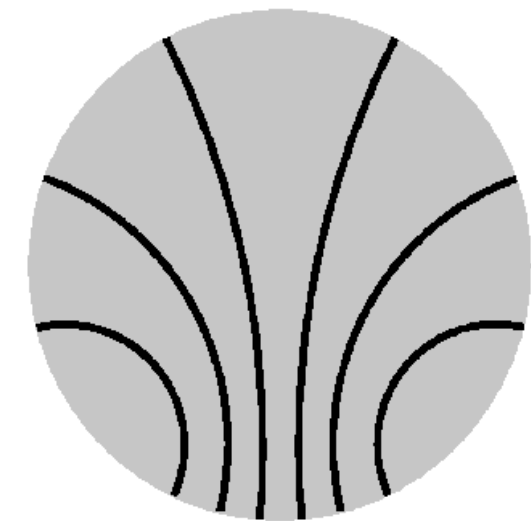
## Probabilistic inference with LIF neurons in the high-conductance state

An increasing number of experiments suggest that the brain performs stochastic inference when dealing with incomplete and noisy sensory information. This, in turn, has led to the development of various theoretical models that attempt to explain how this could be achieved with spiking neural networks. Our theory interprets spiking activity as sampling from distributions over binary random variables [1] and is compatible with the ensemble dynamics of noise-driven LIF neurons in the high-conductance state (HCS) [2, 3].

LIF neuron model $C_m \frac{du}{dt} = g_l(E_l - u) + I^{syn} + I^{ext}$

HCS: $\tau_{eff} = \frac{C_m}{g_{tot}} \ll \tau_{syn}$

100 ms

The free membrane potential can be treated as an Ornstein-Uhlenbeck process

$du(t) = \theta \cdot (\mu - u(t)) + \Sigma \cdot dW(t)$

$\theta = \tau_{syn} \rightarrow$ synaptic memory!
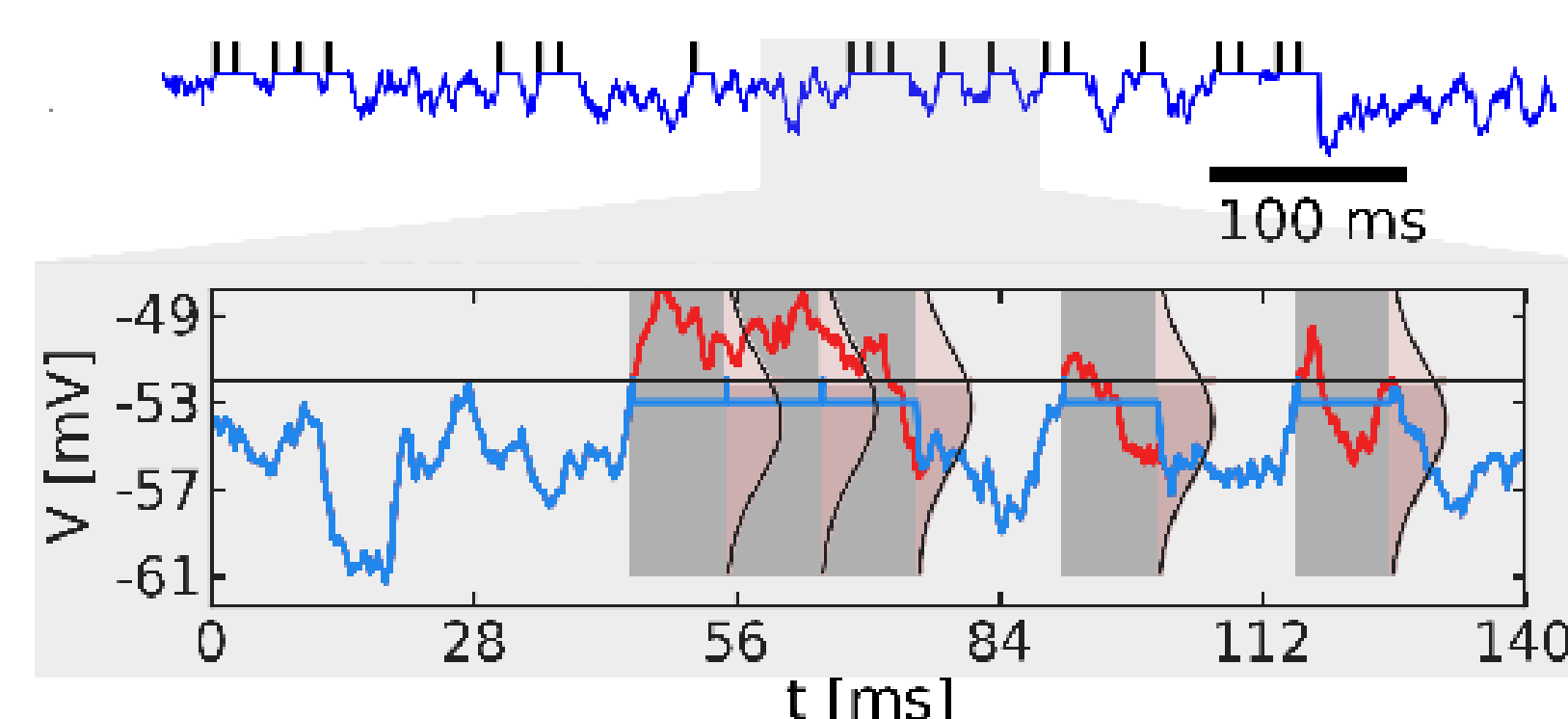
Other theories [4, 5] do not hold in this regime, but it is possible to find a recursive expression for the membrane potential during bursts

$P_n = \left(1 - \sum_{i=1}^{n-1} P_i\right) \int_{\vartheta}^{\infty} du_{n-1} p(u_{n-1}|u_{n-1} \geq \vartheta)$

$\left[\int_{-\infty}^{\vartheta} du_n p(u_n|u_{n-1})\right]$

and during inter-burst intervals, which can be treated as a first-passage-time problem

$T_n = \int_{\vartheta}^{\infty} du_{n-1} p(u_{n-1}|u_{n-1} \geq \vartheta)$

$\left[\int_{-\infty}^{\vartheta} du_n p(u_n|u_n < \vartheta, u_{n-1}) \langle T(\vartheta, u_n)\rangle\right]$

The resulting theoretical prediction accurately matches simulation data [2]

$p(z = 1) = \frac{\sum_n P_n \cdot n \cdot \tau_{ref}}{\sum_n P_n \cdot \left(n\tau_{ref} + \sum_{k=1}^{n-1} \overline{\tau_k^b} + T_n\right)}$
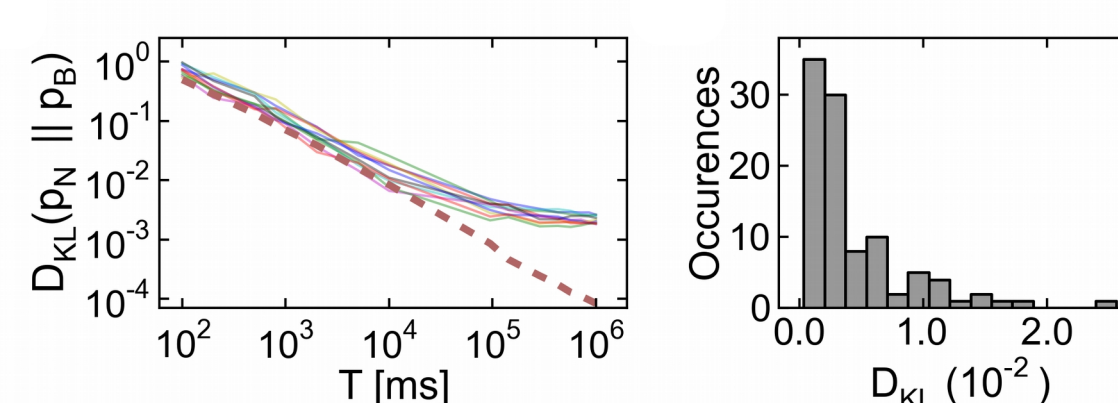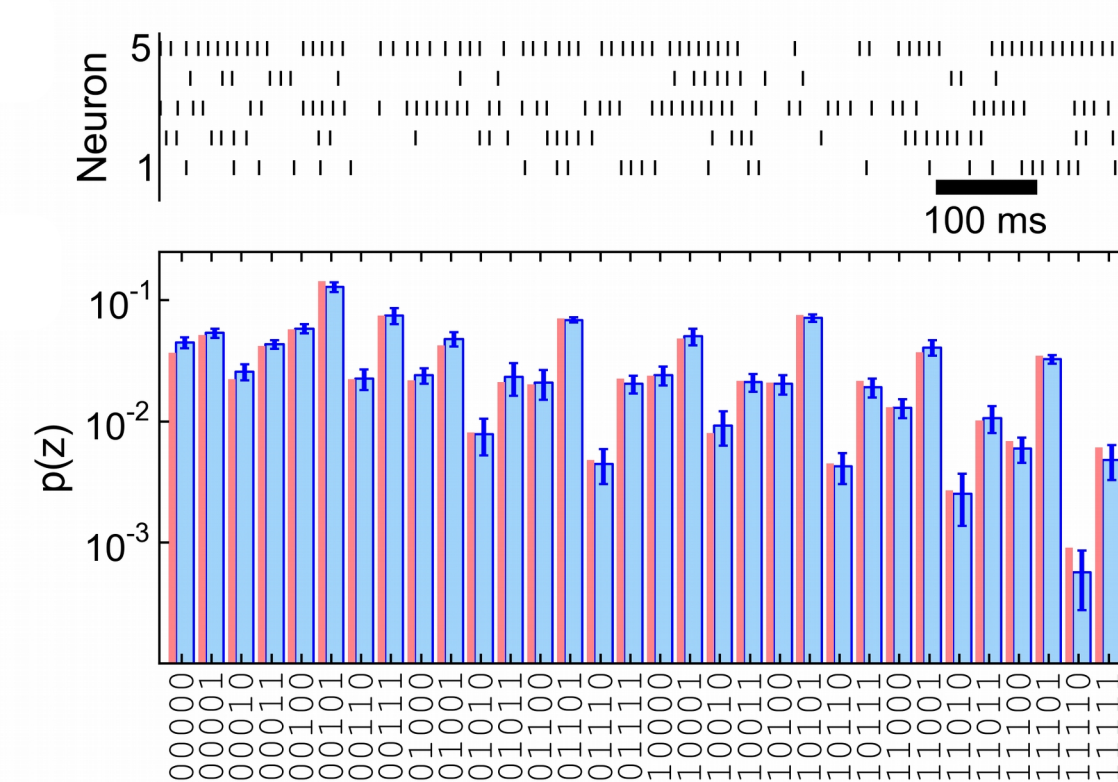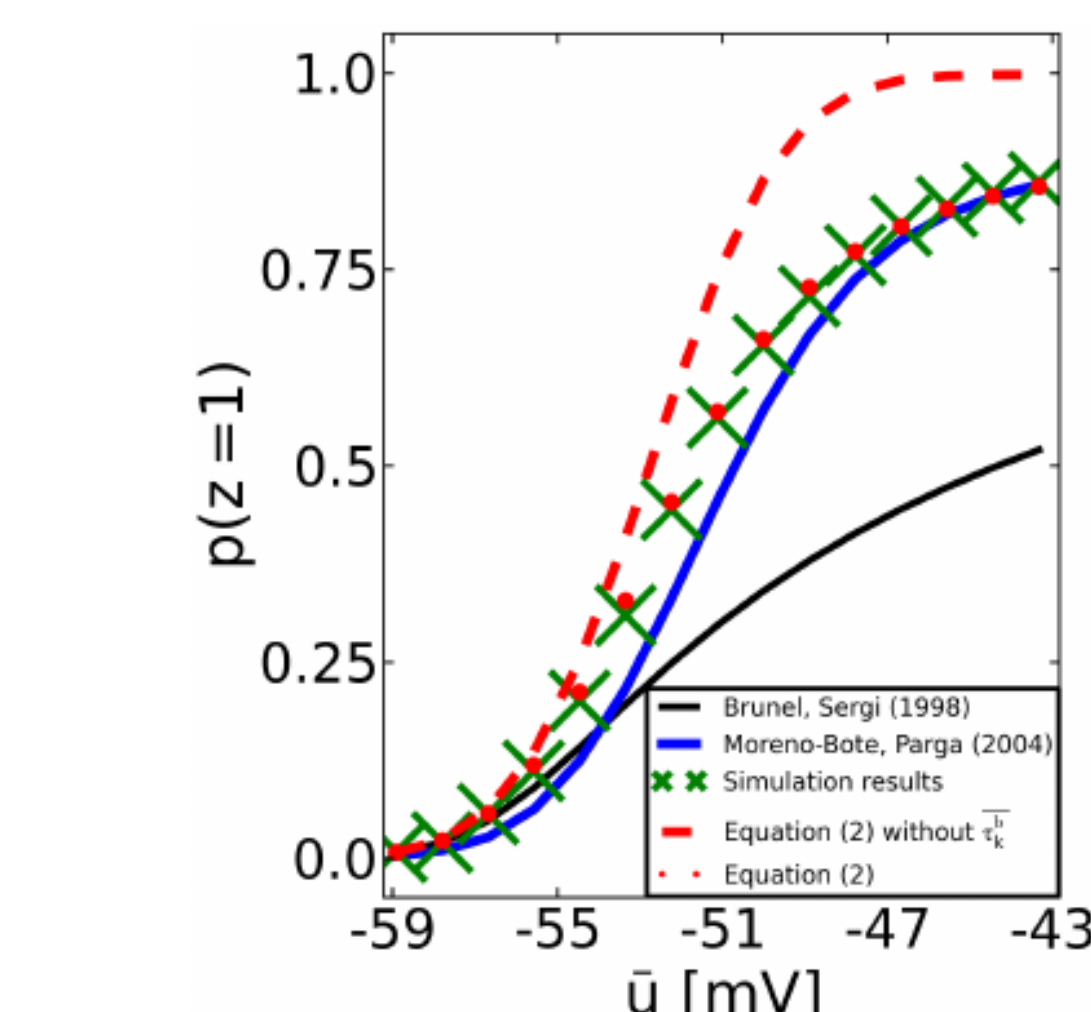
In the HCS, the LIF activation function becomes symmetric, allowing networks to sample from Boltzmann distributions [2]

$u_k(t) = \log \frac{p(z_k = 1|z_{\setminus k})}{p(z_k = 0|z_{\setminus k})}$

$b_k = (\bar{u}_k^b - \bar{u}_k^0)/\alpha$
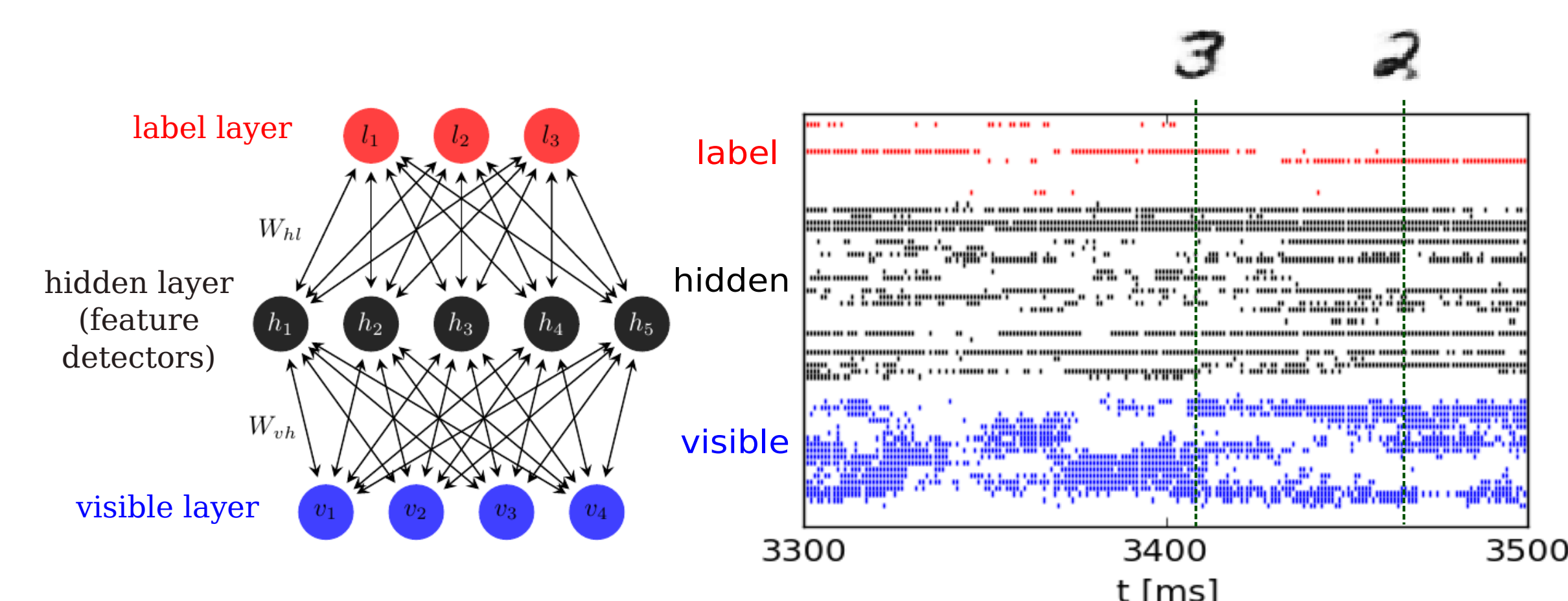
$W_{kj} = \frac{1}{\alpha C_m} \frac{w_{kj}(E_{kj}^{rev} - \mu)}{1 - \frac{\tau_{syn}}{\tau_{eff}}}$

$\left[\tau_{syn}(e^{-1} - 1) - \tau_{eff}\left(e^{-\frac{\tau_{syn}}{\tau_{eff}}} - 1\right)\right]$


Deep spiking neural networks for image classification

Based on our LIF sampling framework, we can construct large-scale spiking neural networks (approx. 2000 neurons) that sample from restricted Boltzmann distributions.

Supervised training (coupled adaptive simulated tempering, [6]) on handwritten digits leads to the formation of receptive fields (both in the hidden layer and the label layer) that capture essential features of the ten different digits. After training, the activity of the deepest layer, which contains one neuron per digit class, can be directly used to classify input data.


Receptive fields of label units
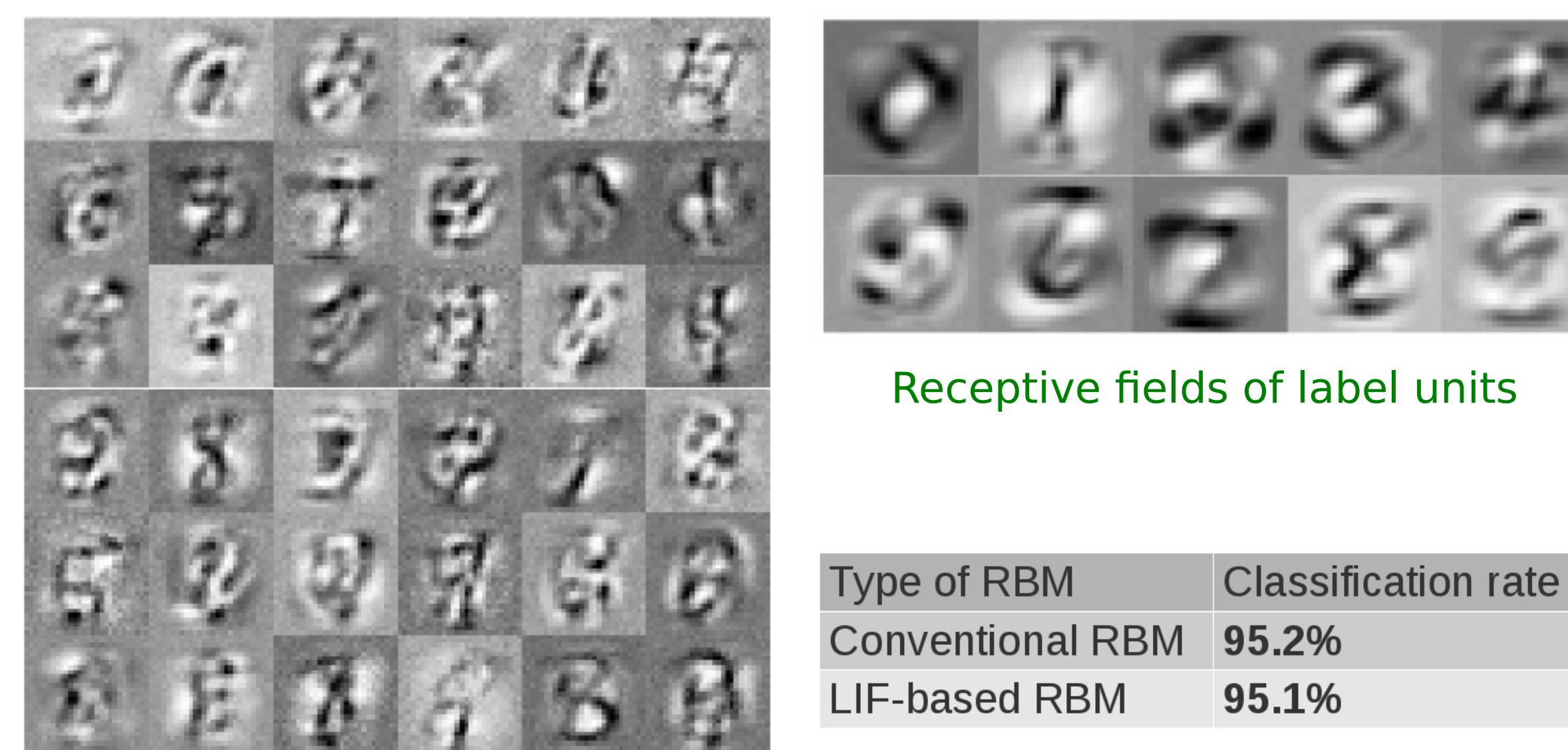
Receptive fields of hidden units

Type of RBM	Classification rate
Conventional RBM	95.2%
LIF-based RBM	95.1%

The achieved performance (classification on the test set of the MNIST [7] dataset, 10,000 handwritten digits, 28×28 grayscale pixels) of deep LIF networks is only slightly lower than the one of classical RBMs [8] trained with the same learning method. The result can be further improved with larger network size and longer fine tuning during the training period.

Generative models and short-term plasticity

In classical RBMs, statistics are typically gathered by Gibbs sampling:

$p(z_k^{(t)} \rightarrow z_k^{(t+1)}) = p(z_k|z_{\setminus k})$

This algorithm has a distinct disadvantage when dealing with high-dimensional multimodal distributions, where it often gets trapped in local minima due to deep troughs in the energy landscape that appear during training. It is for this reason that conventional RBMs that may perform very well as discriminative models are, at the same time, rather poor generative models of the learned data. While methods exist that alleviate this problem [6], they usually come at a highly increased computational cost.
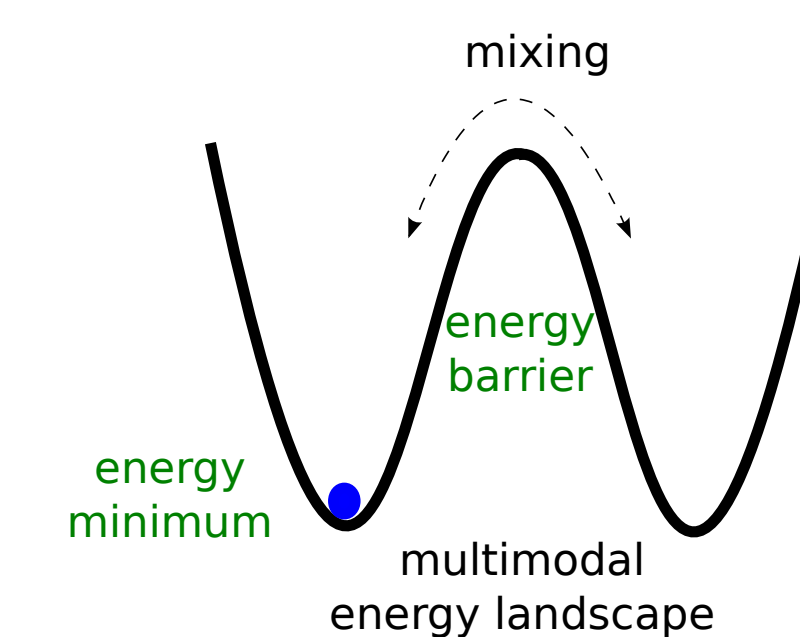
Short-term plasticity [9] allows the activity-dependent modulation of synaptic interaction.

$\frac{dR}{dt} = \frac{(1-R)}{\tau_{rec}} - U_{SE} R \delta(t - t_{sp})$

$\frac{dU_{SE}}{dt} = -\frac{U_{SE}}{\tau_{facil}} + U1 \cdot (1 - U_{SE}) \cdot \delta(t - t_{sp})$

regular · potentiation · depression
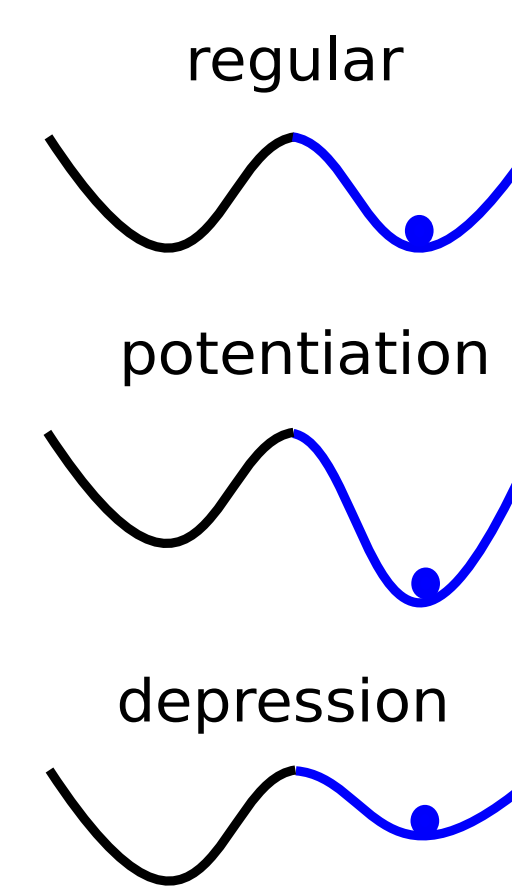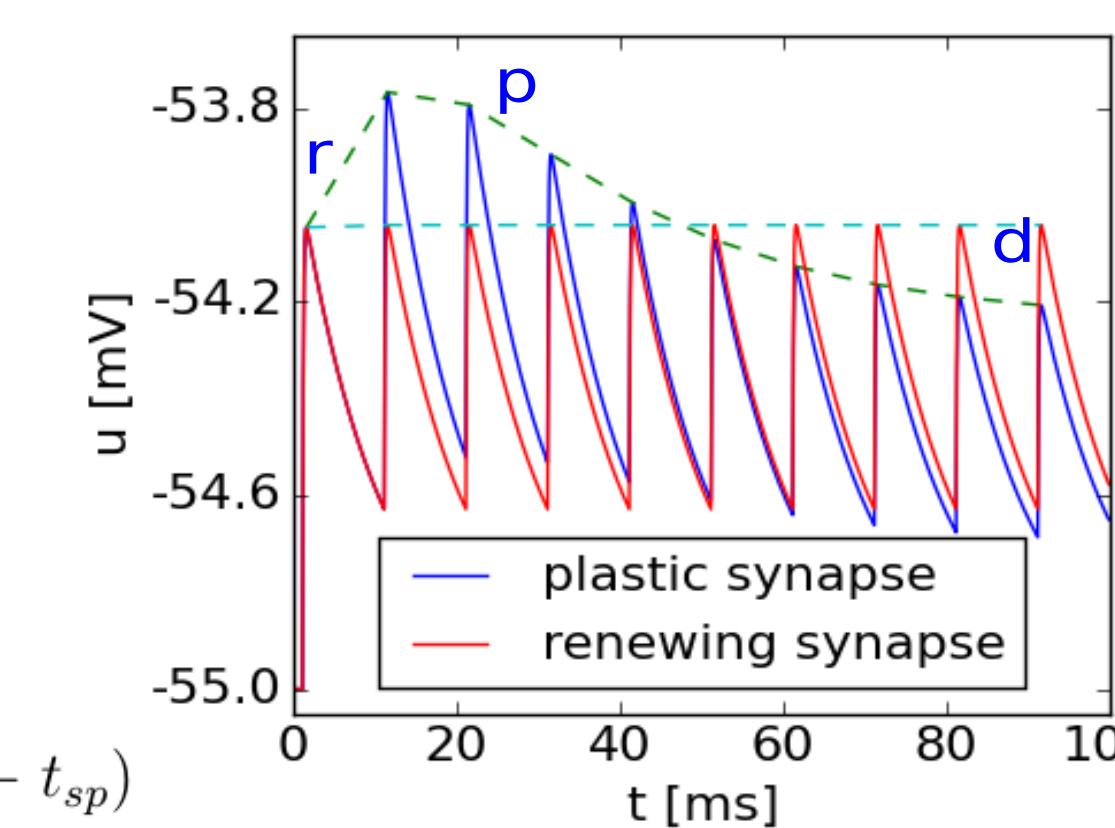
With an appropriate choice of parameters, the expectation value of synaptic weights can be maintained, ensuring that the energy landscape remains, on average, unchanged, while at the same time modulating the strength of the active local attractor of the network. The initial potentiation deepens the local energy well, allowing the visible layer of the network to produce a clearer image. The subsequent depression increases the energy of the network, thereby catapulting it out of the local minimum.

GS · LIF
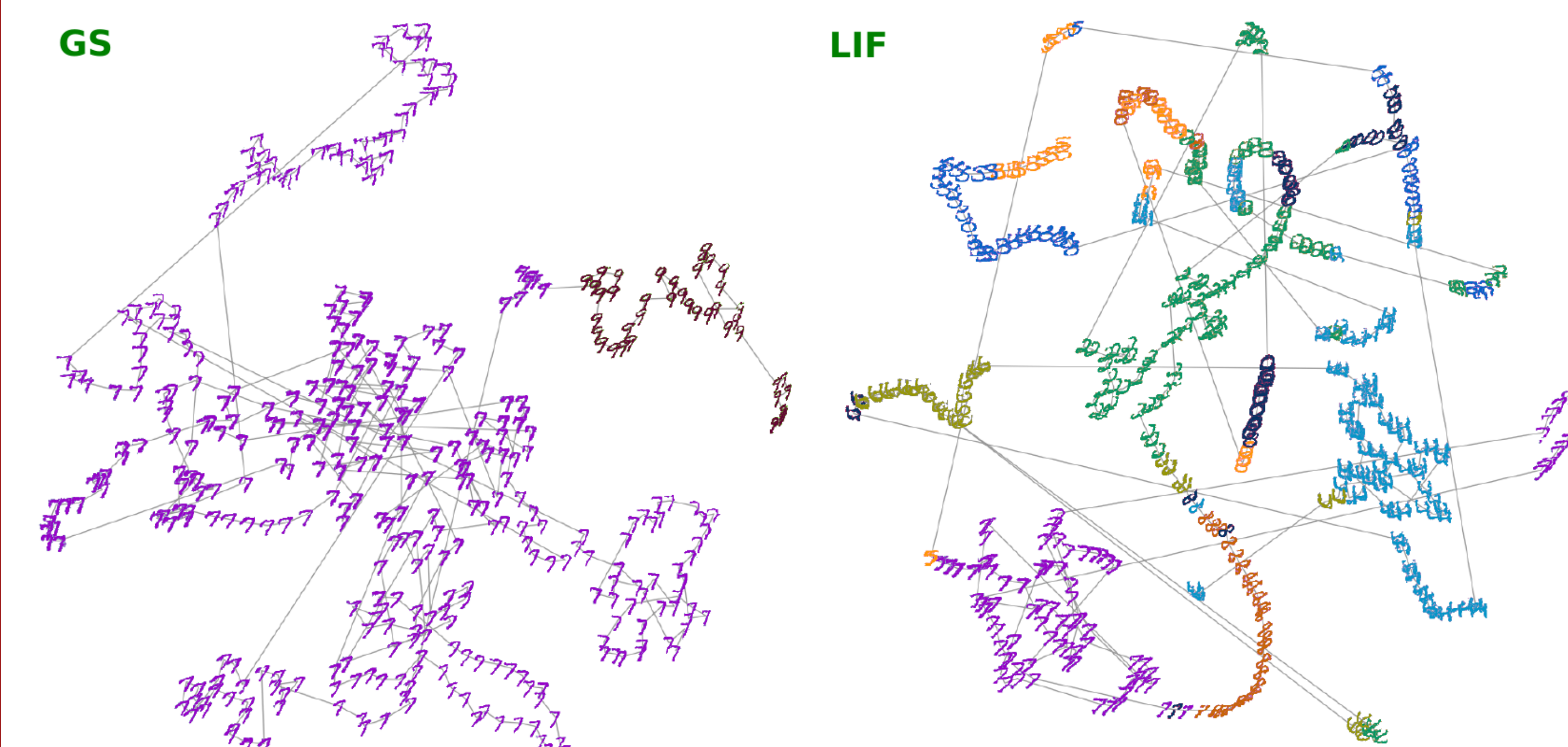
Short-term plasticity enables LIF networks to travel efficiently through the energy landscape and thereby attain a generative performance that significantly surpasses the one achievable by Conventional Gibbs sampling. This distinct feature of biological neural networks allows them to Simultaneously become good generative and discriminative models of learned data. Our finding points towards an important advantage of spike-based computation and communication, which is Relevant in any scenario where spiking neural networks need to be able to escape local attractors.

[1] 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 Computational Biology, 7(11). 2011.
[2] M. A. Petrovici, J. Bill, I. Bytschok, J. Schemmel, and K. Meier, Stochastic inference with deterministic spiking neurons. arXiv preprint arXiv:1311.3211, 2013.
[3] M. A. Petrovici, Bytschok, I., Bill, J., Schemmel, J. & Meier, K. The high-conductance state enables neural sampling in networks of LIF neurons. 24th Annual Meeting of the CNs, 2015.
[4] R. Moreno-Bote and N. Parga. Role of synaptic filtering on the firing response of simple model neurons. Physical Review Letters, 92(2):028102, 2004.
[5] N. Brunel and S. Sergi. Firing frequency of leaky integrate-and-fire neurons with synaptic current dynamics. Journal of Theoretical Biology, 195:87–95, 1998.
[6] R. Salakhutdinov, Learning deep Boltzmann machines with adaptive MCMC, in Proceedings of the 27th International Conference on Machine Learning (ICML-10), pp. 943–950, 2010.
[7] Y. LeCun, and C. Cortes, The MNIST database of handwritten digits, 1998.
[8] G. E. Hinton, Training products of experts by minimizing contrastive divergence. Neural Computation, 14 (8), 17711800, 2002.
[9] G. Fuhrmann, I. Segev, H. Markram, and M. Tsodyks. Coding of temporal information by activity-dependent synapses. Journal of Neurophysiology, 87(1): 140–148, 2002.