Neural coding
Ecological approach to sensory coding: efficient adaptation to the natural environment
Part 2: Single cell - optimization of the transfer function

Jean-Pierre Nadal
CNRS & EHESS

Laboratoire de Physique Statistique
(LPS, UMR 8550 CNRS - ENS – UPMC – Univ. Paris Diderot)
Ecole Normale Supérieure (ENS)

&

Centre d’Analyse et de Mathématique Sociales
(CAMS, UMR 8557 CNRS - EHESS)
Ecole des Hautes Etudes en Sciences Sociales (EHESS)

nadal@lps.ens.fr
Mutual information

\[ I[\theta, X] = \text{[entropy of } X]\]  

\[ = \int \ln Q(X) Q(X) \, d^p X + \int \left[ \int \ln Q(X|\theta) Q(X|\theta) \, d^p X \right] \rho(\theta) d^N \theta \]

\[ = \text{Information that } X \text{ carries about } \theta \]

\[ = \text{Information that } \theta \text{ carries about } X \]

\[ = H(X) + H(\theta) - H(\theta, X) \]

\[ = \int \int \ln \frac{P(\theta, X)}{\rho(\theta) Q(X)} P(\theta, X) \, d^p X \, d^N \theta \]

Output distribution (marginal distribution of \(X\)):

\[ Q(X) = \int Q(X|\theta) \rho(\theta) d^N \theta \]

Joint distribution of \(X\) and \(\theta\):

\[ P(\theta, X) = Q(X|\theta) \rho(\theta) \]
Mutual information

environment → stimulus → neural representation

\[ \rho(\cdot), \theta \quad Q(\mathbf{x} | \theta) \quad \mathbf{x} = \{ x_1, x_2, \ldots, x_p \} \]

\[ I[\theta, \mathbf{x}] \geq 0 \quad (I = 0 \iff \theta \text{ and } \mathbf{x} \text{ are statistically independent}) \]

principle for optimal coding:

\text{Infomax} \quad \rho \text{ given} \quad \text{(environment)}

\[ \max I[\theta, \mathbf{x}] \quad Q \]
Fig. 1. The modular structure of the retina and lamina of muscoid flies. The retina of the compound eye is an array of optical units, the ommatidia. Each ommatidium focuses light from a small solid angle of space onto a group of photoreceptors. Six of these photoreceptors terminate in the first optic neuropile, the lamina. The lamina is subdivided into cartridges: one for every retinal sampling unit in the retina. The diagrammatic cross section through a lamina cartridge shows the six photoreceptor axon terminals surrounding the two central large monopolar cells (LMCs). Each photoreceptor makes over 200 chemical synapses with each LMC. The two LMCs project retinotopically to equivalent cartridges in the next optic neuropile. A lamina cartridge contains a number of other interneurons (not shown), and is surrounded by a glial sheath. (Modified from Refs 3,5,7.)
Response of contrast-sensitive cells in the fly by Laughlin, 1981
Optimization of the transfer function

\[ V = f(S) + \text{noise} \]

\[ \rho(\cdot) \xrightarrow{\text{s}} f(\cdot) \xrightarrow{\rho \text{ given}} V \]

\[ \max_{f(\cdot)} I[V; S] \]

additive noise

\[ I[V; S] = \text{output entropy} - \text{equivocation} \]

equivocation = \text{(average over the input distribution of the)}

entropy of the output \( V \) when the input \( S \) is known

\[ = \text{noise entropy} \quad (= \text{constant}) \]
Optimization of the transfer function

\[ \rho(\cdot) \xrightarrow{\text{f}(\cdot)} S \xrightarrow{\text{f}(\cdot)} V = f(S) + \text{noise} \]

\[ \rho \text{ given} \]

\[ \max_{f()} I[V; S] ? \]

\[ I[V; S] = \text{output entropy} + \text{cst.} \]

vanishing additive noise limit \( \Rightarrow V \approx f(S) \in [0,1] \)

then: max entropy \( \leftrightarrow \) uniform distribution: \( q(V) = 1 \) on \([0,1]\)

but from \( V = f(S) \) one has:

\[ dV = f'(S) dS \]

\[ q(V) dV = \rho(S) dS \Rightarrow q(V) = \frac{\rho(S)}{f'(S)} \]

Hence:

\[ f'(S) = \rho(S) \]
Optimization of the transfer function (bis)

\[ \rho(\cdot) \xrightarrow{\text{S}} f(\cdot) \xrightarrow{\text{V}} V = f(S) + \text{noise} \]

\( \rho \text{ given} \)

\[ \max_{f(\cdot)} I[V; S] \text{?} \]

additive noise \( \Rightarrow I[V; S] = \text{output entropy} + \text{cst.} \)

output entropy:

\[ H = -\int \ln q(V) q(V) \, dV \]

vanishing noise:

\[ \begin{cases} dV = f'(S) dS \\ q(V) dV = \rho(S) dS \end{cases} \]

\[ I = \text{cst.} - \int \ln \frac{\rho(S)}{|f'(S)|} \rho(S) \, dS \]

\[ I = \text{cst.} - L(\rho; f') \]

\( \Rightarrow \max I = \min \text{Kulback divergence} \ (\rho; |f'|) \)

\( \Rightarrow |f'| = \rho \)
Optimization of the transfer function (bis)

\[ \rho(.) \xrightarrow{\text{s}} f(.) \xrightarrow{\text{V}} V = f(S) + \text{noise} \]

\( \rho \) given

\[ \max_{f(.)} I[V; S] ? \]

additive noise \( \Rightarrow I[V ; S ] = \text{output entropy} + \text{cst.} \)

output entropy:

\[ H = - \int \ln q(V) q(V) \, dV \]

vanishing noise:

\[ \begin{cases} dV = f'(S) dS \\ q(V) dV = \rho(S) dS \end{cases} \]

\[ I = \text{cst.} - \int \ln \frac{\rho(S)}{|f'(S)|} \rho(S) \, dS \]

\[ I = \text{cst.} - L(\rho; f') \]  

\[ \Rightarrow \max I = \min \text{Kulback divergence} (\rho; |f'|) \]

\[ |f'| = \rho \]
Optimization of the transfer function (3)

Given the noise level $\varepsilon$, estimation of $S$ from the observation of $V$?

$V = f(S)$

High slope $\rightarrow$ high resolution

Good performance in average $\rightarrow$ high slope where the stimuli are the most frequent

$\Rightarrow f'(S) \sim \rho(S)$
The matched amplification strategy requires a relationship between input intensity and response amplitude in which equal increments in response correspond to equal areas under the probability distribution of input levels (upper figures). Contrast coding in LMCs (lower figure) comes close to this expectation. 

Laughlin, 1981

\[ f'(S) \sim \rho(S) \]
Histogram equalization

$\rho(S) = \frac{1}{N} \sum_{k=0}^{K-1} \rho(k)$

$\rho(S) = \text{distribution of grey levels in original image}$

Grey level at pixel $k$ in new image = $f(\text{grey level at pixel } k \text{ in original image})$

Histogram of grey levels in new image = flat (that is uniform distribution)
Males locate their female mates by the pheromone they release.

Method (as compared to Laughlin, backward approach): from the neural activity statistics, determination of the characteristics of the pheromone plume which are best detected by the male reception system. Results in agreement with plume measurements in the field, so providing quantitative evidence that this system obeys the efficient coding principle.

Image of a smoke plume filmed in a wind tunnel 1 m across and 2 m long with source on the left side. 

High pheromone concentrations can be found relatively far from the source due to the imperfect mixing of odorant with air. The signal detected by both moving and stationary detectors is therefore always intermittent, consisting of pulses of relatively undiluted pheromone.
Multiplicative noise
Statistics of spiking neurons: comparison with a Poisson process

Left: Interspike interval histogram from an MT neuron responding when the monkey is looking at a moving, random-dot image.
Right: Interspike interval distribution from a Poisson model with a stochastic refractory period.

_Figure from Bair, Koch, Newsome & Britten, The Journal of Neuroscience, 1994_
Spikes statistics vs. Poisson statistics

Figure 1. Response variability of a neuron recorded from area MT of an alert monkey. A, Raster and peristimulus time histogram (PSTH) depicting response for 210 presentations of an identical random dot motion stimulus. The motion stimulus was shown for 2 sec. Raster points represent the occurrence of action potentials. The PSTH plots the spike rate, averaged in 2 msec bins, as a function of time from the onset of the visual stimulus. The response modulates between 15 and 220 impulses/sec. Vertical lines delineate a period in which spike rate was fairly constant. The gray region shows 50 trials from this epoch, which were used to construct B and C. B, Magnified view of the shaded region of the raster in A. The spike rate, computed in 5 msec bins, is fairly constant. Notice that the magnified raster reveals substantial variability in the timing of individual spikes. C, Frequency histogram depicting the spike intervals in B. The solid line is the best fitting exponential probability density function. D, Variance of the spike count is plotted against the mean number of spikes obtained from randomly chosen rectangular regions of the raster in A. Each point represents the mean and variance of the spikes counted from 50 to 200 adjacent trials in an epoch from 100 to 500 msec long. The shaded region of A would be one such example. The best fitting power law is shown by the solid curve. The dashed line is the expected relationship for a Poisson point process.
Mean activity given the stimulus: tuning curves

A: Recording from neuron in primary visual cortex (V1, area 17, striate cortex) in monkey when presented with moving bars of light falling over the neuron’s receptive field.

B: Gaussian tuning curve fitted to the responses.

Hubel and Wiesel, 1968; Henry et al 1974; Wandell 1995
Simple(st) model of spiking neuron:

\[ \rho(\cdot) \xrightarrow{S} \nu(S) \rightarrow k \text{ spikes in } [0, t] \]

Poisson process

\[ Q_t(k|S) = \frac{(\nu t)^k e^{-\nu t}}{k!} \]

R. Stein 1967


Large times \( vt \rightarrow \infty \)  
Poisson law \( \rightarrow \) Gaussian law

Observation of \( k \) spikes in the time window \( [0, t] \) \( \approx S \rightarrow V = \frac{k}{t} \)

\[ V = \nu(S) + \sqrt{\frac{\nu(S)}{t}} z \]

(Gaussian) 

\( z = \text{Gaussian noise} \)  
\[ \langle z \rangle = 0 \quad \langle z^2 \rangle = 1 \quad \langle (\text{noise})^2 \rangle = \frac{\nu}{t} \]

\( \nu = \text{mean firing rate} \)
large time limit

\[ S \rightarrow v(S) \sim \text{transfer function as in Laughlin case, } \]
\[ \text{except that here we have a stimulus dependent noise } \]

\[ \max_{\nu} I_t(S, V)? \]
\[ \rightarrow \text{same analysis as for the Laughlin' case, } \]
\[ \text{but with a term coming from the equivocation} \]
\[ I = \text{cst.} - \int \ln \frac{\rho(S) \sqrt{v(S)}}{|v'(S)| \sqrt{t}} \rho(S) dS \]

\[ \rightarrow \text{optimal tuning curve:} \]
\[ (\text{cst.}) \frac{d}{dS} \sqrt{v(S)} = \rho(S) \]
\[ \text{instead of:} \]
\[ \frac{d}{dS} v(S) = \rho(S) \]

Multiplicative noise:

\[ \text{optimal tuning curve/transfer function depends on noise, } \]
\[ \text{even for vanishing multiplicative noise} \]
Single cell - **Spiking neuron** (Poisson process – large times)

Reminder: large times,

\[ I = \text{cst.} - \int \ln \frac{\rho(S) \sqrt{v(S)}}{|v'(S)| \sqrt{t}} \, \rho(S) dS \]

→ behaviour of the mutual information at large times:

sublinear growth

\[ \mu t \to \infty \quad I_t \approx \frac{1}{2} \ln t \]
Next:

- Single cell, optimization of the tuning curve
  - Short times
  - From short to large times

- Redundancy reduction: from infomax to ICA

- Coding by a population of neurons