Statistical models of visual neurons

Anna Sotnikova
Applied Mathematics and Statistics, and Scientific Computation program

Advisor: Dr. Daniel A. Butts
Department of Biology

Project Background

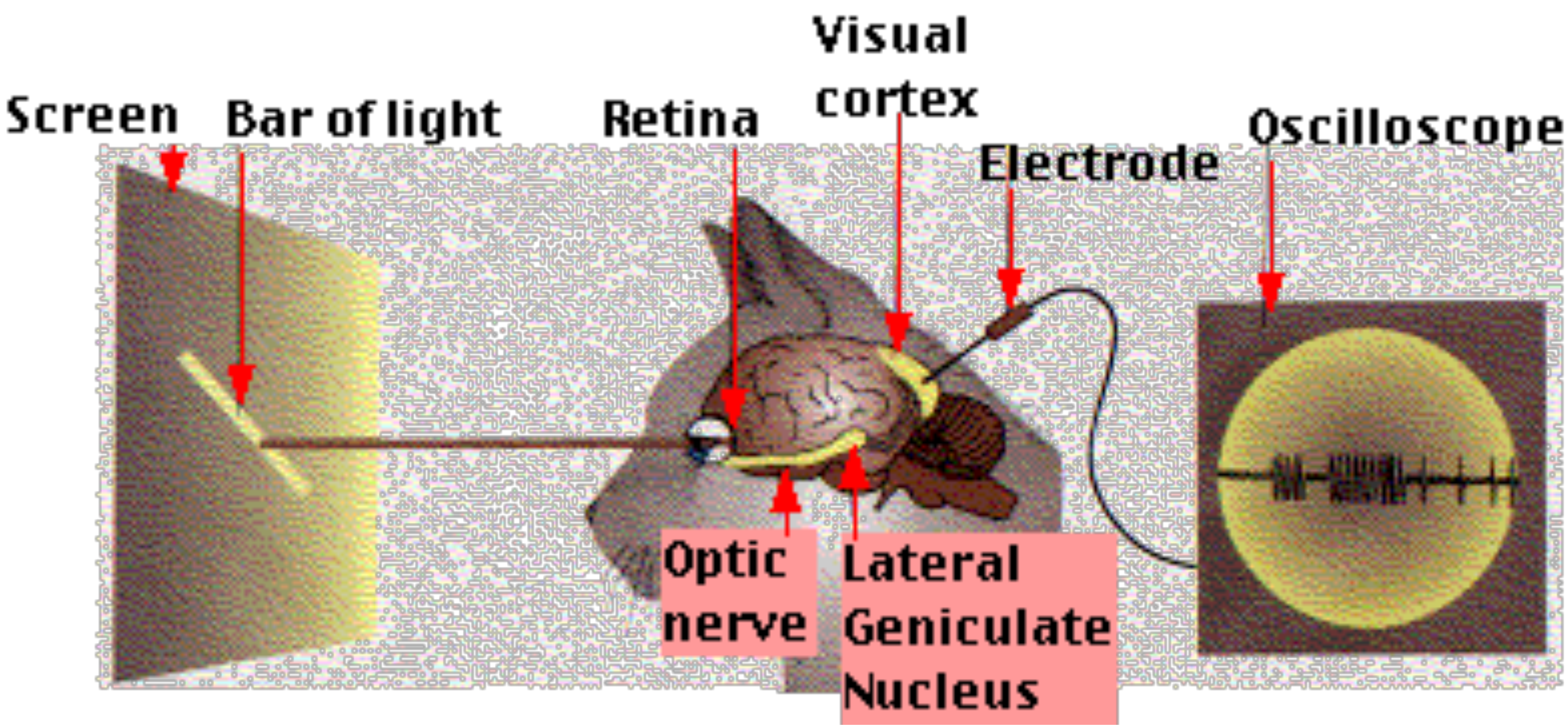
http://viralcircus.com/someone-bought-a-house-which-didnt-turn-out-the-way-they-hoped-its-indescribable-youll-be-speechless/

Project Background

General question : which algorithm did your brain use to analyze the picture?

Project Background

David Hubel and Torsen Wisel “Cat experiment”, 1958

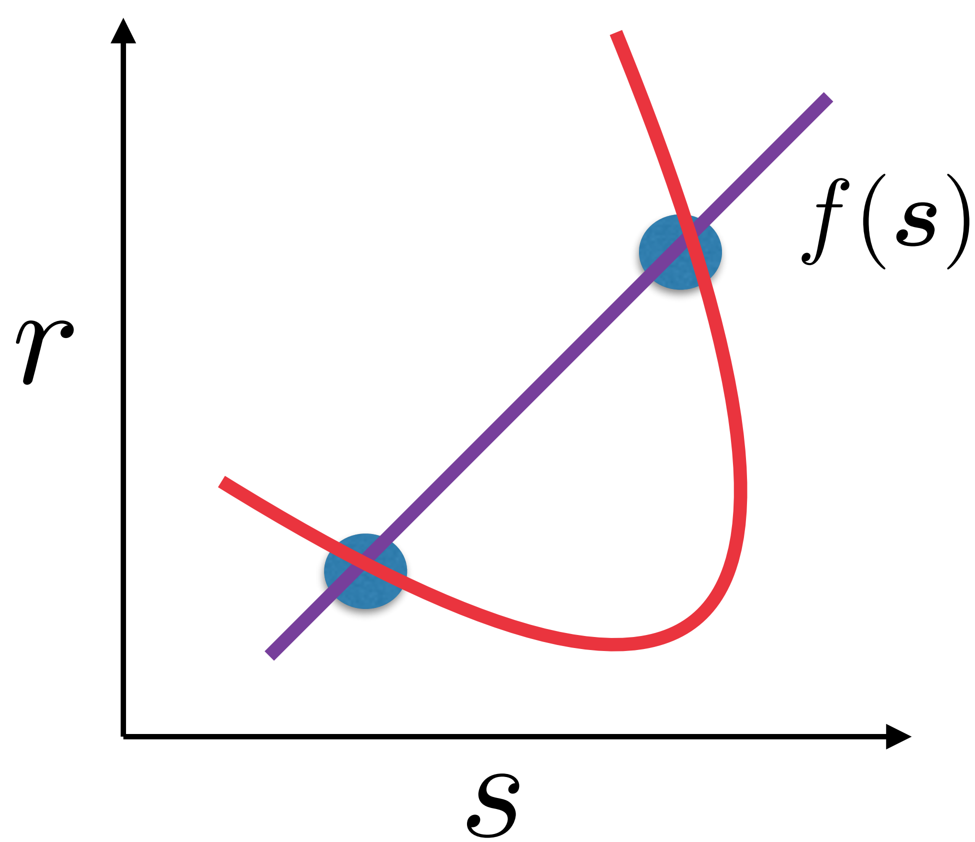


http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/V/VisualProcessing.html

Project Background

Stimulus -> $\boldsymbol{s}(t)$

-> Response $\boldsymbol{r}(t)$



Question : what is the mathematical form of $f(\boldsymbol{s})$?

1. Linear
2. Quadratic
3. Cascade

Project Goals

- Implement 5 models:

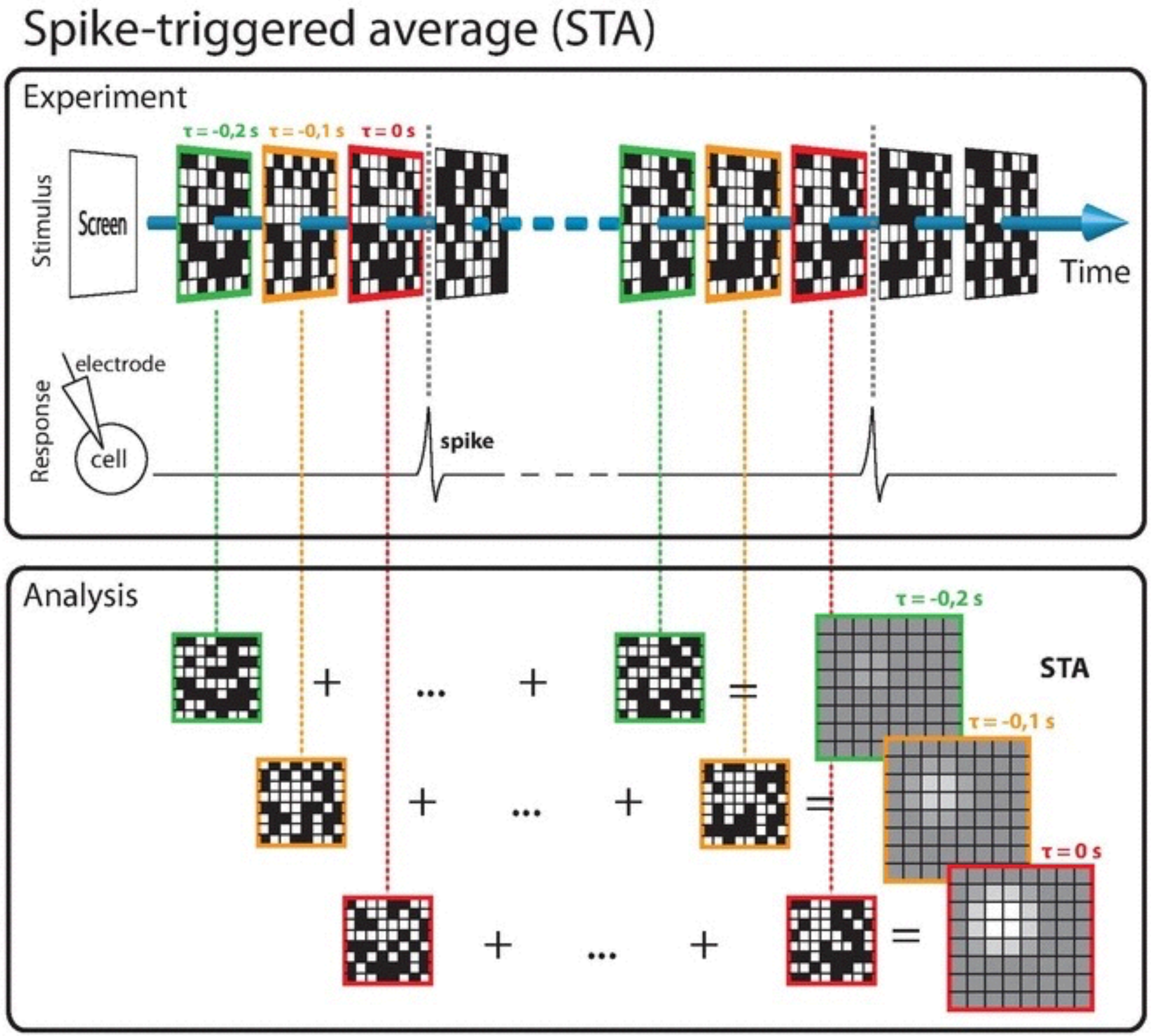
  1. Linear models
     - Spike Triggered Average (STA)
     - Generalized Linear Model (GLM)
  2. Quadratic models
     - Spike Triggered Covariance (STC)
     - Generalized Quadratic Model (GQM)
  3. Cascade model
     - Nonlinear Input Model (NIM)

Do not require parameters fitting

Require parameters fitting

- Test models on 2 data sets:
  1. Retina ganglion cells (RGC) and primary visual cortex(V1) synthetic data
  2. Lateral geniculate body experimental data

Spike Triggered Average



https://en.wikipedia.org/wiki/Spike-triggered_average

STA & STC formulas

STA - Spike Triggered Average

$$\hat{s} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{s}_i$$

N - total number of spikes per experiment time

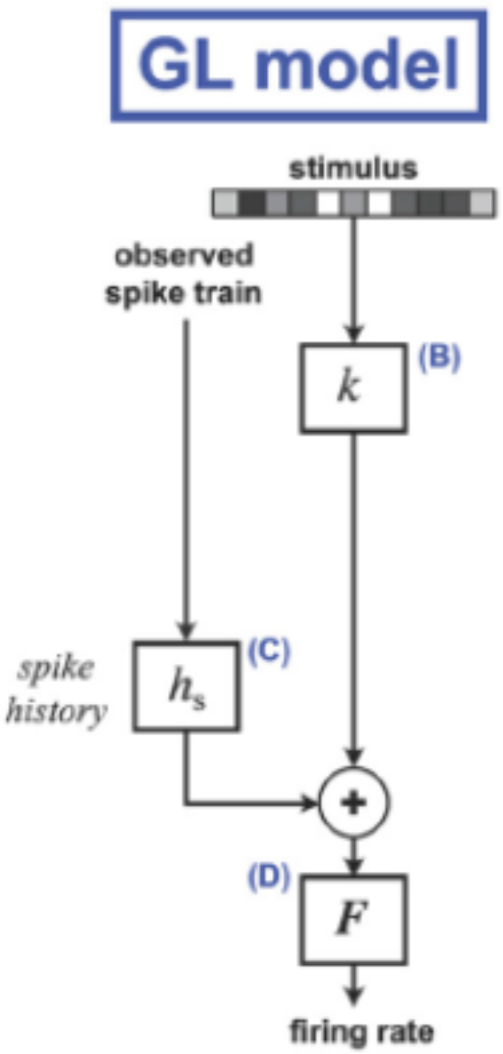
$\boldsymbol{s}_i$ - stimulus block preceding the i*th* spike

STC - Spike Triggered Covariance

$$C = \frac{1}{N-1}\sum_{i=1}^{N} (\boldsymbol{s}_i - \hat{s}) \cdot (\boldsymbol{s}_i - \hat{s})^T$$

Reference [5]

Generalized Linear Model



$$r(t) = F\Big(\boldsymbol{s}(t)\cdot\boldsymbol{k} + h_s\cdot\boldsymbol{R}(t) + \boldsymbol{b}\Big)$$

Notations

$r(t)$ - firing rate

F - spiking non-linearity $F(g) = log(1+e^g)$

$\boldsymbol{b}$ - shift spiking non-linearity

$\boldsymbol{k}$ - filter

h_s - spike history term

$\boldsymbol{R}(t)$ - history of observed spike train

Reference [2]

Generalized Quadratic Model

$$r(t) = F\left(\boldsymbol{k}_L \boldsymbol{s} + \boldsymbol{s}^{\mathrm{T}} \boldsymbol{C} \boldsymbol{s}\right) = F\left(\boldsymbol{k}_L \boldsymbol{s} + \sum_{i=1}^{M} w_i (\boldsymbol{k}_i \cdot \boldsymbol{s})^2\right)$$

Notations

$r(t)$ - firing rate

F - spiking non-linearity

$\boldsymbol{k}_L$ - linear filter

$\boldsymbol{C}$ - quadratic component

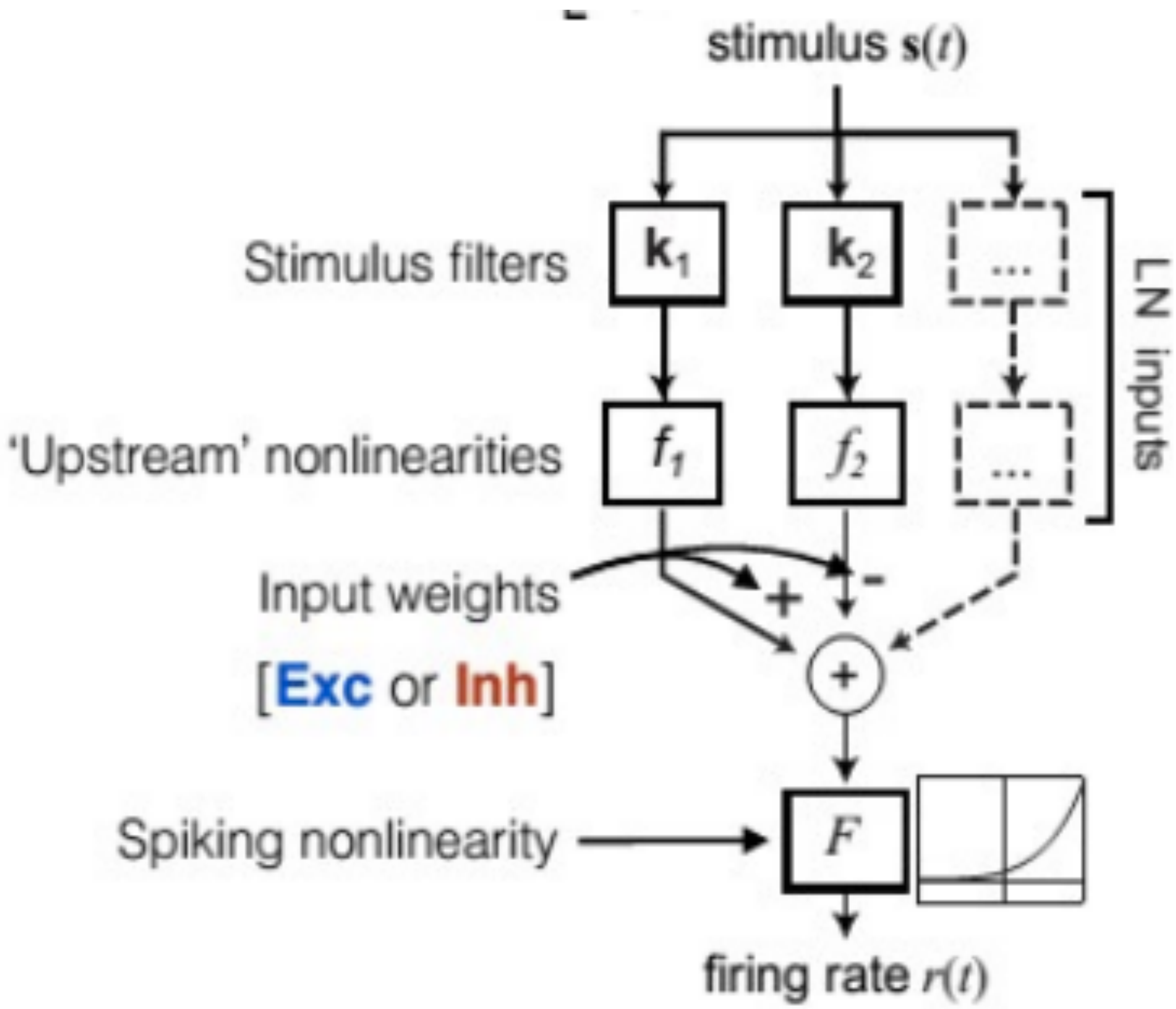
$\boldsymbol{k}_i$ - squared filters

w_i - +1 or -1

Reference [1]

Nonlinear Input Model

$$r(t) = F\left(\boldsymbol{X}(t) \cdot \boldsymbol{h} + \sum_{i=1} w_i f_i(\boldsymbol{k}_i \cdot \boldsymbol{s}(t)) \right)$$



Notations

$r(t)$ - firing rate

F - spiking non-linearity

$\boldsymbol{X}(t) \cdot \boldsymbol{h}$ - additional observables

f_i - 'upstream nonlinearities'

$\boldsymbol{k}_i$ - squared filters

w_i - +1 or -1

Reference [1]

Parameters estimation

1. Estimate filters and non-linear function f_i
   - fix f_i and estimate filters

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ x & \text{otherwise} \end{cases}$$

$$LL = \sum_t (R_{obs}(t) log(r(t)) - r(t))$$

Example for NIM

$$\frac{\partial LL}{\partial k_{i,m}} = \sum_t \left(\frac{R_{obs}(t)}{r(t)} - 1 \right) F'[G(t)] w_i f_i'(g_i(t)) s_m(t),$$

$$G(t) = \sum_{i=1} g_i(t) = \sum_{i=1} w_i f_i(\boldsymbol{k}_i \cdot \boldsymbol{s}(t))$$

Parameters estimation

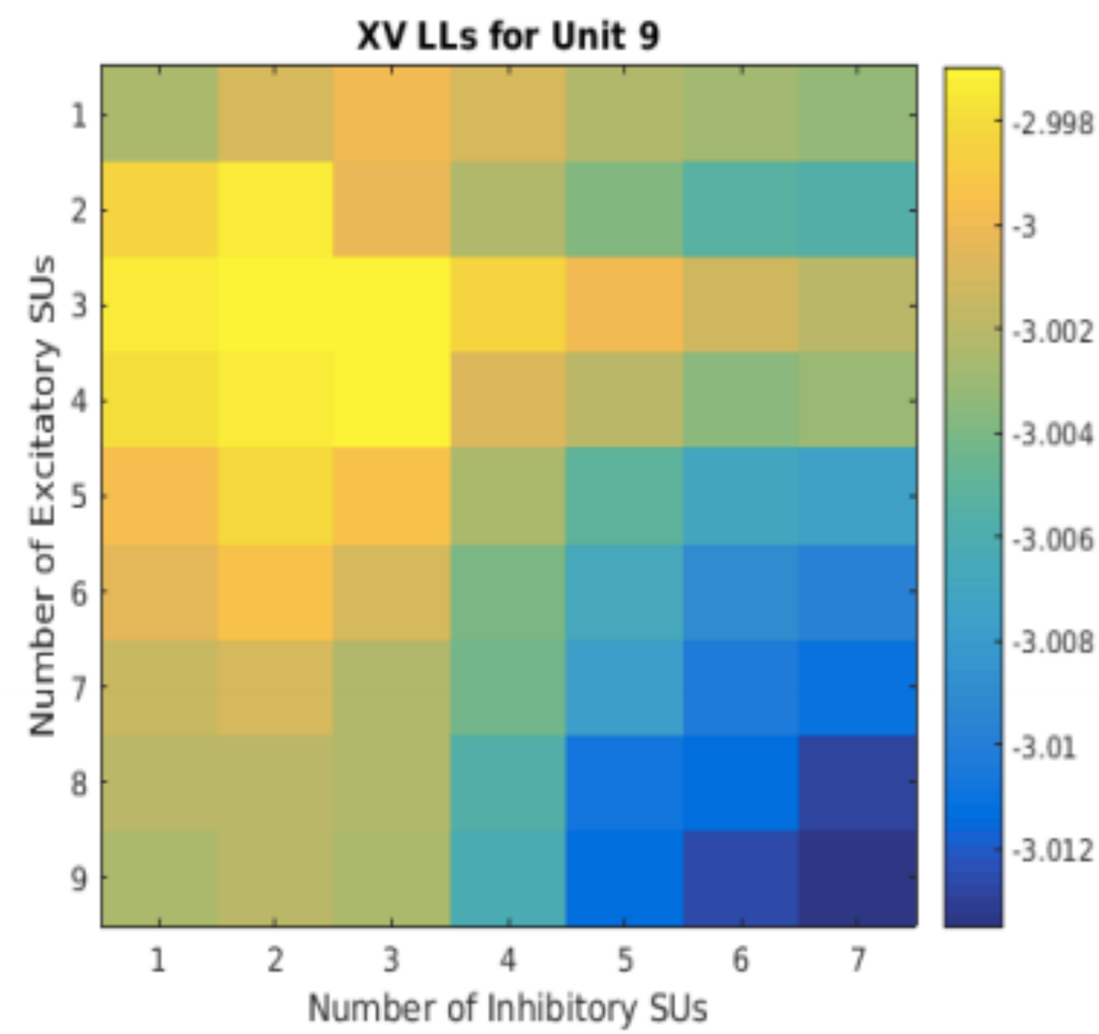
- fix $\boldsymbol{k}_i$ and estimate f_i nonparametrically

$$f_i(g) = \sum_j a_{ij} \phi_j(g)$$

$$\varphi_k(x) = \begin{cases} \dfrac{x - x_{k-1}}{x_k - x_{k-1}} & \text{if } x \in [x_{k-1}, x_k] \\ \dfrac{x_{k+1} - x}{x_{k+1} - x_k} & \text{if } x \in [x_k, x_{k+1}] \\ 0 & \text{otherwise} \end{cases}$$

Parameters estimation

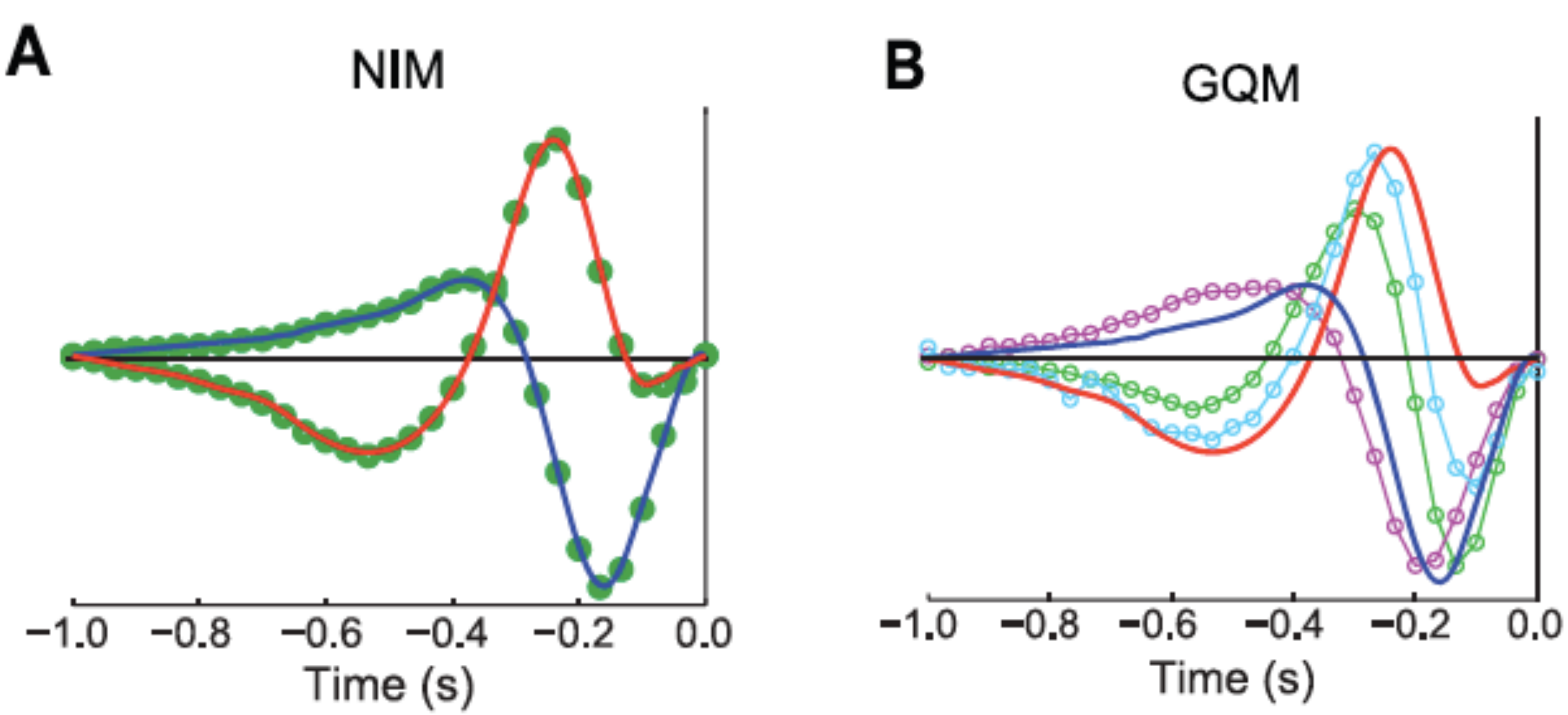
2. Find how many filters do we need to work with the data and w_i sign.

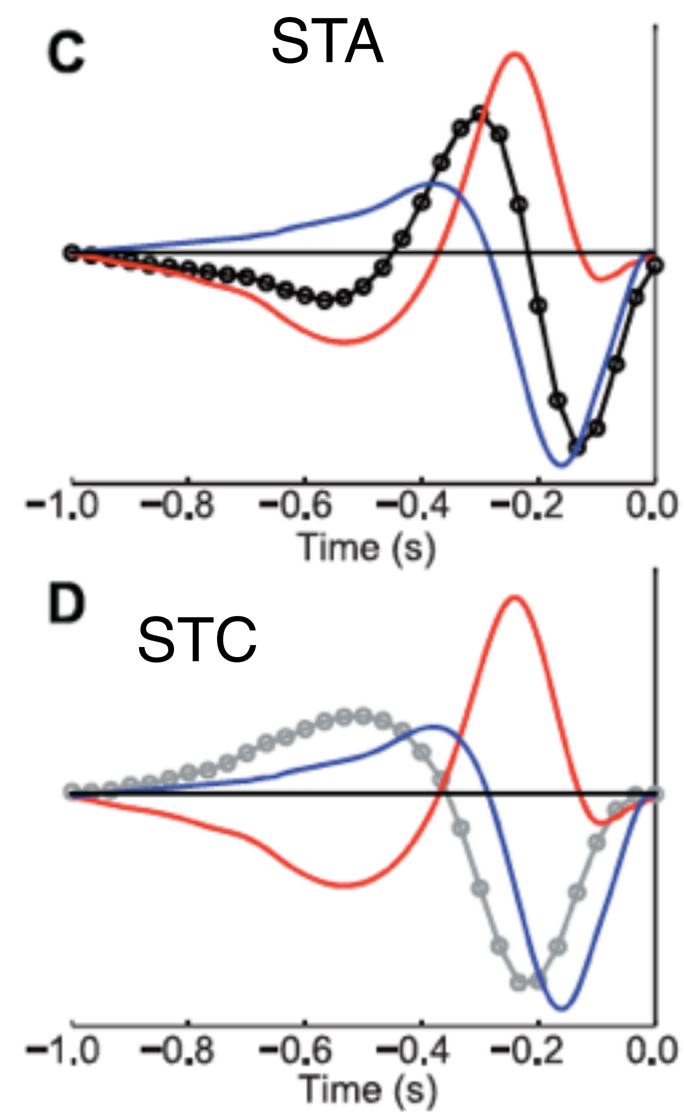


Picture from Felix Bartsch' s presentation

Validation and Testing

- Validate program routines on synthetic data
    - Program should produce known output for retinal ganglion cells

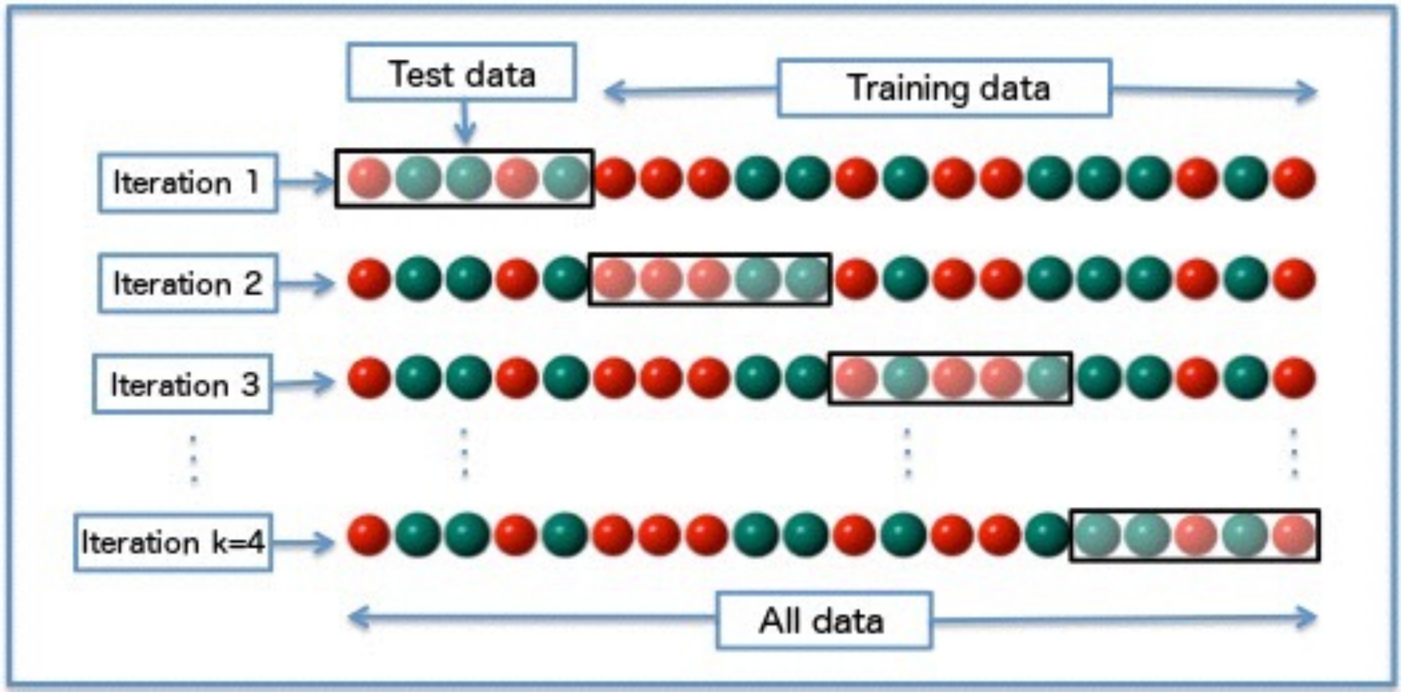




Reference [1]

Validation and Testing

- Use cross-validation for LGN data set
  - Test part of the data for predicting the output and comparing it with the real one



https://en.wikipedia.org/wiki/Cross-validation_(statistics)

Data set description

Data set contains following elements:

- stimulus
- time interval of the stimulus update
- spike times (in units of seconds)

Implementation

Hardware

- MacBook Air, 1.4 GHz Intel Core i5, 4 GB 1600 MHz DDR3

Software

- Matlab_R2015b

Project schedule

October - mid November

- Implement STA and STC models
- Test models on synthetic data set and validate models on LGN data set

November - December

- Implement GLM
- Test model on synthetic data set and validate model on LGN data set

January - March

- Implement GQM and NIM
- Test models on synthetic data set and validate models on LGN data set

April - May

- Collect results and prepare final report

Deliverables

- Matlab code for all 5 models
- List of models' parameters
- Reports and presentations

References

1. McFarland JM, Cui Y, Butts DA (2013) Inferring nonlinear neuronal computation based on physiologically plausible inputs. PLoS Computational Biology 9(7): e1003142.

2. Butts DA, Weng C, Jin JZ, Alonso JM, Paninski L (2011) Temporal precision in the visual pathway through the interplay of excitation and stimulus-driven suppression. J. Neurosci. 31: 11313-27.

3. Paninski L (2004) Maximum likelihood estimation of cascade point-process neural encoding models. Network 15: 243-262.

4. Simoncelli EP, Pillow J, Paninski L, Schwartz O (2004) Characterization of neural responses with stochastic stimuli. In: The cognitive neurosciences (Gazzaniga M, ed), pp 327–338. Cambridge, MA: MIT.

5. Chichilnisky EJ (2001) A simple white noise analysis of neuronal light responses. Network 12:199 –213.