Spike-based inference with correlated noise

Ilja Bytschok[1], Dominik Dold[1], Johannes Schemmel[1], Karlheinz Meier[1], Mihai A. Petrovici[2]

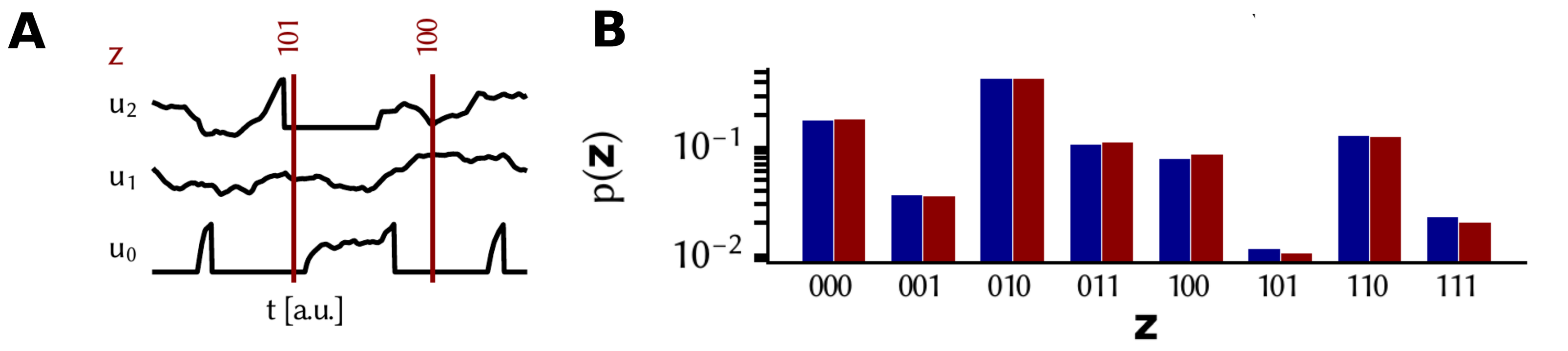
[1]Heidelberg University, Kirchhoff-Institute for Physics; [2]University of Bern, Department of Physiology

The Manfred Stärk Foundation

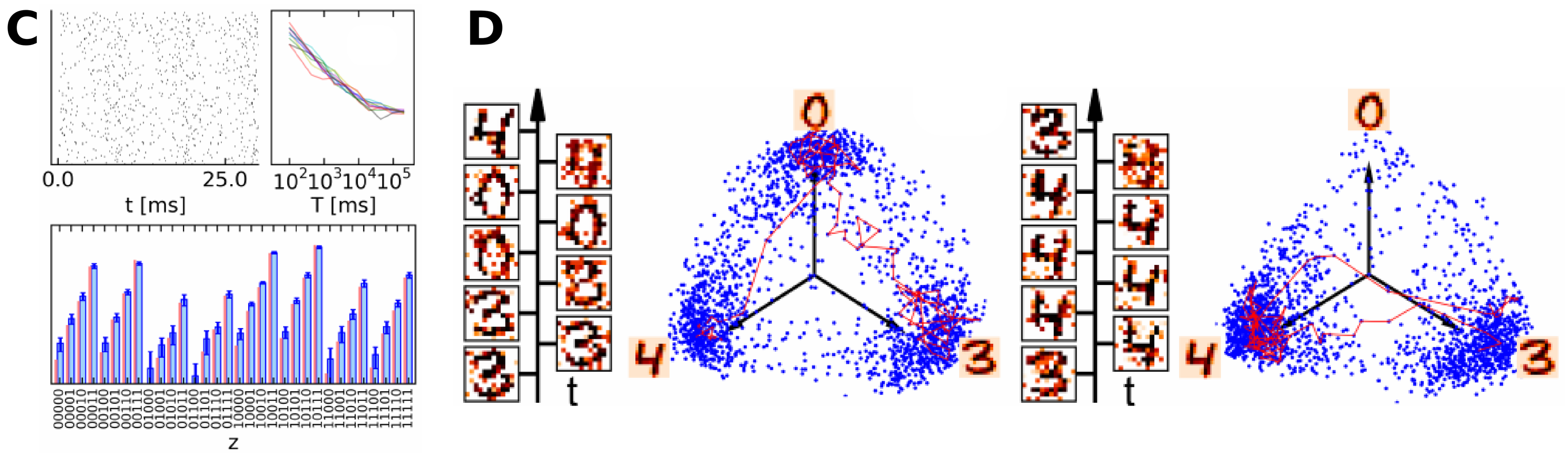
Probabilistic inference in spiking neural networks

Increasing evidence suggests that trial-to-trial variability in neural activity patterns is not merely a byproduct of computation, but rather a hallmark of ongoing probabilistic inference to interpret and respond to sensory input [1,2,3].

The **neural sampling hypothesis** [4] interprets neural states z_i of neurons u_i as binary random variables $z \in \{0,1\}$ and the network firing activity as sampling from underlying probability distributions (Fig. A).



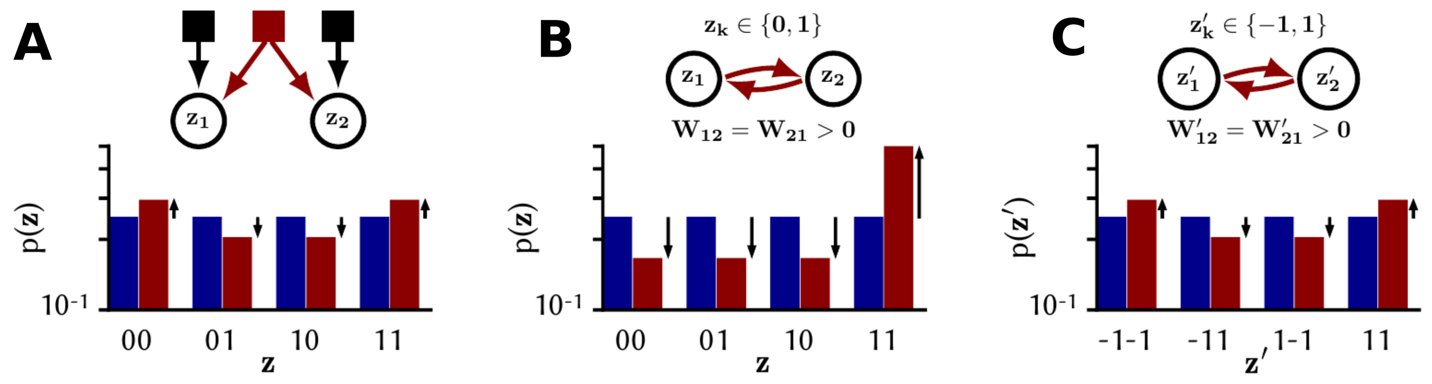
To induce stochastic firing activity in **Leaky Integrate-and-Fire** networks, every neuron is elevated into a **high-conductance state** by high-frequency Poisson input. With an appropriate choice of network parameters, this makes it possible to sample from probability distributions that are either specified explicitly (Fig. B, C) or learned from data (Fig. D) [5, 6, 7].



Impact of shared noise

In cortical networks and on neuromorphic devices, neurons may share significant portions of presynaptic inputs with other neurons, introducing additional correlations.

Example: Two connected LIF neurons with increased probabilities of synchronous states {00} and {11} share noise (Fig. A). The underlying distribution $p_{\mathrm{shared}}(z)$ differs from the one resulting from additional synaptic connections $W_{12} = W_{21}$ (Fig. B). Hence, this effect cannot be compensated only by modifications of the synaptic weight matrix.



To recover sampling from the original distribution, we first transform the state space to $z' \in \{-1,1\}$ without shared noise and find that p(z') is still a Boltzmann distribution by means of a bijective mapping:

$$\{W, b\},\ z \in \{0,1\} \rightarrow \{W', b'\},\ z' \in \{-1,1\}$$

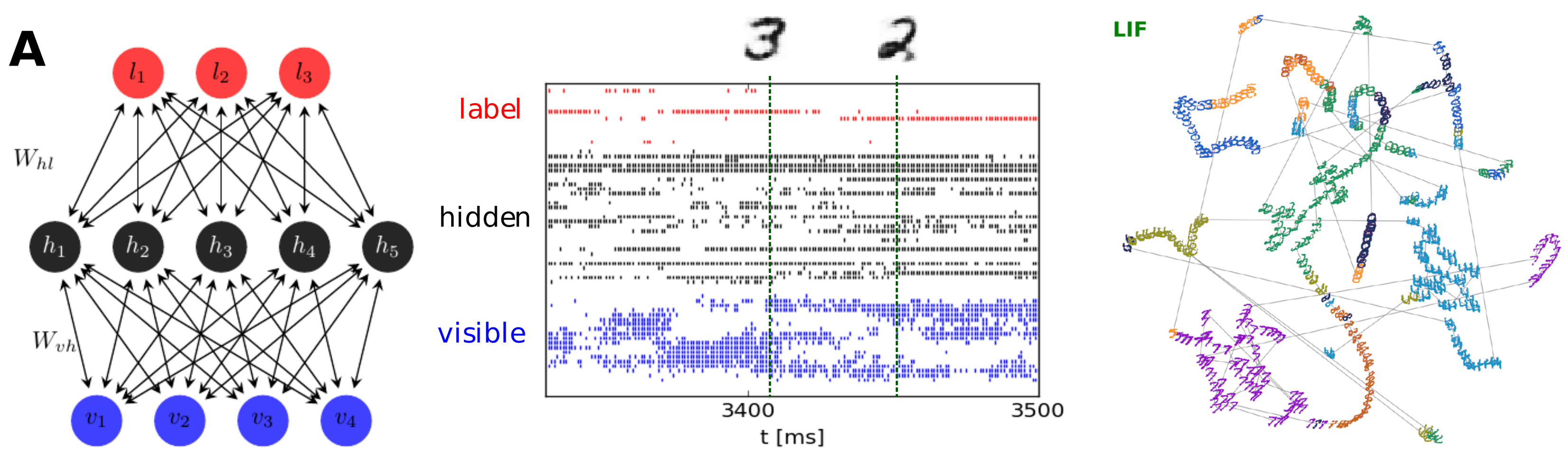
with translation rules

$$W = 4W',\ b = 2b' + 2W',\ z = (z' + \frac{1}{2})$$

The underlying distribution remains in the class of Boltzmann distributions and can be compensated by appropriate changes of the network parameters. In the new domain (Fig. C), we see that synaptic weights W' can be set to produce the same effect on the distribution as in the shared noise case (Fig. A).

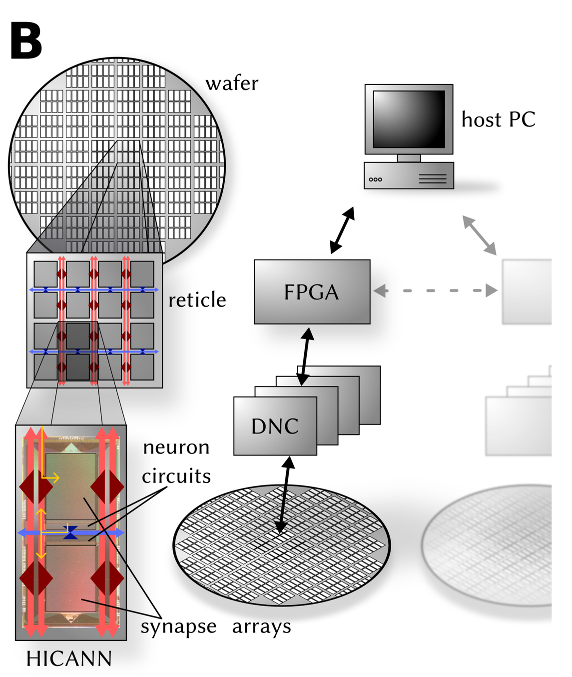
Stochasticity in neuromorphic hardware

It has been shown that layered sampling LIF networks can be trained to generate and classify data sets utilizing diffuse noise (Fig. A). Short-term plasticity enables such spiking networks to simultaneously exhibit good mixing properties and classification rates (Fig. C) [6].



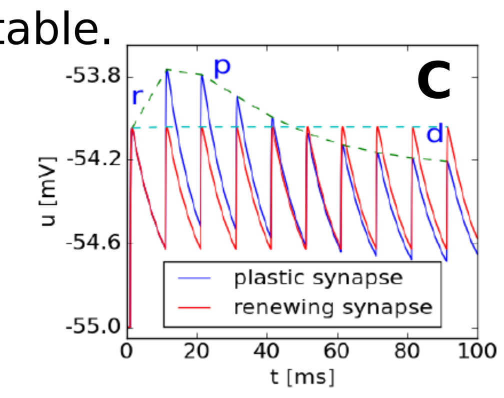
Mixed-signal neuromorphic devices are ideally suited for a physical implementation of such networks. In these systems, neurons and synapses are implemented in silico as analog circuits, while the spike transmission is handled digitally [8] (Fig. B).

Biology-inspired plasticity mechanisms (Fig. C) and the advantage of a $\mathbf{10^4}$ **runtime speed-up** compared to biological networks allows fast inference and learning compared to conventional simulation setups.



Since controllable sources of randomness are difficult to implement in hardware, these devices usually receive noisy stimuli from external sources. Due to **limited on-chip bandwidth**, shared stochastic input among hardware neurons is inevitable.

The resulting shared noise-induced correlations impair the network-sampled distributions.



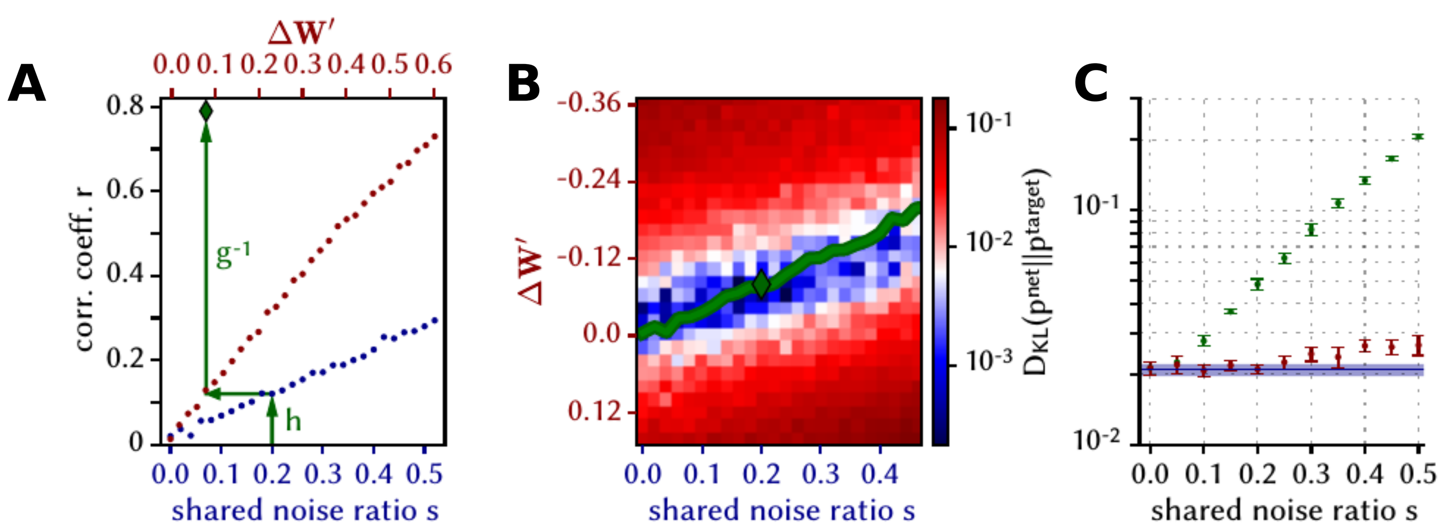
To ensure the convergence towards the desired distribution, we have developed a method to compensate for the harmful effects of shared inputs.

Compensation of correlations

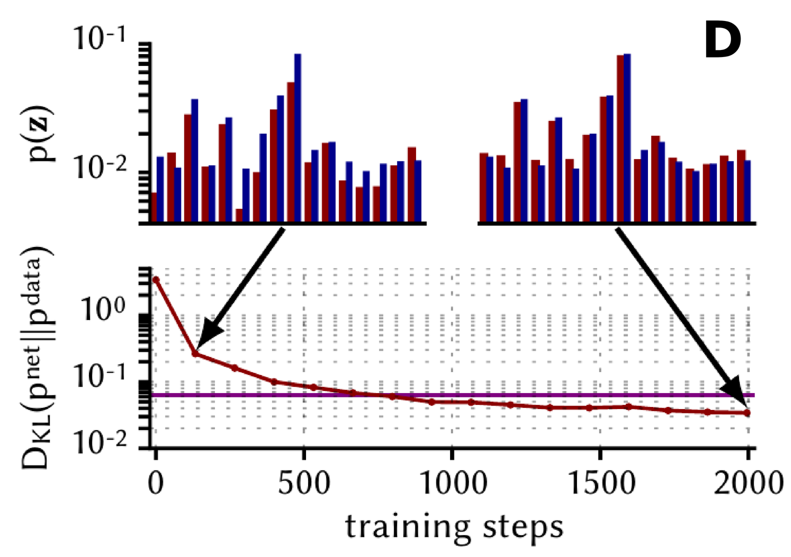
Since shared noise and synaptic weights in the {-1,1} state space produce the same distribution, we can change the weight by $\Delta W'$ to match the correlation coefficient r of a joint firing pattern with shared noise ratio s (Fig. A). We find a **bijective mapping** between the two parameters, $\Delta W' = f(s)$, with

$$f := g^{-1} \circ h,\ h(s) \rightarrow r,\ g(\Delta W') \rightarrow r$$

The $\Delta W'$ represents the sought compensation of the shared noise. We find suitable parameters over a broad range of s (Fig. B), resulting in a good performance even for large portions of shared noise and larger network sizes (Fig. C).



The mapping allows to interpret the presence of shared noise in {0,1} as a parameter shift in the {-1,1} domain, which can be trained from data. This causes differences between the network distribution (Fig. D, red bars) to the desired distribution (blue bars). We **train the network** to match the target distribution and measure the error (Fig. D, red curve).



After training, $p_{\mathrm{shared}}^{\mathrm{train}}(z)$ matches the desired distribution. By further compensating parameter mismatch [8], this additional training phase allows a transfer of network implementations.

This work has received funding from the European Union 7th Framework Programme under grant agreement 604102 (HBP), the Horizon 2020 Framework Programme under grant agreement 720270 (HBP) and the Manfred Stärk Foundation.

Further information: https://arxiv.org/abs/1707.01746

References:

[1] Konrad Körding and Daniel Wolpert. Bayesian integration in sensorimotor learning, Nature, 2004.
[2] József Fizser, Pietro Berkes, Gergő Orbán and Máté Lengyel. Statistically optimal perception and learning: from behavior to neural representations, Trends in Cognitive Sciences, 2010.
[3] Erin L. Rich and Jonathan D. Wallis. Decoding subjective decisions from orbitofrontal cortex, Nature Neuroscience, 2016.
[4] Lars Buesing, Johannes Bill, Bernhard Nessler and Wolfgang Maass. Neural dynamics as sampling: a model for stochastic computation in recurrent networks of spiking neurons, PLoS Comput Biol, 2011
[5] Mihai A. Petrovici, Johannes Bill, Ilja Bytschok, Johannes Schemmel and Karlheinz Meier. Stochastic inference with spiking neurons in the high-conductance state, Physical Review E, 2016.
[6] D. Probst, M. A. Petrovici, I. Bytschok, J. Bill, D. Pecevski, J. Schemmel and K. Meier. Probabilistic inference in discrete spaces can be implemented into networks of LIF neurons, Front. Comput. Neurosci, 2015
[7] L. Leng*, M. Petrovici*, R. Martel, I. Bytschok, O. Breitwieser, J. Bill, J. Schemmel, K. Meier. Spiking neural networks as superior generative and discriminative models, Cosyne 2016
[8] Sebastian Schmitt, Johann Klaehn et al. Neuromorphic Hardware In The Loop: Training a Deep Spiking Network on the BrainScaleS Wafer-Scale System, IEEE International Joint Conference on Neural Networks, 2017