Inference and learning
at the level of single synapses and spiking neurons

PART II

Jean-Pascal Pfister

ETH Zürich

University of Zurich UZH

Roscoff 26.6.2017

Outlook of part II

1. Learning with the GLM
   (Brea at al. J. Neurosc. 2013)

2. Decoding at single synapses
   (Pfister et al. Nat. Neurosc. 2010)

3. Decoding with the GLM
   (with A. Kutschireiter)

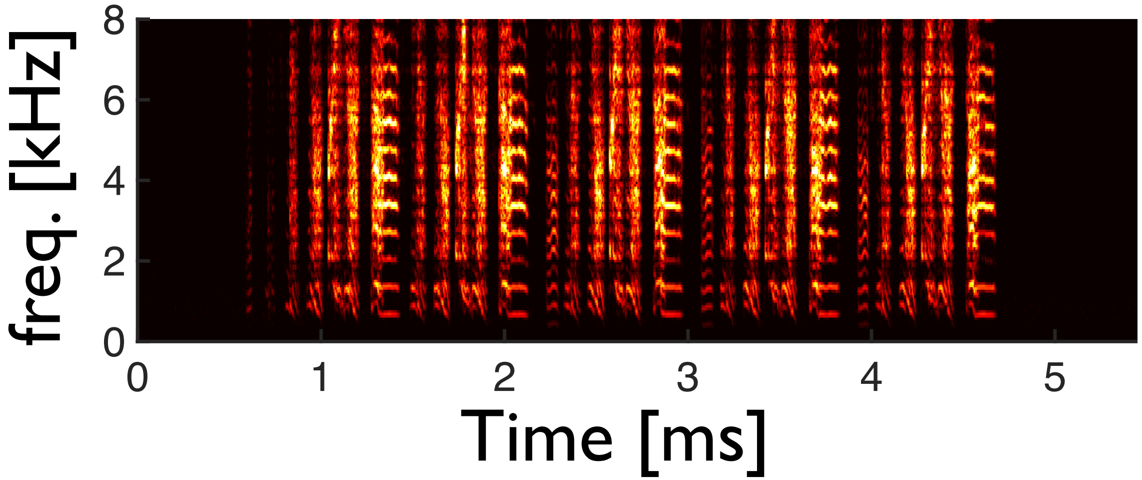
Matching Recall and Storage in Sequence Learning with Spiking Neural Networks. Brea, Senn and Pfister J. Neurosc. 2013

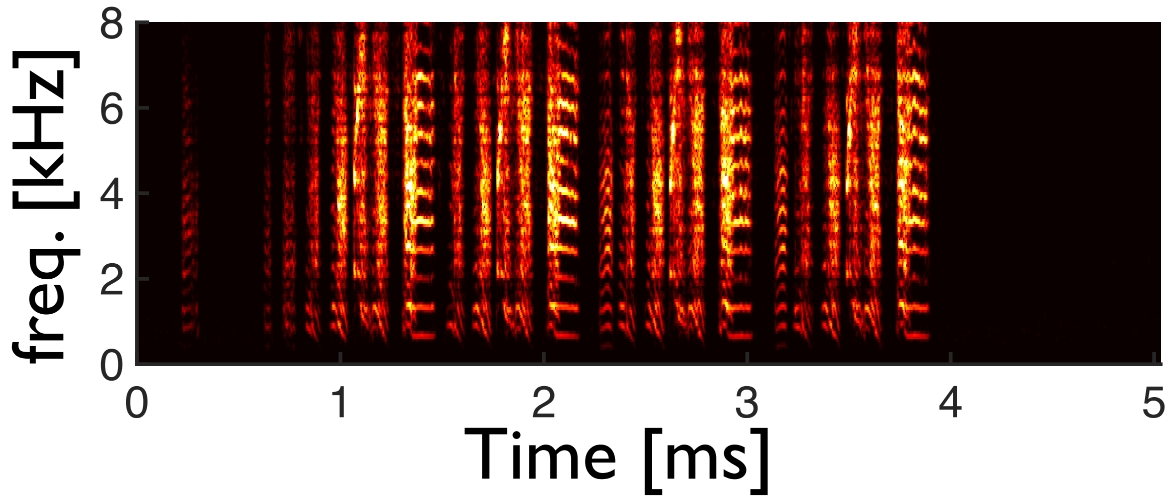
- Motivation for modelling sequence learning
- Definition of the problem
- Network model
- Learning rule
  (with only visible units)
- Learning rule
  (with additional hidden units)

Johanni Brea W. Senn

Roscoff 26.6.2017

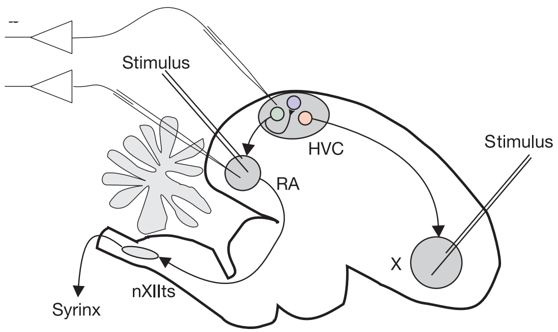
Neural sequences in songbird

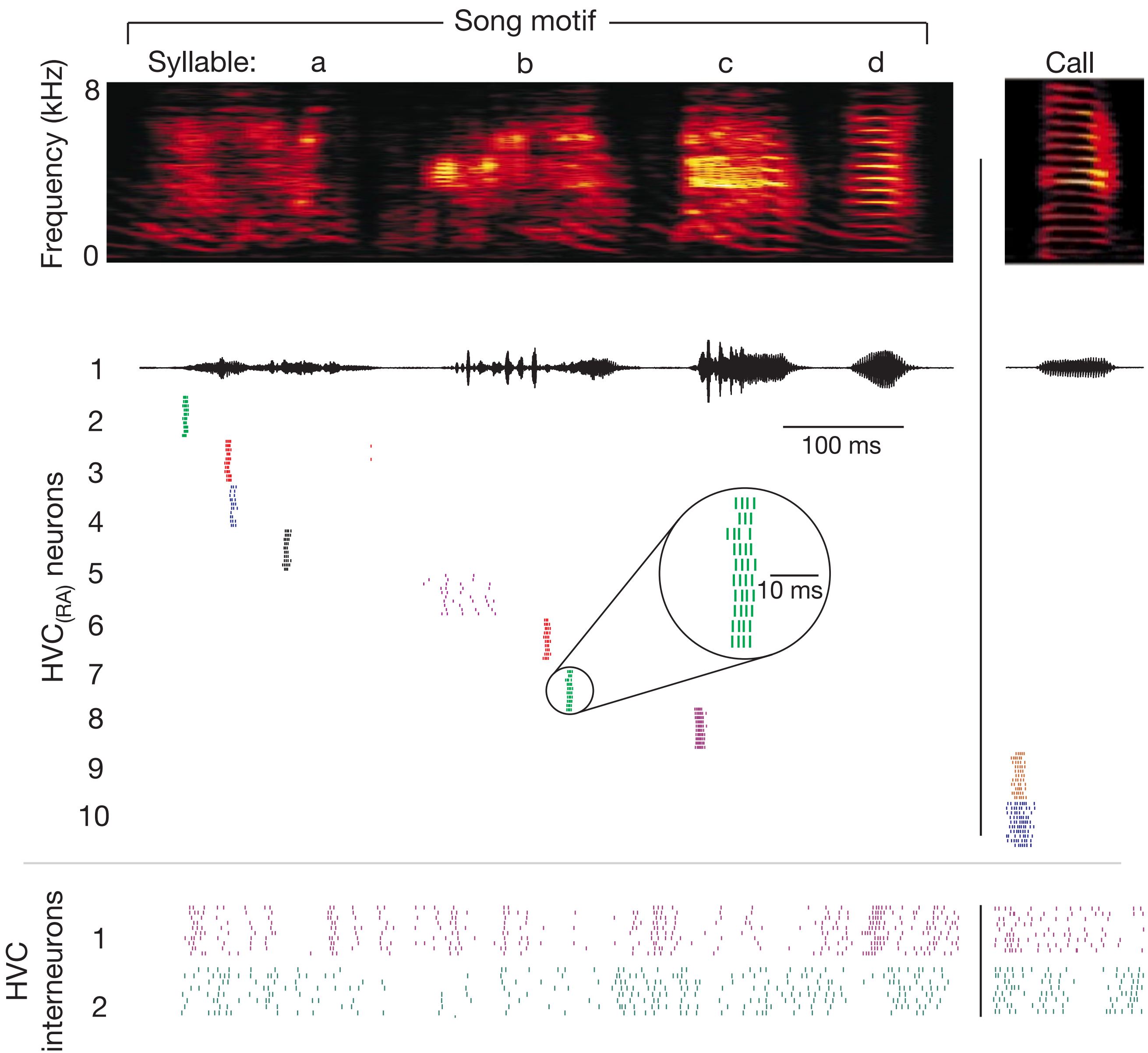




Hahnloser et al., Nature, 2002

Neural sequences in songbird





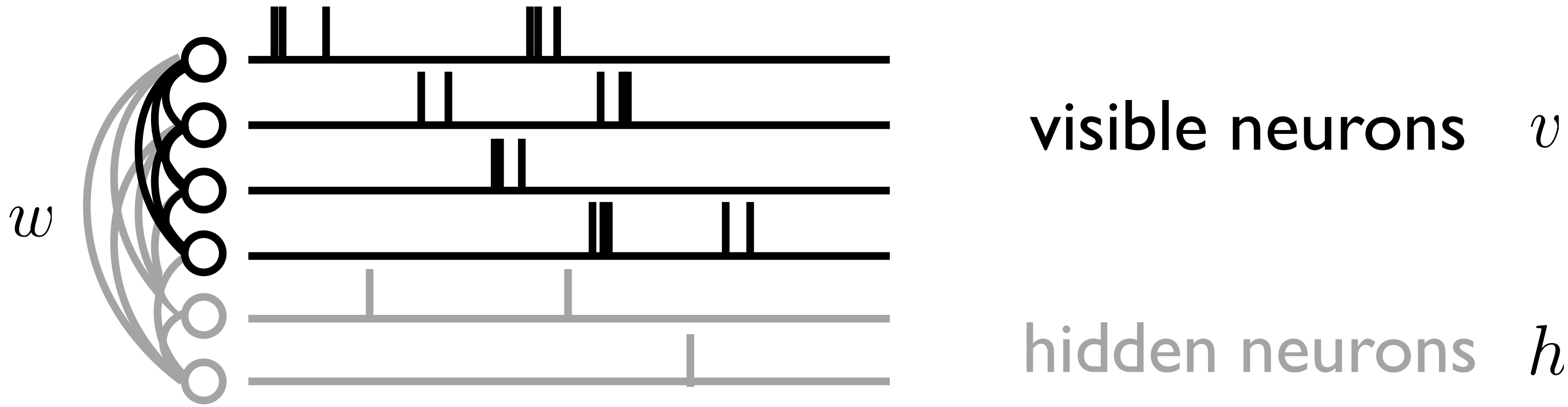
Hahnloser et al., Nature, 2002

Outline

- Motivation for modelling sequence learning
- Definition of the problem
- Network model
- Learning rule (with only visible units)
- Learning rule (with additional hidden units)
- Discussion

Problem description

Learn a **distribution** of **temporal** sequences with a **recurrent** network of spiking neurons with **hidden** units.



$$v = \left\{ v_i(t) = \sum_{t_i^f} \delta(t - t_i^f) \middle| i = 1 \ldots N_{\mathrm{v}}, t \in [0, T] \right\}$$

Problem description

Learn a **distribution** of **temporal** sequences with a **recurrent** network of spiking neurons with **hidden** units.

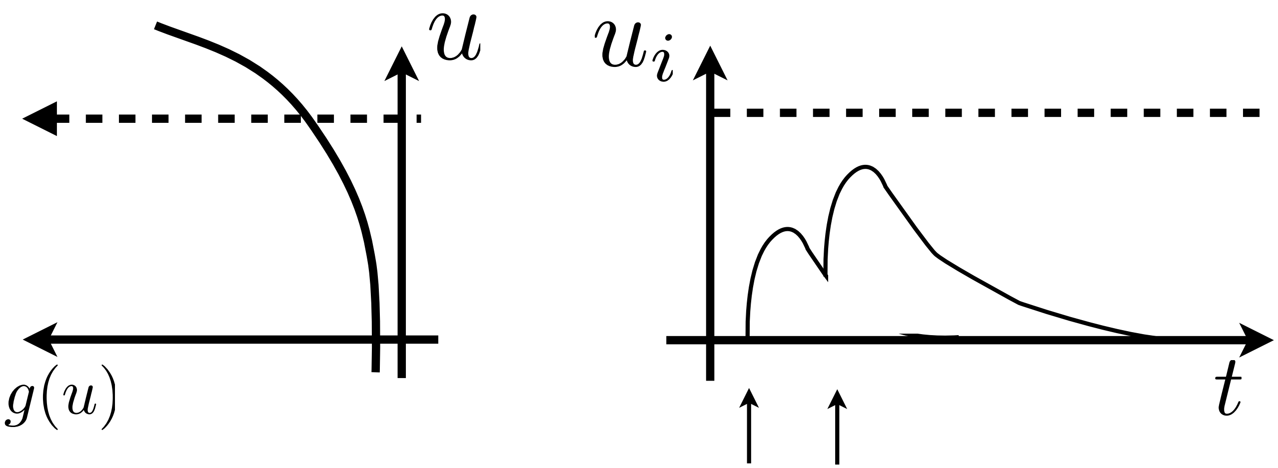
Minimize: $$D_{KL}(P^*(v) \| P_w(v)) = \left\langle \log \left(\frac{P^*(v)}{P_w(v)} \right) \right\rangle_{P^*(v)}$$

Network model

Network spiking activity: $x = (v, h)$

Membrane potential of neuron i: $u_i(t) = \sum_{j \neq i} w_{ij} x_j^{\epsilon}(t) + x_i^{\eta}(t)$

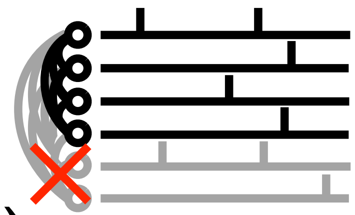
Probability density that neuron i fires at t : $\rho_i(t) = g(u_i(t))$



$$P_w(x) = \exp\left(\sum_{i=1}^{N} \int_0^T x_i(t) \log \rho_i(t) - \rho_i(t) dt \right)$$

Outline

- Motivation for modeling sequence learning
- Definition of the problem
- Network model
- Learning rule (with only visible units)



- Learning rule (with additional hidden units)
- Discussion

Deriving the learning rule (only visible units)

Model pdf of visible pattern v

$$P_w(v) = \exp\left(\sum_{i=1}^{N}\int_0^T v_i(t)\log\rho_i(t) - \rho_i(t)dt\right)$$

Gradient learning rule:

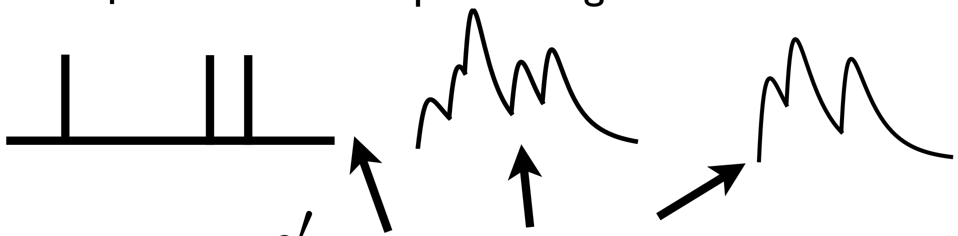
$$\begin{aligned}\Delta w &= -\nabla_w D_{KL}(P^*(v)||P_w(v)) \\ &= \langle \nabla_w \log P_w(v)\rangle_{P^*(v)}\end{aligned}$$

Online learning rule:

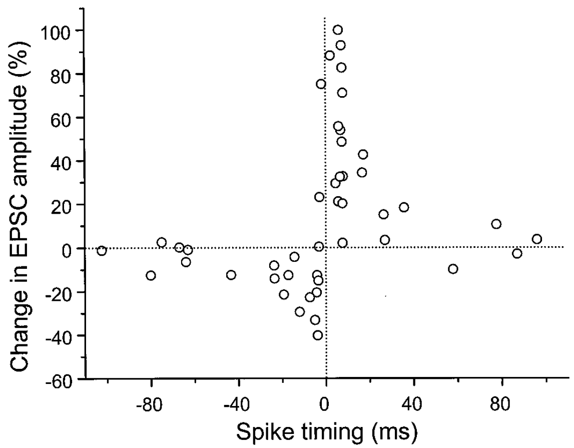
$$\boxed{\dot{w}_{ij} = \frac{\rho_i'}{\rho_i}(v_i - \rho_i)v_j^{\epsilon}}$$

Biological plausibility (1): link to STDP

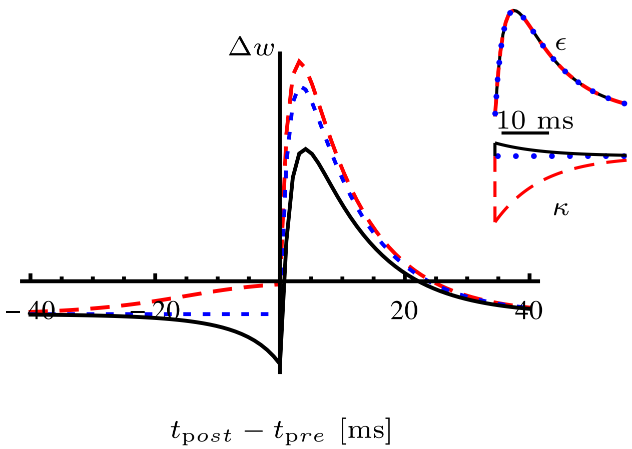
Dirac spike train   Expect. firing rate   EPSP

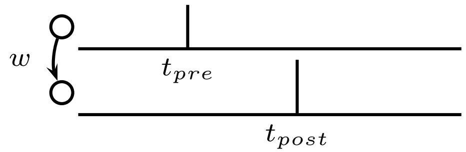


$$\Delta w_{ij} = \frac{\rho_i'}{\rho_i}(v_i - \rho_i)v_j^{\epsilon}$$
$$\approx \epsilon(t_{post} - t_{pre}) - b$$

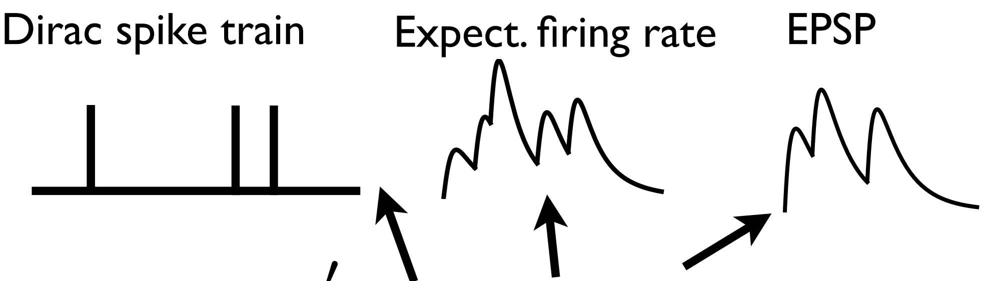


Bi and Poo, 1998





Biological plausibility (2): link to triplet rule

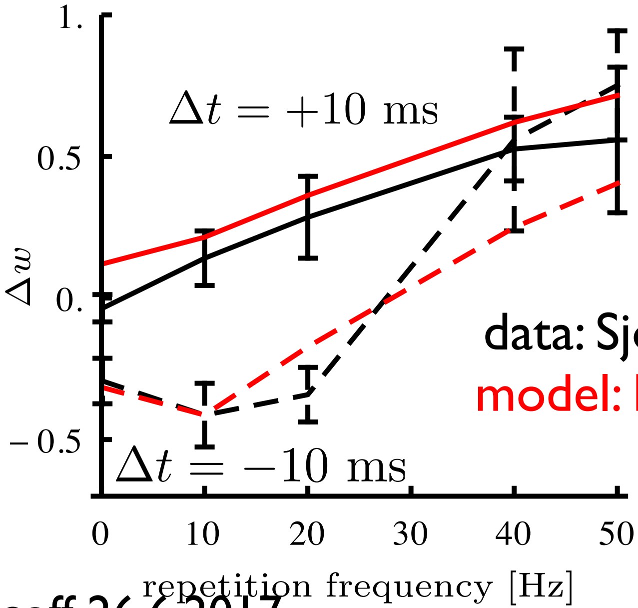


$$\Delta w_{ij} = \frac{\rho_i'}{\rho_i}(v_i - \rho_i)v_j^\epsilon$$
$$= -\rho_i' v_j^\epsilon + \frac{\rho_i'}{\rho_i} v_i v_j^\epsilon$$

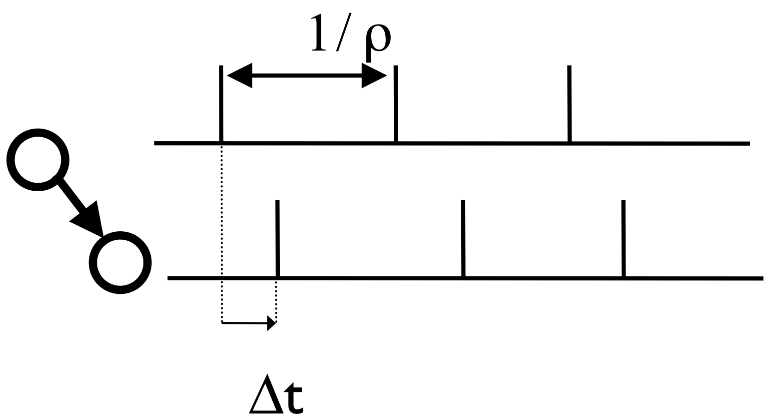
post pre   post post pre

triplet rule
(Pfister et al. 2006)

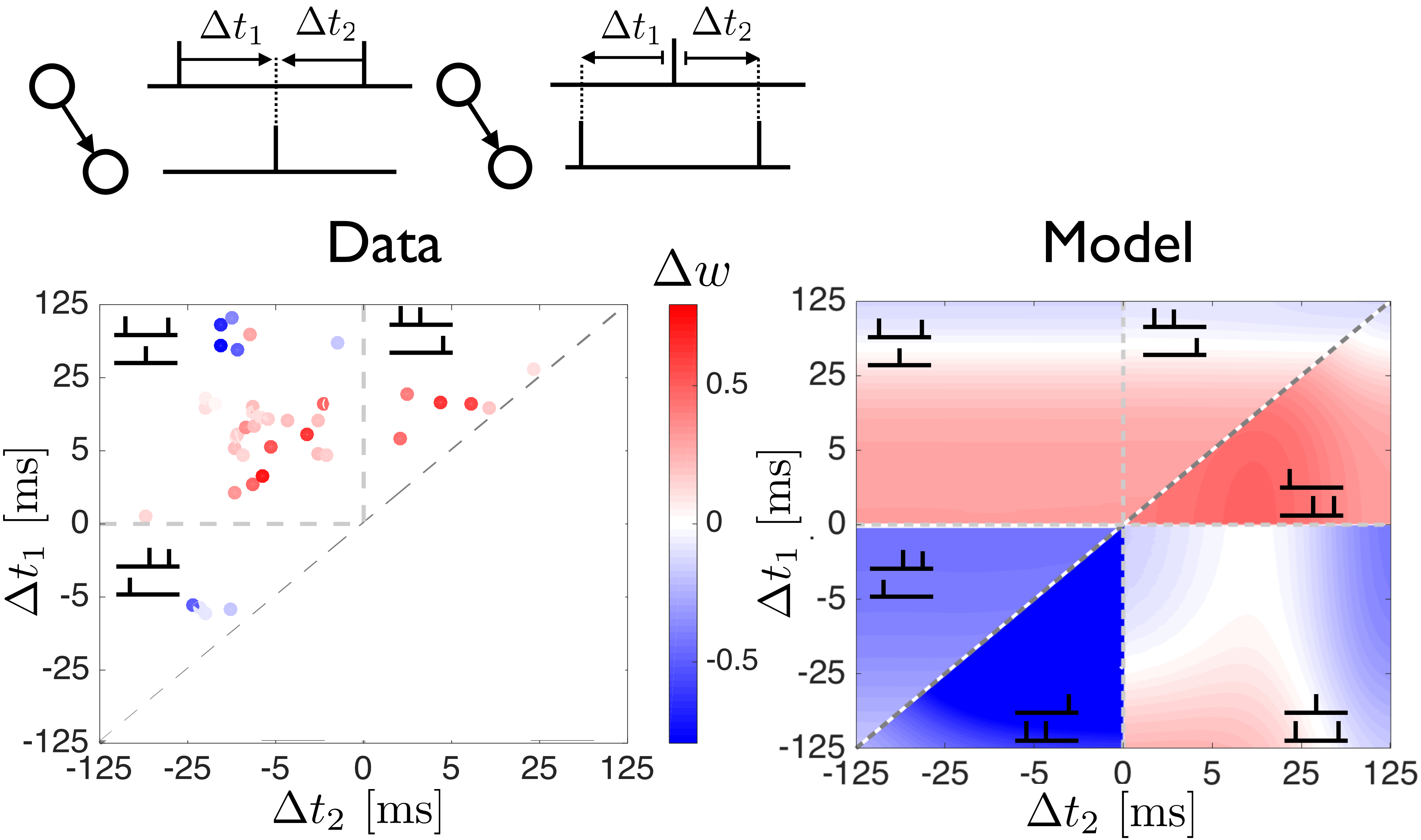
voltage triplet rule
(Clopath et al. 2010)



data: Sjöström et al., 2001
model: Brea et al. 2013



Biological plausibility (3)



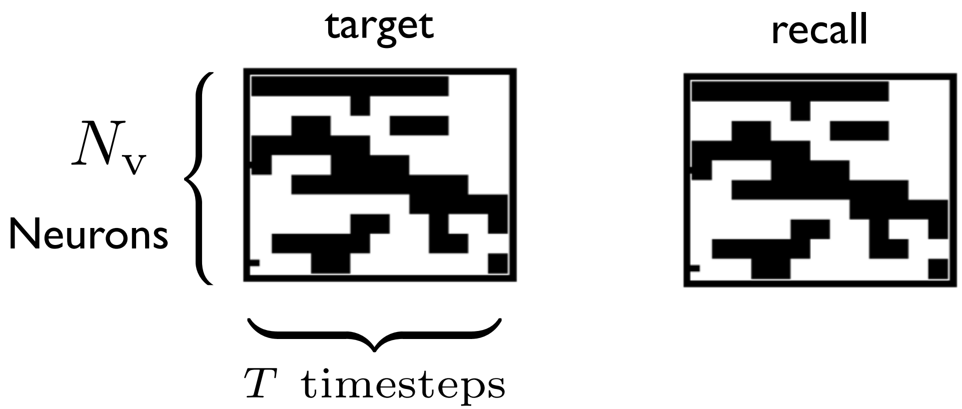
Froemke et al. 2002

Pfister and Dziennik, *in prep.*

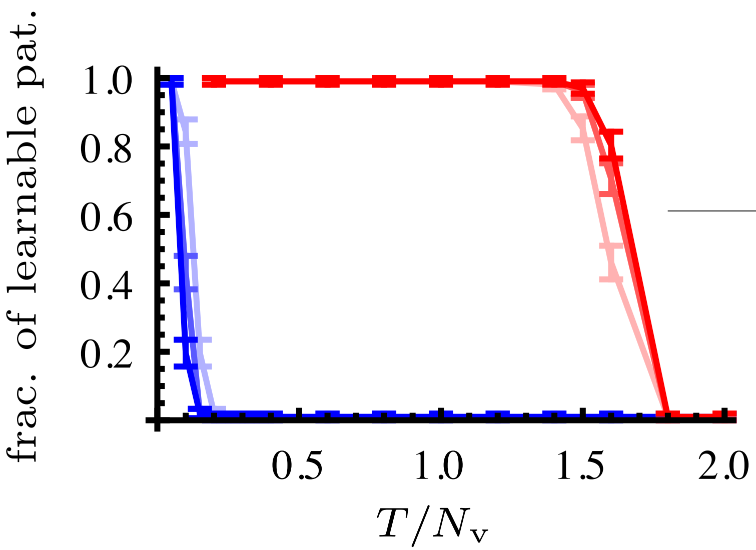
Deterministic pattern: capacity

Assumptions

- $\epsilon(t) =$ 
- g(u): sigmoidal

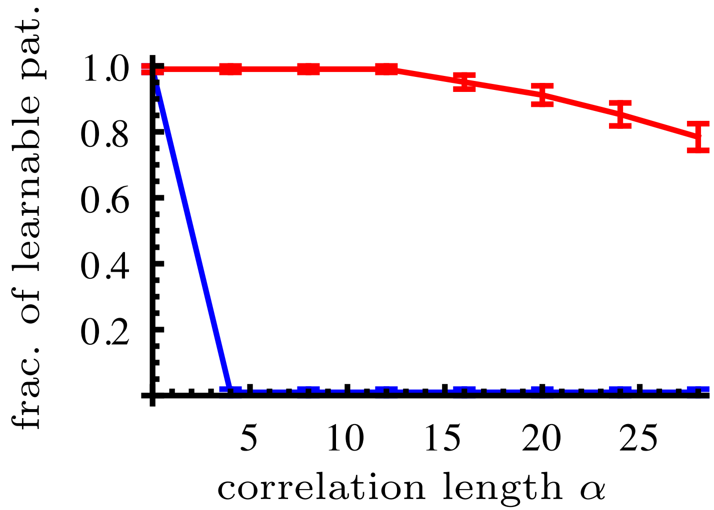


uncorrelated patterns



correlated patterns

$P(v_i(\alpha+1) = v_i(0)|v_i(0)) = 0.5$

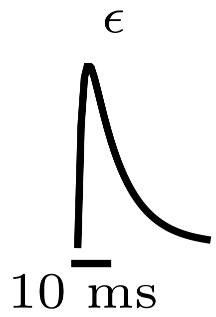


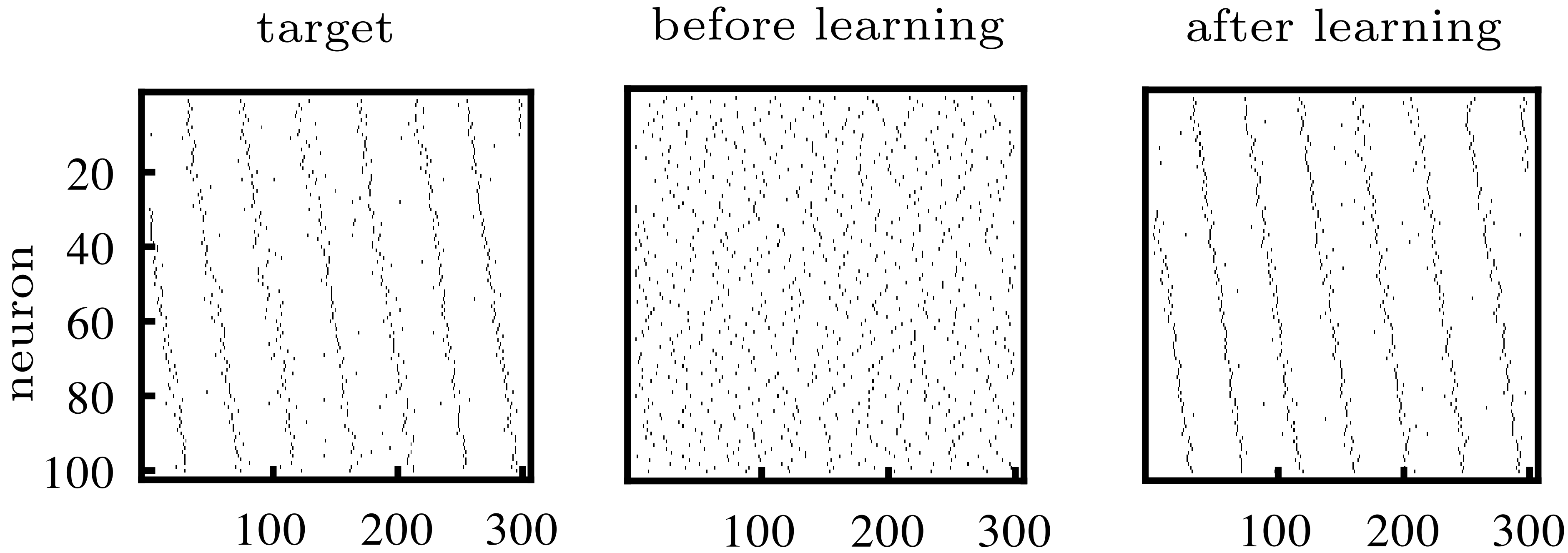
Our learning rule : $\Delta w_{ij} = (v_i(t+1) - \rho_i(t+1))v_j(t)$

Hopfield learning rule : $\Delta w_{ij} = (v_i(t+1) - 0.5)(v_j(t) - 0.5)$

Brea, Senn, Pfister, J Neurosc., 2013

Realistic EPSP

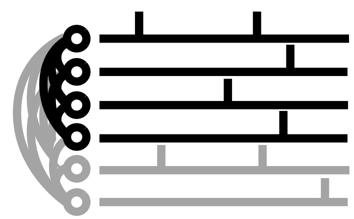




Brea, Senn, Pfister, J Neurosc., 2013

Outline

- Motivation for modelling sequence learning
- Definition of the problem
- Network model
- Learning rule (with only visible units)
- Learning rule (with additional hidden units)
- Discussion



Deriving the learning rule

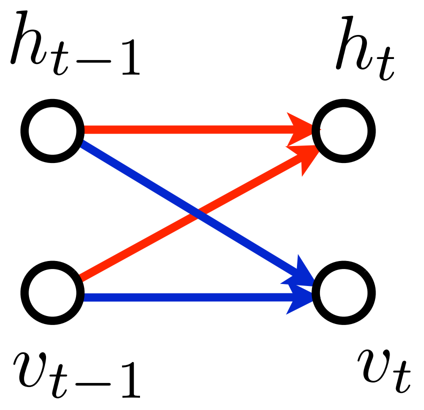
$$
\begin{aligned}
\Delta w &= -\nabla_w D_{KL}(P^*(v)||P_w(v)) \\
&= \langle \nabla_w \log P_w(v) \rangle_{P^*(v)} \\
&= \left\langle \nabla_w \log \sum_h P_w(v,h) \right\rangle_{P^*(v)} \\
&= \left\langle \frac{1}{P_w(v)} \sum_h \nabla_w P_w(v,h) \right\rangle_{P^*(v)} \\
&= \langle \nabla_w \log P_w(v,h) \rangle_{P_w(h|v)P^*(v)}
\end{aligned}
$$

Hard to
sample from

Two tricks to solve the problem

1. Joint distribution can be factorized

$$P(v,h)=\underbrace{\prod_t p(v_t|v_{t-1},h_{t-1})}_{R(v|h)}\;\underbrace{\prod_t p(h_t|v_{t-1},h_{t-1})}_{Q(h|v)}$$

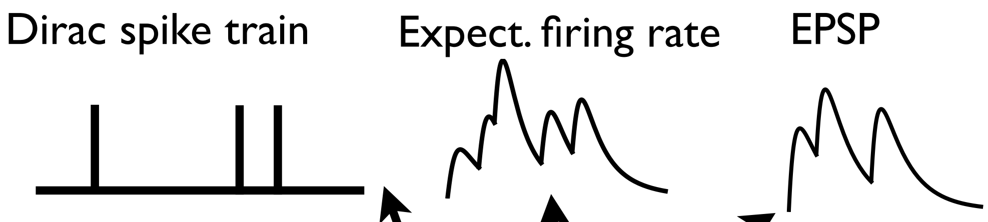


easy to calculate easy to sample from

2. Maximize a lower bound of the log-likelihood

$$\log P(v)=\log\left\langle R(v|h)\right\rangle_{Q(h|v)}\;\geq\;\left\langle \log R(v|h)\right\rangle_{Q(h|v)}$$

Biological relevance



$$\boxed{\Delta w_{ij} = (r - \bar{r})\frac{\rho_i'}{\rho_i}(x_i - \rho_i)x_j^{\epsilon} \qquad i: \ \text{hidden}}$$

internal reward (global)

$$r = \int_0^T \sum_{i=1}^{N_{\mathrm{v}}} \beta u_i(t) x_i(t) - \rho_i(t) dt$$

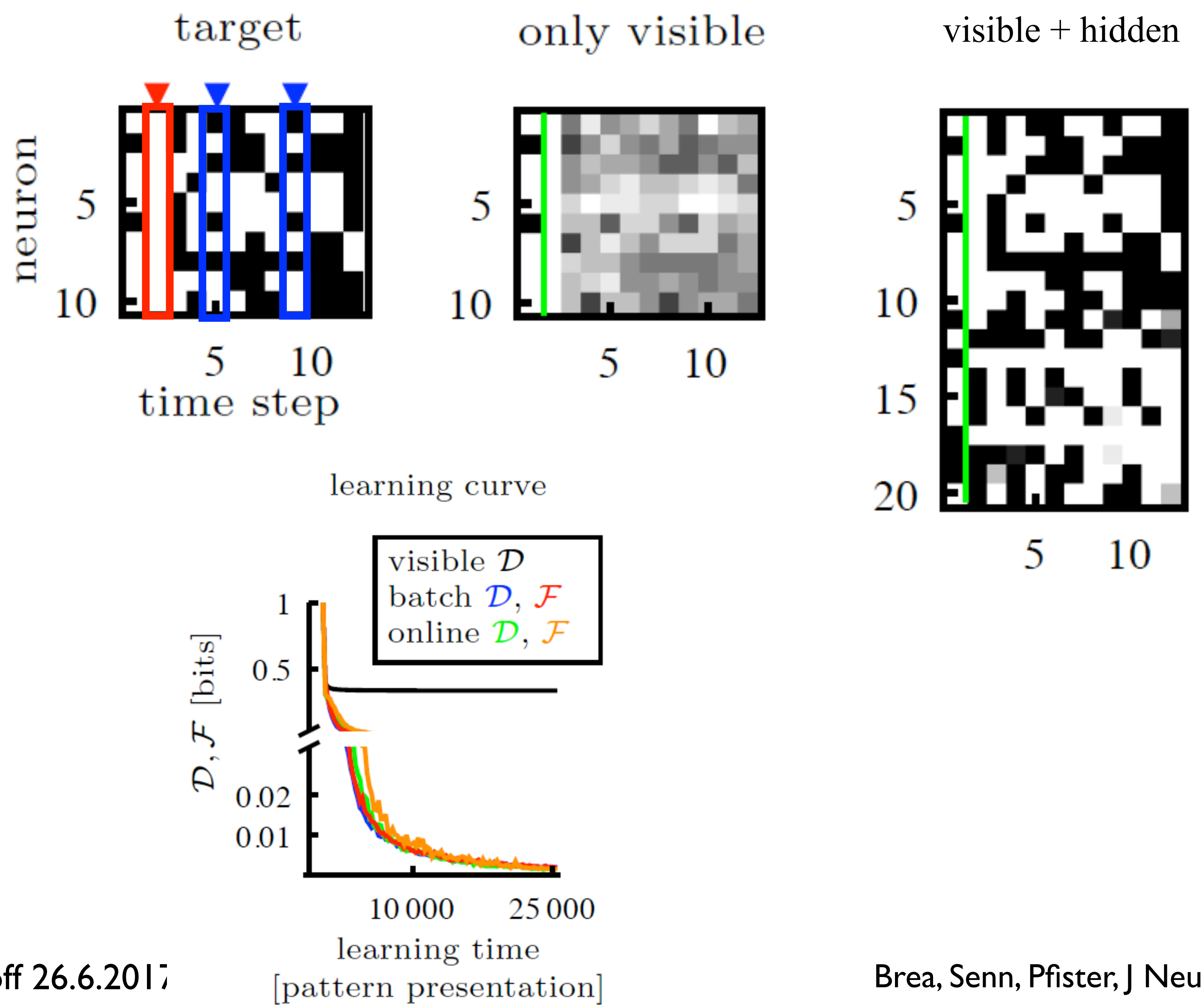
Related to astrocytes ? Neuromodulators ?

➡ Modulates synaptic plasticity (Henneberger et al. 2010)

➡ Can shift LTD to LTP (Panatier et al. 2006)

➡ acts at a slow time constant

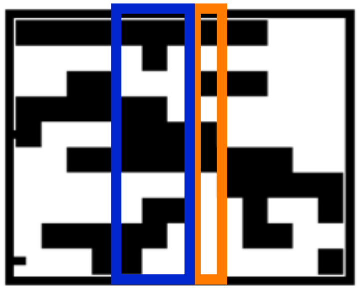
Difficult deterministic patterns can be learned

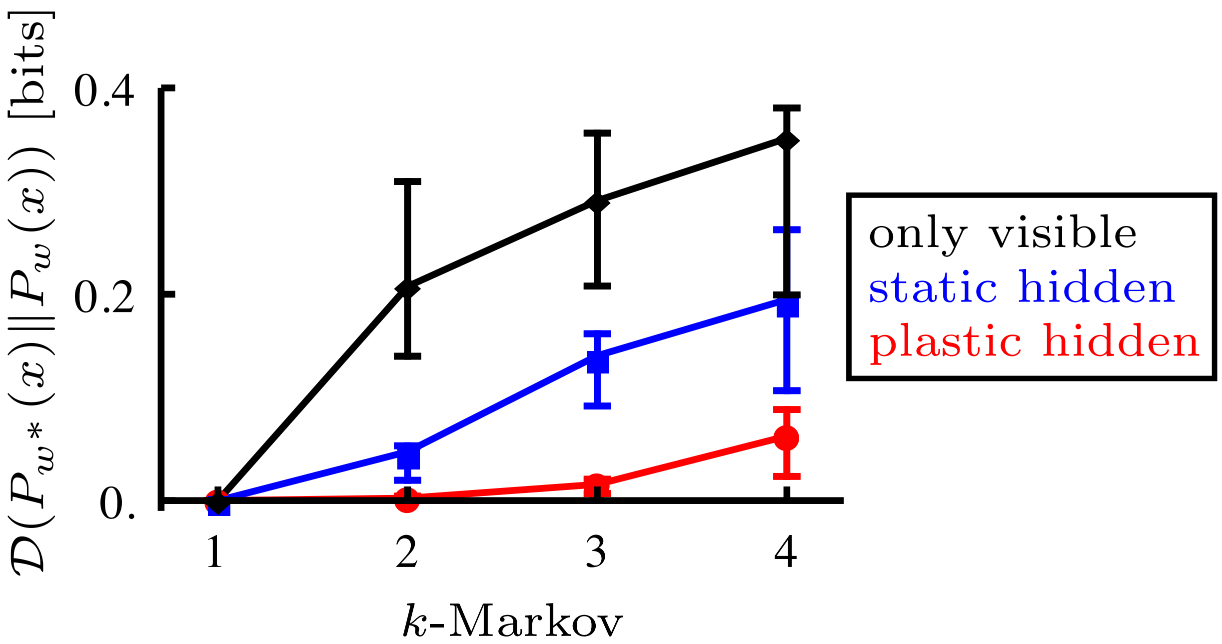


Brea, Senn, Pfister, J Neurosc., 2013

Learnable stochastic patterns

$$P^*(v) = \prod_t P(v_t | v_{t-1}, v_{t-2}, v_{t-3})$$





Brea, Senn, Pfister, J Neurosc., 2013

Conclusion I

- Derived learning rules from first principles
  - ➡ that are causal, local (+ global term)
  - ➡ link with STDP, triplet rule
  - ➡ simple learning rule for the hidden weights
  - ➡ can learn difficult (non-Markovian) sequences of spikes
- learning time is still an issue

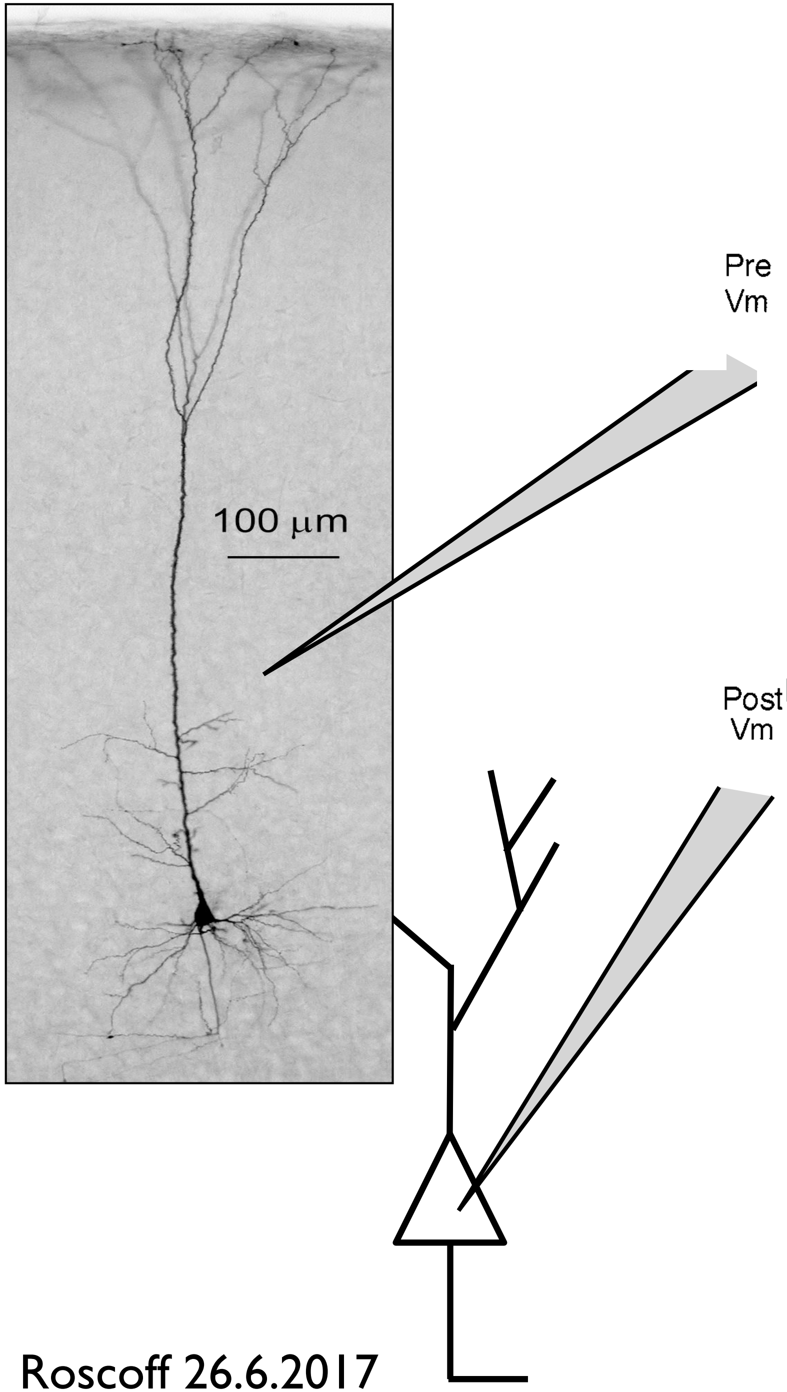
Outlook of part II

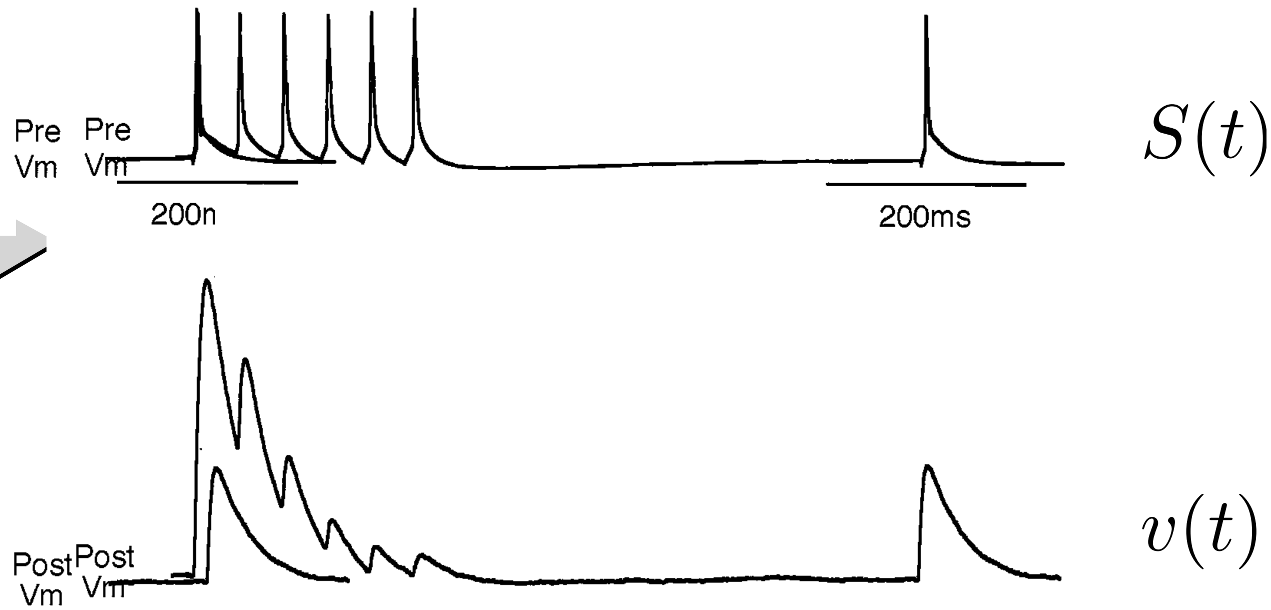
1. Learning with the GLM
   (Brea at al. J. Neurosc. 2013)

2. Decoding at single synapses
   (Pfister et al. Nat. Neurosc. 2010)

3. Decoding with the GLM
   (with A. Kutschireiter)

Short-Term Synaptic Plasticity





Tsodyks and Markram 1997

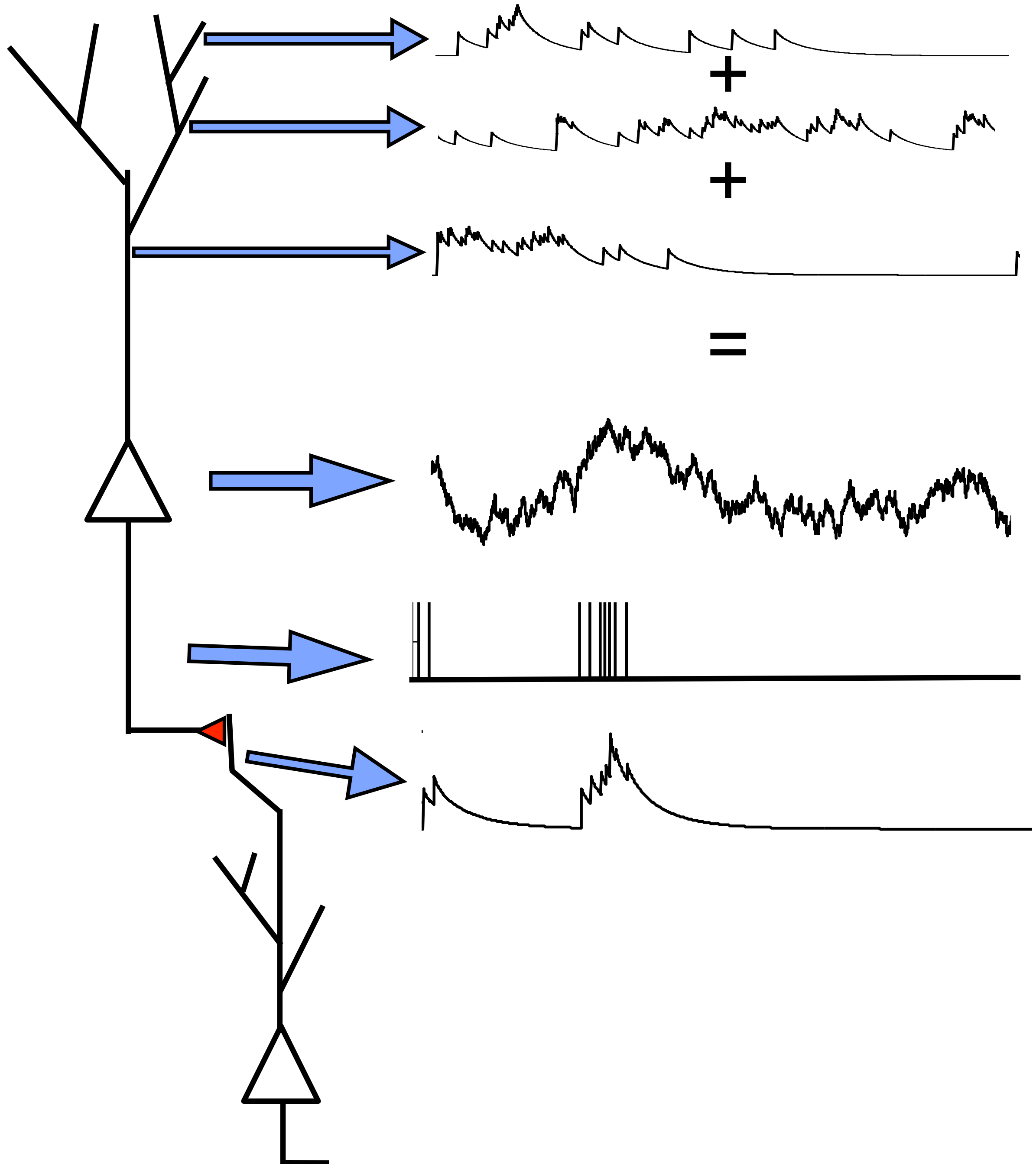
Postsynaptic depolarization

$$\dot{v} = -\tau^{-1}(v - v_0) + JxS(t)$$

resource variable

$$\dot{x} = -\tau_D^{-1}(x - 1) - YxS(t)$$

Analog and digital signaling



Analog

↓

Digital

↓

Analog

Loss of Information !

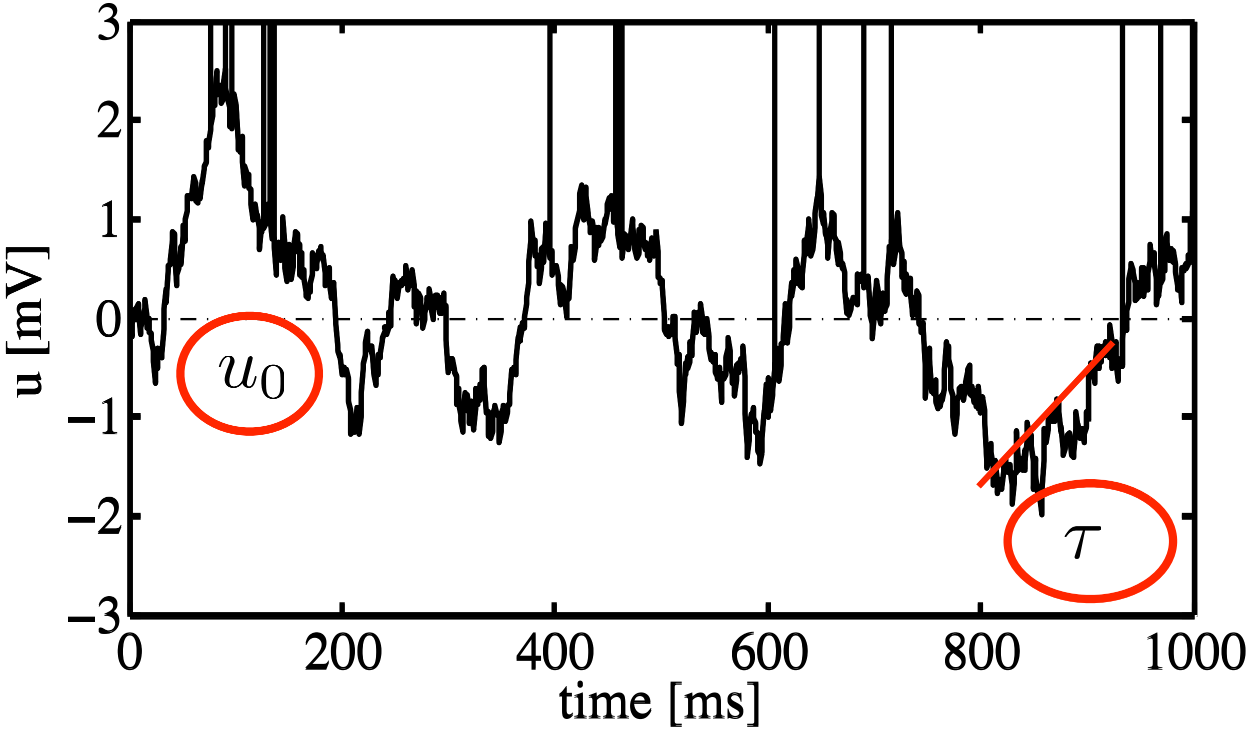
Reconstruct? with STP ?

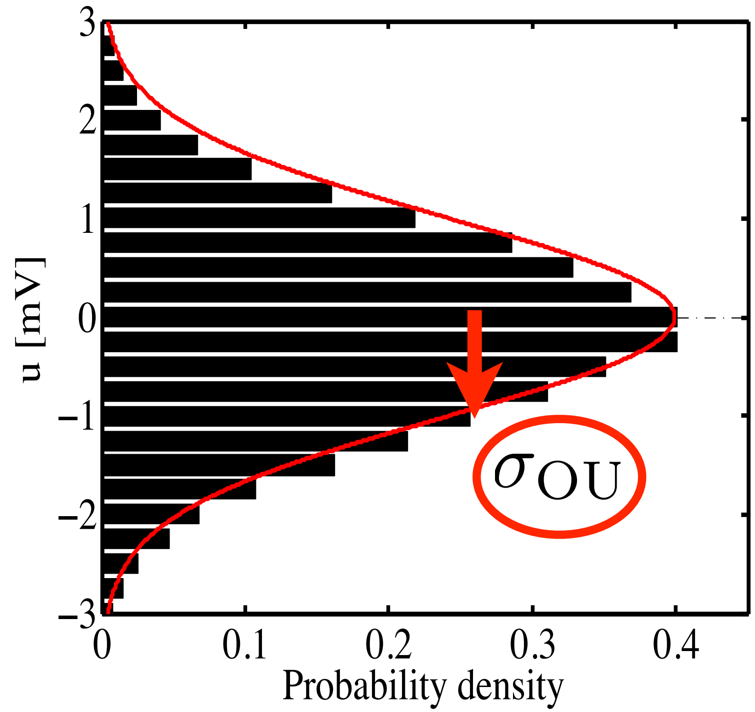
Does he love me?

Mister random mood

Miss detector

Neuronal model

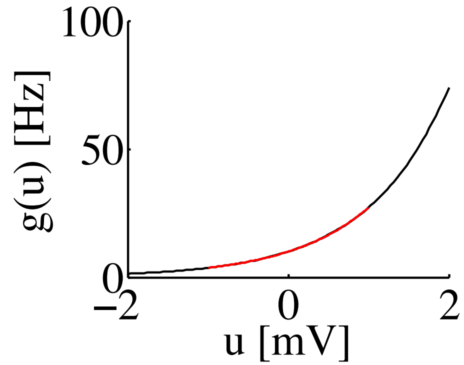




Ornstein Uhlenbeck process: $\dot{u} = -\tau^{-1}(u - u_0) + \sigma\xi(t)$

prob. spike: $p(s_t = 1 | u_t) = g(u_t)\Delta t$

$$g(u) = g_0 \exp(\beta u)$$



Neuronal model

$$
\begin{tikzcd}
u_1 \arrow[r] \arrow[d] & \cdots \arrow[r] & u_{t-1} \arrow[r] \arrow[d] & u_t \arrow[d] \\
s_1 & \cdots & s_{t-1} & s_t
\end{tikzcd}
$$

Ornstein Uhlenbeck process: $$\dot{u} = -\tau^{-1}(u - u_0) + \sigma\xi(t)$$

prob. spike: $$p(s_t = 1 | u_t) = g(u_t)\Delta t$$

Inference of the membrane potential

$$
\begin{tikzcd}
u_1 \arrow[r] \arrow[d] & \ldots \arrow[r] & \boxed{u_{t-1}} \arrow[r] \arrow[d] & \boxed{u_t} \arrow[d] \\
s_1 & \ldots & s_{t-1} & s_t
\end{tikzcd}
$$

Posterior distribution:

$$
\underline{p(u_t|s_{1\ldots t})} \propto p(s_t|u_t) \int p(u_t|u_{t-1})\, \underline{p(u_{t-1}|\, s_{1\ldots t-1})}\, du_{t-1}
$$

$$
\Downarrow \qquad\qquad\qquad\qquad \Downarrow
$$

$$
\simeq \mathcal{N}(u_t|{\color{red}\hat{u}_t}, {\color{blue}x_t}) \qquad\qquad = \mathcal{N}(u_{t-1}|{\color{red}\hat{u}_{t-1}}, {\color{blue}x_{t-1}})
$$

Depressing synapse

Optimal estimator

Postsynaptic depolarization

$$\dot{v} = -\tau^{-1}(v - v_0) + JxS(t)$$

Posterior mean

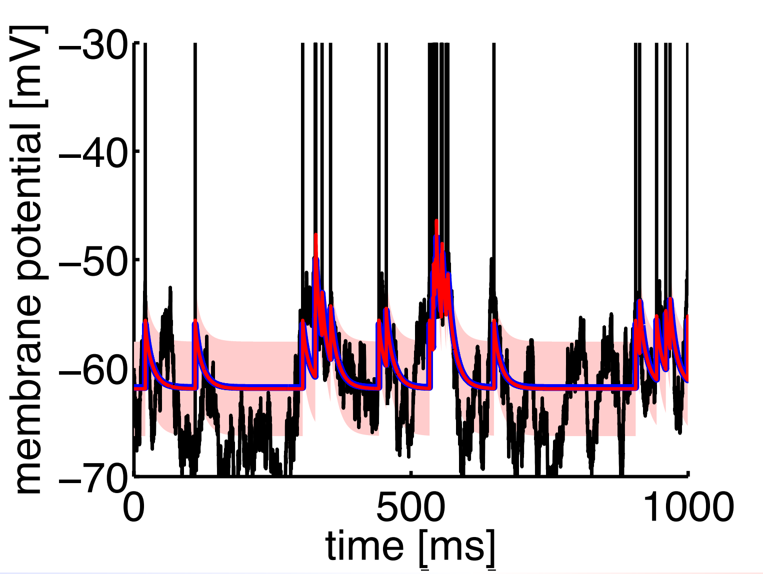
$$\dot{\hat{u}} = -\tau^{-1}(\hat{u} - u_0) + \beta x(S(t) - \gamma(t))$$

resource variable

$$\dot{x} = -\tau_D^{-1}(x - 1) - YxS(t)$$

Posterior variance

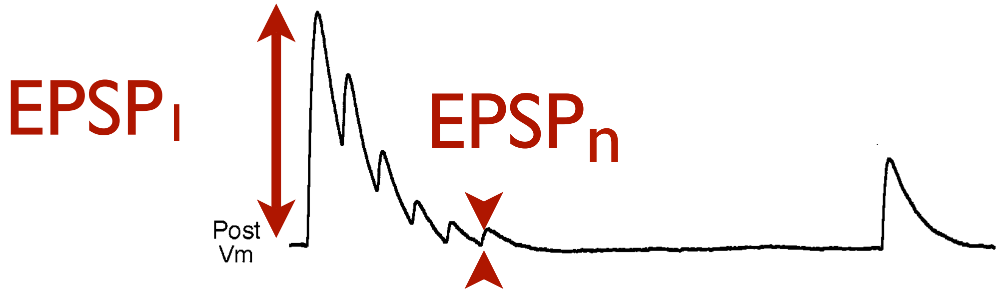
$$\dot{x} = -2\tau^{-1}(x - 1) - \gamma(t)\beta^2 x^2$$



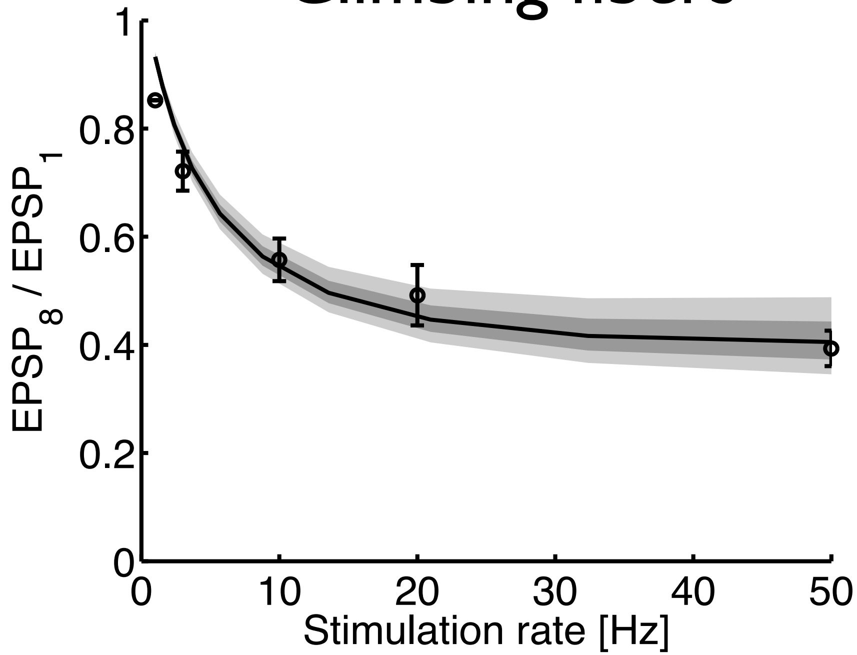
Tsodyks and Markram, 1997

Pfister, Dayan and Lengyel: NIPS 2009, Nature Neuroscience 2010

Steady-state spiking increment



Climbing fibers

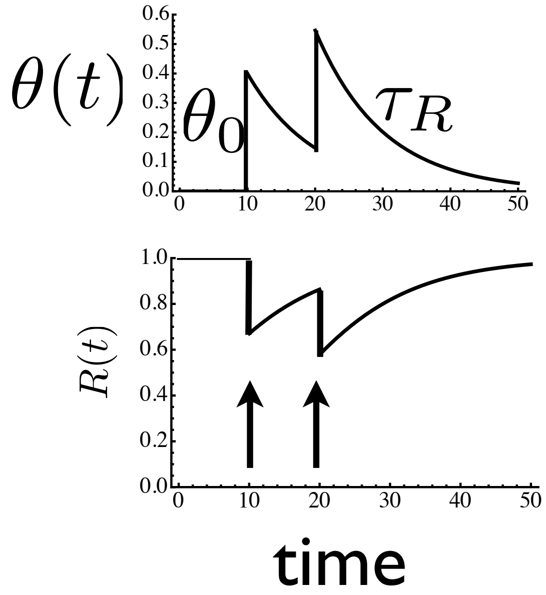


Dittman et al. 2000

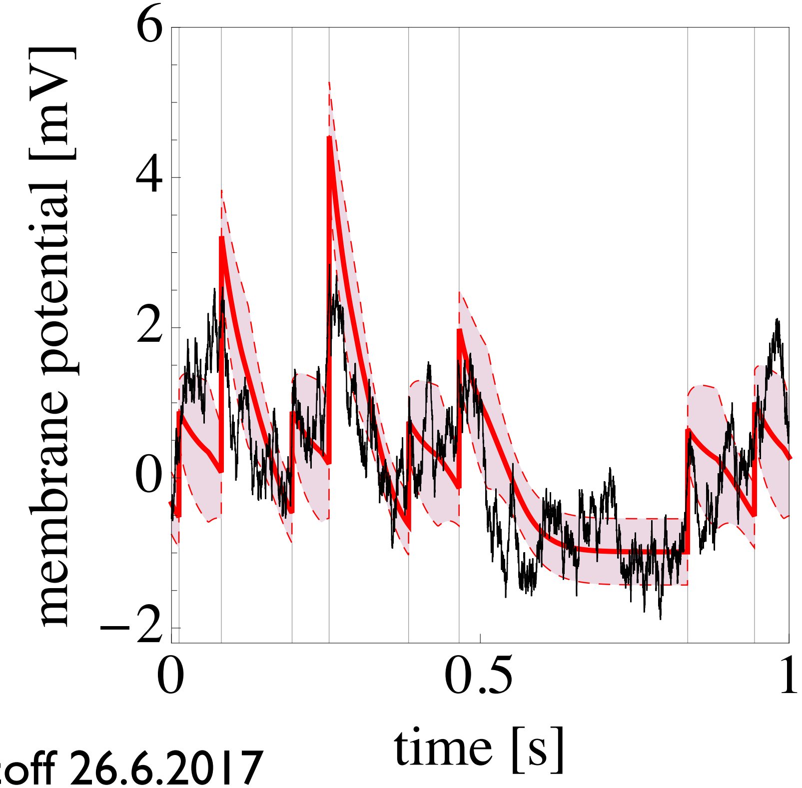
Include adaptation effects

Firing intensity:

$$g(u(t)) \rightarrow g(u(t)) \underbrace{\exp(-\beta\theta(t))}_{R(t)}$$



Facilitation



Intuition:

Spike in the refractory region

➔evidence for higher u

Surace and Pfister, in Prep

Depressing synapse

Postsynaptic depolarization

$$\dot{v} = -\frac{v - v_0}{\tau} + JxyS(t)$$

resource variable

$$\dot{x} = -\frac{x - 1}{\tau_{\mathrm{D}}} - yxS(t)$$

utilization variable

$$\dot{y} = -\frac{y - Y}{\tau_{\mathrm{F}}} + Y_0(1 - y)S(t)$$

Tsodyks and Markram, 1997

Optimal estimator

Posterior mean

$$\dot{\hat{u}} = -\frac{\hat{u} - u_0}{\tau} + \beta x(S(t) - \gamma(t))$$

Posterior variance

$$\dot{x} = -\frac{x - 1}{\tau/2} - \gamma(t)\beta^2 x^2$$

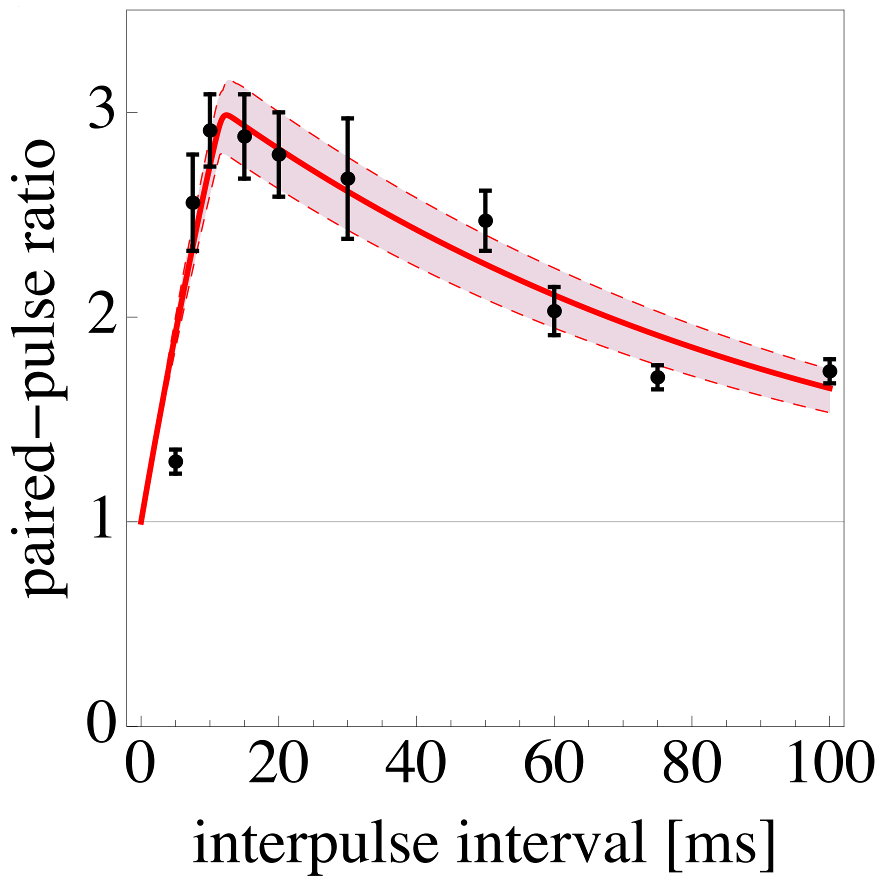
effective "threshold"

$$\dot{\theta} = -\frac{\theta}{\tau_{\mathrm{R}}} + \theta_0 S(t)$$

Surace and Pfister, in Prep

Paired-pulse ratio

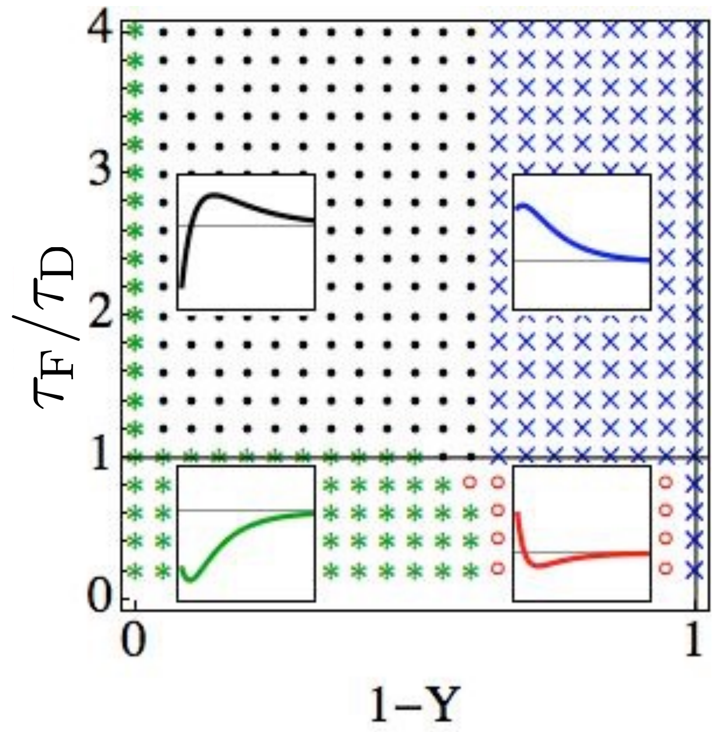
Data: Schaffer collaterals
(Dobrunz et al. 1997)
Optimal estimator



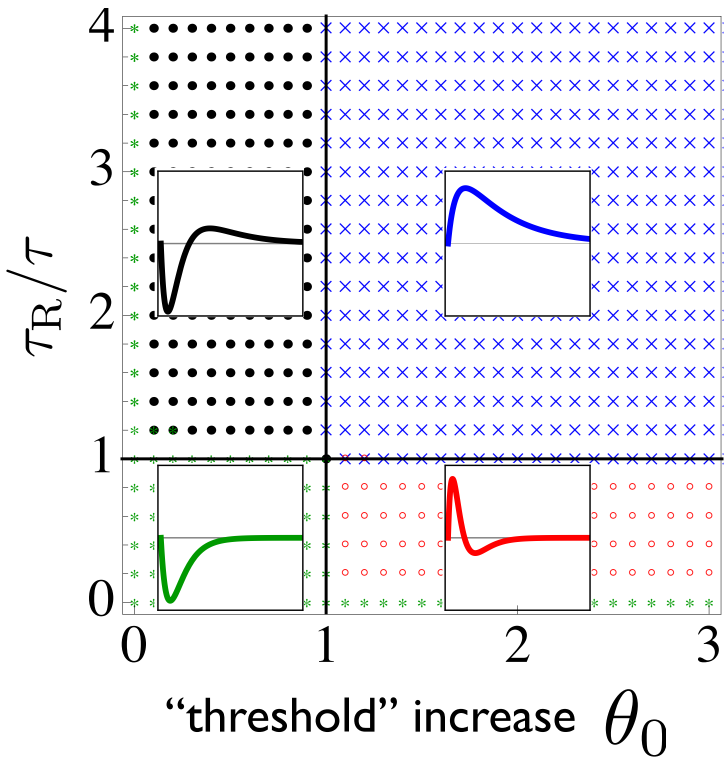
Surace and Pfister, in Prep

Parameter analysis

Tsodyks-Markram Model



Optimal estimator



Surace and Pfister, in Prep

Conclusion II

- Short-term synaptic dynamics is not an undesirable variability but can be useful for computation.

- Synapses with short-term plasticity closely match the behavior of the optimal estimator of the presynaptic membrane potential.

- Predictions of the model
  - STP properties have to be matched to the presynaptic neuron statistics
  - link between facilitation and adaptation

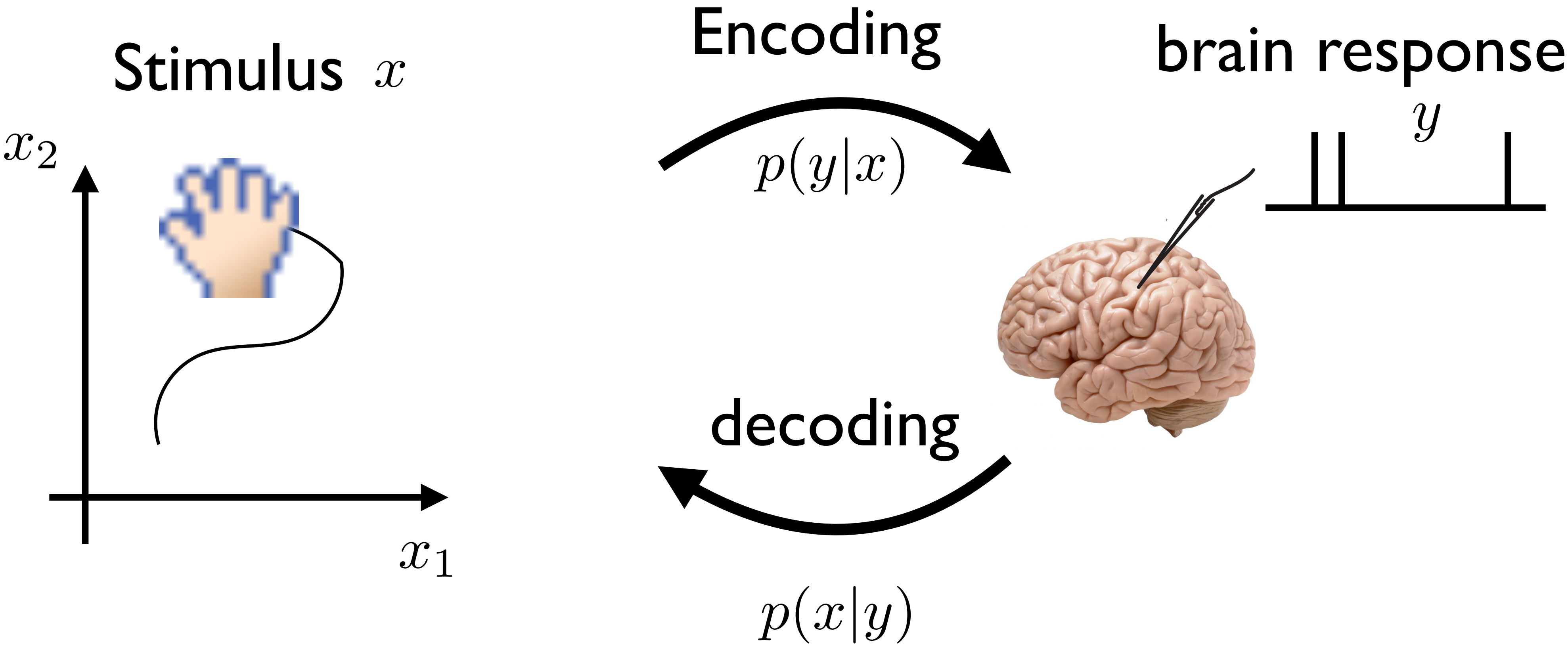
Outlook of part II

1. Learning with the GLM
   (Brea at al. 2013)

2. Decoding at single synapses
   (Pfister et al. 2010)

3. Decoding with the GLM

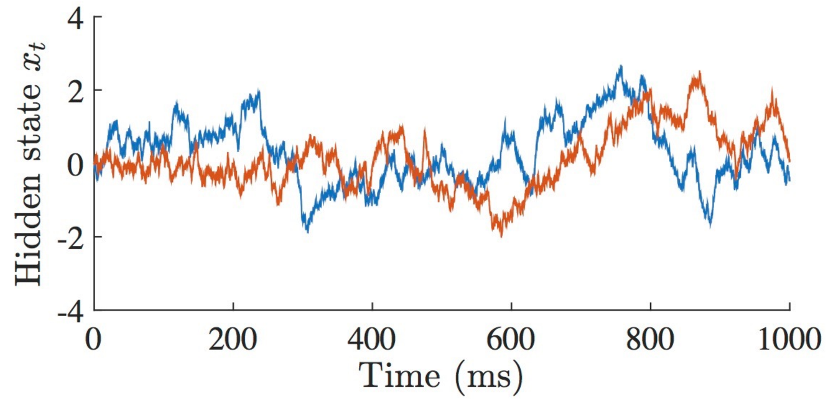
Decoding problem



Generative model

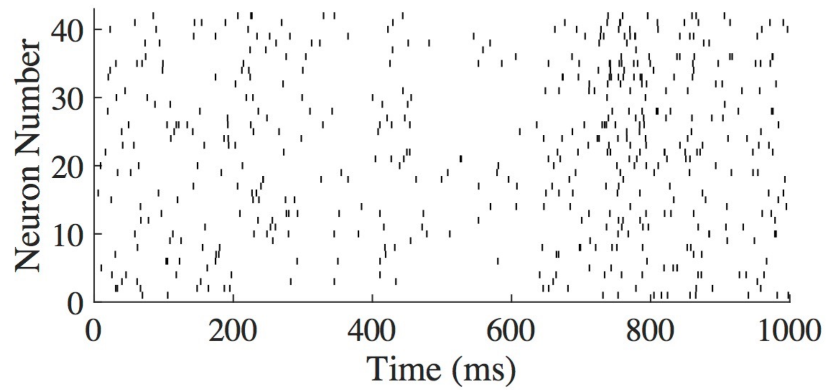
hidden dynamics

$$d\mathbf{x} = \mathbf{f}(\mathbf{x})dt + \Sigma^{1/2}d\omega$$



observation process $\mathbf{y} = \dfrac{d\mathbf{N}}{dt}$

$$d\mathbf{N} \sim \text{Poisson}(\mathbf{g}(\mathbf{u})dt)$$



with membrane potential

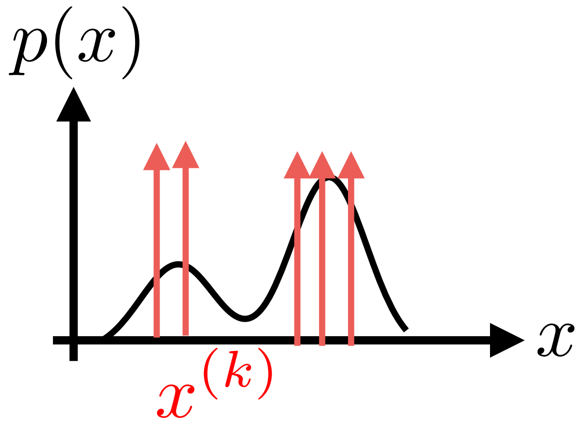
$$\mathbf{u} = W\mathbf{y}^{\epsilon} + V\mathbf{x}$$

task: infer $\mathbf{x}$ from spiking observations $\mathbf{y}$ $\longrightarrow$ $p(\mathbf{x}(t)|\mathbf{y}(0\ldots t))$

Solution

sampling-based representation

$$p(\mathbf{x}_t|\mathbf{y}_{0\ldots t}) = \frac{1}{N}\sum_{k=1}^{N}\delta(\mathbf{x}_t - \mathbf{x}_t^{(k)})$$



spiking neural particle filter

$$d\mathbf{x}^{(k)} = \mathbf{f}(\mathbf{x}^{(k)})dt + \Sigma^{1/2}d\omega + \frac{\mathrm{cov}(\rho(\mathbf{x}),\mathbf{x})}{\langle\rho(\mathbf{x})\rangle}(d\mathbf{N} - \rho(\mathbf{x}^{(k)})dt)$$

with $\rho(\mathbf{x}) = \mathbf{g}(\mathbf{u}(\mathbf{x}))$

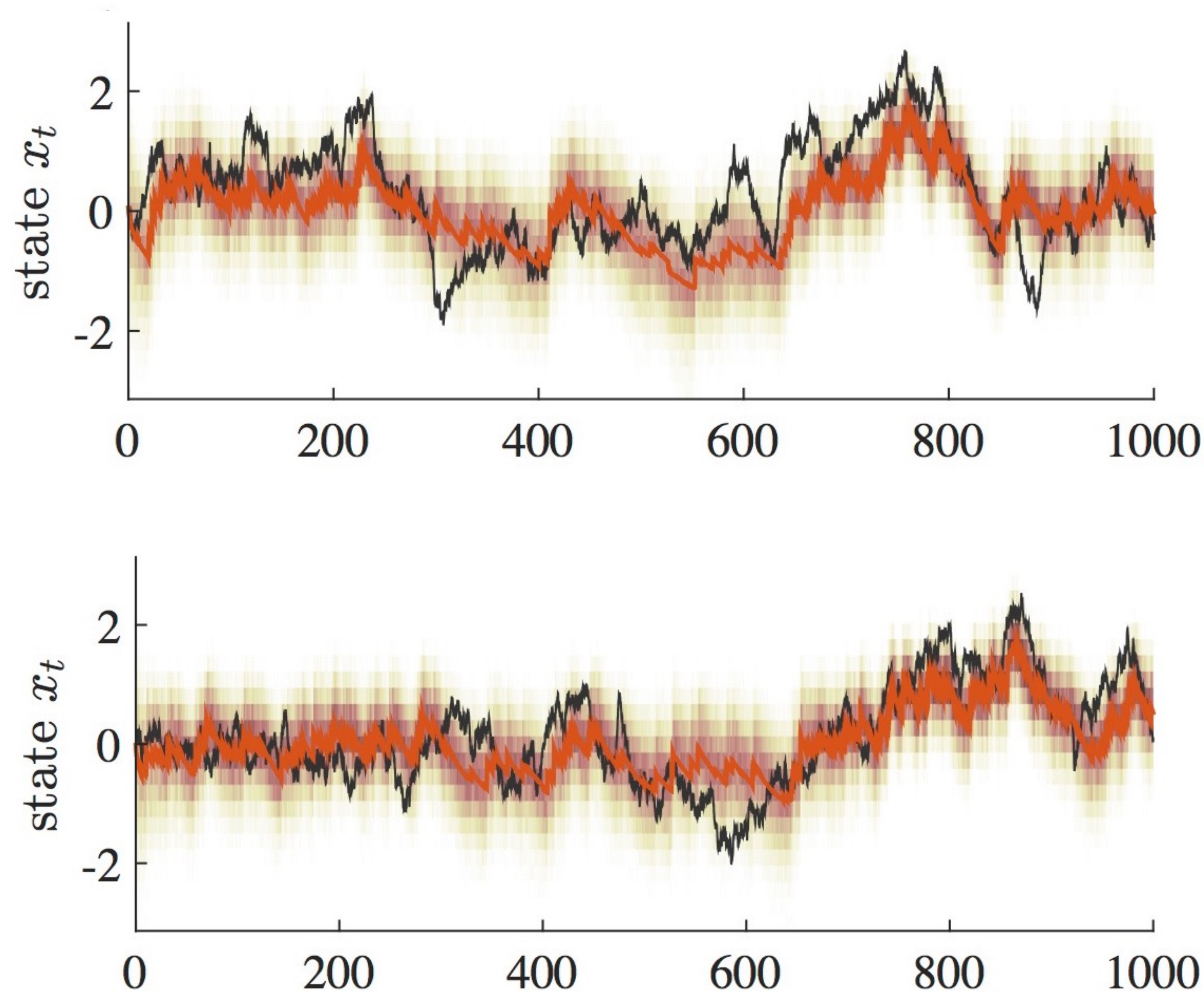
spiking neural particle filter (sNPF)

$$dp = \mathcal{L}^\dagger[p]dt + \left(\frac{g}{\langle g \rangle} - 1\right)(dN - \langle g \rangle dt)p$$

$$d\langle x \rangle = \langle f(x) \rangle dt + \left(\frac{\langle gx \rangle}{\langle g \rangle} - \langle x \rangle\right)(dN - \langle g \rangle dt)$$

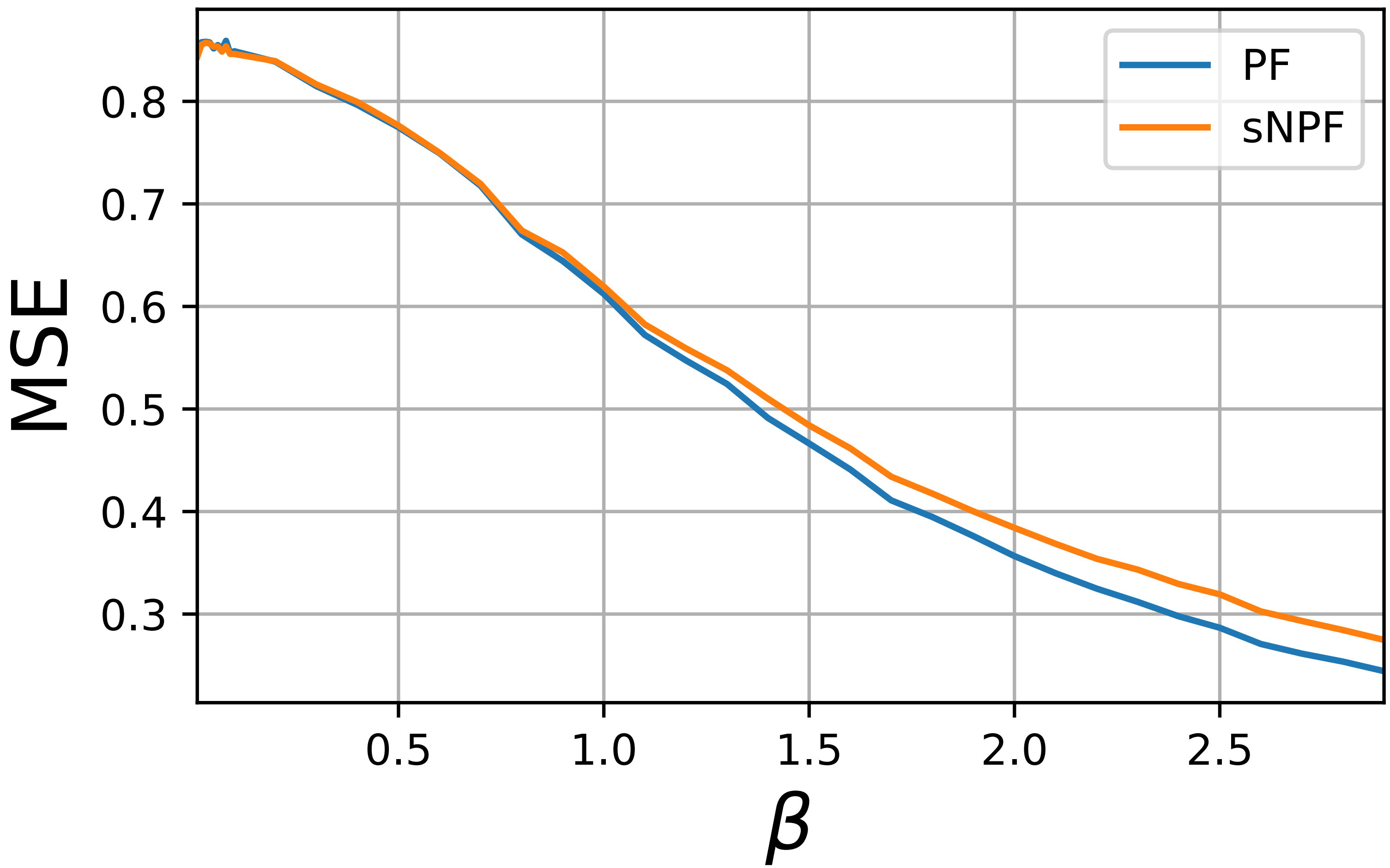
$$dx^{(k)} = f(x^{(k)})dt + \Sigma^{1/2}d\omega + \frac{\mathrm{cov}(g, x)}{\langle g \rangle}(dN - \langle g \rangle dt)$$

Estimated position

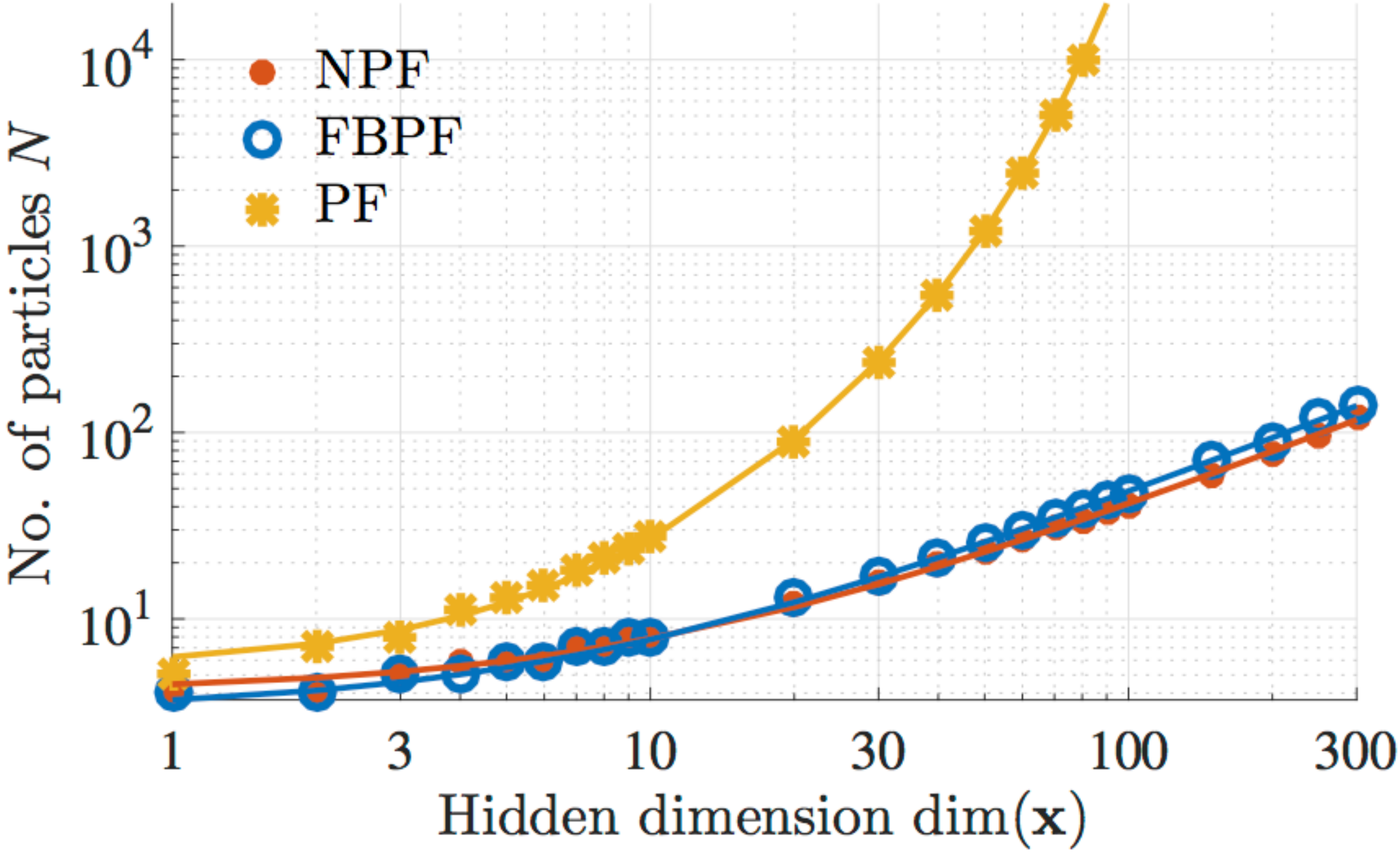


Performance

$$g(u) = g_0 \exp(\beta u)$$



NPF alleviates the curse of dimensionality



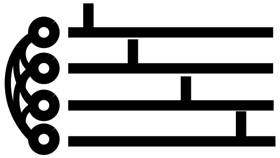
[Note that here simulations are made with a diffusion observation instead of a point emission observation]

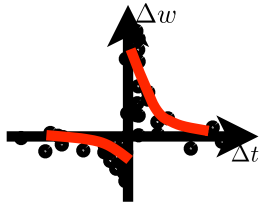
Conclusion III

- Dynamic stimuli can be inferred from spiking recordings
- The performance of the spiking neural particle filter is comparable to standard particle filters
- In higher dimensions unweighted particle filters have the potential to alleviate the curse of dimensionality

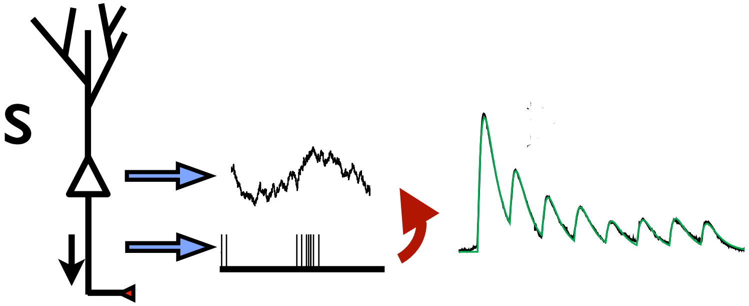
Summary

1. Learning with the GLM
(Brea at al. J. Neurosc. 2013)





2. Decoding at single synapses
(Pfister et al. Nat. Neurosc. 2010)



3. Decoding with the GLM
(with A. Kutschireiter)