Stochastic inference with spiking neural networks

Mihai Petrovici*, Luziwei Leng*, Oliver Breitwieser*, David Stöckel*,
Ilja Bytschok, Roman Martel, Johannes Bill, Johannes Schemmel, Karlheinz Meier
Kirchhoff Institute for Physics, University of Heidelberg, Germany

17.02.2016

Brains are adept at creating an impressively accurate internal model of their surrounding based on incomplete and noisy sensory data. Understanding this inferential prowess is not only interesting for neuroscience, but may also inspire computational architectures and algorithms for solving hard inference problems. Here, we give an overview of our work on probabilistic inference with brain-inspired spiking networks, their advantages compared to classical neural networks and their implementation in neuromorphic hardware.

In the neural sampling framework, we interpret spiking activity as sampling from distributions over binary random variables. By exploiting the dynamics of spiking neurons with conductance-based synapses, we have shown that their activation function can become symmetric in the high-conductance state, which in turn enables Glauber-like dynamics in ensembles of noise-driven LIF networks (Petrovici et al., 2013, 2015). This allows the straightforward construction of LIF networks that sample from previously defined probability distributions.

When the parameters of the distribution are not well-defined, they need to be learned from data. Due to their analogy to classical neural networks such as Boltzmann machines, LIF networks are amenable to the same learning algorithms and can be shown to match the performance of their equally-sized abstract counterparts when trained on classical machine-learning datasets such as MNIST. However, spiking neural networks endowed with short-term plasticity can travel more efficiently through their associated state space, allowing them to simultaneously become good generative and discriminative models of learned data, which is notoriously difficult with conventional techniques such as Gibbs sampling. This finding points towards a distinct 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.

This computational advantage of spiking sampling networks can be further bolstered by emulation on an accelerated neuromorphic substrate. The core idea behind these devices is the direct emulation of biological neuronal dynamics in VLSI circuits. Such hardware can far surpass simulators running on conventional computing architectures both in terms of speed and power consumption, but with the caveat of having limited parameter precision, as well as other sources of disruptive noise (Petrovici et al., 2014; Pfeil et al., 2013). With some additional modifications, we have shown how LIF networks can become robust to certain types of parameter noise – both during training and during operation – thereby making them amenable to a neuromorphic implementation with an acceleration factor of 10^4 compared to biological real-time.

An even more compelling argument for neuromorphic spike-based inference can be made when considering that learning (in particular, the simulation of synaptic plasticity) is by far the most time-consuming factor in simulations. In an effort to make expectation-maximization learning (Nessler et al., 2013) compatible with existing neuromorphic devices, we have developed a network model that can use double-exponential STDP with 4-6 bit weight resolution for learning and spike-based homeostasis (Habenschuss et al., 2012) for stabilization and robustness.

* Authors with equal contributions. This research was supported by EU grants #269921 (BrainScaleS), #237955 (FACETS-ITN), #604102 (Human Brain Project) and the Manfred Stärk Foundation.

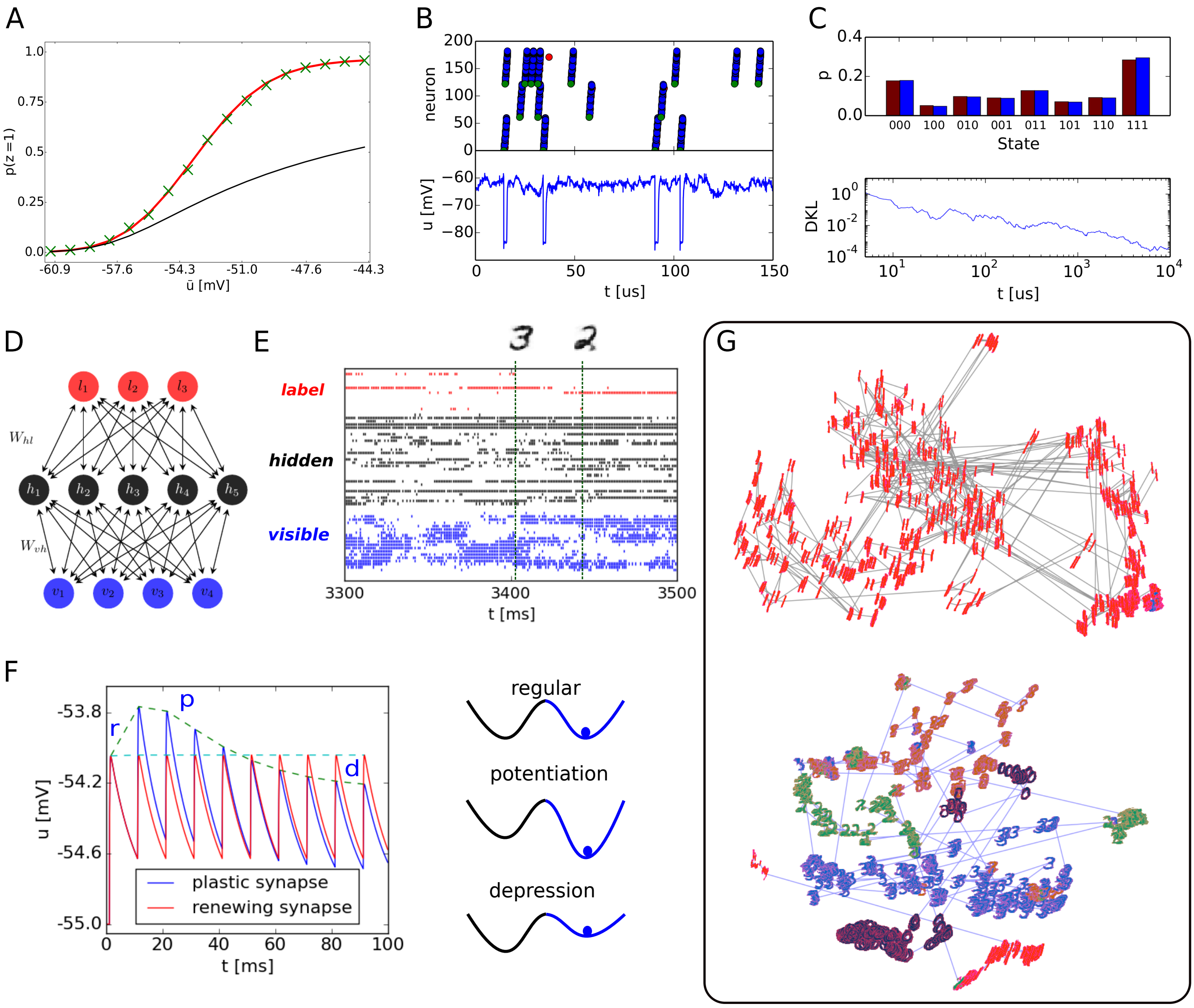

Figure 1: **(A)** Activation function of an LIF neuron for two different background input regimes. The symmetric activation function obtained in the high-conductance state enables neural sampling. **(B)** For a robust implementation on a neuromorphic device, individual neurons are replaced by small populations called "sampling units". The figure shows a spike raster plot and an exemplary membrane potential of a neuron in a sampling network with 3 sampling units. **(C)** Top: comparison of the probability distribution sampled by the neuromorphic chip (red) and target probability distribution (blue) after a wall-clock runtime of 10 ms, which corresponds to a biological runtime of 100 s. Bottom: convergence of the sampled distribution towards the target distribution. **(D)** Hierarchical LIF network structure. After training on the full MNIST dataset, its classification performance (95,1%) closely matched the performance of the equivalent restricted Boltzmann machine (95,2%). **(E)** Selected spike trains from the LIF network. Note the increasing sparseness of the activity in consecutive layers. The network produces different images when the activity in the label layer switches on from one neuron to another. **(F)** Short-term plasticity as modeled by the Tsodyks-Markram mechanism and its effect on the local energy landscape: an initial deepening of the energy trough and sharpening the produced image is followed by a local flattening of the energy landscape which pushes the network state into a different mode. **(G)** Comparison between a sequence of images generated by a conventional RBM with Gibbs sampling (top) and one generated by the LIF network with short-term synaptic plasticity (bottom).

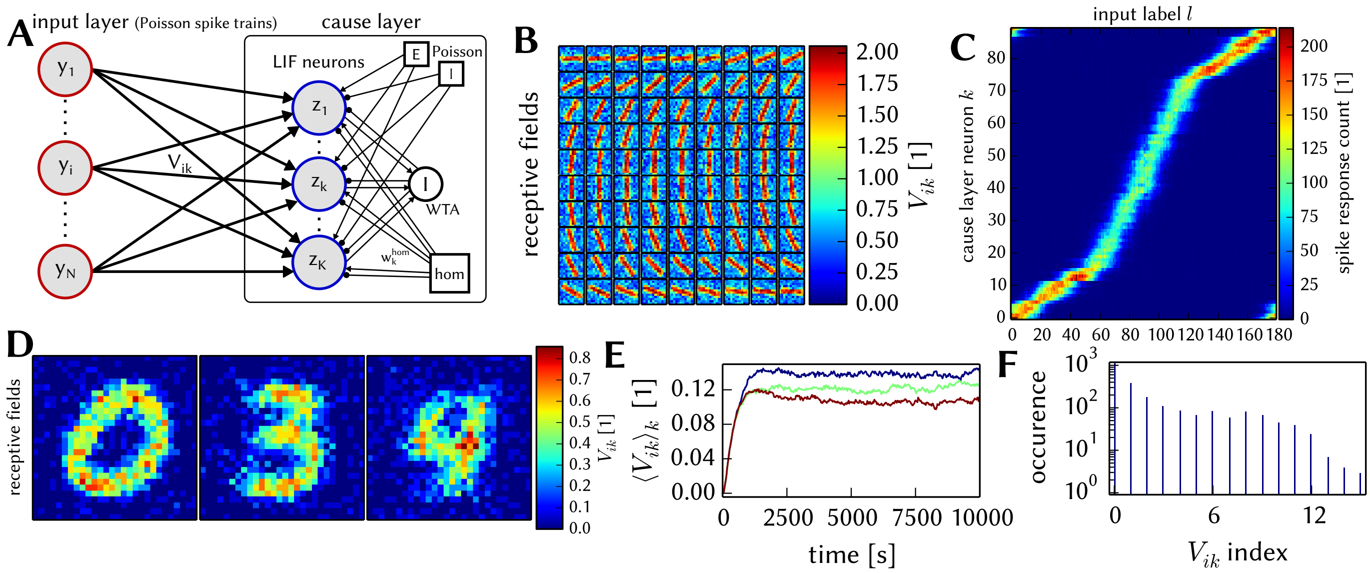

Figure 2: **(A)** Architecture of a stochastic network with homeostatic unsupervised learning. The cause layer consisting of stochastic LIF neurons in the high-conductance state forms a WTA circuit. Common input motifs in these images are learned by modifying the input weights V_{ik} through a double exponential STDP rule. Spike-based homeostasis is implemented by periodic background sources whose efferent weights are modulated by the activity of the cause layer neurons. (**B**) Trained receptive fields of 90 cause layer neurons when presented with images of rotated bars. (**C**) Spike response of the network after learning (the input label l corresponds to the angle of rotation). As orientations 45° to 135° are presented twice as often during learning, more cause layer neurons tend to code for them. (**D**) Receptive fields of a network consisting of three cause layer neurons after learning from a subset of the MNIST dataset. (**E**) Average input weights modulated by homeostasis during learning. (**H**) Distribution of weights after learning on a simulated neuromorphic device with 4-bit synapses. The 16 possible weight values can be set up to cover the required weight range.

References

S. Habenschuss, J. Bill, and B. Nessler. Homeostatic plasticity in bayesian spiking networks as expectation maximization with posterior constraints. *Advances in Neural Information Processing Systems*, 25, 2012.

B. Nessler, M. Pfeiffer, L. Buesing, and W. Maass. Bayesian computation emerges in generic cortical microcircuits through spike-timing-dependent plasticity. *PLoS Computational Biology*, 9(4):e1003037, 2013.

M. A. Petrovici, J. Bill, I. Bytschok, J. Schemmel, and K. Meier. Stochastic inference with deterministic spiking neurons. *arXiv preprint arXiv:1311.3211*, 2013.

M. A. Petrovici, B. Vogginger, P. Müller, O. Breitwieser, M. Lundqvist, L. Muller, M. Ehrlich, A. Destexhe, A. Lansner, R. Schüffny, et al. Characterization and compensation of network-level anomalies in mixed-signal neuromorphic modeling platforms. *PloS one*, 9(10):e108590, 2014.

M. A. Petrovici, I. Bytschok, J. Bill, J. Schemmel, and K. Meier. The high-conductance state enables neural sampling in networks of lif neurons. *BMC Neuroscience*, 16(Suppl 1):O2, 2015.

T. Pfeil, A. Grübl, S. Jeltsch, E. Müller, P. Müller, M. A. Petrovici, M. Schmuker, D. Brüderle, J. Schemmel, and K. Meier. Six networks on a universal neuromorphic computing substrate. *Frontiers in Neuroscience*, 7:11, 2013. ISSN 1662-453X. doi: 10.3389/fnins.2013.00011.