

Neural Maps: Their Function and Development

James A. Bednar and Christopher K.I. Williams
Institute for Adaptive and Neural Computation
The University of Edinburgh, UK

Neural Maps: Their Function and Development

Chapter to appear in
From Brain to Cognition via Computational Neuroscience,
Michael Arbib and James Bonaiuto, Eds.,
MIT Press, 2016.

Acknowledgements:

This work was supported in part by the UK EPSRC and BBSRC
Doctoral Training Centre in Neuroinformatics, under grants
EP/F500385/1 and BB/F529254/1. Thanks to David Willshaw,
David Sterratt, and Stuart P. Wilson for comments and discussions.

January 4, 2016

Neural Maps: Their Function and Development

I invoke the first law of geography: everything is related to everything else, but near things are more related than distant things. (Tobler, 1970)

Overview

In this chapter, we first define a *neural map* as a sheet of neurons systematically related to another population of neurons, including the special but common case of a *topographic map* as a spatially organized mapping of this type. We further define a *feature map* as a topographic mapping of some underlying features or patterns in the input data, beyond the strict anatomical arrangement of the input region.

Next, we review the main biological findings about the properties of these maps in adult animals, which will be expanded in more detail when discussing specific models of development and function later. We use well-established examples from the visual system, but highlight similarities and differences with maps in other sensory modalities and in motor regions.

The remaining sections present models and analyses addressing a series of key questions about neural maps: *Where do neural map patterns come from? How do feature maps arise from neural mechanisms? What is the information-processing goal of neurons in maps?*

Together, the models and data suggest that systematic maps first emerge in sensory areas through genetic specification of initial topography, based on molecular guidance cues. The maps are then shaped by activity-dependent and experience-dependent processes in neurons and their connections that lead a neural population to have good coverage of possible stimuli, with neurons responding to the full range of stimuli encountered during development. These processes also typically result in locally smooth maps, though the functional significance of this organization remains unclear.

Biological background

Many regions of the nervous system of vertebrates have been described as *neural maps*. A neural map is a sheet of neurons whose properties are systematically related either to the external world (e.g., visual or auditory space) or to activities in another set of neurons. This systematicity often takes the form of a *topographic map*, with nearby positions in the input space (or output space, for motor maps) mapping to nearby positions in the neural sheet. In this section, we will describe the main features of such maps as found in adult animals, and in later sections we will focus on models and experimental studies of how such maps develop and

function.

Perhaps the clearest examples of topographic maps in the brain are those corresponding to a sensory receptor surface, such as retinotopic maps (regions whose organization matches that of the retina in the eye), somatotopic maps (organized as a “homunculus” matching the surface of the body), cochleotopic maps (corresponding to the cochlea in the ear), rhinotopic maps (of the olfactory epithelium), or whisker barrel maps (of the whisker pad in rodents). Motor cortex also contains several “homunculus” maps, but mapping to the muscles and joints of the body, or perhaps to motor actions, rather than from touch receptors in the skin.

For a sensory surface that is itself organized in terms of some external quantity, such as the retina (mapping the visual field) or the cochlea (mapping auditory frequency), a topographic map can also be defined in functional terms, such as a visuotopic map (an area mapping the visual field) or a tonotopic map (an area mapping auditory frequency). However, note that distances in topographic neural maps typically correspond to the density of receptors on the receptor surface, not to distances in the external world or even distances across the receptor surface. For instance, retinotopic maps in humans typically devote most of their area to the central region of the eye where most of the axons in the optic nerve originate, and somatotopic

maps have a much greater area devoted to receptor-rich areas like the hands and face, rather than reflecting the skin surface area.

The visuotopic maps found throughout the vertebrate visual system are particularly well-studied examples of topographic maps. In a visuotopic map, the visual *receptive field* (RF) of a neuron (the area of the visual field or the pattern to which it responds; see **chapter 1** and **chapter 8**) tends to be similar to those of its neighbors. A large fraction of the macaque monkey cortex can be divided into separate visuotopic maps, each with a representation of the visual field (figure 14.1). Specificity for visual location and overall organization for visuotopy are strongest in the regions most directly connected to the retina, but often remain clear enough elsewhere to be used as a defining criterion for a visual area (Van Essen & Gallant, 1994).

Other topographic maps cannot be directly traced to the geometry of a sensory surface, such as topographic *feature maps* of an abstract feature of the sensory input (also called *computational maps*). Figure 14.2 shows the oft-studied *orientation map* in the primary visual cortex. Here the largest-scale mapping is for retinotopy/visuotopy, but there is also locally smooth (though patchy) organization of tuning for edge orientation at each retinotopic location. As shown later in figure 14.10, numerous other patchy or stripy feature

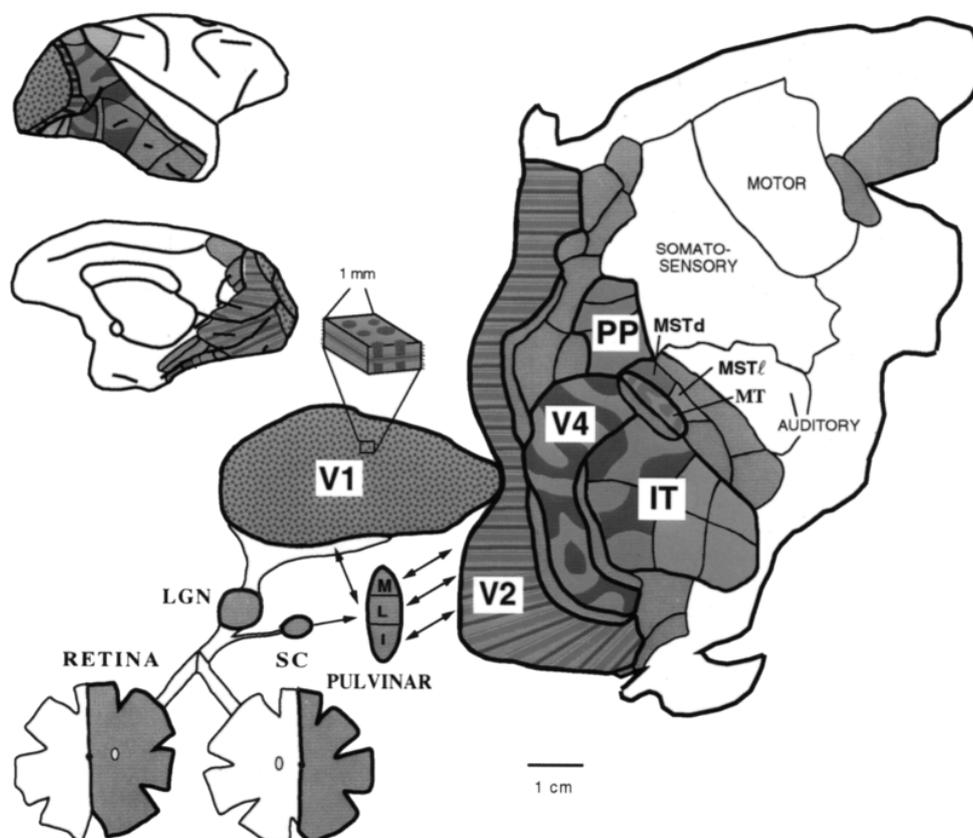


Figure 14.1: **Visual maps in macaque.** The surface of the macaque monkey cortex, viewed laterally (brain, top left), in a sagittal cross-section (brain, middle left), or flattened to show the complete surface of the right hemisphere (middle and right) along with the subcortical paths from the left hemi-retina of each eye (bottom left). When the eye is stimulated visually, there are well-organized responses arranged topographically in the lateral geniculate nucleus (LGN), superior colliculus (SC), pulvinar, and many cortical regions (V1, V2, V4, PP, MST, MT, and to some extent IT). Each of these areas has a map of all or a substantial part of the visual field, with nearby neurons in the map responding to nearby locations on the retina (and thus similar locations in the field of view). Areas farther from the retina (e.g., as measured by latency) have less clear visuotopy, with evidence for organization into more abstract categories like objects, faces, and places, but with some overall visuotopic organization, and moreover local similarity in neural responses. Other sensory cortical areas are similarly mapped to their own receptor surfaces, while primary motor areas are roughly mapped to the body's musculature. Organization of areas distant from the motor and sensory periphery is much less clear (e.g., "association" areas), whether because they have less organization, because they map more abstract quantities or categories, or because they are a combination of sensory inputs and potential motor outputs. Reprinted from Van Essen & Gallant (1994), copyright (1994), with permission from Elsevier.

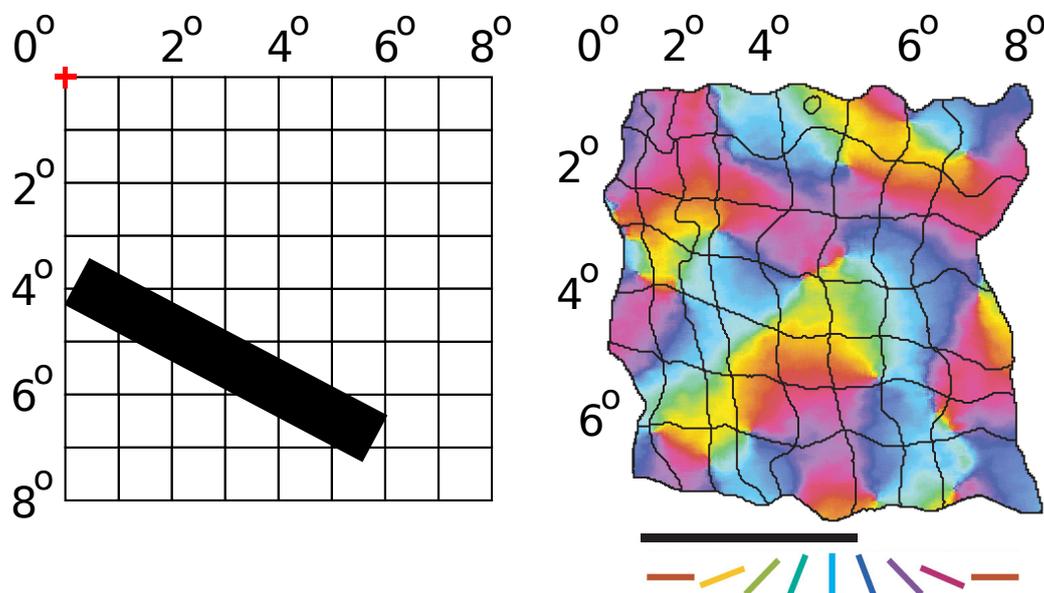


Figure 14.2: **Visuotopic and orientation map in V1. (Color version available in the figure insert.)** Given a particular fixation point (marked with the cross at top left), the visual field seen by an animal can be divided into a regular grid, with each square representing a $1^\circ \times 1^\circ$ area of visual space. In cortical area V1 of mammals, neurons are arranged into a visuotopic map, with nearby neurons responding to nearby areas of the visual field. As an example, the image on the right shows the visuotopic map on the surface of V1 of a tree shrew for an $8^\circ \times 7^\circ$ area of visual space, measured using optical imaging of intrinsic signals (Bosking et al., 2002; scale bar below right is 1mm). A stimulus presented in a particular location in visual space (such as the thick black bar shown at left) evokes a response centered around the corresponding grid square in V1 (3° right, 5.5° down). Which specific neurons respond within that general area, however, depends on the orientation of the stimulus. The V1 map is color coded with the preferred orientation of neurons in each location; e.g. the black bar shown will primarily activate neurons colored in purple in the V1 grid squares around those covered by the stimulus. Similar maps could be plotted for V1 to reveal preference for other visual features, such as motion direction, spatial frequency, color, disparity, and eye preference, as will be shown later in figure 14.10. Adapted by permission from Macmillan Publishers, Ltd: *Nature Neuroscience*, Bosking et al. (2002), copyright (2002).

maps have been measured across the surface of V1 in various species, including motion direction, spatial frequency, color, disparity, and eye preference maps. These maps consist of neurons with receptive fields selective along some or all of these visual feature dimensions, with each map revealing only one type of selectivity amongst the many that could be measured with appropriate stimuli. For instance, a cell might have strong connections (via the LGN) from an oriented pattern of retinal ganglion cells of a certain type in one eye, giving it a specific preference for orientation, pattern size, and eye of origin, and neighboring neurons might have similar but slightly different connections and preferences. Emergent feature maps have similarly been reported for location in auditory space in auditory cortex, along with odotopic maps for odor similarity in the olfactory bulb, and gustotopic maps for taste similarity in gustatory cortex.

Historically, the first clear indications that cortex was organized into topographic maps were from electrode stimulation in epileptic humans by Wilder Penfield in the 1930s. The first evidence for feature maps in sensory cortex came from multiple recordings from an electrode moved along a straight line. These experiments revealed that neurons in somatosensory and visual cortex were locally similar in their responses, but that preferences for features like orientation varied

across the surface of the cortex, with occasional jumps (Hubel & Wiesel, 1962; Mountcastle, 1957). In the 1980s, optical imaging techniques were developed that were able to show organization across a surface area of tens of square millimeters simultaneously, albeit at a low resolution that blurred responses from many adjacent neurons (see example in figure 14.2).

Two-photon imaging of calcium signals developed in the 2000s now allows simultaneous imaging of activity in hundreds to thousands of individual cells across multiple layers (Ohki, Chung, Ch'ng, Kara, & Reid, 2005), albeit for a smaller surface area (typically less than a square millimeter). Studies using this technique have shown that feature maps are largely continuous (i.e., with similar RF) down to the single-neuron level in cat (figure 14.3*b-e*), ferret, monkey, and tree shrew V1, both across the surface of V1 and across the cortical laminae, thus validating the optical imaging results for these species. As first discovered electrophysiologically, the feature maps also contain local discontinuities (regions of quick change in preference) and are patchy overall, with the same feature preference repeated at multiple separate locations in the retinotopic map.

Interestingly, even though rodent V1 has many properties in common with non-rodent mammals, including orientation selectivity and retinotopic maps, there is no apparent continuous topographic or laminar

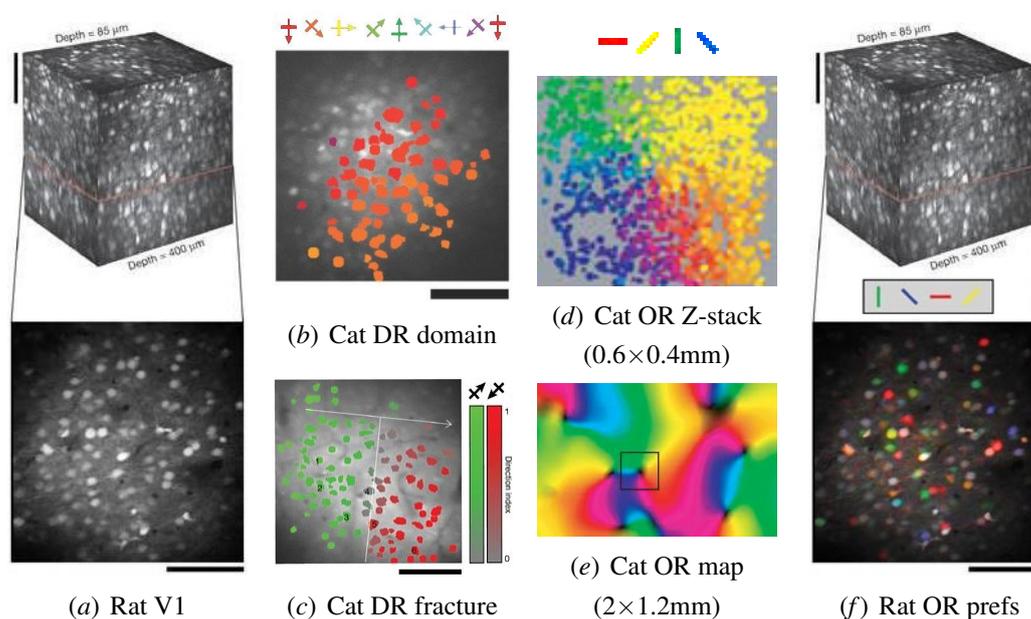


Figure 14.3: Single-cell organization of V1 feature maps. (Color version available in the figure insert.) Two-photon calcium imaging provides single-cell-resolution measurements of neural responses within a volume containing thousands of cells. Plot (a) shows the change in calcium-dye fluorescence in response to visual stimuli, averaged over all stimuli, for rat V1 at each cortical depth. This process labels nearly every neuron in this region, as can be seen in the slice shown for depth $290\ \mu\text{m}$ below the surface. In cat V1, the results show that neural maps are organized exquisitely smoothly with transitions between direction (DR, *b,c*) and orientation (OR, *d*) preferences precise down to the level of single cells, even when all cortical depths are combined together (*d*). These results validate previous low-resolution results from optical imaging (*e*); compare square region in (*e*) with the individually identified neurons in (*d*). Interestingly, despite an overall retinotopic map and strong selectivity for orientation, orientation preference is randomly organized in rodents (*f*), which has not yet been explained adequately in models. Reprinted by permission from Macmillan Publishers, Ltd: *Nature*, Ohki et al. (2005, 2006), copyright (2005,2006); scale bars are 0.1mm.

organization for orientation or other emergent visual features in rodent V1 (figure 14.3*f*), suggesting that the continuity of the map may not be crucial for function. The *coverage* of the maps, i.e., the extent to which they fully tile the multidimensional space of all combinations

of features (e.g., orientation and retinotopic position), seems to be of more functional relevance. Evidence from cats suggests that the maps do provide good coverage

(Swindale, Shoham, Grinvald, Bonhoeffer, & Hubener, 2000), though the methods for evaluating coverage are not necessarily straightforward to apply or interpret (Carreira-Perpiñán & Goodhill, 2002). The models in this chapter will focus on achieving good coverage via locally continuous, topographically organized feature maps, even though continuity does not apply to all maps in all species, because local continuity does provide important constraints on the possible map formation mechanisms. Once map formation has been established in these models, it is possible to account for results in rodents by relaxing these constraints (e.g., by reducing strengths or density of lateral interactions, Law, 2009), but precisely how to do so remains controversial.

Overall, the experimental data from adult animals indicates that neurons throughout much of the cortex and in many subcortical regions are systematically organized into large-scale neural maps, that their response properties together cover some multidimensional input space (and/or output space, in the case of motor neurons), that the maps are usually topographic at some scale, that feature maps are locally smooth and continuous in non-rodent mammals, and

that the precise spatial organization varies depending on the modality, feature dimension, and species.

Questions about neural maps

Despite all of this evidence about the ubiquity and importance of neural maps, it is still not completely clear:

- how and why they develop,
- why they are typically so highly organized,
- what determines the map pattern for different input features,
- what determines the range of values represented in a map,
- why the maps differ between species,
- why they differ between different neural regions processing the same modality,
- to what extent they are specified genetically vs. developing based on experience,
- how multiple overlaid maps relate to each other,
- what precise computations are performed by the circuitry in the maps,
- whether the smooth, topographic organization is functionally significant, or
- to what extent any functional role may be specialized for a specific modality or range of sensory experience.

In each of these cases, there are some data available, and sometimes substantial data, but enough unknowns remaining that computational models can play an important role in formulating concrete hypotheses, evaluating consistency of possible explanations, making predictions to guide experiments, and articulating theoretical principles for interpreting empirical data.

Many of these questions have yet to be answered, but the following sections consider a range of models of various types (see Box 14.1) that address several of the most important ones. First, we explore geometrical developmental models addressing the question “*Where do neural map patterns come from?*” In the following section, we use mechanistic models of maps, receptive fields, and lateral connectivity addressing the question “*How do feature maps arise from neural mechanisms?*” In the last major section, we explore answers to “*What is the information-processing goal of neurons in maps?*” from normative generative models.

We consider each question separately because in recent years, they have been addressed mainly with very different types of models, developed in separate research communities with different goals. In each category, the models we consider in detail are those that are relatively simple, easily understood, and still in current use, rather than either the most complex and realistic models currently available, or models that were

Box 14.1: Descriptions of models

There are many different types of neural map models with different goals and methods. Here we will discuss models of the following non-exclusive types:

Phenomenological: Models designed to reproduce observed behavior, whether or not they work by the same mechanisms as the underlying biological system.

Geometrical: Phenomenological models of the topology or geometric structure of neural maps, not necessarily addressing the behavior of neurons within the map.

Mechanistic: Models with an explicit claim of isomorphism between model elements or mechanisms and the underlying neural hardware.

Developmental: Models explaining how adult-like circuitry or mechanisms can emerge from a simpler starting point in a young animal.

Normative: Models derived not from observed neural elements or circuits, but from an explicit specification that the model should achieve a functional criterion.

Generative: Normative models built with the specific objective of being able to faithfully regenerate the likely sensory input from the map's activity pattern.

seminal yet are no longer used in current publications. Throughout and particularly in the concluding sections, we will suggest how these different models relate to each other and could potentially be combined to provide a comprehensive account of neural maps. The initial models we discuss focus on the most basic observations, of topography and of feature map organization, but in later sections we discuss more complex interdependencies between neurons in the maps that

help reveal their role in sensory processing.

Figure 14.4 shows the basic architecture assumed by most of these models, along with a concrete example of a simple retinotopic map like those found in the lateral geniculate nucleus (LGN) of cats or macaques. In the simplest case, a map model consists of two sets of neurons (**Input** and **Map** (output) in figure 14.4a), each arranged as a 2D neural *sheet* (array of units) with topographic connections between them. More realistic models include multiple sheets and sets of connections (representing interconnected populations of cells in various brain regions), but with the same overall organization in terms of neurons in sheets receiving input from other sheets and from neighboring neurons in the same sheet. To make this type of large-scale model practical, in nearly all cases a neuron is modeled as a single-compartment (point) unit, with the response of a neuron represented by a single number standing for its firing rate. Each unit typically also represents many real units, often collapsing a vertical column of similar neurons into one computational unit for efficiency. Other abstractions are then made in specific cases below, while preserving the overall ability to represent the map computationally. We will first consider the simplest case of mapping a sensory surface, and then consider emergent feature maps and the shape of RFs within them, followed by more detailed studies of how

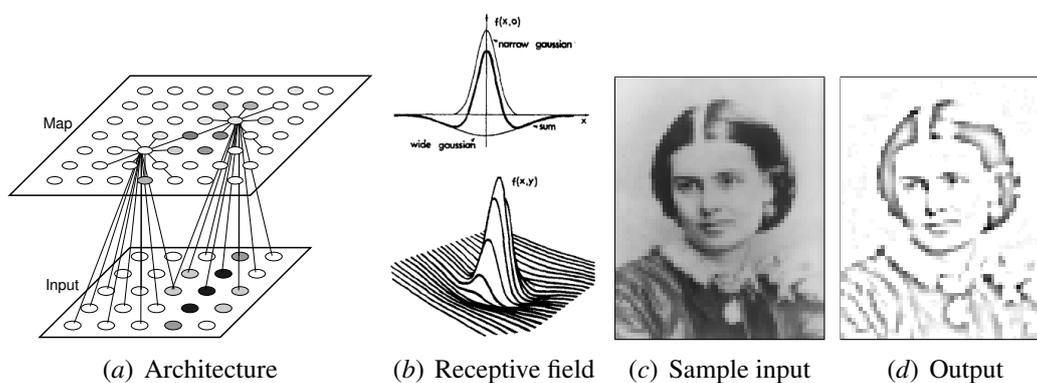


Figure 14.4: Basic map architecture and example. (a) Neurons in a topographic map get input from neurons in the corresponding region of an input sheet, along with input from their close (and possibly more distant) neighbors in the map. Connections to two such units in the map are shown. (b) For each neuron, the incoming connections each have different synaptic weights, which together form a receptive field (RF). For instance, the retinal RF for an LGN neuron measured using reverse correlation can be well approximated using a two-dimensional Difference-of-Gaussians function like the one shown, centered on the appropriate location in the photoreceptors. (Reprinted from Rodieck, 1965, copyright (1965), with permission from Elsevier.) This ON neuron has a positive center and negative surround; the corresponding OFF neuron would have the same RF but with a negative center and positive surround. This function is an abstraction of a complex pattern of connectivity in the retina and LGN, but captures much of the spatial preferences of a given class of LGN neurons (Rodieck, 1965). It is thus a good phenomenological model, even if highly simplified as a mechanistic model. For a simple network with no lateral (intramap) connections and topographically mapped Difference-of-Gaussians input connections, the map performs a parallel convolution (d) of an input image (c), using an OFF RF in this example. I.e., the activation of each output unit in the map is determined by multiplying the difference-of-Gaussians RF with the topographically corresponding input patch, then summing the results to give the value of one pixel in the plot. For this isotropic RF the overall result is simply edge enhancement, but other RF shapes will have different effects, leading to selectivity for specific visual features like orientation. Real neurons also have connections to other neurons across the map surface and feedback from higher areas, which lead to complex interdependencies such as the contextual modulation effects discussed later.

these neurons interact during sensory processing.

Where do neural map patterns come from?

The most obvious question about neural maps is where their large-scale structure comes from – why do they have the patterns that they do? In this section we first focus on topographic maps, with their close relationship to a sensory surface, and then consider feature maps separately.

As a specific topographic-map starting point, how do the retinal ganglion cell axons from the eye form an ordered projection onto the optic tectum of a frog or the superior colliculus of a mammal (a *retinotectal map*)? This question was one of the first about maps to be investigated, and has involved a healthy exchange of ideas between modelers and experimenters ever since. Early experiments by Roger Sperry starting in the 1930s suggested that retinotectal maps develop by chemoaffinity, based on explicit matching of chemical signals between the retina and the target region. The main evidence was that axons growing from each part of the retina appear to target specific corresponding locations in the colliculus/tectum, even when the retinal tissue is moved around or transplanted. Initial ideas of absolute addressing eventually gave way to relative addressing, to account for experiments in which e.g.

half of the retina was ablated and the remaining half innervated the entire tectum. But theoretical considerations suggested that there was also a role for neural activity, and specifically correlations between neighboring retinal cells, as a cue for establishing local neighborhood relationships at the target. Other explanations focused on competition between axons for some fixed synaptic resource at the target. Early models of all three types (e.g., Overton & Arbib, 1982; Prestige & Willshaw, 1975; Willshaw & von der Malsburg, 1976, 1979) helped drive subsequent experiments in this area, by clearly articulating possible alternative explanations.

As reviewed by Flanagan (2006), Hjorth, Sterratt, Cutts, Willshaw, & Eglen (2014), and Huberman, Feller, & Chapman (2008), these later experiments have now given detailed evidence for how these maps develop, and have shown clear roles for all three proposed mechanisms. As illustrated in figure 14.5, the chemical matching process is driven by orthogonal spatial gradients of families of molecules and receptors known as Ephs and ephrins found across the source and target regions, which establish a specific identity for each neuron relative to its neighbors. Once the axons arrive near their targets, neural activity appears to be important for local refinement of the projection, because disruptions to spontaneous retinal

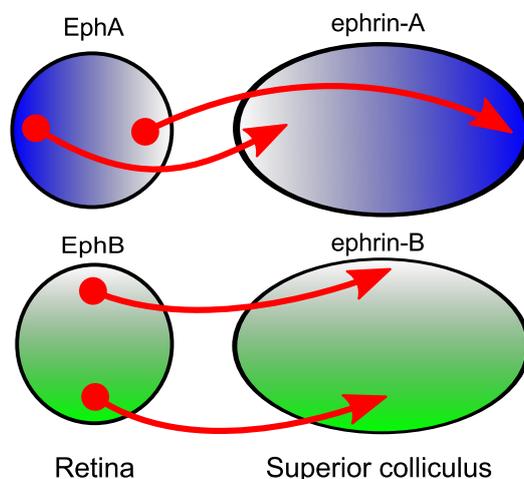


Figure 14.5: **Retinotectal mapping via chemoaffinity.** The retina and superior colliculus of mice establish chemical gradients of signalling molecules and receptors. Neurons in different locations on the retina have different levels of EphA and EphB, forming a two-dimensional retinal coordinate system. The level of EphA or EphB on a growing retinal axon and the level of ephrin-A or ephrin-B it encounters on superior colliculus cells determines how repelled or attracted the axon is to that cell. These molecules and receptors come in families with numerous genetically distinct variants in both source and target, with both attractive and repulsive effects, but only a pair of each is shown here for simplicity. Adapted from Wei et al. (2013).

activity change the specific patterns of connectivity at the topographically mapped location. This process also appears to require some form of competition for targets, such as axon-axon interactions (Triplett et al., 2011).

As reviewed by Hjorth et al. (2014), there are now computational models of retinotectal development that can account for this process in detail, including matching the changes observed in the maps in a wide range of mutant mice that have disruptions to the gradient chemicals, retina structure, and/or spontaneous activity. These models include all three of the mechanisms listed above, i.e. gradient matching,

correlated retinal activity, and axonal competition (e.g. Triplett et al., 2011).

In the next section, we will look in detail at a model containing only the activity-dependent subset of these mechanisms, to highlight similarities and differences between explanations for topographic maps and feature maps. This approach is supported by findings that mechanisms based on activity and chemoaffinity are largely independent (Benjumbeda et al., 2013). We will then consider how activity-dependent processes work alongside chemoaffinity and axonal competition, before investigating feature map development in later sections.

SOM model of topographic map development

The many models that have been proposed for activity-dependent map formation are reviewed in Swindale (1996), and Goodhill (2007). Here we will primarily discuss models that use specific patterns of neural activity, whether spontaneous or visually evoked, rather than more abstract approaches without individual images or specific activity patterns (such as Carreira-Perpiñán, Lister, & Goodhill, 2005; Durbin & Mitchison, 1990; Linsker, 1986; Miller, 1994).

Many activity-dependent models can be traced back to von der Malsburg's original (1973) network, later

extended to address retinotopy (Willshaw & von der Malsburg, 1976). This model and many others closely related to it (e.g., Grossberg, 1976; Kohonen, 1982; Obermayer, Ritter, & Schulten, 1990) use a set of 2D input patterns presented to a network with initially random local connectivity, some form of lateral interactions in the target region that provide competition and cooperation between neurons, and a local Hebbian-based synaptic modification rule (see **chapter 8**).

The variant of von der Malsburg (1973) we will first consider here was proposed by Kohonen (1982), and is called the self-organizing-map (SOM) algorithm. Even though SOM is highly abstract, it has remained in common use (e.g., Obermayer et al., 1990; Yu, Farley, Jin, & Sur, 2005) because it is simple to implement, robust, computationally efficient, and faithfully reproduces the observed map patterns. SOM is thus a good phenomenological, geometrical model, if not a mechanistic model. In later sections we will present a more complex, mechanistic implementation that shares many of the same principles while being more easily related to the cortical circuitry.

The simplest Kohonen SOM model has an architecture like that in figure 14.4*a*, with two sheets of neural units: a map sheet (e.g. the retinal ganglion cells) and an input sheet (e.g., the optic tectum or superior

colliculus), along with a set of connections from the full set of input units to each map unit. Starting with initially random weights (and thus no topography and no feature selectivity), the SOM is designed to develop map units that will respond to input patterns that are seen during development, achieving a balance between good coverage of the full range of input patterns and continuity between input pattern preferences of nearby neurons.

The SOM algorithm has two alternating phases: computing map activity (neural responses), and adjusting weight values. Given a 2D array of input activity values \mathbf{s} , for each map neuron (k, l) the SOM first computes the Euclidean distance v_{kl} between \mathbf{s} and a 2D weight vector (\mathbf{w}_{kl} ; initially random):

$$v_{kl} = \|\mathbf{s} - \mathbf{w}_{kl}\|. \quad (1)$$

In a process meant as an idealization of lateral interactions between neurons in the map, the initial activity of the map neurons is then constructed as a single isotropic “neighborhood kernel” function centered on the neuron (r, s) for which v_{rs} is smallest. Unit (r, s) is considered the “winner”, as it has the weight vector most similar to the input pattern, and is considered to be responding most strongly. The final activity of each map unit (i, j) is then computed as a fixed function of its distance from (r, s) , using a fixed

kernel function (e.g., a 2D Gaussian):

$$h_{rs,ij} = \exp\left(-\frac{(r-i)^2 + (s-j)^2}{\sigma_h^2}\right), \quad (2)$$

Each weight $w_{k,ij}$ from input unit k to map unit (i, j) is then updated by a variant of the Hebbian rule where the input is taken relative to the current weight:

$$w'_{k,ij} = w_{k,ij} + \alpha(s_k - w_{k,ij})h_{rs,ij}. \quad (3)$$

The result of this process is that the winning neuron and its distance-weighted neighbors now have weights that are more similar to this input pattern, a form of positive feedback that gradually drives neurons to become selective, and drives nearby neurons to become more similar. As other neurons elsewhere in the map “win” for different input values, each patch of map neurons in turn becomes selective for some part of the input space. Implementations typically gradually reduce the cortical kernel width σ_h and learning rate α so that neurons can gradually differentiate, with different preferences in different locations in the map. The result provably converges to a smooth and complete mapping of the input space, in the limit of infinitely slow learning rate and radius reduction over the course of self-organization (Erwin, Obermayer, & Schulten, 1992). In practice, the mapping may actually be discontinuous or patchy, depending on the values of these parameters.

Note that if $h_{rs,ij}$ were set to a Dirac δ function (a neighborhood of just the winning neuron), this algorithm would reduce to k -means clustering (Hartigan, 1975). I.e., the network would assign each input pattern to one of the k output neurons, updating its weight vector with an incremental estimate of the mean of all inputs activating that neuron (Bishop, 2006). Thus one way to think of the SOM is as a form of clustering, plus a local smoothing value that achieves a continuous topographic map via lateral interactions.

For an example of this self-organization process, we can specify a set of input patterns with known properties, and then see how the SOM responds. Consider a Kohonen SOM model with a 24×24 input sheet and a 40×40 map. For each input presentation, let us create random spatially localized training pattern images (e.g., small 2D isotropic Gaussians), such as might occur during retinal waves. This input is high dimensional ($24 \times 24 = 576$), but actually varies in only two dimensions, the x and y of the pattern centers. I.e., out of all of the possible image inputs, actual training patterns for this SOM lie only on a 2D *manifold*, a lower-dimensional surface embedded in the higher-dimensional space. SOM (and many other manifold learning algorithms; see Bishop, 2006) can approximate this underlying (“latent”) 2D manifold using a discrete set of neurons arranged in 2D, each

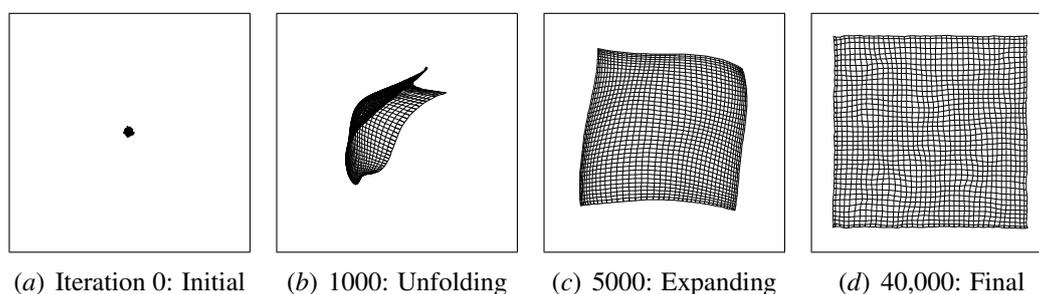


Figure 14.6: Unfolding to represent a 2D space. For this SOM network with a 24×24 input, all input patterns have been chosen to be identical 2D isotropic Gaussian spots, differing only in their (x, y) position. Each of these plots shows the full 24×24 input space, with a point for each map neuron located at the center of gravity of that neuron's weight vector. Each neuron is plotted with a line segment connecting it to its neighbors in the map array, yielding a grid-like plot if the map neurons have centers of gravity that smoothly tile the underlying 2D space of pattern centers. (a) The initial weight vector for each neuron is random, and thus they are all mapped to the center of the input space (as the center of gravity of all of them is approximately equal). (b,c) Over time, the Hebbian learning process applied to winning neurons and their neighbors leads to neurons differentiating to cover the input space, while varying smoothly in their response properties due to the smooth neighborhood kernel. Weight values are adjusted to become more similar to the input patterns each time, and the centers of gravity gradually unfold to cover the 2D manifold on which the inputs lie, with each map unit responding to some portion of the 2D space. (d) The final result by iteration 40,000 is a smooth mapping covering the area of the input space that was encountered during training, with arbitrary final orientation and flipping. Here the mapping is flat, with equal cortical area for each location in the input space, but non-uniform distributions lead to mappings that reflect the probability density of each location in the input space (cf. Recanzone et al., 1992 for similar phenomena in somatosensory cortex). Reprinted from Miikkulainen et al. (2005) with permission from Springer Science and Business Media.

responsible for part of the range of the input space over which the inputs varied during training (figure 14.6).

This example simulation shows that, in principle, neighborhood information provided by locally correlated retinal activity could be sufficient to drive topographic map formation in a target region (as suggested by Willshaw & von der Malsburg, 1976). We

focused on this subset of the mechanisms because, as we will see in the following sections, similar mechanisms can also explain feature map formation. But this example also serves to highlight the limitations of activity-only accounts of topographic map formation: (1) The model provides nothing to ensure the overall orientation of the map (such as whether up on the retina is to the right in the cortex), contrary to experimental observations of consistent orientation. (2) It assumes full initially random connectivity, which is likely only to be feasible for a very small network, and moreover has not been observed experimentally. (3) It does not provide explanations for how maps can develop under activity blockade, nor for the specific changes when genes related to the chemical gradients are knocked out.

As described above, addressing these limitations requires adding mechanisms explicitly incorporating chemical gradients and chemoaffinity processes, which establish the overall orientation of the map and the initial topography. Several such models are discussed by Hjorth et al. (2014). In the following sections, most of the models pick up where the gradient-based explanations leave off, modelling activity-dependent feature map formation once the overall topography has been established. As a first example of feature map development, in the next section we will consider a simple extension of this SOM model to address ocular

dominance.

SOM models of feature map development

Given an overall topographic map, what explains the complex observed maps of neural feature preferences, such as orientation maps and other retinotopic visual feature maps? Some of these maps are already present at eye opening (including orientation and ocular dominance; reviewed in Huberman et al., 2008), suggesting that genetic mechanisms such as chemoaffinity could be involved in establishing feature maps as well. However, the development of feature maps appears to depend very strongly on neural activity and these maps show widespread experience-driven plasticity during a postnatal *critical period* (see Kaas, 1991 for review). Thus most models of feature maps focus on the activity-dependent processes.

To explain the feature maps seen at birth, these models typically assume a prenatal phase of spontaneous activity, such as the spatiotemporally localized patterns of retinal waves. Indeed, disrupting spontaneous activity patterns does disrupt the formation of these maps, though there is not yet consensus on the initial origin of the maps (reviewed in Huberman et al., 2008). In the rest of this chapter, we will focus primarily on models whose inputs could be either

spontaneous or visually evoked activity, which lead to precise connectivity patterns through activity-dependent unsupervised learning rules (see **chapter 8**). As outlined below, the resulting feature map patterns are a close match to those observed in animals, suggesting that a similar process may be occurring in nature.

As a first example, we will again consider the SOM network from the previous section, this time focusing on the mapping between the LGN and V1. We saw previously how a SOM can recover an underlying manifold that is of the same dimensionality as the SOM map – a set of neurons arranged in 2D forming a smooth map of an underlying 2D space over which the pattern centers were chosen. We can now consider what happens when the underlying manifold is of a *higher* dimensionality than the SOM map, and show how this process helps explain the patterns of emergent feature maps, not just the initial topography.

Let us now examine a SOM model where the inputs are chosen from a 3D space of (x, y, e) , such that the pattern is either in the left eye only $e = 0$, the right eye only $e = 1$, or in both eyes at different brightnesses ($0 < e < 1$). Figure 14.7 illustrates the map pattern developed by such a SOM (cf. Miikkulainen et al., 2005; Ritter, Obermayer, Schulten, & Rubner, 1991), which resembles the stripy pattern of ocular dominance (eye preference) maps seen in V1 of macaque monkeys

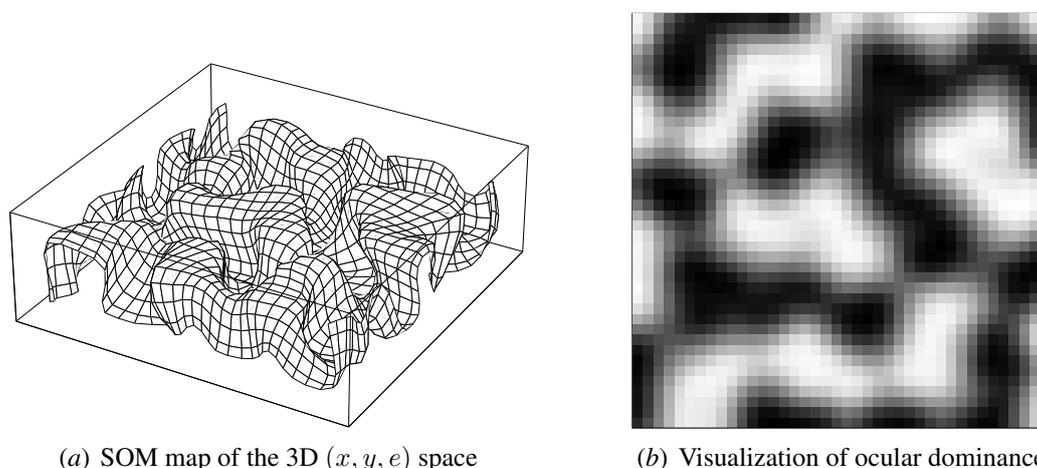


Figure 14.7: **Mapping three dimensions onto two.** This SOM model has developed a two-dimensional cortical map of a three-dimensional input feature parameter space (x, y, e) . The first two dimensions can be interpreted as retinotopy and the third dimension as ocular dominance (Ritter et al., 1991). In (a), the full 3D input space is indicated by the box outline. Weight vectors of the map units are plotted in this space as a grid, with adjacent units connected as in figure 14.6. The 2D cortical space has developed a representation that effectively folds to fill up the 3D input space, with each unit in the SOM map responding best to some portion of the 3D space, just as for the 2D case. (b) The weight value for the height dimension e is visualized for each cortical neuron: Gray-scale values from black to white represent continuously changing values from left to right eye preference. The resulting pattern resembles the ocular dominance stripes found in the visual cortex (compare to “OD, Macaque” in figure 14.10), suggesting that this feature map could be the result of a self-organized 2D mapping of a three-dimensional (or higher) parameter space. Reprinted from Miikkulainen et al. (2005) with permission from Springer Science and Business Media.

(LeVay, Hubel, & Wiesel, 1975).

SOMs are a type of dimension-reduction model, where the cortical surface is (typically) of a lower dimensionality than the input space (Durbin & Mitchison, 1990; Hastie & Stuetzle, 1989; Ritter et al., 1991). While folding to represent eye preference results in stripy or patchy patterns like ocular dominance maps, SOM folding to represent *cyclic*

quantities like orientation generates realistic pinwheel-based orientation maps, in part due to geometric constraints on achieving map continuity for cyclic quantities in a 2D plane (Obermayer et al., 1990). Interactions between multiple simultaneous features also appear to be realistic in SOM models, with feature values interacting such that areas with high selectivity and relatively low gradient (change in feature value across the map) for one feature are areas of low selectivity and high gradient for the others (Yu et al., 2005). Similar results have been found using other abstract geometric models using similar principles, such as the Elastic Net (Carreira-Perpiñán et al., 2005; Goodhill & Willshaw, 1990), and for SOM models of other modalities (e.g., the somatosensory homunculus, Ritter et al., 1991).

The results with SOM and related models have been very useful for understanding the geometrical process of map formation, in terms of input and output spaces. However, because of the abstract level at which these models are formulated, it is difficult to relate them to biological processes at any level of detail. For instance, the SOM process of picking a global winner requires unrealistic global knowledge of the network state, the process of neighborhood interaction is based on isotropic connectivity rather than the patchy connectivity seen in the visual cortex, the radius of

lateral connectivity has not been found to decrease over time in animals, and the Euclidean distance response function is difficult to relate to the mechanisms of input summation in real neurons. The following section will consider how some of the same dimension-reduction and folding principles could be implemented in models that can be related more directly to the circuitry of the cortex, in the hopes of revealing detailed explanations for cortical phenomena other than the map patterns themselves.

How do feature maps arise from neural mechanisms?

The type of model considered in the previous section focused on explaining feature map patterns in the abstract, glossing over the specific details of map formation and map function in animals. Including more of those details to build a functioning model sensory system leads to *mechanistic* models that can explain not just the map patterns, but how the neurons in the map function when processing realistic inputs. For instance, models that can process natural images (e.g., Burger & Lang, 1999; Miikkulainen et al., 2005; Stevens, Law, Antolik, & Bednar, 2013) can be tested with any visual pattern used in an experiment, in order to determine how closely the model map matches the behavior of an animal's cortex. Models that include

specific, modifiable lateral connections (Burger & Lang, 1999; Miikkulainen et al., 2005; Sirosh & Miikkulainen, 1994; Stevens et al., 2013) can be used to investigate phenomena like contextual modulation and surround suppression that arise through interactions between large numbers of neurons in maps. Yet as shown below, these models can still develop feature map patterns closely resembling the results of SOM and other more abstract models, unifying explanations of map formation and map function.

Several closely related mechanistic models of this type have been proposed, each using firing-rate single-compartment neurons, local Hebbian synaptic learning rules, and explicit patterns of connectivity rather than global picking of winners or optimization of abstract quantities (Barrow & Bray, 1992; Burger & Lang, 1999; Miikkulainen et al., 2005; Sirosh & Miikkulainen, 1994; Stevens et al., 2013). The idea behind each of these is to show how simple neuron models, plausible wiring patterns, and observed synaptic plasticity mechanisms could together lead to the observed properties of neural maps.

GCAL model of map development and function

Here we will focus on the GCAL (Gain-Control/Adaptive/Lateral) model of cortical

topographic feature maps, which has both a simple version suitable for analysis here (Stevens et al., 2013) and more complex implementations more closely tied to the neuroanatomy (Bednar, 2012). GCAL is a generalized variant of the LISSOM family of cortical map models (Miikkulainen et al., 2005; Sirosh & Miikkulainen, 1994), replacing implausible mechanisms inherited from SOM with well-established local mechanisms such as homeostatic plasticity (described below). The resulting model shares general mechanisms with many other models, and can be considered a more-detailed elaboration of the original von der Malsburg (1973) model.

The simplest GCAL model of the pathway to V1 (Stevens et al., 2013) uses four 2D sheets of neurons: Photoreceptors, Lateral Geniculate Nucleus/ Retinal Ganglion Cells (LGN/RGC) ON, LGN/RGC OFF, and V1 (figure 14.8). Activity of the Photoreceptors is initialized using a bitmap image from an image database or from a synthetic distribution. Feature maps will only develop for features present in these input images; e.g. if the input has only small, isotropic Gaussian patterns, no orientation map will develop. The ON and OFF sheets receive input from the photoreceptors and from neurons in the same sheet, as shown in figure 14.4a. These cells are meant to represent the complete pathway to V1's input layer phenomenologically, including

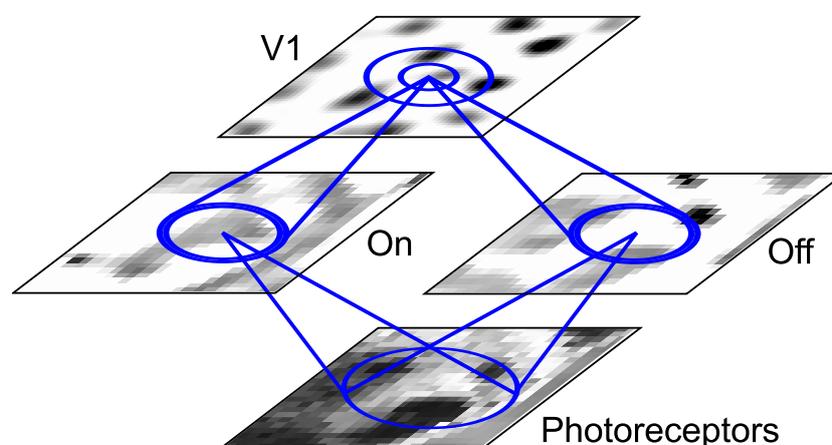


Figure 14.8: **Basic GCAL model architecture.** The simplest GCAL model, for developing a retinotopic orientation map. The model consists of four neural sheets and 8 separate projections between them. Each projection is illustrated with an oval showing the extent of the connection field in that projection. Afferent projections are shown with lines converging on the target of the projection, while lateral projections in the ON, OFF, and V1 layers connect within the same sheet. The single photoreceptor sheet represents input from one eye, processed by hardwired center-surround ON and OFF RGC/LGN cells, and driving a V1 sheet. The initial V1 response to an image is a blurred version of the contents of the ON and OFF sheets, because of the initially random weights, but V1's lateral connections settle the activity into discrete bubbles (multiple “winning” patches). Connections to the neurons remaining active strengthen via Hebbian learning, causing local patches of V1 to become selective for particular input patterns. Adapted from Stevens et al. (2013).

processing in both the retina and in the LGN. The V1 sheet receives input from the ON and OFF sheets, as well as lateral connections from within the sheet. A full GCAL model of the known V1 feature maps is shown in figure 14.9, which includes numerous sheets representing different subcortical and cortical cell types and laminae (Bednar, 2012), but the simpler four-sheet version described here illustrates the key principles.

For GCAL, the subcortical pathways are treated as

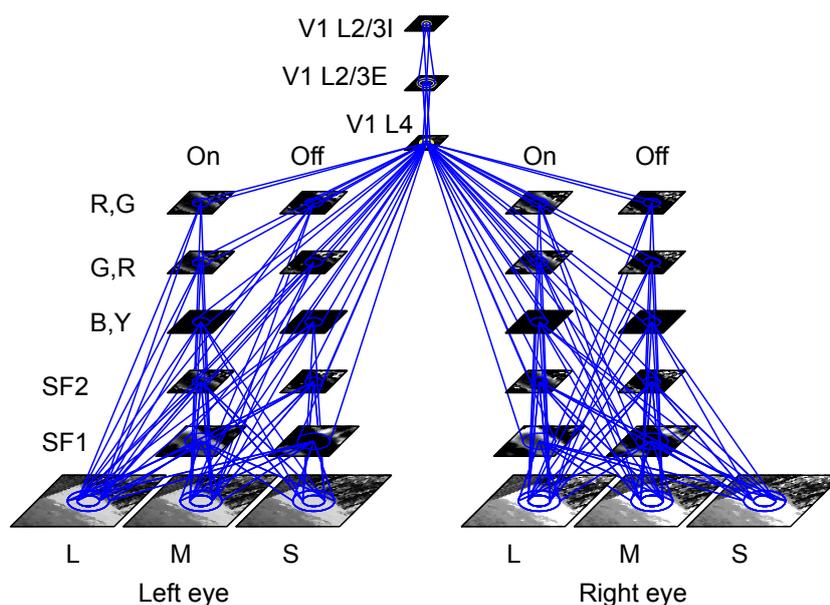


Figure 14.9: **Comprehensive GCAL model architecture.** GCAL model for simple and complex cells with surround modulation and maps for retinotopy, orientation, ocular dominance, disparity, motion direction, temporal frequency, spatial frequency, and color. The model consists of 29 neural sheets and 123 separate projections between them. The six L, M, and S sheets represent the three photoreceptor cone types in each eye, the twenty On and Off sheets are hardwired collections of different types of center-surround RGC/LGN cells, and the three V1 sheets represent cells of different V1 layers and/or cell types. Connections to V1 neurons adapt via Hebbian learning just as in the simpler GCAL model, allowing initially unselective V1 neurons to exhibit the full range of response types seen experimentally, by differentially weighting each of the subcortical and lateral inputs. Adapted from Bednar (2012).

given, with each known type of ON or OFF receptive field implemented using a convolutional model from figure 14.4*c-d* with a different RF. This processing allows natural image input by reducing large areas of constant activation, as in figure 14.4*d*, and provides V1 with appropriate patterns of input activity that model outputs from various LGN cell types.

A GCAL V1 cell at location (k, l) receives afferent input from the ON or OFF cells near that topographic

location, just as the ON or OFF cells receive input from the photoreceptors. I.e., retinotopy is built in; GCAL simulates only the activity-dependent local reorganization in connectivity after the gradient-based processing described in the previous section has completed. In addition to the topographically mapped afferent connections, the V1 cells also receive lateral excitatory and lateral inhibitory input from their neighbors. Activity at a given time is computed from each of these sources of input:

$$v_{kl} = \sigma\left(\sum_p \gamma_p \mathbf{x}_{pkl} \mathbf{w}_{pkl}\right). \quad (4)$$

Here $p \in \{A_{on}, A_{off}, E, I\}$ indexes one of the four connection fields for this neuron in the simplest model (from the afferent ON, afferent OFF, lateral excitatory, and lateral inhibitory projections), γ_p is a fixed multiplicative factor scaling that projection, \mathbf{w}_{pkl} is the weight vector in projection p to unit (k, l) , and \mathbf{x}_{pkl} is the topographically corresponding portion of the input vector. For a lateral projection, $\mathbf{x}_{pkl} = \mathbf{v}_{pkl}(t - 1)$, i.e., the activity of the topographically corresponding portion of this sheet in the previous time step.

σ is a half-rectifying function to ensure that activities are positive, with a fixed gain of 1.0 but an adaptive threshold θ that is automatically adjusted to maintain a fixed target activity for each neuron (Stevens et al.,

2013):

$$\theta_{kl}(t) = \theta_{kl}(t - 1) + \lambda(\overline{\eta_{kl}}(t) - \mu) \quad (5)$$

where $\lambda = 0.01$ is the homeostatic adaptation learning rate, $\overline{\eta_{kl}}$ is an average of neuron (k, l) 's recent activity, and $\mu = 0.024$ is a target average firing rate.

In the simplest GCAL model (Stevens et al., 2013), as in nearly all other activity-dependent models of feature map development, the lateral connections have a Mexican-hat configuration (short-range excitation, longer-range inhibition). The result of such lateral interaction is to focus an initially diffuse pattern of activation around one or more hotspots that receive more excitation than inhibition, gradually resulting in spatially discrete “bubbles” of activity in response to a slowly changing input. Similar processes occur in the variants of the model that have more realistic connectivity patterns with long-range excitation, as long as disinaptic inhibition dominates at high contrasts (Bednar, 2012). These interactions are a locally implemented version of the global winner-picking and neighborhood kernel from the SOM, and just as for the SOM, they ensure that the maps achieve good coverage and continuity. Unlike the SOM, there can be many spatially distinct “winning” patches simultaneously (see V1 sheet in figure 14.8), each learning properties from the retinotopically corresponding inputs.

As for a SOM, the weight values \mathbf{w}_{pkl} in each projection (whether afferent or lateral) are initially isotropic or random, developing selective patterns only through Hebbian synaptic plasticity. Specifically, after the postsynaptic activity v_{kl} for unit (k, l) is computed using equation 4, a new weight vector \mathbf{w}'_{pkl} is computed from the old weight vector \mathbf{w}_{pkl} , the presynaptic activity vector \mathbf{x}_{pkl} , the postsynaptic activity v_{kl} , and the learning rate α :

$$\mathbf{w}'_{pkl} = \frac{\mathbf{w}_{pkl} + \alpha v_{kl} \mathbf{x}_{pkl}}{\|\mathbf{w}_{pkl} + \alpha v_{kl} \mathbf{x}_{pkl}\|_1} \quad (6)$$

where $\|y\|_1$ is the L1-norm of y , i.e., the sum of all the weight magnitudes. This divisively normalized rule results in connections that reflect correlations between the presynaptic activity and the postsynaptic response, with competition enforced by the normalization so that weights are bounded while becoming selective.

Note that there is strong biological support for the various elements of the more elaborate version of this model (Bednar, 2012; Miikkulainen et al., 2005), but there is also a large number of other similar models that would also be compatible. For instance, there is little agreement about the particular form of divisive normalization used, but there is general agreement that some mechanism of this type is needed to keep values stable. Similarly, whether excitation and inhibition should combine additively, multiplicatively, or in some

other way is not clear, since the underlying operations of real neurons are a complex function of their detailed morphology and physiology. Rather than trying to incorporate all of these details, GCAL focuses on extracting a simple, plausible mechanism that is sufficient to replicate the observed behaviors and can be elaborated as needed when data is available.

When trained on natural images and with well-chosen values of parameters such as the distances and strength of lateral connectivity, this model develops map patterns that are a close match to those found in animals. Figure 14.10 shows GCAL and LISSOM maps from simulations of each of the spatial-feature dimensions for which maps have been found in V1, along with the corresponding experimental data.¹ All of the maps show local continuity and smoothness, which in the model is due to local lateral connectivity that makes nearby neurons respond to similar patterns. The animal maps also each have unique properties that are reflected in the model results, as described below. For instance, orientation is represented by a smooth map interrupted by pinwheels and other discontinuities. Interestingly,

¹Permissions for figure 14.10: Macaque OR (Blasdel, 1992b) and Macaque OD (Blasdel, 1992a), reproduced with permission of Society for Neuroscience. Ferret DR, reprinted by permission from Macmillan Publishers Ltd: *Nature* (Weliky, Bosking, & Fitzpatrick, 1996), copyright (1996). GCAL OR (Stevens et al., 2013), reproduced with permission of Society for Neuroscience. Owl monkey SF, reprinted with permission from (Xu, Anderson, & Casagrande, 2007), copyright (2007) Wiley-Liss, Inc. Macaque CR (Lu & Roe, 2008), reproduced by permission of Oxford University Press. Cat DY, reprinted by permission from Macmillan Publishers Ltd: *Nature*, (Kara & Boyd, 2009), copyright (2009).

the patterns of discontinuities in both animals and the model follow quantitative predictions from abstract principles of geometric pattern formation, rather than representing incomplete self-organization (Kaschube, Schnabel, Löwel, Coppola, White, & Wolf, 2010; Stevens et al., 2013).

Motion direction is mapped similarly, while ocular dominance forms stripes alternating between left eye (white) and right eye (black). Rather than stripes or pinwheels, spatial frequency preferences occur in patches (black: 0.2 cycles/° to white: 2.4 cycles/°). Color responses are plotted as red outlines highlighting areas of strong response to a red-green isoluminant grating, compared to a monochromatic grating. Such color selectivity occurs in isolated patches, typically aligned with less-selective regions of the orientation selectivity map (darker areas of grayscale background), indicating largely separate coding for color and orientation. Data for horizontal disparity preference, estimated here as the preferred phase difference between sine gratings, is less clear. Even so, in both experimental data and the model, disparity-selective cells tend to occur in small, isolated patches (of which only one is shown for the experimental data due to the technique used). In the model, this patchiness is due to horizontal disparity being detectable only for cells with a vertical orientation preference.

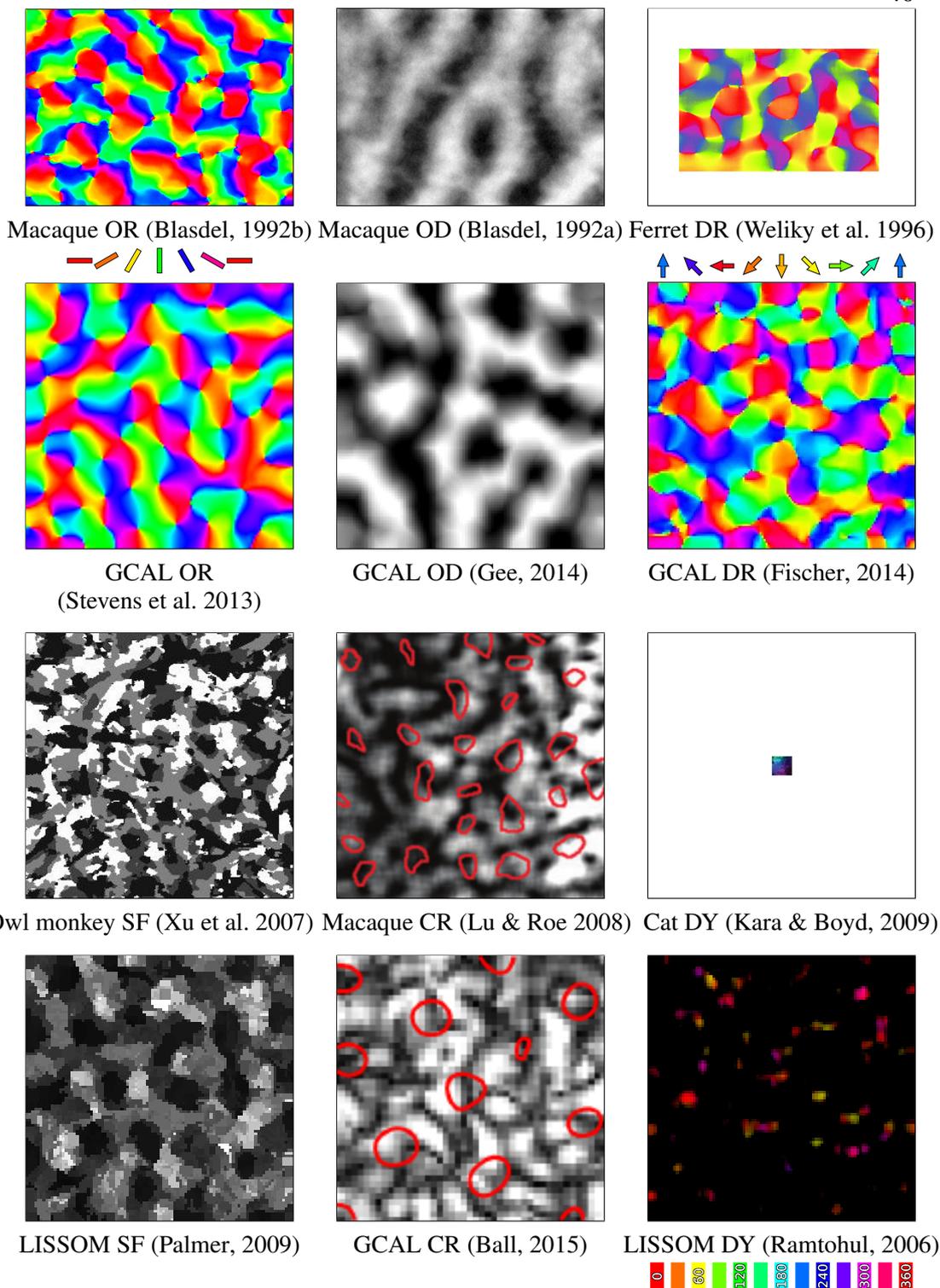


Figure 14.10: **Animal and model V1 maps.** (Color version available in the figure insert.) For V1 in monkeys, ferrets, and tree shrews, maps have been shown for a large number of input features, including: orientation (OR), ocular dominance (OD), motion direction (DR), spatial frequency (SF), color (CR), and disparity (DY). Each of these maps is plotted above in a box representing a 4mm-wide patch of V1, along with the species and citation for the corresponding paper, and a corresponding model plot (below each animal map). See main text for description of each map and for permissions from the sources listed.

Note that each of these maps in principle can consist of the same set of physical neurons, just tested using different input patterns and analysis techniques. The overall result is that neurons are each selective for a small region in a large, multidimensional parameter space (e.g. for a pattern oriented in a specific direction, at a specific location on the retina, with a certain size, color, etc.), and together the neurons cover the space of possible patterns encountered during development (cf: Durbin & Mitchison, 1990). Although each neuron has some level of selectivity for every dimension, selectivities are typically orthogonal, such that a neuron highly selective in one dimension is less selective in others (cf: Yu et al., 2005).

Although not all of the dimensions in figure 14.10 have been tested using a SOM, it is likely that a SOM model could be built to yield similar map patterns, as for many other possible models of this general form. Unlike the SOM or other geometrical models, however, the GCAL model can be considered a functioning visual system, and is thus able to be evaluated for visual illusions, aftereffects, transient responses, surround modulation, and in general tested with any visual pattern or experimental paradigm, as in the next section.

Contextual modulation in neural maps

The preceding sections have focused on how the feature preference and receptive field properties of neurons in neural maps can be generated. But once neurons have developed selectivity, treating them as a set of disconnected spatial filters (as in figure 14.4c-d) is at best a useful first step. Such neurons also exhibit very many “non-classical” phenomena that involve interactions across large areas of the map. E.g., retina and LGN neurons exhibit contrast gain normalization, such that responses are relative to the prevailing activation in that local area, not simply a function of the neuron’s RF (Bonin, Mante, & Carandini, 2005). Cortical neurons also show such normalization, and exhibit complex changes in receptive field and response properties when stimuli are present in locations surrounding the “classical” receptive field: perceptual completion or suppression effects, illusory contours, “pop-out” of certain stimuli from their background, and a variety of other contextual phenomena (for review see Graham, 2011).

These phenomena are generally considered to involve lateral connections from other neurons in the map and/or feedback from neurons in other regions, along with possible “read-out” mechanisms in downstream neurons receiving input from these areas

(Angelucci & Bressloff, 2006). Many of the effects are specific to certain feature values, such as orientation. Importantly, long-range lateral connections have been shown to reflect the underlying feature maps, tending to connect neurons with similar orientation preferences in V1 (Bosking, Zhang, Schofield, & Fitzpatrick, 1997), which is consistent with a role in these effects.

Most models treat these lateral interactions as isotropic, primarily for simplicity and computational efficiency, which is a valid approximation only at very short ranges. The SOM analysis in terms of folding in previous sections suggests that seemingly complex patchy connectivity that respects feature maps could actually be compact, isotropic connectivity in an underlying multidimensional feature space, which then looks patchy only because the neural architecture is organized in only two dimensions. Interestingly, such connectivity emerges automatically from the mechanistic GCAL model in the previous section, driven by Hebbian learning (equation 6) that strengthens connections between neurons with correlated responses (Miikkulainen et al., 2005). Figure 14.11 shows examples of patchy lateral connectivity patterns like those developed by GCAL models, with neurons connected by similarity of responses to the multidimensional input space, not only by locality in the cortical space. Work with related but more complex

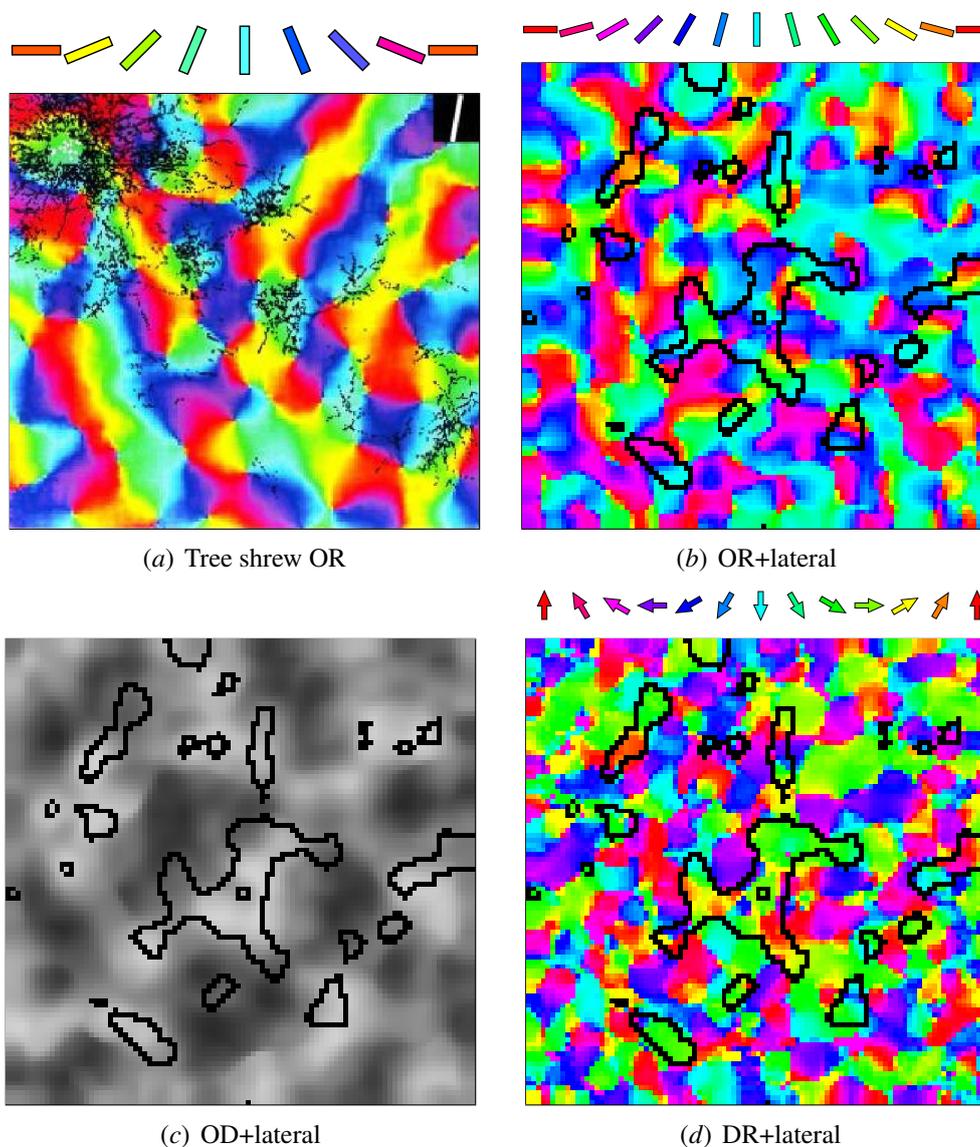


Figure 14.11: **OR, OD, DR lateral connections.** (Color version available in the figure insert.) (a) Reprinted from Bosking et al. (1997) with permission of Society for Neuroscience. Orientation map in 2.5×2 mm tree shrew V1, with black dots indicating areas labeled by a tracer injected in the white-dotted area in the upper-left corner. The patchy connections primarily connect to regions of the map with similar orientation preferences. (b-d) Maps from a joint LISSOM model of orientation, ocular dominance, and direction, each with a contour threshold plot of lateral connection strength for the central neuron (denoted by a black square). The contour outlines enclose the highest-weighted connections to this neuron. This neuron is nearly monocular, and primarily connects to other right-eye-preferring neurons. The same pattern of connectivity can also be seen to respect the direction map, with long-range connections primarily to similar directions, and the orientation map, with long-range connections primarily to similar orientation and direction preferences. These connections reflect correlations between neurons during development, which in turn reflect each of the feature preference values for each neuron. The connections thus respect similarity in a multidimensional feature space, becoming patchy because that space has been projected down onto the 2D cortical surface. Reprinted from Bednar & Miikkulainen (2006), copyright (2006), with permission from Elsevier.

models suggests that correlation-based connections can account for many of the non-classical phenomena listed above (Bednar, 2012), including size tuning, surround modulation, and orientation illusions and aftereffects. Future work will be needed to investigate the role of feedback connections, which could have many of the same effects.

What is the information-processing goal of neurons in maps?

The previous sections outlined how map models could establish an initial topography, how feature map patterns could emerge through folding to represent a multidimensional latent input space, and how these processes could be related to neural mechanisms to provide detailed predictions and tests. The models also suggest explanations for the function of a feature map, such as that it provides good coverage of an underlying feature space (Swindale et al., 2000), that lateral connections help decorrelate the responses, and that contextual modulation and aftereffects are the unavoidable result of these processes (Barlow, 1990; Miikkulainen et al., 2005).

However, in each case the functional aspects of the model are incorporated only indirectly, either feeding into the modeler's design choices or as a *post-facto*

interpretation of the results. For a different class of models that can be considered *normative*, functional criteria are incorporated explicitly into the definition of the model. Doing so allows a hypothesis about function to be evaluated directly, independently of the details of specific neural mechanisms, which are often not known.

An influential series of normative models starts from an assumption that the computational goal of a sensory system is to recover information about the underlying structure of the world, presumably in order to guide actions relating to that structure. Given raw sensory input, the objective of these models would be to *infer* a set of external causes that would explain the observed data (Bell & Sejnowski, 1997; Olshausen & Field, 1996).

The real causes of an image would comprise its constituent set of external 3D objects, surfaces, and lighting sources. However, determining these is a highly complex and non-linear inference process, and full reconstruction is not necessary for most of the tasks that an animal needs to do. Plus, many features of the world can only be discovered through interaction and thus require loops involving the motor system, placing them outside of the present account. A much more tractable goal, which we will focus on in the next two sections because it is locally realizable in V1, is to find a linear combination of *basis vectors* that can generate a small

patch of an image.

Basis vectors for generating images

A *generative* model is one whose activity pattern can be used directly to generate an input patten. Starting with the idea of generating a small image patch is appropriate if the real causes of the scene can be estimated hierarchically, with V1 representations accounted for by causes at a higher level, as suggested by the multiple visuotopic maps in areas along the visual pathway. In this section we will first explain the general principles behind these models using simple examples, and then the following section will consider a specific normative generative model applied to explain neural maps and RFs in V1.

A basis set for an image patch \boldsymbol{x} is a set of basis vectors \boldsymbol{v}_i such that \boldsymbol{x} can be approximated as a sum of the \boldsymbol{v}_i weighted with scalar coefficients c_i :

$$\boldsymbol{x} \approx \sum_i c_i \boldsymbol{v}_i. \quad (7)$$

This model is represented graphically in figure 14.12. The coefficients c_i can be considered estimates of latent (or hidden) variables, specifying how much of each corresponding basis vector \boldsymbol{v}_i is present in this image patch.

The generative approach allows explicit expression of

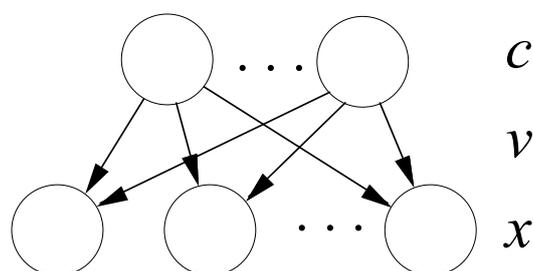


Figure 14.12: **One-hidden-layer generative model.** Here the inferred data vector x is obtained as a linear combination of the basis vectors v , with coefficients c . The arrowheads indicate that the basis vector acts in the “reverse” direction compared to the weights in the models previously discussed (e.g., figure 14.4), specifying the weights that could be used to reconstruct the input image, not the weights by which the coefficients c are calculated given an input image. I.e., the values c have to be inferred approximately, when given an actual set of inputs, rather than calculated directly as in the mechanistic models from previous sections. Once the values c have been estimated, they provide a compact representation for the input image, and can then be used directly for further processing. The actual input need not ever be reconstructed from c during sensory processing, as long as c provides a representation that has captured the important properties of the image patch well enough that it *could* be reconstructed with low error.

an *objective function*, where a representation is iteratively optimized to minimize or maximize a mathematically specified criterion (which is what makes these models normative). For instance, the objective function might specify a low reconstruction error (to ensure that the representation is faithful), low levels of neural activity (to minimize metabolic cost), few neurons active (a *sparse* (mostly zero) representation that could simplify downstream processing), etc., or weighted combinations of any of these. Various automatic (though typically non-neural) techniques can then optimize the representation to achieve the specified objective, approximating the result of processes of

evolution and development in animals. Note that specifying that reconstruction error should be low does not necessarily mean assuming that the visual system actually does reconstruct the input in practice – it simply specifies a criterion that the map's representation of the input should be rich enough to allow that in some form.

Perhaps the simplest generative model would be to set $c_j = 1$ for one particular index j , and set all of the other coefficients $c_i = 0$ (a winner-take-all approach achieving maximum sparsity in the cortical representation). If a suitable objective function were defined, such a model could do *template matching*, where the input is represented in terms of the closest-matching basis vector. For instance, each basis vector could be determined using the *k-means* clustering algorithm (Hartigan, 1975). Given a large set of image patches, *k-means* provides k vectors that are each averages of a cluster of image patches, with the clusters chosen iteratively to maximize within-cluster similarity. A novel input would then be mapped into the nearest cluster, as previously described as a limiting case of the SOM algorithm. Interestingly, such clustering can yield basis vectors similar to receptive fields found in V1 (Barrow, 1987; Coates, Ng, & Lee, 2011). However, a single chosen template (akin to a single V1 neuron firing) is unlikely to be a good match to an arbitrary patch of a natural image, and thus this would be a poor

generative model, with high reconstruction error.

A much richer representation is obtained by allowing multiple c_i coefficients to be nonzero at the same time, forming a *distributed representation* or *componential representation* of the input. I.e., one can select multiple coefficients that *together* could reconstruct the input, echoing how GCAL has multiple “activity bubbles” in response to an image, instead of SOM’s single winning neuron.

Figure 14.13 shows an example of how an image patch can be reconstructed using basis vectors developed by the Olshausen & Field (1996) Sparse Coding model of V1, to be discussed in more detail below. For now, we will just treat the basis set as given, and look at how a 14×14 -pixel patch of an image can be reconstructed with relatively low error given a small number of basis vectors. Because natural images are only a limited subset of all possible patterns of pixels (forming a lower-dimensional manifold, as discussed in the context of dimension reduction above), only a small number of well-chosen basis vectors are needed to represent most natural image patches with high accuracy. For this example, the image patch is easily recognizable after only four basis vectors ($k = 1 \dots 4$) have been summed, and barely distinguishable after 16, which means this specific input can be represented faithfully using 4 or 16 activation values c_i at the output

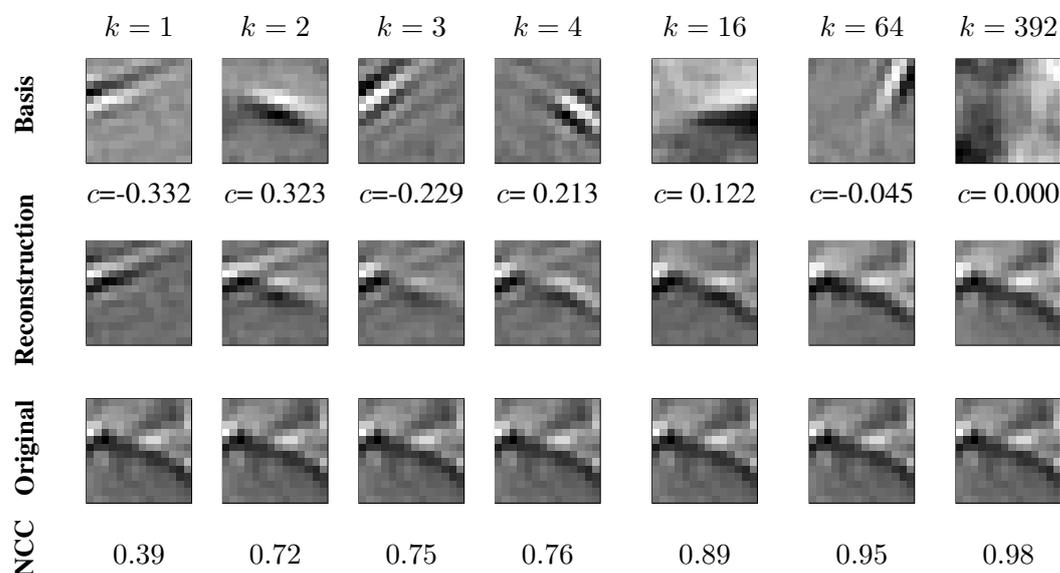


Figure 14.13: **Reconstruction from basis vectors.** The image patch that has been duplicated in each column of the row labeled **Original** can be reconstructed as a weighted linear combination of 392 basis vectors, some of which are shown in row **Basis** with their corresponding coefficient c . The basis vectors were developed by the Olshausen & Field (1996) model discussed below. The result is the reconstructed image patch in the column labeled $k = 392$ of the row **Reconstruction**, which is nearly indistinguishable from the original. Using a single basis vector (column $k = 1$) amounts to template matching; even the best-matching single template is still a poor representation. But most of the coefficients are small, and using e.g., the 16 basis vectors with the largest coefficients suffices for a high-quality reconstruction (high normalized cross correlation, row **NCC**, and low visible difference, even if treating the other coefficients as being zero). The coefficients of those 16 vectors act as a compact and sparse but faithful representation of the original input patch.

layer, rather than the 196 pixel values on the input layer.

In the context of image compression algorithms like JPEG and wavelet coding, achieving a good-quality reconstruction from a small number of coefficients (a *sparse* representation) is the final goal. In the visual system, of course, it is unlikely that images will actually ever need to be reconstructed, but requiring good reconstruction is a practical way of objectively ensuring

that the algorithms have preserved the information in the image. In practice, the idea is that the sparse coefficients c_i themselves, rather than the reconstruction, will be an effective representation suitable for driving action or further computation.

Olshausen and Field model of V1 RFs

To form a full generative model, the basis-function approach can be combined with (1) an objective function that (in part) minimizes reconstruction error, (2) procedures for inferring the coefficients c_i in response to each image, and (3) procedures for developing the basis vectors \mathbf{v}_i for a given image dataset. For instance, the influential Olshausen & Field (1996) model defines an objective function as

$$\|\mathbf{x} - \sum_i c_i \mathbf{v}_i\|^2 + \lambda \sum_i g(c_i) \quad (8)$$

with the first term expressing the reconstruction error (zero for perfect reconstruction), while in the second term $g(c_i)$ is a penalty function corresponding to the negative log sparse prior on c_i , and λ is a trade-off parameter determining the balance between reconstruction error and coefficient sparsity. Eq. 8 is (up to constant terms) the negative log of the posterior probability $p(\mathbf{c}|\mathbf{x})$ of the generative model. Here sparseness means that the prior distribution on the c_i

favors values close to zero, reflecting a belief that “natural images may generally be described in terms of a small number of structural primitives” (Olshausen & Field, 1997). To allow a highly sparse representation, the model uses an *overcomplete basis set*, with more basis vectors than are required for a perfect linear reconstruction.

Given a set of basis vectors and an input image patch, the model must carry out *probabilistic inference* to determine the coefficients $\{\hat{c}_i\}$ that best explain the image, rather than simply computing the map activity directly as in the mechanistic models described in previous sections. The basis vectors associated with these coefficients can be interpreted as estimates of the *latent causes* or *latent variables* in the external world that led to the formation of that image. The basis vectors are learned by an iterative Expectation-Maximization style algorithm (Dayan & Abbott, 2001), where one alternately carries out inference to obtain \hat{c} , the set of c 's that minimize equation 8, then adjusts the basis vectors to minimize $|\mathbf{x} - \sum_i \hat{c}_i \mathbf{v}_i|^2$, eventually settling on a stable set of basis vectors for a given set of training images.

As an alternative to this general optimization technique, Olshausen & Field (1997) also showed how one can design a locally implementable gradient descent scheme on the objective of equation 8 to achieve this

computation. Interestingly, this physically realizable implementation shares important features in common with the mechanistic models like GCAL, suggesting that it may be possible to relate the two to show how a specific goal can be implemented in terms of neural mechanisms. For instance, the physical implementation involves *lateral interactions* between the c -units, somewhat like the correlation-based connections in GCAL, where the strength of the interaction between units c_i and c_j depends on the similarity of their weight vectors $\mathbf{v}_i \cdot \mathbf{v}_j$. This dependence is because two units with similar basis vectors that both could have contributed to the observed image structure need to compete if the sparsity term is to be minimized. Zylberberg, Murphy, & DeWeese (2011) have shown how this method can be implemented (approximately) in a network of spiking neurons. Thus mechanistic interpretations of these ideas do exist, though nearly all normative and generative models are implemented and presented in more abstract, non-mechanistic form.

In any case, once the model has developed such a set of basis vectors, they can be compared to data from animals, to see if this approach can explain anything about the structure and properties of V1 neurons. Because the basis vectors are used as a *projective field* from V1 back onto the LGN in the model, not a *receptive field* computing V1 activations from LGN

activity, they are not directly comparable to V1 RFs. Even so, the model can be treated as a “black box” as in animal experiments, presenting patterns on the input and recording the responses, allowing an RF pattern to be estimated by reverse correlation for each input. The resulting patterns are similar to the projective fields, and so the projective fields (basis vectors) are typically compared as-is to V1 RFs (Olshausen & Field, 1997). In what follows, we will use the basis vector as an approximation to that neuron’s RF, because directly estimated RFs are not provided for most models of this type.

Figure 14.14 shows the basis vectors that resulted from iteratively maximizing the Olshausen & Field (1996) objective function for a set of patches of whitened natural images, comparing them to results from monkey V1. The results resemble the receptive fields shaped like Gabor functions found in V1 in mammals, leading to a hypothesis that primary visual cortex neurons are wired to form a basis set that provides good coverage of the space of natural image patches using a small number of active units at any one time (sparse coding; Olshausen & Field, 1996).

Later work has shown that there are very many other ways to get Gabor-shaped weight matrices by training on natural image patches. For instance, models based on independent components analysis (ICA) also lead to

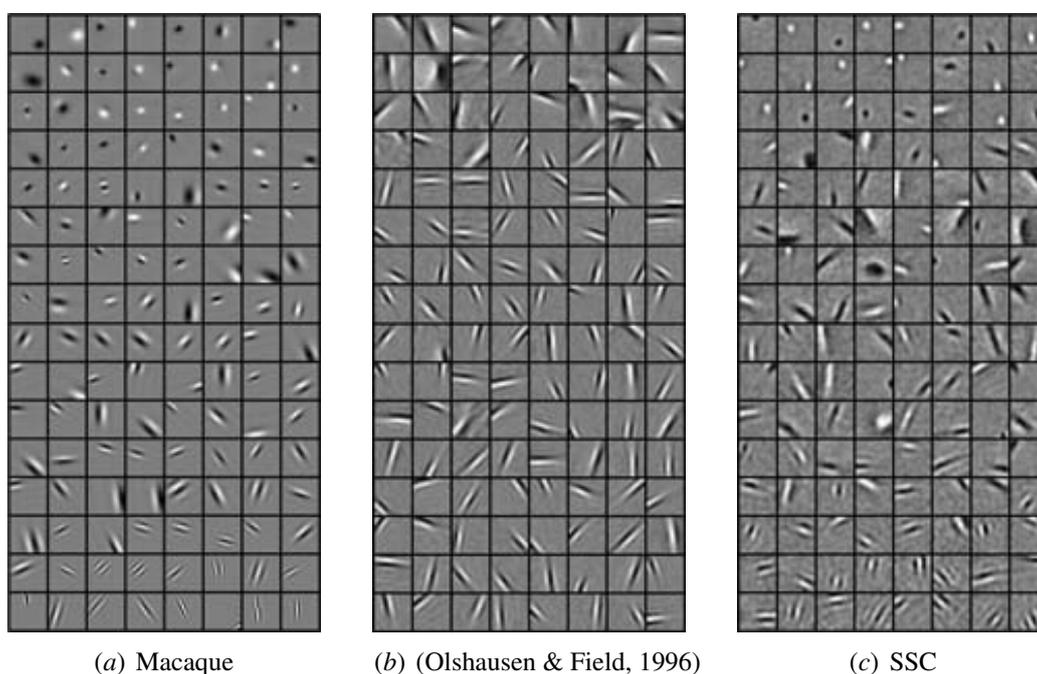


Figure 14.14: Macaque and model V1 RFs. (a) Macaque monkey receptive fields in V1 are typically shaped like Gabor functions (a sine grating masked by a circular Gaussian; Jones & Palmer, 1987), with various RF shapes amounting to different values of the Gabor's frequency, phase, and Gaussian-size parameters. These RFs cover a range of selectivities, orientations, spatial frequencies, and spatial locations. Gabor fits using data from Ringach (2002), adjusted for eccentricity and sorted by RF shape. (Raw data not published, but the authors found that the Gabor fits had low error.) (b) Basis vectors from the Olshausen & Field (1996) sparse coding model are also Gabors, although the range of sizes is more limited, with more multi-lobed patterns. (c) A model with a slightly modified sparseness constraint (SSC) can reproduce the range of Gabor-shaped RFs found in macaque, when the basis vector is taken as an approximation to the RF (Rehn & Sommer, 2007). Reprinted with kind permission from *J. Computational Neuroscience*, Rehn & Sommer (2007), copyright (2007) Springer Science and Business Media.

Gabor-shaped weights (Bell & Sejnowski, 1997). ICA is nearly identical mathematically to the Olshausen & Field (1996) Sparse Coding model, but with a simplifying restriction that the number of input and output units are the same; a complete rather than

overcomplete basis set. In this case, the basis vector weights are adjusted assuming that the underlying causes are statistically independent, hence the name. With this assumption, the inference step can be replaced with a simple feedforward operation, and the receptive fields can be calculated directly from the projective fields by matrix inversion.

However, each of these algorithms results in a relatively narrow range of RF shapes, missing many of the RF types commonly observed in animals (compare figure 14.14*b* with figure 14.14*a*). More recent work has shown how to account for the full range of monkey RF shapes, by using an alternative definition of sparseness (Rehn & Sommer, 2007; figure 14.14*c*) or by very different approaches altogether (e.g., the undirected “energy based” model from Osindero, Welling, & Hinton, 2006).

Note that nearly all of the work using normative models for neural maps has focused on a “bag of neurons” approach with no topography. Some such models have addressed topography as well, notably Topographic ICA (Hyvärinen & Hoyer, 2001; Hyvärinen, Hurri, & Hoyer, 2009), the “energy based” model from Osindero et al., 2006, and a variant of ICA expressed in terms of mutual information (Kozloski, Cecchi, Peck, & Rao, 2007; Linsker, 1989). Each of these models adds some form of lateral

interaction across a map surface, as in SOM and GCAL, causing neurons to become more similar to neighbors. However, so far the normative-model maps have only been simulated at a very coarse spatial scale, leaving it unclear whether the map shapes correspond to those seen in animal maps as has been shown for the geometrical and mechanistic models. Moreover, the required modifications to generate such maps are motivated primarily in terms of matching experimental data or in mechanistic terms, rather than showing how the topography satisfies a normative goal. Thus normative approaches at present can account primarily for receptive field shapes and their diversity for neurons in neural maps, rather than for topographic and feature maps, perhaps because there is no functional role for the topography (as discussed further below).

Non-visual maps

The preceding sections have covered a set of related topics from the early visual system in depth. The visual system is used as a model for map development for many reasons, such as having easily controllable stimuli, a receptor surface with a relatively simple and well understood organization, a small number of synapses between the receptors and the first cortical map, and a cortical map that is easily accessible for imaging. The

hope of much of this work is that similar principles will apply to other non-visual modalities. For instance, there are numerous SOM-based models of other modalities (e.g., somatosensory cortex, Ritter et al., 1991), GCAL/LISSOM-based models of whisker-barrel maps (Wilson, Law, Mitchinson, Prescott, & Bednar, 2010), and sparse-coding or ICA-based models of a wide range of sensory neuron types. There is also corresponding data for maps and plasticity in the motor cortex (reviewed in Harrison & Murphy, 2014) as well as similar self-organizing map modelling approaches for motor cortex (Morasso & Sanguineti, 1995).

As already shown for visual maps, different aspects of the stimuli will lead to very different map shapes and properties (such as ocular dominance stripes versus isolated color blobs). Similarly, different organizations are expected for each modality. For instance, in auditory cortex there is a primary organization by tonotopy (matching the arrangement of receptors on the cochlea), but it is not yet well established which other dimensions of auditory stimuli might be represented. And although modelling has replicated the basic organization of somatosensory cortex, how the various touch, temperature, and pain receptors distributed across the skin surface affect the organizations in somatosensory cortex is not well understood. Similar work in auditory cortex is constrained by the huge number of auditory

subcortical regions before cortical processing. Modelling of map development in these other sensory areas will require significant new experimental data on the subcortical properties and how they connect to the cortex, at which point the methods shown here can be applied and tested. Given the relatively little work done so far on applying such models to the motor system, it remains an open question how much of the same methods will apply, and again suitable experimental data is probably the constraining factor.

Concluding discussion

The models covered in this chapter help explain how topographic mappings are established, how and why feature map patterns develop, how the context of neurons in feature maps affects their function, and what aspects of the information present in images are represented in the activity of neurons in maps. The models span a wide range of approaches, from abstract models focused on uncovering geometrical relationships, to mechanistic models expressed in terms of neural mechanisms, to normative models expressed in terms of their information processing goals (refer to Box 14.1).

Ideally, explanations of neural maps would address each of these levels at once, showing how the geometric

patterns of maps and receptive fields emerge from biophysical mechanisms that satisfy clear information-processing goals. In practice, there are still significant gaps between the models of each type, and so no type has subsumed the others. For instance, there is relatively little data about the visual experience of developing animals, which makes it difficult to build fully mechanistic models of feature map development based on realistic stimuli, and so the more abstract geometrical models remain very relevant for these cases. But in principle there is no serious conflict between the geometrical and mechanistic map models, with the Elastic Net, SOM, and GCAL all developing very similar types of feature map patterns through largely compatible mechanisms.

More interesting and problematic are the relationships between mechanistic and normative models, and between normative models and the neurobiological data. Normative models are primarily formulated at a level that is far removed from neural mechanisms, which achieves their purpose of revealing functional aspects but makes it more difficult to relate them to results from neurobiological experiments. As described above, some progress has been made on building mechanistic interpretations of normative models, with many of the required elements having a clear basis in known mechanisms. However, significant

gaps still exist between all of the current normative models and what we know of neural architecture, which can be addressed in future modelling studies, as described below.

First, due in part to their computational complexity, none of the normative topographic map models have been run with a cortical area large enough to demonstrate a close match to the topographical properties of animal feature maps (other than local smoothness), such as regular periodicity, hierarchical relationships between multiple maps, and various types of systematic topological defects (Blasdel, 1992b). For geometrical and mechanistic approaches, previous models with superficial similarity to the animal data (including smoothness, periodicity, and orientation pinwheels) have later been found to be incompatible with the experimental map patterns (Swindale, 1996), while the geometrical and mechanistic models discussed here have been found to match these properties in detail (Carreira-Perpiñán et al., 2005; Stevens et al., 2013; Yu et al., 2005). Testing normative models in this way as well could help determine whether the mechanisms involved are compatible with those of the other types, and these tests should soon become computationally feasible.

Second, the existing normative topographic map algorithms (including at least Hyvärinen & Hoyer,

2001; Kozloski et al., 2007; Osindero et al., 2006) rely on squaring or full-wave rectification of negative firing rates at some step of the calculation. This step allows these algorithms to avoid grouping by spatial phase so that they can develop complex cells by local pooling, but is difficult to reconcile with physical neural responses based on actual (positive) firing rates.

Third, the process of “development” in all of the normative topographic map models (and also in most of the non-topographic normative approaches) relies on non-neural optimization algorithms that require information distributed throughout the model network, rather than synaptically or neurally local learning rules that have a clear biological implementation. They typically are temporally non-local as well, with results that use batch learning to achieve smooth basis vector patterns (e.g., in Olshausen & Field, 1996), and on temporally separate inference and learning phases with no clear biological interpretation as a continuous system, even in the most mechanistic implementations of sparse coding built so far (Zylberberg et al., 2011). These conflicts between the normative and mechanistic approaches are worth investigating in future studies, because they may reveal either previously unknown neural mechanisms or invalid assumptions about the goals of sensory processing. They could also represent implicit constraints provided by evolution, rather than

actual neural mechanisms, which would also be important to establish. For now, all three types of explanation (geometrical, mechanistic, and normative models) need to be considered together to appreciate the range of phenomena that can be addressed by neural map models.

From a larger perspective, while the ubiquity of smooth, topographic maps suggests that they could be important, it has proven surprisingly difficult to demonstrate any functional advantage to having neurons organized spatially in maps. Conceptually, if one starts with a smooth, topographic feature map and randomly perturbs the position of each cortical neuron without changing the underlying connectivity pattern, then the pattern of connectivity, and hence much of the function, will remain unchanged. However, the lengths (and presumably latencies) of the connections *will* change, however, leading to a hypothesis that maps exist to optimize the total wiring length in the brain (which dates back at least to the seminal work of Santiago Ramón y Cajal). Indeed, a smooth map appears to be at least a local optimum in wiring length, given the observed patterns of connectivity (Chklovskii & Koulakov, 2004).

A deeper issue is why the observed patterns of connectivity, i.e., 2D maps with primarily local connections but patchy long-range, feature-specific

connections as well, occur in the first place. Although any answer to this question must be speculative, perhaps a 2D cortex makes it easier to connect local neurons (using wiring in the third dimension), while not being much worse than the highest-dimensional physical arrangement possible (3D), in comparison to the large, multidimensional space of variation in the external world. In turn, smoothness in the map presumably reflects the fact that properties that can be sensed (objects, sounds, odors, etc.) in the external world tend to be localized in space (as suggested by Tobler's First Law of geography, at the start of this chapter). With a smooth, topographic organization, sensory neurons with local RFs can then process meaningfully distinct underlying (latent) causes, and interact with neurons somewhat more distant to improve discrimination and detection. In this context, smoothness could make the process of development more robust, allowing a large margin of error in axon guidance without affecting the results. As suggested for dimension-reduction models above, the observed patchiness in connectivity would then reflect a map that has extracted an underlying space with a dimensionality greater than 2D.

Finally, model maps have been considered largely in isolation here, and primarily in terms of feedforward processing, but real maps are situated in the context of numerous other maps that are simultaneously involved

in neural processing, including widespread feedforward and feedback connections between them (Van Essen & Gallant, 1994). Numerous models of feedback and attentional processes incorporate static (non-developmental) neural maps, and can help us understand how the maps relate to systems-level phenomena and eventually to behavior. For instance, Rao & Ballard (1999) and Hyvärinen et al. (2009, chap. 14) have investigated the effect of feedback from higher regions on the activations of lower-level units. In their models higher-level units may identify longer-range structure, and feed back information about this to lower-level units, e.g., to model contour integration. Understanding how lateral and feedback interactions affect representations in neural maps is an important topic for current research, because it can help bridge from the map level all the way to behavior.

Summary

- Neural maps are ubiquitous but diverse
- Many maps are topographic
- Topography is established under genetic control with activity-based refinement
- Maps for other features appear to develop due to neural activity

- Feature map patterns appear to be the result of a transformation folding a multidimensional input space onto a 2D cortical space
- Mechanistic implementation of such dimensionality reduction can explain contextual modulation, lateral connectivity, and aftereffects
- Normative models suggest that neurons are uncovering latent causes in sensory inputs, forming a sparse representation suitable for further computation and for action
- Determining relationships between geometrical, mechanistic, and normative models of maps will be important for building complete explanations for neural maps, but is only beginning

References

- Anderson, J. A., & Rosenfeld, E. (Eds.) (1988). *Neurocomputing: Foundations of Research*. Cambridge, MA: MIT Press.
- Angelucci, A., & Bressloff, P. C. (2006). Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons. *Progress in Brain Research*, *154*, 93–120.
- Ball, C. E. (2015). *Modeling the Emergence of Perceptual Color Space in the Primary Visual Cortex*. Doctoral Dissertation, School of Informatics, The University of Edinburgh, UK. Viva held 15 January.
- Barlow, H. B. (1990). A theory about the functional role and synaptic mechanism of visual after-effects. In Blakemore, C. (Ed.), *Vision: Coding and Efficiency* (pp. 363–375). Cambridge, UK: Cambridge University Press.
- Barrow, H. G. (1987). Learning receptive fields. In *Proceedings of the IEEE First International Conference on Neural Networks* (San Diego, CA) (Vol. IV, pp. 115–121). Piscataway, NJ: IEEE.
- Barrow, H. G., & Bray, A. J. (1992). An adaptive neural model of early visual processing. In Aleksander, I., & Taylor, J. G. (Eds.), *Artificial Neural Networks, 2: Proceedings of the 1992 International Conference on Artificial Neural Networks* (pp. 881–884). Amsterdam: North-Holland.
- Bednar, J. A. (2012). Building a mechanistic model of the development and function of the primary visual cortex. *Journal of Physiology (Paris)*, *106*, 194–211.
- Bednar, J. A., & Miikkulainen, R. (2006). Joint maps for orientation, eye, and direction preference in a self-organizing model of V1. *Neurocomputing*, *69* (10–12), 1272–1276.
- Bell, A. J., & Sejnowski, T. J. (1997). The “independent components” of natural scenes are edge filters. *Vision Research*, *37*, 3327.
- Benjumeda, I., Escalante, A., Law, C., Morales, D., Chauvin, G., Muca, G., Coca, Y., Marquez, J., Lopez-Bendito, G., Kania, A., Martinez, L., & Herrera, E. (2013). Uncoupling of EphA/ephrinA signaling and spontaneous activity in neural circuit wiring. *The Journal of Neuroscience*, *33*, 18208–18218.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Berlin: Springer.
- Blasdel, G. G. (1992a). Differential imaging of ocular dominance columns and orientation selectivity in monkey striate cortex. *The Journal of Neuroscience*, *12*, 3115–3138.
- Blasdel, G. G. (1992b). Orientation selectivity, preference, and continuity in monkey striate cortex. *The Journal of Neuroscience*, *12*, 3139–3161.

- Bonin, V., Mante, V., & Carandini, M. (2005). The suppressive field of neurons in lateral geniculate nucleus. *Journal of Neuroscience*, *25*, 10844–10856.
- Bosking, W. H., Crowley, J. C., & Fitzpatrick, D. (2002). Spatial coding of position and orientation in primary visual cortex. *Nature Neuroscience*, *5* (9), 874–882.
- Bosking, W. H., Zhang, Y., Schofield, B. R., & Fitzpatrick, D. (1997). Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *The Journal of Neuroscience*, *17* (6), 2112–2127.
- Burger, T., & Lang, E. W. (1999). An incremental Hebbian learning model of the primary visual cortex with lateral plasticity and real input patterns. *Zeitschrift für Naturforschung C — A Journal of Biosciences*, *54*, 128–140.
- Carreira-Perpiñán, M. Á., & Goodhill, G. J. (2002). Are visual cortex maps optimized for coverage?. *Neural Computation*, *14* (7), 1545–1560.
- Carreira-Perpiñán, M. A., Lister, R. J., & Goodhill, G. J. (2005). A computational model for the development of multiple maps in primary visual cortex. *Cerebral Cortex*, *15* (8), 1222–1233.
- Chklovskii, D. B., & Koulakov, A. A. (2004). Maps in the brain: What can we learn from them?. *Annual Review of Neuroscience*, *27*, 369–392.
- Coates, A., Ng, A., & Lee, H. (2011). An analysis of single-layer networks in unsupervised feature learning. *Journal of Machine Learning Research, Workshop and Conference Proceedings: AISTATS 2011*, *15*, 215–223.
- Dayan, P., & Abbott, L. F. (2001). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. Cambridge, MA: MIT Press.
- Durbin, R., & Mitchison, G. (1990). A dimension reduction framework for understanding cortical maps. *Nature*, *343*, 644–647.
- Erwin, E., Obermayer, K., & Schulten, K. J. (1992). Self-organizing maps: Ordering, convergence properties and energy functions. *Biological Cybernetics*, *67*, 47–55.
- Fischer, T. (2014). *Model of All Known Spatial Maps in Primary Visual Cortex*. Master's thesis, The University of Edinburgh, UK.
- Flanagan, J. G. (2006). Neural map specification by gradients. *Current Opinion in Neurobiology*, *16*, 1–8.
- Gee, A. (2014). *Neural Modelling of Congenitally Abnormal Visual Pathways*. Master's thesis, The University of Edinburgh, UK.
- Goodhill, G. J. (2007). Contributions of theoretical modeling to the understanding of neural map development. *Neuron*, *56* (2), 301–311.
- Goodhill, G. J., & Willshaw, D. J. (1990). Application of the elastic net algorithm to the formation of ocular dominance stripes. *Network: Computation in Neural Systems*, *1* (1), 41–59.
- Graham, N. V. (2011). Beyond multiple pattern analyzers modeled as linear filters (as classical V1 simple cells): Useful additions of the last 25 years.

- Vision Research*, 51 (13), 1397–1430.
- Grossberg, S. (1976). On the development of feature detectors in the visual cortex with applications to learning and reaction-diffusion systems. *Biological Cybernetics*, 21, 145–159.
- Harrison, T. C., & Murphy, T. H. (2014). Motor maps and the cortical control of movement. *Current Opinion in Neurobiology*, 24 (1), 88–94.
- Hartigan, J. A. (1975). *Clustering Algorithms*. Hoboken, NJ: Wiley.
- Hastie, T., & Stuetzle, W. (1989). Principal curves. *Journal of the American Statistical Association*, 84, 502–516.
- Hjorth, J. J. J., Sterratt, D. C., Cutts, C. S., Willshaw, D. J., & Eglen, S. J. (2014). Quantitative assessment of computational models for retinotopic map formation. *arXiv:1408.6132*.
- Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160, 106–154.
- Huberman, A. D., Feller, M. B., & Chapman, B. (2008). Mechanisms underlying development of visual maps and receptive fields. *Annual Review of Neuroscience*, 31, 479–509.
- Hyvärinen, A., & Hoyer, P. O. (2001). A two-layer sparse coding model learns simple and complex cell receptive fields and topography from natural images. *Vision Research*, 41 (18), 2413–2423.
- Hyvärinen, A., Hurri, J., & Hoyer, P. O. (2009). *Natural Image Statistics: A Probabilistic Approach to Early Computational Vision*. Berlin: Springer.
- Jones, J. P., & Palmer, L. A. (1987). The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58 (6), 1187–1211.
- Kaas, J. H. (1991). Plasticity of sensory and motor maps in adult animals. *Annual Review of Neuroscience*, 14, 137–167.
- Kara, P., & Boyd, J. D. (2009). A micro-architecture for binocular disparity and ocular dominance in visual cortex. *Nature*, 458 (7238), 627–631.
- Kaschube, M., Schnabel, M., Löwel, S., Coppola, D. M., White, L. E., & Wolf, F. (2010). Universality in the evolution of orientation columns in the visual cortex. *Science*, 330 (6007), 1113–1116.
- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43, 59–69.
- Kozloski, J., Cecchi, G., Peck, C., & Rao, A. R. (2007). Topographic infomax in a neural multigrid. In Liu, D. e. a. (Ed.), *Advances in Neural Networks – ISSN 2007* (Lecture Notes in Computer Science 4492, pp. 500–509). Berlin: Springer.
- Law, J. S. (2009). *Modeling the Development of Organization for Orientation Preference in Primary Visual Cortex*. Doctoral Dissertation, School of

- Informatics, The University of Edinburgh, UK.
- LeVay, S., Hubel, D. H., & Wiesel, T. N. (1975). The pattern of ocular dominance columns in macaque visual cortex revealed by a reduced silver stain. *Journal of Comparative Neurology*, *159* (4), 559–576.
- Linsker, R. (1986). From basic network principles to neural architecture: Emergence of orientation columns. *Proceedings of the National Academy of Sciences, USA*, *83*, 8779–8783.
- Linsker, R. (1989). How to generate ordered maps by maximizing the mutual information between input and output signals. *Neural Computation*, *1* (3), 402–411.
- Lu, H. D., & Roe, A. W. (2008). Functional organization of color domains in V1 and V2 of macaque monkey revealed by optical imaging. *Cerebral Cortex*, *18* (3), 516–533.
- Miikkulainen, R., Bednar, J. A., Choe, Y., & Sirosh, J. (2005). *Computational Maps in the Visual Cortex*. Berlin: Springer.
- Miller, K. D. (1994). A model for the development of simple cell receptive fields and the ordered arrangement of orientation columns through activity-dependent competition between ON- and OFF-center inputs. *The Journal of Neuroscience*, *14*, 409–441.
- Morasso, P., & Sanguineti, V. (1995). Self-organizing body schema for motor planning. *Journal of Motor Behavior*, *27*, 52–66.
- Mountcastle, V. B. (1957). Modality and topographic properties of single neurons of cat's somatic sensory cortex. *Journal of Neurophysiology*, *20* (4), 408–434.
- Obermayer, K., Ritter, H., & Schulten, K. J. (1990). A principle for the formation of the spatial structure of cortical feature maps. *Proceedings of the National Academy of Sciences, USA*, *87*, 8345–8349.
- Ohki, K., Chung, S., Ch'ng, Y. H., Kara, P., & Reid, R. C. (2005). Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature*, *433* (7026), 597–603.
- Ohki, K., Chung, S., Kara, P., Hubener, M., Bonhoeffer, T., & Reid, R. C. (2006). Highly ordered arrangement of single neurons in orientation pinwheels. *Nature*, *442* (7105), 925–928.
- Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*, 607–609.
- Olshausen, B. A., & Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1?. *Vision Research*, *37*, 3311–3325.
- Osindero, S., Welling, M., & Hinton, G. E. (2006). Topographic product models applied to natural scene statistics. *Neural Computation*, *18* (2), 381–414.
- Overton, K. J., & Arbib, M. A. (1982). The extended branch-arrow model of the

- formation of retino-tectal connections. *Biological Cybernetics*, 45 (3), 157–175.
- Palmer, C. M. (2009). *Topographic and Laminar Models for the Development and Organisation of Spatial Frequency and Orientation in V1*. Doctoral Dissertation, School of Informatics, The University of Edinburgh, UK.
- Prestige, M. C., & Willshaw, D. J. (1975). On a role for competition in the formation of patterned neural connexions. *Proceedings of the Royal Society of London Series B*, 190 (1098), 77–98.
- Ramtohl, T. (2006). *A Self-Organizing Model of Disparity Maps in the Primary Visual Cortex*. Master's thesis, The University of Edinburgh, Scotland, UK.
- Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2, 79–87.
- Recanzone, G. H., Merzenich, M. M., & Dinse, H. R. (1992). Expansion of cortical representation of a specific skin field in primary somatosensory cortex by intracortical microstimulation. *Cerebral Cortex*, 2, 181–196.
- Rehn, M., & Sommer, F. T. (2007). A network that uses few active neurones to code visual input predicts the diverse shapes of cortical receptive fields. *Journal of Computational Neuroscience*, 22 (2), 135–146.
- Ringach, D. L. (2002). Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *Journal of Neurophysiology*, 88 (1), 455–463.
- Ritter, H., Obermayer, K., Schulten, K. J., & Rubner, J. (1991). Self-organizing maps and adaptive filters. In *Models of Neural Networks* (pp. 281–306). Berlin: Springer.
- Rodieck, R. W. (1965). Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Research*, 5 (12), 583–601.
- Sirosh, J., & Miikkulainen, R. (1994). Cooperative self-organization of afferent and lateral connections in cortical maps. *Biological Cybernetics*, 71, 66–78.
- Stevens, J.-L. R., Law, J. S., Antolik, J., & Bednar, J. A. (2013). Mechanisms for stable, robust, and adaptive development of orientation maps in the primary visual cortex. *Journal of Neuroscience*, 33, 15747–15766.
- Swindale, N. V. (1996). The development of topography in the visual cortex: A review of models. *Network: Computation in Neural Systems*, 7, 161–247.
- Swindale, N. V., Shoham, D., Grinvald, A., Bonhoeffer, T., & Hubener, M. (2000). Visual cortex maps are optimized for uniform coverage. *Nature Neuroscience*, 3 (8), 822–826.
- Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46, 234–240.
- Triplett, J. W., Pfeiffenberger, C., Yamada, J., Stafford, B. K., Sweeney, N. T., Litke, A. M., Sher, A., Koulakov, A. A., & Feldheim, D. A. (2011).

- Competition is a driving force in topographic mapping. *Proceedings of the National Academy of Sciences, USA*, 108 (47), 19060–19065.
- Van Essen, D. C., & Gallant, J. L. (1994). Neural mechanisms of form and motion processing in the primate visual system. *Neuron*, 13 (1), 1–10.
- von der Malsburg, C. (1973). Self-organization of orientation-sensitive cells in the striate cortex. *Kybernetik*, 15, 85–100. Reprinted in Anderson & Rosenfeld (1988), 212–227.
- Wei, Y., Tsigankov, D., & Koulakov, A. (2013). The molecular basis for the development of neural maps. *Annals of the New York Academy of Sciences*, 1305, 44–60.
- Weliky, M., Bosking, W. H., & Fitzpatrick, D. (1996). A systematic map of direction preference in primary visual cortex. *Nature*, 379, 725–728.
- Willshaw, D. J., & von der Malsburg, C. (1976). How patterned neural connections can be set up by self-organization. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 194, 431–445.
- Willshaw, D. J., & von der Malsburg, C. (1979). A marker induction mechanism for the establishment of ordered neural mappings: Its application to the retinotectal problem. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 287, 203–243.
- Wilson, S. P., Law, J. S., Mitchinson, B., Prescott, T. J., & Bednar, J. A. (2010). Modeling the emergence of whisker direction maps in rat barrel cortex. *PLoS One*, 5 (1), e8778.
- Xu, X., Anderson, T. J., & Casagrande, V. A. (2007). How do functional maps in primary visual cortex vary with eccentricity?. *Journal of Comparative Neurology*, 501 (5), 741–755.
- Yu, H., Farley, B. J., Jin, D. Z., & Sur, M. (2005). The coordinated mapping of visual space and response features in visual cortex. *Neuron*, 47 (2), 267–280.
- Zylberberg, J., Murphy, J. T., & DeWeese, M. R. (2011). A sparse coding model with synaptically local plasticity and spiking neurons can account for the diverse shapes of V1 simple cell receptive fields. *PLoS Computational Biology*, 7 (10), e1002250.

Index terms

basis vector
chemoaffinity
componential representation
computational map
continuity
coverage
critical period
developmental model
dimensionality reduction
distributed representation
feature map
folding
Gabor
GCAL
generative model
Hebbian learning
ICA
 k -means
latent causes
latent space
latent variables
manifold
mechanistic model
neural map
normative model
objective function
overcomplete basis
phenomenological model
positive feedback
probabilistic inference
projective basis vector
projective field
receptive field
retinotectal map
retinotopic map
RF
sheet
SOM

sparse coding
SparseNet
template matching
topographic map
visuotopic map