Stochastic inference

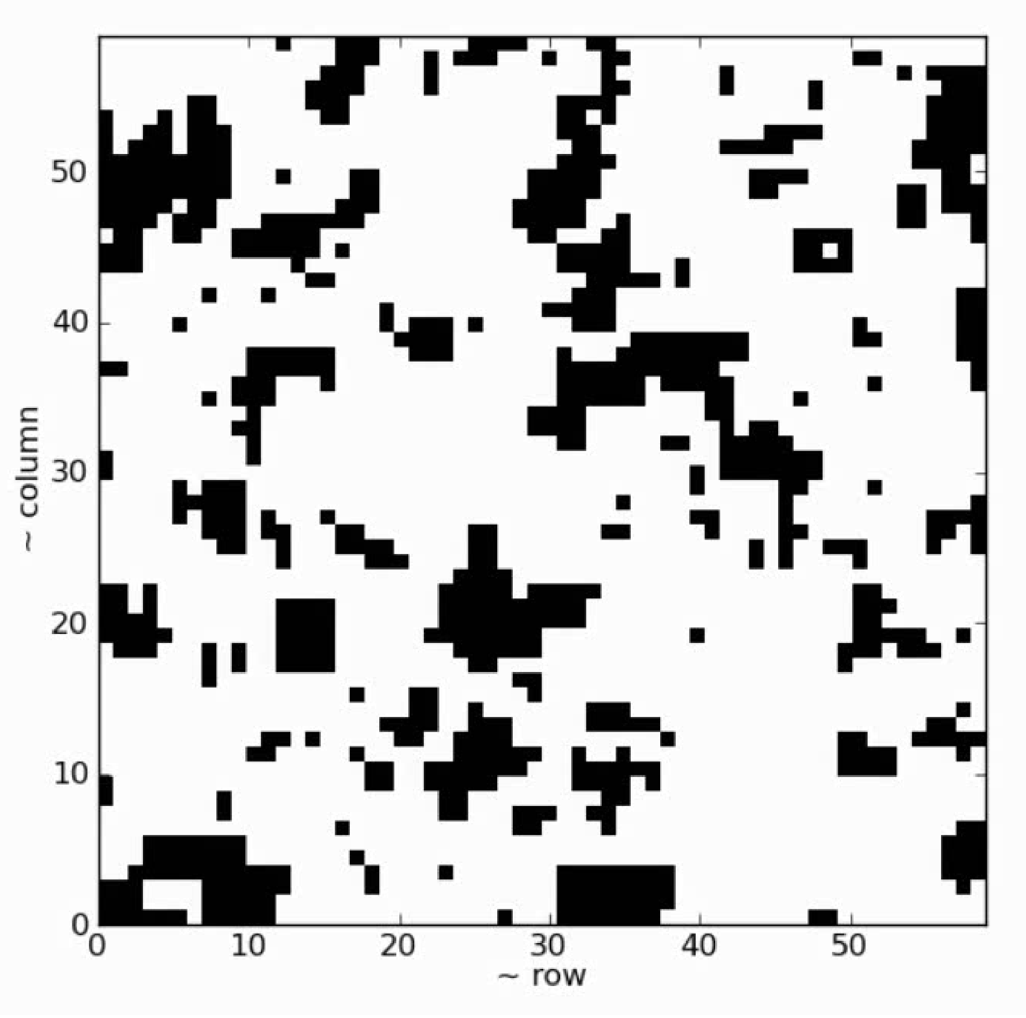
with deterministic

spiking neurons

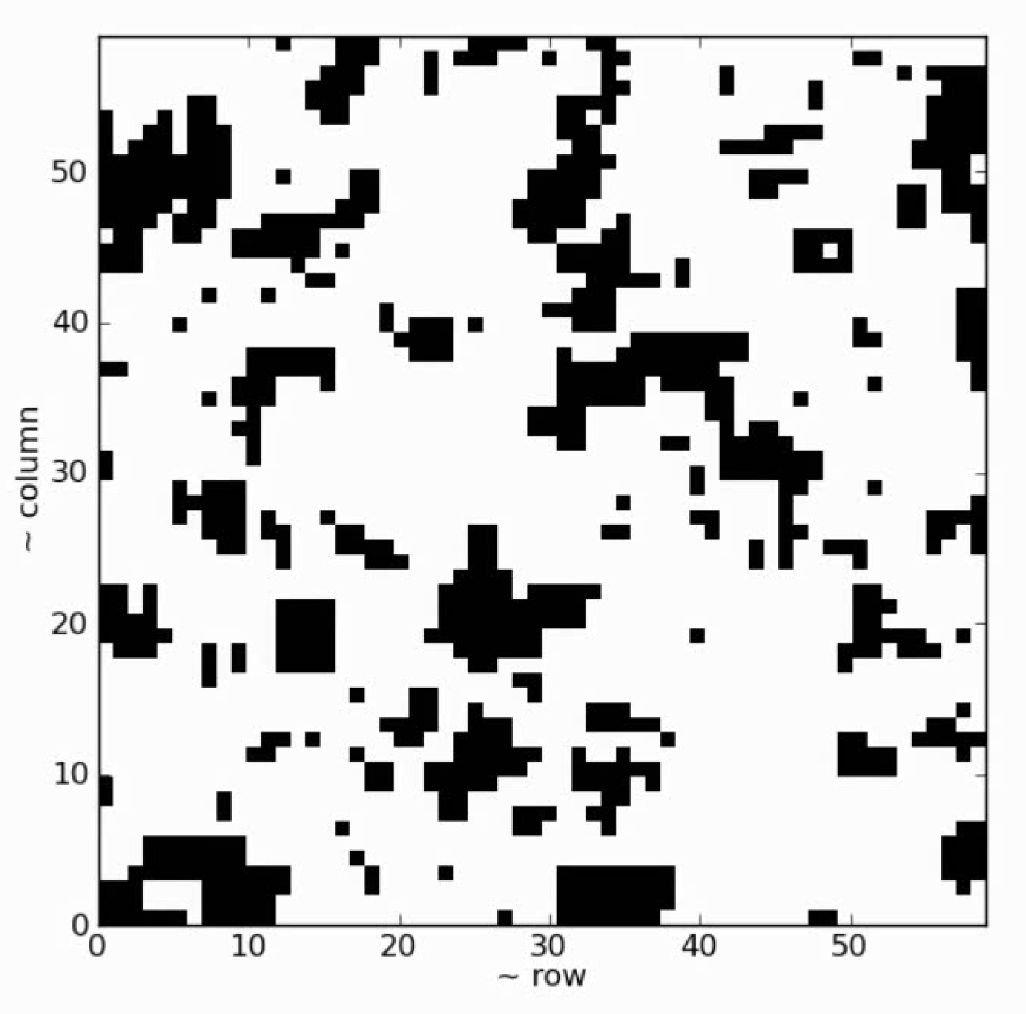
Mihai A. Petrovici
Luziwei Leng, David Stöckel, Oliver Breitwieser, Ilja Bytschok, Jakob Jordan
Roman Martel, Johannes Bill, Johannes Schemmel, Karlheinz Meier

Electronic Vision(s)
Kirchhoff Institute for Physics
University of Heidelberg

Ensemble dynamics



Ensemble dynamics



Ising model

$$H(\boldsymbol{\sigma}) = -\sum_{ij} J_{ij}\sigma_i\sigma_j - \sum_i b_i\sigma_i$$

$$p_\beta(\boldsymbol{\sigma}) = \frac{e^{-H(\boldsymbol{\sigma})/k_B T}}{Z_\beta}$$

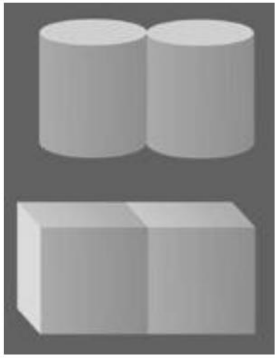
Glauber dynamics

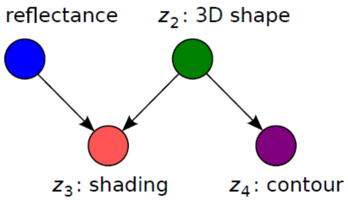
$\rightarrow$ sampling !

Bayesian inference & learning

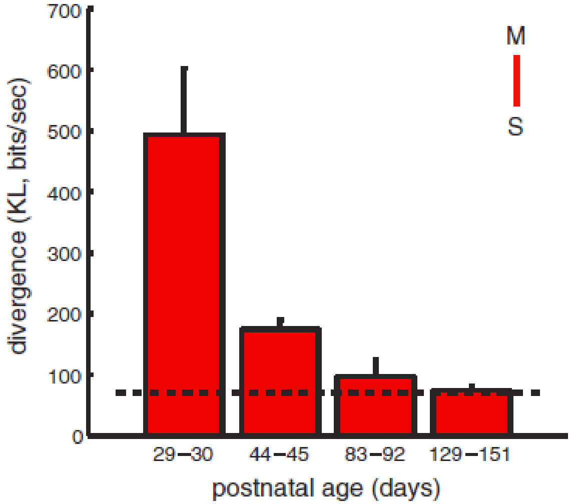
decision problem $\leftrightarrow$ inference problem

p (reality | observations)





Knill & Kersten (1991)

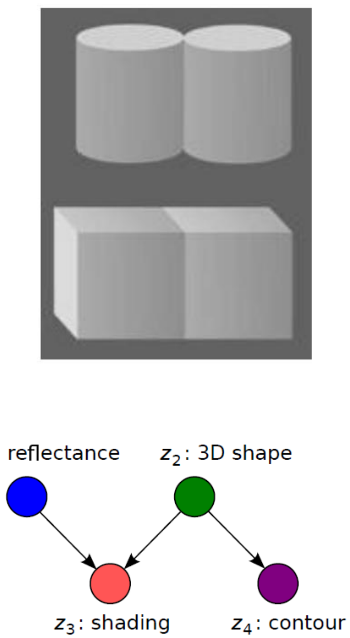


Berkes et al. (2011)

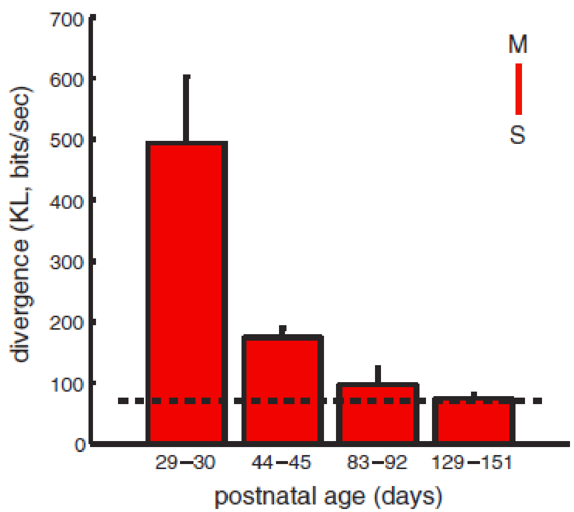
Bayesian inference & learning

decision problem ↔ inference problem

p (reality | observations)



Knill & Kersten (1991)



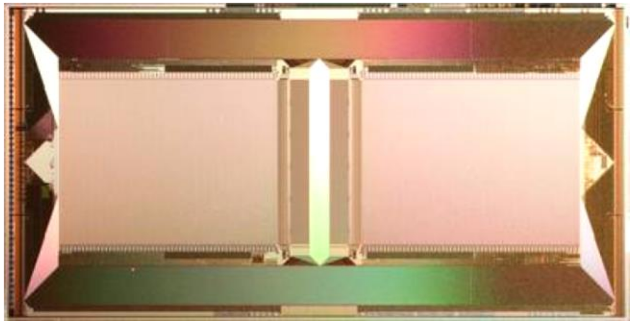
Berkes et al. (2011)

Matthew et al. (ca. 85)

Why spikes ?

understand biology

exploit efficient hardware

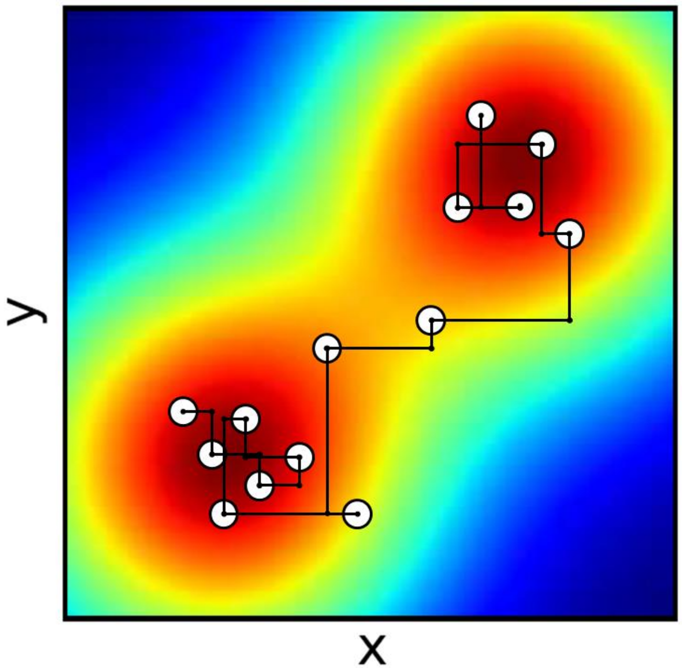


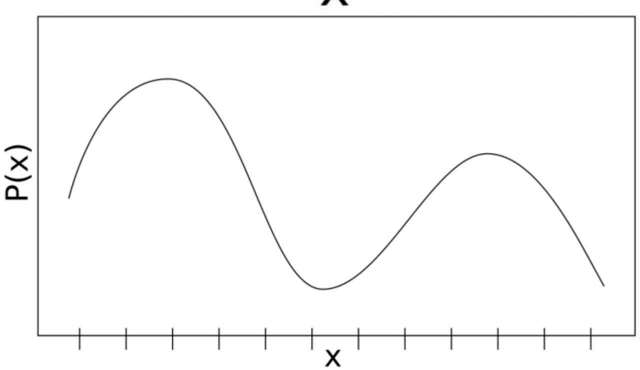
low power, high speed, ...

... and maybe also gain a computational advantage

Neural sampling

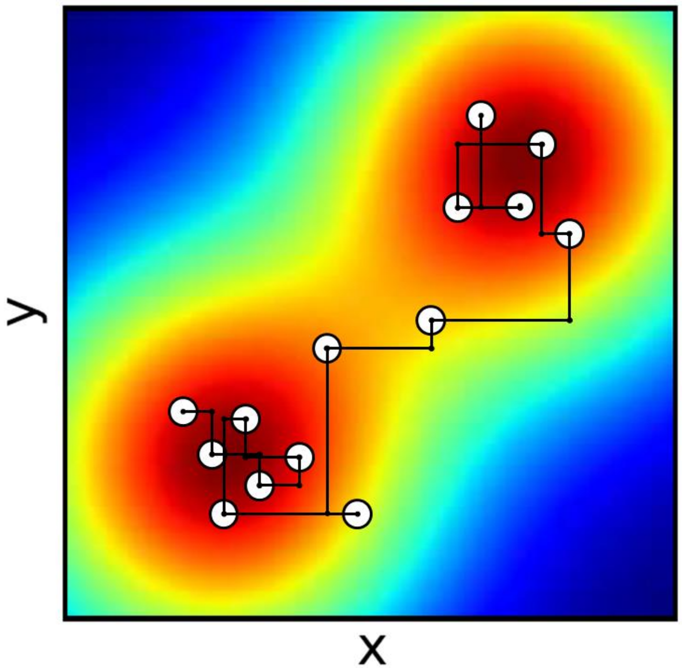
(Gibbs) sampling

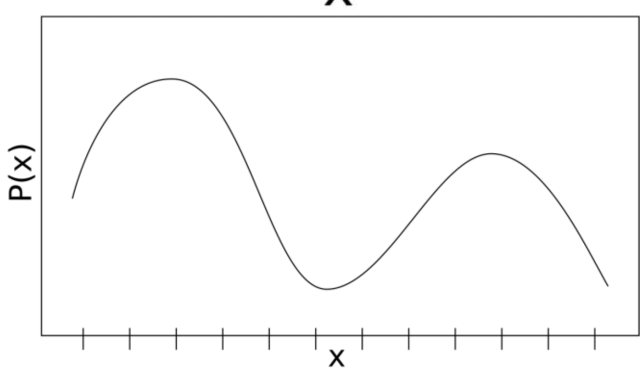




Neural sampling

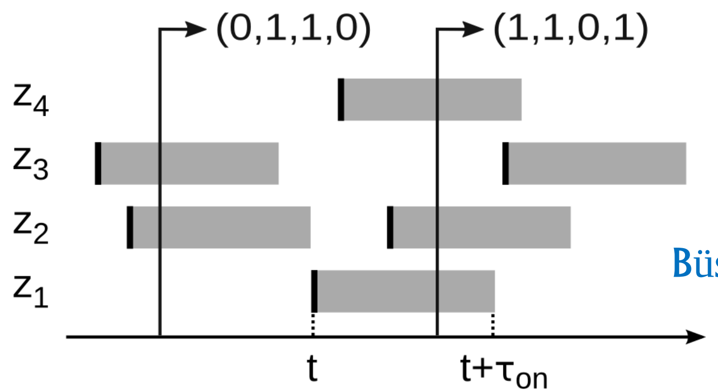
(Gibbs) sampling





assumptions: $z_k \in \{0,1\}$

$z_k = 1 \;\Leftrightarrow\;$ neuron has spiked in $[t-\tau, t)$



Büsing et al. (2011)

information transfer between neurons:

$$u_k = \log \frac{p(z_k = 1 \,|\mathbf{z}_{\setminus k})}{p(z_k = 0 \,|\mathbf{z}_{\setminus k})}$$

$$\Leftrightarrow \qquad \nu_k = p(z_k = 1) = \frac{1}{1+\exp(-u_k)}$$

response/activation function

The LIF activation function

noise source: Poisson spike trains $\Rightarrow$ membrane as Ornstein-Uhlenbeck process

$$du(t) = \Theta \cdot [\mu - u(t)]dt + \sigma\, dW(t)$$

Brunel & Sergi (1998)

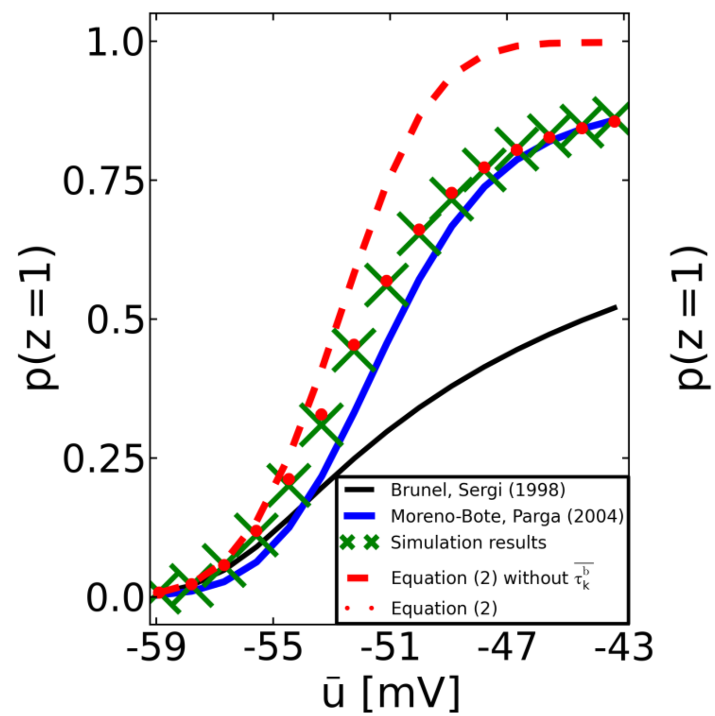
assumption: $\tau_{\mathrm{syn}} \ll \tau_{\mathrm{m}}$

$$\langle T \rangle = \tau\sqrt{\pi} \int_{\rho - \mu/\sigma}^{\vartheta_{\mathrm{eff}} - \mu/\sigma} [1 + \operatorname{erf} x] \exp(x^2)\, dx$$

Moreno-Bote & Parga (2004)

assumption: $\tau_{\mathrm{ref}} \approx \tau_{\mathrm{syn}} \gg \tau_{\mathrm{m}}$

$$\nu = \int_{\vartheta}^{\infty} \left(\tau_{\mathrm{m}} \frac{\varrho - \tilde{u}}{\vartheta - \tilde{u}} \right)^{-1} p(\tilde{u})\, d\tilde{u}$$



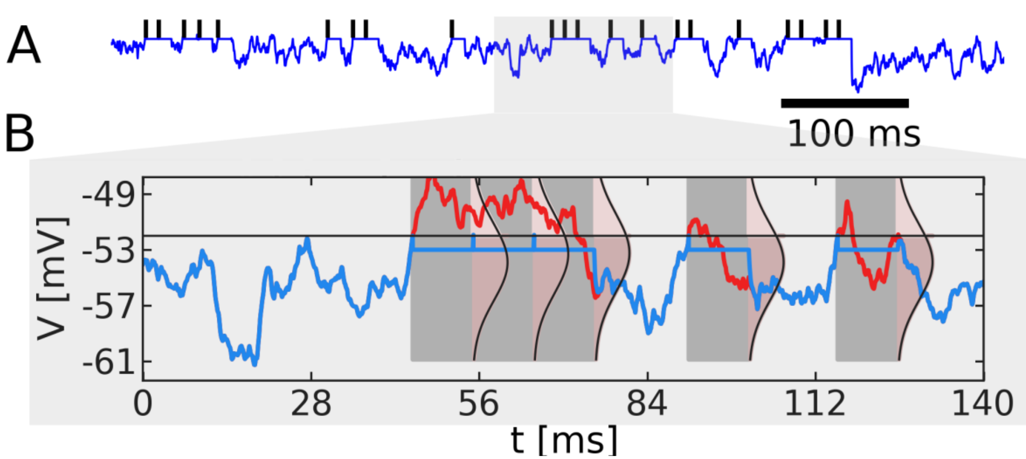
The membrane autocorrelation propagation

$$p(z_k = 1) = \frac{t_{k,\,\text{refractory}}}{t_{\text{total}}} = \frac{\sum_n P_n\, n\, \tau_{\text{ref}}}{\sum_n P_n \cdot \left(n\, \tau_{\text{ref}} + \sum_{k=1}^{n-1} \overline{\tau_k^b} + T_n\right)}$$

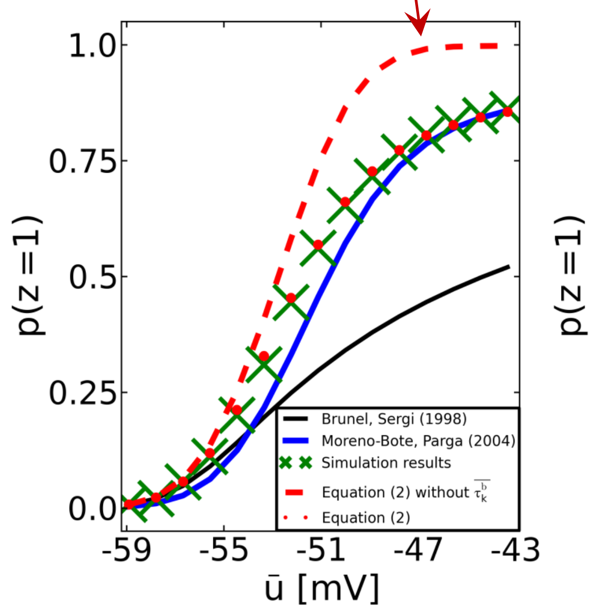
$$P_n = \left(1 - \sum_{i=1}^{n-1} P_i\right) \cdot \int_{V_{\text{thr}}}^{\infty} dV_{n-1}\; p(V_{n-1}|V_{n-1} > V_{\text{thr}}) \left[\int_{-\infty}^{V_{\text{thr}}} dV_n\; p(V_n|V_{n-1})\right]$$

$$T_n = \int_{V_{\text{thr}}}^{\infty} dV_{n-1}\; p(V_{n-1}|V_{n-1} > V_{\text{thr}}) \left[\int_{-\infty}^{V_{\text{thr}}} dV_n\; p(V_n|V_{n-1}) \langle FPT(V_{\text{thr}}, V_n)\rangle\right]$$

$$\overline{\tau_k^b} = \int_{\vartheta}^{\infty} du_k\; \tau_{\text{eff}} \ln\left(\frac{\varrho - u_k}{\vartheta - u_k}\right) p(u_k|u_k > \vartheta, u_{k-1})$$

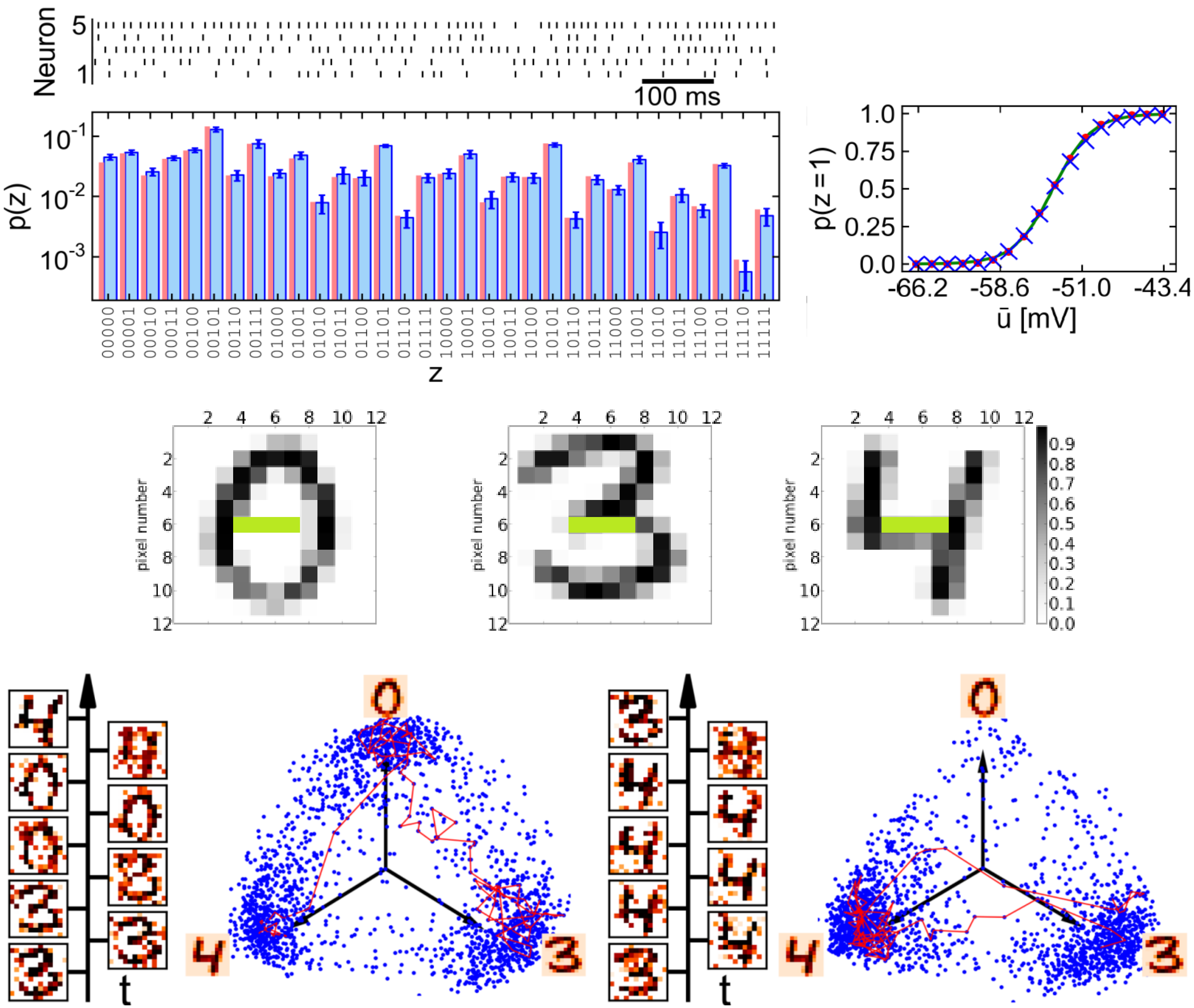


HCS !



Petrovici & Bill et al. (2013)

(Fully visible) LIF-based Boltzmann machines

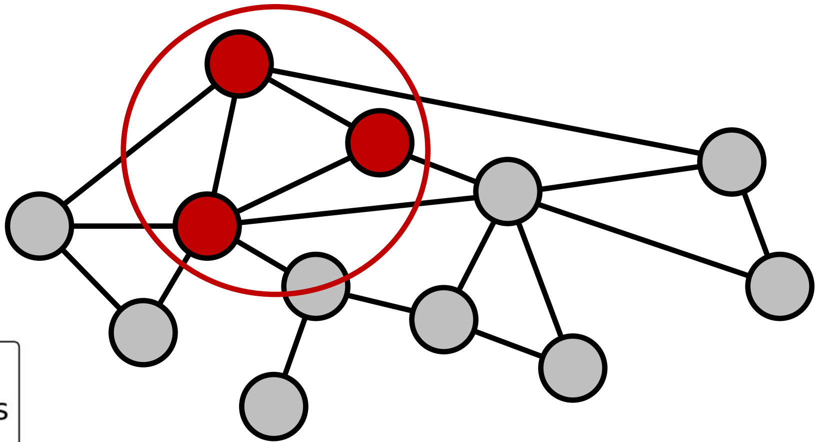


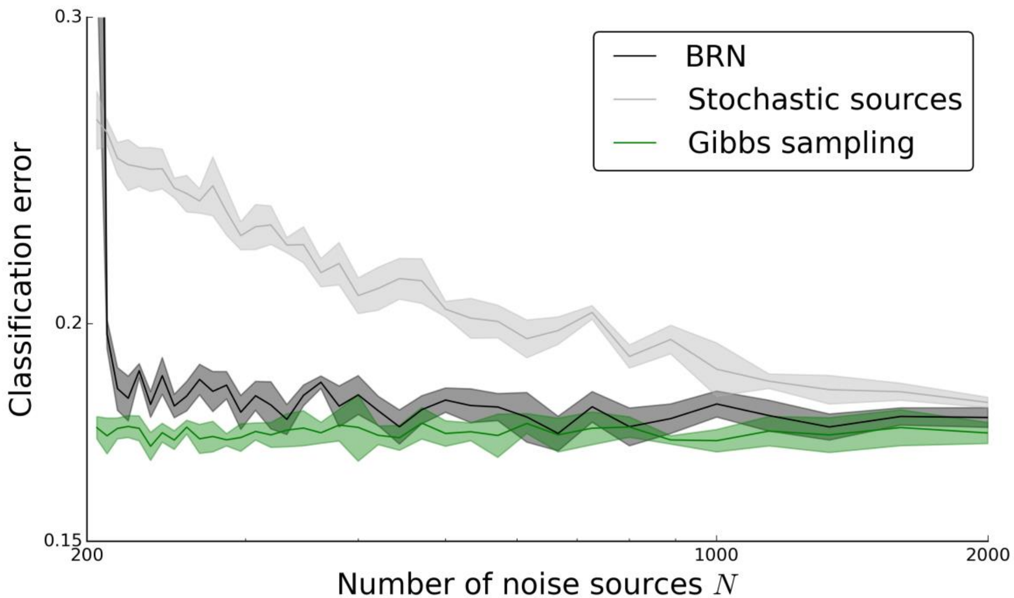
Petrovici & Bill et al. (2013)

… so where does the noise come from ?

- in simulations: from independent Poisson sources
- more realistic: from being embedded in larger networks
- but: noise correlations hurt !

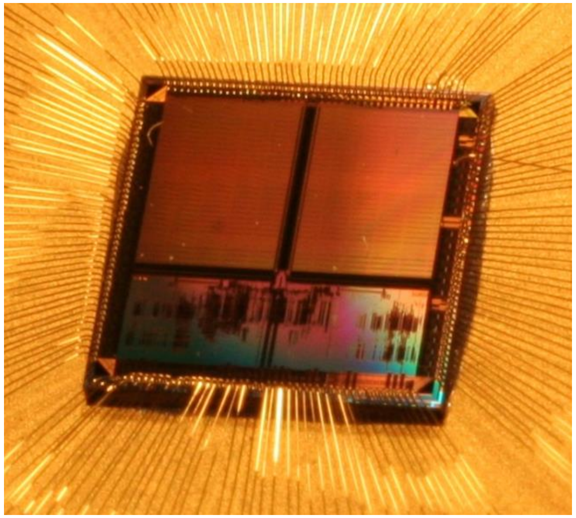
  → decorrelation by lateral inhibition





Jordan et al. (2015)

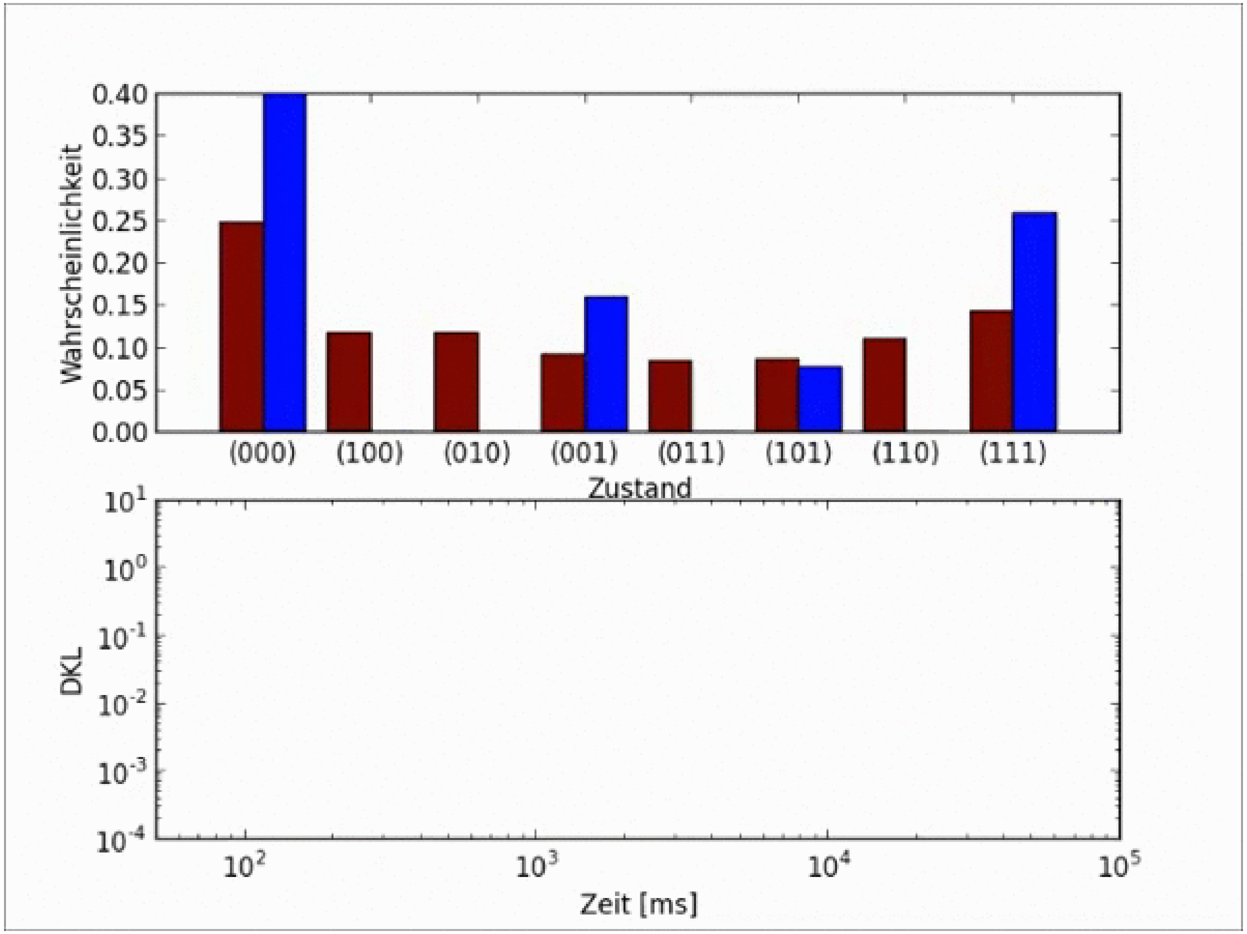
LIF sampling on accelerated neuromorphic hardware



„Spikey“ chip

Schemmel et al. (2006)

Pfeil et al. (2013)

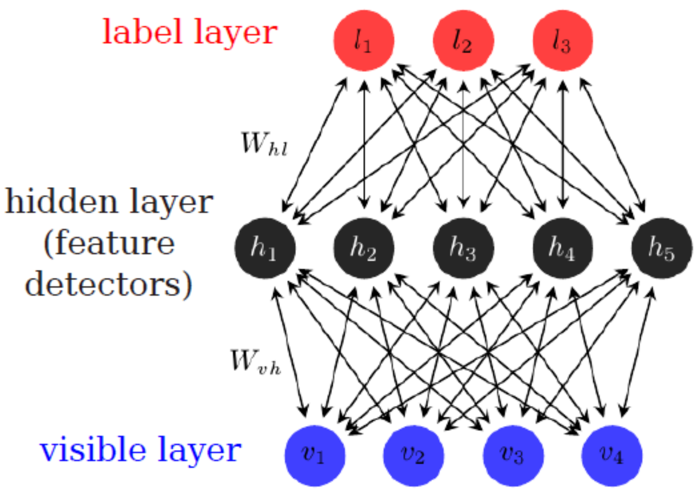


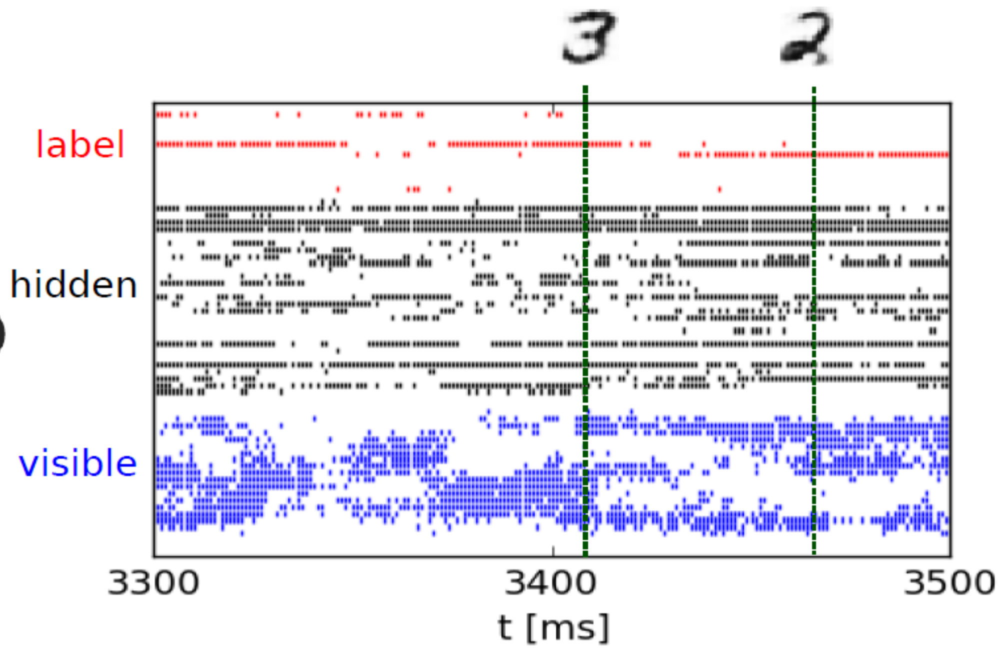
bio: 100 s → software simulation: 10 s

bio: 100 s → neuromorphic emulation: 10 ms

Petrovici & Stöckel et al. (2015)

Deep learning architectures



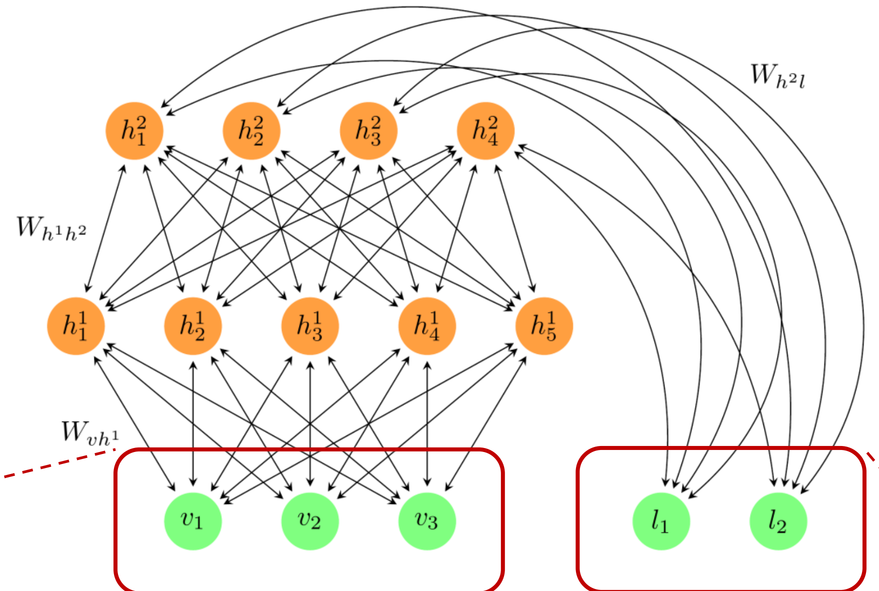


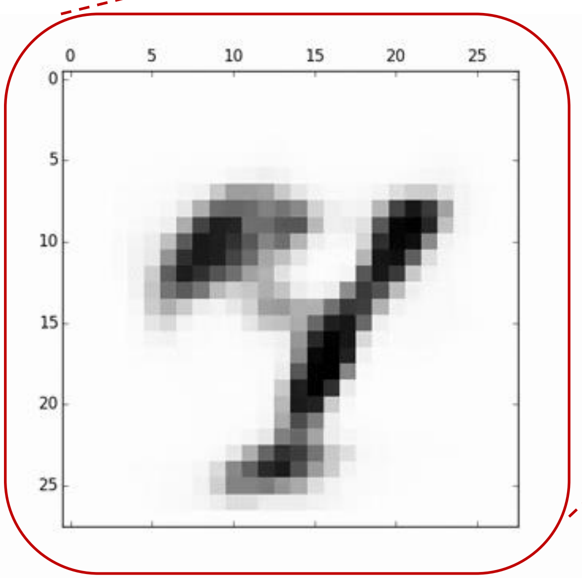
Leng & Petrovici et al. (2016)

Deep learning architectures

Training of RBMs/DBMs on MNIST:

- contrastive divergence
- coupled adaptive tempering

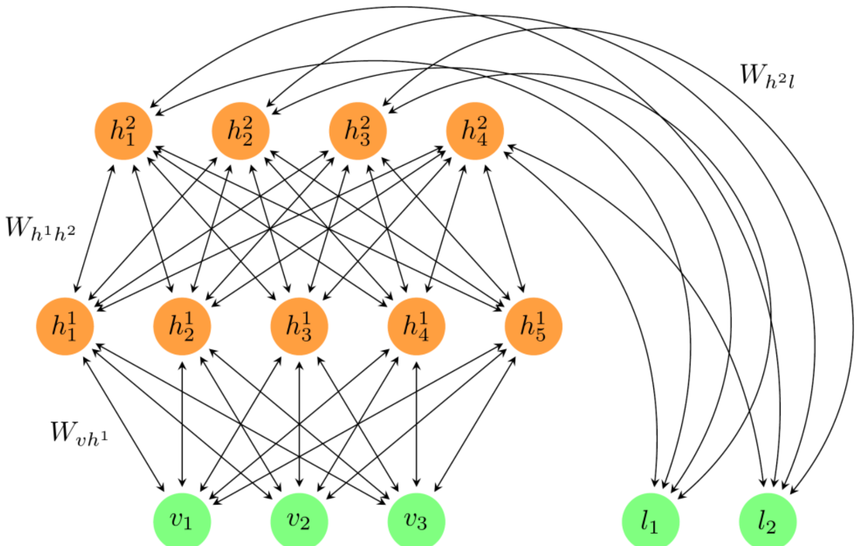


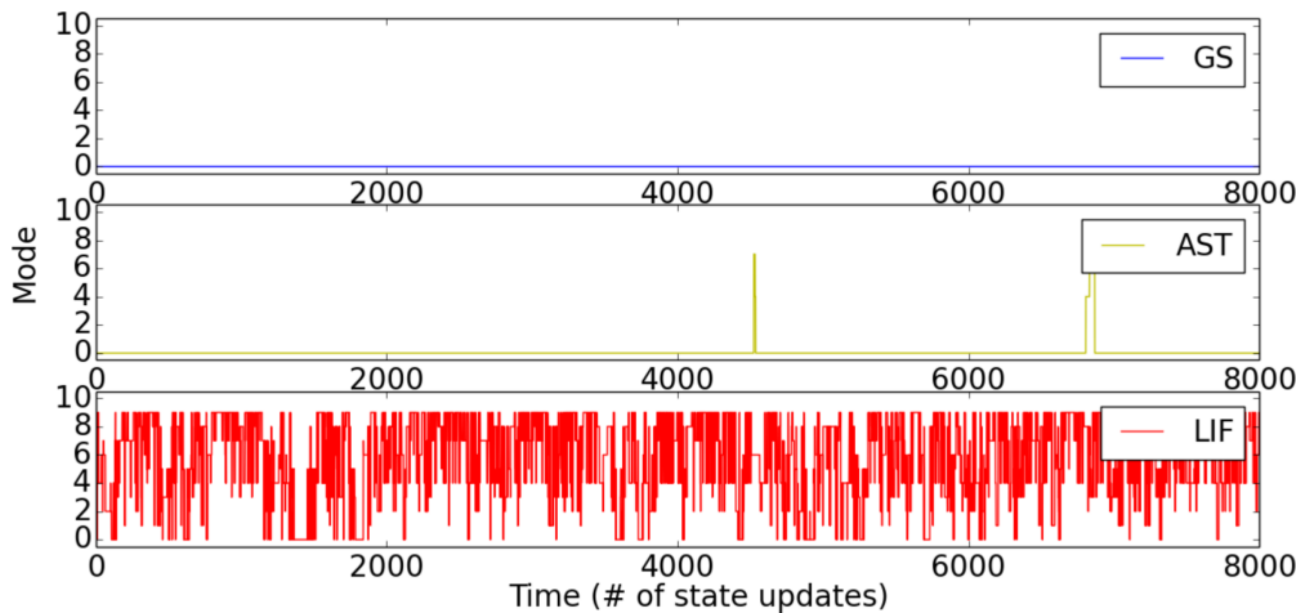


96.9 % correct classification

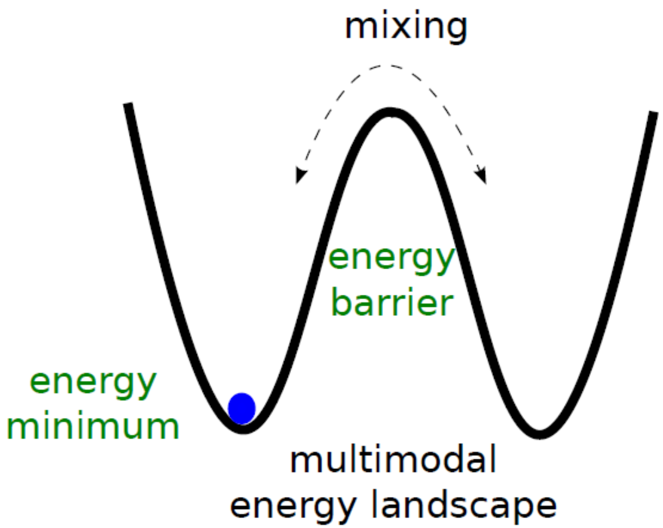
Leng & Petrovici et al. (2016)

Mixing in deep spiking networks

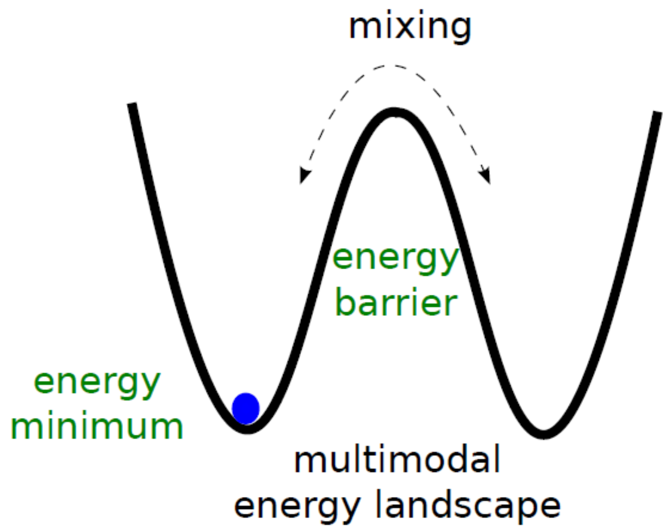




Mixing in deep spiking networks



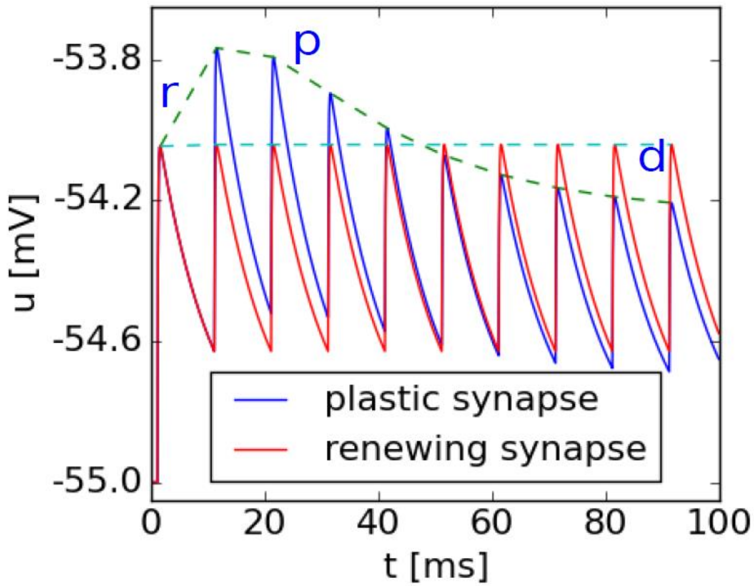
Mixing in deep spiking networks

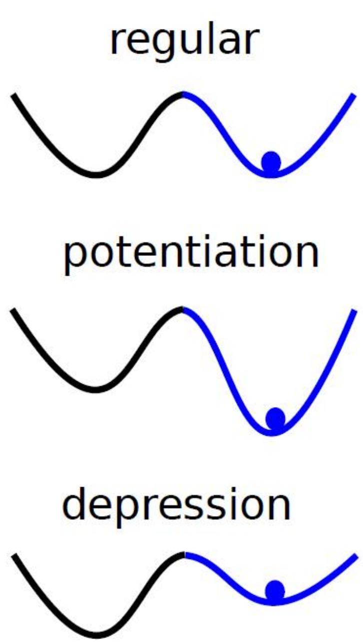


Tsodyks-Markram short-term plasticity

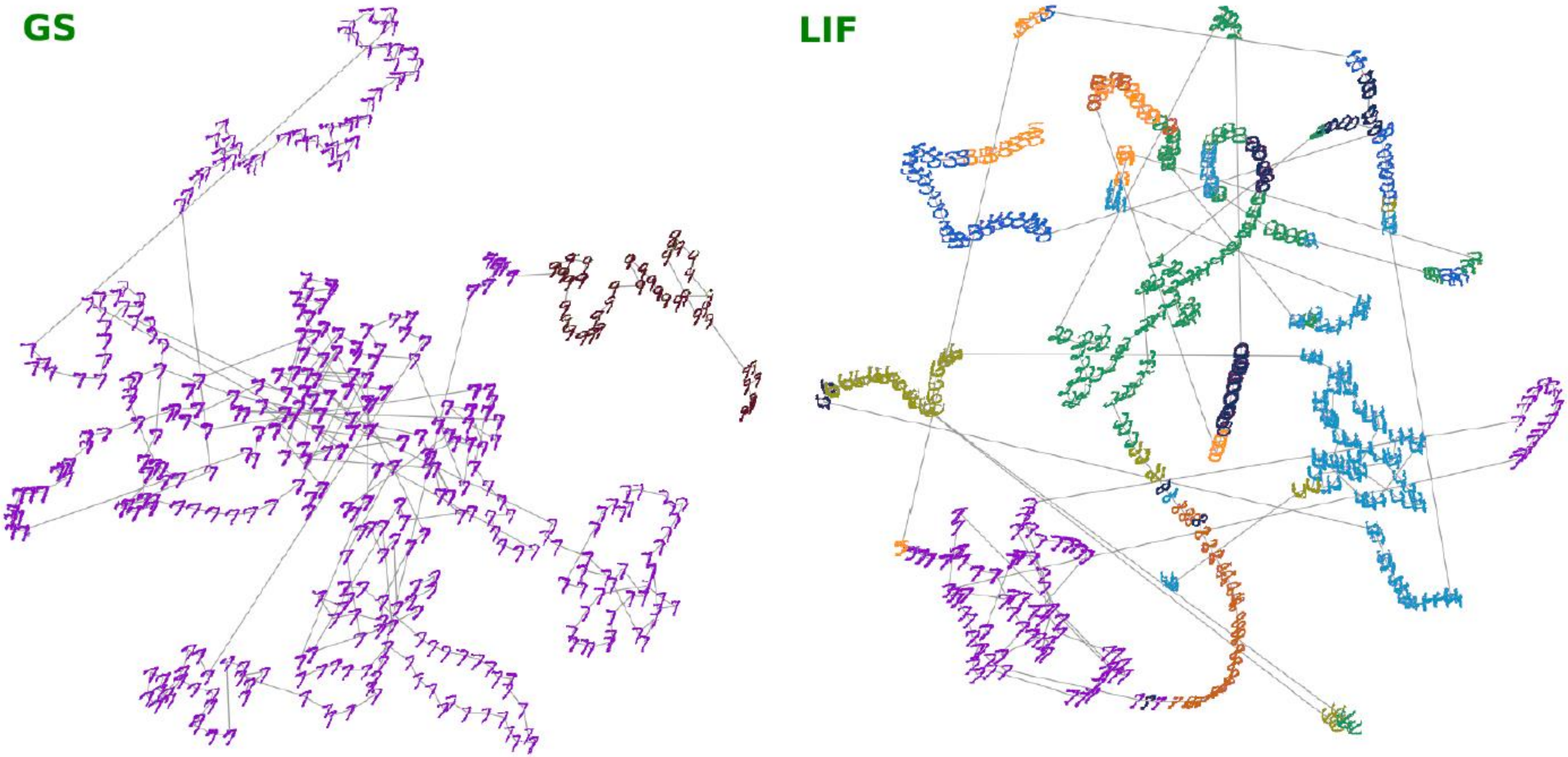
$$\frac{dR}{dt} = \frac{1-R}{\tau_{rec}} - U_{SE}R\delta(t-t_{sp})$$

$$\frac{dU_{SE}}{dt} = -\frac{U_{SE}}{\tau_{facil}} + U_0(1-U_{SE})\delta(t-t_{sp})$$





Spiking networks as superior generative and discriminative models



Leng & Petrovici et al. (2016)

Summary

BIOLOGY

- functional role of HCS
- functional role of STP
- uncorrelated noise is surprisingly easy to generate

HARDWARE

- design of a robust spike-based sampling model
- extremely fast convergence to target distribution

THEORY & MODELS

- model of LIF sampling
- analytical description of LIF activation function
- deep spiking networks as simultaneous discriminative and generative models