The Retina

The retina is the part of the CNS that sends visual information from the eye to the brain. It very efficient at capturing and relaying as much visual information as possible, under a great range of conditions, from starlight to dazzling sunlight. In many ways it approaches the theoretically optimal limits set by physics and information theory, and does so using a great range of strategies at the molecular, subcellular, cellular and circuit levels. We will only be able to touch on a few of these mechanisms.

There are 5 main types of cell, arranged in 5 layers: light first strikes the most superficial ganglion cell layer, then travels through layers of amacrine cell, bipolar cells, horizontal cells and finally photoreceptors, where about 1/3 of it is absorbed. (More could be absorbed if outer segments were longer or there were extra layers of photoreceptors – but it would be difficult to ensure correct pooling of signals from such multiple layers; in nocturnal animalsa reflective layer beyond the photoreceptors helps capture even more photons). The first 4 layers are transparent. The blood vessels supplying the retina do form shadows, but since they do not normally move they are not normally seen.
However, each of these main classes is comprised of many subclasses, of which only the on/off subclasses of bipolar and ganglion cells are shown.

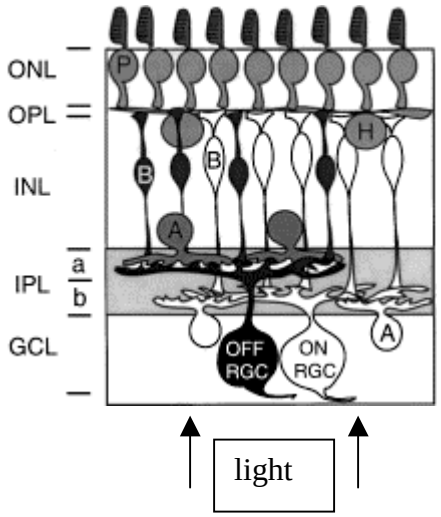

Fig 1. Summary of retinal cells. P = photoreceptor; ONL Outer nuclear layer; OPL outerplexiform layer; H horizontal cell; B bipolar cell; A amacrine cell; IPL inner plexiform layer; A displaced amacrine cell; GCL ganglion cell layer. Note the off cells (black) form synapses in the outer sublamina of the IPL while the on cells (white) form synapses in the inner sublamina; also note the photopigment containing outer segments of the cones at the top of the diagram (rods not shown).

The first step in vision is phototransduction. Curiously, the image is focused at the level of *inner* segments. Cone inner segments have waveguide properties that channel photons into the outer segment. Light is absorbed by the photopigment rhodopsin (which is almost the same in *Homo* and *Drosophila* despite a billion years of evolution) which is a seven-

spanning G- protein coupled protein similar to odorant binding proteins and GPCRs. There are about 0.14 billion rhodopsin molecules per photoreceptor. Activation of rhodopsin activates the G-protein transducin which activated phosphodiesterase which lowers the outer segment cGMP concentration. As a result cyclic nucleotide gated channels in the outer segment shut, hyperpolarizing the photoreceptor. In rods a single photon causes a millivolt hyperpolarisation; cones are much less sensitive, but respond more rapidly. Photoreceptor terminals continuously release glutamate onto bipolar cells; hyperpolarisation reduces this release. Because glutamate hyperpolarizes on-bipolars (via a metabotropic receptor), light depolarizes on-bipolars (hence the name). Glutamate depolarises off-bipolars, which therefore hyperpolarize in response to light. Glutamate release from bipolars then depolarizes ganglion cells, which fire action potentials which are then sent to the brain via long axons.

The foregoing basic picture is complicated by the requirement to operate efficiently over an enormous range of photon arrival rates. In daylight the cone pathway operates, basically in the sequence sketched above. In twilight the cones become silent, but there are still enough photons to smoothly modulate the membrane potential of rods. Because random variations in the arrival rate of photons make individual rods rather noisy, rod signals are pooled. This is done by electrically coupling (through "electrical synapses" see note below) many rods terminals to individual cone terminals; cone terminals are well suited to transmitting to transmitting smoothly modulated voltages to bipolars because they make multiple synapses. Thus in twilight the rods piggy back on the cone pathway, by parasitising cone terminals. This means they can take advantage of the special features of the cone pathway without needless duplication. However in starlight most rods are silent, with occasional well separated single photon events. The great majority of silent rods would be a noisy drag on the few active rods if they were all electrically coupled, so in twilight rods disconnect from cone bipolars and talk directly, via chemical synapses, to rod bipolars. These rod bipolars combine(or "pool") the discrete (single photon) signals from many rods, so their electrical signals are relatively smooth, and thus require many synapses to be made on their targets, the A2 amacrine cells. The A2 amacrine cells make electrical synapses on the terminals of cone bipolars, which as usual make multiple chemical synapses onto ganglion cells). Thus in twilight rods parasitise cone terminals, and in starlight rods (indirectly) parasitise cone bipolar terminals. The reason why the ganglion cells, the sole output of the retina, can be used under three completely different conditions (daylight, twilight and starlight) without confusion is simply that these 3 conditions never occur together. Thus the retina exploits a regularity of the natural world: the visual seen is either seen in daylight, twilight or starlight, and not complicated patchy mixtures of the 3 (which would require three sets of ganglion cells, one wired for each condition). This is a major theme in neuroscience: brains works because the world is a regular place; the brain is complicated because these regularities are often extraordinarily subtle; if the world operated randomly, brains would be useless.

[Electrical Synapses: in addition to the type of chemical synapse we considered earlier in this course, cells are also often coupled by "electrical synapses", where the pre and postsynaptic membranes are physically linked by "gap junction" channels which form a pore that links the interior of the presynaptic cells with the interior of the postsynaptic cell. These channels are hexamers of 12 connexin proteins (6 presynaptic and 6

postsynaptic), and each plasma membrane has a hemichannel, sometimes called a connexon)

Another example of these theme arises in color vision. Why do our cones have just 3 color pigments (4 in new world monkeys, 2 or less in color blindness)? This restricts the analysis of color to a 3 dimensional color space, unlike the thousand dimensional space of olfaction. But in principle even 2 color dimensions would be sufficient to uniquely define every colored surface, provided these surfaces were always illuminated by light of constant spectral composition. The light arriving at our cones differs in spectral composition according to 2 factors (1) the selective reflectance properties of the surface and (2) the spectrum of the illuminant. If the latter was always the same (eg "white light") the relative proportion of 2 different wavelengths of arriving light could be used to distinguish all intrinsic colors (the yellowness of egg yolk versus the blueness of a robin's eggshell). Because natural daylight only varies slightly in composition from that of sunlight (being greenish in under the forest canopy, and reddish at sunset) one extra color dimension suffices. However, humans have invented light sources whose spectral composition varies greatly from that of sunlight, so we have difficulty in recognizing our blue car in the sodium lights of the parking lot. Also, central color processing mechanisms further improve the discrimination of intrinsic color.

It is interesting to compare the "pooling" mechanisms that the retina uses in twilight or starlight to the circuitry in the olfactory bulb. An individual glomerulus combines signals from many sensory neurons that respond to the same odorant. However, each sensory neuron typically provides a very noisy estimate of the concentration of the odorant molecule, since the actual numbers of odorant molecules arriving at a cilium is very small and subject to Poissonian fluctuations (indeed, we may be able to detect single odorant molecules, just as we can detect single photons). By adding together many noisy estimates of a signal one averages out the noise and reveals the common underlying signal. The bulb "knows the noise in the nose".

Spatiotemporal Tuning of Retinal Ganglion Cells

Let us now consider the properties of retinal ganglion cells, which provide the sole output of the eye. What visual stimuli are ganglion cells designed to detect? For what visual prototype are these cells tuned? In other words, what visual stimulus provkes the biggest response from these cells? Why? – in both senses of the word: what circuits generate these responses (and how do these circuits develop?) and why (from an information processing perspective) do they respond in this way?

The first question however, is, how do we measure to which visual stimuli the cells are tuned? One way would be to try all possible visual stimuli, and see which gave the biggest response. This can be done by applying random visual stimuli, and remembering (in a computer) those that give the biggest responses. Actually, since we are really interested in the stimuli that are most likely to trigger spikes, we can store just the average of any stimulus that triggered a spike (together with the time between that

stimulus and the spike). This procedure is called spike triggered averaging or reverse correlation.

For the most part retinal ganglion cells only respond to visual stimuli that fall quite close to the position of the ganglion cell on the retina. This region is called its receptive field. Typically, the receptive fields have a concentric center-surround organization. They fall into 2 classes. “On” cells have RF centers where a brief light increase causes, after a brief delay, a brief increase in spike probability followed by a decrease. Off cells do the reverse.

The circuitry involving these responses is well understood. They are generated by input from “on” or “off” bipolar cells that gather signals from a small patch of photoreceptors . The central photoreceptors in this patch (in the fovea in primates, just 1 bipolar and 1 photoreceptor) create the center, and the surrounding photoreceptors create the surround, by subtracting from the net input to bipolars (the subtraction is done at the level of the photoreceptor synaptic terminals, by feedback from horizontal cells that collect from the surround).

The basic reason for this arrangement is “decorrelation” or “whitening” –an idea related to PCA. If one examines different pixels in different visual images (i.e. the statistics of natural scenes) one finds that if they are close together they tend to be similar but this correlation dies off the further apart the pixels are. Also, the optics of the eye blur the image to some extent, further increasing the correlations between neighboring pixels. Furthermore, pixel values in different successive snapshots of the world are also correlated (e.g. successive frames of a movie), to an extent that dies off with increasing temporal separation. Just as PCA identifies and removes correlations in inputs, so the center-surround receptive field structure removes these visual correlations. If there were no local visual correlations, the best receptive field structure would be a pure narrow center (it would be as narrow as a single photoreceptor if there was no photon noise – this is basically the situation for foveal cones). But because surrounding pixels are typically similar to the central pixel, their signal should be subtracted, which will emphasize the unusual but significant cases where neighboring pixels have very different values. Similarly, because pixel values are temporally correlated, the receptive field structure at early times should be opposite to the structure at later times. This is why the RFs reverse from early to late times (from on center to off center, etc).
If there is a great deal of photon noise (for example in dim light), the need to average signals from neighboring pixels outweighs the need to decorrelate the ganglion cell signals, and the functional circuitry changes, eliminating the surrounds and broadening the centers (and likewise, eliminating the temporal RF reversal). Finally, there are 2 types of ganglion cell (on and off) because they each efficiently represent the positive and negative parts of the image. Thus the spikes of the ganglion cells provide an optimal encoding of visual images, without making any assumptions about image statistics (other than the radially symmetric fall-off of pixel correlations). It should be noted that the overall aim is not for the ganglion cells to send a perfect “image” of the visual world to the rest of the brain, but to efficiently encode the visual information using the limited capacity of their axons. If the brain wanted, it could then use that encoded information to

"reconstruct" the visual scene, but, since there is no "homunculus" in the brain to view that reconstructed scene, it does not do so. Instead, it further manipulates that information so that the aspects that were merely 'implicit" in the retinal encoding become "explicit" in the new cortical re-encoding, and easily available to guide decisions.

[There is one slight complication. There are in fact 2 major types of ganglion cells, small linear ones called midgets or P-cells and large nonlinear ones called parasol or M-cells. This arises because the simultaneous requirements for high spatial and temporal resolution are contradictory: high temporal resolution inevitably requires a greater photon capture rate, which can be achieved only by broadening the collecting area for ganglion cells. This P/M distinction is maintained at many stages of visual processing. Also, of course, resolution varies greatly from fovea to peripheral retina.]

Although the ganglion cells "represent" the visual world in an efficient manner, they do NOT perform principle component analysis. One can identify the principal components of natural scenes but these do not look anything like the localized center/surround structure of ganglion cell receptive fields – for example, they are *global*, combining information from all locations in the retina. Of course if the brain could do accurate PCA, it could in principle "reconstruct" the entire retinal image from a PCA representation by ganglion cells, but this is not much use, since the reconstructed image would then have to be analysed in terms of edges, objects, causes etc (see Neocortex lecture). This further, neocortical, analysis requires an explicitly local representation of the visual world. As we will see, the really interesting stuff (shapes, objects etc) is NOT contained in the correlations between pairs of pixel values (i.e. the sort of representation that PCA generates) but in the "higher-order" correlations (e.g. the fact that one pixel depends on another only if a third pixel has a particular value). We will consider this point further in the lecture on ICA.