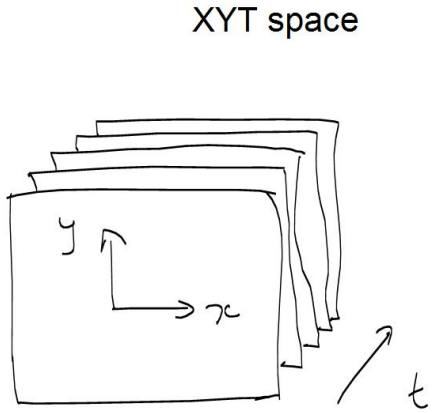
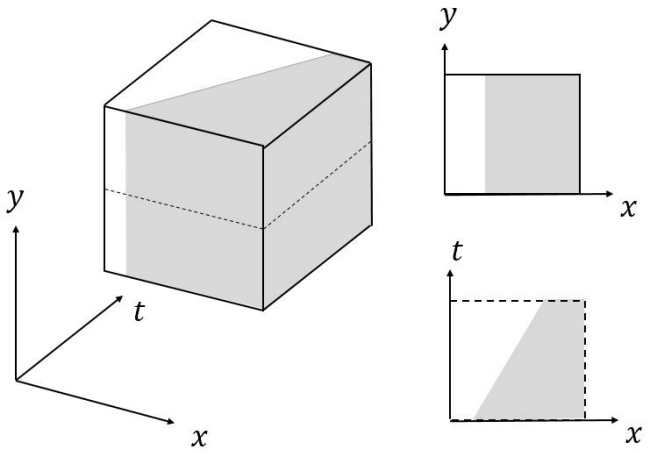
Time varying images

Up to now we have said very little about how images vary over time. But of course they often do. Let's think of an image as a function of x, y and t, namely $I(x, y, t)$.



A video is a sequence of image frames.

An an example image motion, consider a vertical intensity edge drifting to the right over time. The figure below shows a small space-time cube through which the edge passes, and it shows an XY slice and an XT slice through the cube. This edge drifts to the right with speed v_x so v_x is the slope of the edge in the XT slice (where slope is measured $\frac{dx}{dt}$, not $\frac{dt}{dx}$). As an aside for now, note that there could be a motion component in the y direction. However, this component would be impossible to *measure* since the image intensity does not vary in the y direction.

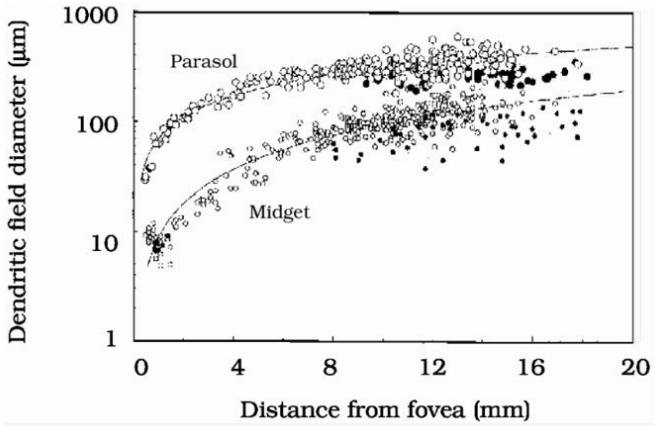


See the lecture slides for two other examples of $I(x, y, t)$. One is just a moving bar instead of a moving edge. The second is more interesting and shows a real video of a person walking from left to right. An XT slice reveals the motion pattern of the person's legs.

Retinal receptive fields and time-varying images

To model how the visual system estimates image motion, we need use model components that build on cells that respond to time varying images. Let's begin in the retina. Photoreceptors measure light intensity continuously over time (unlike digital video cameras which take discrete samples). A photoreceptor does not respond instantaneously, however. Rather there is a delay in the response. There is also temporal blurring, namely if we shine a very brief pulse of light on a photreceptor then the duration of its response will be longer than the pulse.

Retinal ganglion cells also have a temporal dependent response. It turns out there are there are two classes of ganglion cells. These two classes differ in several ways. One is the size of their receptive fields. As the figure below shows, the first class ("midget") of cells have dendrite (bush) diameters that are roughly factor of 10 smaller than the second class ("parasol") of cells. Notice that the sizes of both classes of cells increase steadily as we goes from the center of the field of view into the periphery. Think of the σ of the DOG functions as increasing with eccentricity. Both the difference between the sizes of midget versus parasol and the increase in size with eccentricity are big effects. Note the "x axis" (abscissa) in the figure is on a linear scale whereas the "y axis" (ordinate) is on a log scale.



The response (rate) of a ganglion cell at any time t will depend on the image in some local spatial neighborhood and on some local time interval *in the past.* Consider the XT slice for the cell shown below. Its temporal receptive field lies in the range $t < 0$ and this is meant to illustrate the receptive field weights for determining the response (firing rate) at time $t = 0$. The receptive field can be positive only for $t < 0$ since the cell's response cannot depend on something that hasn't happened yet.

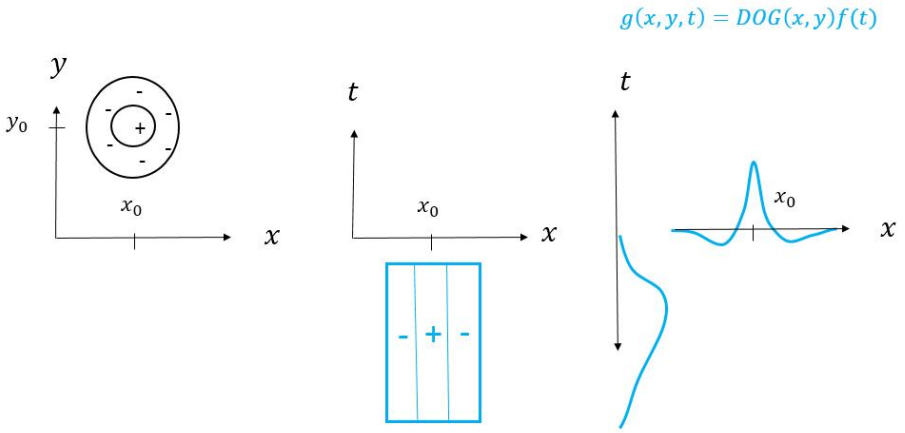
Note that the XT slice for this cell is shown for the slice through the center of the cell. A YT slice through the center of the cell would look similar. Think of rotating the cell's receptive field around its central vertical axis parallel to the T dimension. The cell has a cylinder shape in XYT.

I have given this cell a *separable* response function, namely a DOG in XY and a function $f(t)$ to describe the temporal dependence. Retinal cells do not have separable responses, in general. Intuitively, think of the DOG(x,y) profile as resulting from an excitatory effect of one spatial diameter and an inhibitory effect of a difference spatial diameter, and think of the excitory effect as

having some temporal dependence and the inhibitory effect as have a different temporal dependence. In that case, we might have instead

$$g(x, y, t) = G_{excite}(x, y) f_{excite}(t) - G_{inhib}(x, y) f_{inhib}(t)$$

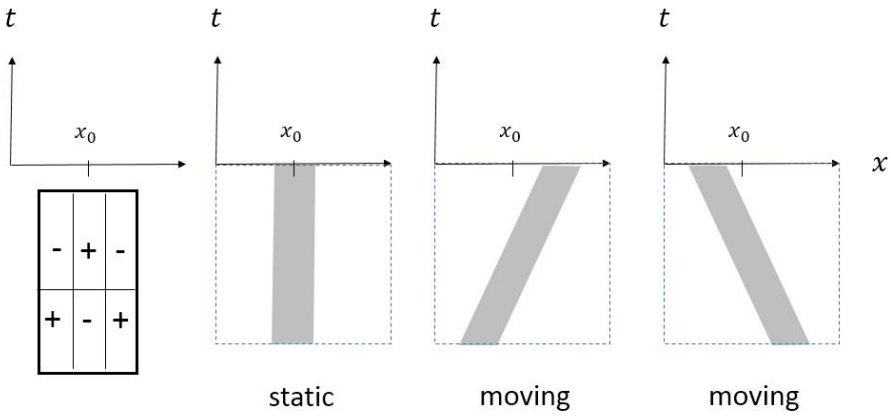
and in this case $g(x, y, t)$ would not be separable, even though the two terms that define it would be separable.



The cell below on the left (only XT shown) *is* sensitive to changes in the intensity over time. As shown in the slides, this cell also could be defined as a separable function. Here the dependence on time could have an excitatory part and an inhibitory part.

This cell would not respond well to a static intensity pattern since at each (x, y) position the pattern would be constant over time (by definition) and the cell's negative and positive weights would cancel. But notice that the cell would give a response to patterns that move over time, and the motion could be either to the left or right. For example, if the motion is at a particular slope in XT, it could cut across the + regions, or it could cut across only the - regions. If the cell's receptive field were stretched or shrunken over time, then the cell would be more sensitive to slow speeds or fast speeds, respectively (see slide).

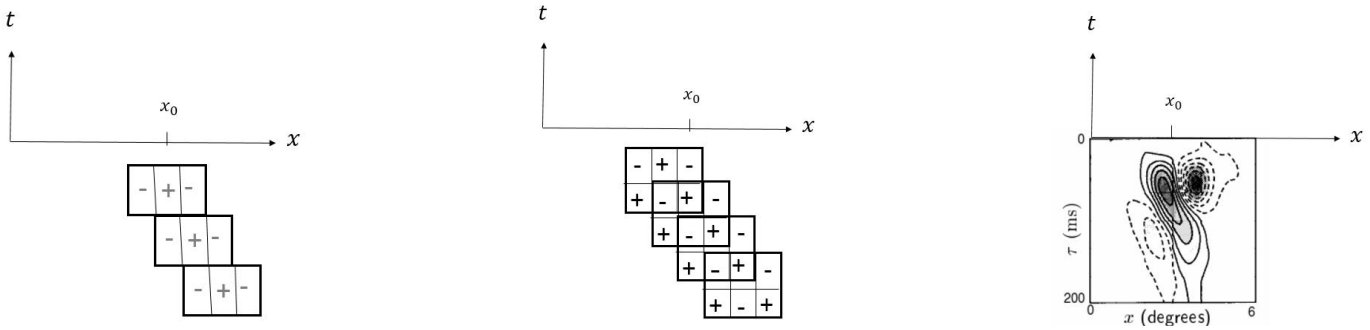
The arguments I am making here are in XT space only. If we consider the Y dimension also, then the arguments require a bit more work to understand and visualize. Let's not go there, since at this point I just want to make the basic point that variation in temporal sensitivity over time can result in sensitivity to motion direction. To really understand the motion system, we need to go beyond the retina (and LGN). Let's do that next.



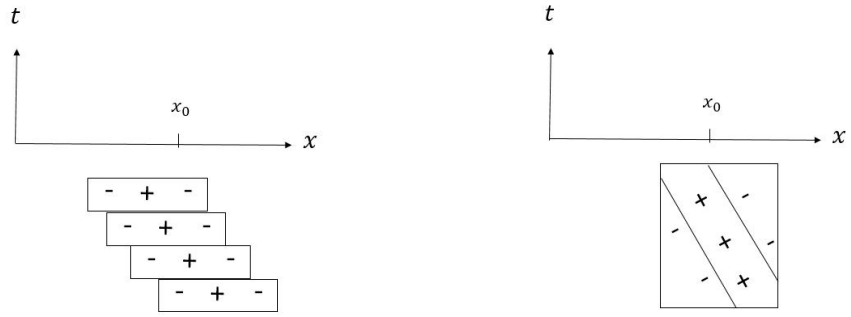
Directionally selective cells in V1

Many cells in V1 – both simple and complex – are sensitive to motion direction, and these cells are also sensitive to orientations (XY). How can the responses of such cells be modeled?

The first idea is that you can define a motion selective cell by summing up the outputs of cells in a time dependent way, namely by delaying the inputs of some cells relative to others. This is the idea of the *Reichart* motion detector that was proposed in the 1950's. The basic idea is illustrated below. The example on the left shows an XT slice through three DOGs that each have a short temporal sensitivity (relative to the previous plots – but keep in mind that the scales of these plots are arbitrary). The idea is that this illustrates one motion selective cell whose receptive field profile is defined by three DOG cells that are delayed in time. The example in the middle is a bit more complicated. Here the receptive field profile is composed from four cells that each have a temporal sensitivity (similar to the one shown at the top of this page, but compressed in time). Overlapping the receptive fields in space-time by delaying the cell inputs will again give rise to a cell that is motion selective. The example on the right is from a real cell. The + and - regions are indicated by iso-contour plots with solid curves indicating + and dashed curves indicating -. *Note that all these cells prefer motion to the left only,* not like the cell's on the previous page which responded to motion either to the left or right.



Let's next consider how to model cells that are both motion direction selective and orientation selective. One way is to stack together XY slices such that the receptive field is shifted by some amount (v_x, v_y) per time slice. (The slice can be thought of as have some duration Δt.) Another way is for the receptive field to be fixed over time, but have the + and - regions within the receptive field shift over time. See below.



One can show (and I will do in a future lecture) that cells whose XY receptive field slices are selective for particular orientations can only detect motion that is perpendicular to that orientation. For example, suppose a cell is sensitive to vertical orientations – e.g. either a cosine or sine Gabor whose underlying sinusoid varies in x only. If the image at that receptive field location contained a vertical line or edge and if that line or edge were moving vertically, then there would be no change in the image across the receptive field regardless of the speed of the line. As such, the cell would be blind to the vertical component of the motion. If, however, the line were to move horizontally instead, then the cell's response would *depend* on the speed of that horizontal motion, in particular, it would depend on how the line or edge fell on excitatory and inhibitory regions in the various XY slices over different times t.

One can model such orientation and motion sensitive cells using Gabor functions. As in the figure above, we could stack together identical Gabors that are shifted over time or we could stack together Gabors that have the same spatial receptive field over time but shift the phase of the Gabor over time, that is, gradually go from a sine to a cosine Gabor over time. (This is a new idea, which I did not mention in the lecture. But hopefully you see the intuition of the idea from the figure above right.)

Another way to define a Gabor is in terms of a sine or cosine function in XYT. Consider a 3D cosine function

$$\cos(\frac{2\pi}{N}(k_0 x + k_1 y) + \frac{2\pi}{T}\omega t)$$

where k_0 and k_1 are fixed integers between 0 and $N-1$, and ω is an integer between 0 and $T-1$. Note that we are sampling time discretely just as we are sampling space.

To understand this function, note the expression in the cosine's argument has a constant value c along a plane in XYT, namely

$$\frac{2\pi}{N}(k_0 x + k_1 y) + \frac{2\pi}{T}\omega t = c.$$

The value of the cosine changes with c and one goes from plane to other plane. Another way to think of it is in terms of a video. Fixing t corresponds to a single frame, and gives a 2D cosine function of (x, y). This cosine has k_0 cycles per N pixels in the x direction and k_1 cycles per N pixels in the y direction. For fixed pixel (x, y), the video changes like a cosine over time t, with temporal frequency ω cycles per T frames. (As an Exercise, figure out the speed of the wave as it travels over time.)

To make a 3D Gabor function, we multiply the 3D cosine or sine by a 3D Gaussian:

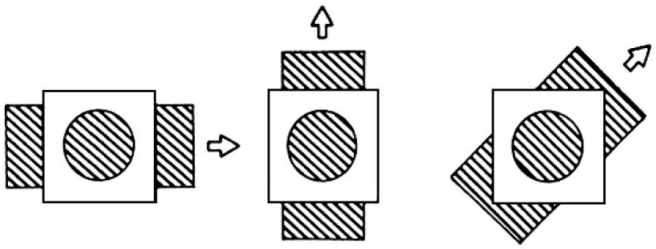
$$G(x, y, t; \omega, \sigma, \sigma_t) \ \cos(\frac{2\pi}{N}(k_0 x + k_1 y) + \frac{2\pi}{T}\omega t)$$

This Gabor is centered at the origin in XYT.

Aperture problem

A space-time Gabor cell will give its best response to an XYT image pattern that matches the Gabor profile. Roughly speaking this will be a moving bar or edge, depending on whether we have a cosine or sine Gabor, respectively. As discussed earlier, such a cell will be most sensitive to line or bar motions that are perpendicular to the spatial orientation of the cell. In particular the cells will be blind to motions that are parallel to the spatial orientation of the cell. I will be technically more precise about these claims later, but for now I just want to familiarize you with it.

The figure below illustrates the issue another way. Suppose we have an image consisting of parallel lines. (The same argument holds for just one line or edge.) Suppose we view that image through an aperture, which we can think of as the receptive field of some cell. In this aperture, we cannot distinguish several different motion vectors (v_x, v_y). We can only "see" the component of motion that is perpendicular to the orientation of the lines. This is known as the *aperture problem*.



The subtlety in the above discussion – and a possible source of confusion – is that we just discussed the orientation both of the cell receptive field and of the underlying image, and those seem like two very different things. They are. However, as we will understand better when we learn about linear image *filtering*, if we are only looking at the outputs (responses) of the Gabor functions then all we get to measure is the image component that has the same pattern as the response: only that component is able to pass through the (Gabor) filter.

