Generalized Linear Models vs Back Propagation Through Time

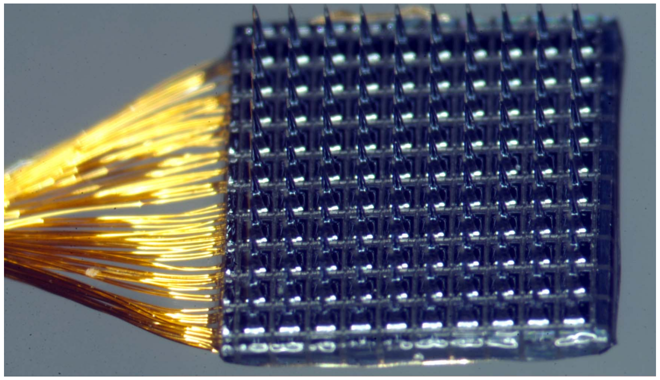
Sara A. Solla

Northwestern University

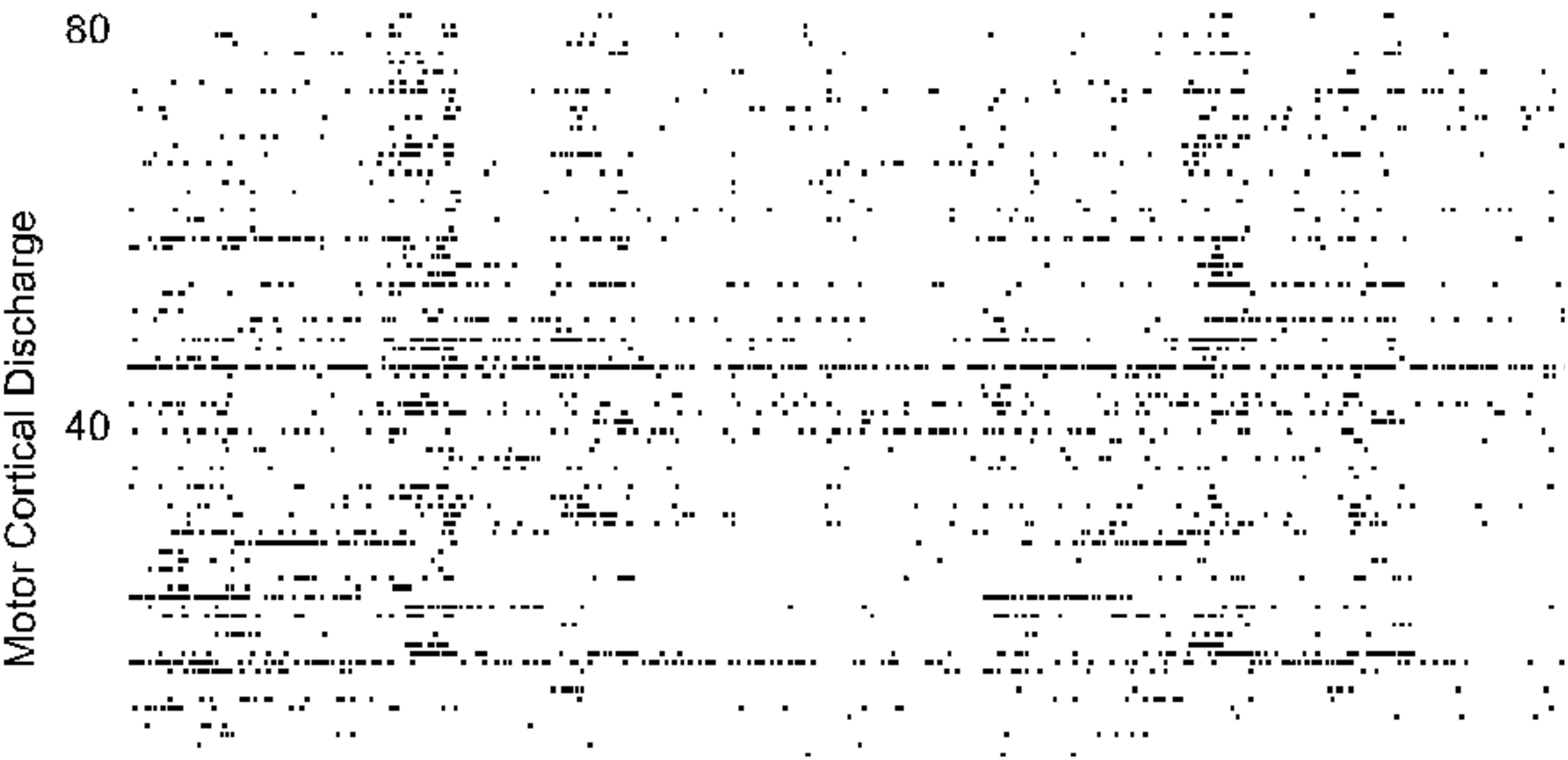
Statistical Physics of Machine Learning Summer School

Les Houches, July 4-29, 2022

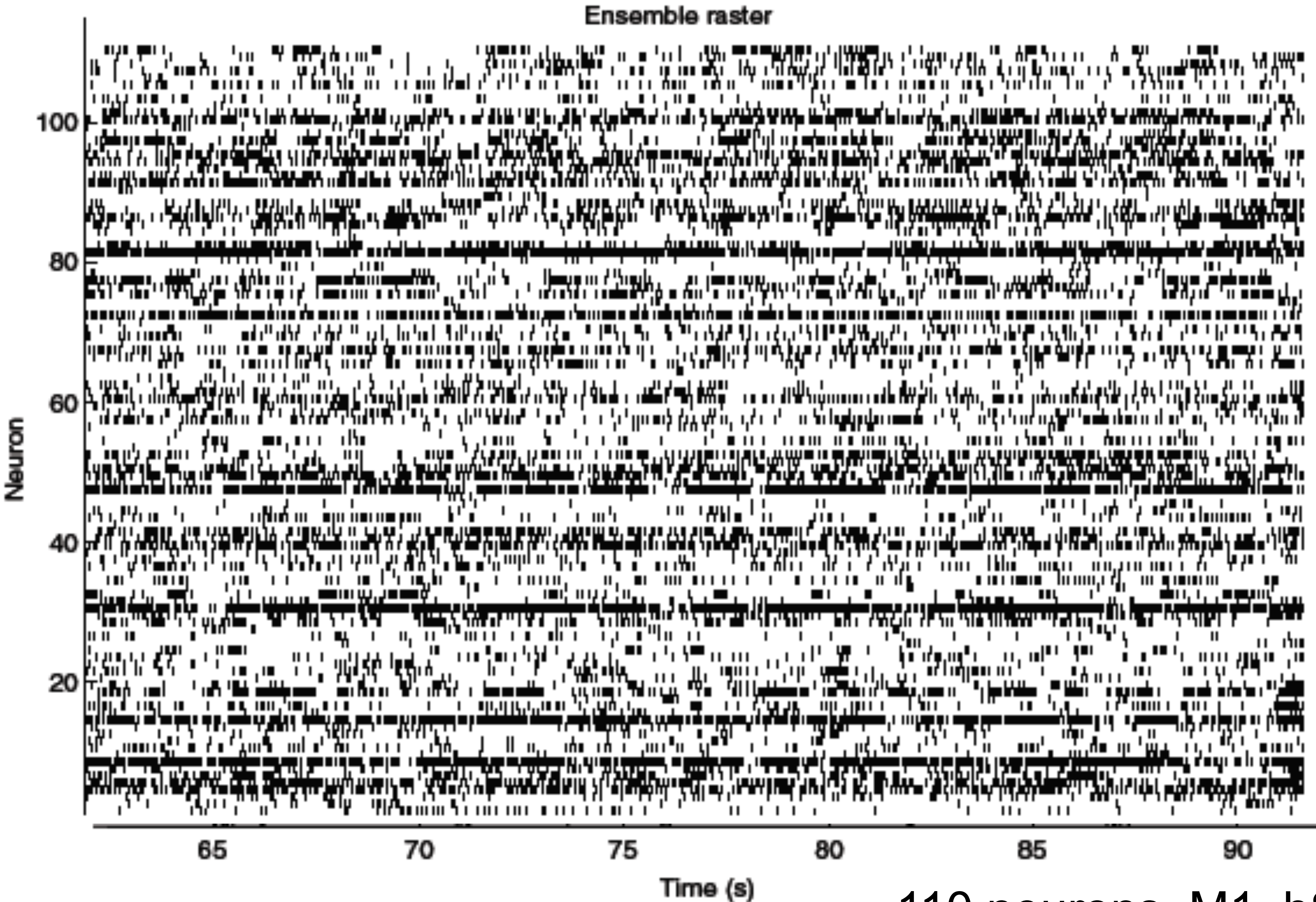
Neural population activity



- High-density multi-electrode arrays
- Record simultaneously one hundred neurons
- Characterize the dynamics of neural circuits.



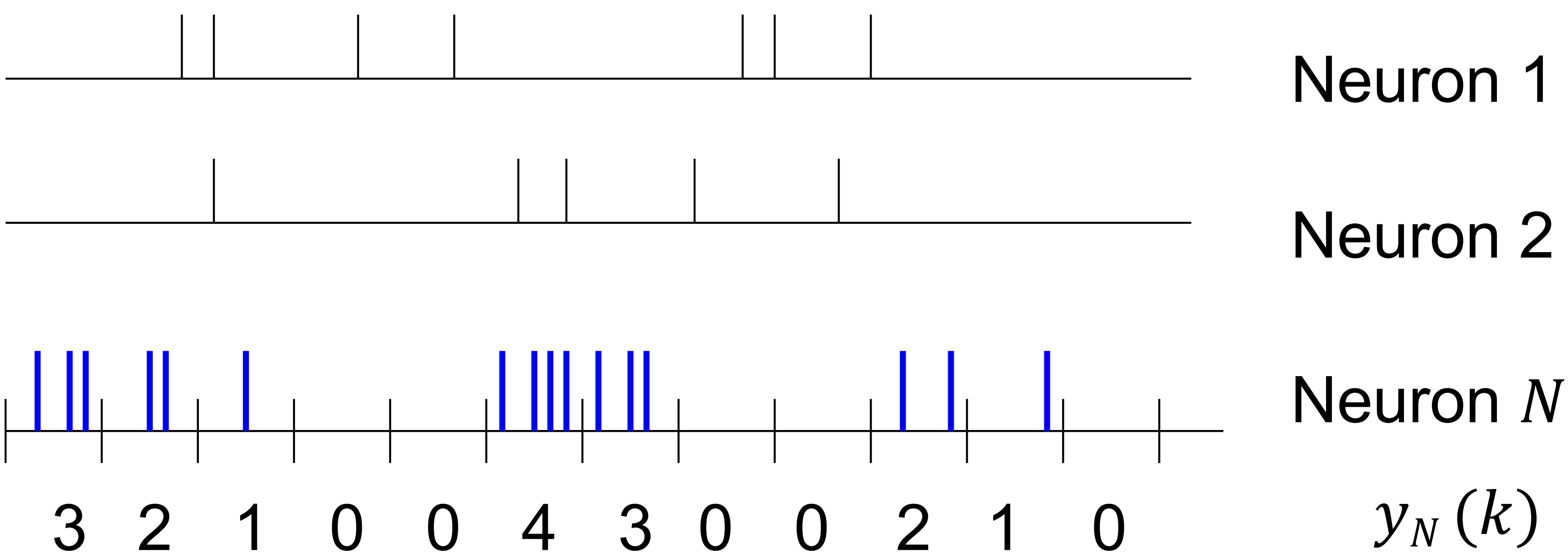
Neural population activity



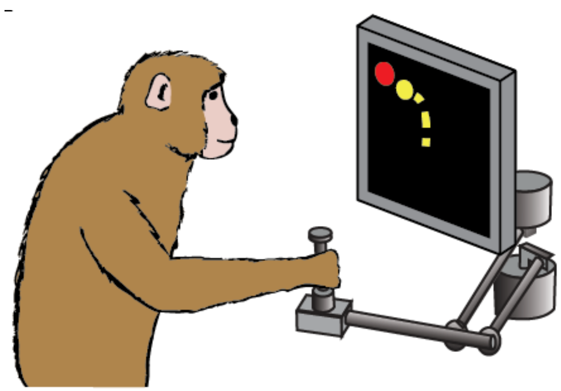
110 neurons, M1, hS3

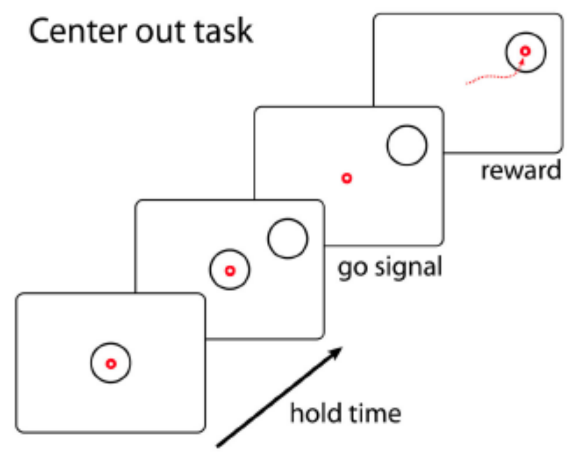
Analysis of neural activity

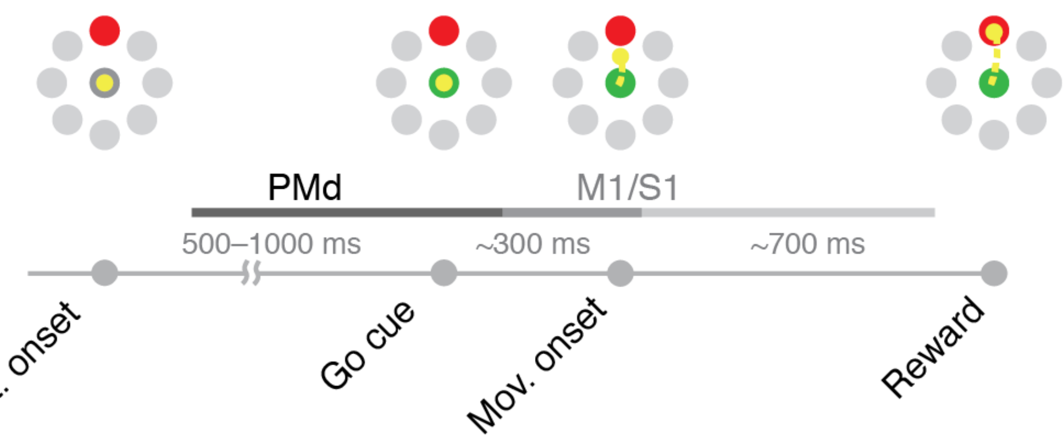
- Consider a population of N neurons whose spiking activity is observed during a time interval $(0, T]$.
- The interval is divided into K bins of size $\Delta = T/K$, labeled by an index $1 \leq k \leq K$.
- In each interval k we observe the number of spikes $y_i(k)$ emitted by neuron i, for all $1 \leq i \leq N$.



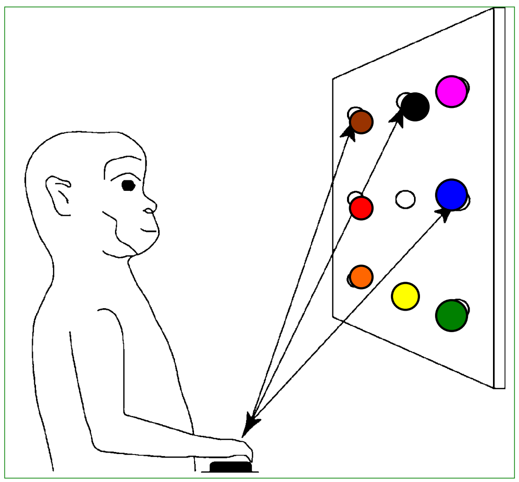
A simple motor task: center-out reaches
Instructed delay center-out reaching task

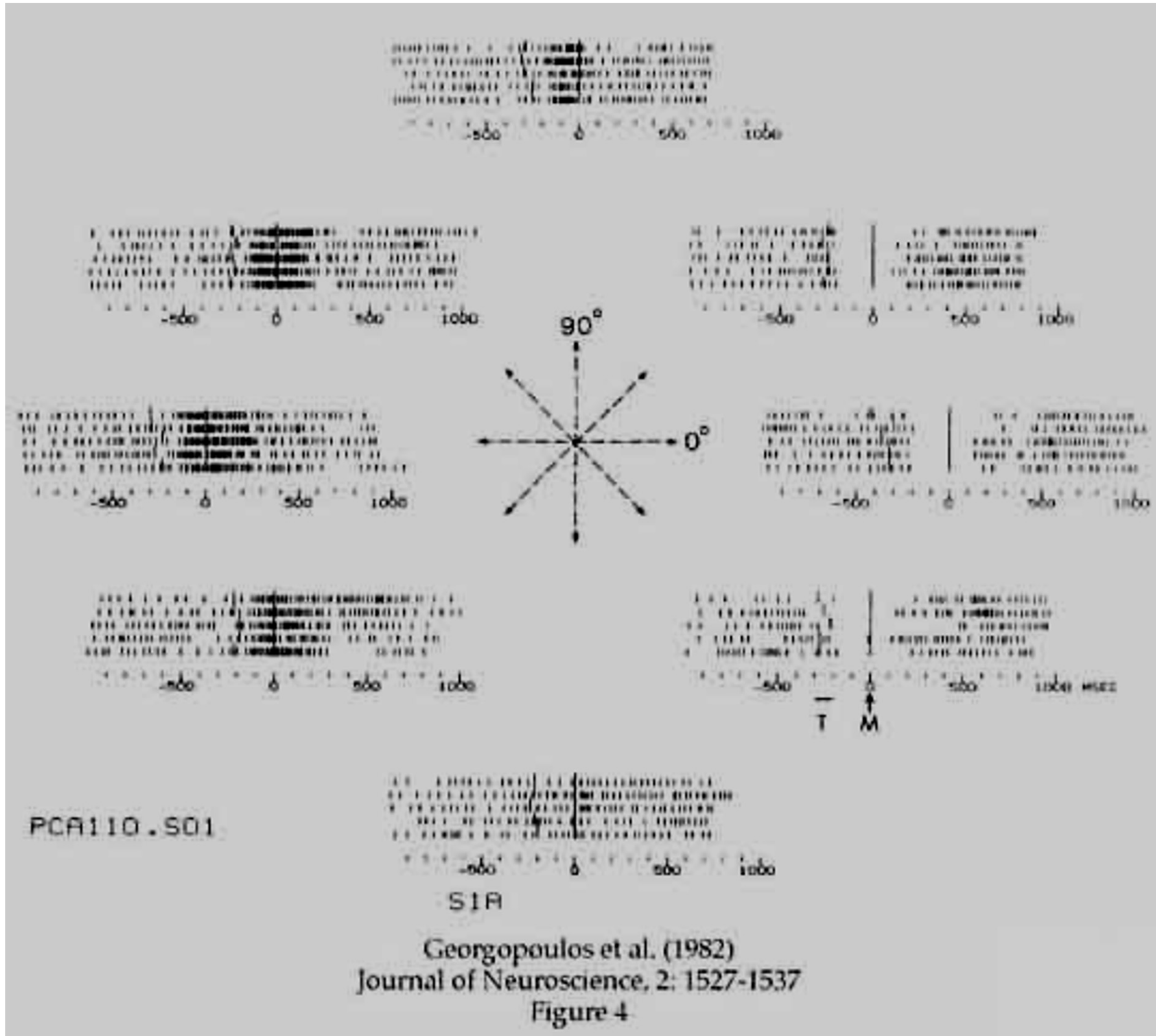






Neural activity: variability and specificity





Georgopoulos, Kalaska, Caminity, Massey *J. Neurosci.* (1982)

Poisson distribution

$$\rho(y|\lambda) = \frac{\lambda^y \, e^{-\lambda}}{y!}$$

The distribution is properly normalized:

$$\sum_{y=0}^{\infty} \frac{\lambda^y \, e^{-\lambda}}{y!} = e^{-\lambda} \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} = e^{-\lambda} \, e^{+\lambda} = 1$$

and has moments:

$$\mathrm{E}(y) = < y > = \lambda$$

$$\mathrm{Var}(y) = < (y - < y >)^2 > = \lambda$$

and Fano factor = $\dfrac{\mathrm{Var}(y)}{\mathrm{E}(y)} = 1$

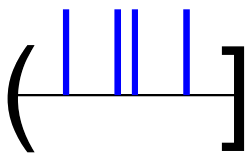
Variability and specificity

For Poisson statistics, the parameter $\lambda_i(k)$ is the mean or expectation value of the random variable $y_i(k)$. The trial-to-trial fluctuations of $y_i(k)$ about its mean $\lambda_i(k)$ describe the variability of neural activity.

The time-dependent parameter $\lambda_i(k)$ provides a tool for specificity: we will model $\lambda_i(k)$ through its relation to sensory stimuli, motor output, and the spiking activity of other neurons.

Neural activity as a Poisson process

$\left(t_{k-1}, t_k\right]$ $\quad 1 \le k \le K$, K bins of size $\Delta = T/K$

 $\quad y_i(k)$ = with probability $P\left(y_i(k)\right)$

- Data: $\{y_i(k)\}$, for $1 \le k \le K$ and for $1 \le i \le N$.
- The spiking activity of neuron i at time interval k is modeled as a Poisson process with mean $\lambda_i(k)$.
- The probability of observing precisely $y_i(k)$ spikes emitted by neuron i at time k is given by:

$$P\left(y_i(k) \middle| \lambda_i(k)\right) = \frac{\left(\lambda_i(k)\right)^{y_i(k)}\ e^{-\lambda_i(k)}}{y_i(k)!}$$

Likelihood of observed spikes

What is the probability of the observed data $\{y_i(k)\}$ given the parameters $\{\lambda_i(k)\}$?

$$P_T(\{y_i(k)\} | \{\lambda_i(k)\}) = \prod_{i=1}^{N} \prod_{k=1}^{K} \frac{(\lambda_i(k))^{y_i(k)} \; e^{-\lambda_i(k)}}{y_i(k)!}$$

Consider the *log-likelihood*, the logarithm of the probability:

$$L_T\left(\{y_i(k)\} | \{\lambda_i(k)\}\right) = \ln\left\{ \prod_{i=1}^{N} \prod_{k=1}^{K} \frac{(\lambda_i(k))^{y_i(k)} \; e^{-\lambda_i(k)}}{y_i(k)!} \right\} =$$

$$= \left\{ \sum_{i=1}^{N} \sum_{k=1}^{K} \left(y_i(k) \ln \lambda_i(k) - \lambda_i(k) - \ln\left(y_i(k)! \right) \right) \right\}$$

Likelihood of observed spikes

$$L_T\left(\left\{y_i(k)\right\}\right) = \left\{\sum_{i=1}^{N}\sum_{k=1}^{K}\left(y_i(k)\ln\lambda_i(k) - \lambda_i(k) - \ln\left(y_i(k)!\right)\right)\right\}$$

Here, $\{y_i(k)\}$ is the data while the neuron specific and time specific firing rates $\{\lambda_i(k)\}$ are the parameters of the model.

QUESTION: if you were to estimate the parameters $\{\lambda_i(k)\}$ to maximize the likelihood of the data $\{y_i(k)\}$, what would you obtain?

Model for $\{\lambda_i(k)\}$

GOAL: find a model for the $\{\lambda_i(k)\}$, to relate their values to sensory stimuli, motor outputs, and the activity of other neurons in the network.

QUESTION: How to model the $\{\lambda_i(k)\}$?

APPROACH: Generalized Linear Models (GLMs)

But first, a detour into statistics: EXPONENTIAL FAMILY OF PROBABILITY DISTRIBUTIONS.

Exponential family

The exponential family encompasses probability distributions of the form:

$$\rho(y \mid \delta, \varphi) = \exp\left\{\frac{y\delta - b(\delta)}{a(\varphi)} + c(y, \varphi)\right\}$$

Here, y is the random variable whose probability density function is given by ρ. The distribution is parametrized by the canonical parameter δ and the dispersion parameter φ. The functions a(.), b(.), and c(.,.) need to be specified, and define the various distributions within the family.

The term $c(y, \varphi)$ plays an important role: it provides a normalization function that guarantees $\int dy\, \rho\left(y \middle| \delta, \varphi\right) = 1$ for all δ, φ.

Exponential family

$$\rho_y(y \mid \delta, \varphi) = \exp\left\{ \frac{y\delta - b(\delta)}{a(\varphi)} + c(y, \varphi) \right\}$$

Since $\int dy\, \rho_y(y \mid \delta, \varphi) = 1$ for all δ, φ then:

$$\frac{\partial}{\partial \delta} \int dy\, \rho_y(y \mid \delta, \varphi) = 0 \qquad \Rightarrow \qquad \mathrm{E}(y) = b'(\delta)$$

$$\frac{\partial^2}{\partial \delta^2} \int dy\, \rho_y(y \mid \delta, \varphi) = 0 \qquad \Rightarrow \qquad \mathrm{Var}(y) = a(\varphi)\, b''(\delta)$$

Note that the canonical parameter δ fully determines the mean E(y) through $b(\delta)$, while the variance Var(y) requires additional information provided by the dispersion parameter through $a(\varphi)$.

McCullagh, Nelder, *Generalized Linear Models* (1989)

Exponential family

Consider the family of canonical exponential distributions with canonical parameter δ and dispersion parameter φ:

$$\rho(y \mid \delta, \varphi) = \exp\left\{\frac{y\delta - b(\delta)}{a(\varphi)} + c(y, \varphi)\right\}$$

WHY CARE? because the normal, Bernoulli, binomial, multinomial, Poisson, gamma, geometric, chi-square, beta, and a few other distributions are all exponential distributions.

POISSON DISTRIBUTION:

$$\rho\left(y \middle| \lambda\right) = \frac{\lambda^y \, e^{-\lambda}}{y!} = \exp\left\{y \ln\lambda - \lambda - \ln\left(y!\right)\right\}$$

with $a(\varphi) = 1,\ \delta = \ln\lambda,\ b(\delta) = \lambda,$ and $c(\varphi, y) = -\ln\left(y!\right)$

Poisson distribution: member of the exponential family

The relations: $$\begin{cases} \mathrm{E}(y) = b'(\delta) \\ \mathrm{Var}(y) = a(\varphi)\; b''(\delta) \end{cases}$$

hold for any probability density function within the exponential family. When applied to the Poisson case, they imply:

$$\mathrm{E}(y) = b'(\delta) = \lambda$$

$$\mathrm{Var}(y) = a(\varphi)\; b''(\delta) = \lambda$$

For Poisson statistics, $\mathrm{E}(y) = \mathrm{Var}(y) = \lambda$ implies:

$$b'(\delta) = a(\varphi)\; b''(\delta) \quad \Rightarrow \quad \begin{cases} a(\varphi) = 1 \\ b(\delta) = e^{\delta} = \lambda \end{cases}$$

GLM: Poisson distribution

In a generalized linear model for a probability distribution that is a member of the exponential family, the expectation value E(y) is related to the canonical parameter δ via a nonlinear link function ***g***:

$$g\big(\mathrm{E}(y)\big) = \delta \qquad \mathrm{E}(y) = g^{-1}(\delta)$$

This is the only nonlinearity in the model, as the canonical parameter δ is constructed as a linear combination of all observed variables that can *explain* the random variable y.

In the Poisson case, $\delta = \log \lambda = \log (\mathrm{E}(y))$, and the nonlinear link function ***g*** is the logarithm!

$$\lambda = \mathrm{E}(y) = g^{-1}(\delta) = \exp(\delta)$$

$$\delta = g(\lambda) = \ln(\lambda)$$

Generalized Linear Model for spikes

$$L_T\left(\{y_i(k)\}\right) = \left\{\sum_{i=1}^{N}\sum_{k=1}^{K}\left(y_i(k)\ln\lambda_i(k) - \lambda_i(k) - \ln\left(y_i(k)!\right)\right)\right\}$$

The parameter $\lambda_i(k)$ is the time-dependent mean of a Poisson process. In a GLM for a Poisson distribution, it is the logarithm (link function) of the mean that is expressed as a linear combination of all observed variables that can be used to explain the observed firing rates.

Internal covariates: preceding neural activity
(hidden neurons)

External covariates: sensory stimulus (input)
direction of motion (output)

Internal covariates: spiking history

We know the spiking history of the ensemble of N neurons up to the current time t. We denote this as the spiking history of the ensemble:

$$H(t) = \left\{ \left\{ y_i(t') \right\}_{i=1}^{N}, t' \leq t \right\}$$

Given this information, what is our expectation of the number of spikes that neuron i will fire in the interval $(t, t + \Delta)$? This is the conditional intensity $\lambda_i\left(t \mid H(t)\right)$, a strictly positive function that provides a history-dependent generalization of the time dependent rate of an inhomogeneous Poisson process.

Model for $\lambda_i\left(t \mid H(t)\right)$: generalized linear model (GLM)

Truccolo, Eden, Fellows, Donoghue, Brown, *J. Neurophysiol.* (2005)

GLM: internal covariates

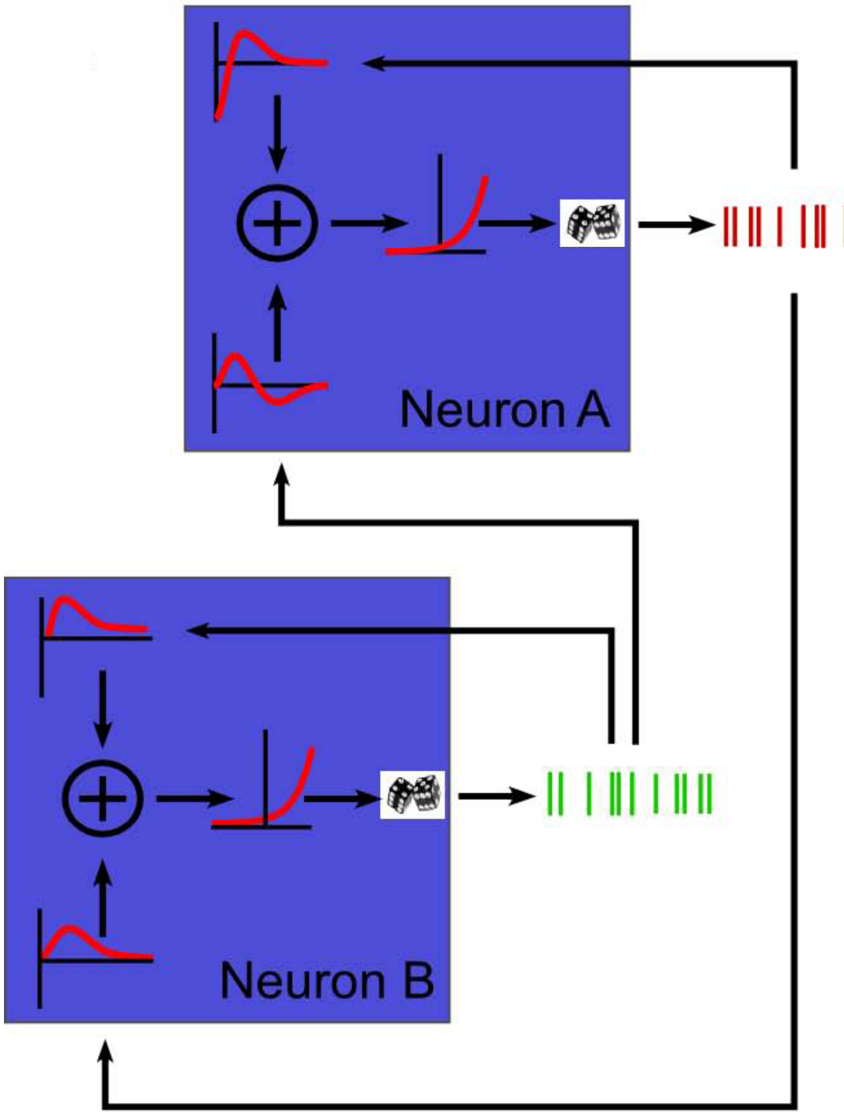
$$\delta_i\big(t \mid H(t),\{\alpha\}\big) = \ln \lambda_i\big(t \mid H(t),\{\alpha\}\big)$$

$$= \alpha_{i0} + \sum_{j=1}^{N} \sum_{m=1}^{\tau_N} \alpha_{ij}(m)\, y_j(t-m)$$

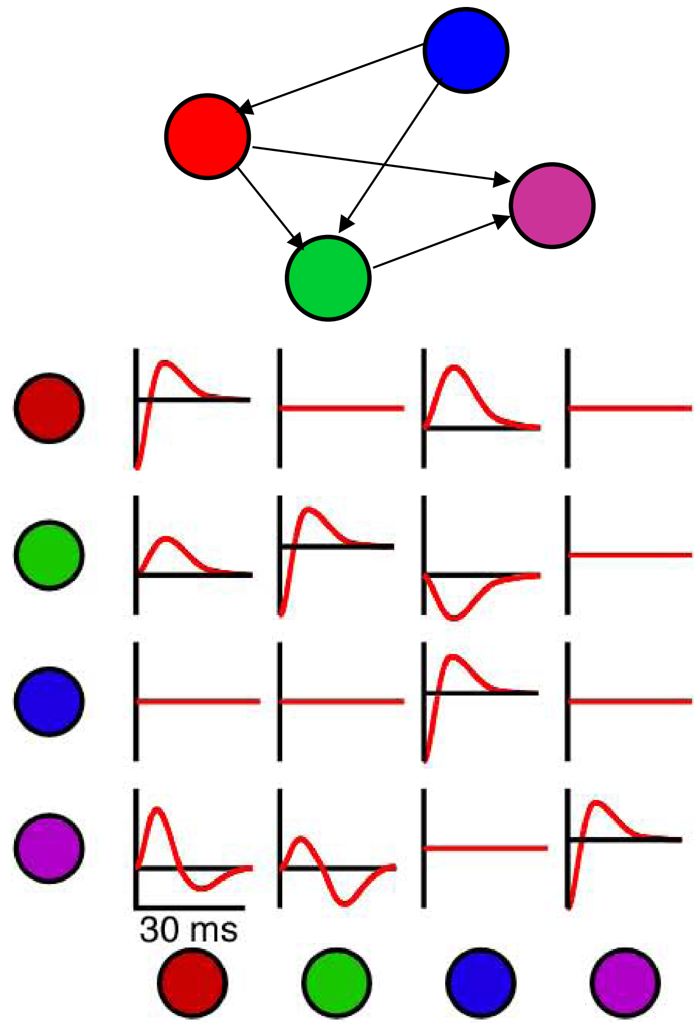
$$\lambda_i\big(t \mid H(t),\{\alpha\}\big) = \exp\left\{\alpha_{i0} + \sum_{j=1}^{N} \sum_{m=1}^{\tau_N} \alpha_{ij}(m)\, y_j(t-m)\right\}$$

Linear-Nonlinear (NL) model! Here, the kernel parameter $\alpha_{ij}(m)$ quantifies the effect that the spiking activity of neuron j at time bin $(t-m)$ has on the spiking activity of neuron i at time bin t.

GLM: $$\lambda_i\left(t \mid H(t),\{\alpha\}\right) = \exp\left\{\alpha_{i0} + \sum_{j=1}^{N}\sum_{m=1}^{\tau_N}\alpha_{ij}(m)\,y_j(t-m)\right\}$$



Generative model



Functional connectivity

GLM: maximum likelihood

Given the data $\{y_i(k)\}$, find the parameters $\{\lambda_i(k)\}$ that maximize

$$L_T\left(\{y_i(k)\}\right) \propto \left\{ \sum_{i=1}^{N} \sum_{k=1}^{K} \left(y_i(k) \ln \lambda_i(k) - \lambda_i(k) \right) \right\}$$

A term that does not depend on $\{\lambda_i(k)\}$ has been dropped.

MODEL: $$\ln \lambda_i(k) = \alpha_{i0} + \sum_{j=1}^{N} \sum_{m=1}^{\tau_N} \alpha_{ij}(m)\, y_j(k-m)$$

Given the data $\{y_i(k)\}$, find the parameters $\{\alpha\}$ that maximize

$$L_T\left(\{y_i(k)\} \mid \{\alpha\}\right) \propto \left\{ \sum_{i=1}^{N} \sum_{k=1}^{K} \left(y_i(k) \left[\alpha_{i0} + \sum_{j=1}^{N} \sum_{m=1}^{\tau_N} \alpha_{ij}(m)\, y_j(k-m) \right] - \right.\right.$$

$$\left.\left. - \exp\left[\alpha_{i0} + \sum_{j=1}^{N} \sum_{m=1}^{\tau_N} \alpha_{ij}(m)\, y_j(k-m) \right] \right) \right\}$$

Iterative gradient ascent

Consider a network of N neurons. The data is of the form $\{y_i(k)\}$, for $1 \le k \le K$, $1 \le i \le N$. The GLM for the likelihood of the data is:

$$
\begin{aligned}
L_T\left(\{y_i(k)\} \mid \{\alpha\}\right) \propto \Bigg\{ \sum_{i=1}^{N} \sum_{k=1}^{K} \Bigg(& y_i(k) \left[\alpha_{i0} + \sum_{j=1}^{N} \sum_{m=1}^{\tau_N} \alpha_{ij}(m)\, y_j(k-m) \right] - \\
& - \exp\left[\alpha_{i0} + \sum_{j=1}^{N} \sum_{m=1}^{\tau_N} \alpha_{ij}(m)\, y_j(k-m) \right] \Bigg) \Bigg\}
\end{aligned}
$$

We want to find the maximum likelihood parameters $\{\alpha^*\}$ using an iterative gradient ascent method with adaptive step.

To implement this algorithm, we need to compute the first and second derivatives of the likelihood with respect to the parameters $\{\alpha\}$.

Likelihood: first derivative

The gradient that drives the uphill search is given by:

$$\left.\frac{\partial L_T\left(\{y_i(k)\} \mid \{\alpha\}\right)}{\partial \alpha_{ij}(m)}\right|_{(\mu)} = \sum_{k=1}^{K} \left\{ \left[y_i(k) - < y_i(k) >^{(\mu)} \right] y_j(k-m) \right\}$$

The update of the parameter $\alpha_{ij}(m)$ is given by the product of the activity $y_j(k-m)$ of the *presynaptic* neuron j at time lag m and the difference between the actual activity $y_i(k)$ of the *postsynaptic* cell and our current estimate of it at iteration (μ). The rule presynaptic activity x postsynaptic error is a famous learning rule, called the Delta Rule.

I have italicized *presynaptic* and *postsynaptic* because I do not mean to imply that the parameter $\alpha_{ij}(m)$ is an actual synaptic strength.

Likelihood: second derivative

The components of the Hessian matrix of second derivatives that controls the size of the uphill steps are given by:

$$\left.\frac{\partial^2 L_T\left(\{y_i(k)\} \mid \{\alpha\}\right)}{\partial\alpha_{ij}(m)\,\partial\alpha_{ij'}(m')}\right|_{(\mu)} = -\sum_{k=1}^{K}\left\{< y_i(k) >^{(\mu)} \; y_j(k-m)\; y_{j'}(k-m')\right\}$$

Now there are two *presynaptic* neurons: neuron *j at time lag* m *a*nd neuron *j '* at time lag m*'.* Their activities are multiplied, and this product is weighted by our current estimate of the activity of the *postsynaptic* neuron.

Note the overall minus sign! The variables $\{y\}$ represent number of spikes emitted during a bin of size Δ. These variables and their averages are always non-negative.

Every component of the Hessian matrix is negative - the surface is everywhere convex!

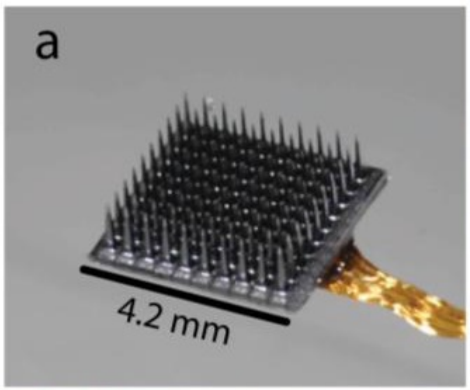
Likelihood maximization

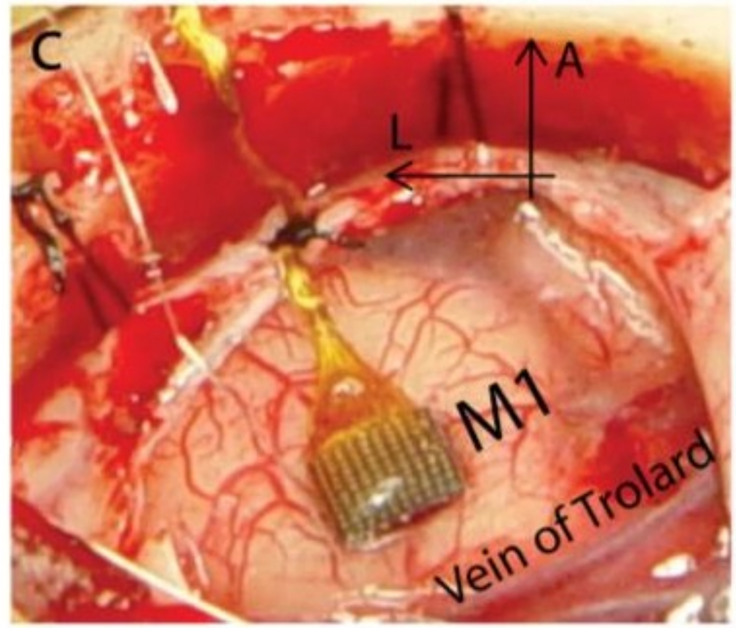
The algorithm can now be written as follows:

$$\vec{\alpha}^{(\mu+1)} = \vec{\alpha}^{(\mu)} + \mathrm{E}^{(\mu)}\vec{\nabla}L^{(\mu)}$$

Here, $\vec{\alpha}$ is a listing of all the parameters needed to specify the model; $\vec{\nabla}L$ is the gradient of the likelihood function L, obtained by taking a derivative of L with respect to every parameter in $\vec{\alpha}$; and E is the matrix of step sizes, obtained by inverting the Hessian matrix of second derivatives of the likelihood function. If the model requires p parameters, then both $\vec{\alpha}$ and $\vec{\nabla}L$ are p-dimensional vectors, and E is a p x p matrix.

GLM: internal covariates





Data: two human clinical trial participants with tetraplegia. M1 recordings in humans while performing a center-out task under neural guidance.

Truccolo, Hochberg, Donoghue, *Nature Neuroscience* (2009)

GLM: internal covariates

What is the probability that neuron i spikes at bin k, conditioned on the spiking history H_k of the ensemble during the preceding 100 ms?

$$\ln \lambda_i(k \mid H_k) = \alpha_{i0} + \sum_{m=1}^{\tau_N} \alpha_{ii}(m) y_i(k-m) + \sum_{j=1, j\neq i}^{N} \sum_{m=1}^{\tau_N} \alpha_{ij}(m) y_j(k-m)$$

Here, Δ=1 ms and τ_N=100. Data is used to fit the conditional intensity $\lambda_i(t)$ and obtain the background level α_{i0} of spiking activity, the kernel $\alpha_{ii}(m)$ related to intrinsic history effects, and the kernels $\alpha_{ij}(m)$ related to ensemble history effects.

Once the instantaneous spiking model is fitted, the estimated probability of a spike at any time bin can be computed.

Uphill iteration

Given the data $\{y_i(m)\}$ and the current value $\{\alpha^{(\mu)}\}$ of the parameters, construct:

1]
$$< y_i(k) >^{(\mu)} = \exp\left[\alpha_{i0}^{(\mu)} + \sum_{j'=1}^{N} \sum_{m'=1}^{\tau_N} \alpha_{ij'}^{(\mu)}(m')\, y_{j'}(k-m')\right]$$

Once the estimates $< y_i(k) >^{(\mu)}$ have been computed, the parameters are no longer needed. We can now build the components of the gradient vector:

2]
$$\left.\frac{\partial L_T\left(\{y_i(k)\} \mid \{\alpha\}\right)}{\partial \alpha_{ij}(m)}\right|_{(\mu)} = \sum_{k=1}^{K} \left\{\left[y_i(k) - < y_i(k) >^{(\mu)}\right] y_j(k-m)\right\}$$

Uphill iteration

Build the components of the Hessian matrix:

3] $$\left.\frac{\partial^2 L_T\left(\{y_i(k)\} \mid \{\alpha\}\right)}{\partial\alpha_{ij}(m)\,\partial\alpha_{ij'}(m')}\right|_{(\mu)} = -\sum_{k=1}^{K}\left\{< y_i(k) >^{(\mu)}\; y_j(k-m)\; y_{j'}(k-m')\right\}$$

Invert the Hessian matrix of second derivatives to obtain the matrix Epsilon of step sizes:

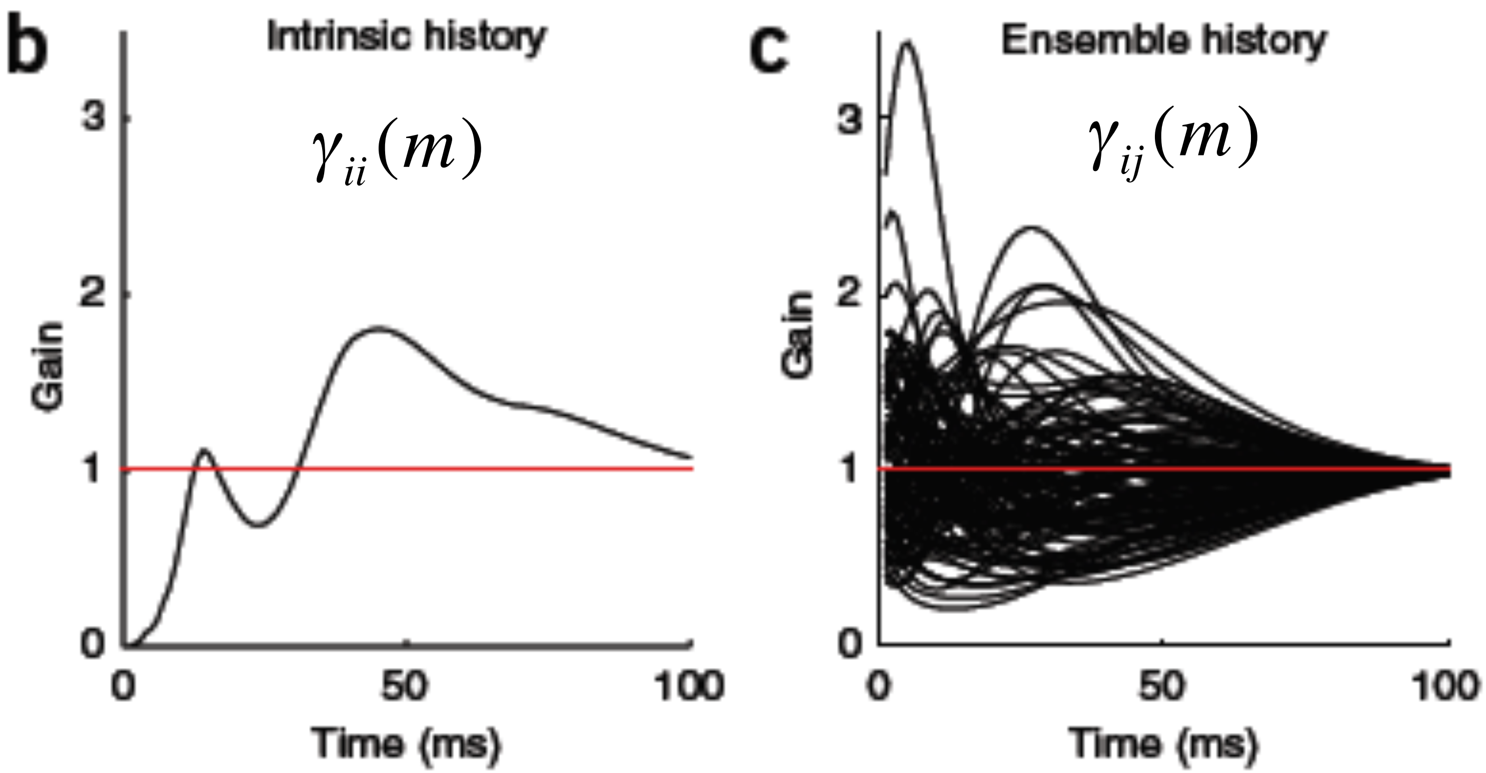
4] $$\mathrm{E} = -\frac{1}{H}$$

Multiply the matrix E and the gradient $\vec{\nabla} L$ to obtain the update:

5] $$\vec{\alpha}^{(\mu+1)} = \vec{\alpha}^{(\mu)} + \mathrm{E}^{(\mu)}\,\vec{\nabla} L^{(\mu)}$$

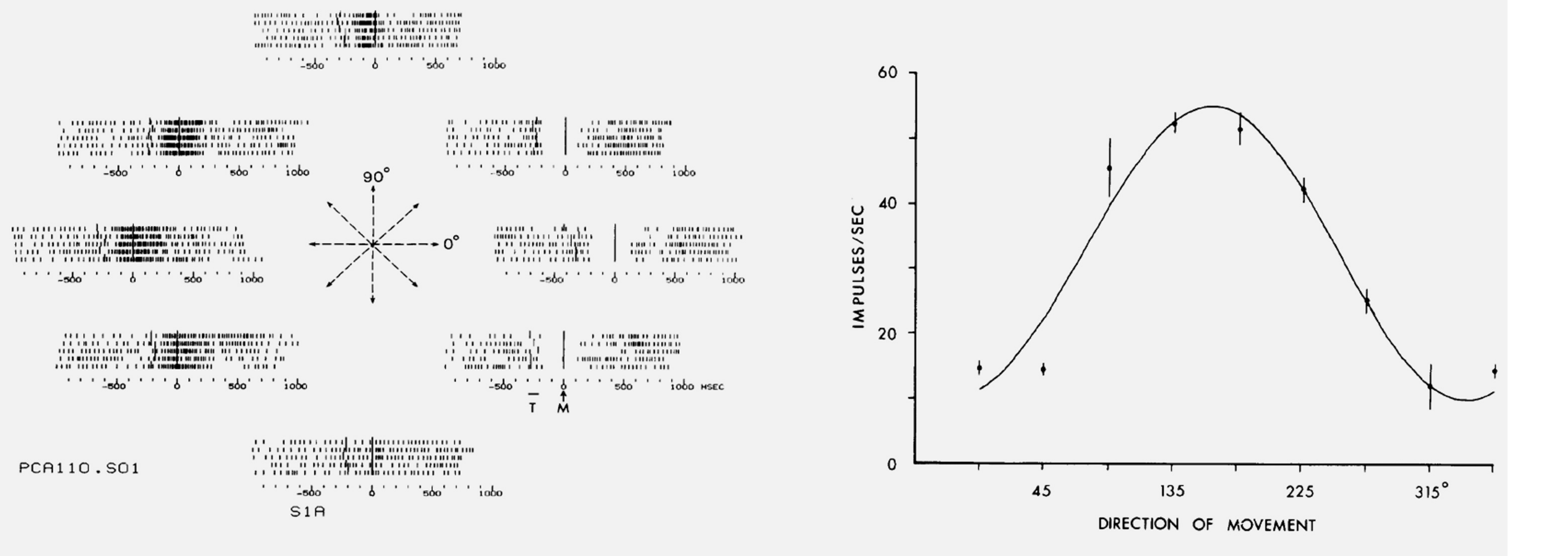
Temporal filters for spiking activity

Define $\gamma_{ij}(m) = \exp\big(\alpha_{ij}(m)\big)$



$i = 34$ $\quad\quad$ $i = 34,\ j \neq i$

External covariates: direction of motion



Georgopoulos et al. (1982)
Journal of Neuroscience, 2: 1527-1537
Figure 4

A reach characterized by (r, θ) corresponds to an activity:

$$f_i = b_i + r a_i (1/2)[1 + \cos(\theta - \theta_i)]$$

where b_i is the background activity, a_i is amplitude of activity modulation, and θ_i is the preferred direction of neuron i.

GLM for direction of motion

Independent cosine-tuned neurons:

$$\ln \lambda_i(t) = \alpha_{i0} + (1/2)\alpha_{iR}\, r(t+\tau_R)\left[1+\cos\left(\theta(t+\tau_R)-\theta_i\right)\right]$$

History-dependent independent cosine-tuned neurons:

$$\ln \lambda_i(t) = \alpha_{i0} + (1/2)\alpha_{iR}\, r(t+\tau_R)\left[1+\cos\left(\theta(t+\tau_R)-\theta_i\right)\right] + \\ + \sum_{m=1}^{\tau_N} \alpha_{ii}(m)\, y_i(t-m)$$

History-dependent interacting cosine-tuned neurons:

$$\ln \lambda_i(t) = \alpha_{i0} + (1/2)\alpha_{iR}\, r(t+\tau_R)\left[1+\cos\left(\theta(t+\tau_R)-\theta_i\right)\right] + \\ + \sum_{j=1}^{N} \sum_{m=1}^{\tau_N} \alpha_{ij}(m)\, y_j(t-m)$$

GLM for direction of motion

MEA (MultiElectrode Array) recordings in arm area of primary motor cortex (M1) of awake and behaving monkeys.

Task: two-dimensional tracking of a smoothly and randomly moving visual target. Target tracked by moving a two-link low friction manipulandum. Hand movement constrained to the horizontal plane. Hand position (X,Y) digitized and resampled at 1KHz. Low-pass filtered finite differences of position used to obtain velocity.

Model 1: velocity model

Model 2: velocity model plus autoregressive spiking history

Paninski, Fellows, Hatsopoulos, Donoghue, *J Neurophysiol* (2004)

Truccolo, Eden, Fellows, Donoghue, Brown, *J Neurophysiol* (2005)

GLM: velocity model

$$\lambda_i\left(t\middle|v(t+\tau_R),\,\theta(t+\tau_R),\,\{\alpha_{i0},\,\alpha_{iX},\,\alpha_{iY}\}\right)=$$

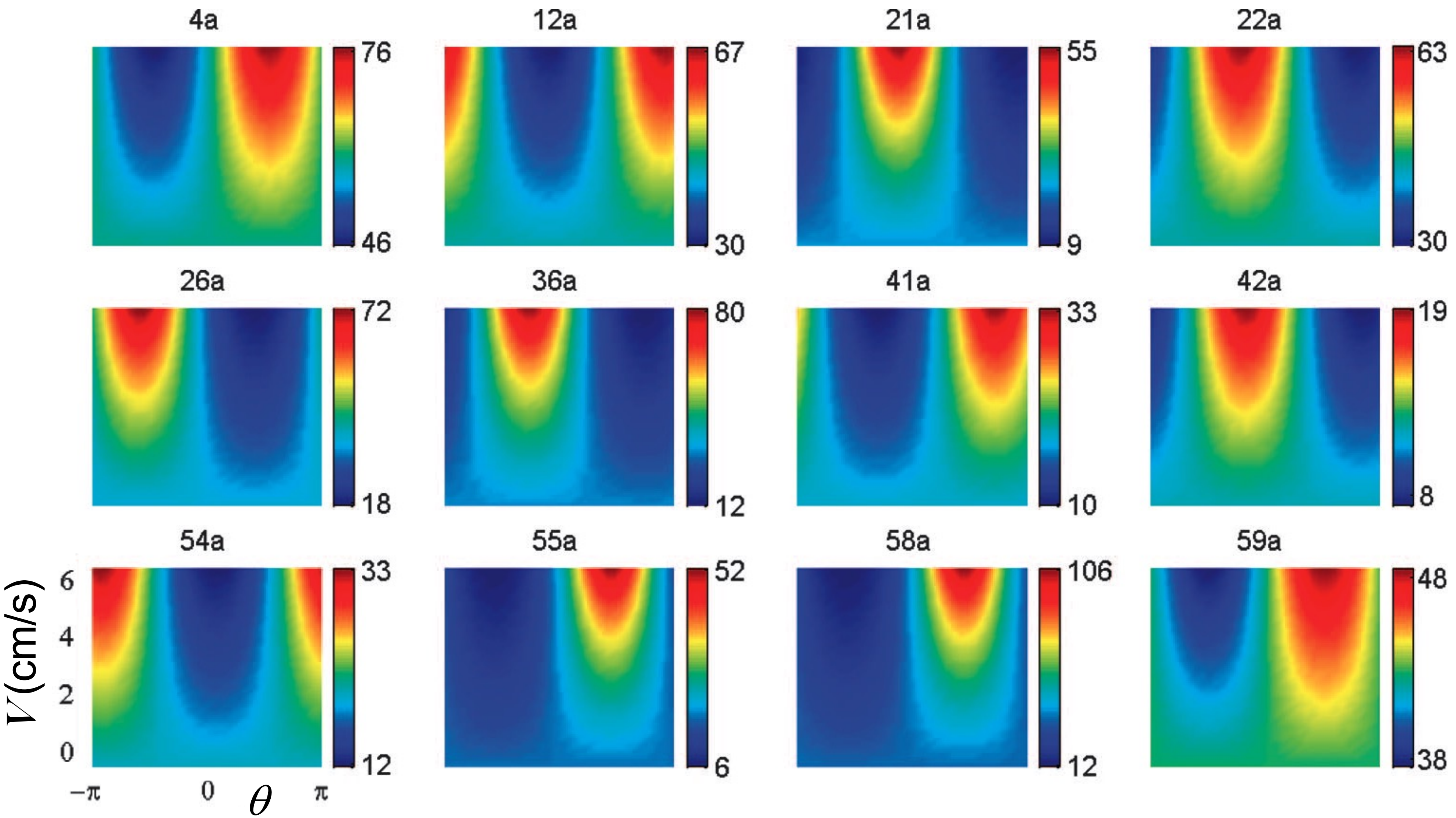
$$=\exp\left\{\alpha_{i0}+\alpha_{iX}\,v_X(t+\tau_R)+\alpha_{iY}\,v_Y(t+\tau_R)\right\}=$$

$$=\exp\left\{\alpha_{i0}+v(t+\tau_R)\left[\alpha_{iX}\cos\left(\theta(t+\tau_R)\right)+\alpha_{iY}\sin\left(\theta(t+\tau_R)\right)\right]\right\}$$

Note: no sum over time lags! A single time shift, with τ_R= 150 ms. There are only three parameters to be determined through a maximum likelihood fit to the spiking data of each neuron: $\{\alpha_{i0},\alpha_{iX},\alpha_{iY}\}$.

Once the values for these parameters have been specified, the conditional intensity $\lambda_i(t)$ can be plotted as a function of the subsequent velocity in polar coordinates: $v(t+\tau_R),\theta(t+\tau_R)$.

GLM: encoding



Velocity tuning functions for 12 different cells. Values of the expected number of spikes λ are color coded.

GLM: velocity plus autoregressive model

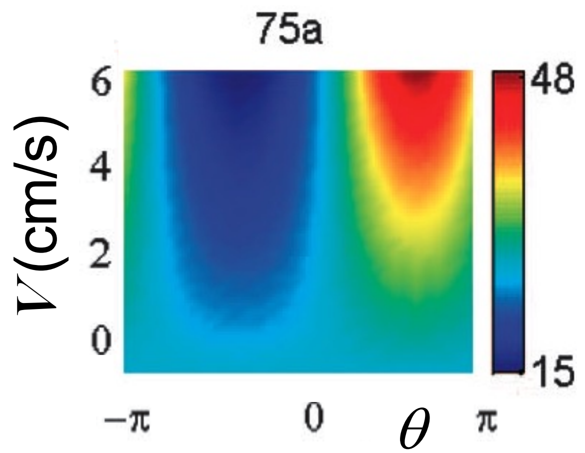
$$\lambda_i\left(t \middle| H_i(t), v(t+\tau_R), \theta(t+\tau_R), \{\alpha_{i0}, \{\alpha_{ii}(m)\}, \alpha_{iX}, \alpha_{iY}\}\right) = \exp\Big\{\alpha_{i0} +$$

$$+v(t+\tau_R)\left[\alpha_{iX}\cos\left(\theta(t+\tau_R)\right) + \alpha_{iY}\sin\left(\theta(t+\tau_R)\right)\right] + \sum_{m=1}^{\tau_N}\alpha_{ii}(m)\,y_i(t-m)\Big\}$$

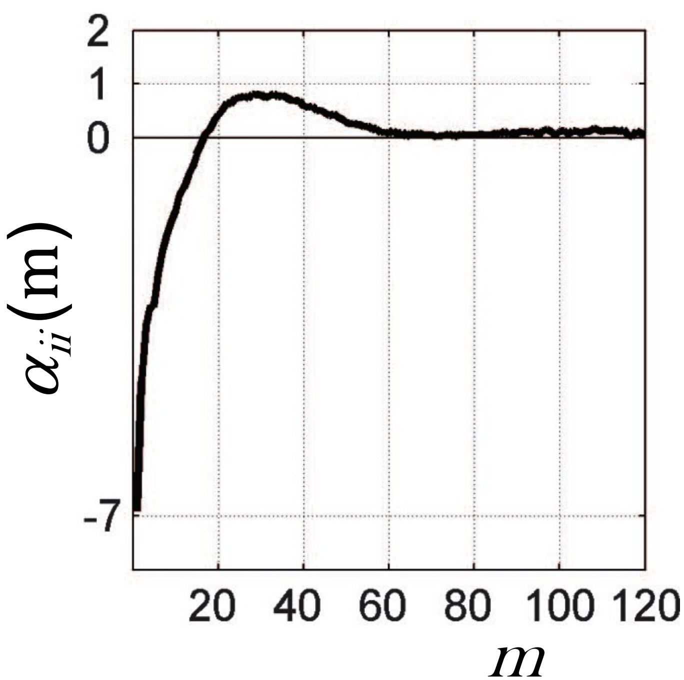
In addition to the three parameters $\{\alpha_{i0}, \alpha_{iX}, \alpha_{iY}\}$, the model for $\lambda_i(t)$ has parameters $\{\alpha_{ii}(m)\}$ for the autoregressive filter.

The data is binned at Δ=1 ms. At this time resolution, the number of spikes $y_i(t-m)$ can only be 0 or 1. The maximum temporal length of the filter is τ_N=120.

GLM: velocity plus autoregressive model

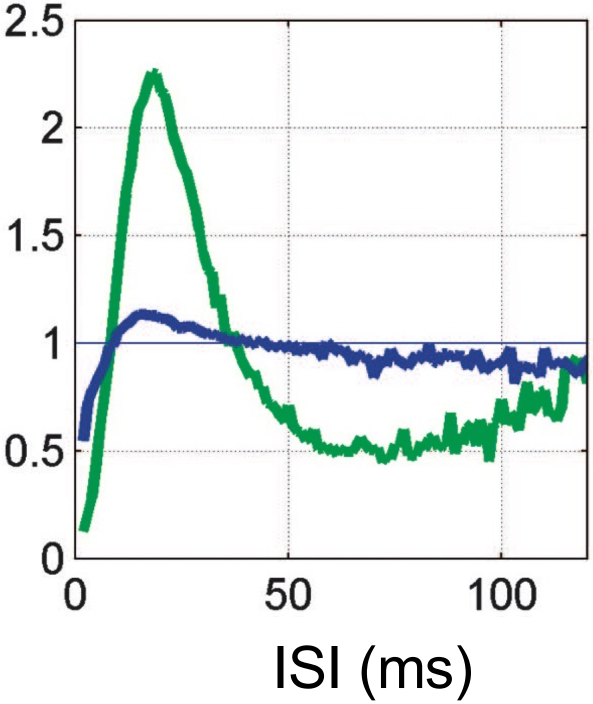


Velocity tuning function for cell 75a.



Autoregressive coefficients for cell 75a. Significant history effects extended only 60 ms into the past. Recovery period (negative coefficients) lasts about 18 ms after the cell spikes. The firing probability then increases and peaks at about 25 ms after a spike.

GLM: velocity plus autoregressive model



Ratio of observed to expected values of the rescaled times:

$$z_j = 1 - \exp\left\{ -\int_{t_{j-1}}^{t_j} \lambda\big(u \,|\, H(u)\big)\, du \right\}$$

Green: velocity model. Blue: autoregressive spiking history plus velocity model.

The velocity model overestimates λ for up to 10 ms after a spike, and it underestimates λ for between 10 ms and 40 ms after a spike.

The negative autoregressive coefficients significantly decrease the early overestimation, while the positive autoregressive coefficients almost eliminate the subsequent underestimation of the conditional density λ .

Summary

- Generalized linear models provide a principled and systematic approach to modeling the time-dependent rate of inhomogenous Poisson processes that describe the expected firing activity of a neural ensemble.

- The logarithm of the time-dependent rate for each neuron is modeled as a linear combination of intrinsic (the preceding firing activity of all neurons in the ensemble) and extrinsic (the preceding input stimulus or subsequent output activity) observables.

- The likelihood of the data is log-convex; optimal model parameters follow from an unambiguous gradient ascent algorithm.