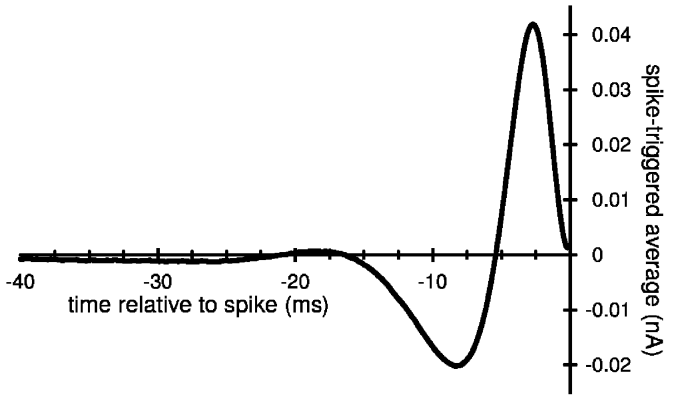
10 Optimal stimuli and receptive fields

We consider the simplest measure of determining the stimulus that is most likely to cause a neuron to fire. This can be expressed through the spike-triggered average.. We consider this first for a point stimulus and then for a spatially extended stimulus.

10.1 The optimal input to drive spiking

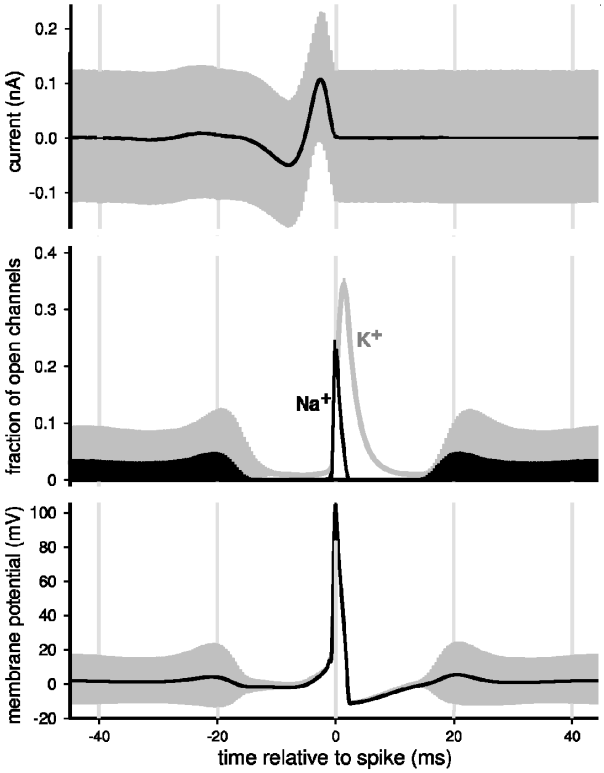
We previously considered that a step input leads to spiking and that noise can lead to spiking. But details of the ionic membrane currents should lead to a definitive input as the best driver of a neuron. While we have not explored such currents so far, the need to de-inactivate the inactivating component of the sodium channel suggests that a brief hyperpolarization before a depolarization is optimal. Indeed, using the spike-triggered agaeraged correlation technique, described next,with a computer model of a Hodgkin Huxley cell, Blaise Aguera y Arcas and Adrienne Fairhall numerically determined the optimal current to drive a neuron (Figure 1). The combination of inhibitory and excitatory components to the input current suggests that this indicates the necessity of coordinating inhibitory and excitatory inputs in brain circuits. Interestingly, the literature speaks of "feed-forward inhibition" and, as we shall see next, the tuning curves for inhibitory neurons in sensory and motor brain regions typically match those for excitatory neurons.

Figure 1: The optimal linear transfer function, i.e., current waveform, to elicit a spike. From Aguera y Arcas, Fairhall and Bialek (2003).



Another issue concerns the variability around the time of the spike. Consistent with an optimal input, Aguera y Arcas and Fairhall observed that the variability in membrane voltage is quenched at the time of a spike (Figure 2). This implies that the optimal input will reduce the jitter of spike timing.

Figure 2: Spike-triggered average of a Hodgkin-Huxley action potential with standard deviations for (top) the input current I, (middle) the fraction of open K^+ and Na^+ channels, and (bottom) the membrane voltage V, for the input current parameters mean $I_o = 0$ and spectral variance $S = 6.5 \times 10^{-4} nA^2 s$. From Aguera y Arcas, Fairhall and Bialek (2003).



10.1.1 Transfer function of the spike response

Typically one measures two time series in recording from a neuron in an animal

- $V^k_{app}(t)$ is the applied stimulus or motor output for the k-th trial.
- $S^k_{meas}(t) = \sum_{spike\ times} \delta(t - t^k_r)$ is the measured spike time for the k-th trial.

How well can we reconstruct the stimulus from the spikes" We use the measured information to construct a linear filter that allows us to predict the stimulus for an unknown spike train. In a sense. This idea is old, although it came into use only at the time of WW II, when there was a big push at the MIT Radar Laboratory to formulate the mathematics of optimal filtering and prediction. The procedure is as follows:

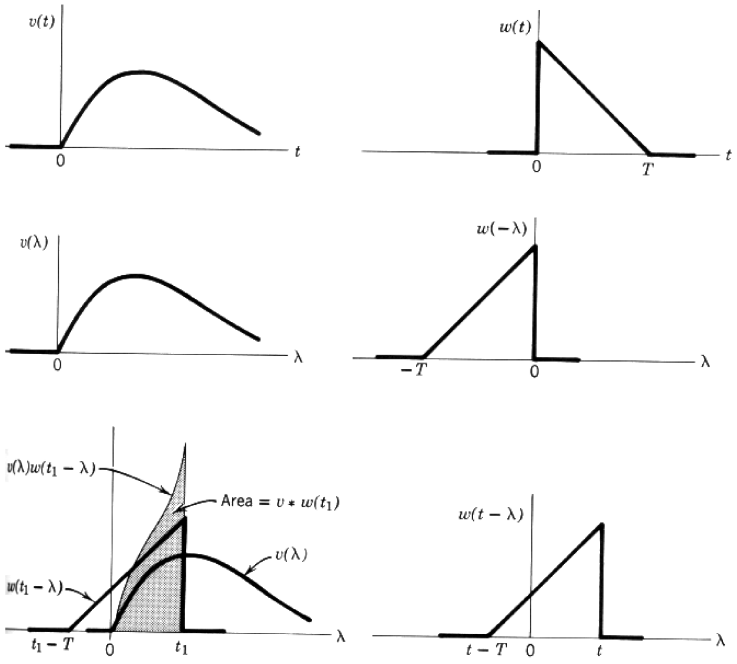
- We define $T(t)$ as the sought after transfer function.
- We define $V^k_{pred}(t)$ as the predicted stimulus for the k-th trial, based on the measured spike train, where

$$\begin{aligned} V^k_{pred}(t) &= \int_{-\infty}^{t} dt' T(t-t') S^k_{meas}(t') \\ &= \int_{-\infty}^{t} dt' T(t-t') \sum_s \delta(t' - t^k_S) \end{aligned} \tag{10.1}$$

$$= \sum_{\text{spike times, } S} \int_{-\infty}^{t} dt' T(t-t')\delta(t'-t_S^k)$$
$$= \sum_S T(t-t_S^k)$$

is the predicted output, given by a convolution integral (Figure 3).

Figure 3: Pictorial guide to the convolution of $v(t)$ with $u(t)$.



To determine $T(t)$, we minimize the difference between the actual and the predicted stimulus, averaged over all trials and time, i.e.,

$$Error = \sum_k \int dt \left(V_{pred}^k(t) - V_{app}^k(t)\right)^2 \tag{10.2}$$
$$= \sum_k \int dt \left(\int_{-\infty}^{t} dt' T(t-t') S_{meas}^k(t') - V_{app}^k(t)\right)^2$$

The error is computed in terms of measured quantities, except for $T(t)$, which we find by the criteria that we choose $T(t)$ to minimize the error. This is much easier to solve in the frequency domain, where convolutions turn into products. We consider the Fourier transformed variables:

$$V_{app}^k(t) \iff \tilde{V}_{app}^k(f) \tag{10.3}$$
$$S_{meas}^k(t) \iff \tilde{S}_{meas}^k(f) \tag{10.4}$$
$$V_{pred}^k(t) \iff \tilde{V}_{pred}^k(f) \tag{10.5}$$
$$T(t) \iff \tilde{T}(f) \tag{10.6}$$
$$\tag{10.7}$$

where

$$\tilde{T}(f) = \int_{-\infty}^{\infty} dt e^{i2\pi ft} T(t) \tag{10.8}$$

$$T(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} df e^{-i2\pi ft} \tilde{T}(f) \tag{10.9}$$

so that (ignoring causality for the moment) the convolution becomes

$$\int_{-\infty}^{\infty} dt' T(t-t') X(t') = \tilde{T}(f)\tilde{X}(f) \tag{10.10}$$

and we recall Parseval's theorem, effectively a conservation of energy, i.e.,

$$\int_{-\infty}^{\infty} dt |T(t)|^2 = \int_{-\infty}^{\infty} df |\tilde{T}(f)|^2 \tag{10.11}$$

where$|\tilde{T}(f)|^2 = \tilde{T}(f)\tilde{T}^*(f)$. We put the above together to write:

$$\begin{aligned}
Error &= \sum_k \int df \left| \tilde{V}^k_{pred}(f) - \tilde{V}^k_{app}(f) \right|^2 \qquad (10.12)\\
&= \int df \sum_k \left(\left| \tilde{V}^k_{pred}(f) - \tilde{V}^k_{app}(f) \right|^2 \right) \\
&= \int df \sum_k \left(\left| \tilde{T}(f)\tilde{S}^k_{meas}(f) - \tilde{V}^k_{app}(f) \right|^2 \right) \\
&= \int df \sum_k \left(\tilde{T}(f)\tilde{S}^k_{meas}(f) - \tilde{V}^k_{app}(f) \right) \left(\tilde{T}^*(f)\tilde{S}^{k*}_{meas}(f) - \tilde{V}^{k*}_{app}(f) \right) \\
&= \int df \sum_k \left(\tilde{T}(f)\tilde{T}^*(f)|\tilde{S}^k_{meas}(f)|^2 - \tilde{T}(f)\tilde{S}^k_{meas}(f)\tilde{V}^{k*}_{app}(f) \right. \\
&\quad \left. -\tilde{V}^k_{app}(f)\tilde{T}^*(f)\tilde{S}^{k*}_{meas}(f) + |\tilde{V}^k_{app}(f)|^2 \right) \\
&= \int df \tilde{T}(f)\tilde{T}^*(f) \sum_k |\tilde{S}^k_{meas}(f)|^2 - \int df \tilde{T}(f) \sum_k \tilde{S}^k_{meas}\tilde{V}^{k*}_{app}(f) \\
&\quad - \int df \tilde{T}^*(f) \sum_k \tilde{V}^k_{app}\tilde{S}^{k*}_{meas}(f) + \int df \sum_k |\tilde{V}^k_{app}(f)|^2
\end{aligned}$$

The next step is to minimize the error with respect to the transfer function. We compute the function derivative

$$\frac{\partial(Error)}{\partial \tilde{T}^*(f)} = 0 \tag{10.13}$$

so that

$$\begin{aligned}
0 &= \int df \tilde{T}(f) \sum_k |\tilde{S}^k_{meas}(f)|^2 - \int df \sum_k \tilde{V}^k_{app}\tilde{S}^{k*}_{meas}(f) \qquad (10.14)\\
&= \int df \left(\tilde{T}(f) \sum_k |\tilde{S}^k_{meas}(f)|^2 - \sum_k \tilde{V}^k_{app}\tilde{S}^{k*}_{meas}(f) \right)
\end{aligned}$$

The expression for $T(f)$ must be valid at each frequency. Thus the frequency representation of the transfer function is

$$\tilde{T}(f) = \frac{\sum_k \tilde{V}^k_{app}(f)\tilde{S}^{k*}_{meas}(f)}{\sum_k |\tilde{S}^k_{meas}(f)|^2} \tag{10.15}$$

This is the central result.

For the case of measured signal that is a spike train,

$$\tilde{T}(f) = \frac{\sum_k \tilde{V}^k_{app}(f) \sum_s e^{i2\pi f t^k_S}}{\sum_k \sum_{S,S'} e^{i2\pi f (t^k_S - t^k_{S'})}} \tag{10.16}$$

In the time domain, this is

$$T(t) = \frac{1}{2\pi} \int df e^{-i2\pi f t} \frac{\sum_k \tilde{V}^k_{app}(f) \sum_s e^{i2\pi f t^k_S}}{\sum_k \sum_{s,s'} e^{i2\pi f (t^k_S - t^k_{S'})}}. \tag{10.17}$$

Ugly! But this has a simple form when the spike arrival times may be taken to be a random, e.g., Poisson variable. This occurs if the spike rate is not too high, so that the refractory period plays no role. In this case the denumerator is just

$$\sum_k \sum_{S,S'} e^{i2\pi f (t^k_S - t^k_{S'})} \approx N \tag{10.18}$$

where N is the total number of spikes across all trials, and the numerator is just

$$\begin{aligned} \frac{1}{2\pi} \int df e^{-i2\pi f t} \sum_k \tilde{V}^k_{app}(f) \sum_S e^{i2\pi f t^k_S} &= \sum_k \sum_S \frac{1}{2\pi} \int df e^{-i2\pi f (t - t^k_S)} \tilde{V}^k_{app}(f) \\ &= \sum_k \sum_S V^k_{app}(t - t^k_S) \end{aligned} \tag{10.19}$$

Thus $T(t)$ is just the spike triggered average of the stimulus waveform, i.e.,

$$T(t) \approx \frac{1}{N} \sum_k \sum_S V^k_{app}(t - t^k_S) \tag{10.20}$$

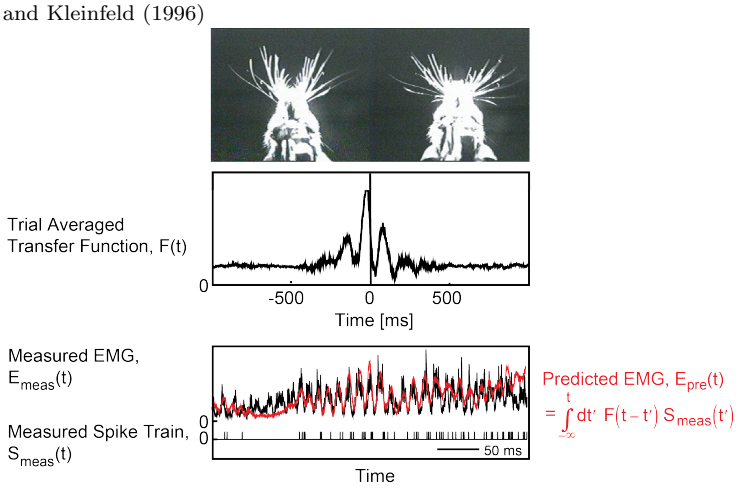
and finally we see that all that happens is that the transfer function reports the waveform of the stimulus that is most likely to cause the neuron to fire.

This procedure has led to an understanding of optimal stimuli for many systems beyond a single neuron (Figure 1). As one example, the optimal self-motion signal as a rat whisks can be based out f recordings from primary vibrissa somatosensory cortex. This defines a movement receptive field (Figure 4).

10.2 General description of receptive fields

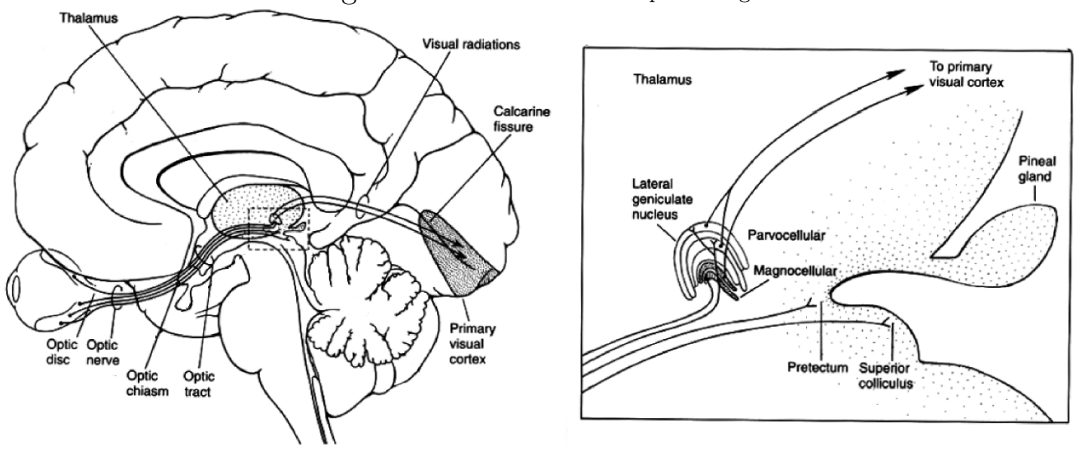
We generalize the phenomenological description of the stimulus that causes a neuron to fire to space as well as time. As a matter of practice, it is convenient to think in terms of the visual system (Figure 5) and visual objects (Figure 6), *i.e.*, a pattern of illumination

Figure 4: The spiking response in vibrissa S1 cortex for whisking without touch for a rat and the calculated optimal filter. From Fee, Mitra and Kleinfeld (1996)



that evolves over time and space. The receptive field forms a kernel, of filter, such that the spike rate of the cell is the temporal convolution of the stimulus with the receptive field and the spatial overlap of the stimulus with the receptive field. The way to think of this is that the inputs to cell comprise a set of photoreceptors, and each receptors has an accompanying synaptic weight and time dependence. This is a lot of information to specify. We shall see that in proactive there is typically only one or two time dependences, each with an accompanying weight matrix.

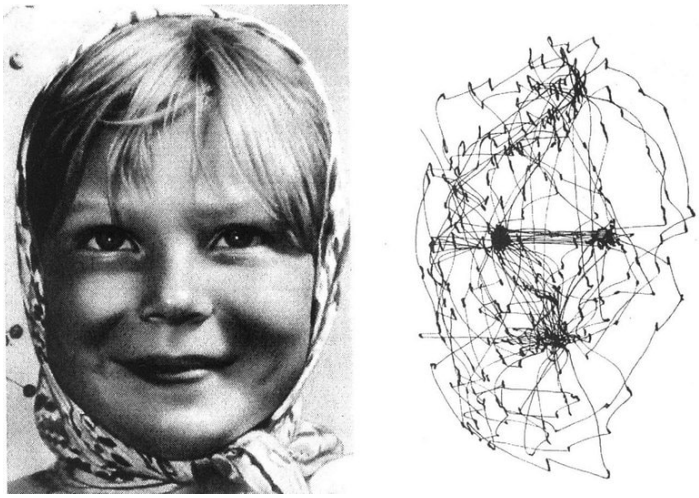
Figure 5: Overview of visual processing



.

We define the inhomogeneous spike rate as $r(t)$. This is the rate that goes into a Poisson rate expression where, for example, the probability of no spikes in the interval

Figure 6: Focal attention on faces causes the visual gaze to be maintained at key locations for 100 ms. From Yarbus



$[0,t]$ and one spike in the interval;$(t,t+dt]$ is $r(t)\cdot exp\left(-\int\limits_0^t dt' r(t')\right)$. Then

$$r(t) = f\left[I_o + \int\limits_{-\infty}^{\infty} d^2\vec{x} \int\limits_{-\infty}^{t} dt' I(\vec{x},t') R(\vec{x},t-t')\right] \tag{10.21}$$

where $f[\cdot]$ is the nonlinear input-output relation, $I(\vec{r},t)$ is the stimulus or input, $R(\vec{x},t)$ is the receptive field with $\vec{x}$ the two-dimensional spatial vector, and I_o is the baseline input.

When the stimuli driven part of the input is small compared to I_o, we can expand $g[\cdot]$ in a Taylor series and write

$$r(t) \simeq r_o + f' \int\limits_{-\infty}^{\infty} d^2\vec{x} \int\limits_{-\infty}^{t} dt' I(\vec{x},t') R(\vec{x},t-t') \tag{10.22}$$

where $r_o = f[I_o]$ and

$$f' = \left.\frac{df}{dI}\right|_{I=I_o} \tag{10.23}$$

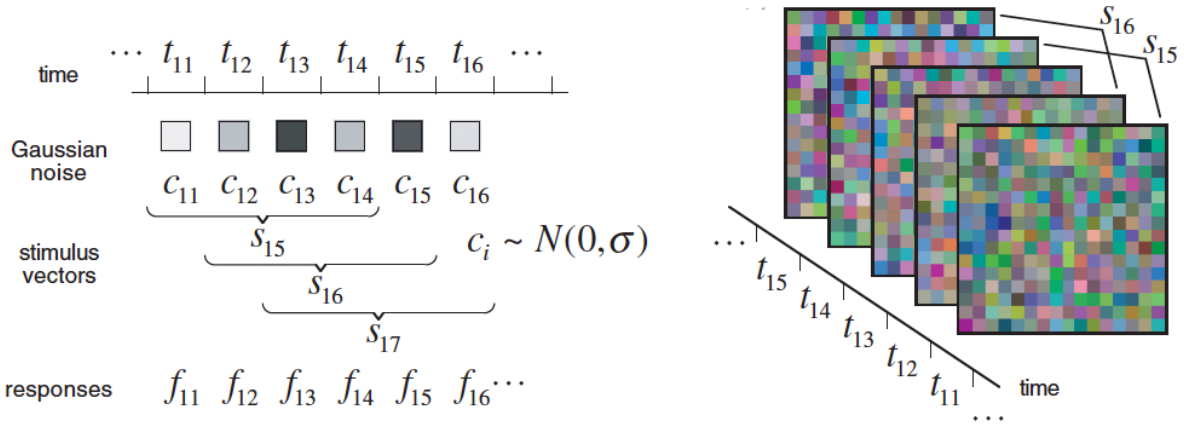
so that the firing rate is a linear function of the stimulus. This allows us to focus on the receptive field without worrying about the nonlinearity $f[\cdot]$. Reviews by Chichilnisky (2001) and by Aljadeff, Lansdell, Fairhall and Kleinfeld (2017) addresses the assignment of both $R(\vec{x},t)$ and $f[\cdot]$ when the stimulus driven part of the input is not small compared to I_o. Aljadeff et al. (2016) also address high-order statistical descriptions of neuronal data.

The simplest procedure to define a receptive field is to compute the spike triggered average, which corresponds to the cross-correlation between the stimulus and the time of an action potential (Figure 7). This defines the space-time receptive field, as the averaging occurs for all lag times. It is illustrated for two classes of neurons in visual thalamus (Figure 8), magnocellular (fast, luminance) versus parvocellular (slow, chromatic

plus luminance sensitive). Note that the receptive field does not have to be simple nor understandable in simple terms!

Figure 7: Receptive field mapping. From Chichilnisky, 2001

.



As a technical issue, any measure of activity can be used to define a receptive field - and that includes calcium signals from cells in cortex (Figure 9). In fact, this method permits the mapping of receptive fields from very many neurons in cortex in the same imaging field (Figure 10).

To gain some insight into the general response properties of neurons, we recall that a matrix can always be expanded in terms of its eigenvectors by a singular valued decomposition (Box 1). In terms of the receptive field, we have

$$R(\vec{x},t) \equiv \sum_{n=1}^{rank(R)} \lambda_n u_n(\vec{x}) v_n(t) \tag{10.24}$$

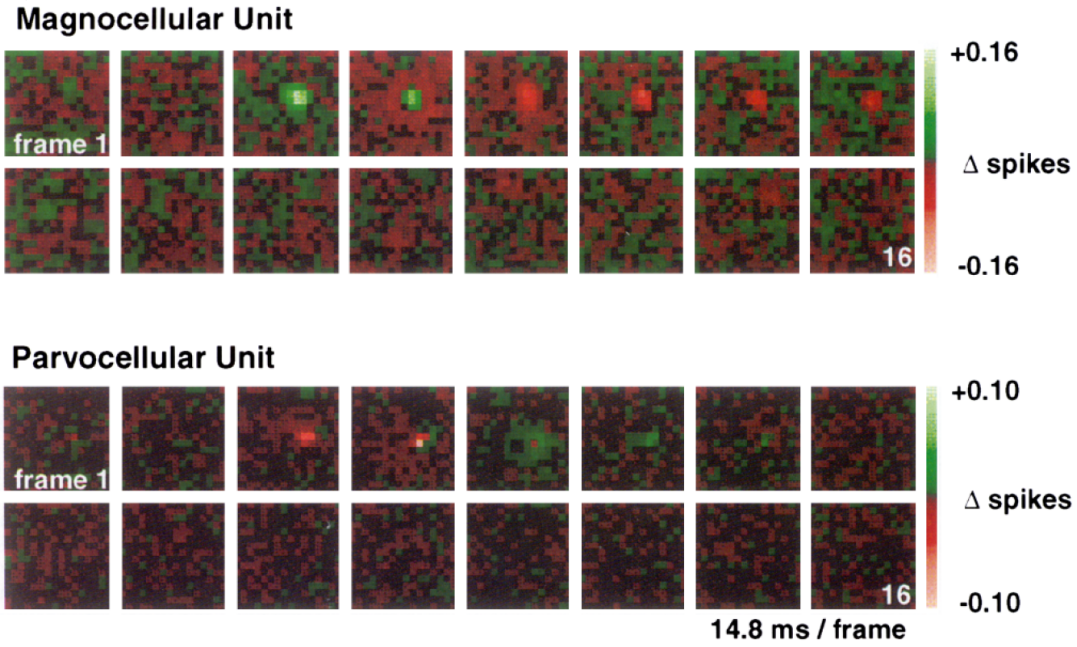
where the functions $u_n(\vec{x})$ form an orthonormal basis set in space and $v_n(t)$ for an orthonormal basis set in time. The eigenvalues for these basis sets are given by λ_n^2 and. of course, are ordered so that $\lambda_1 > \lambda_2 > \lambda_3 \cdots$. When λ_1 is the only significant term the receptive field is said to be separable, as the spatial and temporal functions factor (Figure 11).

10.3 Singular value decomposition

In the expansion

$$R(\vec{x},t) \equiv \sum_{n=1}^{rank(R)} \lambda_n u_n(\vec{x}) v_n(t) \tag{10.25}$$

Figure 8: Spacetime receptive fields for thalamic (LGN) neurons in cat. From Golomb, Kleinfeld, Reid, Shapley and Shraiman, 1994



.

the functions satisfy the orthonormality constraints

$$\int_{-\infty}^{\infty} d^2\vec{x}\; u_n(\vec{x})u_m(\vec{x}) = \delta_{nm} \tag{10.26}$$

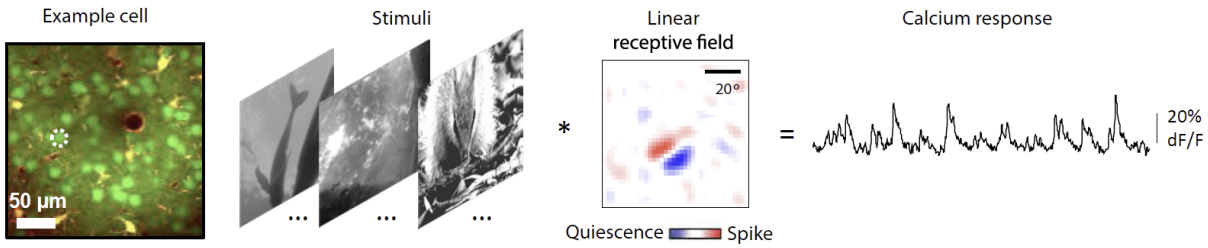
and

$$\int_{-\infty}^{\infty} dt'\; v_n(t')v_m(t') = \delta_{nm}. \tag{10.27}$$

We now consider the contraction of the receptive field matrices to form a symmetric correlation matrix, *i.e.*,

$$\begin{aligned} C(t,t') &\equiv \int_{-\infty}^{\infty} d^2\vec{x} R(\vec{x},t)R(\vec{x},t') \qquad (10.28)\\ &= \sum_{n=1}^{rank(R)} \sum_{m=1}^{rank(R)} \lambda_n\lambda_m \int_{-\infty}^{\infty} d^2\vec{x}\; u_n(\vec{x})u_m(\vec{x})\, v_n(t)v_m(t') \\ &= \sum_{n=1}^{rank(R)} \sum_{m=1}^{rank(R)} \lambda_n\lambda_m\, \delta_{nm}\, v_n(t)v_m(t') \\ &= \sum_{n=1}^{rank(R)} \lambda_n^2\, v_n(t)v_n(t'). \end{aligned}$$

Figure 9: Spatial receptive field for a L2/L3 neuron in mouse visual cortex measured from the neuronal Ca^{2+} response. The temporal dimension has been collapsed. From Cossell, Iacaruso, Muir, Houlton, Sader, Ko, Hofer and Mrsic-Flogel, 1994



Then $v_n(t)$ solves the eigenvalue equation

$$\begin{aligned} \int_{-\infty}^{\infty} dt' C(t,t') v_n(t') &= \sum_{m=1}^{rank(R)} \lambda_m^2 v_m(t) \int_{-\infty}^{\infty} dt' v_n(t') v_m(t') \\ &= \lambda_n^2 \, v_n(t) \end{aligned} \tag{10.29}$$

and the $u_n(\vec{x})$ are found from

$$\begin{aligned} \int_{-\infty}^{\infty} dt' R(\vec{x},t') v_n(t') &= \sum_{m=1}^{rank(R)} u_m(\vec{x}) \int_{-\infty}^{\infty} dt' \, v_m(t') v_n(t') \\ &= u_n(\vec{x}). \end{aligned} \tag{10.30}$$

In general the receptive field is not separable, as first discussed by the work of McClean and Palmer (1989) (Figure 12) and analyzed in some detail by Golomb, Kleinfeld, Reid, Shapley and Shraiman (1994)(Figure 13; this is an analysis of the data in Figure 8). Then

$$r(t) \simeq r_o + f' \sum_{n=1}^{rank(R)} \lambda_n \int_{-\infty}^{\infty} d^2\vec{x} u_n(\vec{x}) \int_{-\infty}^{t} dt' I(\vec{x},t') v_n(t-t'). \tag{10.31}$$

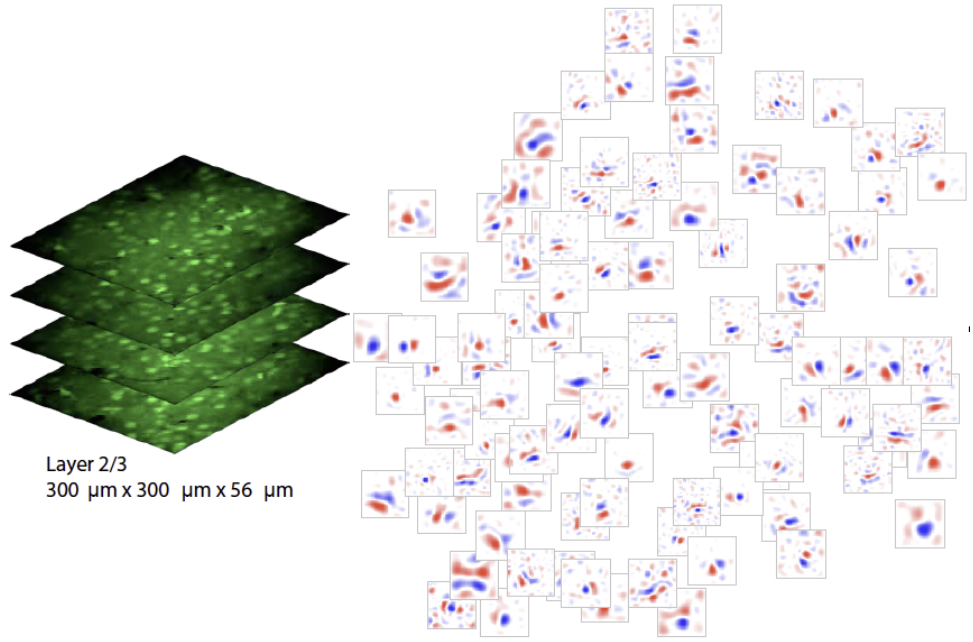
Now suppose that the stimulus is separable, as is often the case in primary sensory areas. For example, in vision our eyes shift from position to position about five times a second. In this case we may write

$$I(\vec{x},t) \equiv X(\vec{x}) T(t). \tag{10.32}$$

so that

$$r(t) \simeq r_o + f' \sum_{n=1}^{rank(R)} \lambda_n \int_{-\infty}^{\infty} d^2\vec{x} X(\vec{x}) u_n(\vec{x}) \int_{-\infty}^{t} dt' T(t') v_n(t-t'). \tag{10.33}$$

Figure 10: Spacetime receptive fields for multiple L2/L3 neurons. From Cossell, Iacaruso, Muir, Houlton, Sader, Ko, Hofer and Mrsic-Flogel, 1994



The spatial part of the stimulus that each mode "sees" is given by the overlap integral of the spatial pattern of the stimulus with the spatial pattern of each mode, *i.e.*,

$$U_n = \int_{-\infty}^{\infty} d^2\vec{x}\; X(\vec{x})\; u_n(\vec{x}). \tag{10.34}$$

where the U_n are scalars. In this case the $u_n(\vec{x})$ act as the weights and the U_n are the output of say a dendritic branch as opposed to the entire cell.

The time dependence of the stimulus is convoluted with each of the associated temporal modes to form the temporal evolution for that mode, *i.e.*,

$$V_n(t) = \int_{-\infty}^{t} dt'\; T(t')\; v_n(t-t'). \tag{10.35}$$

where the $V_n(t)$ are functions. We thus find

$$r(t) = r_o + f' \sum_{n=1}^{rank(R)} \lambda_n\; U_n\; V_n(t). \tag{10.36}$$

so that each temporal waveform is weighted by the expansion coefficient for the receptive field and the spatial overlap of the mode with the stimulus. The point is that the temporal response of the neuron, given by $r(t)$, depends on the spatial pattern of the input as well as the temporal evolution of the stimulus. This is what some call a "temporal code", *i.e.*, the coding of different stimuli, even quasi-static stimuli, by different temporal patterns of

Figure 11: Separable visual receptive field. From Chichilnisky, 2001

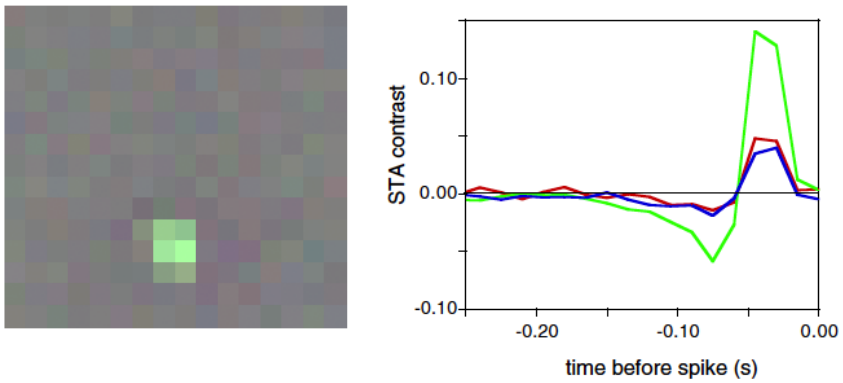
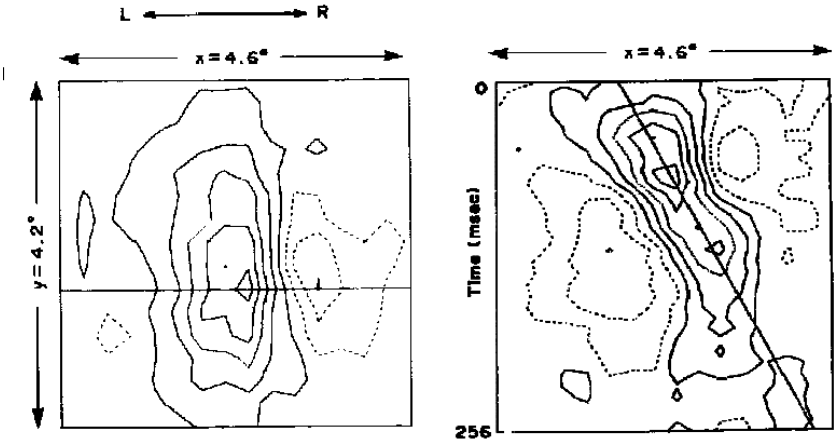


Figure 12: Spacetime receptive fields. From McClean and Palmer, 1989



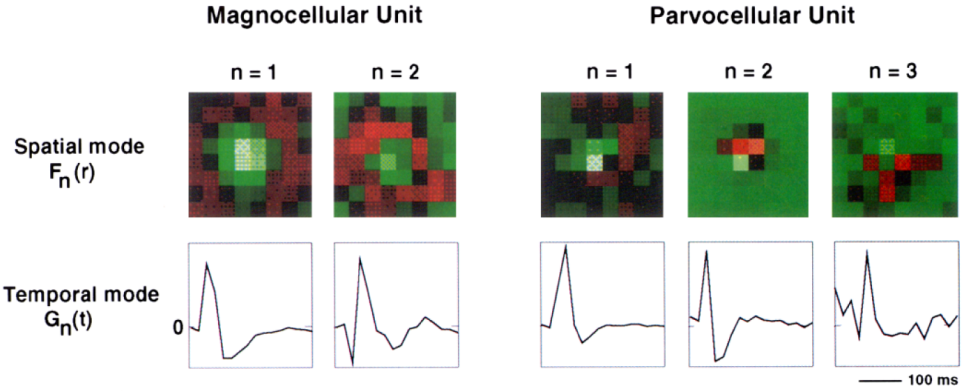
spike rates. The inhomogeneous rate $r(t)$ may evolve in time as fast as the response of the sensory cells, such at retinal ganglion cells for the case of vision.

A final point is that the summation over modes rarely contains more than a few terms, not the full rank of the matrix R. The spatial coefficient U_n has a signal-to-noise ratio that varies in proportion to λ_n for the $n-th$ mode. Thus the above series is cut off after two or three terms as the signal dives below the noise. The SVD expansion can be used as a data compression scheme in the description of the receptive field. For magnocellular cells,

$$r(t) \approx r_o + [f'\lambda_1 U_1]\ V_1(t) + [f'\lambda_2 U_2]\ V_2(t). \tag{10.37}$$

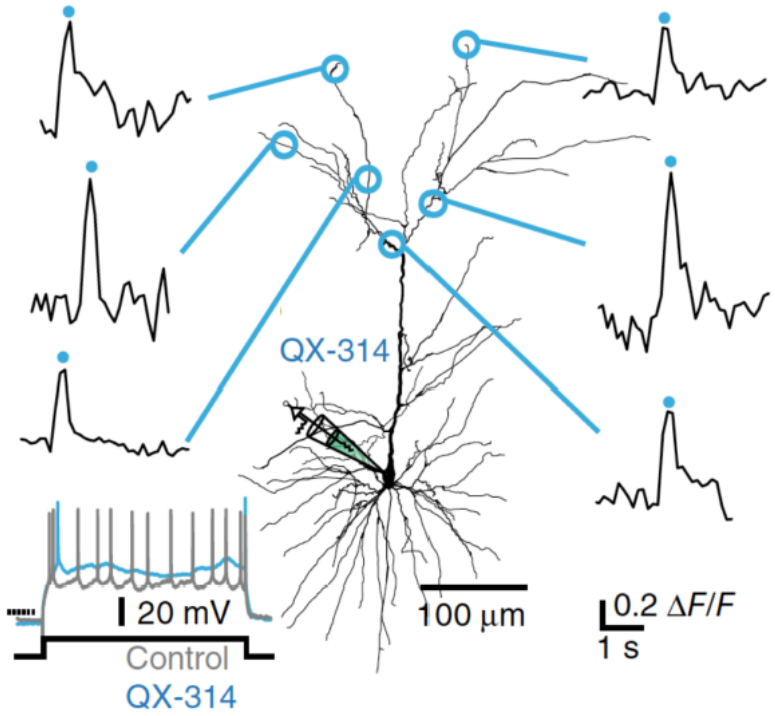
Another interpretation is that each mode corresponds to the input to a different dendrite of a spatially extended neuron (Figure 14).

Figure 13: SVD modes of receptive fields from thalamus. From Golomb, Kleinfeld, Reid, Shapley and Shraiman, 1994



.

Figure 14: Near independent integration of inputs along different dendritic branches. From Palmer, Shai, Reeve, Anderson, Paulsen and Larkum, 2014



.