A modified kinetic inverse Ising method for the inference of synaptic spatial structure and characteristic times

Federico Ricci-Tersenghi
Physics Department
Sapienza University, Roma
in collaboration with
Cristiano Capone, Paolo Del Giudice (ISS)
Carla Filosa, Guido Gigante (ISS)

Outline of the talk

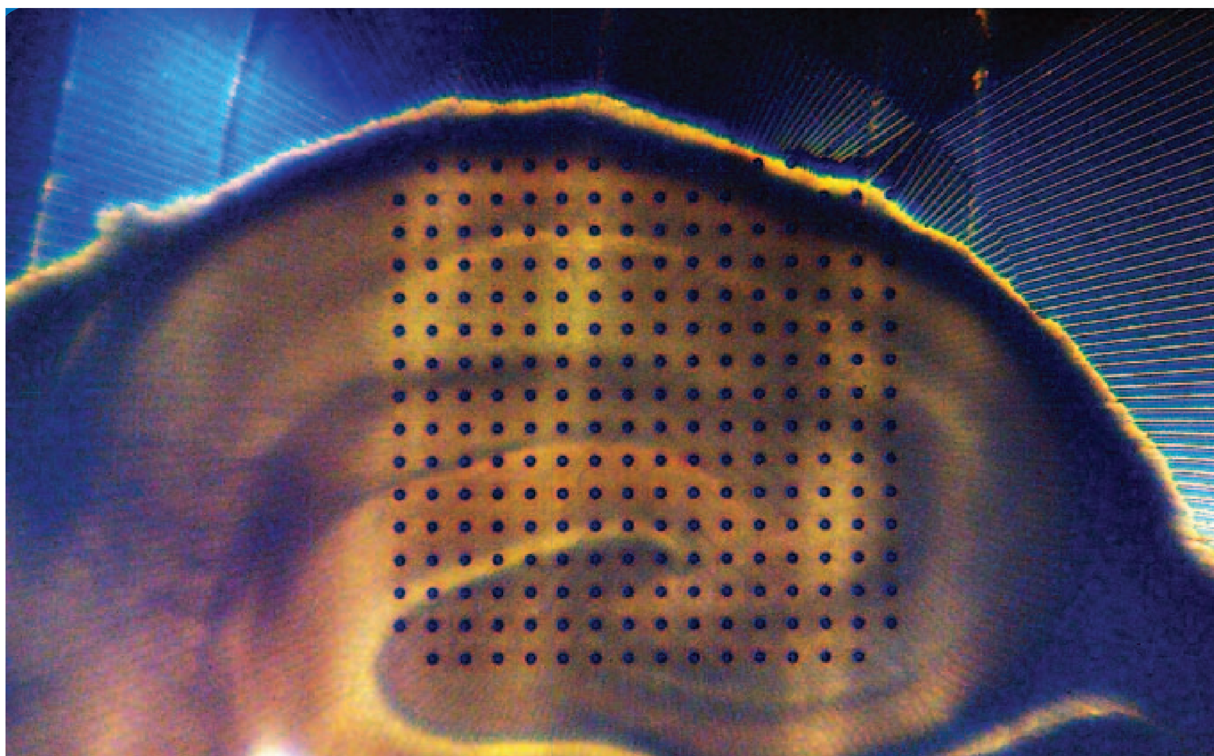
- Kinetic inverse Ising problem
- Mean field approximation
- Neural network simulator (LIF)
- Time discretization vs. interaction delays
- Kinetic inverse Ising problem with delays
- Relation between true and inferred couplings
- Application to more realistic data

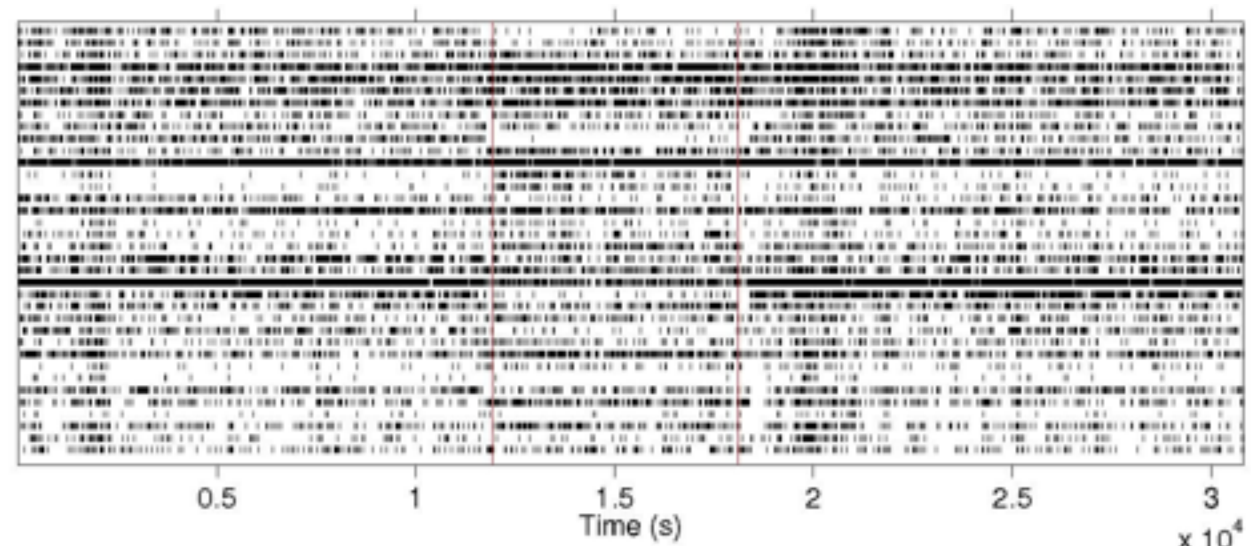
Motivations

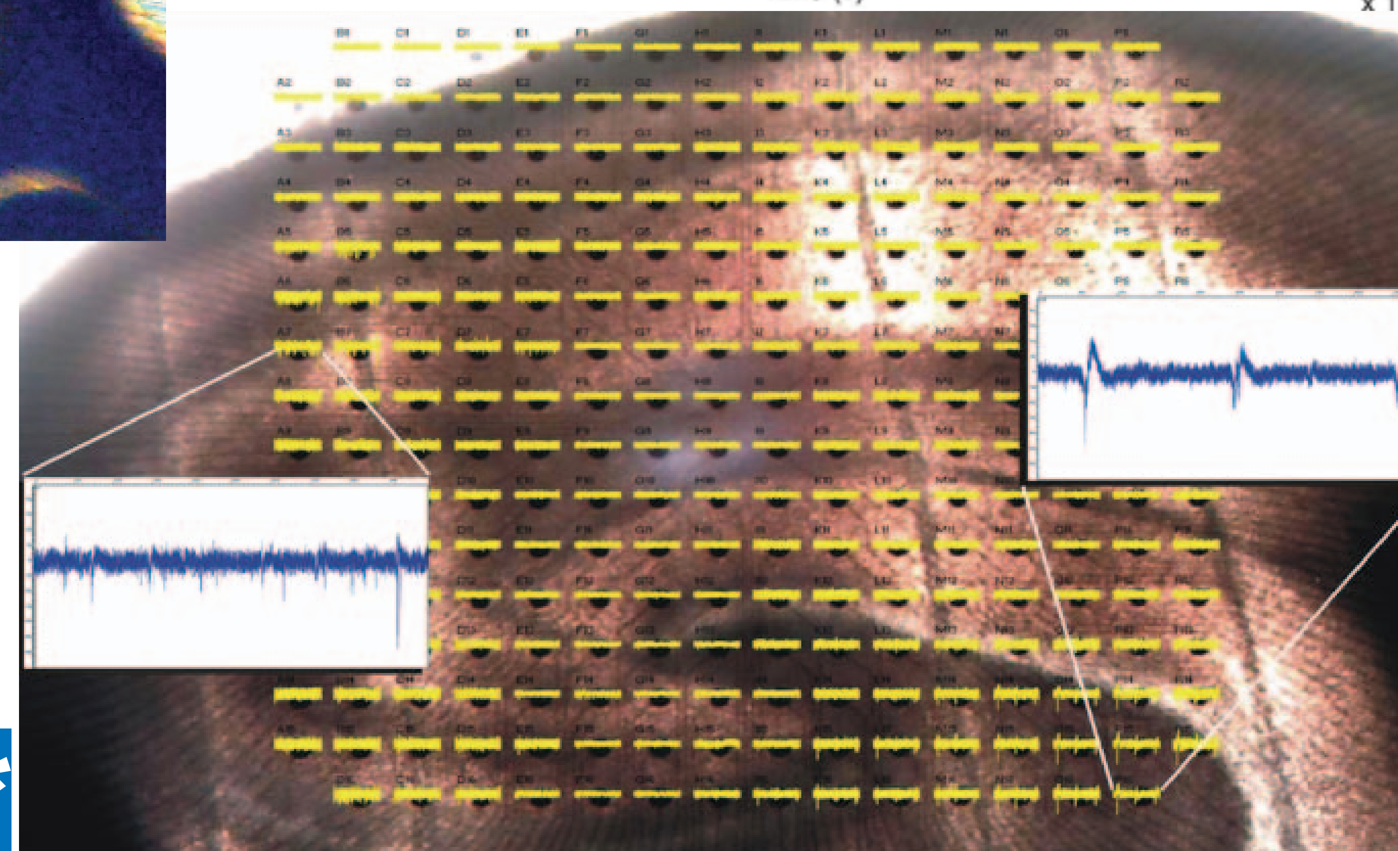
- Big data from biology -> interesting inference problems
- Not only "data fitting", but "data modeling"
  -> choice of the right interaction terms
  -> explanation for the collective behavior
- Under-sampling
  -> hidden variables
  -> effective interactions
- Statistical physics of complex systems can inspire us...
  but we should make our models more realistic

Big data in neurophysiology

- Multi-recording arrays (up to 256 channels)





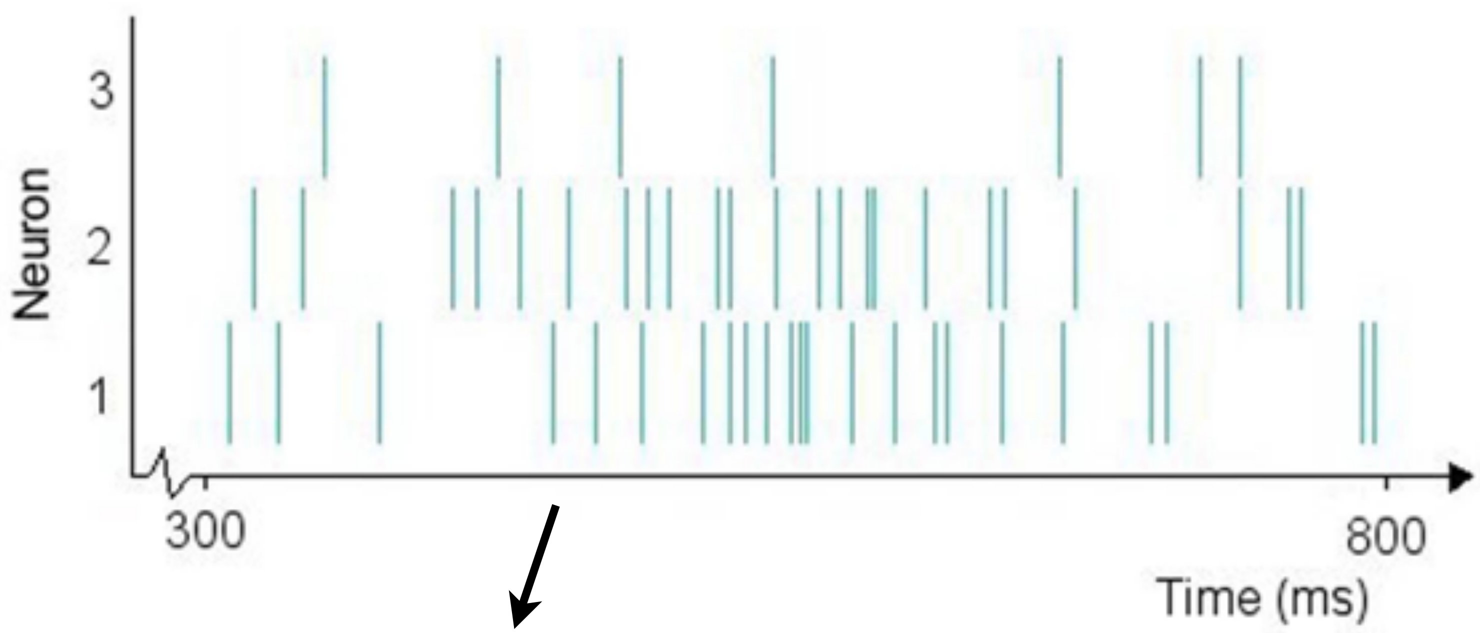


- each electrode can stimulate and record

multichannel systems

An interesting inference problem

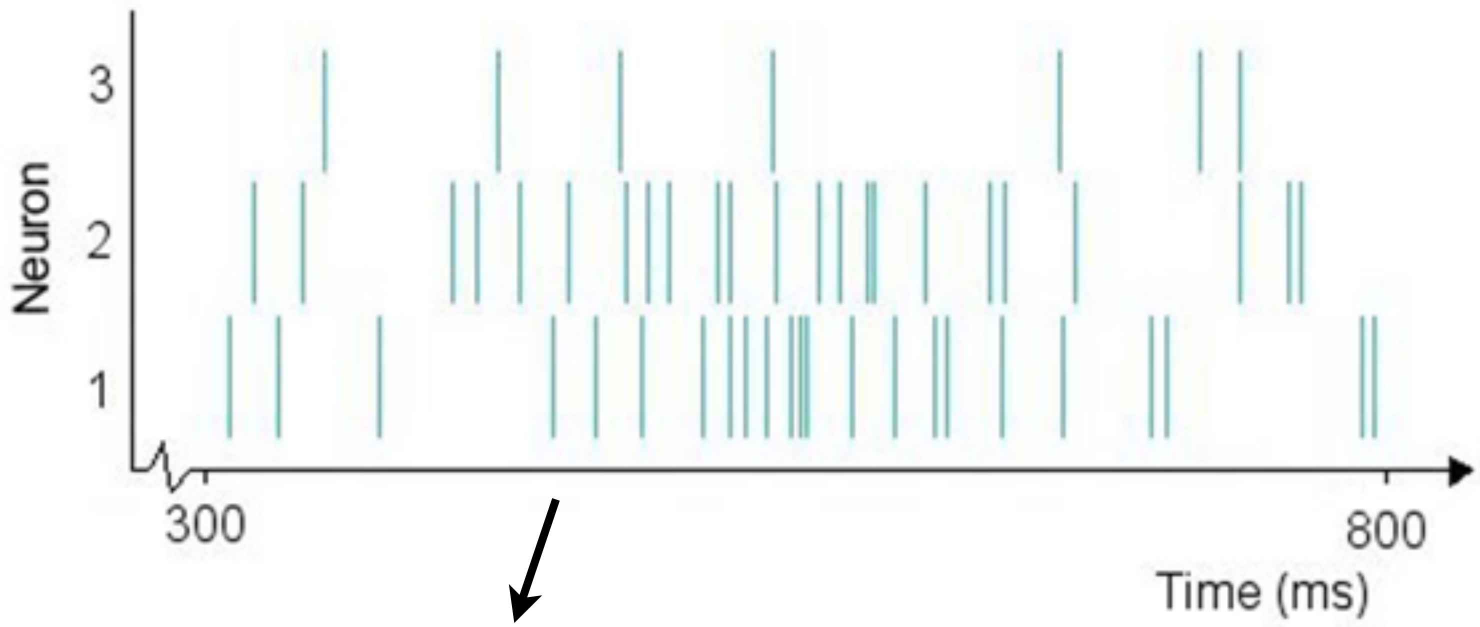
- Infer synaptic couplings from correlations of the sampled neural activities
- Data are discretized in bins of length dt



...0000100000100001000...
...001001000001001000... $\longrightarrow \{b_i(t)\}_{i=1,\ldots,N}$
...010010010000010000...

An interesting inference problem

- Infer synaptic couplings from correlations of the sampled neural activities
- Data are discretized in bins of length dt



...0000100000100001000...
...0010010000001001000...
...0100100100000010000...

$$\{b_i(t)\}_{i=1,\ldots,N}$$

$$S_i(t) = (-1)^{1-b_i(t)}$$

Maximum entropy solution

- From data compute average values

$$\langle S_i \rangle_{\text{data}} \qquad \langle S_i S_j \rangle_{\text{data}}$$

- Look for the most probable model (maximum entropy) within a class of interacting models (e.g. pairwise)

$$P[\mathbf{S}] = \frac{1}{Z} \exp\left[\tfrac{1}{2}\sum_{ij} J_{ij} S_i S_j + \sum_i h_i S_i\right]$$

such that

$$\langle S_i \rangle_{J,h} = \langle S_i \rangle_{\text{data}}$$
$$\langle S_i S_j \rangle_{J,h} = \langle S_i S_j \rangle_{\text{data}}$$

- Maximum entropy = minimum free-energy =
= maximum likelihood (with flat prior)

Why only pairwise interactions ?

- Compute more averages, e.g.

$$\langle S_i S_j S_k \rangle_{\rm data}$$

- Enlarge the class of interacting models, e.g. adding

$$\sum_{i,j,k} J_{ijk} S_i S_j S_k$$

- High order couplings turn out to be very small...
- Maybe are important in some applications as high order corrections, but pairwise interactions seem to be able to reproduce most of the data

Finding optimal model parameters

- Find coupling and fields {J,h} satisfying

$$\langle S_i \rangle_{J,h} = \langle S_i \rangle_{\mathrm{data}}$$
$$\langle S_i S_j \rangle_{J,h} = \langle S_i S_j \rangle_{\mathrm{data}}$$

- Very well known problem (Boltzmann machine learning)
Learn parameters according to following rules

$$\begin{aligned} \delta h_i &= \eta(\langle S_i \rangle_{\mathrm{data}} - \langle S_i \rangle_{\mathbf{J},\mathbf{h}}), \\ \delta J_{ij} &= \eta(\langle S_i S_j \rangle_{\mathrm{data}} - \langle S_i S_j \rangle_{\mathbf{J},\mathbf{h}}) \end{aligned}$$

- Exact computation is exponentially slow

- Even Monte Carlo is not very useful
(one MC run for each learning step)

Mean-field approximations

- Naive mean-field $m_i = \tanh\left(h_i + \sum_j J_{ij} m_j\right)$

$$h_i = \tanh^{-1} m_i - \sum_j J_{ij} m_j$$

$$\chi_{ij}^{-1} = \frac{\partial h_i}{\partial m_j} = \frac{\delta_{ij}}{1 - m_i^2} - J_{ij}$$

$$C_{ij} = \langle (S_i - m_i)(S_j - m_j) \rangle$$

$$J_{ij} = -(C^{-1})_{ij}$$

Not bad for weakly interacting systems

Kinetic Ising model

Hertz, Roudi and Tyrcha in "Principle of Neural Coding" 2013

- Limits of the Gibbs equilibrium measure:
  -> configurations at close-by times are correlated
  -> non symmetric interactions
  -> non stationary regimes

- Kinetic Ising model is based on a stochastic dynamics where configuration $\mathbf{S}(t+1)$ depends on $\mathbf{S}(t)$

$$P[S_i(t+1)|\mathbf{S}(t)] = \frac{\exp[S_i(t+1)H_i(t)]}{2\cosh H_i(t)}$$

$$H_i(t) = h_i(t) + \sum_j J_{ij} S_j(t)$$

J_{ij} may be non symmetric

Stationary case

- Fields are time-independent
  Model parameters to be inferred: $N\ h_i\ ,\ N(N-1)\ J_{ij}$
  Maximize the log-likelihood

$$\mathcal{L} = \sum_{it} S_i(t+1)H_i(t) - \log 2\cosh H_i(t)$$

  by e.g. gradient ascent

$$\begin{aligned} \delta h_i &= \eta\left[\langle S_i(t+1)\rangle_t - \langle \tanh H_i(t)\rangle_t\right] \\ \delta J_{ij} &= \eta\left[\langle S_i(t+1)S_j(t)\rangle_t - \langle \tanh H_i(t)S_j(t)\rangle_t\right] \end{aligned}$$

  Easier than equilibrium case: averages can be quickly evaluated from the data

Naive MF approx. for KIM

- Assume averages satisfy naive MF equation

$$m_i = \tanh\Big(h_i + \sum_j J_{ij} m_j\Big)$$

Study fluctuations around mean values $S_i = m_i + \delta S_i$

$$\langle S_i(t+1) S_j(t)\rangle_t = \langle \tanh H_i(t) S_j(t)\rangle$$

$$\Downarrow$$

$$\langle \delta S_i(t+1)\delta S_j(t)\rangle_t = (1-m_i^2)\sum_k J_{ik}\langle \delta S_k(t)\delta S_j(t)\rangle$$

Naive MF approx. for KIM

- Assume averages satisfy naive MF equation

$$m_i = \tanh\Big(h_i + \sum_j J_{ij} m_j\Big)$$

Study fluctuations around mean values $S_i = m_i + \delta S_i$

$$\langle S_i(t+1) S_j(t) \rangle_t = \langle \tanh H_i(t) S_j(t) \rangle$$

$$\Downarrow$$

$$\langle \delta S_i(t+1) \delta S_j(t) \rangle_t = (1 - m_i^2) \sum_k J_{ik} \langle \delta S_k(t) \delta S_j(t) \rangle$$

$$D = A\,J\,C$$

MF approx. for KIM

$$\boxed{J = A^{-1}\, D\, C^{-1}}$$

$$A_{ij} = a_i\, \delta_{ij}$$

$$C_{ij} = \langle \delta S_i(t) \delta S_j(t) \rangle_t$$

$$D_{ij} = \langle \delta S_i(t+1) \delta S_j(t) \rangle_t$$

$$h_i = \tanh^{-1} m_i - \textstyle\sum_j J_{ij} m_j$$

- Naive MF $\quad a_i = 1 - m_i^2$
- TAP $\quad a_i = (1 - m_i^2)\, [1 - (1 - m_i^2)\Delta_i]$
- Exact for fully asymmetric SK model

$$\Delta_i = \sum_j J_{ij}^2 (1 - m_j^2)$$

$$a_i = \int Dx \left[1 - \tanh^2 \left(h_i + \sum_j J_{ij} m_j + x \Delta_i \right) \right]$$

Simulated neural networks

- There are many, with very different complexity levels
- We have used Leaky Integrate and Fire (LIF) dynamics

$$\tau_m \frac{dV(t)}{dt} = -V(t) + R\,I(t) \qquad\qquad \tau_m = 20\,ms$$

- If $V > \theta \;\; (\theta = 1)$ neuron fires and potential is reset
- Two populations of neurons:
  - **internal** sparsely connected with parameters $\{J_{ij}, \delta_{ij}\}$
  - **external** sustaining the network activity
    Poisson process with parameters $J_{ext},\, \nu_{ext}$

$$\mathbb{P}[I_{ext} = kJ_{ext}; dt] = \frac{(\nu_{ext} dt)^k}{k!} e^{-\nu_{ext} dt}$$

Simulated neural networks

- The incoming current on neuron i is

$$I_i(t) = I_{ext} + \sum_{j=1}^{N} J_{ij} \sum_k \delta[t - (t_{j,k}^{fire} + \delta_{ij})]$$

ΔV_i induced by the firing of neuron j

interaction delay

instantaneous spike

- Interaction delay is an effective delay.
  More realistic models have non-zero synaptic integration times ranging between 1 ms and 100 ms

Simulated neural networks

- Sparse topology
  typically only 10% of interactions are present $J_{ij} \neq 0$
- Excitatory synapses $J_{ij} > 0$
  Inhibitory synapses $J_{ij} < 0$
- Event-driven numerical simulation
  (no time discretization)

The choice of the binning time dt

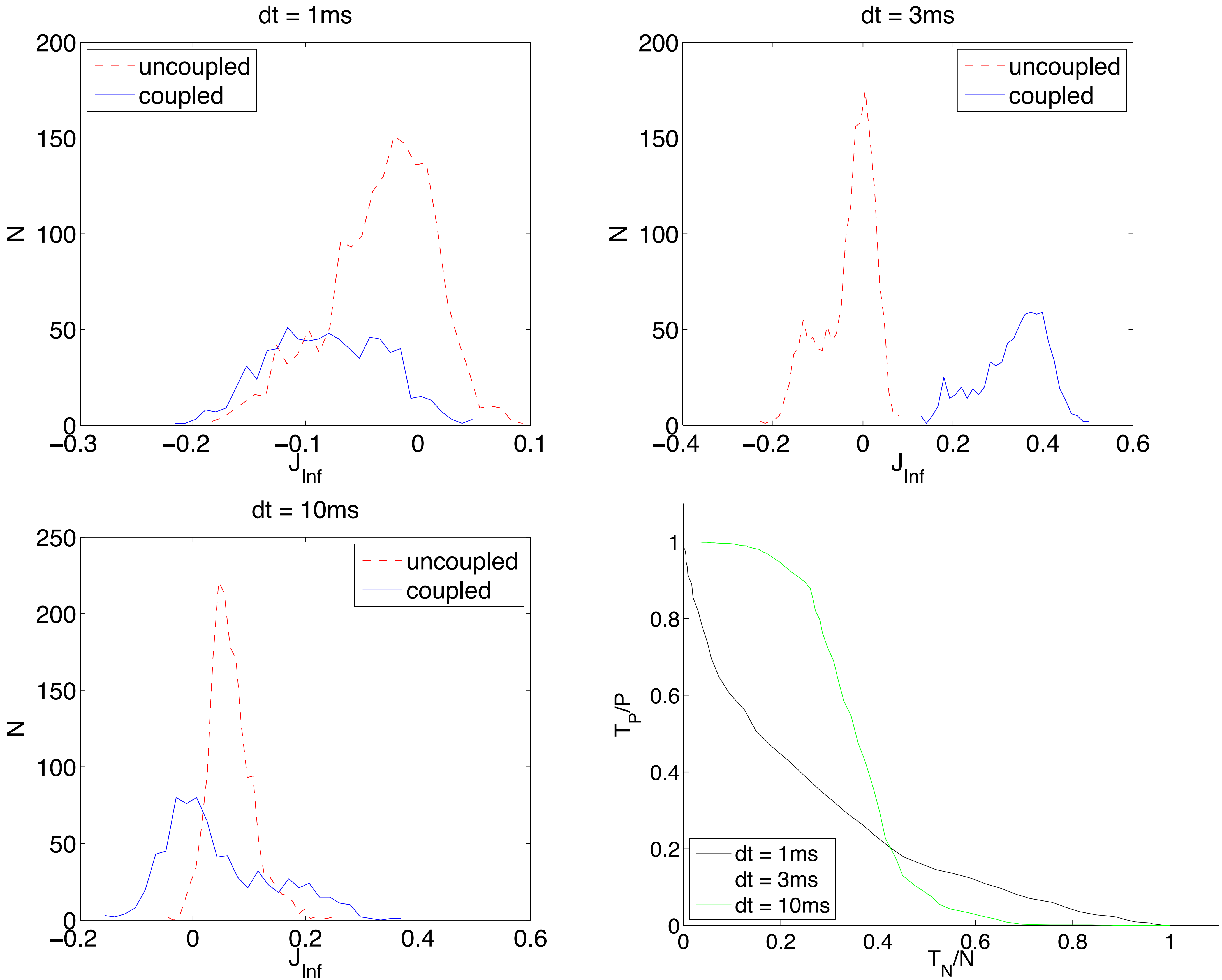
- Simulation outcome: firing times for each neuron $\{t_{i,k}^{fire}\}$
- Free to choose the binning time dt as long as the probability of having two spikes in the same time bin is negligible
- The choice affects correlations

$$D = \langle \delta S_i(t+dt) \delta S_j(t) \rangle_t$$

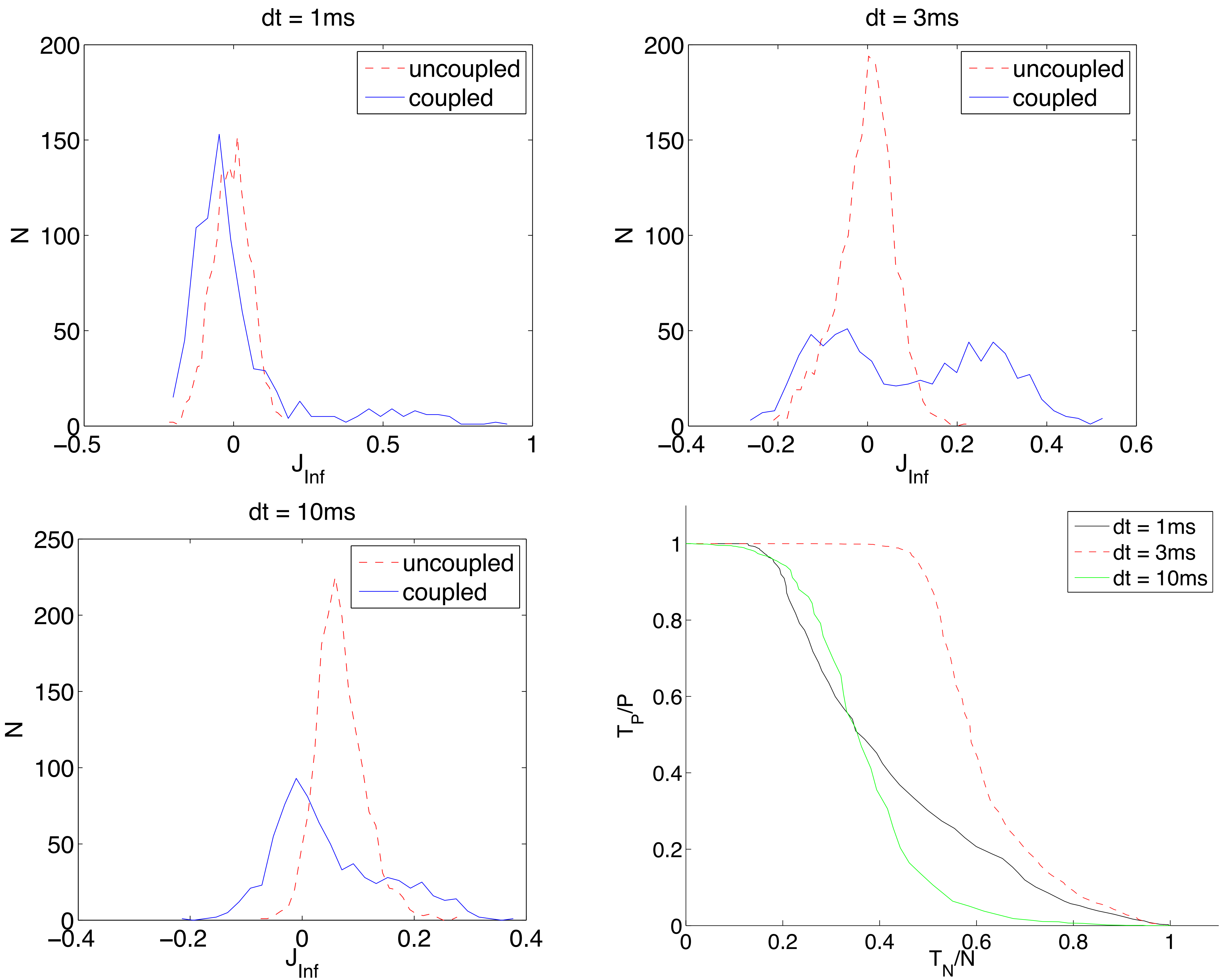
and coupling estimates through

$$J = A^{-1}\, D\, C^{-1}$$

$$N = 50,\ J_{ext} = 0.05,\ \nu_{ext} = 1\,ms^{-1},\ J_{ij} = 0.05,\ \boxed{\delta_{ij} = 3\,ms}$$

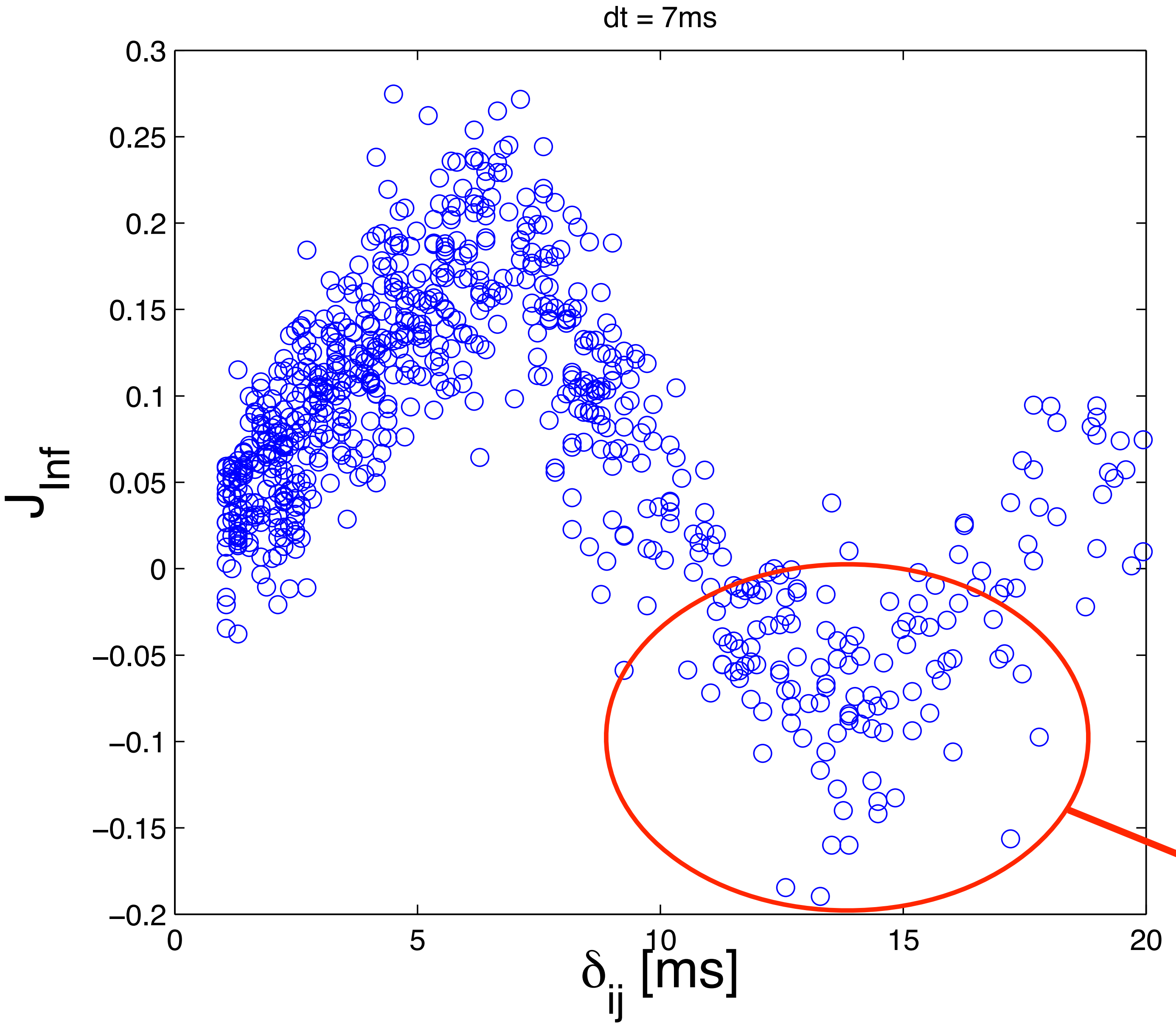


$$N = 50,\ J_{ext} = 0.05,\ \nu_{ext} = 1\,ms^{-1},\ J_{ij} = 0.05,\ \boxed{\delta_{ij} \in [1, 20]\,ms}$$



Couplings of wrong sign ?!

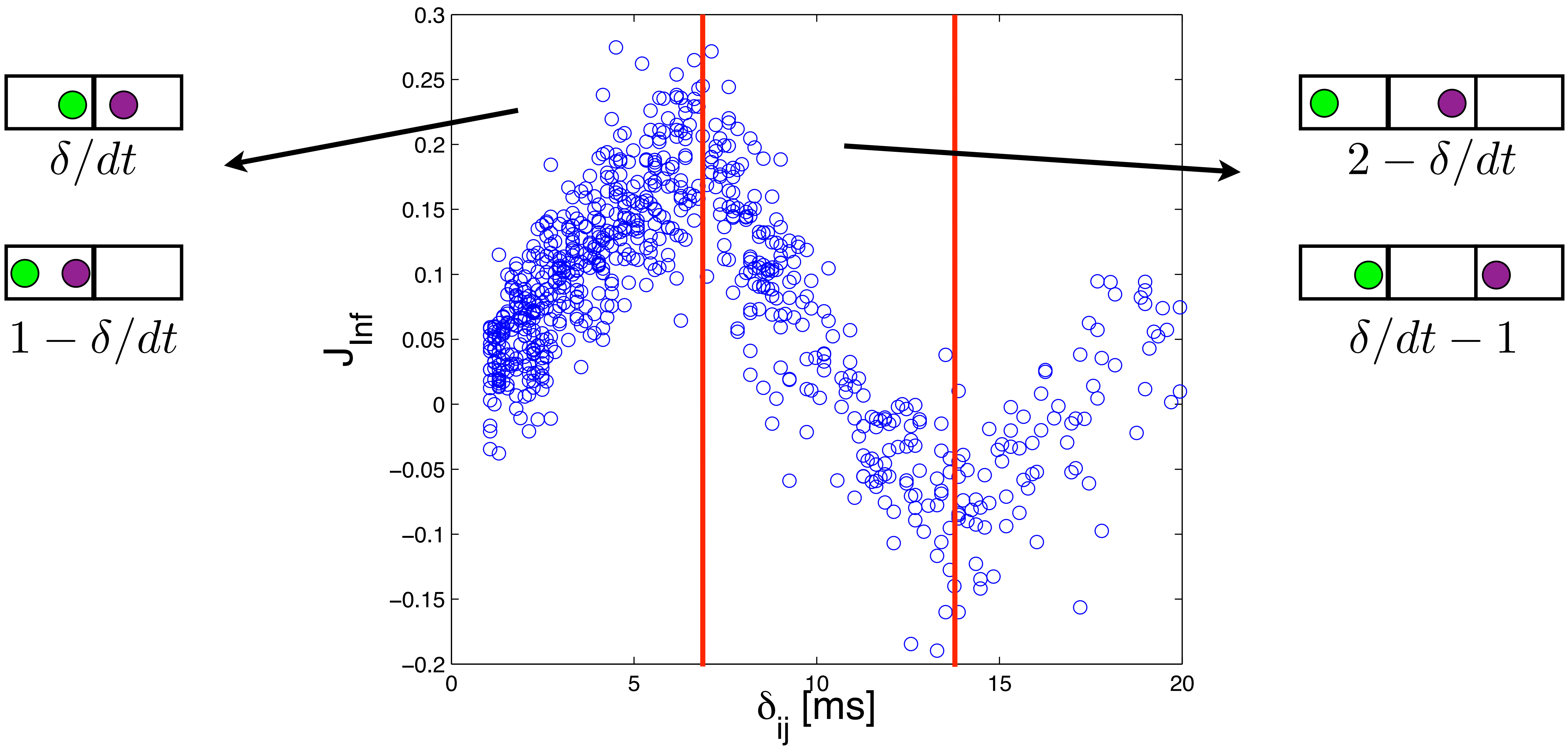
$$dt = 7\,ms$$



Couplings inferred with a wrong sign

Couplings of wrong sign ?!

$$dt = 7\,ms$$



KIM with time delays

- Stochastic dynamics as in KIM

$$P[S_i(t+1)|\mathbf{S}(t)] = \frac{\exp[S_i(t+1)H_i(t)]}{2\cosh H_i(t)}$$

but with delays

$$H_i(t) = h_i(t) + \sum_j J_{ij} S_j(t+1-\delta_{ij})$$

- Maximize the log-likelihood

$$\mathcal{L} = \sum_{it} S_i(t+1)H_i(t) - \log 2\cosh H_i(t)$$

with respect to couplings and delays

KIM with time delays

$$\frac{\partial \mathcal{L}}{\partial J_{ij}} = 0 \implies D_{ij}(\delta_{ij}) = (1 - m_i^2) \sum_k J_{ik} M_{kj}^{(i)}$$

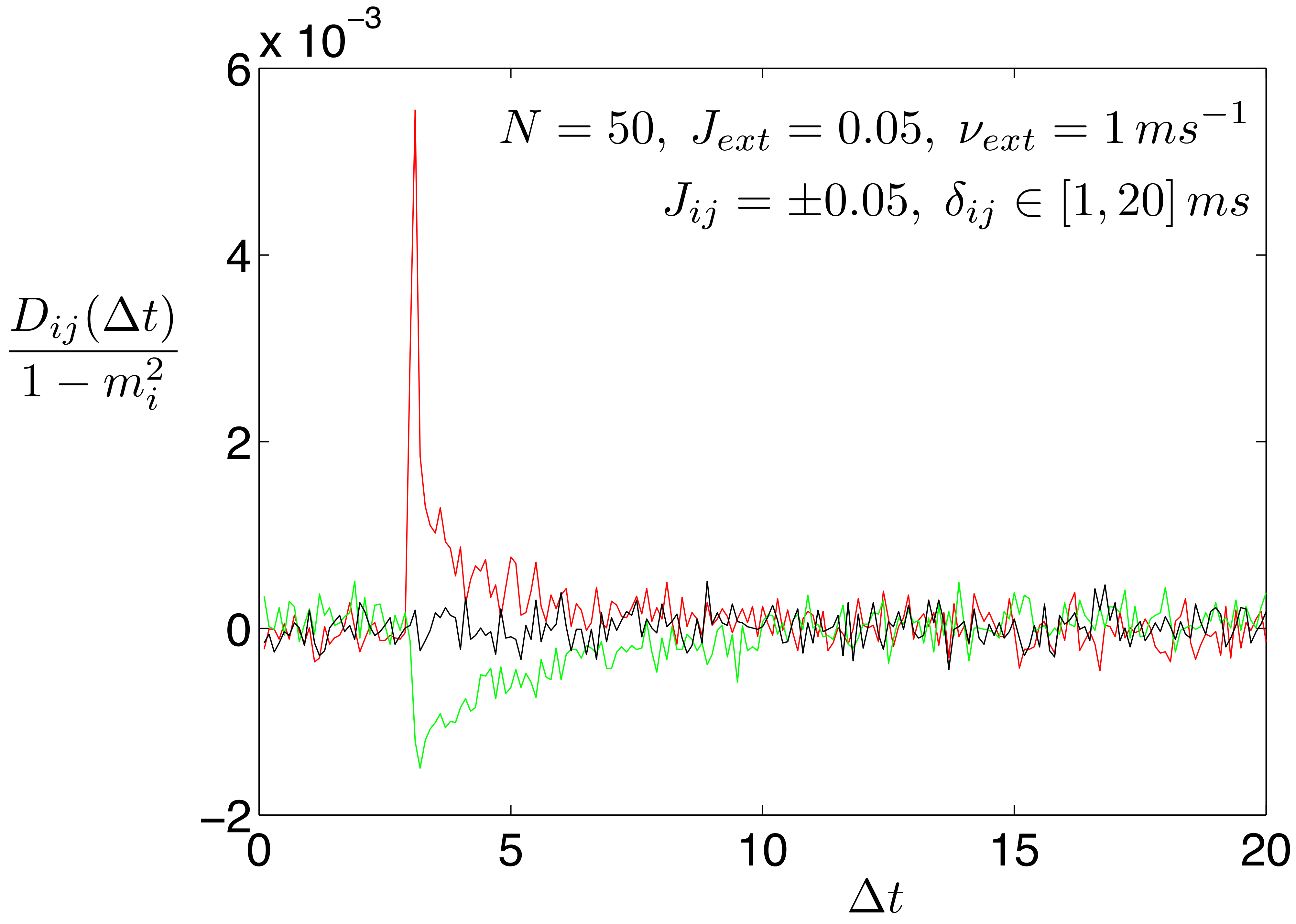
$$D_{ij}(\Delta t) = \langle \delta S_i(t + \Delta t)\, \delta S_j(t) \rangle_t$$

$$M_{kj}^{(i)} = \begin{cases} D_{kj}(\delta_{ij} - \delta_{ik}) & \text{if } \delta_{ik} < \delta_{ij} \\ D_{jk}(\delta_{ik} - \delta_{ij}) & \text{if } \delta_{ik} > \delta_{ij} \\ C_{kj} = C_{jk} & \text{if } \delta_{ik} = \delta_{ij} \end{cases}$$

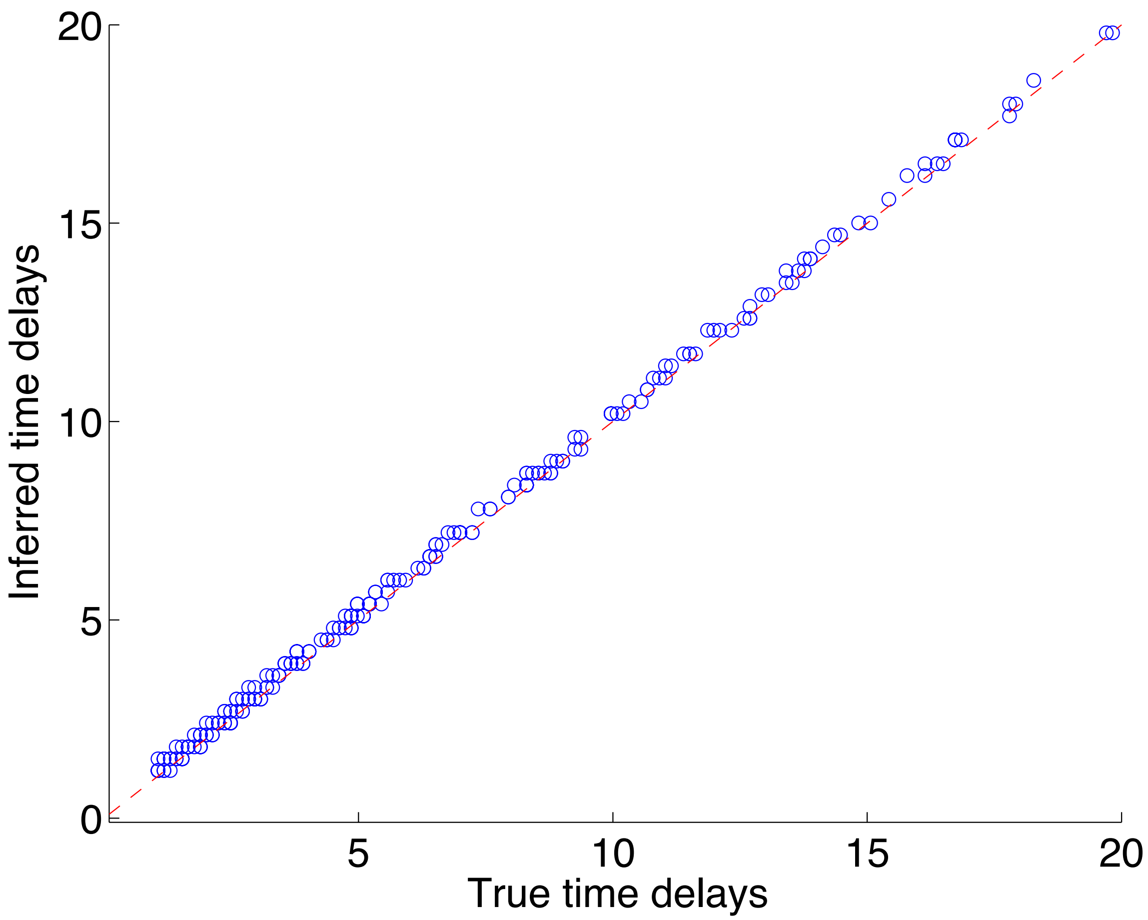
$$\boxed{J_{ij} = \frac{1}{1 - m_i^2} \sum_k D_{ik} (M^{(i)})^{-1}_{kj}}$$

But finding the right delays by $\frac{\partial \mathcal{L}}{\partial \delta_{ij}} = 0$ it is not easy

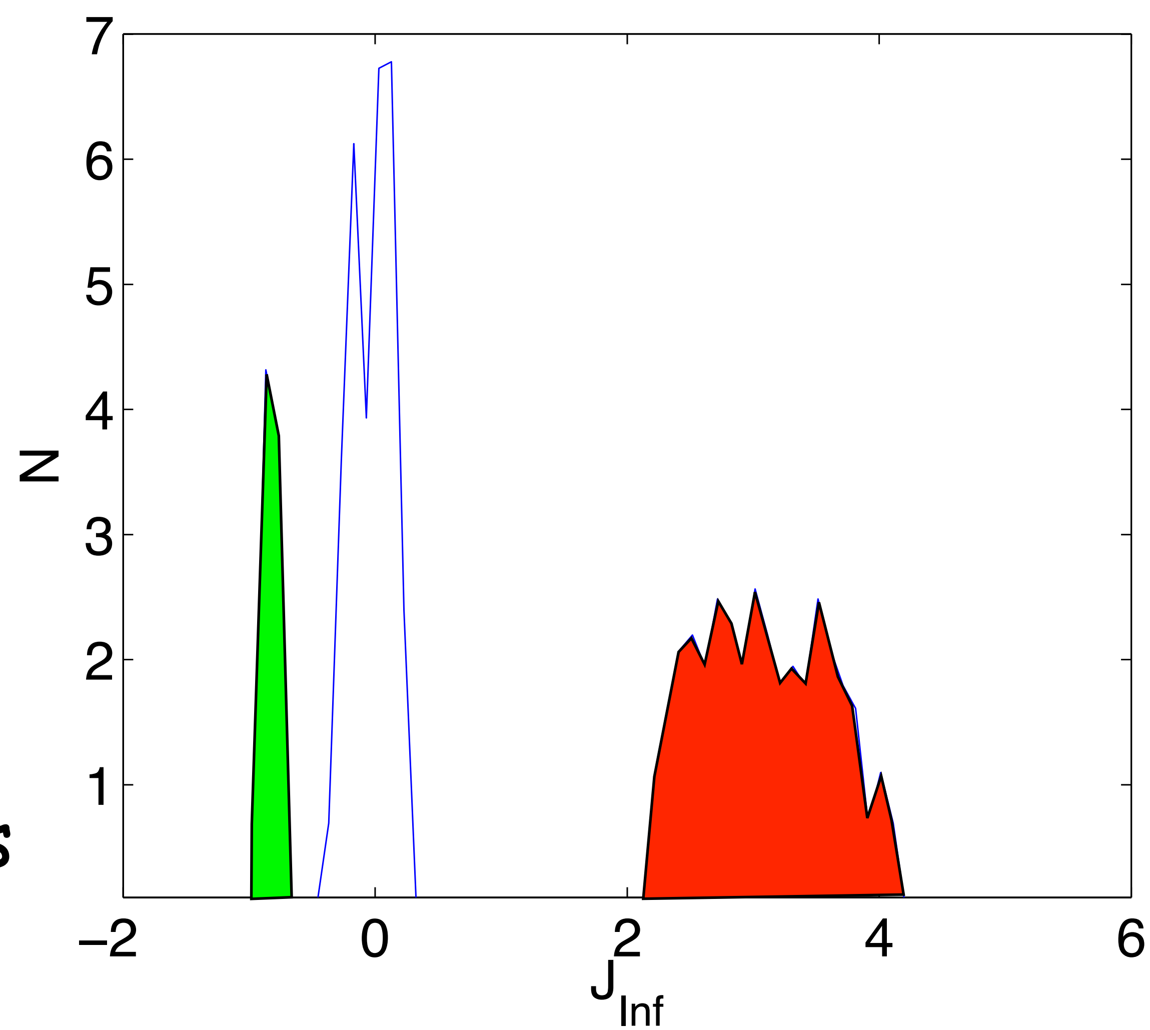
Looking for the time delays



Inference by the KIMTD



$$N = 50,\ J_{ext} = 0.05,\ \nu_{ext} = 1\, ms^{-1}$$
$$J_{ij} = \pm 0.05,\ \delta_{ij} \in [1, 20]\, ms$$



Strong asymmetry in the inferred couplings

Inferred vs. true couplings

- Inferred values for excitatory and inhibitory synaptic efficacies/couplings are different
- Excitatory synapses induce extra spikes
  -> can be inferred from **correlations** between spikes
- Inhibitory synapses can only retard spikes
  -> must be inferred from **lack** of spikes
- The 2 processes are not symmetric!

Inferred vs. true couplings

- Stationary state: $p(v)$ pdf of the potential
- Excitatory synapsis

$$\mathbb{P}(\mathrm{spike}, dt=0) = \int_{\theta-J}^{\theta} p(v)dv = \int_{\theta-J}^{\theta} p'(\theta)(v-\theta)dv \propto J^2$$

Diffusion with adsorbing wall at $v=\theta$

- In a finite time window also the external current can help

$$\mathbb{P}(\mathrm{spike}, dt) \propto J^2 + J\, J_{ext}\, \nu_{ext} dt$$

to linear order in dt

Inferred vs. true couplings

- Inhibitory synapsis: You need to have a chance to fire to measure the lack of a spike!
  -> the signal is proportional to dt

- Since $J^{infer} \propto \mathbb{P}[\mathrm{spike}, dt]/dt$ then

$$J_{ij}^{infer} \simeq \frac{1}{\nu_{ext}\, dt} \frac{J_{ij}^2}{J_{ext}^2} + O(1) \quad \text{if } J_{ij} > 0$$

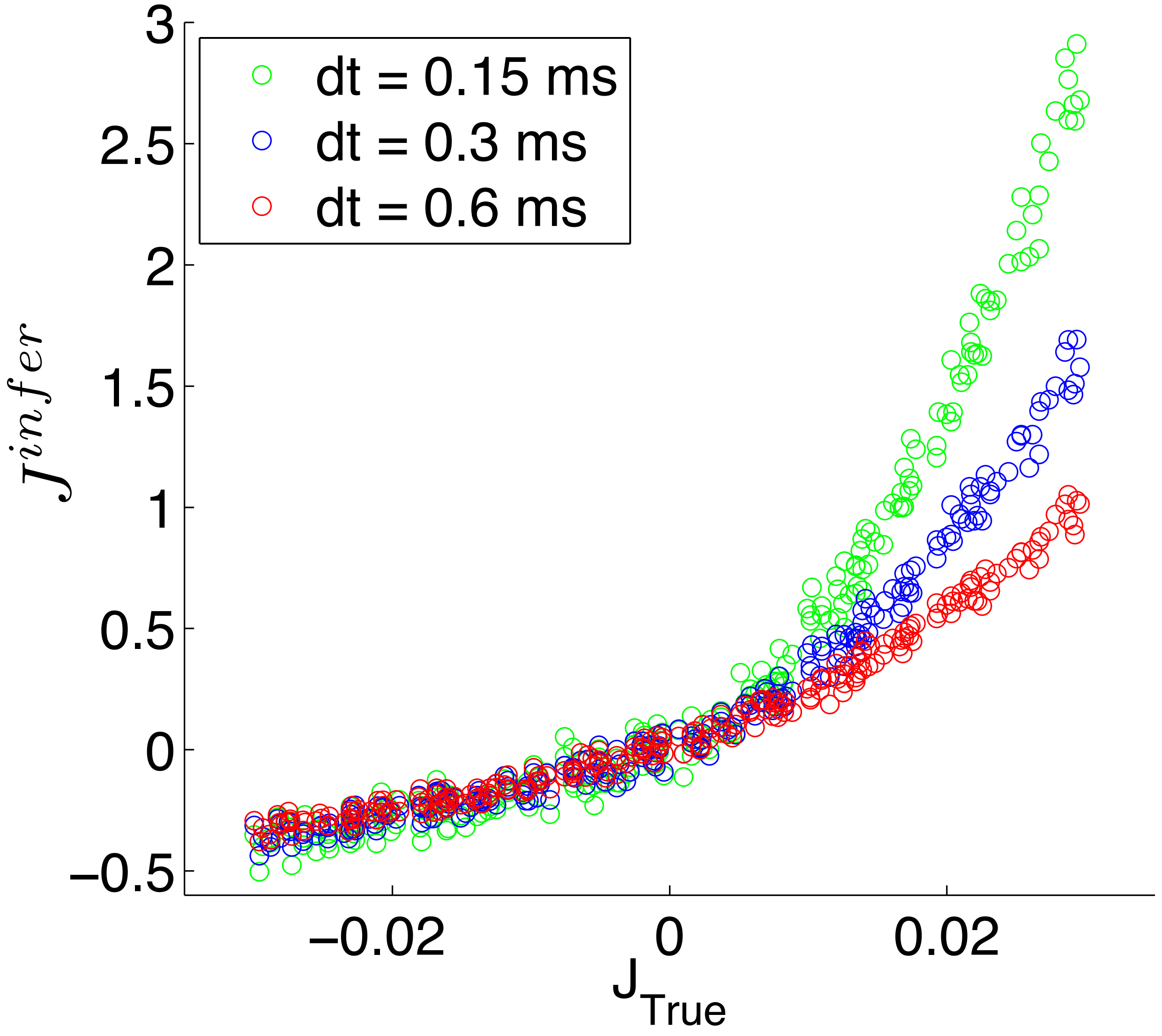
$$J_{ij}^{infer} \simeq \frac{1}{J_{ext}^2} \left(J_{ext}\, J_{ij} + \frac{J_{ij}^2}{2}\right) + O(dt) \quad \text{if } -J_{ext} \leq J_{ij} < 0$$

$$J_{ij}^{infer} \simeq -\frac{1}{2} + O(dt) \quad \text{if } -2\, J_{ext} \leq J_{ij} < -J_{ext}$$

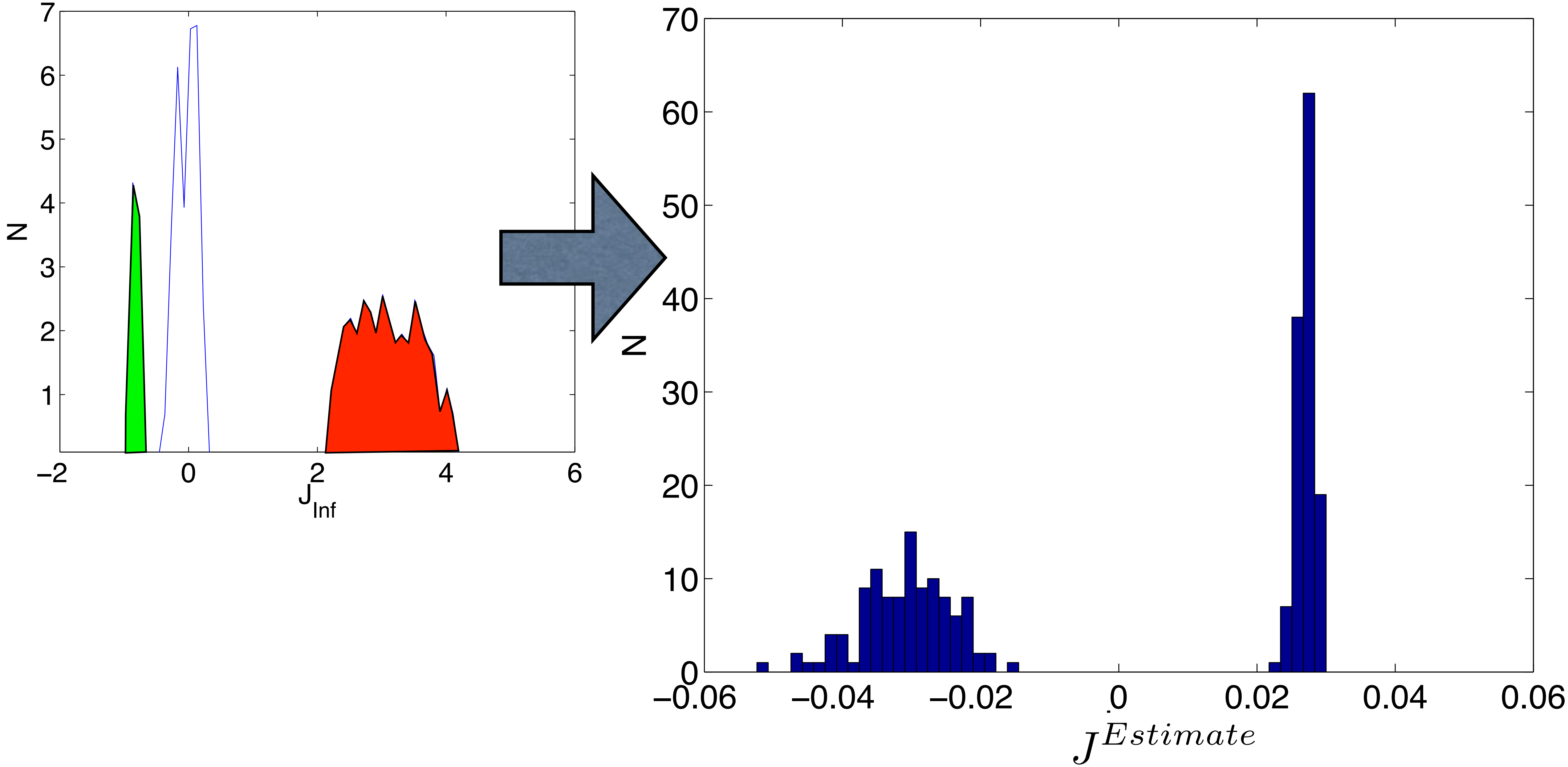
Inferred vs. true couplings

$$J_{ij}^{infer} = O(1/dt) \qquad r_{ij} > 0$$

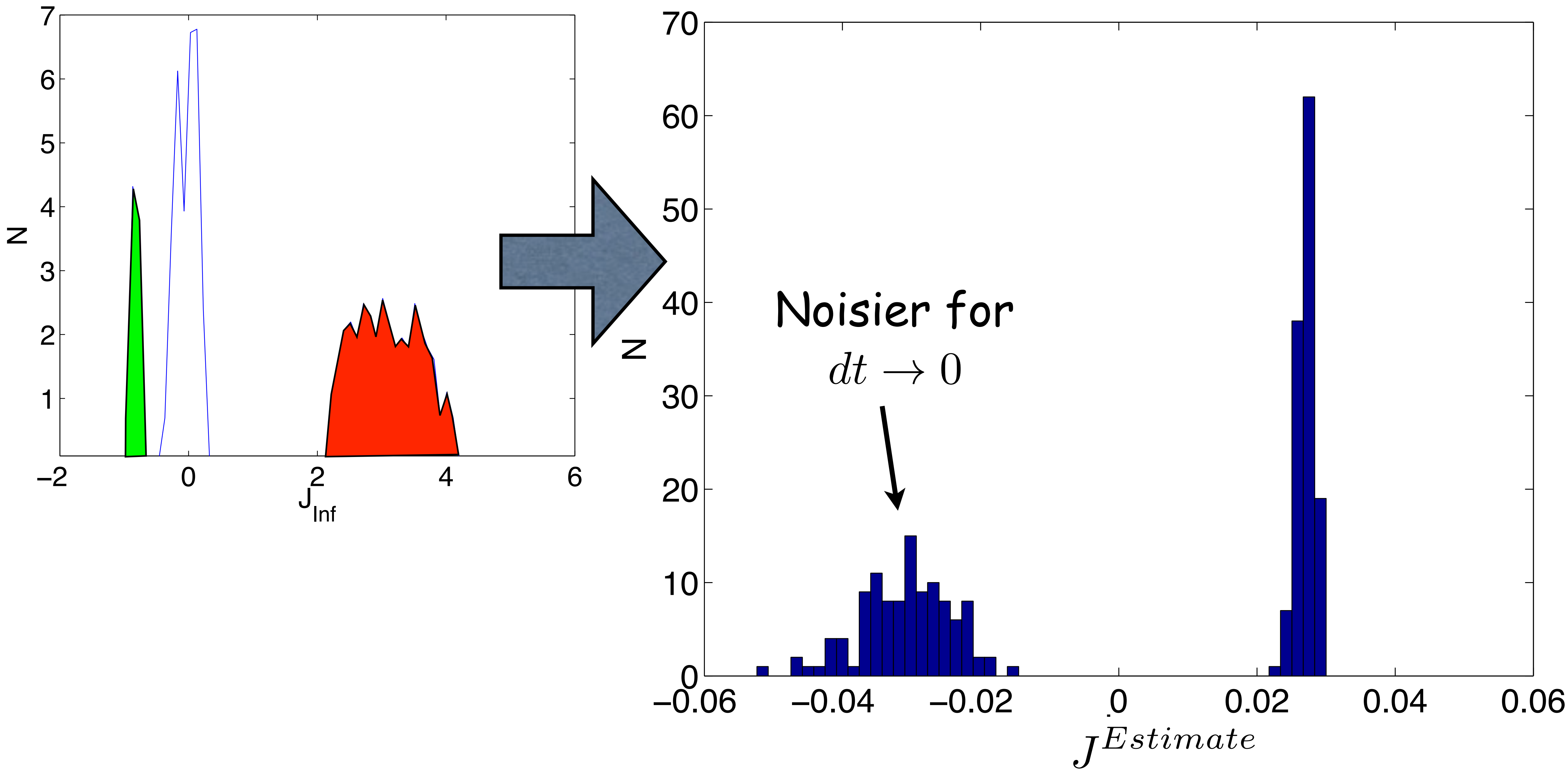
$$J_{ij}^{infer} = O(1) \qquad r_{ij} < 0$$



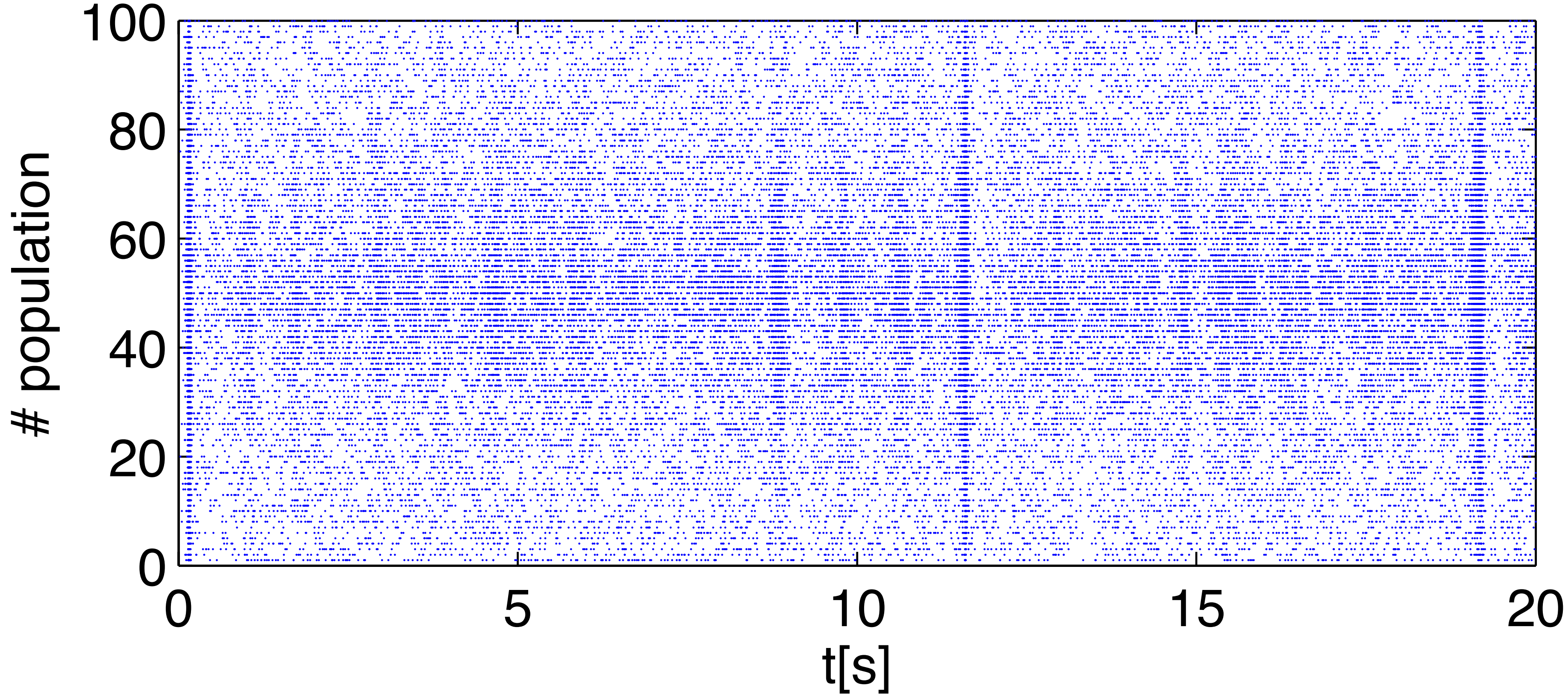
Inferred vs. true couplings

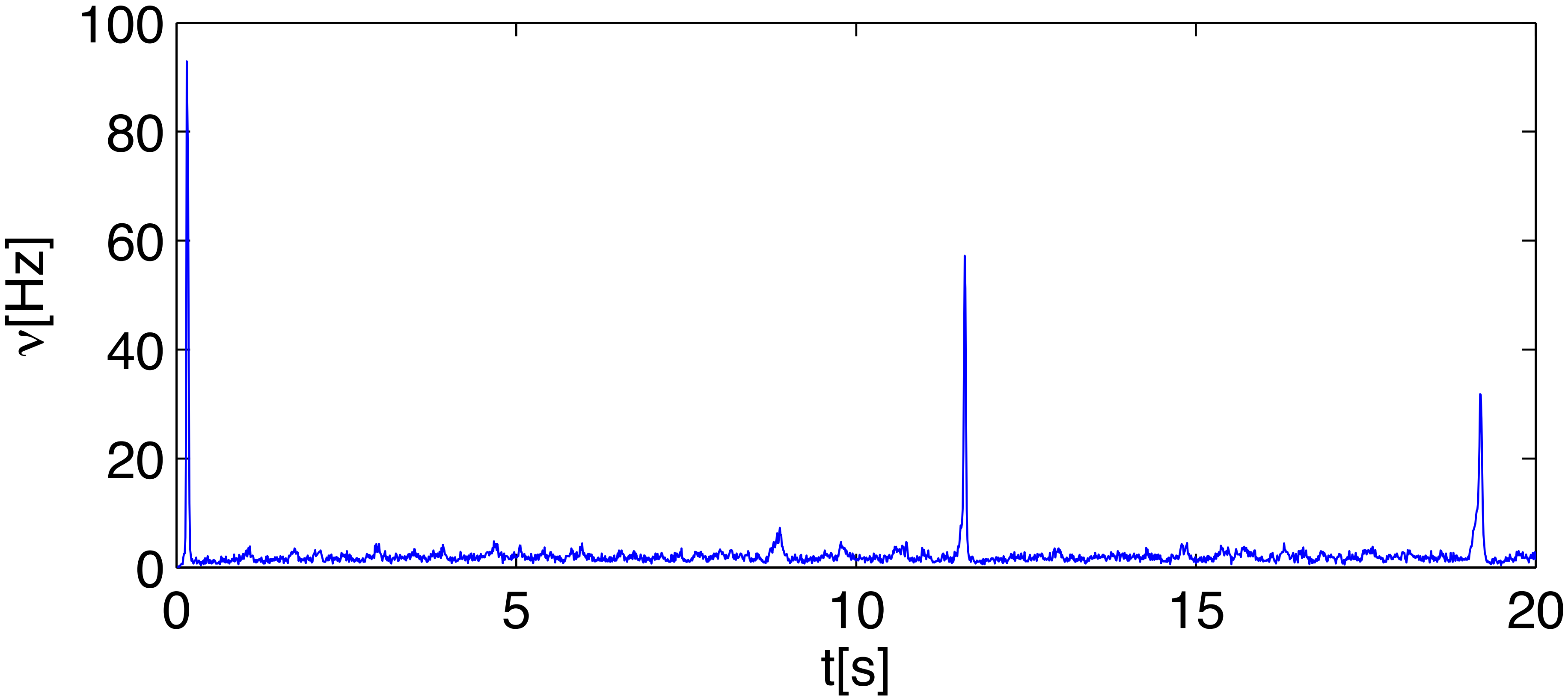


Inferred vs. true couplings



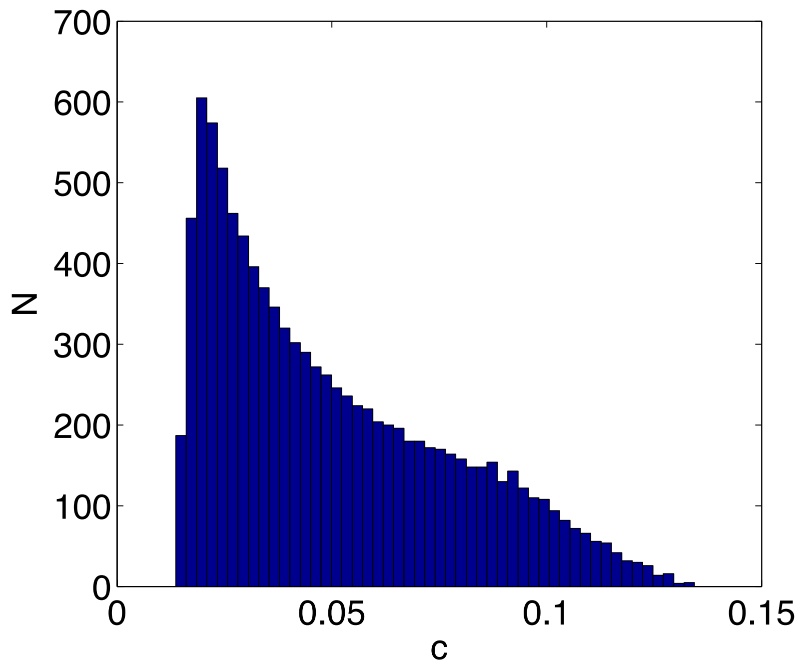
Bursting neural network

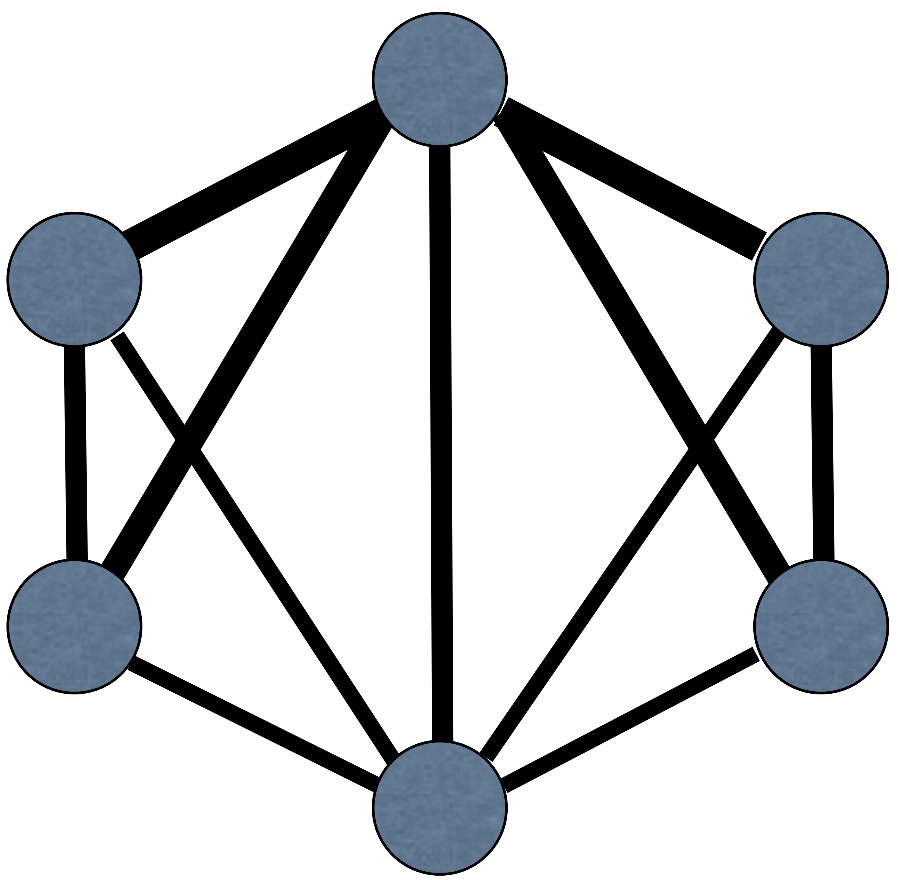




Bursting neural network

- 100 populations of 10 neurons each
  sparse and heterogenous connectivity, P(J) are broad

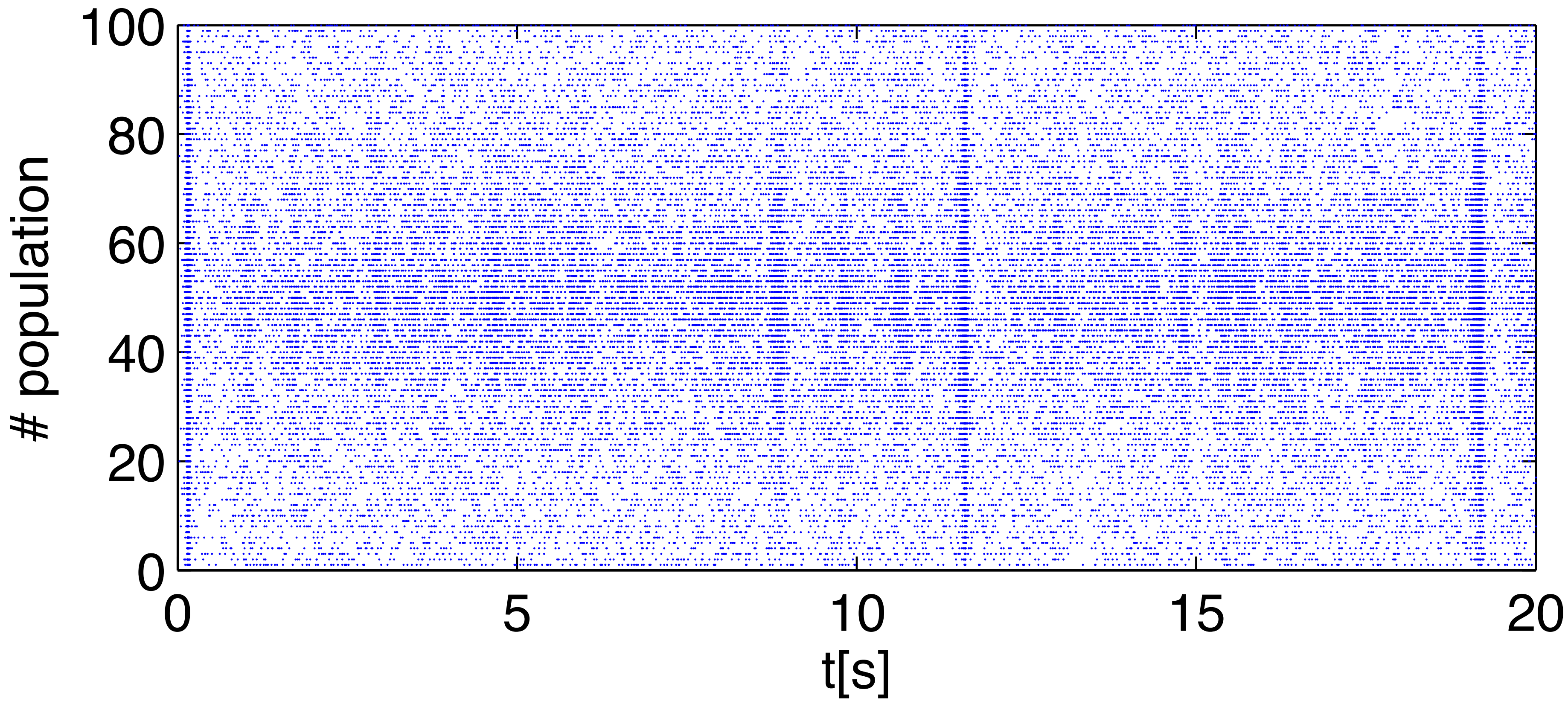




- delays $\delta_{ij} \in [0.1, 15]\, ms$

- simulates short-term depression

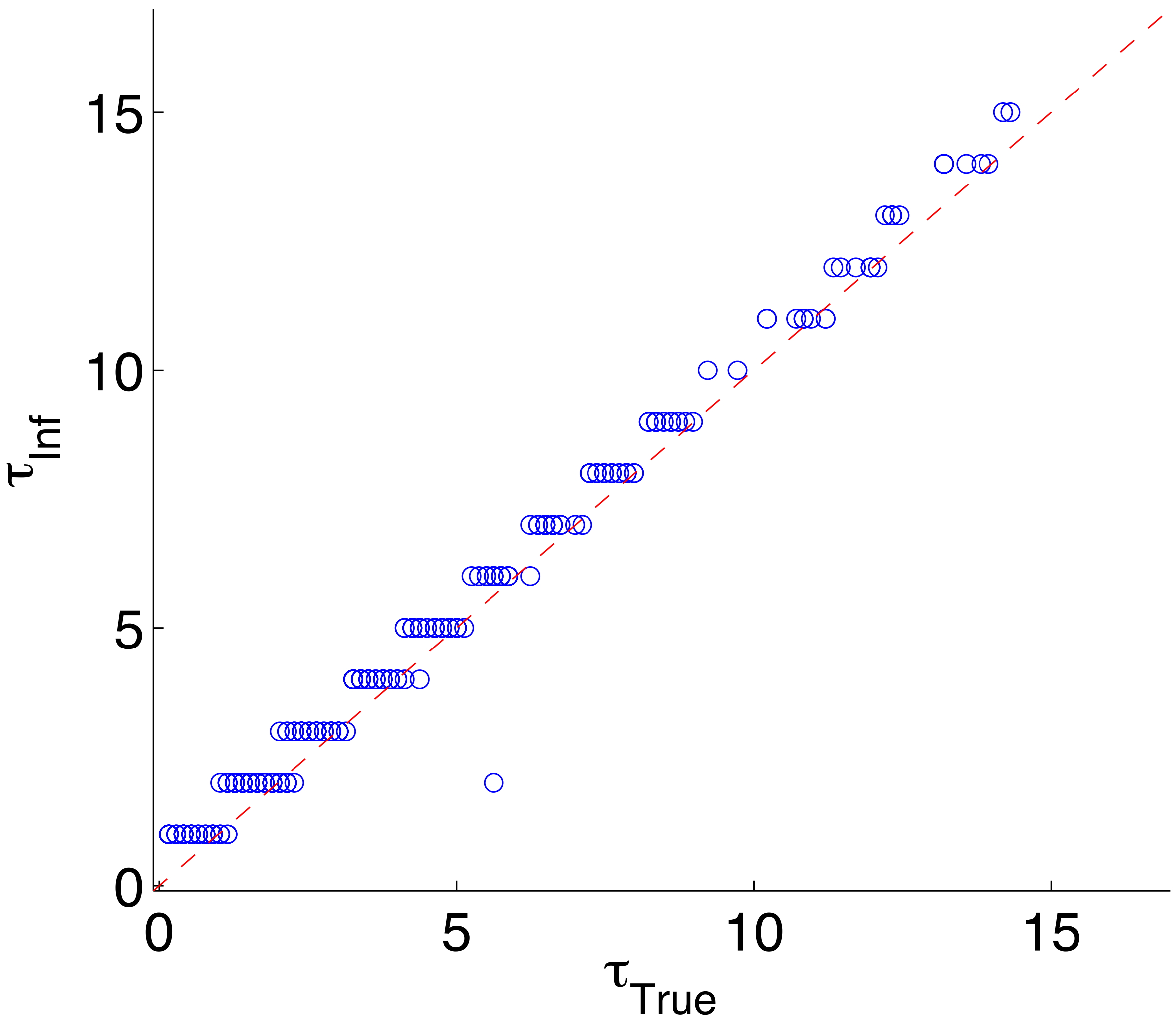
Bursting neural network

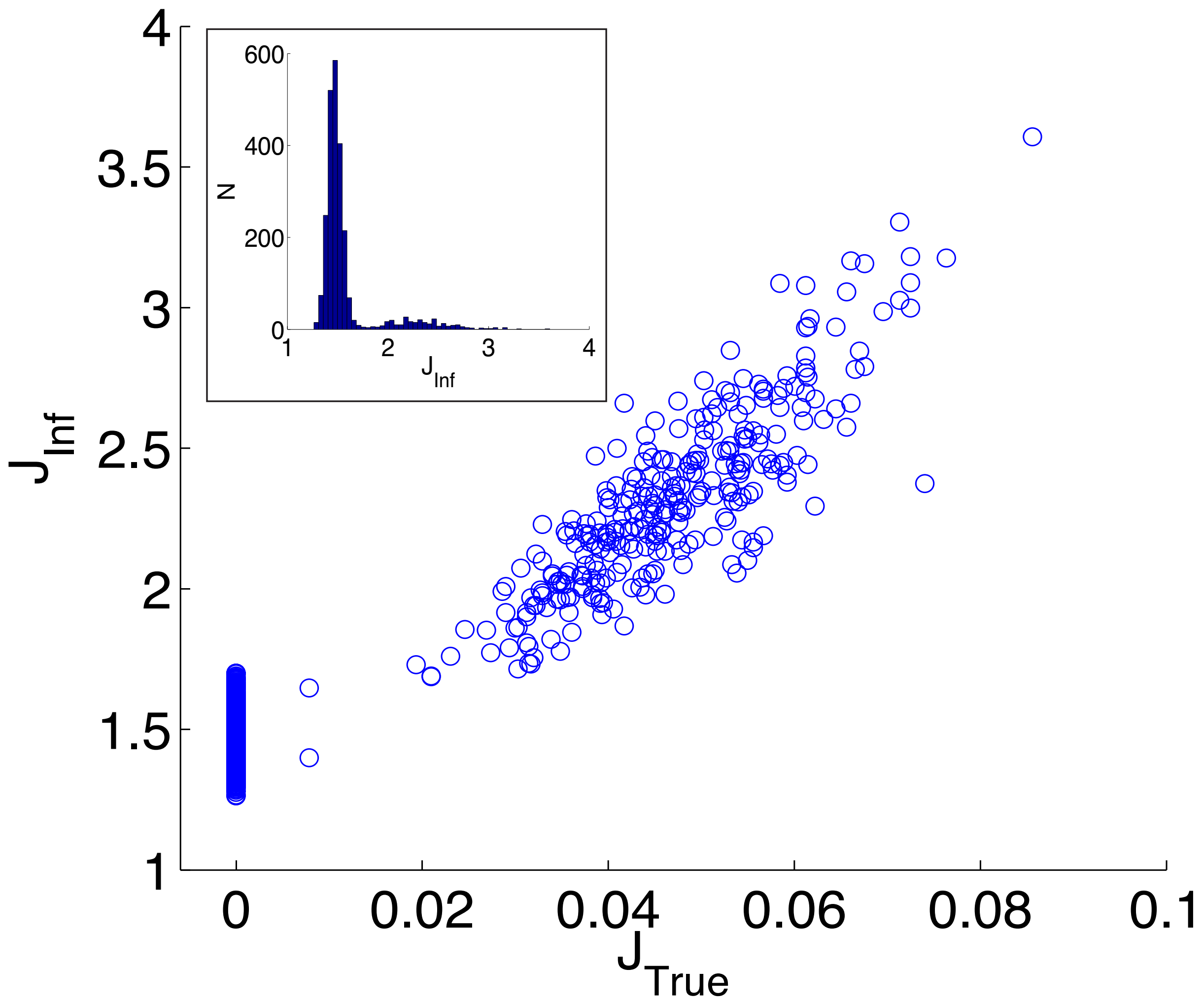


- Non-stationary, non-periodic, non-homogeneous
- We focus on the most active 50 neurons and try to infer model parameters by KIMTD

Bursting neural network

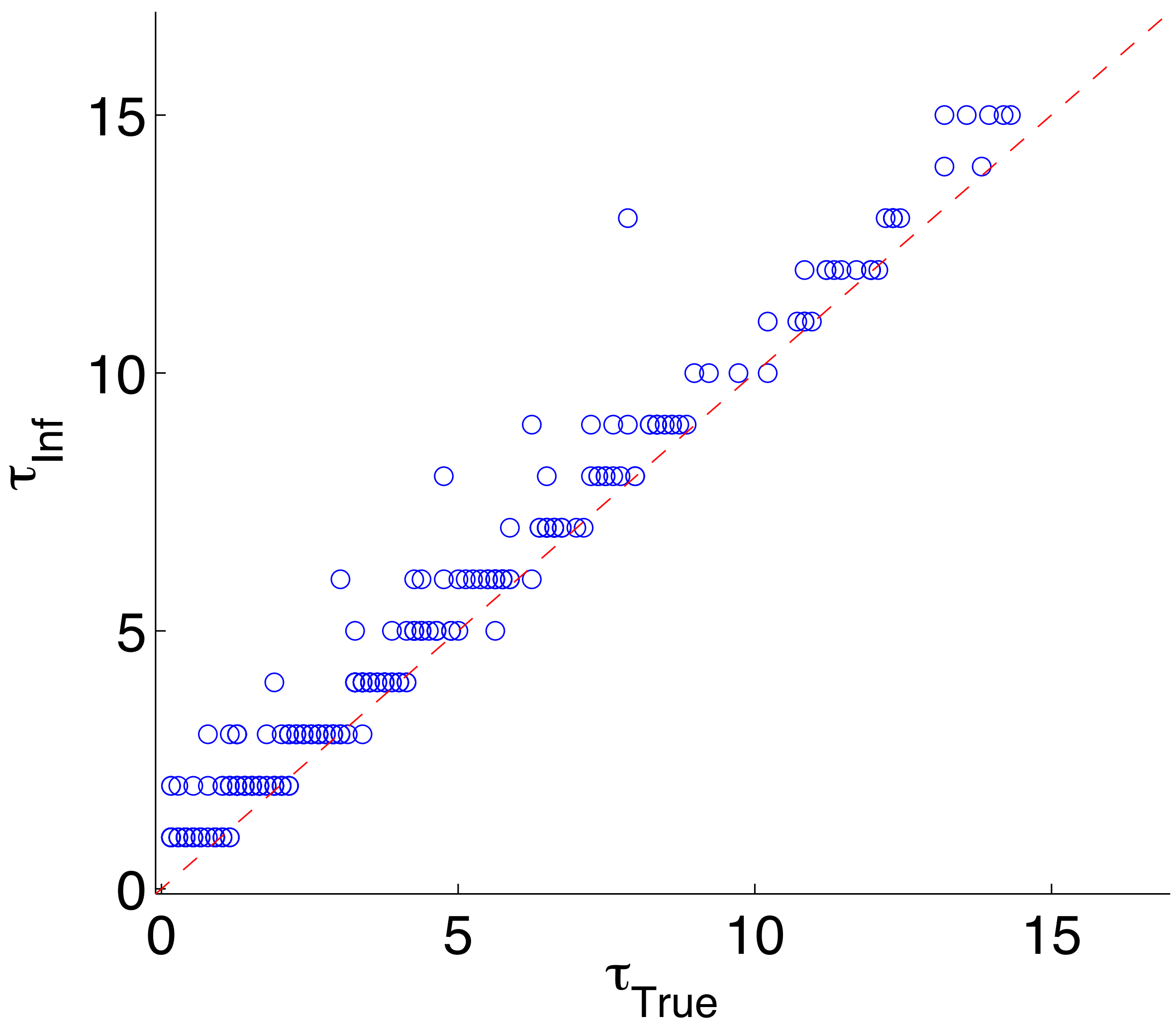
- Inference by KIMTD (preliminary results)
  measuring time **11 hours** (dt = 1 ms, ~4e7 bins)

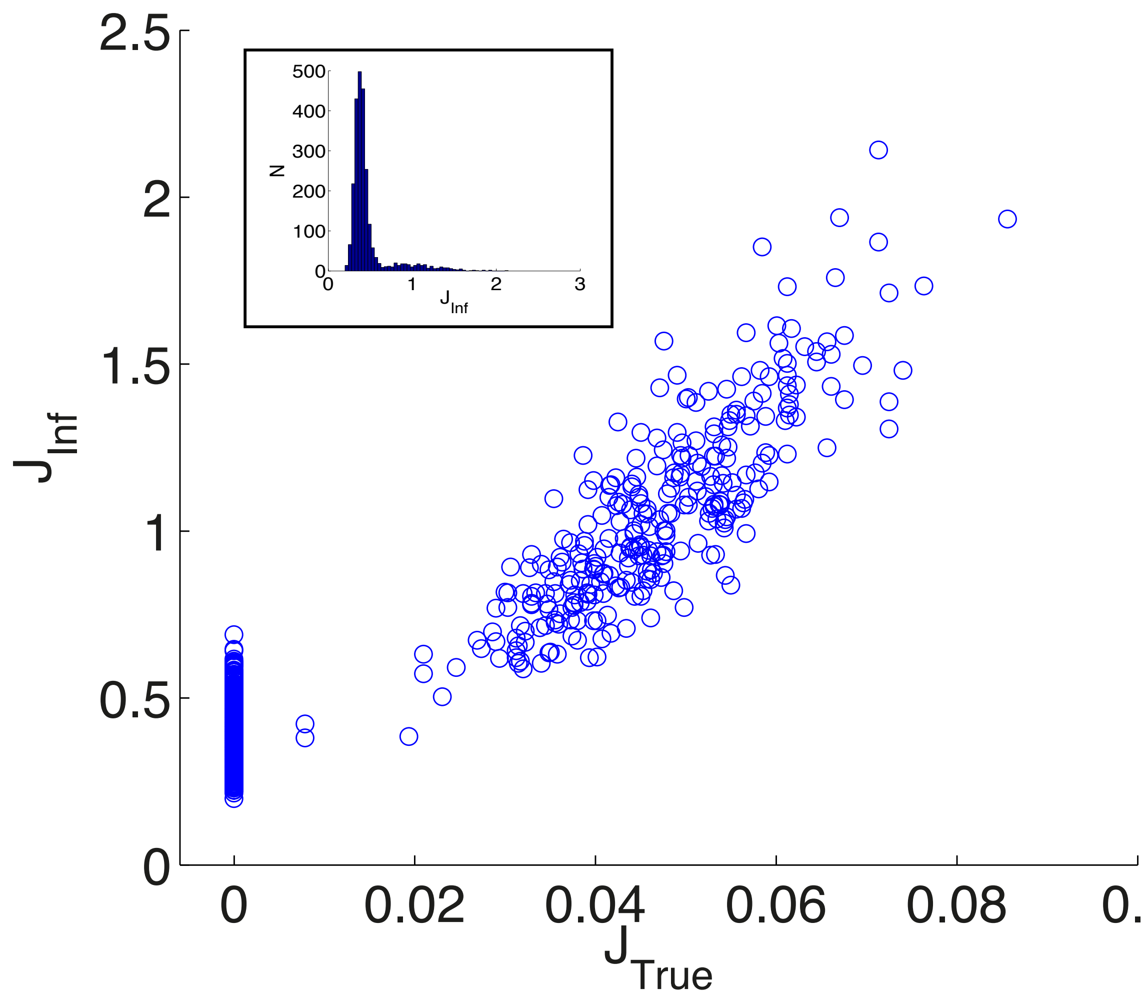




Bursting neural network

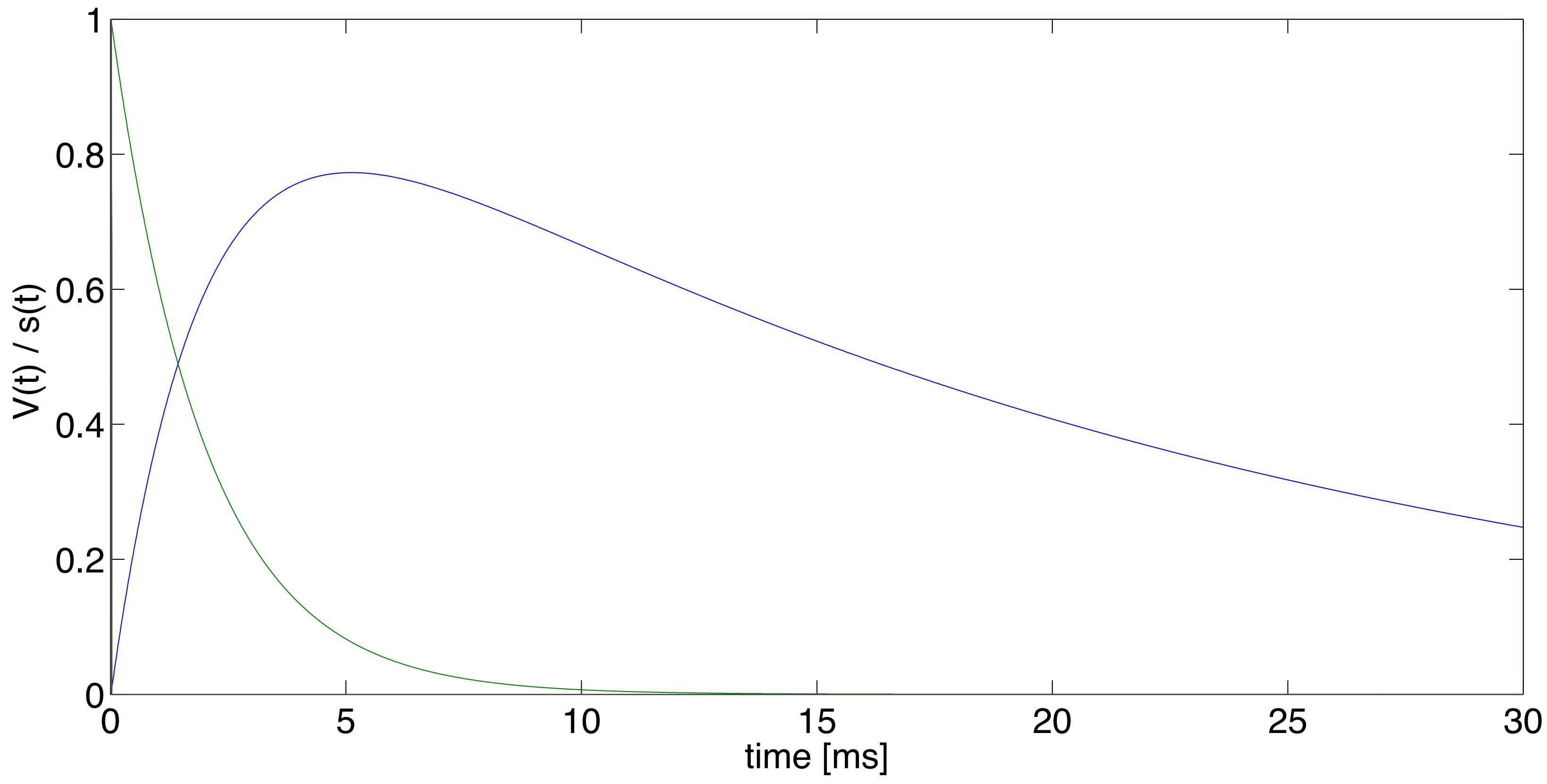
- Data filtering: using **only 2 hours** (1/5 than before) from the quiescent regimes





A more realistic neural network

- Synaptic transmission is not instantaneous
  Synaptic channels have their own dynamics with closing timescales between 1 and 100 ms

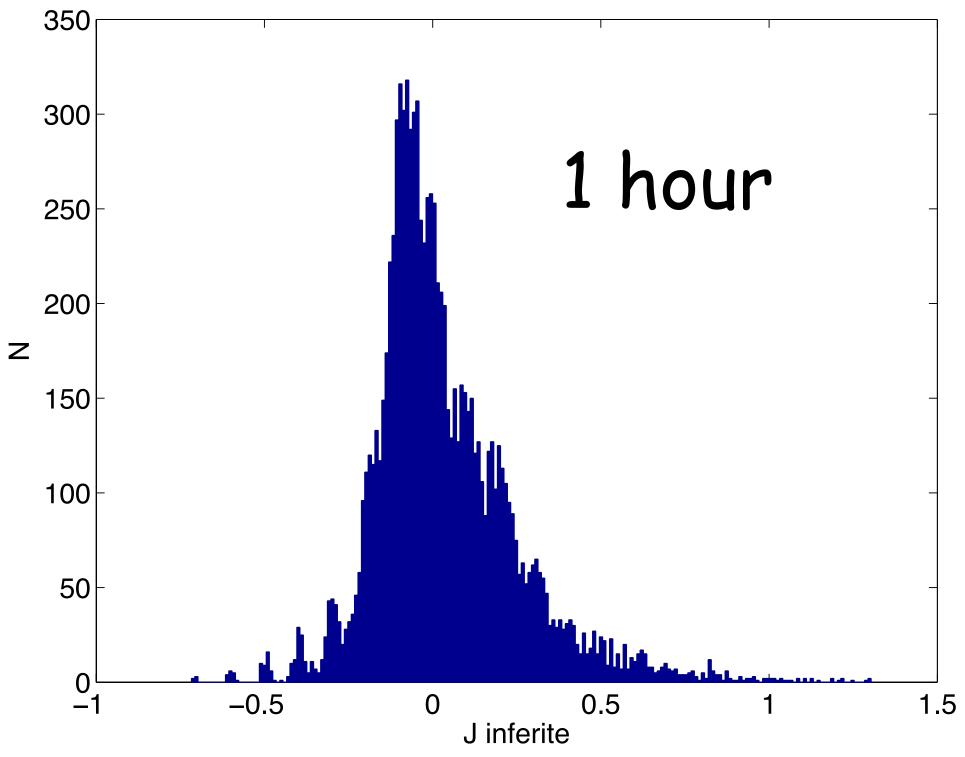


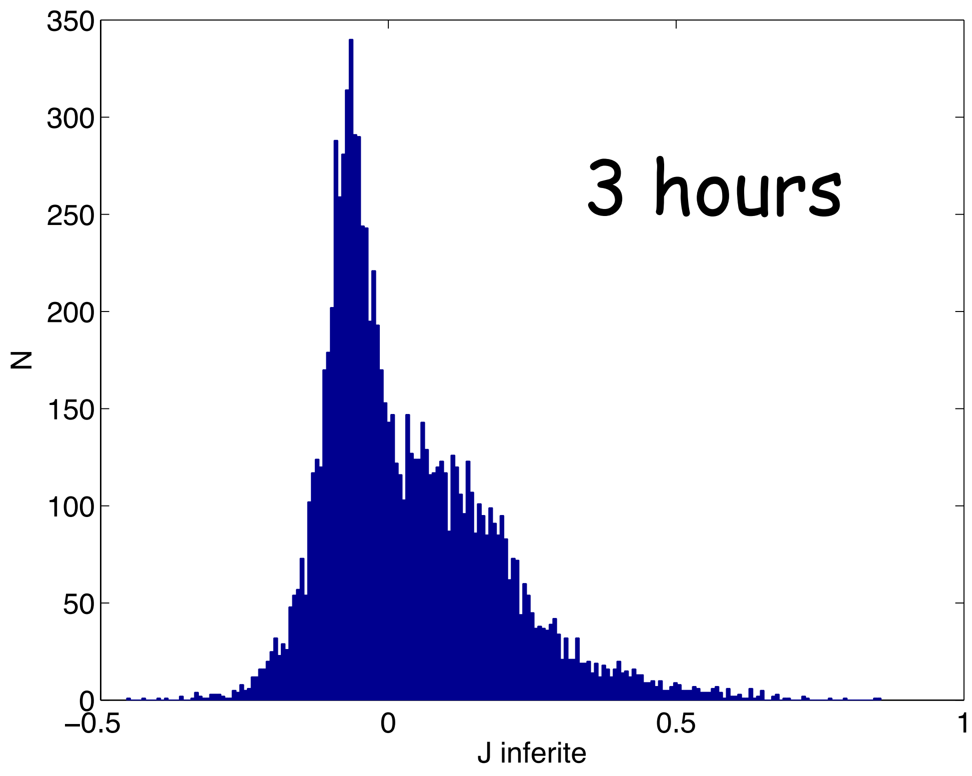
A more realistic neural network

- Fully connected topology
- 4 populations of neurons
- 10 different kind of synapsis, both excitatory and inhibitory
- No synaptic delay, but synaptic integration times between 1 and 100 ms
- Preliminary results...

A more realistic neural network

- Coupling inferred by standard KIM (dt = 2 ms)





KIM with memory

- Non-instantaneous synaptic transmission implies $S_i(t)$ may depend on $S_j(t-\Delta t)$ at several previous times
- Memory term to keep effectively interactions taking places at several time differences

$$\widehat{S}_j(t) \propto \sum_{\Delta t} S_j(t-\Delta t) e^{-\Delta t/\tau}$$

- Easy to compute (running averages)
- Mean values are unchanged $\langle \widehat{S}_i(t) \rangle_t = \langle S_i(t) \rangle_t$

KIM with memory

- Maximize $\mathcal{L} = \sum_{it} S_i(t+1) H_i(t) - \log 2 \cosh H_i(t)$

with $H_i(t) = h_i(t) + \sum_j J_{ij} \widehat{S}_j(t)$

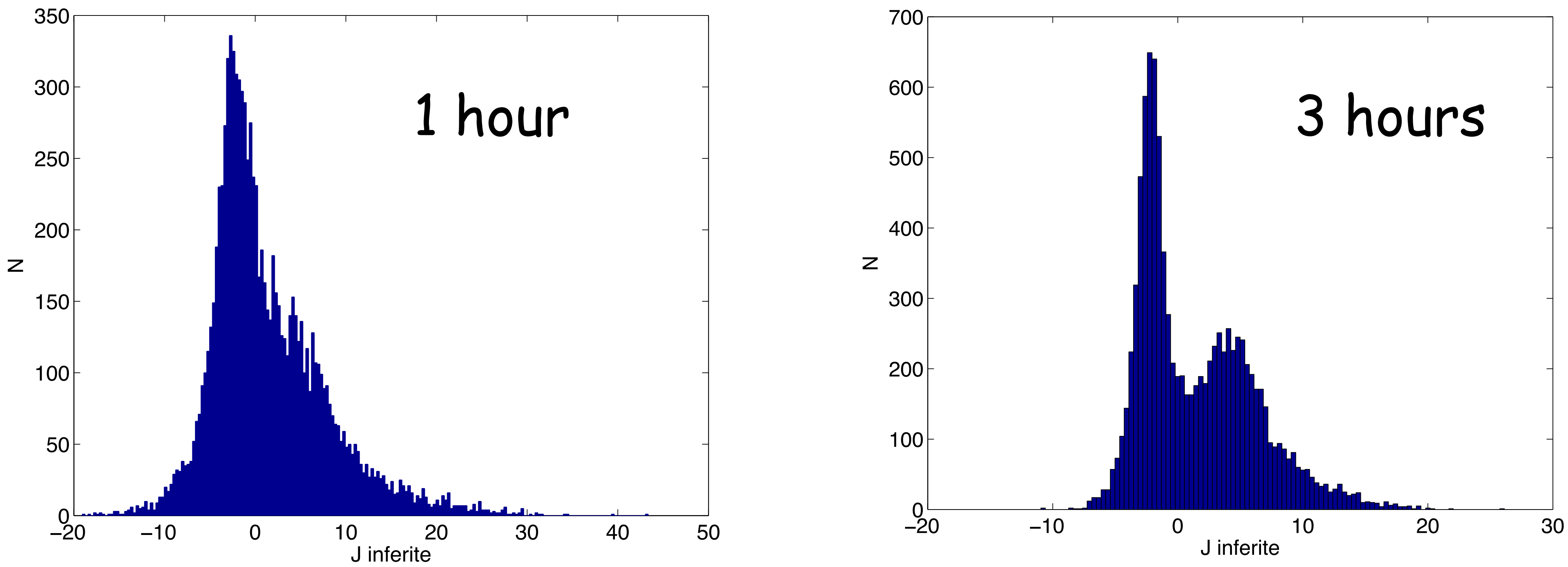
$$\frac{\partial \mathcal{L}}{\partial J_{ij}} = 0 \qquad \Longrightarrow \qquad J = A^{-1}\, \widehat{D}\, \widehat{C}^{-1}$$

$$\widehat{C}_{ij} = \langle \delta \widehat{S}_i(t)\, \delta \widehat{S}_j(t) \rangle_t$$

$$\widehat{D}_{ij} = \langle \delta S_i(t+1)\, \delta \widehat{S}_j(t) \rangle_t$$

A more realistic neural network

- Couplings inferred by KIM with memory, $\tau = 5\,ms$



Conclusions

- Simple MF approximations for solving the Kinetic inverse Ising problem seems to work also on “realistic” neural networks, if complemented with
  -> time delays
  -> finite integration times
  -> filtering out a quasi-stationary regime

- Choosing the right time binning is mandatory, otherwise inferred coupling may even have the wrong sign!

- Relation between true and inferred couplings is asymmetric and depends on the time binning
  -> a too small time-bin does not allow the separation of negative and zero couplings