Adaptation Dynamics and Inference of Mesoscopic Neuron models

André Longtin
Physics, Faculty of Science
Cellular & Molecular Medicine, Faculty of Medicine
Centre for Neural Dynamics
Brain and Mind Research Institute
University of Ottawa, Canada

Dà Lat, Vietnam, July 2023

Funding: NSERC, CIHR Canada

OUTLINE

1) Neuron firing models: Linear, ReLu, Tanh, Type 1, Type 2, etc…

→ Threshold or current adaptation: brief review

WHY ADAPTATION? One frontier in deep learning

2) Computing with adaptation: precise temporal coding

3) Computing with adaptation: time sequence prediction

4) Inferring adaptive neural circuitry from microscopic data

5) A word about Stochastic Optimal Control of Neurons

→ **produces reliable spiking output for real learning tasks?**

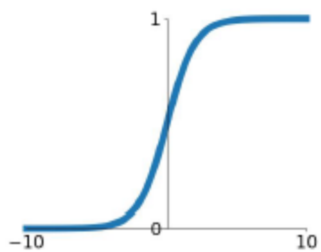
Neural Firing functions used in deep learning

Input = total synaptic currents

Output = mean firing rate

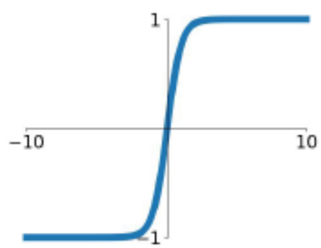
Sigmoid

$\sigma(x) = \frac{1}{1+e^{-x}}$



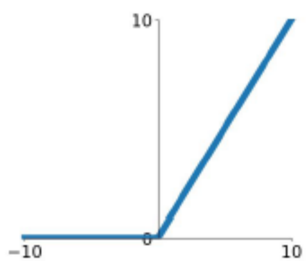
tanh

$\tanh(x)$



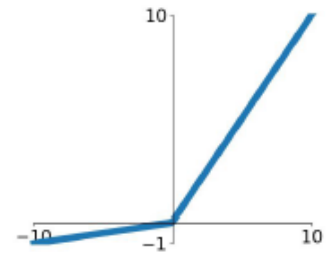
ReLU

$\max(0, x)$



Leaky ReLU

$\max(0.1x, x)$

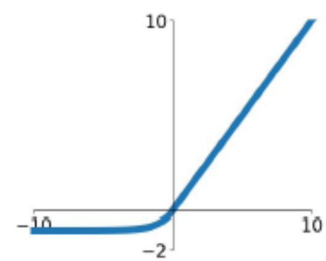


Maxout

$\max(w_1^T x + b_1, w_2^T x + b_2)$

ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$

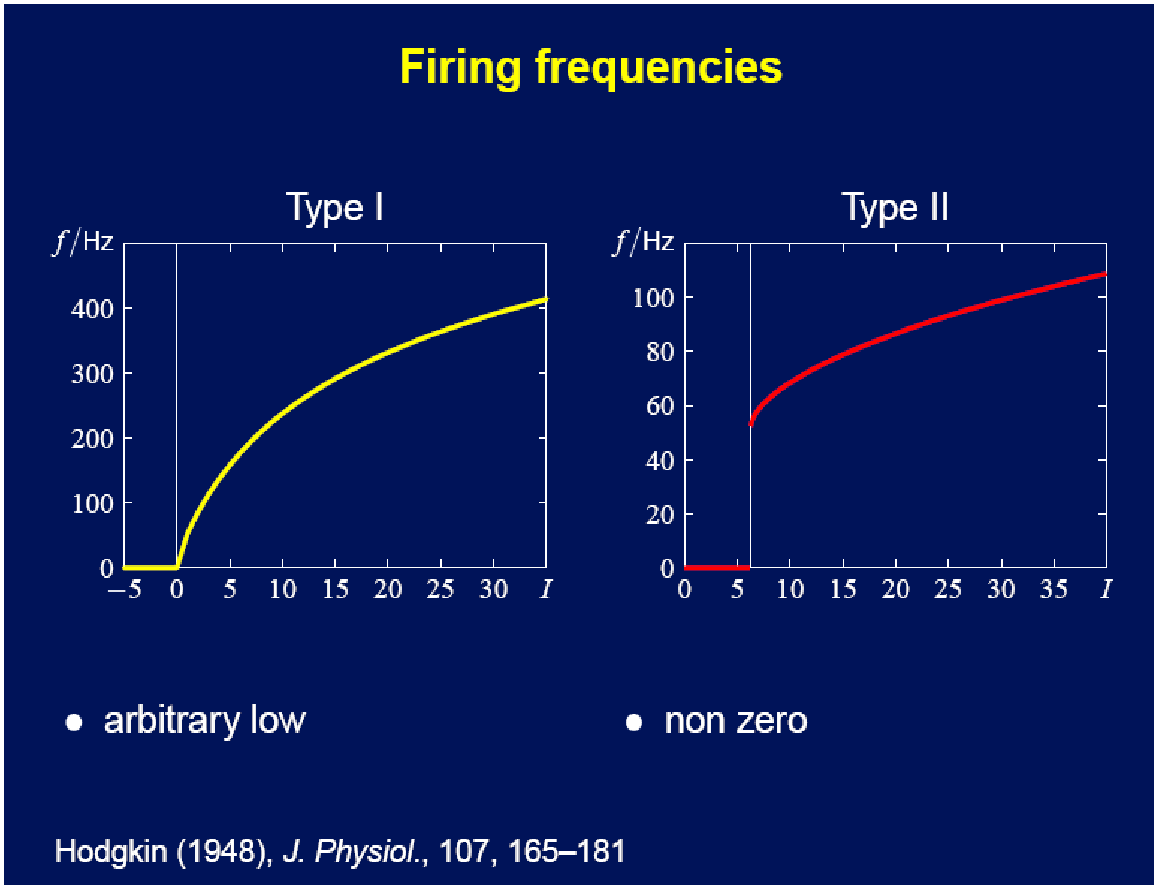


https://aman.ai/cs231n/training-neural-nets-I/

Two main "physiological" firing functions

x axis: input current

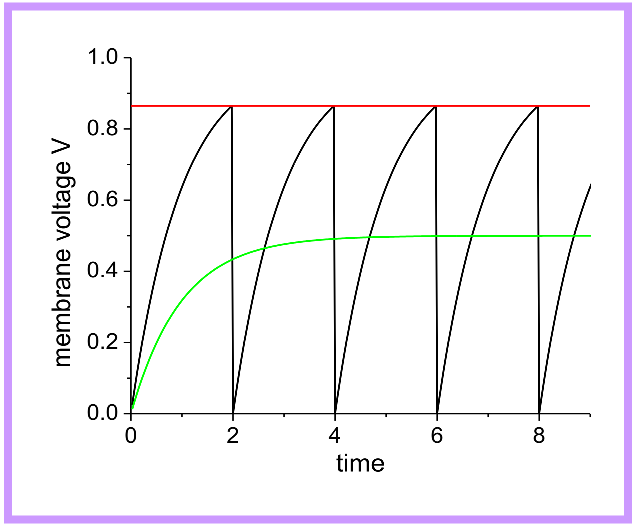
y axis: (steady state) mean firing rate

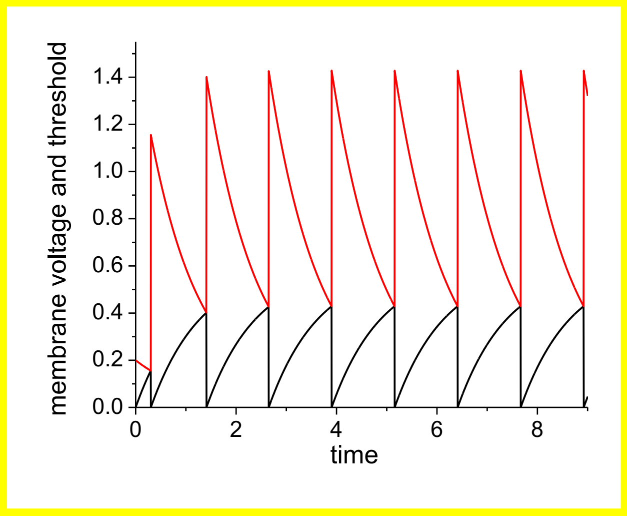


Neuron: fires when voltage meets threshold

Leaky Integrate-and-Fire model with adaptive threshold ("generalized LIF")

Leaky Integrate-and-Fire model:





Ex: Leaky integrate-and-fire dynamics inside a recurrent network

$$\frac{d}{dt}V_i(t) = \mu - V_i(t) + \sqrt{2D_i}\xi_{i,bg}(t)$$
$$+\beta I_i(t) + \frac{g}{N}\sum_j K_{\tau_d} * y_j(t)$$

Internal noise and other background noise

External input

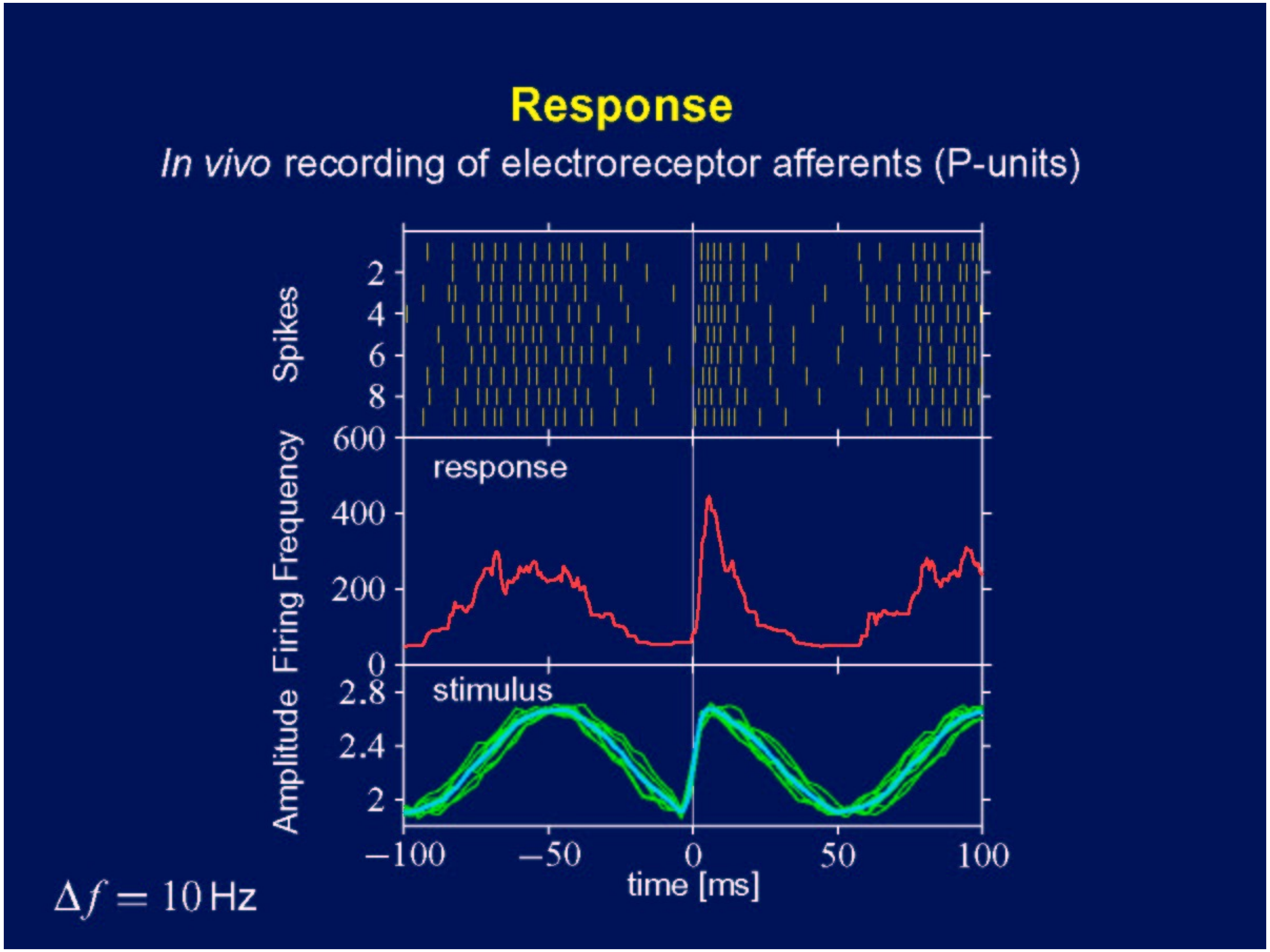
Spike train from neuron j

Feedback response kernel

Noises are “additive”

“Adaptation” signals “change”

→ Encodes time derivative of input: high-pass property



Good model:
Leaky Integrate-and-Fire + Adaptation + Gaussian white noise

$$\begin{aligned}\dot{v} &= -v + \mu - a + \sqrt{2D}\xi(t), \\ \dot{a} &= -a/\tau + A \sum_{t_j \in \mathcal{T}} \delta(t - t_j),\end{aligned}$$

$$v(t) = \Theta_0 \;\; \Rightarrow \;\; t_i \doteq t \;,\; i \rightarrow i+1 \;,\; \text{and } v(t^+) = v_R$$

Some theoretical work on ISI correlations in single neurons and networks
(series of papers by Schwalger + Lindner)

Escape-time process is non-Markovian - Evolution of correlations
(Braun, Thul, Longtin PRE2017)

Part 2: Adaptation produces sequential correlations + precise temporal firing

- Correlations increase regularity of stochastic firing process

→ regularity reduces noise, improves signal-to-noise ratio

→ similar to what a refractory period does

Adaptation is a latent variable: optimal representation of information in individual neurons (Nesse, Maler, Longtin, PNAS 2010)

Adaptation increases number of network firing patterns (in prep)

$$\tau_V \frac{\mathrm{d}V}{\mathrm{d}t} = -V + R \cdot [I(t) - A]$$

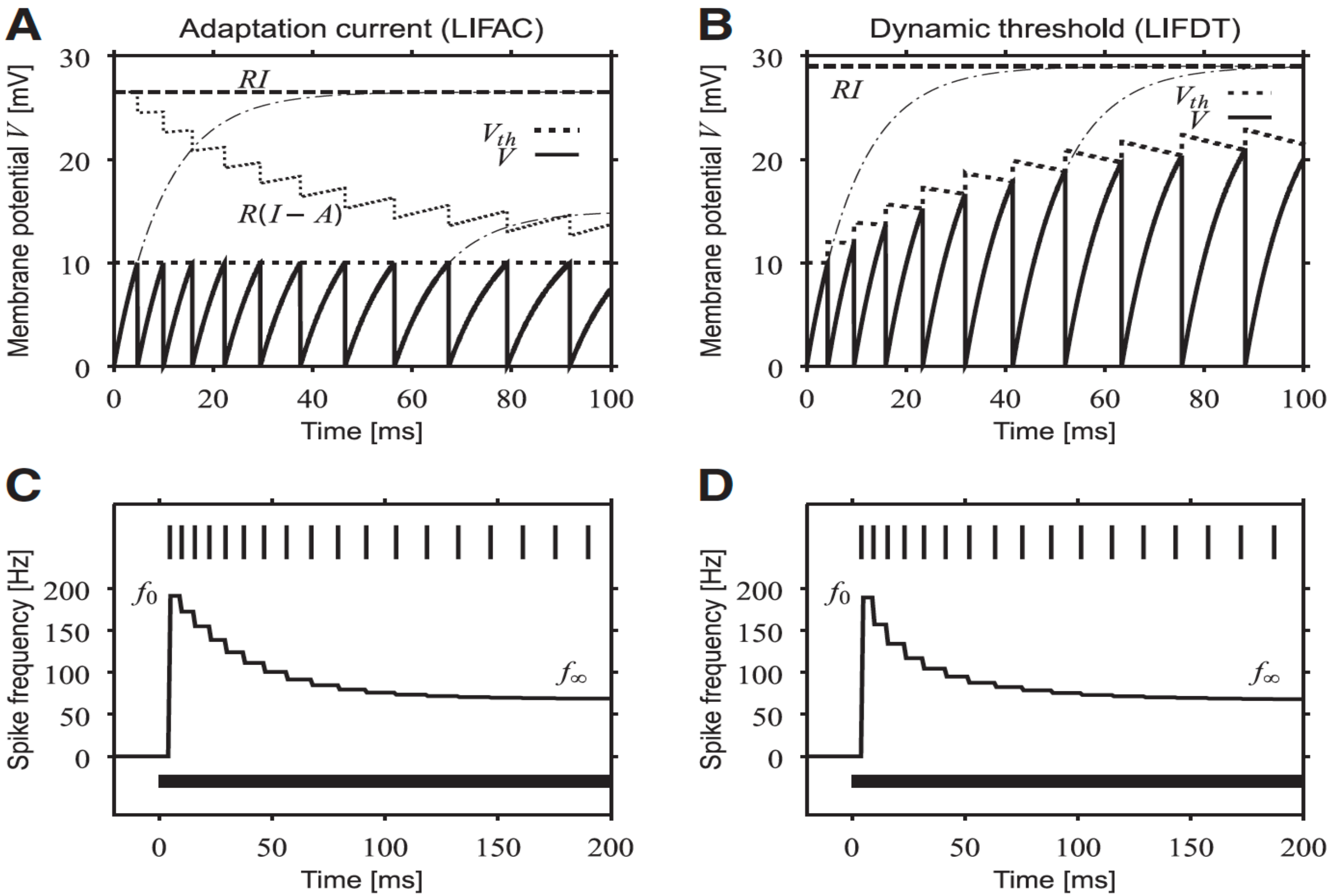
$$\tau_A \frac{\mathrm{d}A}{\mathrm{d}t} = -A + \Delta\ \delta(t - t_i)$$

firing occurs when $V(t) = V_{thr}$ with V_{thr} is fixed

$$\tau_V \frac{\mathrm{d}V}{\mathrm{d}t} = -V + R \cdot I(t)$$

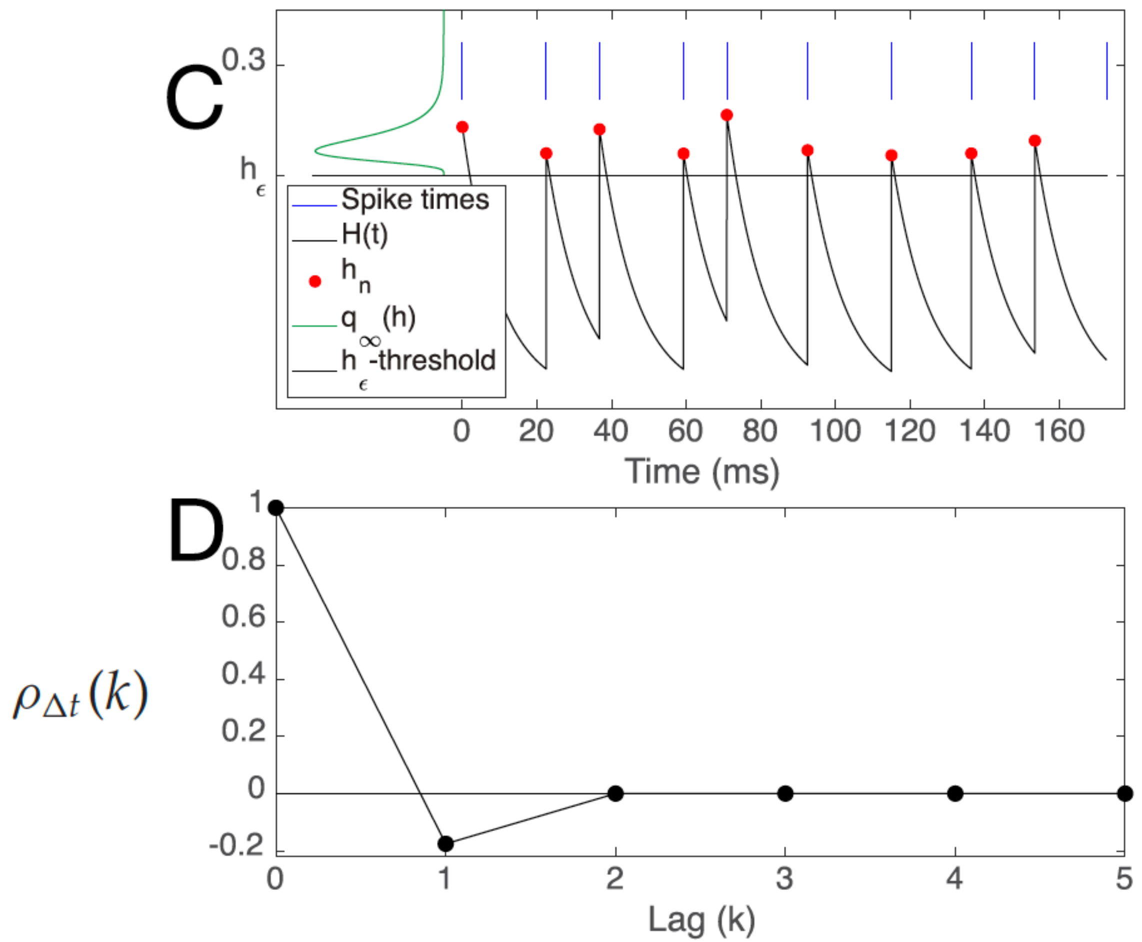
$$\tau_A \frac{\mathrm{d}A}{\mathrm{d}t} = -A + V_{th} + \Delta\ \delta(t - t_i)$$

firing occurs when $V(t) = A(t)$



Benda, Maler, Longtin, J. Neurophysiol. 2010

Correlations between firing intervals



Adaptation H(t)

Autocorrelation of successive intervals between spikes

Nesse, Maler, Longtin, PNAS 2010; NECO 2021

Serial correlations between Interspike Intervals

do not imply

Serial correlations between adaptation states giving rise to these intervals

Adaptation states are quasi-independent:
→ enhanced encoding properties

conditions for quasi-independent adaptation states to give rise to correlated ISI's

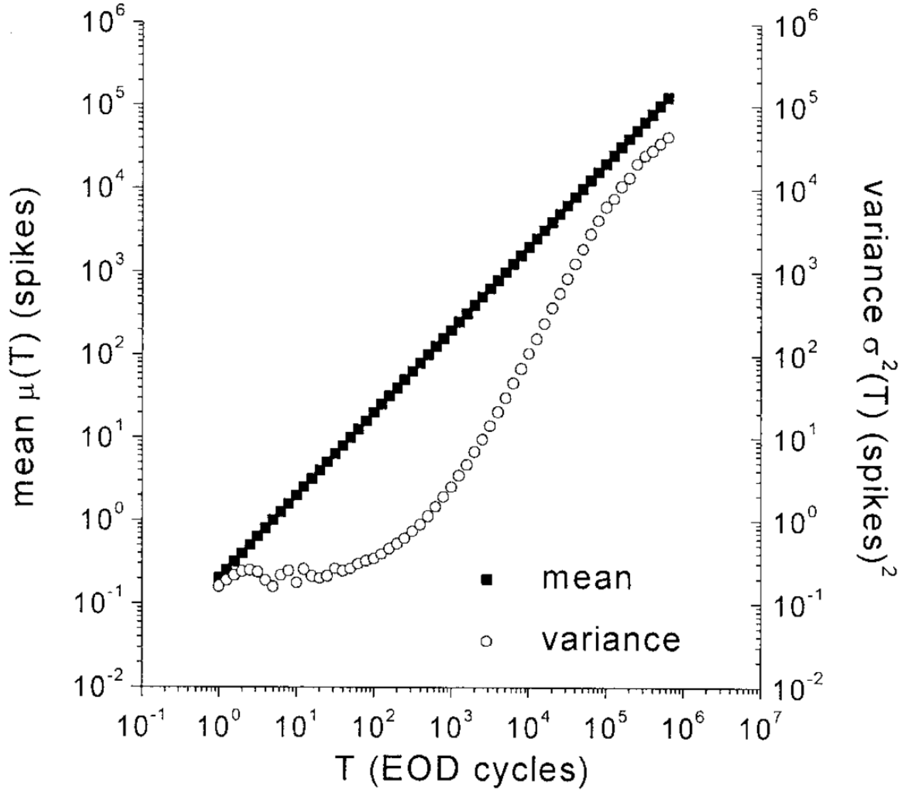
you basically need a mechanism that impedes the next spike, like a refractory period

Proof uses a time coordinate change to an internal "adaptation time"

Nesse, Maler, Longtin, NECO 2021

Adaptation makes computing more temporally precise

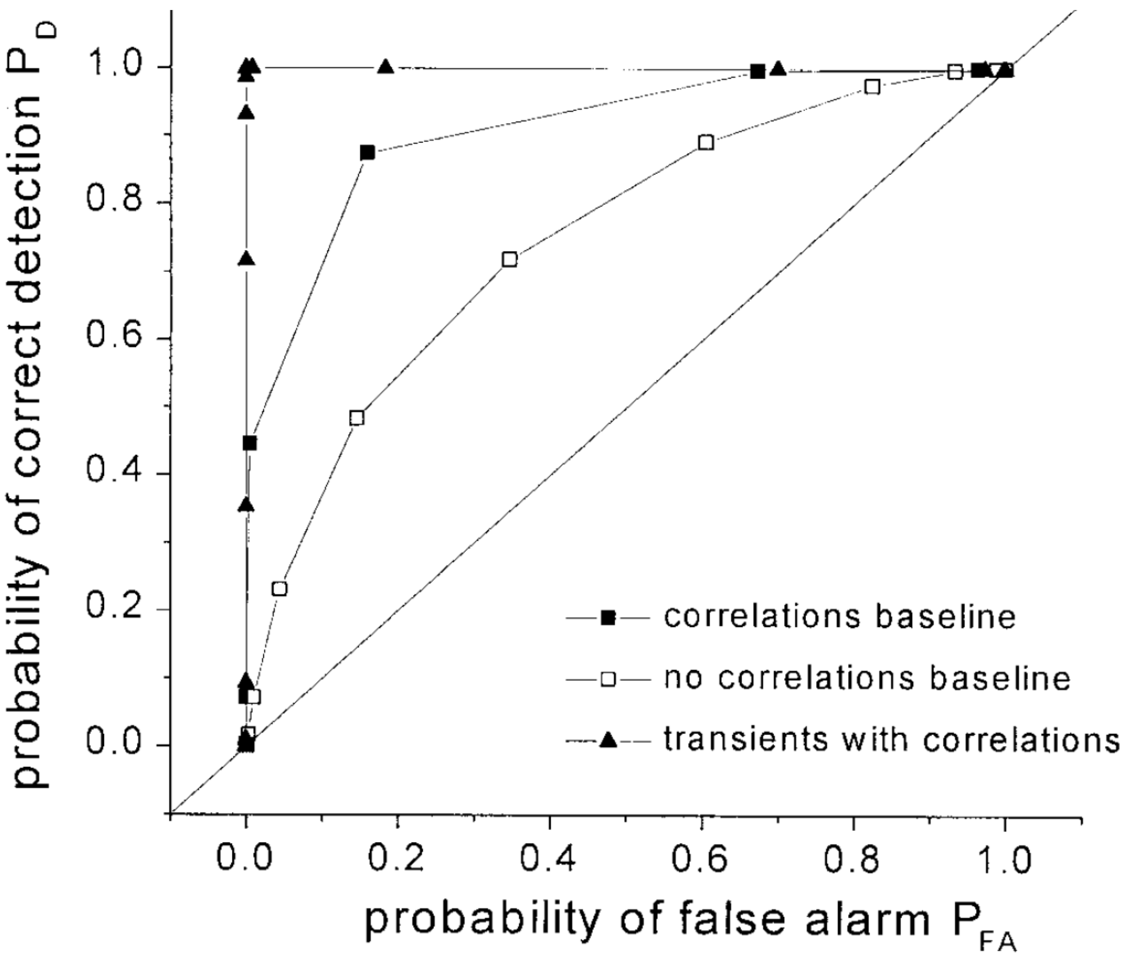
→ based on mean and variance of spike count during windows of duration T



Deviation from Poisson process is due to correlations:
lower variance implies enhanced information transmission

Chacron, Longtin, Maler, J. Neurosci. 2001

Another point of view: ROC analysis
→ the further from the diagonal, the better



Chacron, Longtin, Maler, J. Neurosci. 2001

$$\tau_V \frac{\mathrm{d}V}{\mathrm{d}t} = -V + R\cdot[I(t) - A]$$

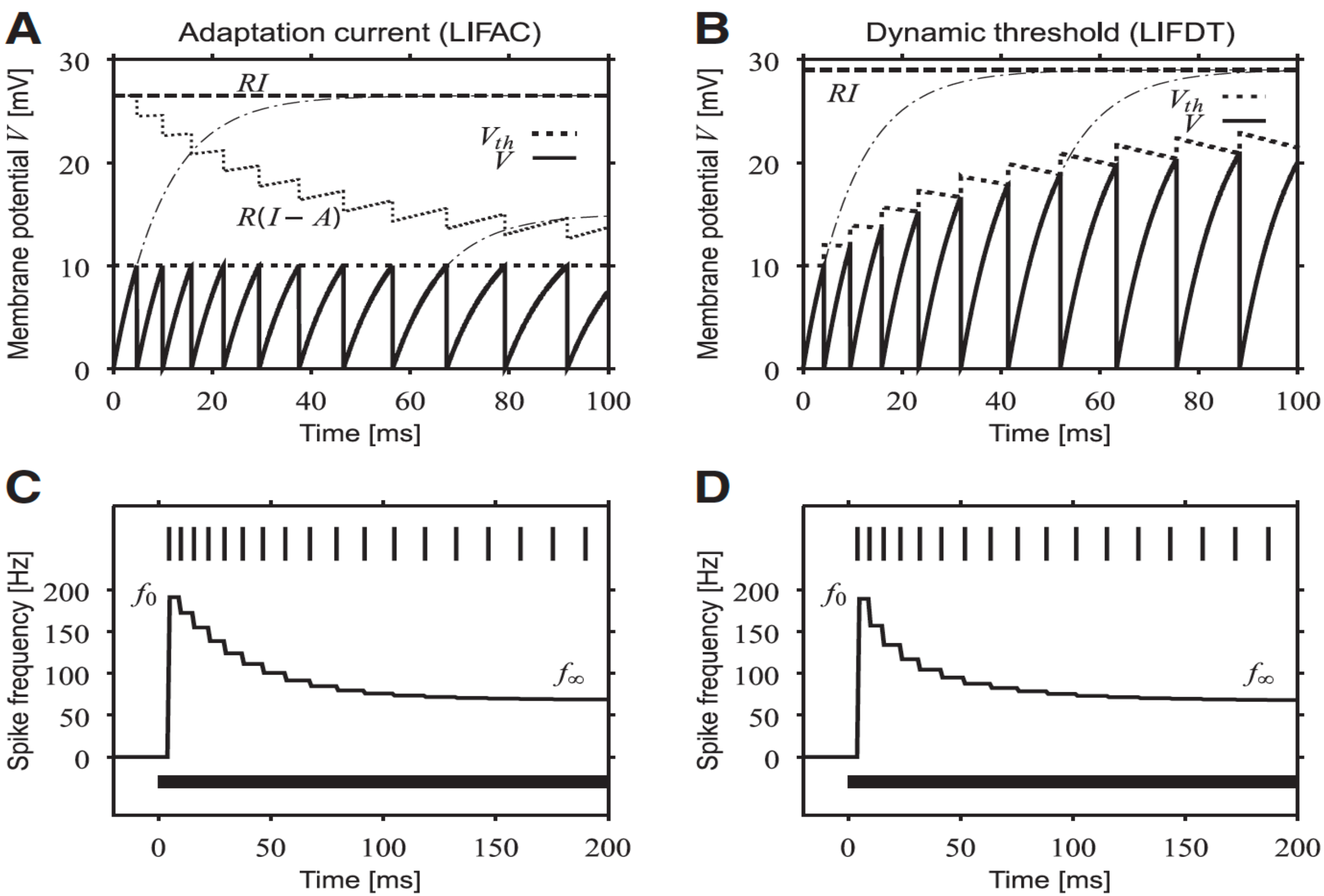
$$\tau_A \frac{\mathrm{d}A}{\mathrm{d}t} = -A + \Delta\ \delta(t - t_i)$$

$$\tau_V \frac{\mathrm{d}V}{\mathrm{d}t} = -V + R\cdot I(t)$$

$$\tau_A \frac{\mathrm{d}A}{\mathrm{d}t} = -A + V_{th} + \Delta\ \delta(t - t_i)$$

firing occurs when $V(t) = V_{thr}$ with V_{thr} is fixed

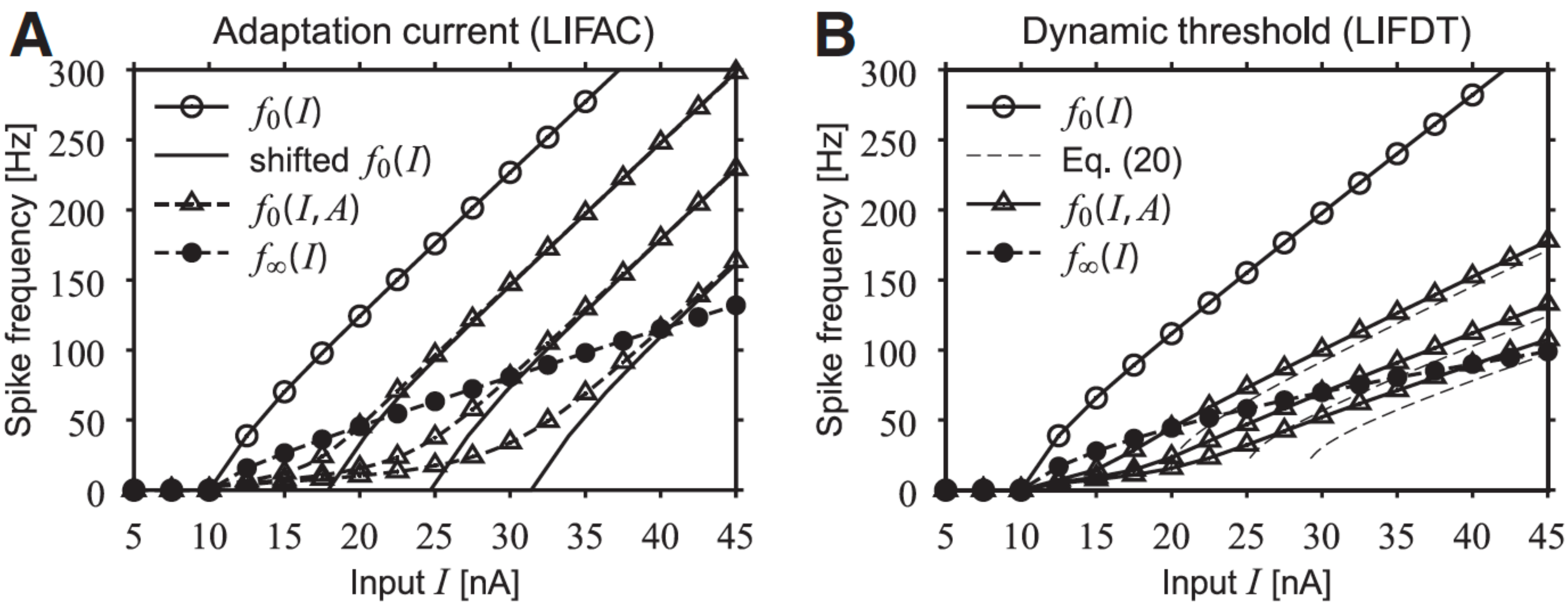
firing occurs when $V(t) = A(t)$



Benda, Maler, Longtin, J. Neurophysiol. 2010

Level of adaptation before stimulus changes the instantaneous firing function:

SUBTRACTION (right shift) VS DIVISION (slope decrease)

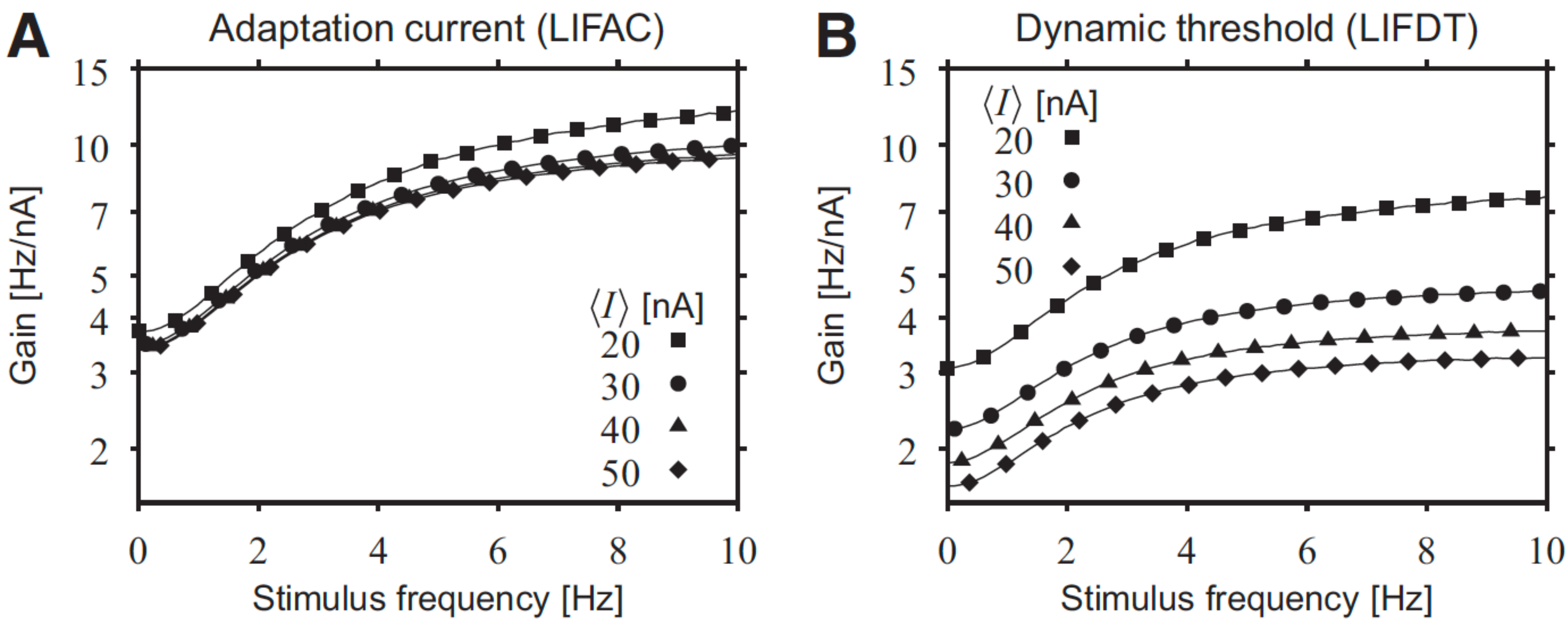


Many conductance-based models (more realistic) behave like LIFAC

Many neurons display both LIFAC and LIFDT

Benda, Maler, Longtin, J. Neurophysiol. 2010

Dependence of transfer function on level of adaptation at time of stimulus



Benda, Maler, Longtin, J. Neurophysiol. 2010

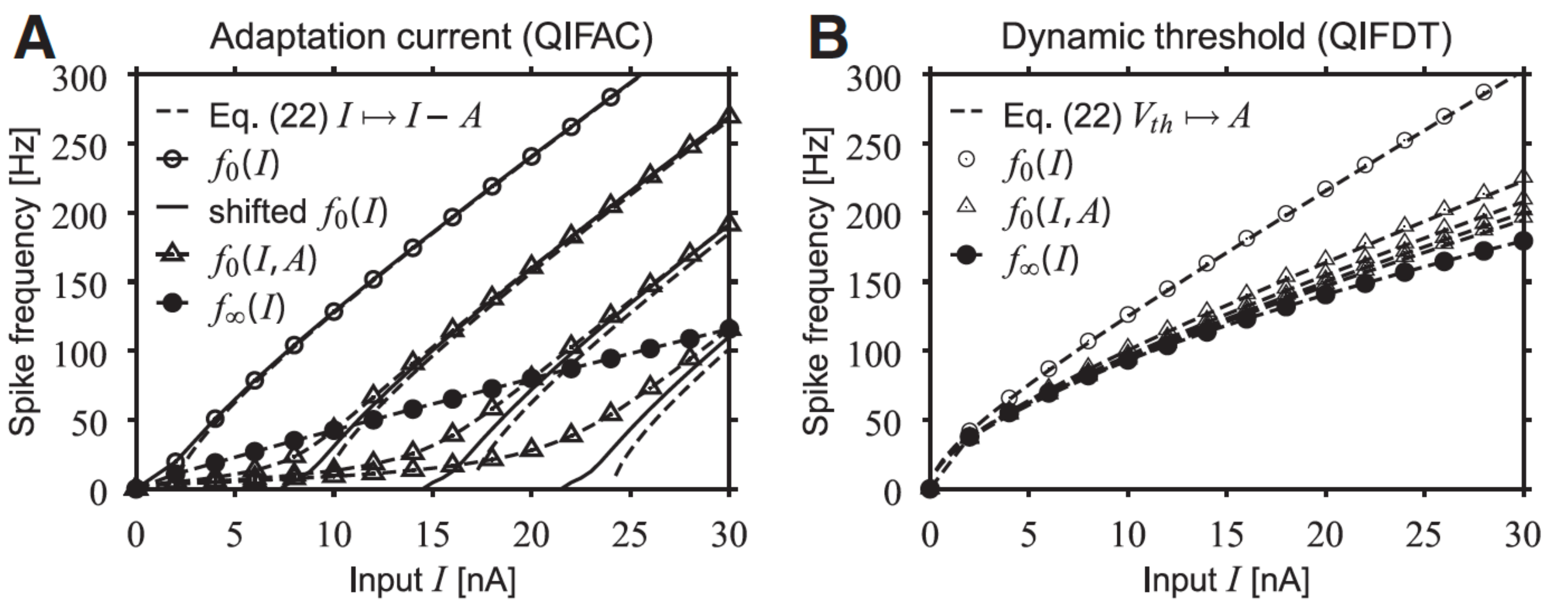
Quadratic integrate-and-fire model
“type 1 excitability”

$$\tau_V \frac{\mathrm{d}V}{\mathrm{d}t} = \frac{V^2}{2\Delta_T} + RI$$

Firings ?

Subthreshold: V goes beyond unstable fixed point

Suprathreshold: V crosses a fixed phase on a limit cycle



Benda, Maler, Longtin, J. Neurophysiol. 2010

Exponential integrate-and-fire with adaptive threshold

excellent model for many cells in cortex

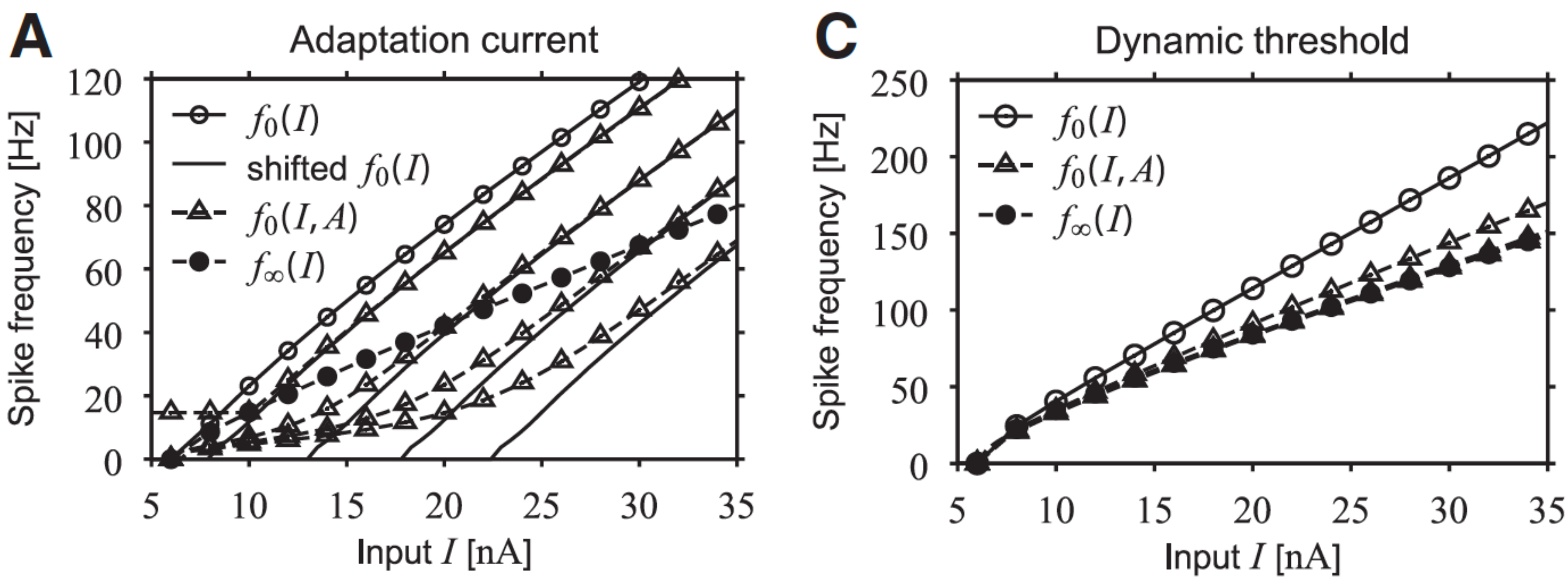
$$\tau_V \frac{dV}{dt} = -V + \Delta_T e^{(V-A)/\Delta_T} + RI$$

$$\tau_A \frac{dA}{dt} = -A + V_T$$

fires when

Subthreshold: V goes beyond unstable fixed point

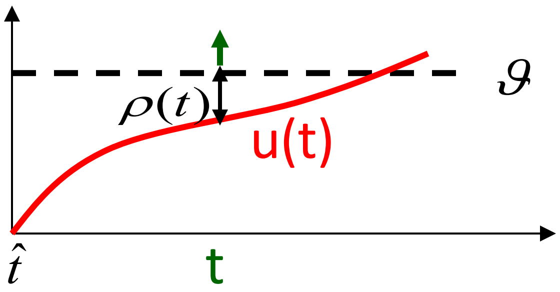
Suprathreshold: V crosses a fixed phase on a limit cycle



Benda, Maler, Longtin, J. Neurophysiol. 2010

Noise models

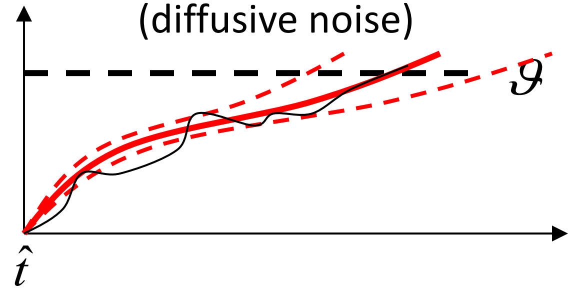
escape process,
stochastic intensity



escape rate

$$\rho(t) = f(u(t) - \vartheta)$$

stochastic spike arrival
(diffusive noise)



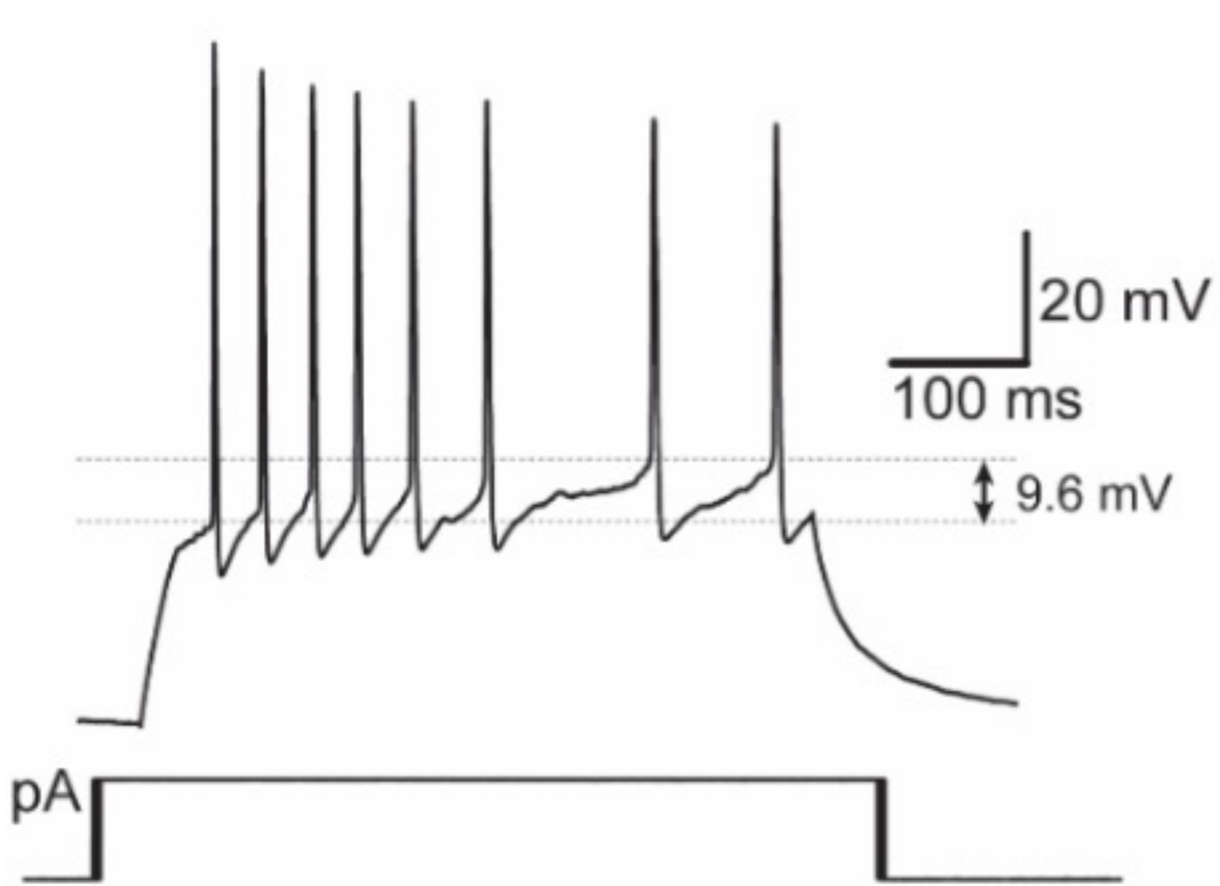
noisy integration

$$\tau \cdot \frac{du_i}{dt} = -u_i + RI + \xi(t)$$

Probability of firing VS. first passage time of stochastic diff. eq.

Gerstner et al. ND book

Adaptation of Mossy cells of the hippocampus Hilus (between DG and CA3)



Trinh, Girardi-Schappo, Harvey-Girard, Beique, Longtin, Maler, J. Physiol. (in press 2023)

Part 3: Adaptation to Compute Time Sequences

1) Context: Timing of encounters with landmarks during navigation

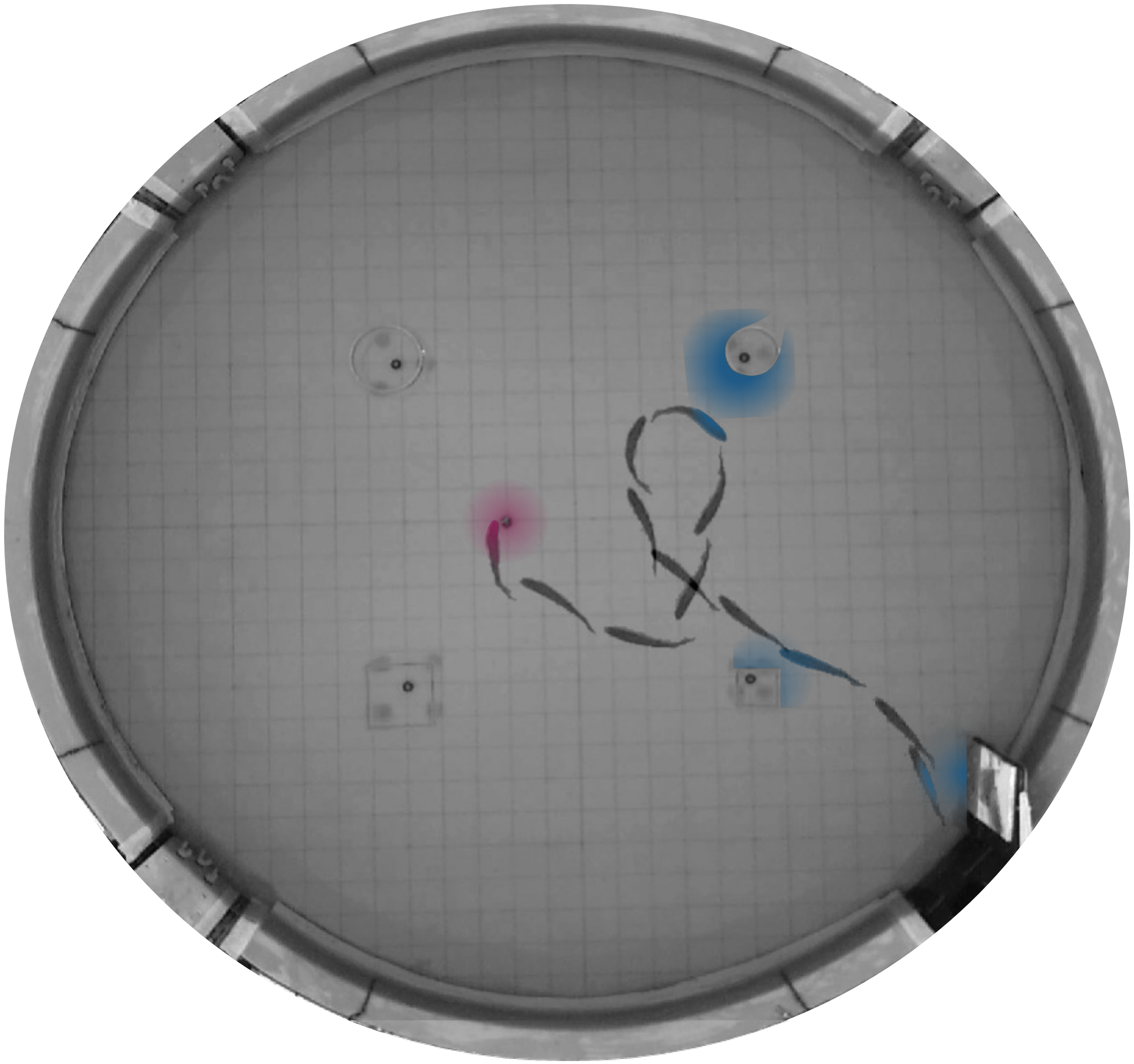
2) Encounters produce bursts of spikes targeting memory circuits

3) Using burst size to infer time since last encounter

4) Using burst size to infer sequence of past encounters

*Wallach, Harvey-Girard, Jun, Longtin, Maler, **eLife 2018**;*
Lafond-Mercier, Wallach, Maler, Longtin (in prep)

Adaptation to represent time between spatial encounters



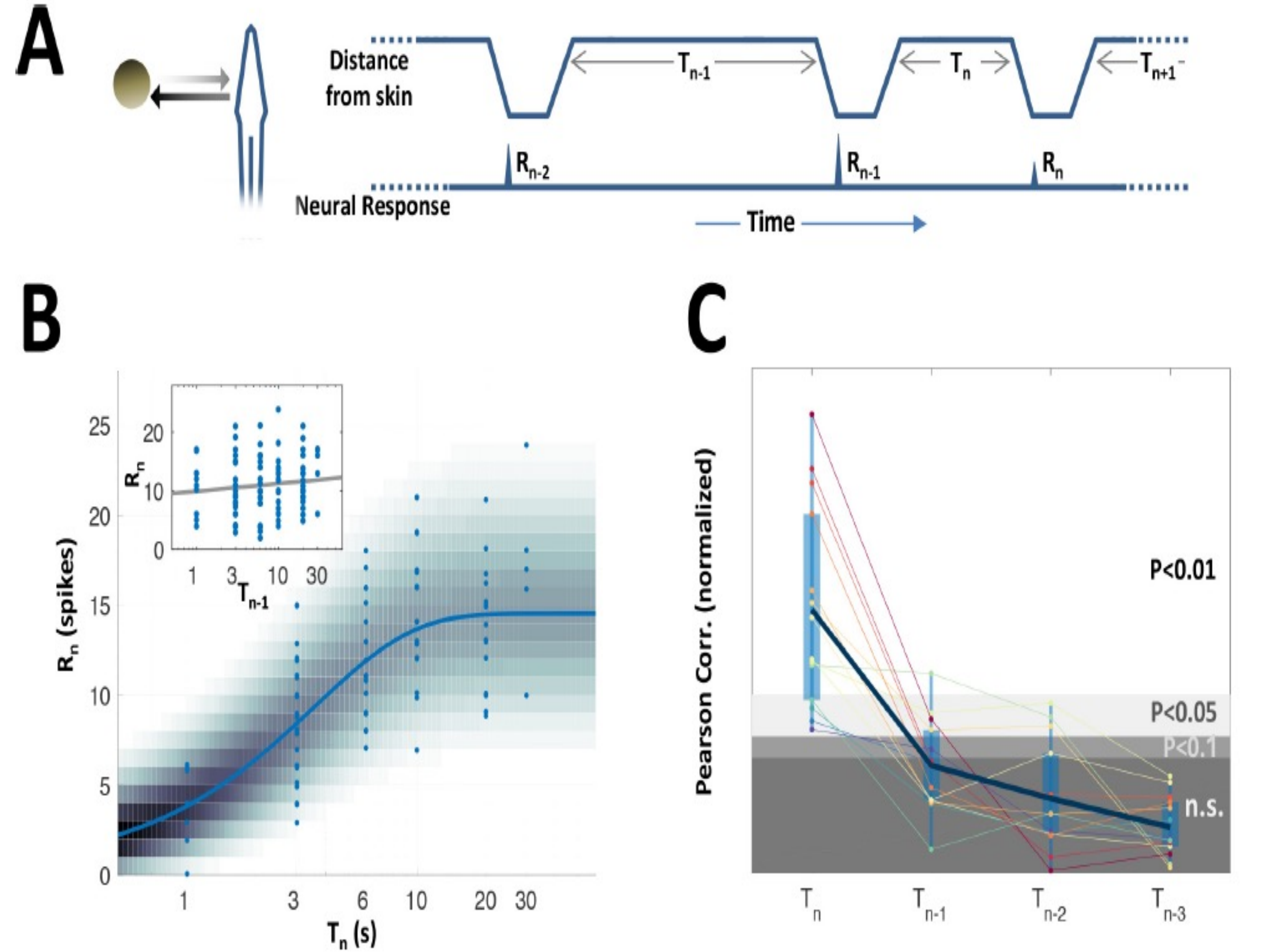
IDEA: Assume neurons can compute time between last encounter (OBJECT 1) and new encounter (OBJECT 2).

Then:

distance between OBJECT 1 and OBJECT 2

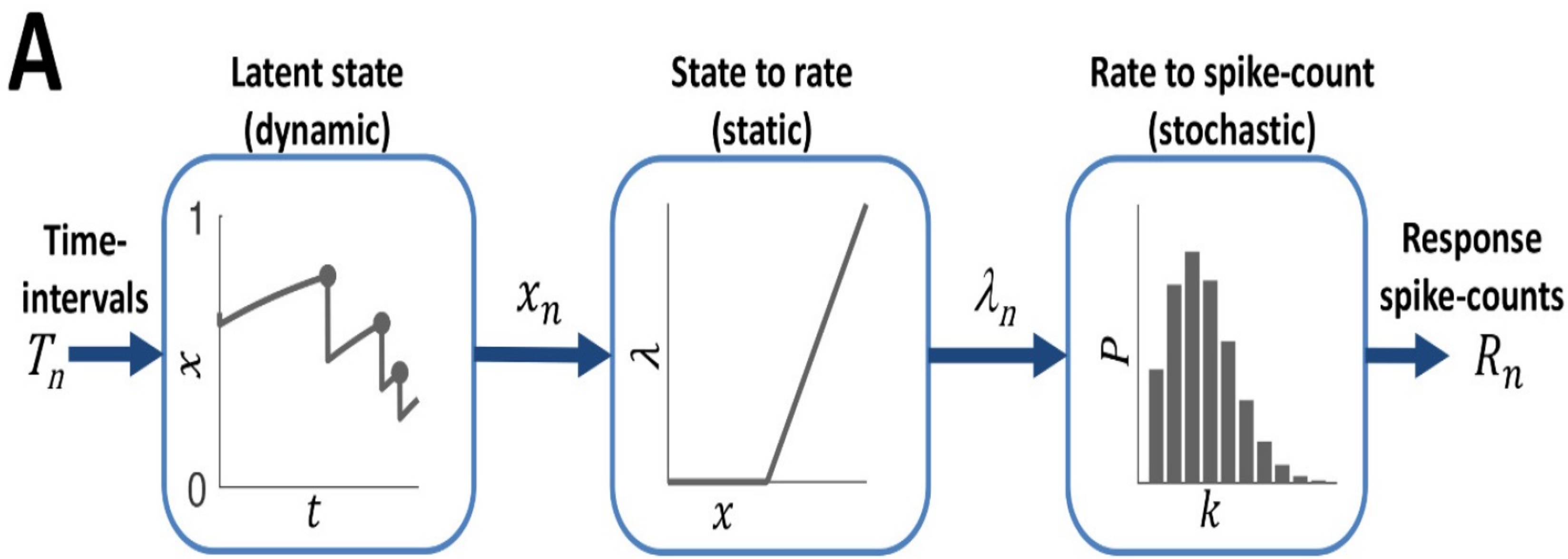
=

(mean) travel speed X elapsed time


A
Distance
from skin
T_{n-1}
T_n
T_{n+1}
R_{n-2}
R_{n-1}
R_n
Neural Response
Time
B
R_n (spikes)
T_n (s)
R_n
T_{n-1}
C
Pearson Corr. (normalized)
P<0.01
P<0.05
P<0.1
n.s.
T_n
T_{n-1}
T_{n-2}
T_{n-3}

Computational model

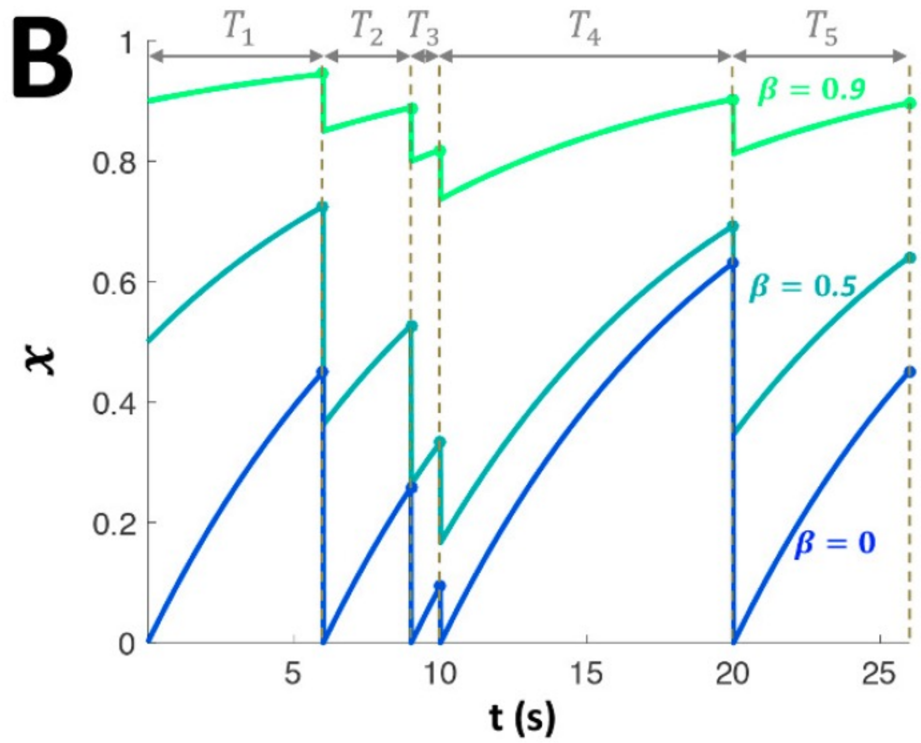
→ Maximum likelihood estimation (MLE) from population; compute Cramer-Rao bound



$$x_n = 1 - e^{-\frac{T_n}{\tau}}(1 - \beta x_{n-1}) \qquad \lambda_n = [a x_n + c]_+ \qquad P(R_n = k) = \left(\lambda_n{}^k e^{-\lambda_n}\right)/k!$$

ReLu

Latent "adaptation" dynamics
50% of the cells have beta > 0 (i.e. longer memory)



Maximum Likelihood Approach

$$L(T_n|\{R_n^j\}) = P(\{R_n^j\}|T_n) = \prod_{j=1}^{N} \frac{\left(\left[a_j\left(1-e^{-\frac{T_n}{\tau_j}}\right)+c_j\right]_+\right)^{R_n^j} e^{-\left(\left[a_j\left(1-e^{-\frac{T_n}{\tau_j}}\right)+c_j\right]_+\right)}}{R_n^j!}$$

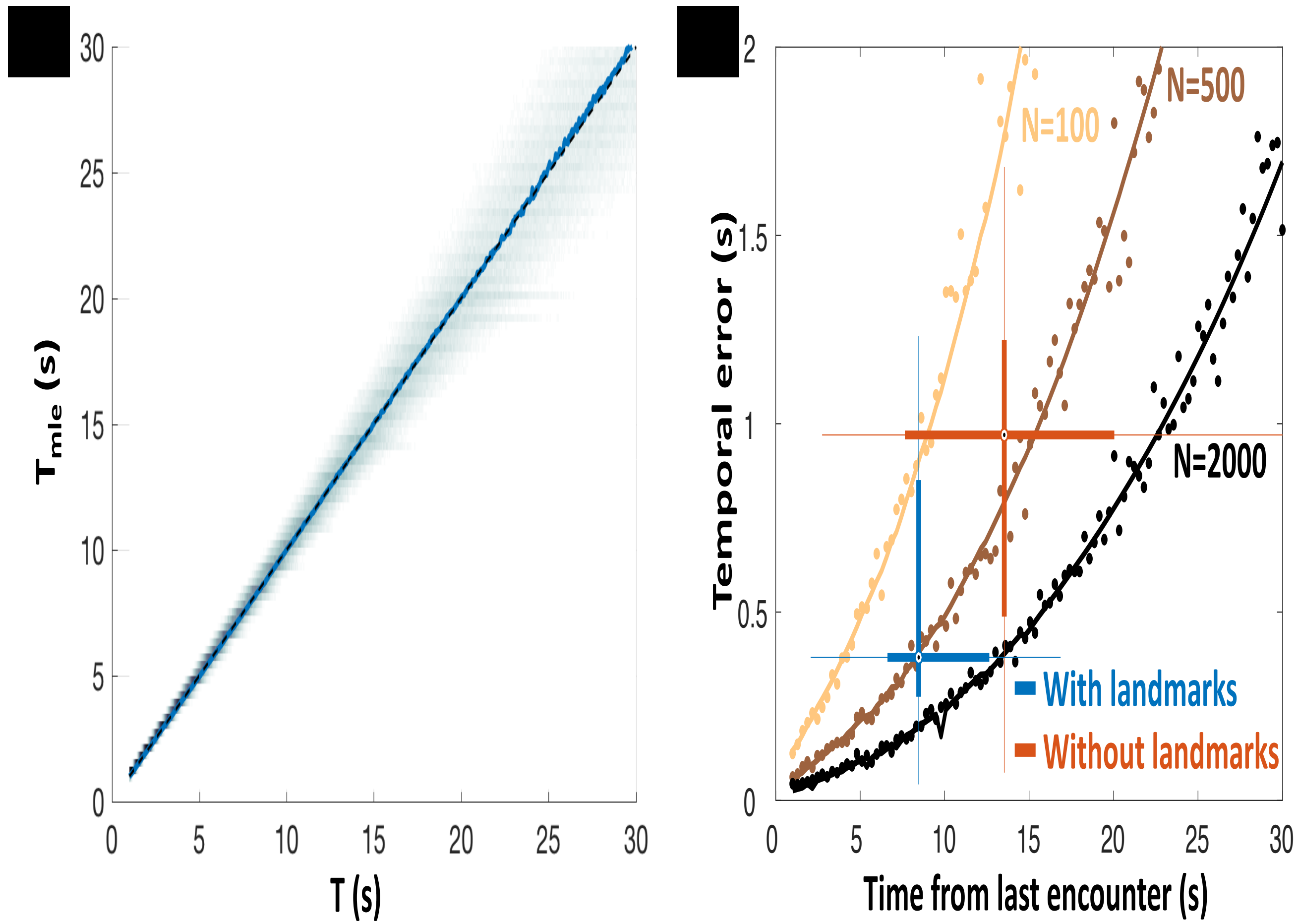
$$T_n^{MLE} = \underset{T_n>0}{\operatorname{argmax}}\left(\sum_{j=1}^{N} R_n^j \log\left(\left[a_j\left(1-e^{-\frac{T_n}{\tau_j}}\right)+c_j\right]_+\right) - \left[a_j\left(1-e^{-\frac{T_n}{\tau_j}}\right)+c_j\right]_+\right)$$

This maximum was found numerically for each generated time interval.

For homogeneous population, assuming rates > 0 :

$$T_n^{MLE} \cong \tau\left(\frac{\frac{1}{N}\sum_{j=1}^{N} R_n^j - c}{a}\right)$$

MODEL: 500 cells suffice to encode time and account for experimental error (good, since 9000 are available!)



Representation of more than one past interval: non-trivial !

$$x_n = 1 - e^{-\frac{T_n}{\tau}}(1 - \beta x_{n-1})$$

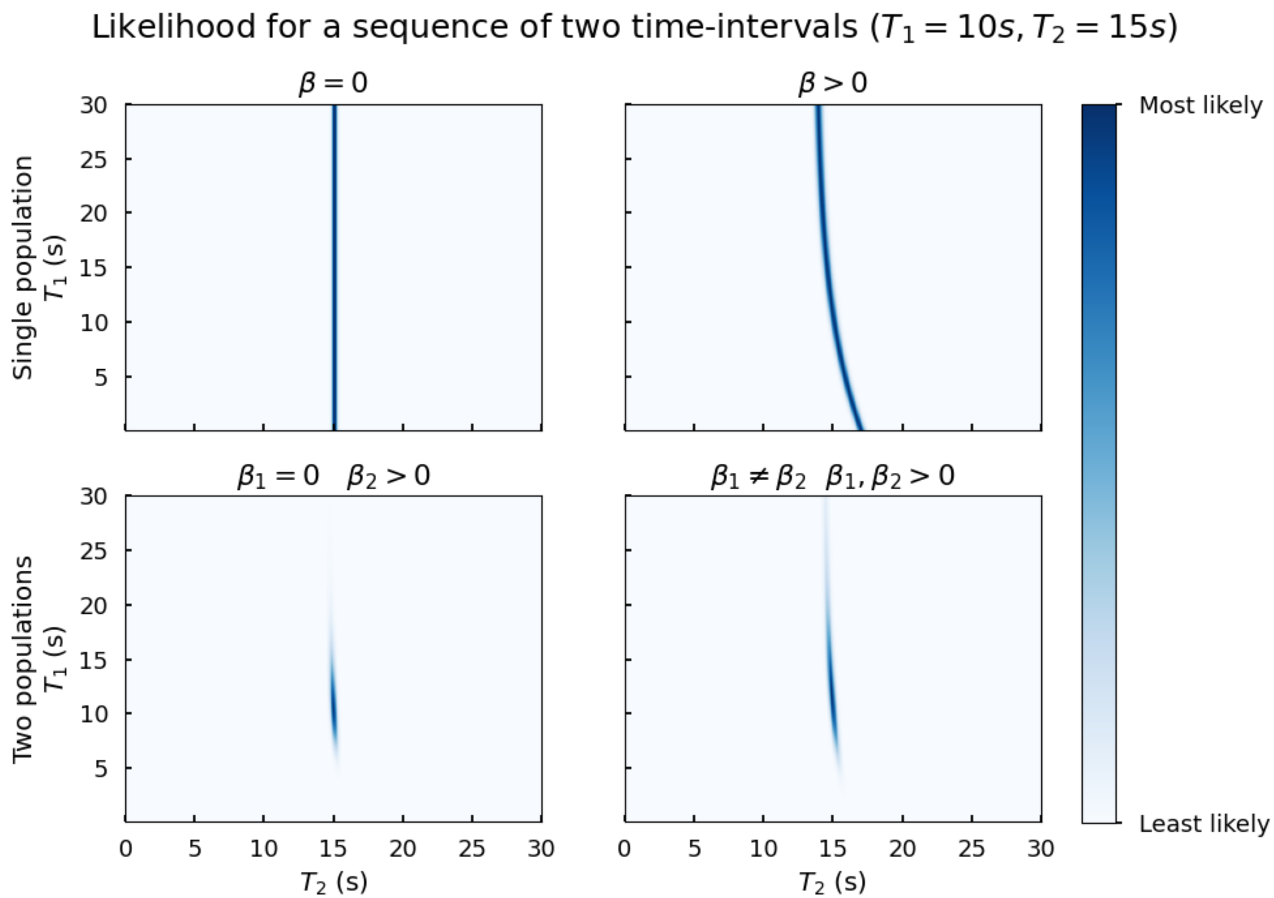
$$L\left(T_n \middle| \left\{R_n^j\right\}\right) = P\left(\left\{R_n^j\right\} \middle| T_n\right) = \prod_{j=1}^{N} \frac{\left(\left[a_j\left(1 - e^{-\frac{T_n}{\tau_j}}\right) + c_j\right]_+\right)^{R_n^j} e^{-\left(\left[a_j\left(1-e^{-\frac{T_n}{\tau_j}}\right)+c_j\right]_+\right)}}{R_n^j!}$$

This maximum was found numerically for each generated time interval

Computational model for 2 intervals:
Different values of β (i.e. 2 or more populations) are needed

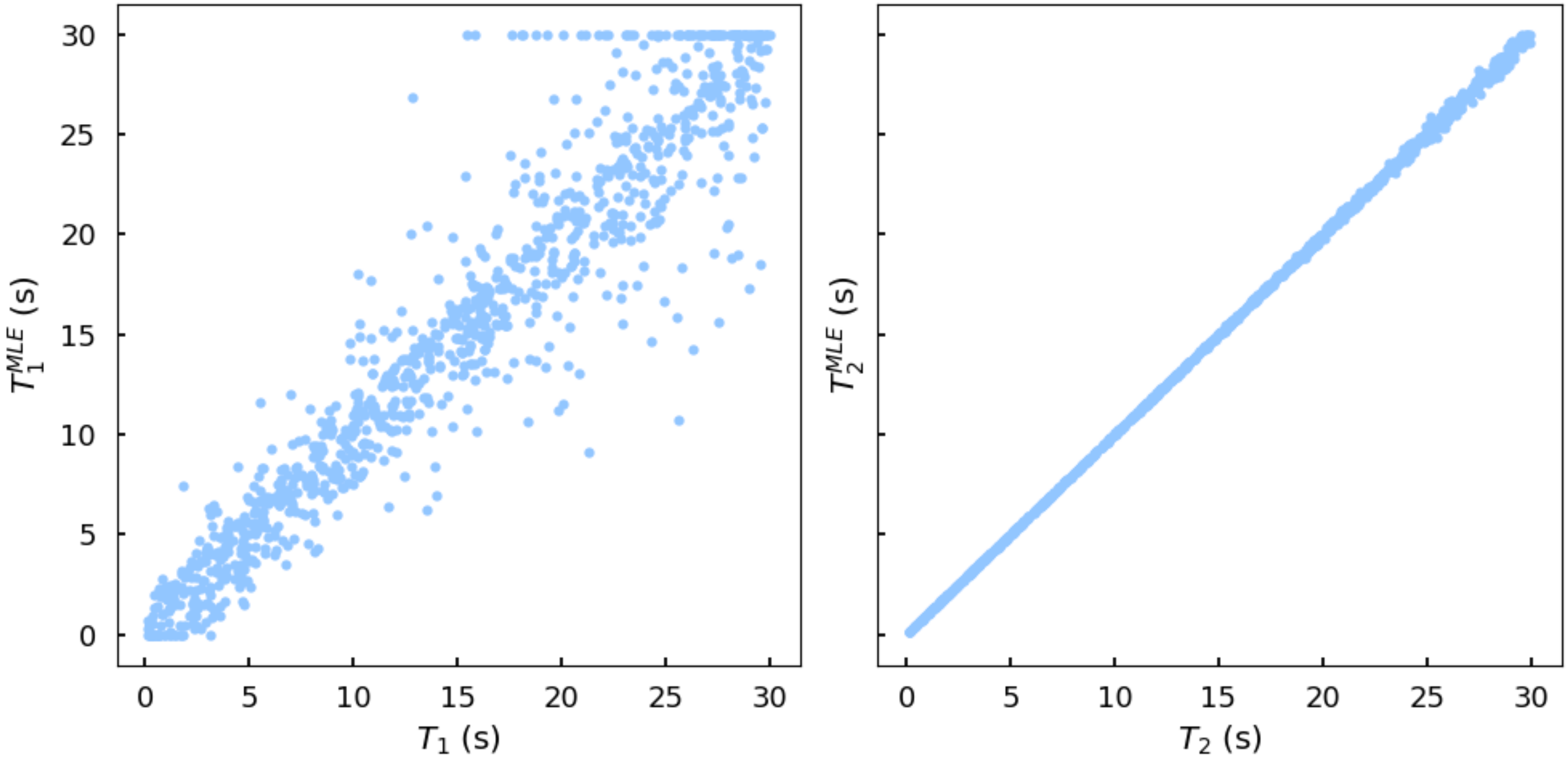
1

2



Computational model

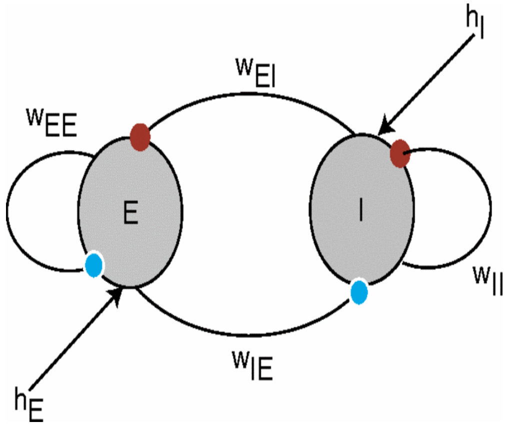
Time interval estimations using 10,000 cells:
$\beta_1 = 0$ (5,000 cells) and $\beta_2 > 0.39$ (5,000 cells)



Conclusion Part 3

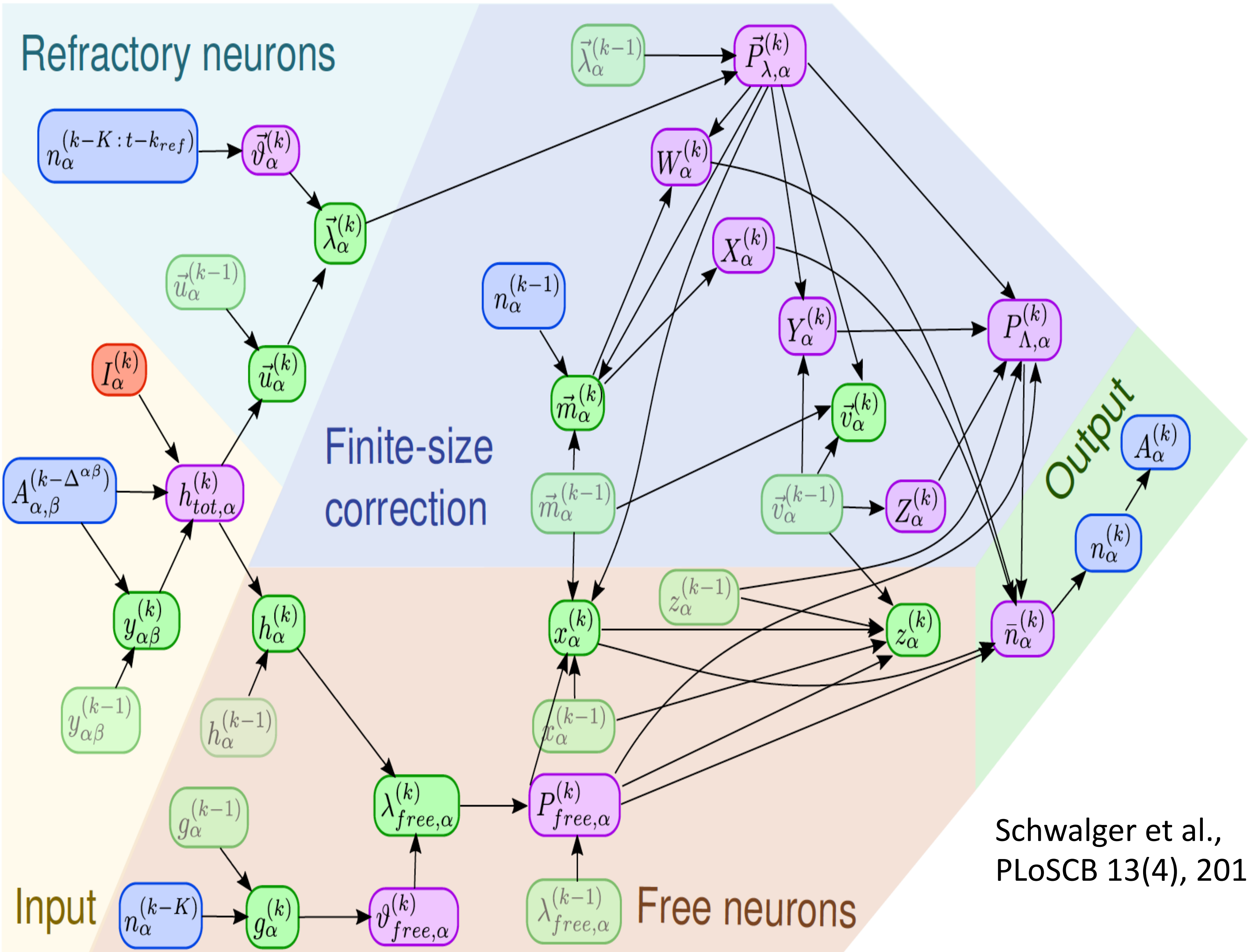
- Adaptation acting on long time scales (beyond 0.5 sec) can represent sequence of intervals between encounters
- This could support path learning and path integration
- Complements RNNs and LSTMs for sequence learning?

PART 4: INFERENCE OF CIRCUITS



- Two principal types of neuron: Excitatory (E) and Inhibitory (I)
- Autonomous rhythms via Synchronization of E and I
- Information stored in the timing of rhythms: not yet including in deep learning frameworks
- Much deterministic modeling
- REAL RHYTHMS ARE STOCHASTIC/CHAOTIC: HOW TO INFER MODEL FROM DATA?

→ stochastic dynamical systems AND machine learning



Schwalger et al.,
PLoSCB 13(4), 2017

What if you don't have the right model

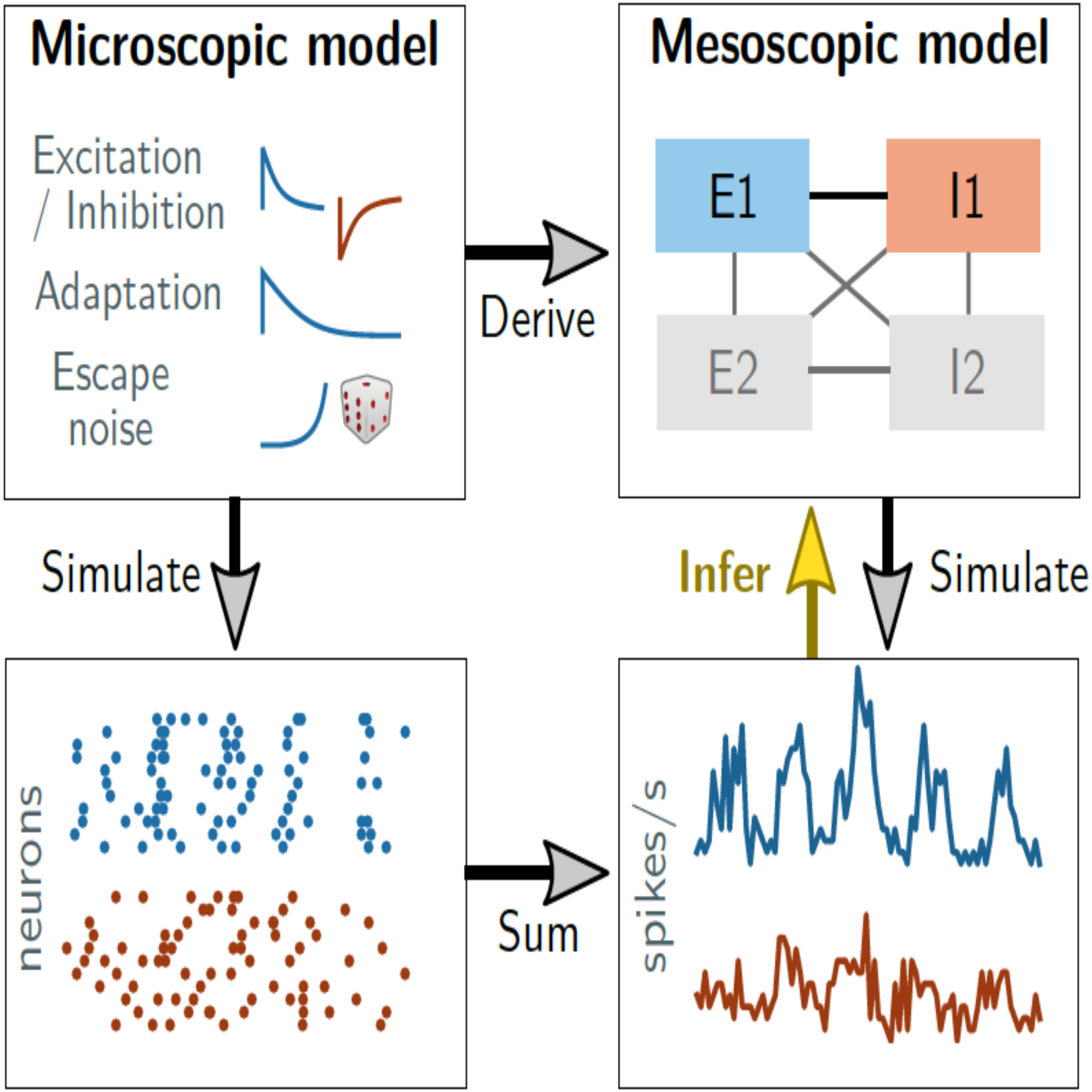
Estimate an "effective" model

See how far that gets you…

Inference of a mesoscopic population model from population spike trains

Alexandre René (U. Ottawa) + Jakob Macke (U. Tuebingen)

A. René, A. Longtin, J. Macke, Neural Comput. 2020 (article)


Microscopic model
Excitation
/ Inhibition
Adaptation
Escape
noise
Mesoscopic model
E1
I1
E2
I2
Derive
Simulate
Infer
Simulate
neurons
Sum
spikes/s

IDEA:

You have microscopic data, e.g. individual spike time from many cells in different populations

Use these data to construct population responses

Need a theory relating population responses (mesoscopic level) to single cell responses (microscopic level)

Need NOISE to write a likelihood of observing the data.

Minimize this likelihood to fit the parameters of the mesoscopic model

Add e.g. adaptation or other phenomena of interest if theory exists.

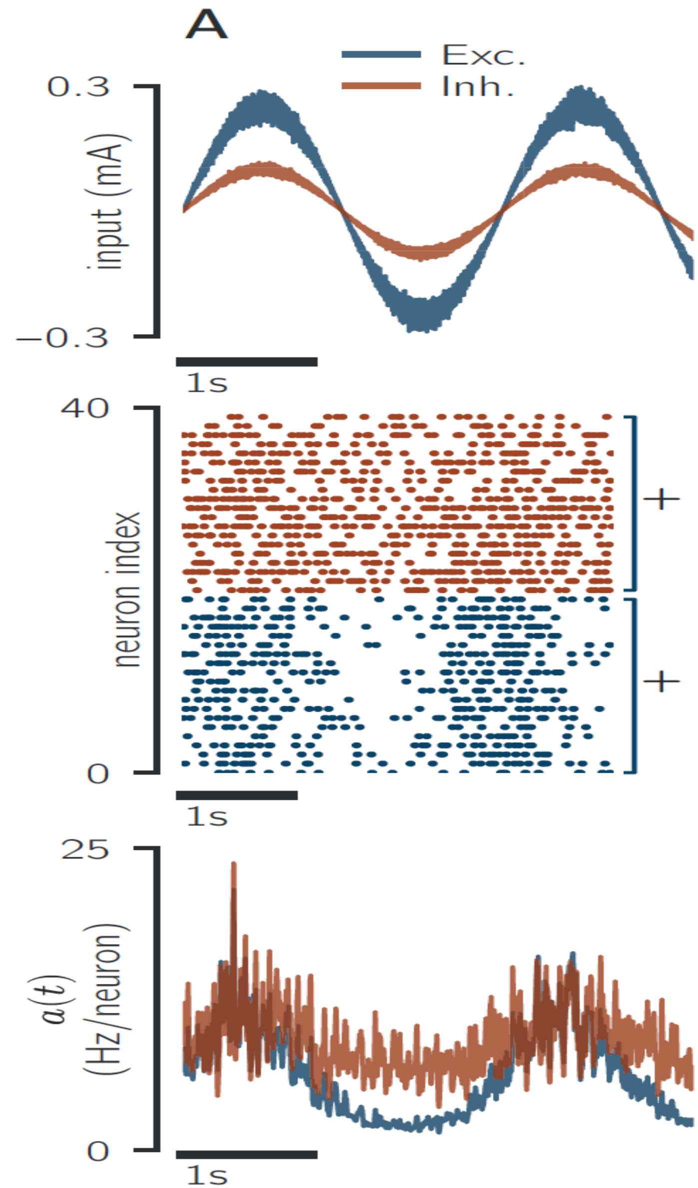
Generating the microscopic data: 2 population model (E-I)

→ noisy sinusoidal input

$$I_{\text{ext}}(t) = B\sin(\omega t)\cdot(1 + q\xi(t))$$

Raster plots of generated spikes

Inference algorithm sees only summed activity

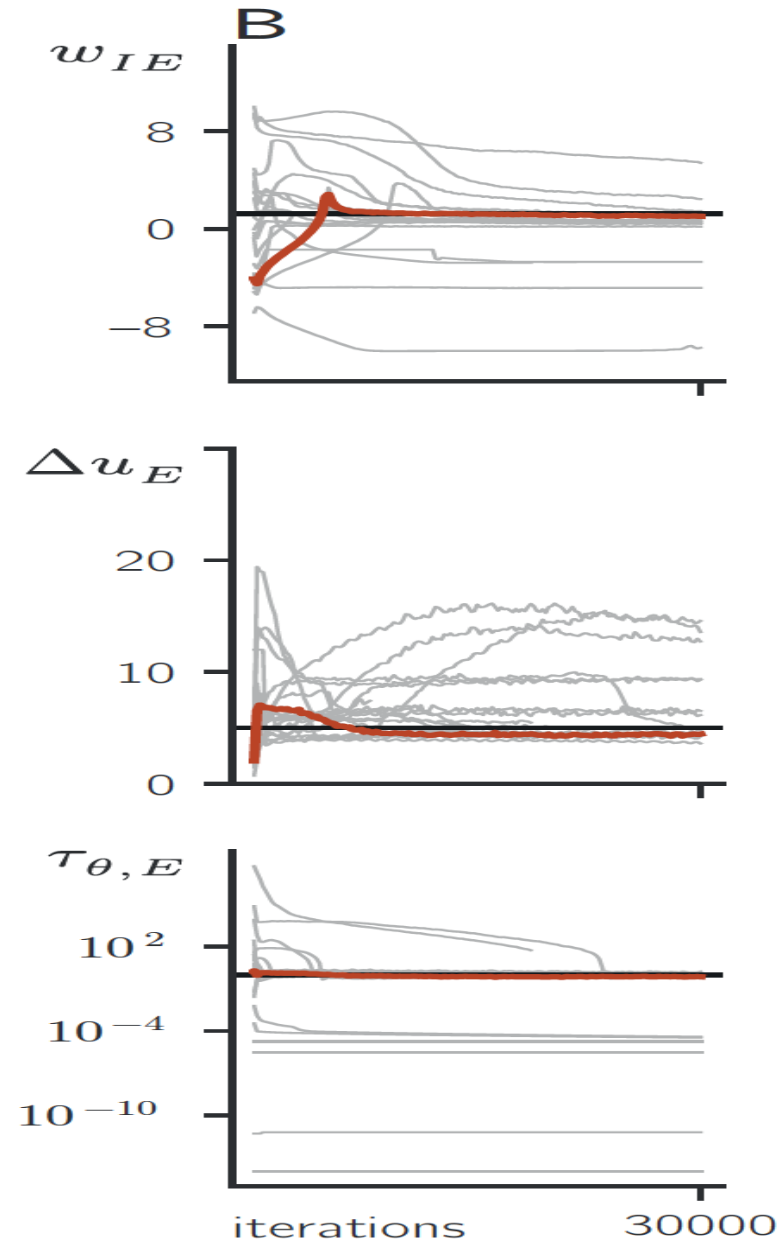


Run 25 fits, keep the one with highest likelihood (in red)

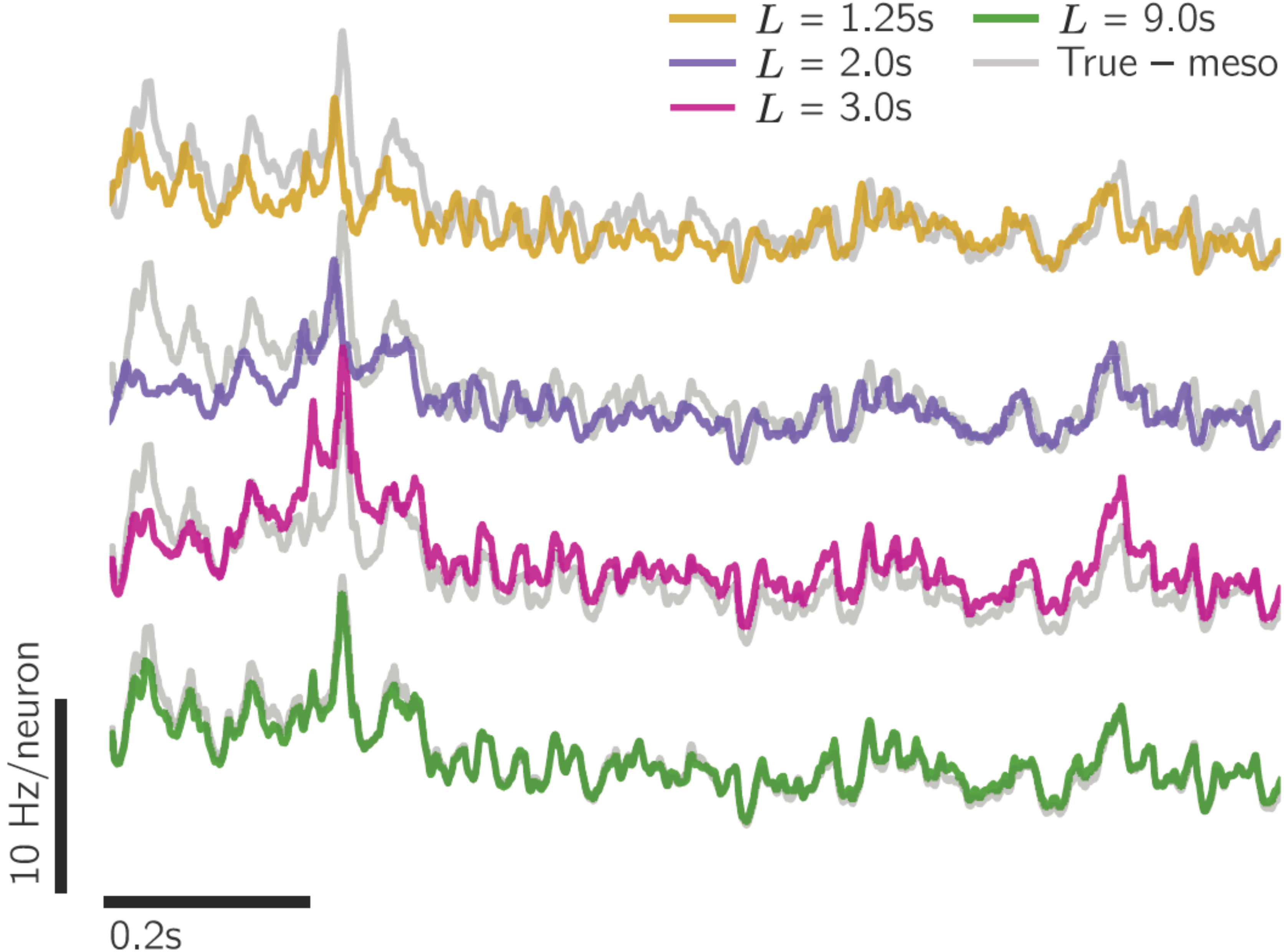
Black: mesoscopic model groundtruth, based on microscopic model groundtruth

Grey: different fits

Red: best fit



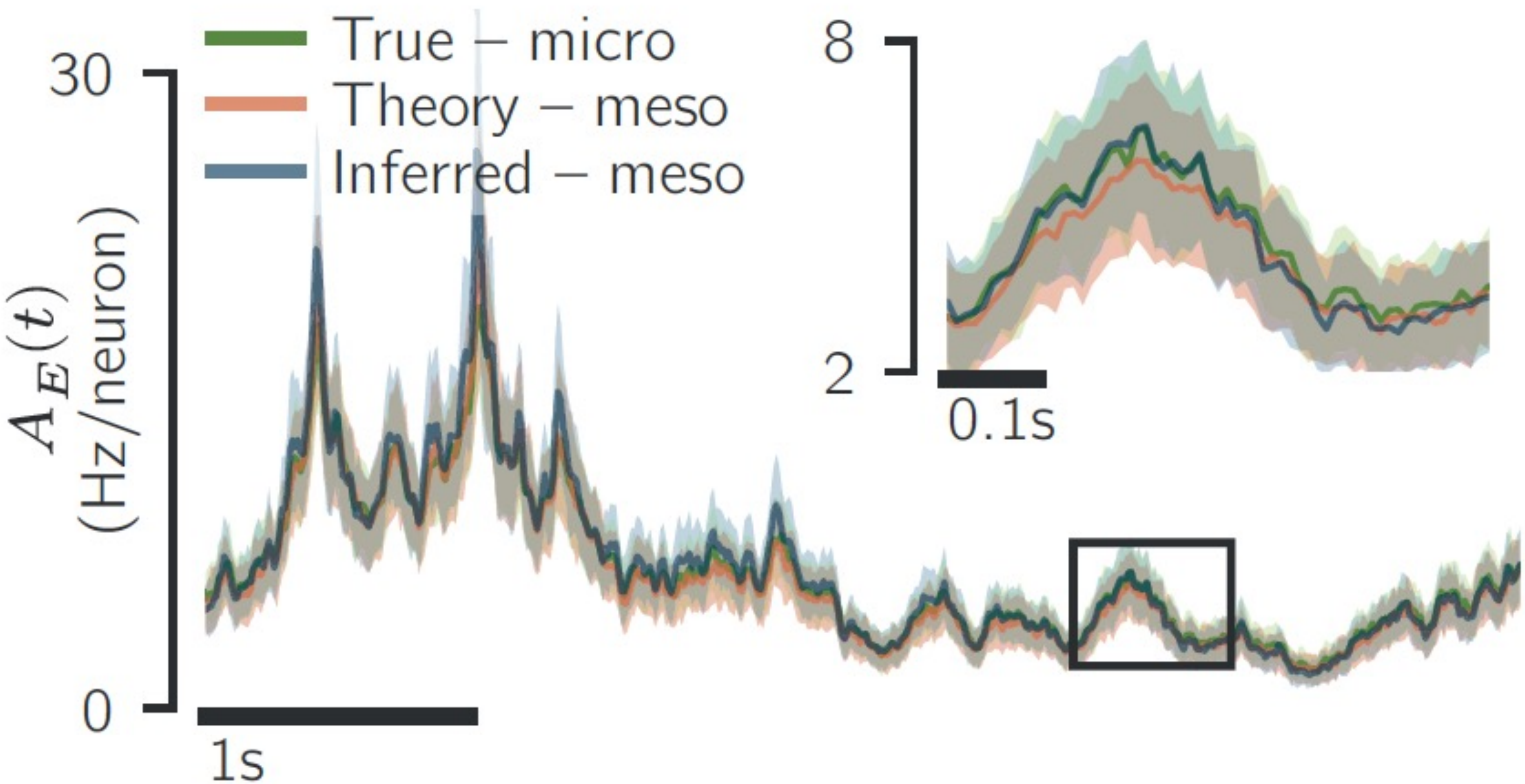
Convergence after about 20,000 spikes



Ability to generalize:

train on one kind of input, test on another

Here: train on noisy sine, testing on lowpass-filtered noise (frozen)

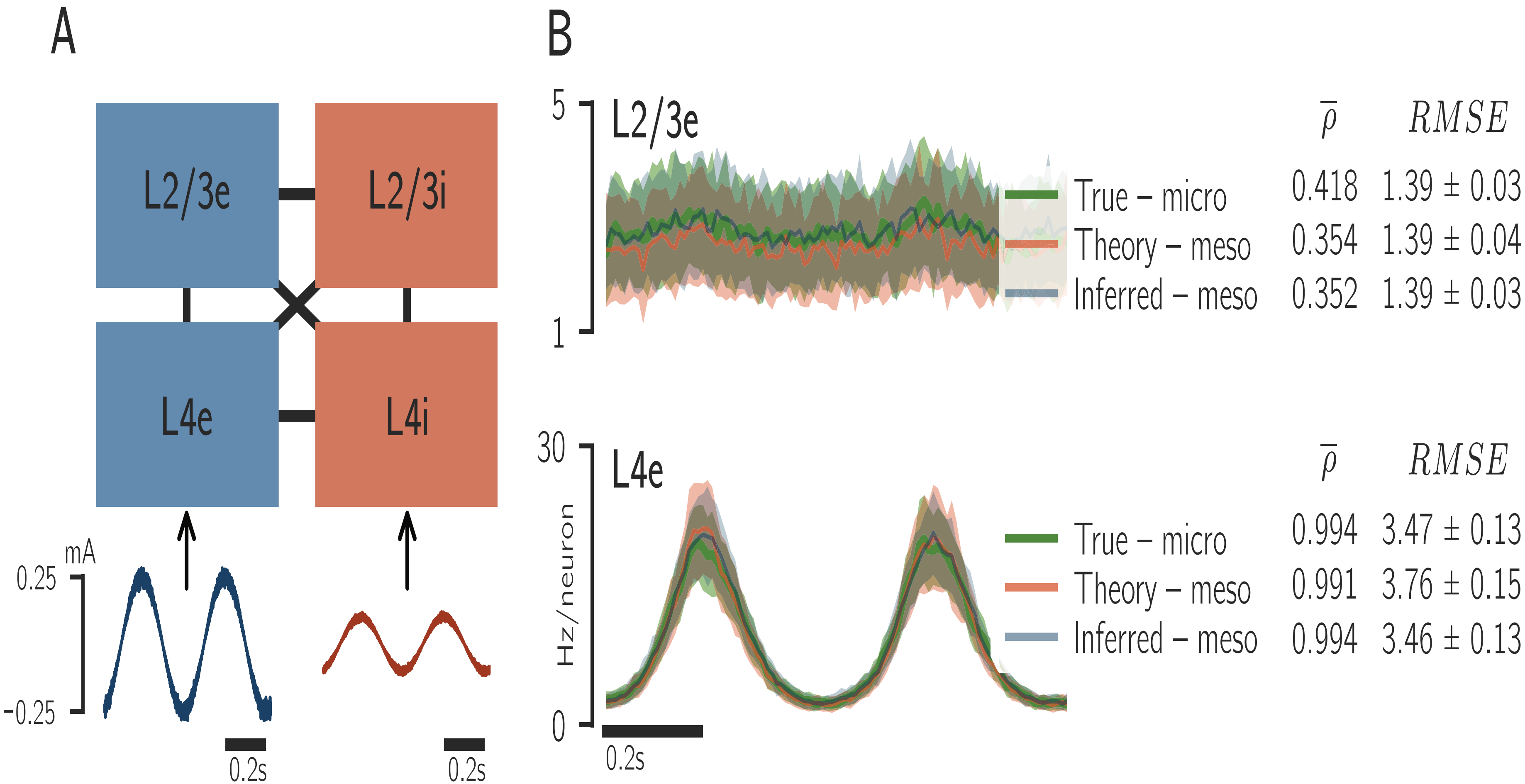


$\bar{\rho} = 0.950, 0.946, 0.918$ and $\mathrm{RMSE} = 3.42 \pm 0.07, 3.55 \pm 0.09, 3.40 \pm 0.08$

Inference of a 4-population model, 36 free parameters

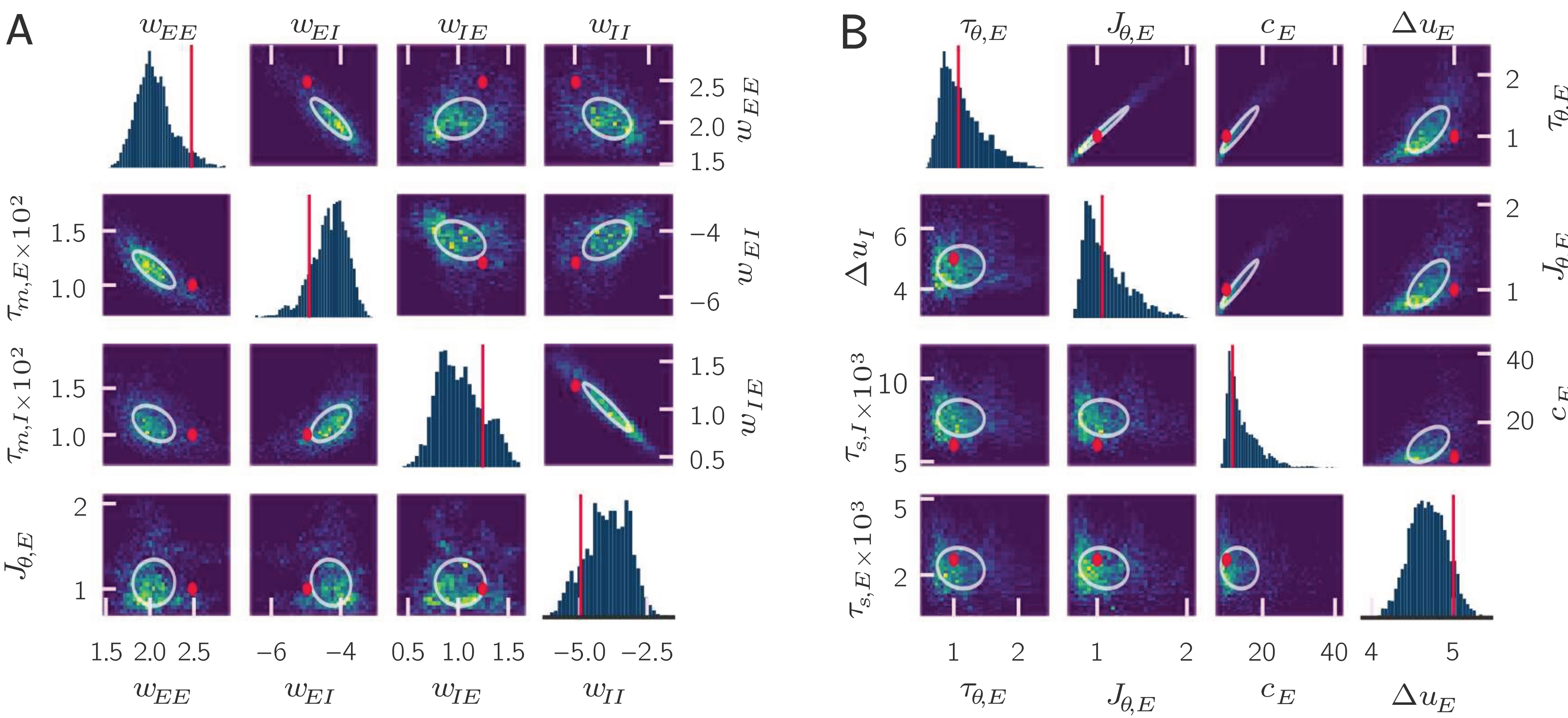
Potjans and Diesmann, 2014

Training: sine input only to L4



Bayesian inference used to obtain best fits

→ Hamiltonian Monte Carlo sampling
for parameter uncertainty and correlations



challenges

What if the model is incorrect?

What if we have only partial information?

Can one do any of this online during an experiment?

Currently applying the method to epileptic and Parkinsonian data

PART 5: Stochastic Optimal Control of Neural Firing Times

- knowing the state variable V(t): intracellular recordings
  → dynamic programming

OR

- knowing only last firing time: extracellular recordings
  → maximization principle (less info, worse performance)

References

Stochastic Optimal Control of Neural Firing

Iolov, A, Ditlevsen S and Longtin, A **(2014)** Stochastic optimal control of single neuron spike trains. **J. Neural Engineering** 11, 046004

Iolov A, Ditlevsen S, Longtin A **(2016)** Stochastic optimal control of spike times in single neurons. In: **Closed-loop Neuroscience**, A. El-Hady, ed. (Elsevier, San Diego)

Iolov A, Ditlevsen S, Longtin A **(2016)** Optimal design for estimation in diffusion processes from first hitting times. **SIAM/ASA J. Uncertainty Quantification** 5(1), 88–110.

General Context

A stochastic (drift-diffusion) process evolves in time.

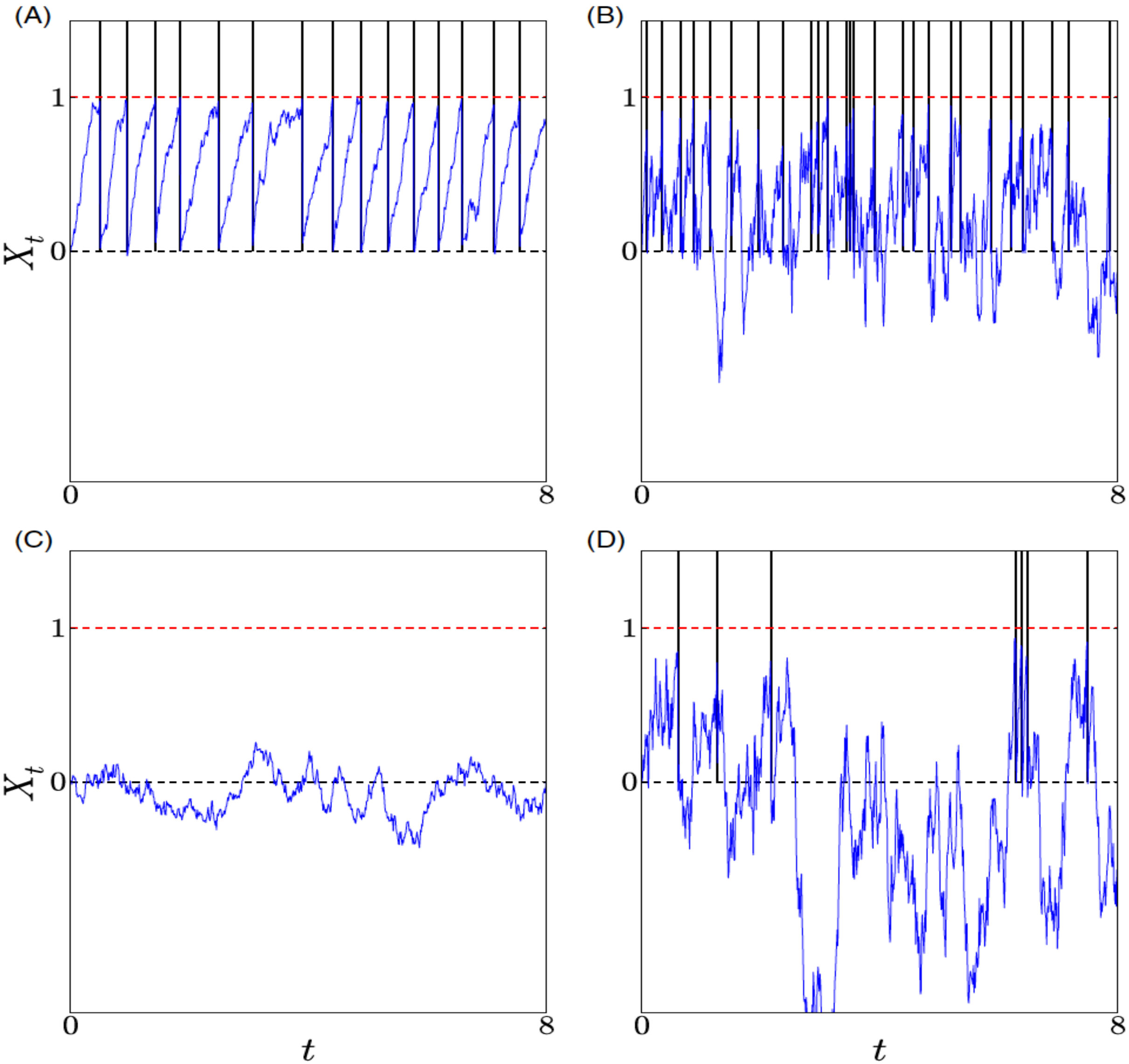
Can we optimally control the time(s) at which it crosses a threshold?

→ Stochastic Optimal Control of "Hitting" Times

REGIMES

SupraThreshold

SubThreshold



Low Noise

High Noise

Leaky Integrate-and-fire Neural Model

$$\mathrm{d}X(t) = \left(I_{\mathrm{ext}}(t) - \frac{X(t)}{\tau_c} \right) \mathrm{d}t + \beta\, \mathrm{d}W$$

$$X(0) = 0,$$

$$X(T_{\mathrm{sp}}) = x_{th} \Rightarrow \{X({T_{\mathrm{sp}}}^{+}) = 0.$$

This is an Ornstein-Uhlenbeck process with an absorbing threshold.

$$I_{\mathrm{ext}}(t) = \mu + \alpha(t)$$

μ is a bias term that sets distance of equilibrium to threshold

α is the <u>external control</u>

Goal:

apply control to achieve a specified threshold crossing time t^*

Desired spike time: t^*

Realized spike time: $T = \inf\{t > 0 : X_t \geq S\}$

Control: $\alpha(\cdot) = \arg\min_{\alpha(\cdot)} \left(E(T - t^*)^2\right)$

Note: System evolution is governed by a stochastic differential equation, so control can only be achieved in a statistical sense

Two Costs

More generally, we seek an optimal solution that minimizes the cost function:

$$J[\alpha(\cdot)] = E\left[\epsilon \int_0^T \alpha^2(s)ds + (T - t^*)^2\right]$$

The first term controls the total injected current to the cell (i.e. delivered energy):

→ we want that to be low…

We seek an optimal control, α^*, that solves

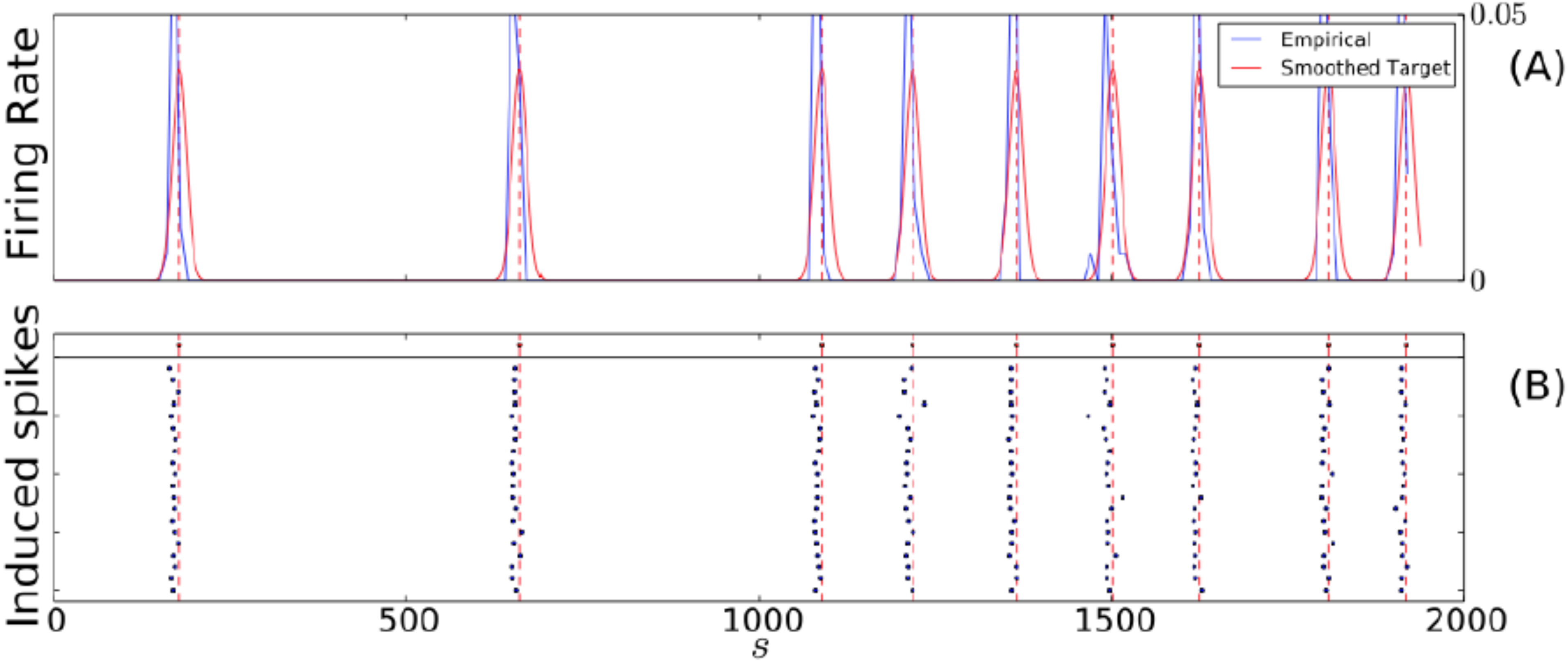
$$J[\alpha(\cdot)] = E\left[\epsilon \int_0^T \alpha^2(s)ds + (T - t^*)^2\right]$$

$$\alpha^*(\cdot) = \arg \min_{\alpha(\cdot)} J[\alpha(\cdot)].$$

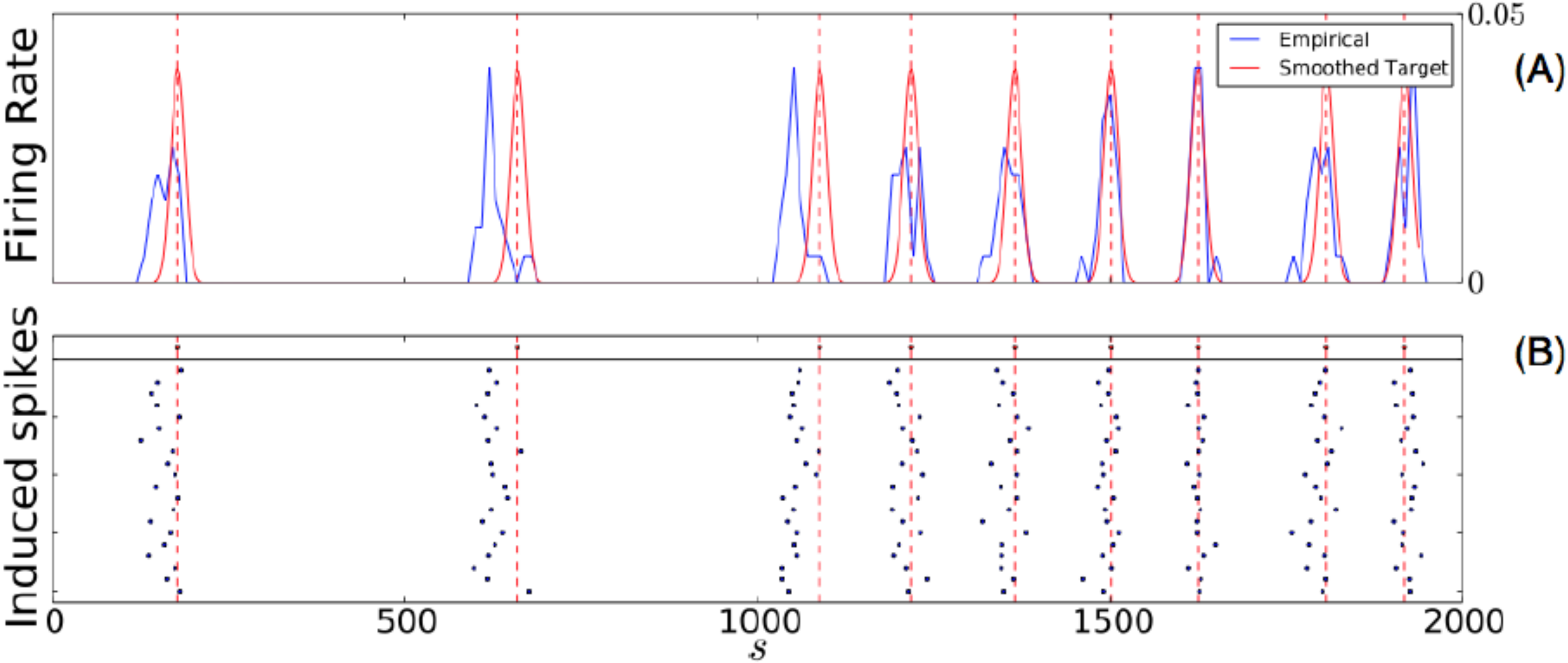
where ϵ scales the penalty on energy expenditure.

- **Closed-loop:** $\alpha = \alpha(x, t)$ (Access to membrane potential)
- **Open-loop:** $\alpha = \alpha(t)$ (Access to spike times)

Closed-loop control of Morris-Lecar neural model



Open-loop control, Morris-Lecar



Conclusion part 5: Stochastic Optimal Control

1. Strategy to control Hitting times in a drift-diffusion process
2. Ornstein-Uhlenbeck process with absorbing boundary (good model for noisy neural firing)
3. Works in closed loop (Hamilton-Jacobi-Bellman eq.) and open loop (optimization of transition density from Fokker-Planck eq.)
4. Works in different regimes: sub or suprathreshold, low or high noise.
5. Can be generalized to more elaborate neuron models

→ used for decision/classification dynamics in "real" neural nets?