Neural Computation

Mark van Rossum (2017 version)
2018 edits by Matthias Hennig

Lecture Notes 2018/19[1]

[1]December 3, 2018

Introduction: Neural Computation

The brain is a complex computing machine which has evolved to give the"fittest" output to a given input. Neural computation has as goal to describe the function of the nervous system in mathematical and computational terms. By analysing or simulating the resulting equations, one can better understand its function, research how changes in parameters would effect the function, and try to mimic the nervous system in hardware or software implementations.

Neural Computation is a bit like physics, that has been successful in describing numerous physical phenomena. However, approaches developed in those fields not always work for neural computation, because:

1. Physical systems are best studied in reduced, simplified circumstances, but the nervous system is hard to study in isolation. Neurons require a narrow range of operating conditions (temperature, oxygen, presence of other neurons, ion concentrations, ...) under which they work as they should. These conditions are hard to reproduce outside the body.
   Secondly, the neurons form a highly interconnected network. The function of the nervous systems depends on this connectivity and interaction, by trying to isolate the components, you are likely to alter the function.

2. It is not clear how much detail one needs to describe the computations in the brain. In these lectures we shall see various description levels.

3. Neural signals and neural connectivity are hard to measure, especially, if disturbance and damage to the nervous system is to be kept minimal.

Perhaps Neural Computation has more in common with trying to figure out how a complicated machine, such as a computer or car works. Knowledge of the basic physics helps, but is not sufficient. Luckily there are factors which perhaps make understanding the brain easier than understanding an arbitrary complicated machine:

1. There is a high degree of conservation across species. This means that animal studies can be used to gain information about the human brain. Furthermore, study of, say, the visual system might help to understand the auditory system.

2. The nervous system is able to develop by combining on one hand a only limited amount of genetic information and, on the other hand, the input it receives. Therefore it might be possible to find the organising principles and develop a brain from there. This would be easier than figuring out the complete 'wiring diagram'.

3. The nervous system is flexible and robust, neurons die everyday. This stands in contrast with computers. This might be helpful in constraining models.

Neural computation is still in its infancy which means many relevant questions have not been answered. My personal hope is that in the end simple principles will emerge. We will see a few examples of that during this course.

More reading

- Although no knowledge of neuroscience is required, there are numerous good neuroscience text books which might be helpful: Kandel, Schwartz, and Jessel (2000), Shepherd (1994), and Johnston and Wu (1995). Finally, Bear, Connors, and Paradiso (2000) has nice pictures.
- There is a decent number of books dealing with neural computation:
  - Dayan and Abbott (2002): Advanced level, yet readable text, not too much math. Wide range of up-to-date subjects.
  - Sterratt et al. (2011): Neural biophysics and beyond, many code examples.

  - Hertz, Krogh, and Palmer (1991): Neural networks book; the biological relevance is often speculative. Fairly mathematical. Still superb in its treatment of abstract models of plasticity.
  - Rieke et al. (1996): Concentrates on coding of sensory information in insects and what is coded in a spike and its timing.
  - Koch (1999): Good for the biophysics of single neurons.
  - Arbib(editor) (1995): Interesting encyclopaedia of computational approaches.
  - Churchland and Sejnowski (1994): Nice, non-technical book. Good on population codes and neural nets.

- Journal articles cited can usually be found via`www.pubmed.org`or `google.scholar.co.uk`
- If you find typos, errors or lack of clarity in these lecture notes, please tell me so they can be corrected.

Contents

Chapter 1

Important concepts

1.1 Anatomical structure of the brain

We briefly look at some common structures in the brain.

1.1.1 The neocortex

The neocortex is the largest structure in the human brain. The neocortex is the main structure you see when you remove the skull. Cortex means bark in Greek, and indeed, the cortex lies like a bark over the rest of the brain, Fig 1.1. It is in essence a two-dimensional structure that is a few millimetres thick.

The human cortical area contains about 2×10^{10} neurons. It is a few mm thick and has an area of 2000 cm^2 (equivalent to 18x18 inch). To fit all this area into the skull, the cortex is folded (convoluted) like a crumbled sheet of rubber. In most other mammals (except dolphins...) the convolution is less than in humans as there is less cortex to accommodate. Because it is so large in humans, cortex is thought to be essential for the higher order processing unique to humans. In contrast, even animals like dogs are remarkable functional without a cortex (see youtube).

Brodmann divided the cortex into areas based on small differences in the layout of the cells (using Nissl staining, below), see Fig. 1.2left. These areas correspond largely to different functions. Functionally, we can distinguish roughly in the cortex the following parts, Fig. 1.2right: 1) In the back there is the occipital area, this area is important for visual processing. Removing this part will lead to blindness, or loss of a part of function (such as colour vision, or motion detection). A very large part (some 40%) of the human brain is devoted to visual processing, and humans have compared to other animals a very high visual resolution.

2) The medial part. The side part is involved in higher visual processing, auditory processing, and speech processing. Damage in these areas can lead to specific deficits in object recognition or language. More on top there are areas involved with somato-sensory input, motor planning, and motor output that is send to the muscles via the spinal cord.

3) The frontal part is the"highest" area. The frontal area is important for short term or working memory (lasting up to a few minutes). Planning and integration of thoughts takes place here. Complete removal of the frontal parts in patients with psychiatric disorders (frontal lobotomy[1]) turns people into zombies without long term goals or liveliness. Limited damage to the pre-frontal has varied high-level effects, as illustrated in Fig. 1.3.

In lower animals, the layout of the nervous system is more or less fixed. In insects the layout is always the same: if you study a group of flies, you find the same neuron with the same dendritic tree in exactly the same place in every fly. But in mammals the precise layout of functional regions in the cortex can vary strongly between individuals. Left or right-handedness and native language will also be reflected in the location of functional areas. After a stroke, other brain parts can take

[1]There was even a Nobel Prize awarded to Moniz for this work

Figure 1.1.3. Comparing cortices of different mammals. Sections through pieces of cerebral cortices of various mammals drawn on 1:1 scale. Although cortical thickness varies only slightly, brain size (figure to the right of each cortical piece) varies considerably. Brain size is given as a ratio between a brain's weight and the weight of the mouse brain. [From C. U. A. Kappers, G. C. Huber, and E. C. Crosby, *The Comparative Anatomy of the Nervous System of Vertebrates, Including Man,* 1936, with permission of Macmillan & Co., Ltd.]

Figure 1.1: Left: Dissected human brain, seen from below. Note the optic nerve, and the radiation to the cortex.
Right: The cortex in different animals. Note the relatively constant thickness of the cortex across species. From Abeles (1991).

over the affected part's role. Especially in young children who had part of their brain removed or damaged, spared parts of the brain can take over many functions, see e.g. Battro (2000). There is even the case of a patient with very small brain (due to too much fluid around the brain), who nevertheless had a normal IQ (Feuillet, Dufour, and Pelletier 2007).

Lesions of small parts of the cortex can produce very specific deficits, such as a loss of colour vision, a loss of calculation ability, or a loss of reading a foreign language. (A good book with case stories is Sacks (1985), see also Andrewes (2002)). fMRI studies confirm this localisation of function in the cortex. This is important for us because it shows that computations are distributed over the brain. This is unlike a conventional Von Neumann computer, where most computations take place in the CPU (although this oversimplified for most modern computers). Similarly, long-term memory seems distributed over the brain. This bears a hopeful message for us trying to understand the computations; maybe we don't need to understand everything at once, but can divide the brain up in modules.

1.1.2 The cerebellum

Another large structure in the brain is the cerebellum (literally the small brain). The cerebellum is also a highly convolved structure. The cerebellum is a beautiful two-dimensional structure. The cerebellum contains more than half of all the neurons in the brain (Herculano-Houzel 2009). The cerebellum is involved in precise timing functions, in particular of motor functions. People without a cerebellum apparently function normally, until you ask them to dance or to tap a rhythm, in which they will fail dramatically[2]. Other roles of the cerebellum are less clear.

[2]See youtube for videos of cerebellar patients

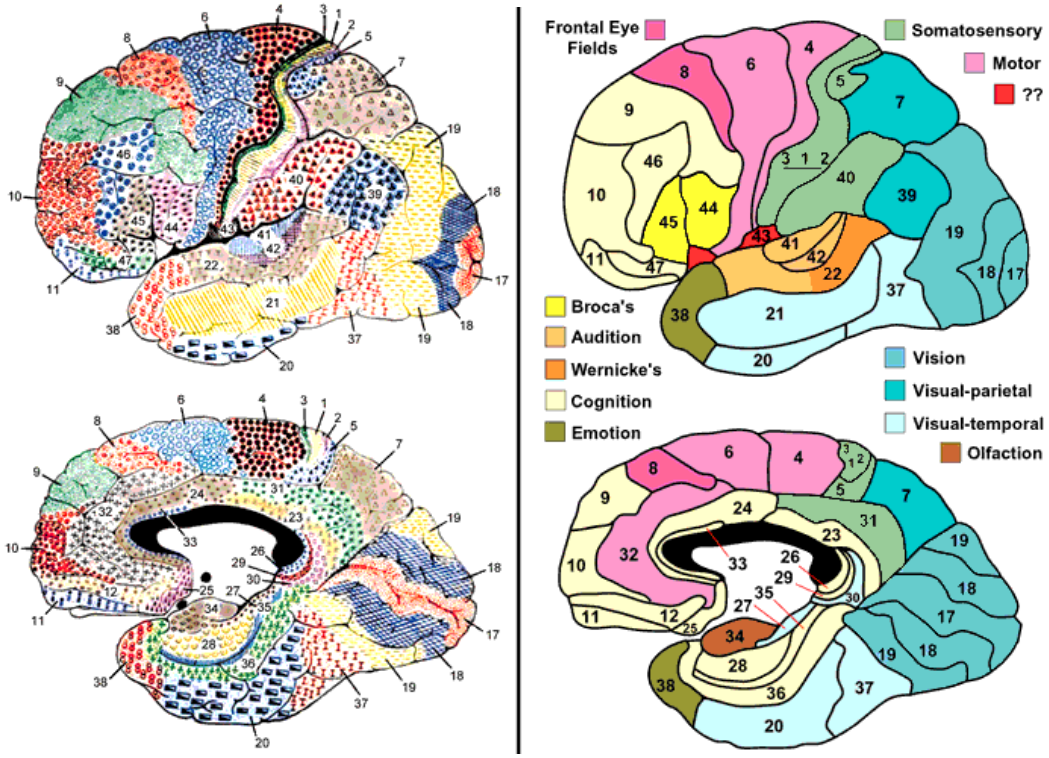


Figure 1.2: The Brodmann areas (left) and the various functions associated to them (Wikipedia has a nice clickable map). From http://spot.colorado.edu/~dubin/

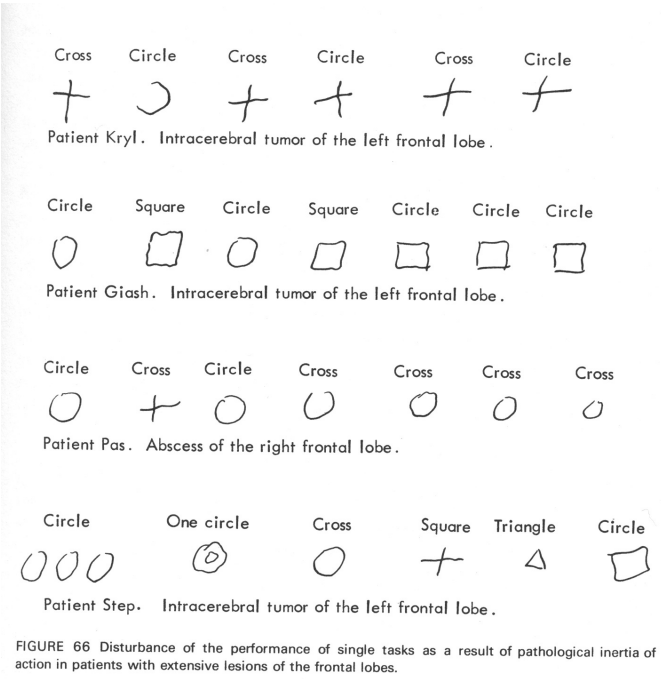


Figure 1.3: Pre-frontal damage. Patients are asked to draw a figure; command in upper line, response below. FromLuria 1966.

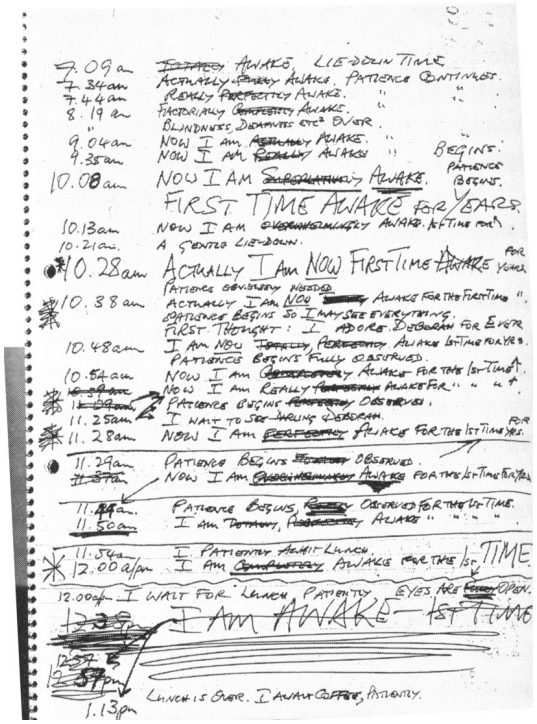


Figure 1.4: Amnesia in a patient whose hippocampus was damaged due to a viral infection. FromBlakemore 1988.

A typical laboratory timing task is classical conditioning (see below). Here a tone sounds and one second later an annoying air-puff is given into the eye. The animal will learn this association after some trials, and will then close its eye some time after the tone, right before the puff is coming. The cerebellum seems to be involved in such tasks and has been a favourite structure to model.

1.1.3 The hippocampus

The hippocampus (Latin for sea-horse) is a structure which lies deep in the brain. The hippocampus is sometimes the origin of severe epilepsy. Removal of the hippocampus, such as happened in the famous case of patient H.M., leads to (antero-grade) amnesia. That is, although events which occurred before removal are still remembered, new memories are not stored any more. Therefore, it is assumed that the hippocampus works as an association area which processes and associates information from different sources and temporally memorizing it before storing it in the cortex. In Fig. 1.4 another case of hippocampal damage is shown.

In rats some hippocampal cells are so-called place cells. When the rat walks in a maze, such a cell is active whenever the rat visits a particular location. The cells thus represent the location of the rat. Interference with the hippocampus will lead to a deteriorated spatial memory in rats.

Also much studied in the hippocampus are its rhythms: the global activity in the hippocampus shows oscillations. The role of these oscillations in information processing is not known.

1.2 Neurons and network

Like most other biological tissue, the brain consists of cells. One cell type in the brain are the so-called glial cells. These cells are widely believed to not do any computation, but provide support to the neurons. They suck up the spilt over neurotransmitter, and others provide myelin sheets around the axons of neurons.

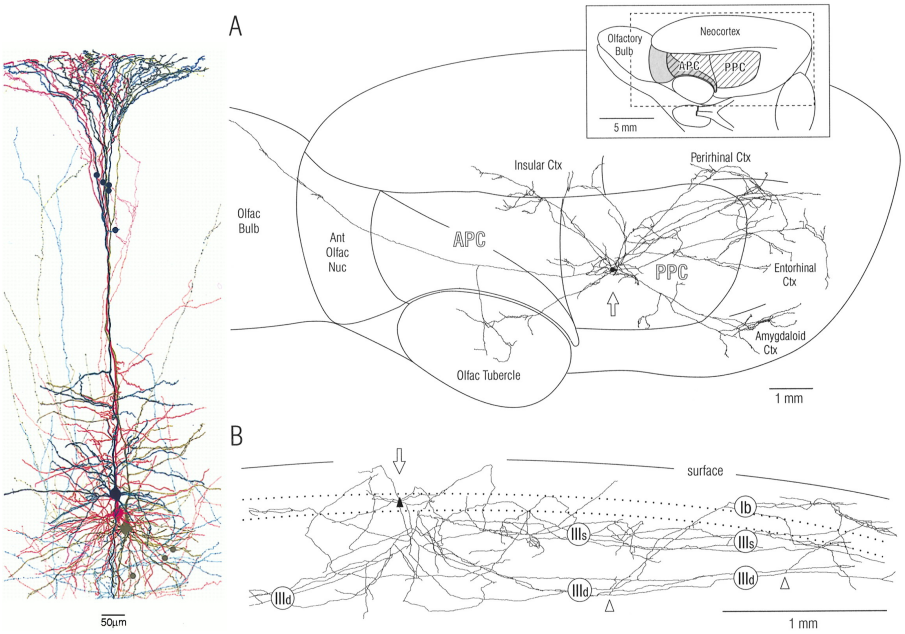


Figure 1.5: Left: dendritic tree of three cortical pyramidal neurons, from Markram, Wang, and Tsodyks 1998. Part of the axons are also shown (in faint colour). Circles denote (putative) synapses. Right: extensive axon of a neuron from Johnson et al. 2000. Note difference in scale.

More important for us are the neurons. There are some 10^{11} neurons in a human brain. Like other biological cells, a neuron has a nucleus, mitochondria and other organelles. In other aspects neurons are different from other cells. For instance, only in a few brain areas new neurons are generated after birth, while in most other parts of the body birth and death of cells is an ongoing process. Secondly, the brain is separated from blood by the so called blood-brain barrier in the blood vessels. This protection prevents many substances, harmful and medical alike, to enter the neural tissue.

The basic anatomy of the neurons is shown in Fig. 1.6: Every neuron has a cell body, or soma, contains the nucleus of the cell. The nucleus is essential for the cell, as here the protein synthesis takes place making it the central factory of the cell. The neuron receives its input through synapses on its dendrites (dendrite: Greek for branch). The dendritic trees can be up to one millimeter long, can be very elaborate and often receive more than 10,000 synapses, Fig. 1.5. The synapses are where the neuron connects to other neurons.

Neurons mainly communicate using SPIKES, these are a brief (1ms), stereotypical excursions of the neuron's membrane voltage. Spikes are thought to be mainly generated in the axon-hillock, a small bump at the start of the axon. From there the spike propagates along the axon. The axon can be very long (up to one meter or so when it goes to a muscle). To increase the speed of signal propagation, long axons have a fatty insulator (myelination sheet) around them. The cortical regions are connected to each other with myelinated axons. This is the white matter, because the layer of fat gives the neurons a white appearance, see Fig. 1.2. The axon ends in many axon terminals (about 10,000 of course), where the connection to next neurons in the circuit are formed, Fig. 1.5. From the soma the action potential also propagates back into the dendrites. This provides the synapses with a signal that an action potential was fired.

In general the dendritic tree has been more studied is probably more interesting than the axonal structure. The dendrites collect input, and inputs can interfere or support each other, while in the axon the signal is everywhere the same: a spike or not.

A basic distinction between neurons is between the excitatory and inhibitory ones, depending on whether they release excitatory or inhibitory neurotransmitter. (The inhibitory neurons I

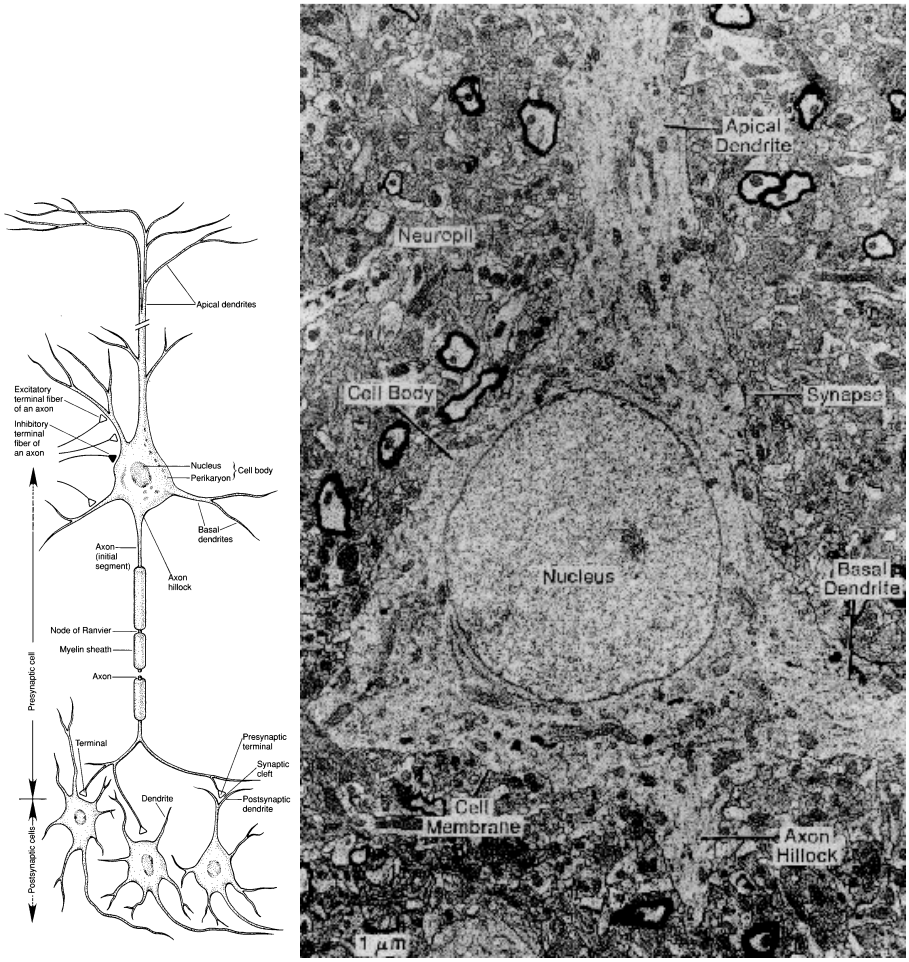


Figure 1.6: Sketch of the typical morphology of a pyramidal neuron. Right: Electron micrograph of a neuron's cell body. From Shepherd 1994.

personally call "neuroffs", but nobody else uses this term. Yet...). There are multiple sub-types of both excitatory and inhibitory neurons. How many is not well known, at least tens or hundreds. As genetic markers become more refined, more and more subtypes of the neurons are expected to appear. It is not clear if and how these different subtypes have different computational roles. In most experiments determine the precise types is too difficult, and all groups are lumped together; even excitatory and inhibitory neurons are often hard to distinguish.

Finally, in reading biology literature one should remember there are very few hard rules in biology: There are neurons which release both excitatory and inhibitory neurotransmitter, there are neurons without axons, not all neurons spike, etc...

1.2.1 Cortical layers

Each piece of cortex will contains some 100.000 neurons beneath 1 mm^2 of surface area, independent of cortical thickness. When one looks at an unstained slice of brain under the microscope one does not see much, but after applying a stain, a regular lay-out of the cells becomes visible. There is a variety of stains, see Fig. 1.7left, and nowadays also genetic markers are often used; these for instance leads to certain cells becoming fluorescent once they have adopted the genetic code for the dye.

One has classified 6 layers in the cortex. The cortical circuitry does not appear to vary much

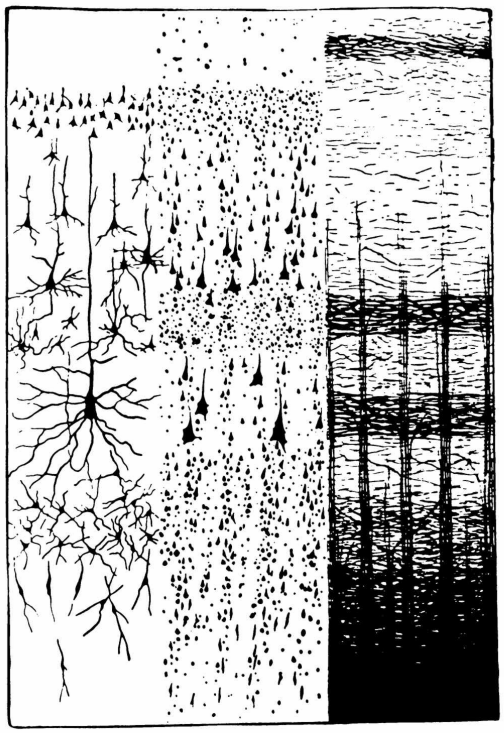


Figure 1.3.1. Cortical slices stained by the Golgi, Nissl, and Weigert methods. [From H. Gray, *Anatomy of the Human Body,* 26th ed., 1954, with permission of Lea & Febiger. After Brodman: from *Luciani's Physiology,* Macmillan & Co., Ltd.]

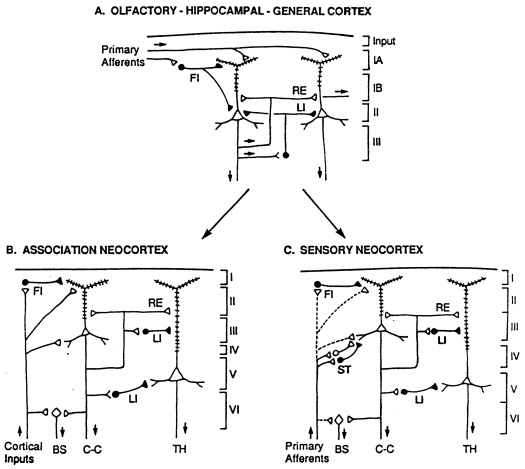
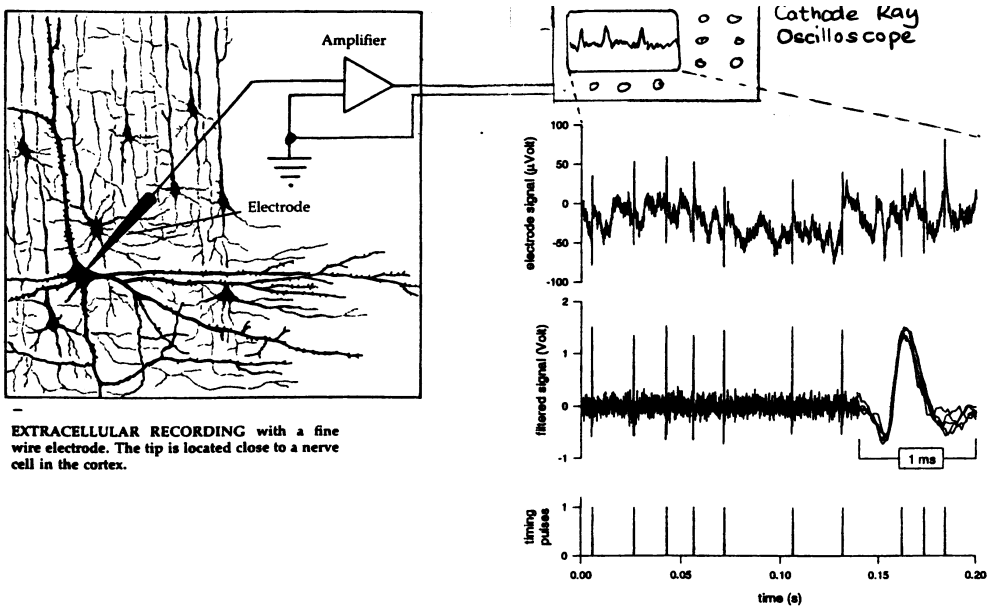


Figure 1.7: Left: Layers in the cortex made visible with three different stains. The Golgi stain (left) labels the soma and the thicker dendrites (only a small fraction of the total number of cells is labelled). The Nissl stain shows the cell bodies. The Weigert stain labels axons. The vertical dimension is one or two millimetres. FromAbeles 1991. Right: very approximate circuitry in various cortical areas. FromShepherd 1994.

across brain area, while the computations that the areas are doing are quite different (e.g. visual, sound, motor,...). Fig. 1.7right shows an attempt to draw circuit diagrams for the cortex. As shown, the input and output layers arranged in a particular manner. This holds for cortical areas with functions a diverse as lower and higher vision, audition, planning, and motor actions. Why this particular layout of layers of cells is so common and what its computational function is, is not clear. It should be noted though that these diagrams are far from complete and that while computationalists will emphasize the commonalities, experimentalists will typically emphasize the differences across brain areas.

The connectivity within the cortex is large (about 100.000 km of fibers). Anatomically, the cortex is not arranged like a hierarchical, feed-forward network. Instead, a particular area receives not only feed-forward input from lower areas, but also many lateral and feedback connections. The feedback connections are usually excitatory and non-specific.

It would be nice to know the wiring diagram between the neurons, when we try to figure out how it works. However, this turns out to be very difficult and labor intensive task. Automated experiments and analysis, such as done by the Alan Brain Institute will hopefully bring resolve.



EXTRACELLULAR RECORDING with a fine wire electrode. The tip is located close to a nerve cell in the cortex.

Figure 1.8: A set-up to measure extracellular activity. After a band-pass filter the spikes can be clearly extracted from the small electrical signal. Electrodes can remain implanted for years. Note the signal is quite weak (μV).

1.3 Measuring neural activity

Many ways have been developed to measure neural activity. Ideally, one would be able to measure from millions of neurons individually (high 'spatial resolution'), for very long times, with high temporal resolution and without disturbing the brain. It would be even better if we can also stimulate the brain. However, no technique is ideal:

- Both EEG (electro encephalogram) and ERP (event related potential) measure the potential on the scalp. It is the oldest technique to measure neural activity. This has very poor spatial resolution, but temporal resolution is good. Non-invasive, so can be used on humans. Good for behavioural reaction times, e.g. Thorpe, Fize, and Marlot (1996). A related technique is MEG that measures the tiny magnetic fields accompanying the electric activity in the brain.
- fMRI (functional magnetic resonance imaging) measures increased blood oxygenation level. The signal is related to neural activity in a not fully understood manner. It seems to correlate better to synaptic activity than to spike activity (Logothetis et al. 2001). Resolution: about 1mm, 1 minute. Non-invasive. Good for global activity and localisation studies in humans.
- Extracellular electrodes: When a tungsten electrode is held close enough to a firing cell, the cell's spikes can be picked up. Alternatively, the slow components of the voltage can be analysed, this is called the field potential and corresponds to the signal from ensembles of synapses from the local neighbourhood of cells. Extracellular recordings can be done chronically and in awake animals. A newer trend is to use many electrodes at once (either in an array, or arranged in tetrodes). Limitations of this technique: no access to precise synaptic current, difficult to control stimulation, requires spike sorting (an electrode receives usually signals from a couple of neurons, which is undesirable; the un-mixing is called spike sorting) . Extracellular electrodes can also be used to stimulate neurons.
- Intracellular: Most intracellular recording are now done using the patch clamp technique. A glass pipette is connected to the intracellular medium, Fig. 1.9. From very small currents

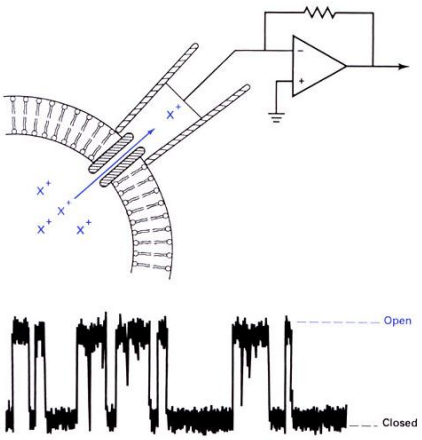


Figure 1.9: Patch-clamp recording. The inside of the cell is connected to the pipette. This allows the measurement of single channel openings (bottom trace).

(single channel) to large currents (synaptic inputs and spikes) can be measured. Secondly, the voltage and current in the cell can be precisely controlled. Access to the intracellular medium allows to wash in drugs that work from the inside of the cell. However, the access can also lead to washout (the dilution of the cell's content). Limitations: harder *in vivo* (due to small movements of the animal even under anaesthesia), and limited recording time (up to 1 hour).

- A relatively new method is to use optical imaging. Dyes sensitive to Ca (or, less common, to voltage) can be added, and activity can be measured, from increasingly many neurons at once. Light activated ion channels have been engineered that also allow for optical stimulation of neurons.

Technique:	Intracell. electr.	Extracell. electr.	Field potential	EEG	fMRI	Optical signal
# channels	few	few to 1000s	up to ~100	~50	very many	100-1000
Time resol.	~0.1 ms	~0.1 ms	~1ms	10ms	>1 sec	10ms-1s
Spatial resol.	single neuron	single neuron	within ~100 μm^2	large areas	1-5mm	single neuron and smaller
Signal	voltage/currents	(voltage) spikes	extracellular pot. mainly synaptic	sum of many extracell. pot.	blood oxygenation	fluorescent indicator
Stimulation	yes	yes, imprecise	yes, imprecise	no	no	yes

1.4 Preparations for measuring from neurons

In order to study the nervous system under controlled conditions various preparations have been developed. Most realistic, and therefore most ideal, would be to measure the nervous system *in vivo* (alive) without anaesthesia. However, this is not always possible due to both technical and ethical problems. Under anaesthesia the technical problems are less, but reliable intracellular recording is still difficult. And, of course, the anaesthesia itself changes the functioning of the nervous system.

For *in vitro* ('in glass') studies a widely used method is to prepare slices of brains, about 1/2 mm thick. Slices allow for accurate measurements of single cell or few cell properties. However, some connections will be severed, and it is not clear how well the *in vivo* situation is reproduced, as the environment (transmitters, energy supply, temperature, modulators) will be different.

Finally, it is possible to culture the cells from young brain. The neurons will by themselves form little networks. These cultures can be kept alive for months under the right conditions. Also here the relevance to *in vivo* situations is not always clear.

More reading: Quantitative anatomy Braitenberg and Schüz (1998), neuropsychological case studies Sacks (1985) and Luria (1966)

Chapter 2

Passive electric properties of neurons

2.1 Resting potential

A living neuron maintains a voltage drop across its membrane. One commonly defines the voltage outside the cell as zero. At rest the inside of the cell will then be at about -70mV (range -90..-50mV). This voltage difference exist because the ion concentrations inside and outside the cell are different. The main ions are K^+ (potassium, or *kalium* in many languages), Cl^-(chloride), Na^+ (sodium), and Ca^{2+} (calcium).[1]

Consider for instance for Na, the concentration outside is 440mM and inside it is only 50mM (taking the squid axon as an example). If the cell membrane were permeable to Na, it would flow in: first, because of the concentration gradient (higher concentration outside than inside), second, because of the attraction of the negative membrane to the positive Na ions. Because of these cooperation two forces, Na influx does not stop when the voltage across the membrane is zero. Only if the voltage across the membrane reaches +55 mV, net Na inflow would stop. This +55mV is called the REVERSAL POTENTIAL of Na. Likewise K has a reversal potential of -75mV (outside 20 mM and inside 400mM), and Cl of -60mV (outside 500mM and inside 50mM). The reversal potential can be calculated from the Nernst equation

$$E = \frac{58mV}{z} \log_{10} \frac{[X]_{outside}}{[X]_{inside}}$$

which follows from the condition that diffusive and electrical force should cancel at equilibrium. The valence of the ion is represented by z (e.g. +2 for Ca).

However, at rest the Na channels are largely closed and only very little Na will flow into the cell.[2]The K and Cl channels are somewhat open, together yielding a resting potential of about -70mV. By definition no net current flows at rest (else the potential would change). The concentration gradient of ions is actively maintained with ion-pumps and ion-exchangers. These proteins move ions from one side of the membrane to the other. Because this typically goes against the forces this costs energy (ATP).

[1]For those familiar with electronics: In electronics, charge is mostly carried by electrons. Here, all ions in solution carry charge, so we have at least 4 different charge carriers, all of them contributing to the total current in the cell.

[2]When multiple ion species contribute to the potential, the Goldman-Hodgkin-Katz voltage equation should be used to calculate the potential (Hille 2001). In that case the equilibrium needs to be actively maintained by ion-exchangers in the membrane.

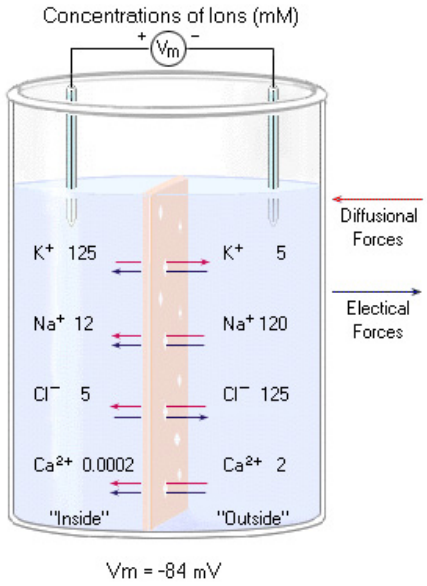


Figure 2.1: The typical ion concentration inside and outside neurons. The concentrations listed here are for mammals, whereas those in the text are for squid.

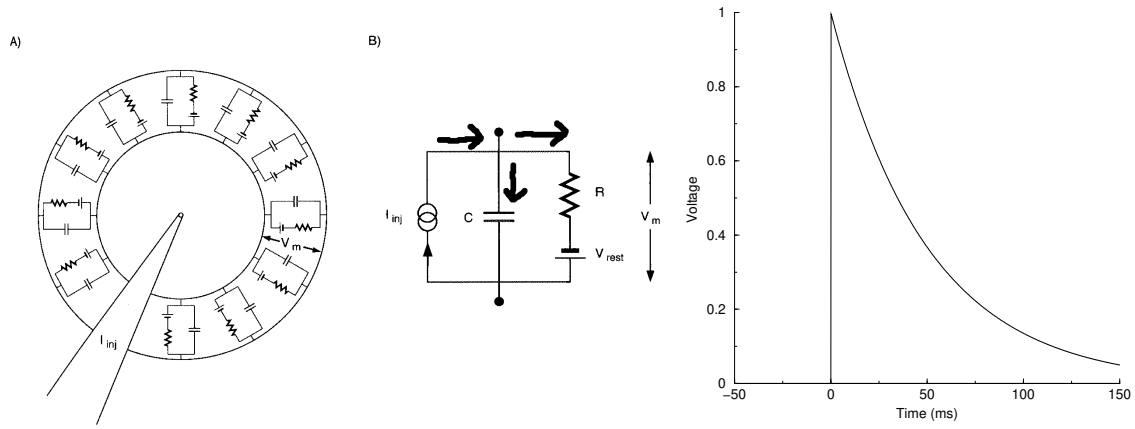


Figure 2.2: Left: RC circuit to model a single compartment cell. Middle: schematic model. The injected current is I_{inj}, the total capacitance is C and R is the total membrane resistance. Right: Response of the voltage to a step in the stimulus current.

2.2 Passive neuron models

If one electrically stimulates a neuron sufficiently it will spike. But before we will study the spiking mechanism, we first look at the so-called passive properties of the cell. We also ignore for now that the various ions listed above that act as charge carriers. We model the cell by a network of passive components: resistors and capacitors. This approximation is reasonable when the cell is at rest and the membrane voltage is around the RESTING voltage. Most active elements only become important once the neuron is close to its firing threshold. The simplest approximation is a passive single compartment model, shown in Fig. 2.2a. This model allows the calculate the response of the cell to an input. KIRCHHOFF'S (FIRST) LAW tells us that the sum of the currents at any point in the circuit should be zero.

$$\boxed{\sum I = 0}$$

What are the different contributions to the current? The current through the resistor is given by OHM'S LAW[3]

$$I_{resistor} = \frac{\Delta V}{R} = \frac{V_m - V_{rest}}{R_m}$$

where V_m is the membrane potential that we like to know, and V_{rest} is a battery voltage, as we will see in the absence of any external current V_m will go to V_{rest}. It is also useful to introduce the CONDUCTANCE, the conductance is simply the inverse of the resistance

$$g = 1/R$$

The larger the conductance, the larger the current. Conductance is measured in Siemens (symbol S). In biophysics, conductance is a bit easier to think about than resistance, because the conductance is proportional to the number of open channels.

Similarly, there is a current associated to the capacitance. This current flows for instance when initially the voltage across the capacitor is zero, but suddenly a voltage is applied across it. Like the battery of your cell-phone, a current flows only until the capacitor is charged up (or discharged). The current into the capacitor is

$$I_{cap} = C\frac{dV_m(t)}{dt}$$

It is important to note that no current flows when the voltage across the capacitor does not change over time. (Alternatively, if you want, you can describe the capacitor using a complex impedance, using $Z = \frac{1}{i\omega C}$, where ω is the frequency of the periodic signals in which the voltage can be decomposed. This trick, common to engineers, simplifies calculations, as the time derivative is replaced by a simple multiplication - a common trick when analyzing linear differential equations- , but will require complex calculus).

Finally, we assume an certain external current is injected, for instance via a pipette. This is sometimes called a CURRENT CLAMP (to be compared to the voltage clamp introduced next chapter).

As stated the sum of the currents should be zero. We have to fix the signs of the currents first: we define currents flowing away from the point to be negative. Now we have $-I_{resistor} - I_{cap} + I_{inj} = 0$. The circuit diagram thus leads to the following differential equation for the membrane voltage.

$$\boxed{C\frac{dV_m(t)}{dt} = -\frac{1}{R_m}[V_m(t) - V_{rest}] + I_{inj}(t)} \tag{2.1}$$

In other words, the membrane voltage is given by a first order differential equation. (The voltage low-pass filters the current).[4]

It is always a good idea to study the steady state solutions of differential equations first. This means that we assume I_{inj} to be constant and $dV/dt = 0$. We find for the membrane voltage $V_\infty = V_{rest} + R_m I_{inj}$. If the current increases the membrane voltage ($I_{inj} > 0$) it is called DE-POLARISING; if it lowers the membrane potential it is called HYPER-POLARISING.

How rapidly is this steady state approached? If the voltage at t=0 is V_0, one finds by substitution that $V(t) = V_\infty + [V_0 - V_\infty]\exp(-t/\tau_m)$. So, the voltage settles exponentially. The input the low-pass filtered; fast fluctuations are filtered out. The product $\tau_m = R_m C$ is the time constant of the cell. For most cells it is between 20 and 50ms, but we will see later how it can be smaller under spiking conditions. The time-constant determines how fast the sub-threshold membrane voltage reacts to fluctuations in the input current. The time-constant is independent of the area of the cell. The capacitance is proportional to the membrane area ($1\mu F/cm^2$), namely, the bigger the membrane area the more charge it can store.

[3]Ohm's law says that current and voltage are linearly related, and their coefficient is the resistance. As soon as the linearity is lost, Ohm's law is broken. This happens for instance in diodes or neon tubes, in that case we have a *non-Ohmic* conductance.

[4]You a can also think of filling a leaky bath. Voltage is the water level.

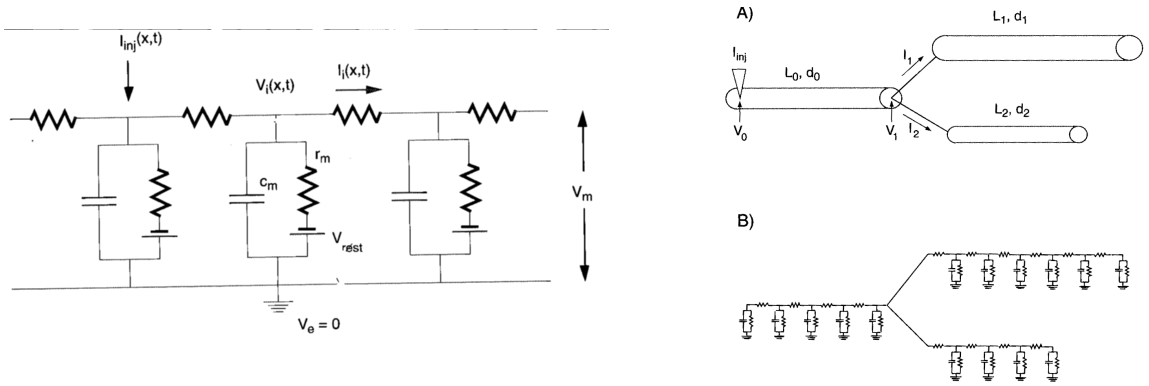


Figure 2.3: Top: Electrical equivalent of a small cable segment. Bottom: Also branched cables can be calculated in this formalism. FromKoch 1999.

It is useful to define the specific resistance, or resistivity, r_m as

$$r_m = AR_m$$

The units of r_m are therefore $\Omega.cm^2$. The resistance is inversely proportional to membrane area (some $50k\Omega.cm^2$), namely, the bigger the membrane area the more leaky it will be. The product of membrane resistance and capacitance, i.e. the time constant, is independent of area. Also the conductance is often expressed per area (units S/cm^2), likewise, the specific capacitance is defined as $c_m = C/A$.

Because it is a linear system, the voltage can be solved for an arbitrary time varying current

$$V(t) = V_0 \exp(-t/\tau) + \frac{R_m}{\tau} \int_0^t \exp[-(t-t')/\tau] I_{inj}(t') dt'$$

as can be shown be substitution in Eq. 2.1. Engineers might recognize this as the convolution of the input current with the exponential kernel that represents the RC filter, $V(t) = (K * I)(t)$, where we ignored the initial condition.

Note, this section dealt with just an approximation of the behaviour of the cell. Such an approximation has to be tested against data. It turns out to be valid for small perturbations of the potential around the resting potential, but at high frequencies corrections to the simple RC behaviours exist (Stevens 1972), and of course for larger perturbations the neuron will fire an action potential.

2.3 Cable equation

The above model lumps together the cell into one big soma. But neurons have extended structures (dendrites and axons) which are not perfectly coupled, therefore modelling the cell as a single compartment has only limited validity. The first step towards a more realistic model is to model the neuron as a cable. A model for a cable is shown in Fig. 2.3. One can think of it as many cylindrical compartments coupled to each other with resistors. The resistance between the cables is called axial resistance. The diameter of the cylinder is d, its length is called h. Thus the cross-section has an area $A = \pi d^2/4$. Now the resistance between the compartments is $r_i h/A = 4r_i h/(\pi d^2)$, where r_i is the intracellular resistivity (somewhere around 100 $\Omega.cm$). The outside area of the cylinder is πhd. As a result the membrane resistance of a single cylinder is $r_m/(\pi hd)$. The capacitance of the cylinder is $c_m \pi hd$.

The voltage now becomes a function of the position in the cable as well. Consider the voltage at the position x, $V(x,t)$ and its neighbouring compartments $V(x+h,t)$ and $V(x-h,t)$. For

convenience we measure the voltage w.r.t. V_{rest}. Again we apply Kirchhoff's law and we get

$$c_m \frac{dV_m(x,t)}{dt} = -\frac{1}{r_m} V_m(x,t) + \frac{d}{4h^2}\frac{1}{r_i}[V_m(x+h,t) - 2V_m(x,t) + V_m(x-h,t)] + \frac{1}{\pi d} I_{ext}(x,t)$$

(where the external current is now defined per unit length.) Now we take the limit of small h, i.e. we split the cable in very many small elements, and get the passive cable equation. Use that the derivative is defined as $\frac{df(x)}{dx} = \lim_{h\to 0} \frac{1}{h}[f(x+h) - f(x)]$, so $\frac{d^2 f(x)}{dx^2} = \lim_{h\to 0} \frac{1}{h^2}[-2f(x) + f(x+h) + f(x-h)]$.

$$\boxed{c_m \frac{dV_m(x,t)}{dt} = \frac{d}{4}\frac{1}{r_i}\frac{d^2 V_m(x,t)}{dx^2} - \frac{1}{r_m} V_m(x,t) + \frac{1}{\pi d} I_{ext}(x,t)}$$

The cable equation[5] describes how a current locally injected to the cable propagates and decays.

2.3.1 Steady state solution to cable equation

First, check that when there are no lateral currents (iso-potential cable), the cable equation reduces to Eq. 2.1. Next, consider the steady state, i.e. dV/dt is zero. Suppose a constant current is injected only at $x = 0$. This current injection can be written as a delta function $I = I_0\delta(x)$. (A delta function is zero everywhere, except at zero, where it is infinite. Its total area is one, $\int_{-\infty}^{\infty} \delta(x)dx = 1$. For practical purposes, you can think of the delta function as a very, very narrow Gaussian, or a very narrow but tall square pulse.)

In the steady state, the cable equation then yields $0 = \frac{d}{4}\frac{1}{r_i}\frac{d^2V(x)}{dx^2} - \frac{1}{r_m}V_m(x) + \frac{1}{\pi d} I_0\delta(x)$. Everywhere away from $x = 0$, we simply have $\frac{d^2V(x)}{dx^2} = \frac{1}{r_m}V_m(x)$, this has an exponential solution. The solution to the steady state equation when $x \neq 0$ is

$$V_m(x) = a_0 \exp(-x/\lambda) + b_0 \exp(x/\lambda)$$

where

$$\lambda = \sqrt{dr_m/4r_i}$$

is called the SPACE-CONSTANT of the cell, or ELECTROTONIC LENGTH. It determines how far within the neuron a constant input reaches. For typical neuron dendrites it is between 100 μm and 1 mm. Note that it depends on the diameter of the cable. Also note, that it is comparable to the typical size of a dendritic tree. If the space constant were much longer, we would not need to take any cable effects into account, the cell would act like a single compartment. If the space constants were much shorter, the most distal synapses would exert no effect on the somatic potential.

To find the behavior around $x = 0$, integrate this over a narrow region around $x = 0$ (i.e. apply $\lim_{\epsilon\to 0}\int_{-\epsilon}^{\epsilon}$ on both sides) and you find $\frac{dV(\epsilon)}{dx} - \frac{dV(-\epsilon)}{dx} = -\frac{4r_i}{\pi d^2} I_0$. In other words, the spatial derivative of $V(x)$ makes a jump at $x = 0$, hence $V(x)$ itself will have a cusp, Fig. 2.4.

The theory of differential equations says that constants a_0 and b_0 have to be determined from the boundary conditions: If the cable is finite and sealed at the ends, the derivative of V_m is zero at the ends (as no current can flow out). If the cable is infinite, only the solution which gives finite values for V_m is allowed, i.e. $V_m(x) = c\exp(-|x|/\lambda)$, where the constant c follows from I_0.

Time dependent solutions

Let's now return to time-dependent cable equation. If we stimulate the cable with a brief current pulse at $x = 0$, $I_{ext}(x,t) = I_0\delta(t)\delta(x)$. Assuming a infinitely long cable, the voltage is

[5] Sometimes this form is easier

$$\tau \frac{dV_m(x,t)}{dt} = \lambda^2 \frac{d^2 V_m(x,t)}{dx^2} - V_m(x,t) + \frac{r_m}{\pi d} I_{ext}(x,t)$$

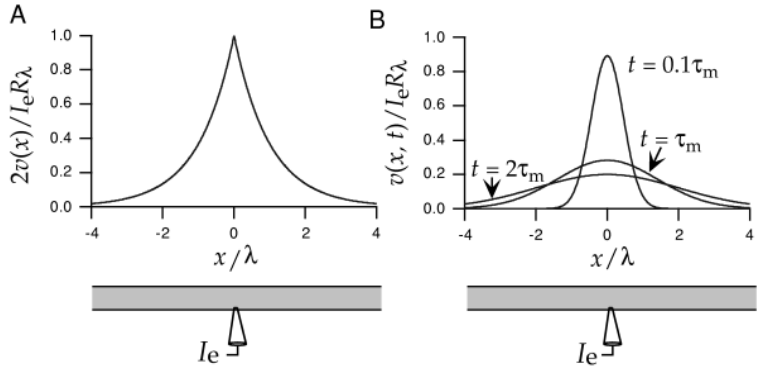


Figure 2.4: Solution to the cable equation in an infinite cable. A) Steady state solution when a constant current is injected at x=0. B) Decaying solution when a brief current pulse is injected at x=0 and t=0. Taken fromDayan and Abbott 2002.

$$V_m(x,t) = \frac{I_0 r_m}{\sqrt{4\pi t/\tau_m}} \exp(-\frac{x^2}{4\lambda^2}\frac{\tau_m}{t}) \exp(-t/\tau_m) \tag{2.2}$$

The solution is plotted in Fig. 2.4B. Mathematically, the cable equation is a diffusion equation with an absorption term. In the cable equation the voltage V is linear in the current I and hence solutions like Eq, 2.2 are linear in I_0[6]. Again, as in the single compartment case, in principle this solution can be used to calculate the solution for arbitrary input currents by decomposing the current into delta-pulses and adding the responses together (using Green's functions for the experts). This will turn out to be very different for the active, spiking solution discussed in the next chapter, as then the membrane behaves strongly non-linearly, and doubling of the current will not simply double the response in the voltage.

In the passive cable there is no wave propagation, the input only "blurs". Therefore it is a bit tricky to define velocities and delays. One way is to determine when the response reaches a maximum at different distances from the injection site. For instance, in Fig. 2.4 right the voltage at $x/\lambda = 1$ reaches a maximum around $t = 1.\tau$. In general one finds with Eq. 2.2 that

$$t_{max} = \frac{\tau_m}{4}[\sqrt{4x^2/\lambda^2 + 1} - 1] \approx \frac{1}{2}x\frac{\tau_m}{\lambda}$$

For large x this means a "speed" $v = 2\lambda/\tau_m$.

It is important to realise that because of the blurring, high frequencies do not reach as far in the dendrite as low frequencies. In other words, the response far away from the stimulation site is a low-pass filtered version of the stimulus. This is called CABLE FILTERING. Mathematically this can be described using a frequency dependent space constant, $\lambda(f)$, which is for higher frequencies than for low or zero (DC) frequency.

In the derivation we assumed a long cable and homogeneous properties. Real neurons have branches and varying thicknesses. The voltage can be analytically calculated to some extent, but often it is easier to use a computer simulation in those cases. Under the assumption that the properties of the cables are homogeneous, this is fairly straightforward using the cable equation. However, the validity of assuming homogeneity is not certain. This important question has to be answered experimentally.

Analogy with fluid flow The flow of electricity can for our purposes be understood in analogy with the flow of liquids. The voltage difference is the difference in height between the water level in a

[6]Definition: a function $f(x, y, \ldots)$ is linear in x when $f(ax_1 + bx_2, \ldots) = af(x_1, \ldots) + bf(x_2, \ldots)\ \forall_{a,b,x_1,x_2}$, where x_1 and x_2 can be vectors. Non-linear is anything else.

water tank and the water level in a bucket (the cell). The current that flows is proportional to the level difference (voltage) and the opening of the tap (the conductance). The capacitance is the size of the bucket. Leak is equivalent to holes (channels) in the bucket. A cable becomes a leaky garden hose.

More reading: Koch (1999)
Table with various notation conventions:

	membr. resistance	spec. membr. resistance	intrac. resistivity	spec. capa.
Here	R_m	r_m	r_i	c_m
Koch book		R_m	R_i	C_m
Dayan book	R_m, C_m	r_m	r_L	c_m
NEURON		g_pas	Ra	cm

Chapter 3

Active properties and spike generation

We have modelled a neuron with passive elements. This is a reasonable approximation for sub-threshold effects and might be useful to describe the effect of far away dendritic inputs on the soma. However, an obvious property is that most neurons produce action potentials (also called spikes). Suppose we inject current into a neuron which is at rest (-70mV). The voltage will start to rise. When the membrane reaches a threshold voltage (about -50mV), it will rapidly depolarise to about +10mV and then rapidly hyper-polarise to about -70 mV. This whole process takes only about 1ms. The spike travels down the axon. At the axon-terminals it will cause the release of neurotransmitter which excites or inhibits the next neuron in the pathway.

From the analysis of the passive properties, it seems that in order to allow such fast events as spikes, the time-constant of the neuron should be reduced. One way would be the reduce the membrane capacitance, but this is biophysically impossible. The other way is to dramatically increase the conductance through the membrane, this turns out to be the basis for the spike generation. The fact that the spike exceeds 0mV, indicates that multiple ion species are involved.

The magnificent series of papers of Hodgkin and Huxley in 1952 explains how spike generation works (Hodgkin and Huxley 1952).

3.1 The Hodgkin-Huxley model

The membrane of neurons contains voltage-gated ion-channels, Fig. 3.1. These channels let through only one particular type of ion, typically Na or K, with a high selectivity. For instance, the channel for Na is called a Na-channel. Due to an exquisite mechanism that relies on conformational changes of the channel, the open probability of the channel depends on the voltage across the membrane. When an action potential initiates, the following events happen: 1) close to the threshold voltage a few Na channels start to open. 2) Because the sodium concentration is higher outside the cell and the sodium is positively charged both electric and diffusive force cooperate (its reversal potential is +40mV). The sodium starts to flow in, de-polarising the cell. 3) This positive feedback loop will open even more Na channels and the spike is initiated. 4) However, rapidly after the spike starts, sodium channels close again and now K channels open. 5) The K ions starts to flow out the cell, hyper-polarising the cell, roughly bringing it back to the resting potential.

We now describe this in detail. Consider a single compartment, or a small membrane patch. As before, in order to calculate the membrane potential we collect all currents. In addition to the leak current and capacitive current, we now have to include Na and K currents. Let's first consider the sodium current. The current (per area) through the sodium channels is

$$I_{Na}(V,t) = g_{Na}(V,t)[V(t) - E_{Na}] \tag{3.1}$$

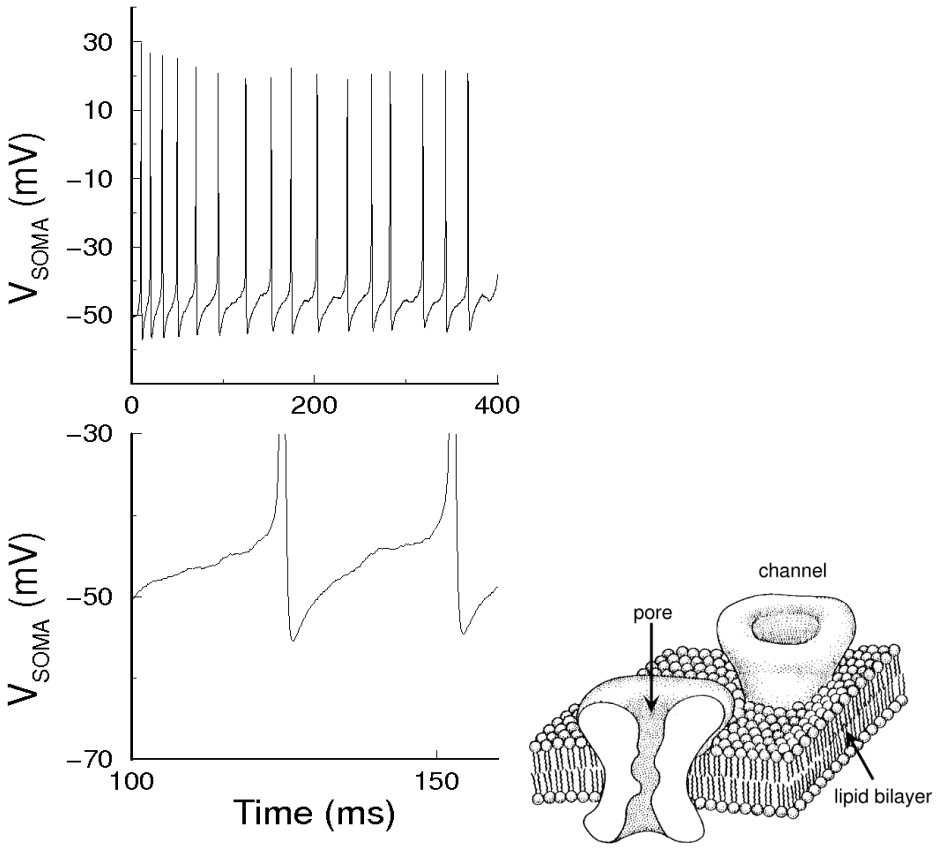


Figure 3.1: Left: Sample spikes from a retinal ganglion cell injected with current. Right: Voltage gated channels populate the cell membrane. The pores let through certain ions selectively, and open and close depending on the membrane voltage.

The current is proportional to the difference between the membrane potential and the Na reversal potential. The current flow will try to make the membrane potential equal to the reversal potential.[1][2]

The total conductance per area through the channels is given by the number of open channels

$$g_{Na}(V,t) = g^0_{Na}\,\rho_{Na} P_{open}(V,t)$$

where g^0_{na} is the open conductance of a single Na channel (about 20 pS), and ρ_{Na} is the density of Na channels per area. The Na channel's open probability turns out to factorise as

$$P_{open}(V,t) = m^3(V,t)h(V,t)$$

where m and h are called gating variables. We just state this equation here; you can imagine that a lot of work of Hodgkin and Huxley went into finding this relation.

Microscopically, the gates are like little binary switches that switch on and off depending on the membrane voltage and correspond to the conformation of the channel. The Na channel has 3 switches labelled m and one labelled h. In order for a sodium channel to conduct all three m and

[1]The amount of Na that flows in during a single action potential causes a very small change in the concentrations inside and outside the cell and hence reversal potential. However, in the long run, ion pumps and ion exchangers maintain the correct concentrations (an energy costly process).

[2]There a small corrections to Eq. 3.1 due to the physical properties of the channels, given by the Goldman-Hodgkin-Katz current equation (Hille 2001).

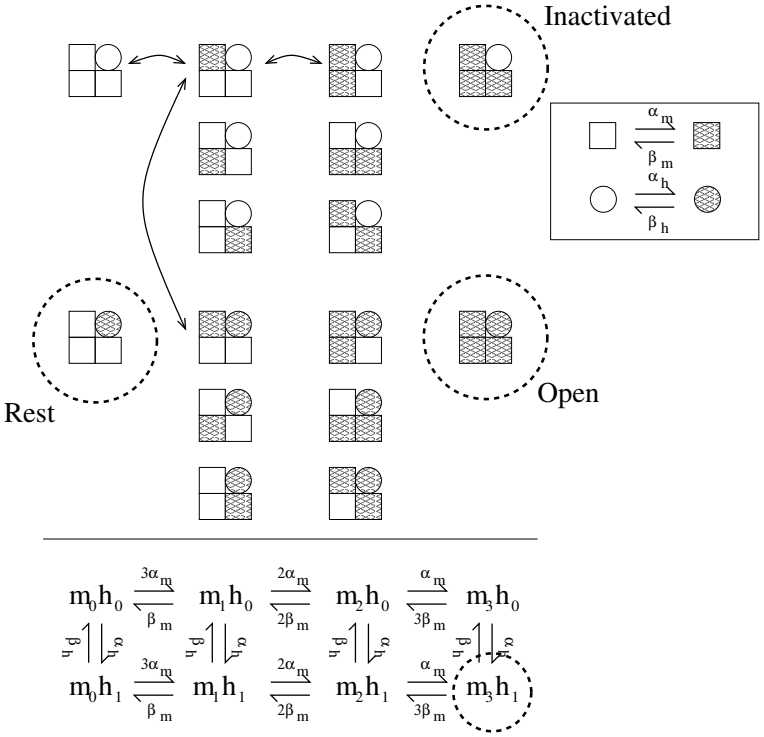


Figure 3.2: Upper: The 16 possible states of the Na channel. The 4 gating variables can each be in off-state (white symbols) or on-state (grey symbols). There are 3 'm' gates (squares) and 1 'h' gate (circle). In the transitions between the states, one symbol changes colour. For clarity all possible transitions of only one state are shown. All states but one correspond to a closed channel. Bottom: The Markov diagram can be simplified to 8 distinct states. Note that the rate-constants have to be changed as indicated to account for the combinatorics (check for yourself).

the h have to be switched on. The probability of a switch m to go from 0 to 1 in a time-step is $\alpha_m(V)$, and from 1 to 0 it is $\beta_m(V)$. We will describe α and β in detail below.

We can write down a Markov state diagram for a single Na channel. There are 4 gates in total (3 m's, and 1 h), which each can be independently in the up or down state. So in total there are $2^4 = 16$ states, where one of the 16 states is the open state, Fig 3.2 top. However, in the diagram it makes no difference which of the m gates is activated, and so it can be reduced to contain 8 distinct states, Fig. 3.2 bottom.

3.1.1 Many channels

The channels that carry the current open and close stochastically. The Markov diagrams are useful when we think of microscopic channels and want to simulate stochastic fluctuations in the current. However, when there is a large number of channels in the membrane, it is possible to average all channels and study just the mean conductance.

Intermezzo:

Consider a simple reversible chemical reaction in which molecule A is turned into an different form A* and back into A. If there are a lot of molecules it makes sense to consider the concentration $[A]$

$$[A] \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} [A^*]$$

The rate equation for the reaction is: $d[A^*]/dt = -\beta[A^*] + \alpha[A]$. Normalising without loss of generality such that $[A] + [A^*] = 1$, we have: $d[A^*]/dt = -\beta[A^*] + \alpha(1 - [A^*])$. This is very similar to what we

have for the gating variables. The solution to this differential equation is exponential, like for the RC circuit. If at time 0, the concentration of A^* is $[A^*]_0$, it will settle as follows

$$[A^*](t) = [A^*]_\infty + ([A^*]_0 - [A^*]_\infty)\exp(-t/\tau)$$

where the final concentration $[A^*]_\infty = \alpha/(\alpha+\beta)$, and the time-constant $\tau = 1/(\alpha+\beta)$ determines how quickly it equilibrates. (Check for yourself)

In the limit of many channels one still has

$$g_{Na}(V,t) = \bar{g}_{Na}\bar{m}^3(V,t)\bar{h}(V,t) \tag{3.2}$$

where $\bar{g}_{Na}$ is the total maximal sodium conductance per area, i.e. when all channels were open ($0.12\ S/cm^2$ in HH). $\bar{m}$ denotes the average over many m. The gating variables now describe the probability that a gate is in the 'on' state, with values ranging between 0 and 1. They evolve as

$$\begin{aligned} \frac{d\bar{m}(V,t)}{dt} &= \alpha_m(V)[1-\bar{m}(V,t)] - \beta_m(V)\bar{m}(V,t) \\ \frac{d\bar{h}(V,t)}{dt} &= \alpha_h(V)[1-\bar{h}(V,t)] - \beta_h(V)\bar{h}(V,t) \end{aligned} \tag{3.3}$$

Note that the gating variables depend both on time and voltage, but the transition rates only depend on voltage.

This is the way the original HH theory was formulated; the gating variable are continuous real quantities, not switches. In contrast, in the stochastic version, one has a discrete number of channels each with activation variables (h, m, n) that flip between on and off states, this in turns leads to a flickering of the conductances. The rate constants give the probability that they flip. In the limit of a large number of channels, the stochastic description matches of course the continuous one. The importance of the channel noise on the spike timing reliability is not fully known. It is not a big effect, but is probably not completely negligible (Rossum, O'Brien, and Smith 2003; O'Donnell and Van Rossum 2014). As seen in Fig 3.4, already for 100 channels the noise is quite small.

If the rate-constants were true constants, the equilibrium would be reached and that would be it (assuming that there was no other way to change, remove or add m or h). The interesting part for the voltage gated channel is that the rate constants depend on the voltage across the membrane. Therefore, as the voltage changes, the equilibrium shifts and the gating variables will try to establish a new equilibrium. The equilibrium value of the gating variable is

$$\bar{m}_\infty(V) = \frac{\alpha_m(V)}{\alpha_m(V)+\beta_m(V)}$$

and the time-constant is

$$\tau_m(V) = \frac{1}{\alpha_m(V)+\beta_m(V)}$$

The rate constants need to be established empirically. For the squid axon they were determined by Hodgkin and Huxley to be (in units of 1/ms; voltage in mV relative to $V_{rest} = -65$mV)

$$\begin{aligned} \alpha_m(V) &= \frac{25-V}{10[e^{0.1(25-V)}-1]} \\ \beta_m(V) &= 4e^{-V/18} \\ \alpha_h(V) &= 0.07e^{-V/20} \\ \beta_h(V) &= \frac{1}{1+e^{0.1(30-V)}} \end{aligned} \tag{3.4}$$

In Fig. 3.3 the equilibrium values and time-constants are plotted. You can see in the right panel that the m variable is much faster than the others (and for practical purposes you can even replace

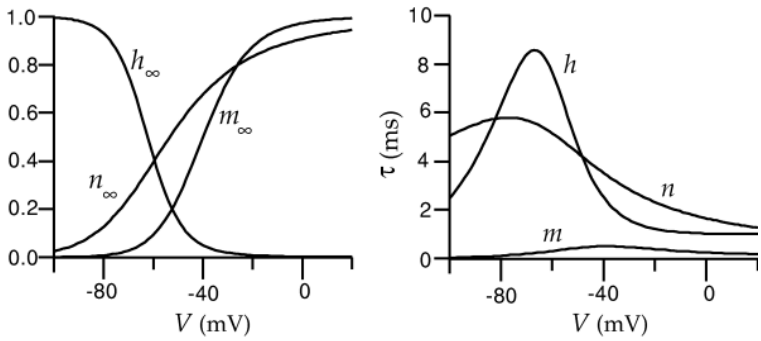


Figure 3.3: Left: The equilibrium values of the gating variables. Note that they depend on the voltage. Also note that the inactivation variable $\bar{h}$, switches off with increasing voltage, whereas $\bar{m}$ switches on. Right: The time-constants by which the equilibrium is reached. Note that $\bar{m}$ is by far the fastest variable. FromDayan and Abbott 2002.

it by its equilibrium value at each voltage). Note that the channel dynamics are prescribed either by specifying $\{\alpha_m(V), \beta_m(V)\}$ and using Eq. 3.3, or $\{\bar{m}_\infty(V), \tau_m(V)\}$ and $\tau_m \frac{d\bar{m}}{dt} = \bar{m}_\infty - \bar{m}$. These descriptions are equivalent as you should be able to easily show.

Importantly, the $\bar{m}$ opens with increasing voltage, but $\bar{h}$ closes with increasing voltage, Fig. 3.3 left. The m is called an activation variable and $\bar{h}$ is an inactivating gating variable. The inactivation causes the termination of the Na current. But because the inactivation is much slower than the activation, spikes can grow before they are shut down.

The K channel works in a similar way as the Na channel, but it hyperpolarizes the membrane. Its conductance can be written as

$$g_K(V,t) = \bar{g}_K \bar{n}^4(V,t)$$

The corresponding state diagram is in Fig. 3.4. The difference with the Na channel is that the K current does not inactivate. As long as the membrane potential remains high, the K conductance will stay open, tending to hyper-polarise the membrane. The gating variable $\bar{n}$ also obeys $d\bar{n}/dt = \alpha_n(1-\bar{n}) - \beta_n \bar{n}$. The rate constants are

$$\begin{aligned} \alpha_n &= \frac{(10-V)/100}{\exp[0.1(10-V)]-1} \\ \beta_n &= 0.125\,\mathrm{e}^{-V/80} \end{aligned}$$

As can be seen in Fig. 3.3, the K conductance is slow. This is important as it allows the Na to raise the membrane potential before the K kicks in and hyper-polarises the membrane. In the Hodgkin-Huxley model there is substantial overlap between the inward Na and outward K current, this is wasteful as the voltage does not change much, but ions will have to be pumped out anyway. It turns out that in actual neurons this overlap, and hence energy waste is very small (Alle, Roth, and Geiger 2009).

3.1.2 A spike

To obtain the voltage equation, we apply Kirchhoff's law again and collect all currents: the capacitive, the leak, Na, and K currents. The Hodgkin-Huxley equation for the voltage is therefore

$$c_m \frac{dV(t)}{dt} = -g_{leak}[V(t) - E_{leak}] - g_{Na}(V,t)[V(t) - E_{Na}] - g_K(V,t)[V(t) - E_K] + I_{ext}$$

or, filling in the definitions of g_{Na} and g_K

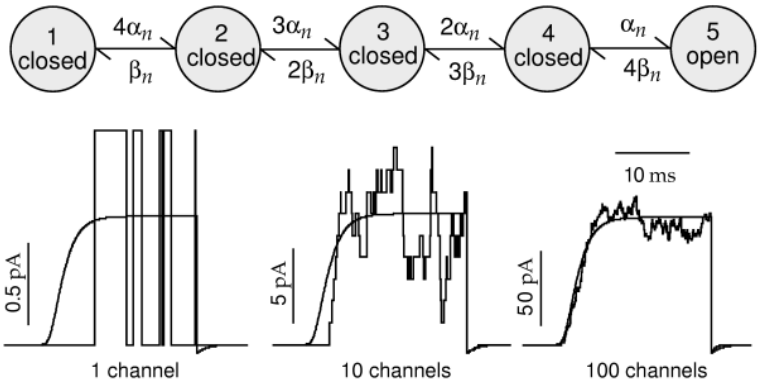


Figure 3.4: Top: State diagram for the K channel. Bottom: the K current for a limited number of channels. Left: 1 channel; right 100 channels. The smooth line is the continuous behaviour. Taken fromDayan and Abbott 2002.

$$c_m \frac{dV(t)}{dt} = -g_{leak}[V(t) - E_{leak}] - \bar{g}_{Na} m^3(V,t) h(V,t)[V(t) - E_{Na}] - \bar{g}_K n^4(V,t)[V(t) - E_K] + I_{ext}$$

Note that the g_{leak} replaces $1/R_m$ of the previous chapter. The electrically equivalent circuit is shown in Fig. 3.5, right. In this circuit diagram a lot of the nasty detail is hidden in the variable conductances that as we have seen depend on voltage and time.

The full Hodgkin Huxley model consists of this equation, the equations for the Na current: Eqs. (3.2),(3.3) and (3.4), and the similar equations for the K current! The Hodgkin-Huxley model is complicated, and no analytical solutions are known. It is a four-dimensional, coupled equation (the dimension are V, h, m, n). To solve it one has to numerically integrate the equations. The simplest (but not the most efficient or accurate) way would be: initialise values for voltage (and the gating variables) at time t. Next, we like to calculate the voltage a little bit later, $t + \delta t$. First calculate the rate constants at the current voltage, $\alpha(V(t))$, $\beta(V(t))$. From this calculate the new values of the gating variables, e.g. $m(t + \delta)$. From this follows the Na and K conductance. Now the new value of the potential $V(t + \delta t)$ can be calculated. Repeat this for the next time-step.

Fig. 3.5 shows a simulated spike. Note the sequence of the processes: first m, later followed by n and h. Remember that the passive membrane is slow. The spike generation can be so fast, because the membrane is made very leaky during a short time. The currents through the voltage gated channels are much larger than the stimulus current. At rest small currents are sufficient to change the voltage, but during the spike, the membrane is leaky and much larger currents flow.

Despite the complex mathematics, spikes have practically always the same amplitude. This is not mathematically strictly true, and also data shows that the height is not perfectly the same across spikes, and during continued stimulation sodium inactivation typically reduces the amplitude (Fig. 3.1). But in practice we don't think such variations matter much.

The above description is for a single compartment. If one wants to describes propagation of the spike in the axon, one has to couple the compartments like we did in the cable equation. The equation holds for each compartments in the axon, so for N compartments we have $4N$ equations...

To simplify the phenomena, Hodgkin and Huxley inserted a thin silver wire along the axon, this ensured that the voltage was equal across the full axon ($r_i = 0$). That way the spatial dependence was eliminated (sometimes called a SPACE-CLAMP). Furthermore, the voltage of the axon was controlled, a so-called VOLTAGE-CLAMP. In a voltage clamp the voltage is controlled with a feedback circuit: if the voltage starts to deviate from the desired value, current is injected to bring it back. This 'clamp' current compensates to leak, Na and K currents at a given voltage. Without the regenerative mechanism in place, the rate constants could be measured accurately.

When the silver wire was removed, the axon is allowed to run"free". The spikes now travel on

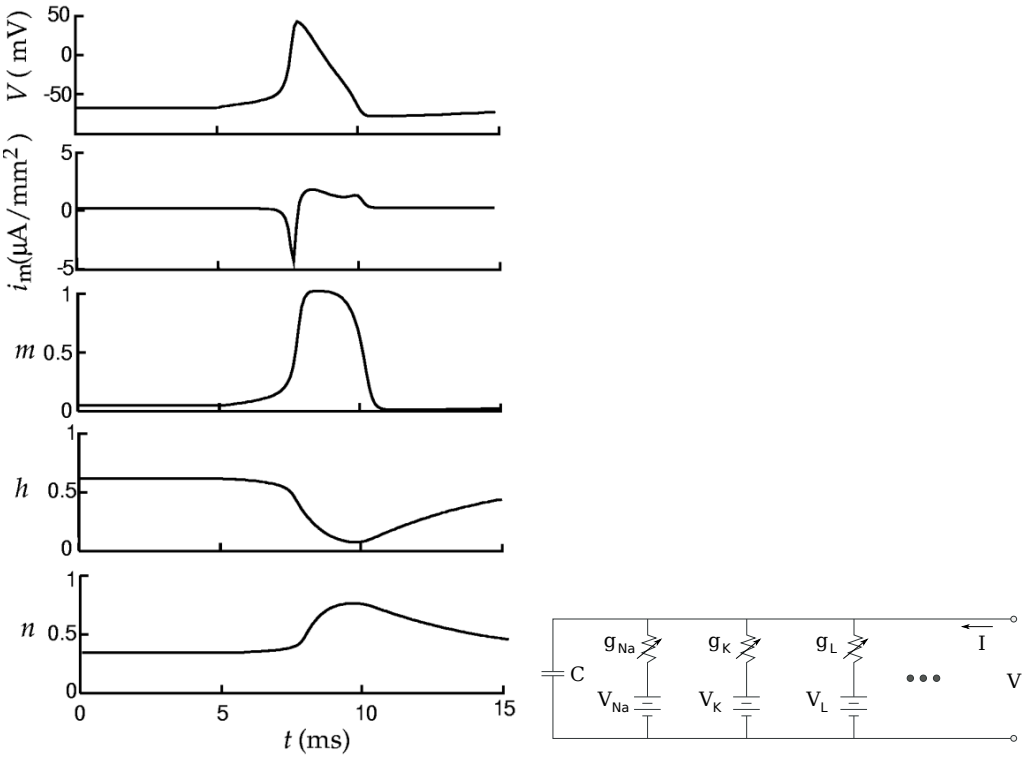


Figure 3.5: Left:Spike generation in a single compartment Hodgkin-Huxley model. The cell is stimulated starting from $t = 5$ms. Top: voltage trace during a spike. Next: Sum of Na and K current through membrane (stimulus current not shown). Bottom 3: the gating variables. Note how m changes very rapidly, followed much later by h and n. FromDayan and Abbott 2002.
Right: equivalent circuit of the HH equations (from Wikipedia). A lot of the complexity is hidden here, because the conductances (g_{na} and g_K) depend on voltage and time.

the axon. The equations predict the speed of the propagation of the spike and this speed closely matched the speed observed. This provided an important check of the model.

Final comments:

- The kinetics of the HH model depend strongly on the temperature. The original experiments were carried out at 6.3C. Usually temperature dependence is expressed with Q_{10}, which describes how much quicker a certain reaction goes when the temperature is 10C higher. A Q_{10} of 3 means that when going from 6C to 36C, the reaction goes 27 times faster. To understand the effect on the spike generation, consider voltage-component of the space-dependent Hodgkin-Huxley equations:

$$c_m \frac{dV(x,t)}{dt} = \frac{d}{4}\frac{1}{r_i}\frac{d^2V(x,t)}{dx^2} - \sum_j g_j(V,t)[V(x,t) - E_j] - g_{leak}[V(x,t) - E_{leak}] + I_{ext} \quad (3.5)$$

  where r_i is the axial resistivity and g_j are the various ionic conductances densities. At 36^oC the channel dynamics speed up to $g'_j(V,t) = g_j(V,qt)$ with $q = 27$, that is, spikes are faster. But in terms of this scaled time qt, the capacitance term becomes larger, $C\frac{dV}{dt} \rightarrow qC\frac{dV}{d(qt)}$, and as a result the spikes are damped more (Huxley 1959). In other words, at high temperature there is no time to charge to membrane. This means that higher channel densities are required for a model to work both at 6C and at 35C than a model adjusted to work just at 6C. Consistent with this the firing rate increase at higher temperatures. In addition the spike

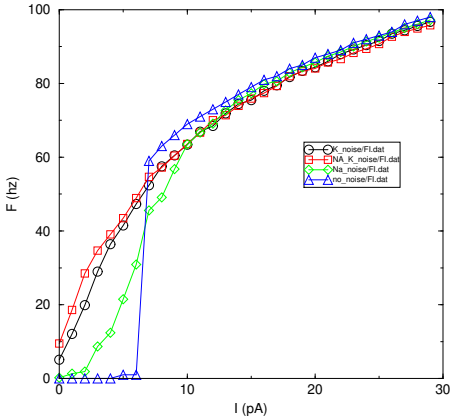


Figure 3.6: FI curve for the Hodgkin Huxley model. Also shown the effect of channel noise on the FI curve. This a small (100 μm^2) patch, so channel noise is substantial. (MvR unpublished).

amplitude decreases at higher temperatures. This can easily be checked in simulations. Note, this is not only quantity depending on temperature, single channel conductances are for instance also temperature dependent.

- We have approximated the gates as operating independently. There is no a priory reason why this should be the case, but it works reasonably. Nevertheless, more accurate models of the channel kinetics have been developed. Readable accounts can be found in Hille's book (Hille 2001).

3.1.3 Repetitive firing

When the HH model is stimulated with the smallest currents no spikes are produced. For slightly bigger currents, a single spike is produced. When the HH model is stimulated for a longer time, multiple spikes can be produced in succession, Fig. 3.1. Once the current is above threshold, the more current, the more spikes. The firing frequency vs. input current (the FI-curve) shows a sharp threshold. In more elaborate neuron models and in physiology one finds often a much more linear FI curve, usually without much threshold (i.e. firing frequencies can be arbitrary low). Noise is one way to smooth the curve and obtain such FI curves, see Fig. 3.6, but additional channel types, and the cell's geometry can also have an effect.

3.2 Other channels

Although the Na and K channels of the HH model are the prime channels for causing the spike, many other channel types are present. The cell can use these channels to modulate its input-output relation, regulate activity its activity level, and make the firing pattern history dependent (by adaptation). As the knowledge of neurons becomes more refined, more channels are being discovered. There appears to be a great variety of K channels in particular (some 100 types vs some 10 types of Na channels).

3.2.1 KA

The KA current (I_A) is a K current that inactivates at higher voltages. This seems a bit counter-productive as one could expect that the higher the membrane voltage, the more important it is to counteract it with a K current.

The effect of KA current on the firing is as follows: Suppose the neuron is at rest and a stimulus current is switched on. At first the KA currents still are active, they keep the membrane

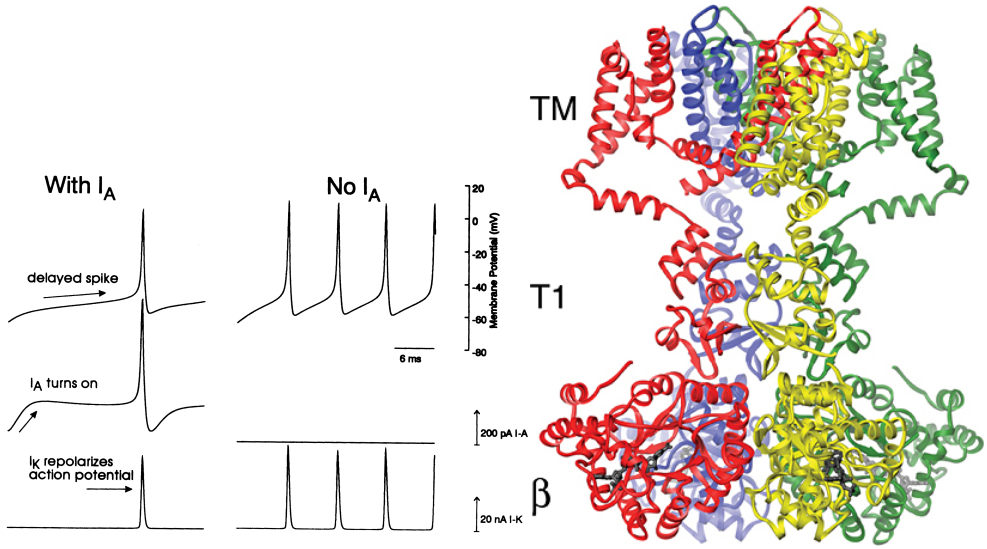


Figure 3.7: Left: The effect of KA currents on spike generation.
Right: protein structure of a potassium ion-channel. The top will be sticking out of the membrane.

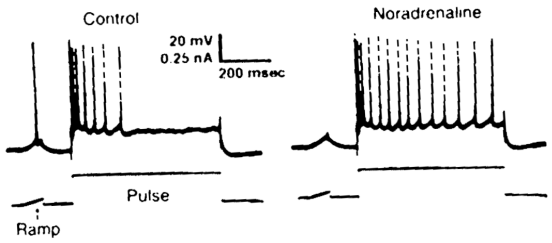


Figure 3.8: Spike frequency adaptation. Left: Normal situation, a step current is injected into the soma; after some 100ms spiking stops. Right: Noradrenaline reduces the KCa current, thus limiting the adaptation.

relatively hyper-polarised. As the KA channels inactivate, the voltage increases and the spike is generated. The KA current can thus delay the spiking. Once repetitive spiking has started, the same mechanism will lower the spike frequency to a given amount of current.

3.2.2 The I_H channel

The I_H channel carries a mainly K current. But the channel also lets through Na and therefore its reversal potential is at about -20mV. Interestingly, the channel only opens at hyper-polarised (negative) membrane voltages, thus preventing prolonged hyper-polarisation. Because it activates slowly (hundreds of ms), it can help to induce oscillations: If the neuron is hyper-polarised it will activate after some delay, depolarising the cell, causing spikes, next, I_h will deactivate and as other currents such as KCa (below) will hyper-polarise the cell, and spiking stops. Now the whole process can start over again.

3.2.3 Ca and KCa channels

Apart from K and Na ions, also Ca ions enter the cell through voltage-gated channels. A large influx occurs during a spike. There is a large variety of Ca channels. As we will see below, one important other role for the Ca channel is to cause Ca influx at the pre-synaptic terminal. But

also in the rest of the cell, the Ca signal appears as an important activity sensor. As the Ca concentration at rest is very low, the Ca concentration changes significantly during periods of activity (this contrasts to Na concentration, which does not change much during activity). In the long run, the Ca is pumped out of the cell again. The time-constant of the extrusion is some 50 ms. This makes the Ca concentration an activity sensor.[3]One important conductance in the cell are the so-called Ca-activated K channels (or KCa channels). These channels all require internal Ca to open, but in addition some have a voltage dependence. The KCa currents cause SPIKE FREQUENCY ADAPTATION. As the internal Ca concentration builds up, the KCa current becomes stronger, hyper-polarising the cell, the spiking becomes slower or stops altogether, Fig. 3.8.

From a theoretical point of view adaptation is interesting: In first approximation it is a high-pass filtering effect. But it is not purely a linear high-pass filtering of the firing rates. If that were the case, cells responding with a low firing rate would adapt as much as fast firing cells. But because adaptation depends on the amount of Ca that flows in, fast firing cells adapt more strongly (see Benda and Herz (2003) and Liu and Wang (2001) for a mathematical model). The precise consequences remain to be examined.

Most excitatory neurons have strong spike frequency adaptation (inhibitory ones much less so). On many levels one can observe that the brain likes change: When an image is projected steadily on the retina, the percept disappears after a few minutes. (By continuously making eye-movements, known as saccades, this effect does not occur in daily life.) Also on the single cell level, changes usually have a larger effect on the output than steady state conditions.

3.2.4 Bursting

Some cells, for example in the thalamus, show bursting behaviour. In a burst, a few spike are generated in a short time. The burst often continues after the stimulation is stopped. Ca channels are thought to be important for burst generation. The computational role of bursts is not clear.

3.2.5 Leakage channels

You might not be surprised that also the leak conductance is partly mediated through a channel. These are mainly Cl channels without much voltage dependence. Without any channels, the membrane would have a very high resistivity, i.e. a large r_m.

3.3 Spatial distribution of channels

Different voltage gated channels populate not only different types of neurons but also different locations of the neuron. It used to be thought that the dendrites where largely passive cables, but more recently one has observed active channels on the dendrites. Thick dendrites can be patched, and differential distributions of channels have been found (Hoffman et al. 1997). Active dendrites can reduce the attenuation of distal inputs on the soma, and also can support back-propagating spikes which are important for plasticity.

The standard picture for an action potential in a cortical neuron is as follows: Synaptic input arrives in the dendrites. Signals travel sub-threshold (i.e. without causing a spike) to the soma. It is possible that the EPSPs activate or deactivate some currents; this can lead to a boosting of the EPSP. Once enough input has been collected, a spike is generated in the axon-hillock. It now appears that Na channels at the axon hillock have a lower threshold than Na channels on the rest of the neuron (Colbert and Pan 2002). Furthermore, the density of channels there is actively regulated (Grubb and Burrone 2010). This elects the axon-hillock as the prime spike generation site. The spike travels down the axon, and also back into the dendrite. The back-propagation action potential (BAP) is probably not full-fledged and can have a substantial Ca component.

[3]The large fluctuations in Ca concentration require a special treatment. Buffers and the diffusion of Ca inside the cell should be taken into account. More details in (Koch 1999; Schutter and Smolen 1998).

3.4 Myelination

Axons that have to transmit their spikes over large distances are covered with a sheet of myelin. This highly insulating fat layer, reduces both the leak and the capacitance by a factor of about 250, speeding up the propagation. The myelin covering is interrupted ever so often along the axon. At these nodes one finds a very high density of sodium and potassium channels that boost the spike. For unmyelinated axons the propagation speed is about 1m/s and proportional to $\sqrt{d}$, while for myelinated fibres the speed is about 50 m/s (and proportional to diameter d). The actual speed depends in a complicated manner on the parameters, but can be found numerically. The proportionality to the diameter can be found using dimensional considerations (Koch 1999).

3.5 Remarks on computational models

To model the Hodgkin Huxley equations in axon or any arbitrary structure, the first step is to discretize space (similar to what we did to derive the cable equation. Each compartment should be so small that the voltage in it can be considered as the same (typical $\lambda/20$). Next, one needs to discretize time so that the change in voltage (in each compartment) per timestep is small, Given that the voltage during the action potential chances rapidly, a small time-step is required (0.1ms at the most). Of course, there is a trade-off between accuracy and simulation speed here.

A major problem for modellers is to put all these channel types in the neuron model. For the physiologist it is a lot of work to extract all the parameters the modeller would like to know. The rate constants are particularly difficult to extract from the data. This is easy to see for the Na channel: the only thing which is easy to measure is the kinetics of one open state. If we assume independent gates, the problem is still tractable, but if we drop the independence assumption of the gating variables we have many more rate constants. Secondly, channels can be modulated (changes in sensitivity for instance) by internal mechanisms and neuro-modulators such as dopamine and acetylcholine. The neuron thus has ways to change its excitability, probably not only globally (Desai, Rutherford, and Turrigiano 1999), but also locally (Watanabe et al. 2002), thus giving inputs at particular locations preference. This change of neural properties by changing the channels complements the change in synaptic strength (so called synaptic plasticity) discussed below and has been explored only very sparsely. The variety of voltage gated channels might also be tuned to make the neuron more sensitive to a particular temporal input. (It is interesting to think about learning rules for the channel composition (Stemmler and Koch 1999).)

A third challenge for the modeller is that the distribution of the channels on the neurons is very hard to determine experimentally. And finally, there is large heterogeneity between neurons. As a result it will be very hard to come up with a precise, gold-standard model of 'the neuron'. It is hoped that it will be possible to address these questions in the near future.

On the other hand when we drop our quest for the standard model, perhaps the modelling situation is not that depressing: The spikes are usually quite robust (you can try parameter variation in the tutorials) and the spikes are an all-or-none event. So that models are not sensitive to slightly incorrect parameter settings.

More Reading: For mathematical aspects of biophysics, Koch (1999) and Johnston and Wu (1995),
everything about channels including an excellent review of the work of Hodgkin and Huxley, see Hille (2001).

Chapter 4

Synaptic Input

So far we have studied artificially stimulated neurons, which responded because we injected a current into them. But except for primary sensory neurons, such as photo-receptors or hair-cells (all using specialized channels to react to stimuli), input comes from other cells in the brain. Synaptic contacts communicate the spikes in axon terminals of the pre-synaptic cell to the dendrites of the post-synaptic cell. The most important are CHEMICAL synapses, in which transmission is mediated by a chemical, called a NEURO-TRANSMITTER.

Another type of connection between neuron are the so called GAP-JUNCTIONS, also called ELECTRICAL SYNAPSES. These synapses are pore-proteins, i.e. channels, and provide a direct coupling of small molecules and ions (and hence voltage!) between cells. Although gap junctions are abundant in some brain areas, their function is less clear. One clear role for gap-junctions exists in the retina where they determine the spatial spread of electric activity. At different light levels, dopamine regulates the gap junction conductivity and hence the spatial filtering of the signal (see below for adaptation). Another role has been implied in the synchronisation of spikes in hippocampal inter-neurons, e.g. (Tam?s et al. 2000). We will not consider gap-junctions any

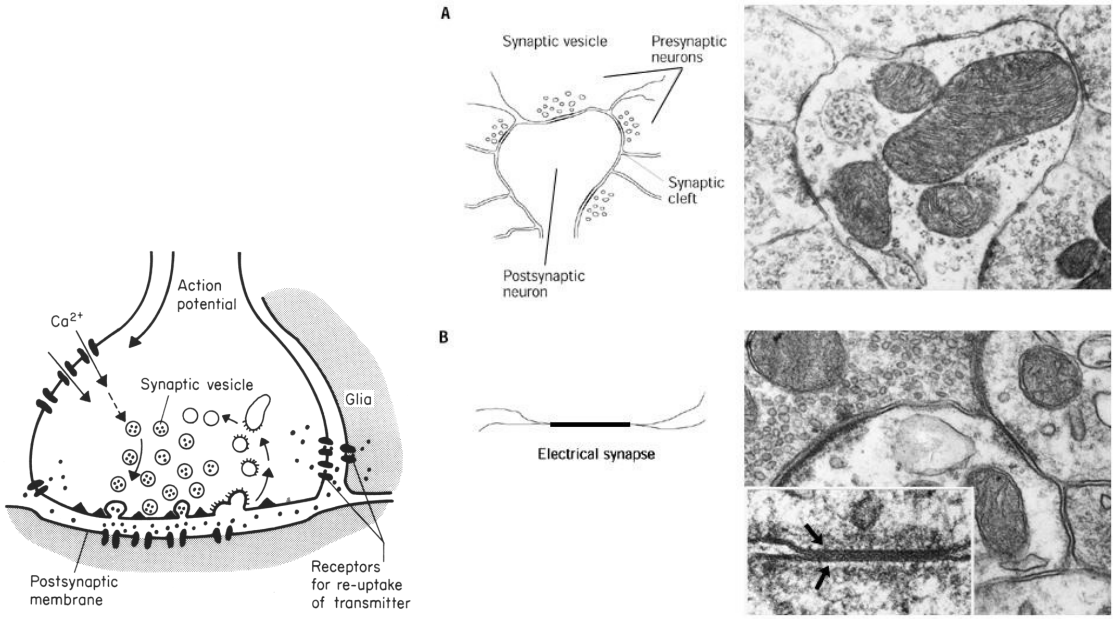


Figure 4.1: Left: Schematic of a chemical synapse. Right top: Electron micrograph of a (chemical) synapse. Right bottom: Electron micrograph of an electrical synapse or gap-junction.

further and follow the common convention that by synapse we mean actually a chemical synapse.

Because synapses are the main way neurons communicate and how the networks dynamically change, it is important to know their behaviour.

4.1 AMPA receptor

We look in detail at the AMPA receptor, which is the main excitatory synapse type in the brain. On the pre-synaptic side of the synapse we find a pool of vesicles, Fig. 4.1. Suppose a spike is generated in the soma of the pre-synaptic cell and travels along the axon. When the spike arrives at the axon terminal, the spike in the membrane voltage causes the opening of voltage gated Ca channels (discussed in last chapter). The Ca flows into the terminal and triggers the release of one or a couple of vesicles. The vesicles are little balls filled with neurotransmitter. When released, the vesicle membrane fuses with the cell membrane the neurotransmitter goes into the extracellular space. The transmitter diffuses into the synaptic cleft and binds to the synaptic receptors on the post-synaptic side.[1]

Above we saw that the excitability of neurons is due to voltage-gated channels. Synaptic input to cells is mediated through channels as well: synaptic channels. Synaptic channels are opened when they bind neurotransmitter. Like the voltage-gated channels, these channels let through ions selectively. On the post-synaptic side excitatory synapses have channels that let through both K and Na; the reversal potential is about 0mV. Inhibitory synapses contain mostly Cl channels, with reversal of -60mV. Unlike the voltage-gated channels important for spike generation, the opening and closing of these channels does not depend on membrane voltage but instead on the binding of neurotransmitter to the channel. The excitatory transmitter is the amino-acid glutamate, Glu (you can buy glutamate in Chinese supermarkets under the name MSG, mono-sodium glutamate). The name AMPA receptor comes from the artificial compound AMPA that also opens these receptor channels.

The diffusion of transmitter across the cleft is very fast. Binding and opening of the AMPA channel are also fast so that the dynamics of the response are largely determined by the unbinding of transmitter from the receptor. In Fig. 4.2 we show an approximate state diagram for the AMPA receptor (these state diagrams are difficult to measure, the shown version fits the data well but is probably not the final word). Few remarks are in order. The transition in the state diagram in the rightward direction depend on the presence of transmitter (cf. the voltage-gated channels where the voltage changed the transition rates). Two glutamate molecules are needed to open the channel. This is called cooperativity. If the glutamate concentration is far from saturation, the probability of opening will therefore be quadratic in the glutamate concentration. Secondly, the AMPA receptor has desensitised states. These becomes most apparent when a long lasting puff of glutamate is delivered to the synapse with a pipette. The initial current is large, but the current desensitises and reduces to a lower steady state value. The same might occur *in vivo* when the synapse is stimulated repeatedly.

The jump in transmitter concentration is very brief (about 1ms). The time course of the synapse is mainly determined by the time-constant of Glu unbinding from the synapse. Model and data are shown in Fig. 4.3. The whole process can reasonably modelled by a single exponential with a time constant τ_{AMPA}. Its value depends on the temperature used in the recording which is often below the physiological temperature (the higher, the faster) and the age of the animal (the older, the faster), values between 1 and 5ms are common (Roth and Rossum 2009). We thus have for the current, assuming a pulse of Glu at $t = 0$,

$$I_{AMPA} = g(t)\,(V_{mem} - E_{AMPA}) \approx g_0 \mathrm{e}^{-t/\tau_{AMPA}}(V_{mem} - E_{AMPA})$$

where E_{AMPA} is the AMPA reversal potential, which is, conveniently, about $0mV$. The g_0 is the synapse's peak conductance. The peak conductance is depends on a number of factors we discuss below.

[1] The distinction pre- and post-synaptic is useful when talking about synapses. However, note this term does not distinguish cells; most neurons are post-synaptic at one synapse, and pre-synaptic at another synapse.

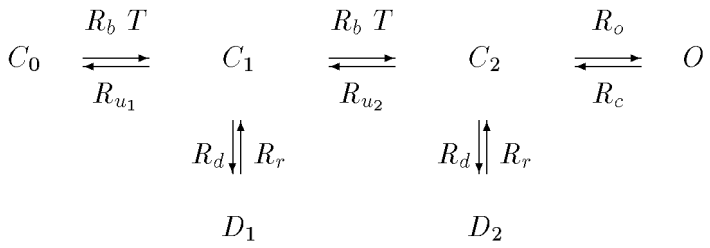


Figure 4.2:
State diagram for AMPA receptor. Its response is shown in the next figure. Parameters for the state diagram are: $R_b = 13 \times 10^6 M^{-1}s^{-1}$, $R_d = 900s^{-1}$, $R_r = 64s^{-1}$, $R_o = 2.7 \times 10^3 s^{-1}$ and $R_c = 200s^{-1}$. T denotes the transmitter. O the open state, C closed states, and D de-sensitised states.From Destexhe, Mainen, and Sejnowski 1998.

Like the voltage-gated channels, the channel openings and closings are stochastic events; they occur randomly. Using patch-clamp recording these fluctuations can be measured, and the single channel conductance has been determined and is about 10-100 pS for most channel types (Hille 2001). The number of receptors is usually somewhere between 10 and a few hundred per synapse for central synapses.

4.2 The NMDA receptor

Like the AMPA receptor, the NMDA receptor opens in the presence of glutamate. Commonly, one finds the AMPA and NMDA receptor at the same synapse and both have a reversal potential of around $0mV$. However, there are a few important differences:

1) The current through the NMDA channel depends non-linearly on the voltage, Fig. 4.4. The reason for this is Mg block. At voltages close to the resting potential an extracellular Mg ion blocks the pore of the NMDA channels. At higher membrane voltages, the attraction from the membrane potential on the Mg ion is less strong. The block is partly relieved and current can flow (the current shows short flickers). Alternatively, one can remove all Mg from the solution and the block is absent, Fig. 4.4. An important consequence is that the post-synaptic current depends on the post-synaptic cell voltage. This makes NMDA a coincidence detector, only opening when pre-synaptic activity is combined with post-synaptic activity. This will turn out to be important for plasticity, Chap. 13.7. The Mg-block can be fitted to

$$B(V) = \frac{1}{1 + \exp(-0.062V)[Mg]/3.57}$$

where the Mg concentration is measured in mM and voltage is measured in volts.

2) The time course of the NMDA current is much longer than AMPA, Fig. 4.3. The NMDA response comes on, after the AMPA response has largely decayed. It decays back with a time-constant of some 100ms. This long time-constant has helped modellers to build networks which have slow dynamics. The slower dynamics help to stabilise attractor states, see Chap. 10.

3) The difference in the dynamics is caused by a much slower unbinding of Glu from the receptor (the binding of Glu to the receptor is similar). As a result, and some NMDA receptors can still be occupied when the next Glu pulse arrives, which can cause saturation. Another consequence is that the NMDA receptor is much more sensitive than AMPA when long puffs of Glu are applied.

4.2.1 LTP and memory storage

It is believed that memory is stored in the synaptic strengths. There is a variety of electrophysiological protocols which cause long term changes in synaptic strength. For instance, long

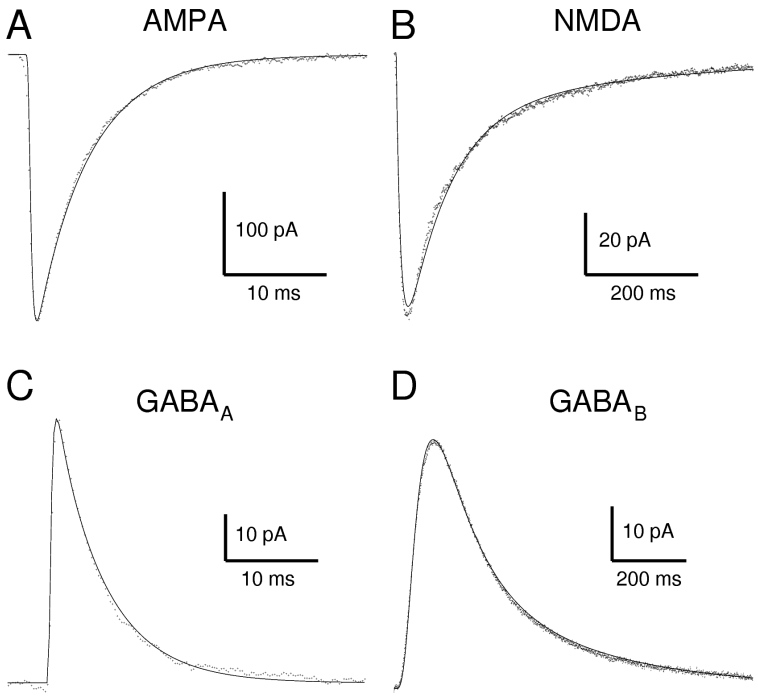


Figure 2:
Best fits of detailed kinetic models to averaged postsynaptic currents obtained from whole-cell recordings. A. AMPA/kainate-mediated currents (obtained from Xiang et al., 1992; recorded at 31 °C). B. NMDA-mediated currents (obtained from Hessler and Malinow, 1993; recorded at 22-25 °C in Mg^{2+}-free solution). C. $GABA_A$-mediated currents. D. $GABA_B$-mediated currents (C-D recorded at 33-35 °C by Otis et al., 1992; 1993). For all graphs, the averaged recording of the synaptic current (noisy trace) is represented with the best fit obtained using the models (continuous trace). Models are described in the text; transmitter release was modeled as in Fig. 1. D modified from Destexhe and Sejnowski, 1995; fitting procedures described in Appendix B).

Figure 4.3:
Comparison of the most important types of synaptic responses. Note the difference in time-scales (both NMDA and GABAb are slow).From Destexhe, Mainen, and Sejnowski 1998

term potentiation (LTP) can be induced for instance by firing the pre- and post synaptic cell at a high frequency (we will discuss models of LTP later).

It is thought that somewhere at the AMPA/NMDA synapse memory storage is expressed. However, if this simply mediated by an increase in the number of receptors is not yet clear. Other possibilities would be: to release more Glu per vesicle, to increase the open time, to increase to open conductance, to increase the open probability, or to increase release probability, change the anatomical structure, etc... It is interesting to note that the AMPA receptors are rapidly recycled, old receptors are taken out of the membrane, new ones inserted (time constant on the order of one day). This seems a challenge for the long term storage of memory in the synaptic weights. Somehow, the rapid cycling does not interfere with storage.

Most channels are constructed of multiple SUBUNITS. For AMPA, four or five subunits together are assembled to build a channel. There are multiple types of subunits, labelled GluR1 to GluR6. By changing the subunits composition the properties of the receptor, such as time-constant and sensitivity, can be modulated. This seems of particular importance during development. But one

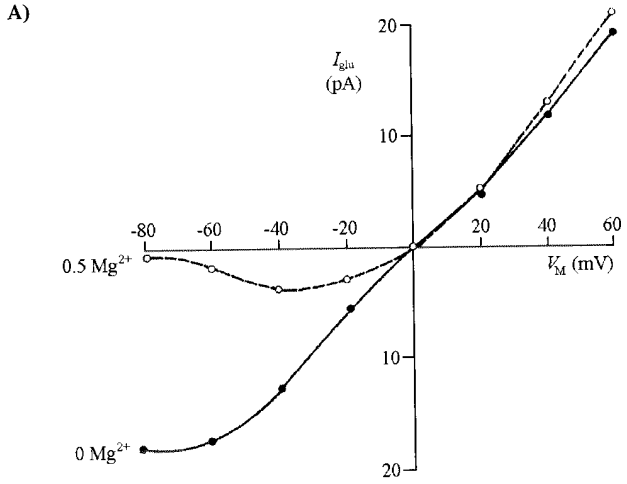


Fig. 4.8 VOLTAGE DEPENDENCY OF THE NMDA CURRENT Current-voltage relationship associated with the NMDA receptor. Because of the action of Mg^{2+} ions in blocking the underlying channel, the associated postsynaptic conductance increase is voltage dependent, different from other fast synaptic inputs. **(A)** Current-voltage relationship of the NMDA current in the absence and presence of 0.5 mM magnesium. While the current behaves relatively ohmic in the absence of magnesium, under physiological concentrations of Mg^{2+} a strong voltage dependency is revealed. Reprinted by permission from Nowak et al. (1984). **(B)** Experimentally recorded normalized NMDA current in a brain slice kept at room temperature (Hessler, Shirke, and Malinow, 1993; heavy jagged line). Superimposed is the current computed from Eq. 4.6, with $\tau_1 = 145.5$ and $\tau_2 = 4.1$ msec. Unpublished data from A. Destexhe, printed with permission.

Figure 4.4: NMDA response is non-linear when Mg is present (which is normally the case). FromKoch 1999.

can not rule out that subunit composition changes during learning later in life. With modern genetic techniques is has become possible to KNOCK OUT certain genes and thereby certain subunits. The channels might still assemble, but their properties will be modified. For instance, it seems that GluR1 is required for LTP induction (Mack et al. 2001). Amazingly, the synaptic channels are turned over and replaced about every hour.

That the NMDA receptor is required for LTP, is much better established. As the NMDA receptor only opens when Glu is present *and* the post-synaptic voltage is high (because of the relieve from the Mg block), the NMDA receptor acts as a coincidence detector, signalling simultaneous activation of both pre- and post-synaptic activity. This is what one would want for LTP. Secondly, the NMDA channel is permeable to Ca. Hence if the channels is open, Ca can flow into the cell. Calcium is thought to be an essential signal for inducing LTP. Blocking the NMDA receptor (with APV) blocks many (but not all) forms of long term plasticity. See (Kandel, Schwartz, and Jessel 2000) for details.

Also the NMDA receptor has varying subunit composition. With the different subunits the time course of the response changes and the amount of Ca influx could change. This has potentially important consequences for LTP threshold.

4.3 GABAa

Inhibitory neurons (neuroffs) only make inhibitory synapses onto other neurons. The fact that neuron make either excitatory or inhibitory synapses, but not both, is called Dale's law. GABA is the main inhibitory transmitter. The transmitter binds to both $GABA_A$ and $GABA_B$ receptors. The $GABA_A$ time-constant is about 5 ms. GABA synapses are often found close to the cell

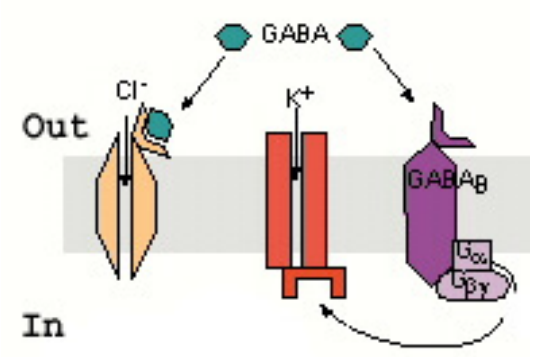


Figure 4.5: Schematic of GABAa and GABAb mechanism. For GABAa the transmitter binds directly to the ion channel which it then opens (left), GABAb uses a second messenger cascade in which the transmitter first binds to a GABAb receptor, which activates a G-protein, which opens a channel (right).

body, whereas excitatory (AMPA) synapses are usually distributed all over the dendritic tree. Counter-intuitively, the inhibitory reversal potential can be close to, or even above the resting potential and still inhibit the cell. This is called shunting inhibition, the inhibitory conductance effectively increases the leak conductance. Because this will make it harder for the neuron the reach the threshold voltage, it has an inhibitory effect.

4.4 Second messenger synapses and GABAb

The GABAb receptors are also inhibitory but work quite differently. The activation of the GABAb receptor does not lead to direct channel opening. Instead, the GABA $_B$ channel activates a G-protein which then diffuses/wanders on the membrane, and opens a potassium channel, Fig. 4.5.

Although second messenger systems don't give fast responses, they are very flexible and can trigger also other events than channel openings. Because they can trigger a cascade of reactions, they can also be very non-linear. The GABAb response is an example of that: longer stimuli give an un-proportionally larger response, Fig. 4.6. Another important second messenger receptor is the metabotropic glutamate receptor mGluR. For reasons not fully clarified, mGluR can often be found in AMPA/NMDA synapses. Furthermore, mGluR receptors can be found *presynaptically*.

4.4.1 Neuro-modulators

Second messengers are common in neurobiology. Many neuro-modulators, such as dopamine, noradrenaline, acetylcholine, serotonin, etc., act through second messenger systems. These modulators are crucial in *modulating* the neural response e.g. through changes in the gain or oscillations as well as regulating processes like plasticity, in accordance to attention, sleep, or reward. Neuromodulators are generally not released in specific neuron-to-neuron connections, instead the sending brain nuclei target most neurons in the receiving area.

4.5 Release statistics

Synapses are stochastic devices in many ways. Even if the pre-synaptic cell is at rest, vesicles are now and then spontaneously released. This leads to small post-synaptic currents. These events are called miniatures, or mini Excitatory Post Synaptic Currents, mEPSC (mIPSC for inhibitory synapses).

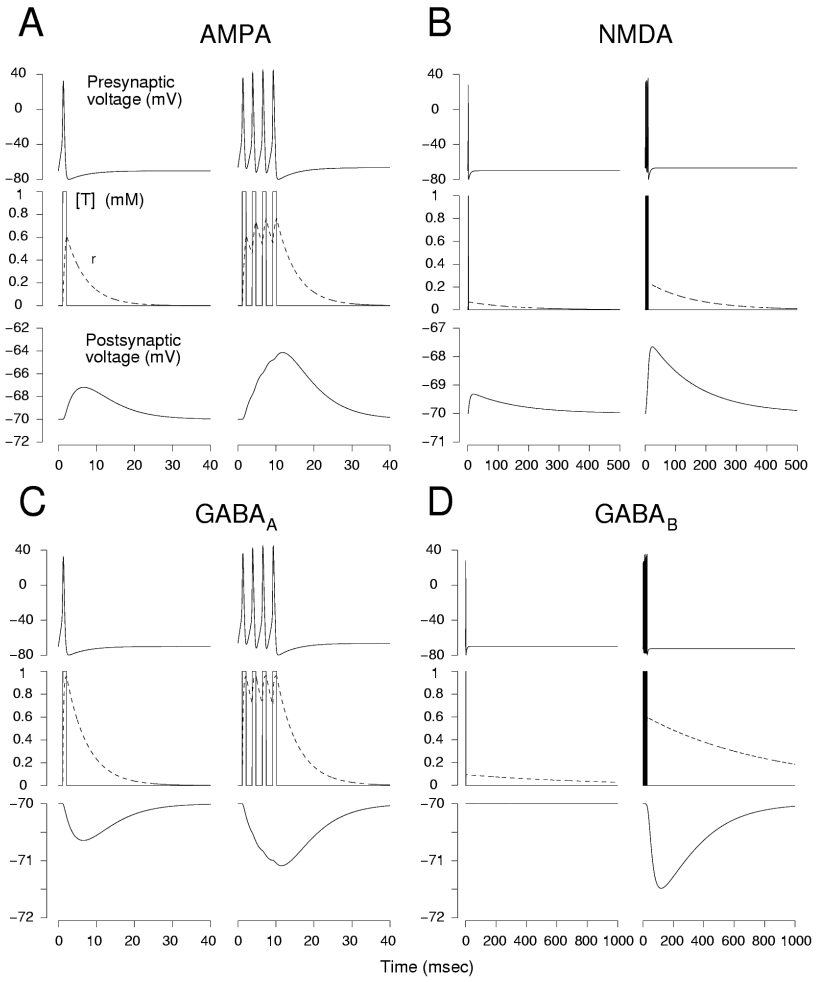


Figure 4:
Summation of postsynaptic potentials in simplified kinetic models of different receptors. A single compartment model (10 μm diameter, 10 μm length, 0.2 mS/cm^2 leak conductance and -70 mV leak reversal) was provided with postsynaptic receptors: A, AMPA/kainate receptors; B, NMDA receptors; C, GABA$_A$ receptors and D, GABA$_B$ receptors. In all cases, the behavior with one presynaptic spike (left panels) is compared with that of a burst of presynaptic spikes at high frequency (300-400 Hz; 4 spikes in A,B,C; 10 spikes in D). All synaptic conductances were of 0.1 nS; other parameters as in Section 4.

Figure 4.6: Comparison of response linearity of the major synapse types. Note that this represents just one aspect of the dynamics. The full dynamics are given by the state diagrams.

When a spike reaches the pre-synaptic terminal, vesicles are released with a much higher probability. But also regular synaptic transmission is random. There are various sources for fluctuations:

- The content and diameter of the vesicle varies. The radii of the vesicles varies (COV on the order of 10%)[2], and hence the amount of transmitter per vesicle varies (assuming a constant concentration). Exercise: Try to figure out the COV in the volume from the radius COV.

[2]COV = coefficient of variation = standard deviation/mean. See chapter 6.

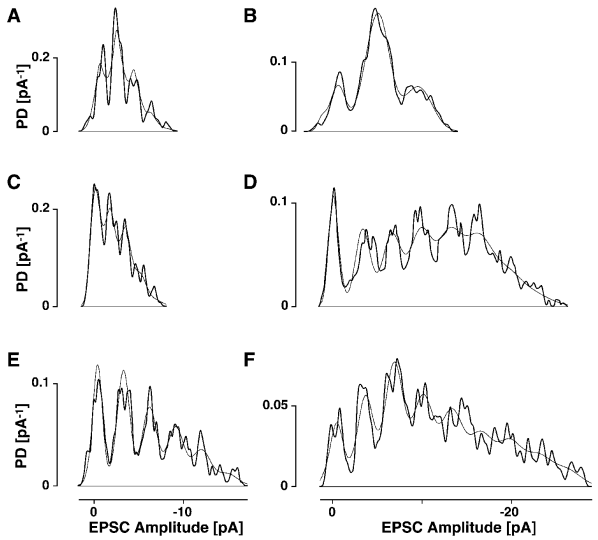


FIGURE 1. A set of 6 excitatory postsynaptic current (EPSC) amplitude distributions (thick line) with clear quantal separations and the best fit of the model to the data (thin line) in rat CA1 pyramidal cells. *A*: quantal separation is 1.9 ± 0.1 pA, with negligible quantal variability. *B*: quantal separation of 4.6 ± 0.2 pA and a variability normalized by the mean (CV) of 0.19 ± 0.03 are shown. *C* and *D*: recordings taken before (*C*) and after (*D*) the induction of LTP. Note the increase in quantal size from 2.7 ± 0.3 to 5.5 ± 0.5 pA. Toward the tail of the distribution in *D*, the amplitude modes start to become less clearly defined (CV 0.19 ± .12). *E* and *F*: same as in *C* and *D*; however, in this set of recordings, the EPSCs were recorded in the presence of 50 μM 2-amino-5-phosphonopentanoate to block *N*-methyl-D-aspartate currents. Note that there is no increase in quantal size (4.2 ± 0.1 vs. 4.4 ± 0.1 pA). In this case, the separation between successive peaks is retained from the control in *E* to the potentiated period in *F*. PD, probability density. Modified from Refs. 14, 16, and 17; reproduced with permission from The Physiological Society, London.

Figure 4.7: Distribution of synaptic response amplitudes averaged over many trials. The different humps are thought to correspond to the release of different numbers of vesicles. This illustrates the quantum hypothesis. From Stricker 2002.

- Per event a varying number of vesicles is released.
- The binding of transmitter post-synaptically and the consequent opening of the channels is stochastic.

The release of a discrete number of vesicles is called the QUANTUM HYPOTHESIS. The quantum hypothesis is one of the classic cornerstones of synaptic transmission. Consistent with this, the distribution of amplitudes can have multiple peaks each associated with a vesicle, Fig. 4.7. The easiest model is to assume a large number of independent release sites. The number of vesicles released per spike, k, is then given by the Poisson distribution

$$P_{Poisson}(k) = \frac{\mu^k \exp(-\mu)}{k!} \tag{4.1}$$

where μ is the average number of vesicles released or QUANTAL CONTENT. This can be used to fit the distribution of amplitudes. Note that sometimes no vesicle will be released at all ($P(k = 0)$). This is called failure. The Poisson distribution predicts that this happens with a probability $\exp(-\mu)$.

If the number of release sites, ACTIVE ZONES, is limited, one needs to replace the Poisson distribution with a binomial distribution

$$P_{bin}(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

when there are n vesicles maximally per release. Most of these models were developed in squid and neuro-muscular junctions, where the distributions are clear. In central synapses (such as in hippocampus or cortex; Br?maud, West, and Thomson (2007)), results are less clear, for review see

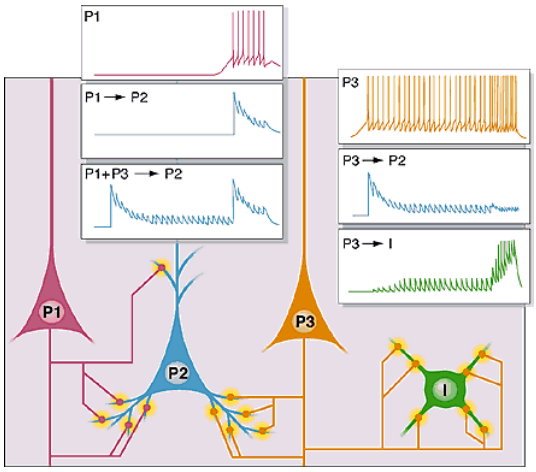


Figure 4.8: Schematic of the effect of synaptic depression. Two excitatory neurons P1 (only firing after a delay) and P3 (continuously firing) provide input to P2. Because of depression, only the transients of the input are transmitted. Synapses onto inhibitory neurons (I) are thought to be more likely to facilitate. The response of I is therefore strongly supra-linear in the firing rate of P3. From comment in Science onAbbott et al. 1997.

Stricker (2002). It has been suggested that a single vesicle could fully saturate the post-synaptic receptors. This would render the above analysis invalid.

Final remark: two neurons might use more than one synapse to contact each other. In the cortex a typical number seems to be between 1 and 10 parallel synaptic contacts per connection. Of course, these synapses are proportionally more reliable and stronger.

4.6 Synaptic facilitation and depression

As we saw, the transmission of synapses is not reliable. The probability for synaptic transmission at central synapses is about 0.1-0.5. But this release probability is not constant and is subject to the recent history of the synapse. This is called SHORT-TERM SYNAPTIC PLASTICITY (short-term because the changes don't last).

As the pre-synaptic cell is stimulated, Ca enter the terminal, causing the release of vesicles (or not). When now a second stimulus is given while the Ca concentration is still enhanced, the release probability will larger. This effect is called PAIRED-PULSE FACILITATION.

However, if stimulation persists, the READILY RELEASABLE POOL OF VESICLES will eventually exhaust causing a decrease in the release probability, we say that the synapse is depressed. In addition, the synaptic response can vary, due to the channel dynamics, as modeled by the state diagrams (e.g. desensitized states in Fig. 4.2). Synapses posses both facilitation and depression.

The depression makes the synapse sensitive to transient inputs. A simple way to model synaptic depression or facilitation is with two terms. One term describes that in the absence of activity the release probability settles to its default value

$$\tau_{depress} \frac{dp_{rel}(t)}{dt} = p_{rel}^{\infty} - p_{rel}(t)$$

The second term describes that every time a vesicle is released the release probability is reduced with a fixed amount

$$\text{if release } p_{rel} = [p_{rel} - \alpha]_+$$

where $[x]_+ = \max(x, 0)$, this rectification prevents the release probability from going below zero. For more models, see (Varela et al. 1997; Dayan and Abbott 2002; Tsodyks and Markram 1997).

The result is a strong non-linearity in the transfer function, dependent on the history of the synapse. In particular in recurrent networks this can yield interesting dynamics, e.g. Rossum et al. (2008).

4.7 Markov description of channels

Like the voltage gated channels, the synaptic channels modeled as a Markov processes. The state, or Markov diagrams, are a convenient method to analyse the channels. And we will see how they can help to derive the kinetics and channel properties from physiological recordings. The Markov diagrams can also be used to accurately simulate synapses, but it should be kept in mind that they only are a model. In a Markov process the transition probability to another state depends only on the current state, there is no history. That does not mean that every response is identical, it is possible by including additional states in the model, e.g. Fig. 4.2, to create for example a channel with desensitization.

The Markov diagrams are directly related to the power spectrum and the auto-correlation of the channel openings. To illustrate this consider a simple state diagram

$$C \underset{\alpha}{\overset{kT}{\rightleftharpoons}} C' \underset{\alpha'}{\overset{k'T}{\rightleftharpoons}} O$$

The first binding of transmitter T move the channel from closed state C to a second closed state C'; binding of another T opens the channel. We consider the case where T is artificially kept constant (e.g. by adding neurotransmitter to the recording bath).

We can write the state of the channel as a vector $\vec{s}(t) = (C, C', O)$. The entries in the vector are the probability to find a certain channel in that state, or, when measuring populations of channels, the average number of channels in that state. The dynamics of the channel can then be written as [3]

$$\frac{d\vec{s}}{dt} = W.\vec{s}$$

where W is a transition matrix between the different states. (This formalism is also called master equation and is closely related to the Chapman Kolmogorov equation.) For our example it is

$$W = \begin{pmatrix} -kT & \alpha & 0 \\ kT & -k'T - \alpha & \alpha' \\ 0 & k'T & -\alpha' \end{pmatrix}$$

The Perron–Frobenius theorems warrant that the W matrix has there is one zero eigenvalue, and the others are negative, but can be complex (e.g. (van Kampen 1992)). The eigenvector with zero eigenvalue is the steady state. In this case $\vec{s}^{\infty} \propto (\alpha\alpha', \alpha' kT, kk'T^2)$. The open probability in this steady state is found by normalising such that the sum of the probabilities is 1 (i.e. $\sum_{i=1}^{3} s_i = 1$). In this case it is $O^{\infty} = kk'T^2/(\alpha\alpha' + \alpha' kT + kk'T^2)$. For small concentration of transmitter, the open probability is proportional to the transmitter concentration squared. Then there is an intermediate regime, while for large concentration, $O^{\infty} = 1$, i.e. the receptor is saturated.[4]

The other two eigenvalues of the transition matrix tell us to how quick the system settles to the steady state. The general solution can be written as

$$\vec{s}(t) = \vec{s}^{\infty} + c_1 \vec{s}^{1} \mathrm{e}^{\lambda_1 t} + c_2 \vec{s}^{2} \mathrm{e}^{\lambda_2 t}$$

where λ_i is eigenvalue i and $\vec{s}^{i}$ is the corresponding eigenvector. The constants c_1 and c_2 are determined by the initial condition at time 0, which is given by $\vec{s}(0) = \vec{s}^{\infty} + c_1 \vec{s}^{1} + c_2 \vec{s}^{2}$.

[3] It is also common to define the Markov process $\vec{s}(t + \Delta t) = M.\vec{s}$. Thus the transition matrix $M = W + I$, the largest eigenvalue of M is one.

[4] In physiology one often defines the effective dose, symbol K_d, EC_{50}, or $K_{0.5}$. When $T = K_d$ the open probability is 50%. One encounter fits to the relation

$$O^{\infty} = \frac{1}{1 + (K_d/T)^n}$$

where n is called the Hill-coefficient. It expresses the cooperativity of binding. Note that this relation will not give a perfect fit in our case.

This describes the average behaviour of many channels. Like with the voltage gated channels, by making the transitions probabilistic, a stochastic model is obtained.

4.7.1 General properties of transition matrices

The matrix W has to obey certain properties to be a valid transition matrix. First, the total probability has to be conserved $\frac{d\sum s_i}{dt} = 0$, this implies that $\frac{d\vec{n}.\vec{s}}{dt} = \vec{n}.W.\vec{s} = 0$, where $\vec{n} = (1, 1, ..., 1)$. As this needs to hold at all time and thus for all allowed values of $\vec{s}$, one has $\vec{n}.W = \sum_i W_{ij} = 0$, which means the sum of every column of W has to be zero.

Secondly, W has to fulfil the *detailed balance* condition. It says that in equilibrium the probability to go around any loop in the diagram in the clockwise direction equals the probability in counter-clockwise direction. Without this condition there would be a constant rotation of occupation numbers in equilibrium. (In diagrams without loops the detailed balance condition is always automatically upheld). In equations it means that for all i, j

$$w_{ij}s_j^\infty = w_{ji}s_i^\infty$$

You check easily check this also holds for the Na-channel state diagram of the previous chapter by calculation clockwise and counter clockwise loop probabilities; this a result of the independence assumption of the gating variables. Detailed balance is a requirement from statistical physics. Without it, we still have a proper Markov process, but one which can not describe a physical process.

4.7.2 Measuring power spectra

Some properties of channels can be deduced from the noise even if single channel events are not resolvable. Suppose we measure the random openings of a single synaptic channel while transmitter is constantly present. In general two quantities are important in characterizing any sort of fluctuation:

1. The PROBABILITY DENSITY FUNCTION (PDF), also called probability distribution. In this case it is simple: the channel is either open or closed. The probability whether the channel is open is determined by the amount of transmitter present. With multiple channels the probability distribution becomes a binomial distribution.

2. The other thing to know is how fast the fluctuations change over time, how often does the channel switch per second. This is expressed in the AUTOCORRELATION FUNCTION. As we shall see, it is closely related to the power spectrum.

The auto-correlation function of a signal $S(t)$ is defined as

$$c(t') = \lim_{t_T \to \infty} \frac{1}{t_T} \int_0^{t_T} S(t)S(t + t')\, dt$$

For most processes $c(t')$ decays for long t', in other words, the process looses its memory. For the simplest Markov scheme with two-states ($C \rightleftharpoons O$), $c(t')$ decays exponentially.

Namely, $c(t') = c\,\mathrm{e}^{-|t|/\tau}$, we call τ the correlation time. In the case of our three-state synaptic channel, the correlation is the sum of two exponentials with time-constants $1/\lambda_1$ and $1/\lambda_2$.

The Fourier transform of the signal is defined as $s(f) = \int_0^T S(t)\mathrm{e}^{-2\pi i f t} dt$. The POWER-SPECTRUM gives how much power the fluctuations generate at each frequency. The power-spectral density, PSD, is given by

$$PSD(f) = \lim_{T \to \infty} \frac{2}{T} |s(f)|^2$$

The WIENER-KHINCHIN theorem says that the power-spectrum and the autocorrelation are related through a Fourier transform

$$\begin{aligned} PSD(f) &= 4\int_0^\infty c(t)\cos 2\pi f t\, dt \\ c(t) &= \int_0^\infty PSD(f)\cos 2\pi f t\, df \end{aligned}$$

(Easy proof by filling in the definitions).

For the simple two-state diagram the PSD is a Lorentzian, which is defined as

$$PSD(f) = \frac{PSD(0)}{1+(2\pi f/\lambda)^2} + c.\delta(f)$$

with $\lambda = 1/\tau$. The additional delta function term comes from the mean of the signal. It is usually removed by subtracting the mean of the signal.

For more complicated state diagrams, the PSD is a sum of such terms, each exponential in the autocorrelation function contributing one. In our three state channel, we have two non-zero eigenvalues and the spectrum is a sum of two Lorentzians. By applying a constant transmitter concentration one can measure a long stretch of fluctuating data, measure its power-spectrum, and fit the time constants. The number of Lorentzians in the spectrum gives the minimal number number of states necessary to describe the kinetics. Also the single channel conductance can be extracted from the power spectrum. This is called STATIONARY NOISE ANALYSIS. The larger the single channel conductance, the larger the fluctuations. The above technique requires the transmitter concentration T to be constant. It can thus be applied in case a long lasting puff of transmitter is applied.

This method can also be applied to voltage gated channels (under voltage clamp). However, with the patch-clamp technique it has becomes possible to measure the individual currents directly and this often provides better data.

4.8 Non-stationary noise analysis

The stochastic nature of the channel openings can also be used to infer the unitary conductance of the channels. Suppose we measure a collection of channels over many trials, the stimulation protocol determines the open probability of the channel as a function of time. Let's call the open probability at any time $p(t)$. The stimulus can be either a voltage step for voltage gated channels, or transmitter jump for synaptic channels. At any instant in time, the distribution of open channels is binomial. Thus the average current, averaged over many trials is

$$\langle I(t)\rangle = i_0 N p(t)$$

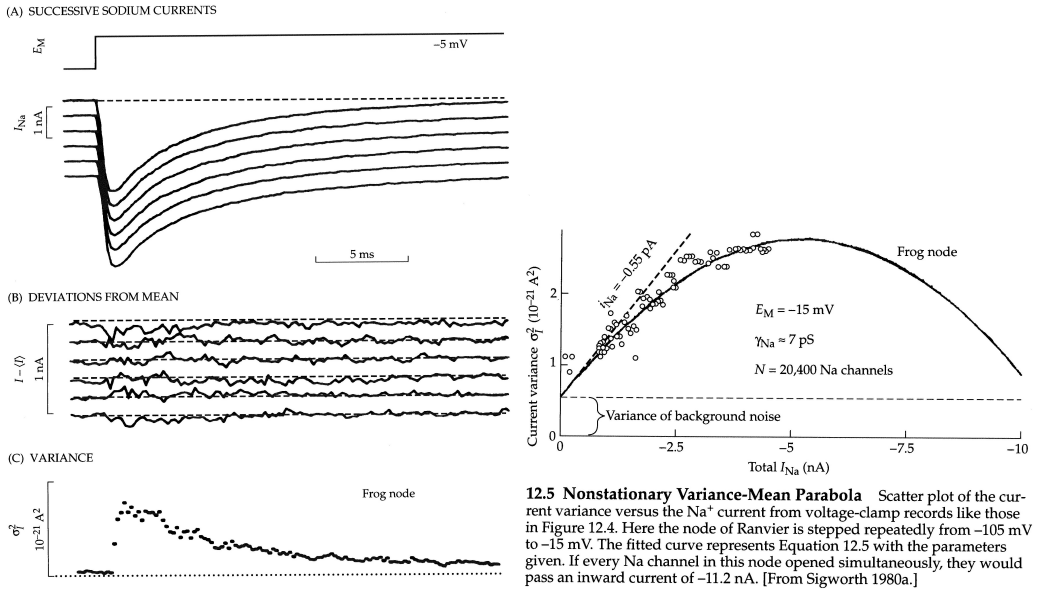
where i_0 is the unitary channel current and N the number of channels. The variance will be

$$\begin{aligned} \langle \delta I(t)^2\rangle &= i_0^2 N p(t)(1-p(t)) \\ &= i_0\langle I(t)\rangle - \langle I(t)\rangle^2/N \end{aligned}$$

The last equation shows that the channel's unitary current can be extracted from plotting the variance versus the mean, Fig. 4.9. In addition the number of channels can be extracted.

4.9 Detailed neuronal models

The work described in the previous chapters can be used to perform simulations of neuronal networks at a very detailed level (see e.g. Markram et al. (2015)). The complications similar to



12.5 Nonstationary Variance-Mean Parabola Scatter plot of the current variance versus the Na^+ current from voltage-clamp records like those in Figure 12.4. Here the node of Ranvier is stepped repeatedly from –105 mV to –15 mV. The fitted curve represents Equation 12.5 with the parameters given. If every Na channel in this node opened simultaneously, they would pass an inward current of –11.2 nA. [From Sigworth 1980a.]

Figure 4.9: Left: Non stationary noise analysis as applied the the sodium channels. A repeated step current leads to an average response. The size of the fluctuations around it are dependent on the unitary conductance.
Right: The resulting parabola when variance is plotted versus the mean. FromHille 2001

those mentioned at the end of last chapter, also hold on the network level: the synaptic properties and the connectivity needs to be guessed or drawn from statistical data. In addition, specification of the input and the output of the network, as well as the objective function of the network are far from trivial.

While having a accurate, detailed brain simulation would be very nice to have, given these problems one can also follow another route and to understand the principles behind neural computation we will abstract away from this detail for the remainder of the course.

Good models for synaptic kinetics can be found in Destexhe, Mainen, and Sejnowski (1998) and Roth and Rossum (2009), more on Markov diagrams in Johnston and Wu (1995).

Chapter 5

Integrate and fire models

Mathematically, the Hodgkin-Huxley description of spike generation is complicated, even in a single compartment model. It is a set of coupled non-linear differential equations, which in the end describe how the membrane voltage reacts to external current injection. The shape or height of the spike hardly depends on the injected current or previous spikes. (And differences in spike shape or spike height are often neglected anyway). The spike is a stereotype event. Not all details of the HH model are therefore necessary to model spike like events. This observations have led to reduced models of spiking neurons. For more on such reductions see Tuckwell (1988) and Koch (1999). These reduced models are typically also easier to fit to experimental data than more complicated models, and there are systematic methods to find their parameters (see NIP course). In principle a full biophysical model is more accurate, provided that all parameters are known, but in practice, given the uncertainties about the dendritic conductances and synaptic inputs, reduced models might actually work better and are of course much more manageable.

A number of reduced models has been developed. The most common reduced model for spiking is the integrate-and fire model. It introduction is often attributed to Lapicque (Lapicque 1907) (but see Brunel and Rossum (2007) for a historic perspective). Because of its easy and efficient implementation, integrate-and-fire neurons are used frequently in modelling studies. The integrate and fire model describes the sub-threshold behaviour with the simple passive RC circuit that we have seen above, Fig. 2.2,

$$\boxed{C\frac{dV_m(t)}{dt} = -\frac{1}{R_m}[V_m(t) - V_{rest}] + I_{ext}(t)}$$

In addition a spike threshold is supplied

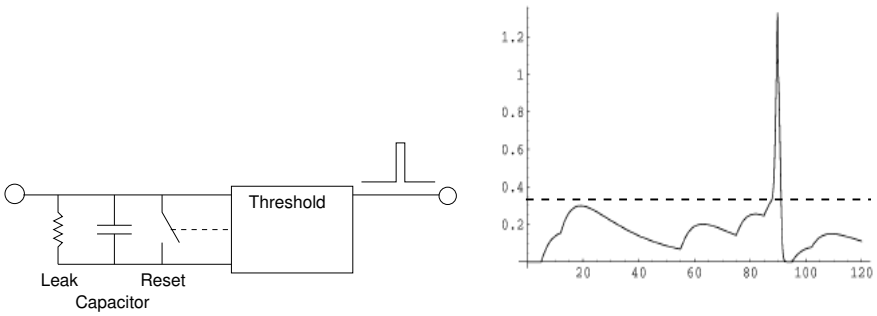


Figure 5.1: Left: The circuit diagram of the integrate-and-fire neuron. Right: example voltage trace. The integrate-and-fire neuron does not generate the nice spike shown, it only generates 'spike-events'. The spike shape was added by hand.

$$\text{if } V > V_{thr} \quad \text{then} \quad \text{Spike} \quad \text{and } V = V_{reset}$$

This is the so called LEAKY INTEGRATE AND FIRE model. The simplification without the R_m-term is called the leak-less I&F model. Usually, the V_{reset} is chosen to be equal to V_{rest}, but this is not necessary. Observe that there is no real spike in the membrane voltage, but instead the spike is an event. Although the integrate-and-fire model appears linear and simple, it actually is not. This is due to the threshold and subsequent reset.

The F/I curve of the I&F neuron is easily calculated: Suppose $V_{reset} = V_{rest}$, and the neuron is driven by a current I_{stim}. Let's assume the neuron has fired at $t = 0$. Its membrane potential is V_{rest}. When will the next spike occur? When is the threshold voltage reached? The voltage will behave as $V(t) = V_{rest} + I_{stim}R_m\left[1 - \exp(-t/\tau_m)\right]$. Solving for t gives

$$\begin{aligned} f &= 0 \qquad (\text{if } I_{stim} < (V_{th} - V_{rest})/R_m) \\ f &= \frac{-1}{\tau_m \log(1 - \frac{V_{th} - V_{rest}}{I_{stim} R_m})} \end{aligned}$$

This curve looks quite like the Hodgkin-Huxley one (Fig. 3.6). There is a threshold, too small input current will not evoke any spikes. For currents above threshold, there is a curved F/I relation, like in Fig. 3.6.

For large stimuli, $f \approx I_{stim}/(C_m V_{thr})$, i.e. the frequency is linear in the input current. It also becomes independent of the leak resistance. This is because the leak current is in this case negligible comparable to the stimulus current. There is no maximal rate as the spikes (which are infinitely thin) can be right after each other. Real neurons show saturation at high stimulus frequencies due to the refractory period. This can be easily introduced in the model by clamping the membrane to the resting potential during the refractory period. *Exercise:* Calculate the F/I curve in that case.

Spike frequency adaptation, Fig. 7.3, can easily be included in integrate and fire models. Hereto we introduce an extra variable in the model which represents the Ca concentration. Every time a spike occurs, the Ca concentration jumps up with fixed amount. In between spikes it decays with a time-constant of, say, 50ms. This represents the pump continuously active to remove the Ca. The Ca can be coupled to a simple KCa-like current, such as

$$I_{Kca} = g_{KCa}[Ca](E_K - V)$$

where g_{KCa} is the strength of the feedback, $[Ca]$ is the calcium concentration.

These extensions to the integrate-and-fire model are useful and easy to do in simulations, but make analytical treatment often much harder. In case of the adaptation, the description of the neuron becomes two dimensional (both $V(t)$ and $[Ca](t)$ are needed to specify the state of a neuron), instead of one dimensional (just $V(t)$).

5.1 Models of synaptic input

When dealing with integrate and fire models, it does not always make sense to have very elaborate synapse models based on ligand binding and Markov diagrams. Instead, one usually takes synaptic currents that are simple functions of time. The simplest choice to take the synaptic current a delta function, which will cause the membrane potential to make an instantaneous jump.

$$I_{syn} = I_0\delta(t_i) \;\Rightarrow\; \Delta V = R_m I_0$$

More realistic is an exponential current, to describe for instance an AMPA synapse.

$$I_{syn} = I_0 \exp(-t/\tau_{syn}) \tag{5.1}$$

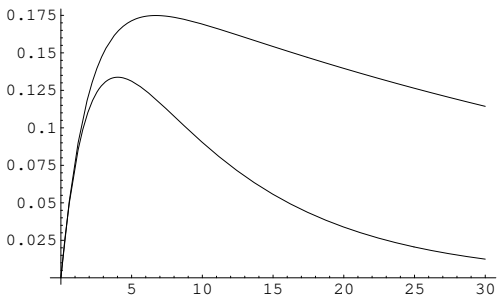


Figure 5.2: Effect of exponentially decaying synaptic current on the membrane potential. Membrane potential vs. time in ms. Parameters: $\tau_{syn} = 2ms$. Upper curve: $\tau_m = 50ms$, lower curve: $\tau_m = 10ms$.

Assuming the neuron doesn't spike, the resulting voltage is a difference of exponentials, the rise time of the voltage is roughly given by τ_{syn}, the decay by τ_m (assuming $\tau_{syn} \ll \tau_m$)

$$\Delta V = I_0 R_m \frac{\tau_{syn}}{\tau_m - \tau_{syn}} [\mathrm{e}^{-t/\tau_m} - \mathrm{e}^{-t/\tau_{syn}}]$$

The voltage trace is shown in Fig. 5.2. As one can see the time-constant of the cell determines the time for membrane to settle back to the equilibrium, but the rise can be much faster. In other words, the synaptic time-constant is more important than the membrane time-constant for the firing dynamics. For example, in the extreme case of a non-leaky I&F neuron one has $\tau_m = \infty$, but the neuron can definitely react faster than that.

Another synapse model often used is an alpha-function: $I_{syn} = I_0\, t/\tau_{syn} \exp(-t/\tau_{syn})$. Such currents can be used to mimic cable filtered synaptic input or more precise synapse models.

Inhibitory inputs can be modelled with a negative current. More realistically, an inhibitory conductance is used. (For excitatory synapses it makes virtually no difference whether the input is modelled as a conductance or a current as the membrane potential is always far from the excitatory reversal potential).

5.2 Shunting inhibition

An important question is how excitation and inhibition inputs are combined in a cell. In particular in spatially extended models, this is a difficult question. However, here we concentrate on a single compartment model. What is the effect of inhibition in these simple models? Consider a single,*passive* compartment with both excitatory g_e with zero reversal potential and inhibitory input g_i with reversal potential V_{inh}.

$$C\frac{dV(t)}{dt} = -g_l(V(t) - V_{rest}) - g_e(V(t) - V_{exc}) - g_i(V(t) - V_{inh})$$

The steady state voltage of the compartment is given by the solution to $C\frac{dV}{dt} = 0$, or

$$V_\infty = \frac{g_l V_{rest} + g_e V_{exc} + g_i V_{inh}}{g_l + g_e + g_i}$$

Usually the excitatory reversal potential V_{exc} is around $0mV$. The inhibitory reversal potential V_{inh} is about the resting potential. This is called SHUNTING INHIBITION: the inhibition does not

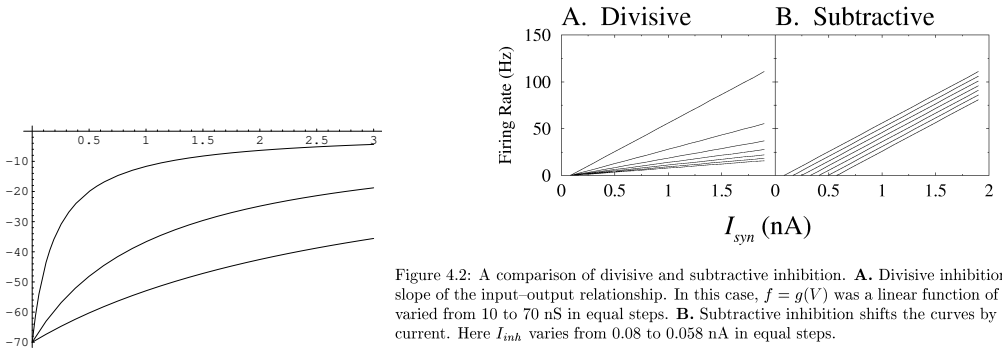


Figure 4.2: A comparison of divisive and subtractive inhibition. **A.** Divisive inhibition changes the slope of the input–output relationship. In this case, $f = g(V)$ was a linear function of V and G was varied from 10 to 70 nS in equal steps. **B.** Subtractive inhibition shifts the curves by subtracting a current. Here I_{inh} varies from 0.08 to 0.058 nA in equal steps.

Figure 5.3: Effect on inhibition. Left: a passive model. The membrane voltage as a function of excitatory input in a non-spiking, passive model for three levels of inhibition, $g_i = 0.1, 1$, and 3 (upper curve to lower curve). Parameters: $V_{rest} = V_{inh} = -70$, $V_{exc} = 0$, $g_l = 0.1$. Right: How perfect divisive and subtractive inhibition would look like in theory.Right plot from Holt 1998.

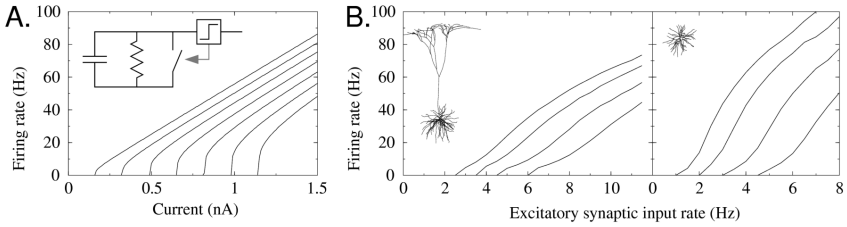


Figure 4.3: Changing g_{leak} has a subtractive rather than a divisive effect on firing rates. **A.** Current discharge curves for the integrate–and–fire model, with g_{leak} varying from 10 to 70 nS (from left to right) in steps of 10 nS. **B.** Fully adapted firing rates of the two cells as a function of excitatory input rate for different inhibitory input rates. From left to right, the curves correspond to a $GABA_A$ inhibitory rate of 0.5, 2, 4, and 6 Hz. Note that in all of these cases, the curve shifts rather than changes slope. In this case, inhibitory synapses were near the soma ("near" configuration in table 4.1), as found in cortical cells.

Figure 5.4: The effect of shunting inhibition in a spiking model cells can be more subtractive than divisive.From Holt 1998.

hyper-polarise the neuron but instead forces the membrane potential toward the resting potential. For strong inhibition ($g_i \gg g_e, g_l$), the difference between resting potential and steady state voltage is $\Delta V \approx g_e(V_{exc} - V_{inh})/g_i$, see Fig. 5.3. The effect of the inhibition is thus roughly divisive if one considers passive neuron models.

However, in spiking neurons shunting inhibition acts rather in a subtractive than divisive manner (Holt and Koch 1997), as is shown in Fig. 5.4. This effect happens both in I&F neurons, as in more realistic models. It highlights the non-linearity of the I&F model. The reason why the steady state argument does not hold, is that the neuron never reaches a steady state when it is firing. If V_{inh} is exactly V_{rest} the inhibitory conductance just effectively increase the leak conductance. The F/I curve is as above but with rescaled R_m. When the inputs are not constant but fluctuating (noisy) the story is more complicated and the inhibition can under some circumstances act divisively (Mitchell and Silver 2003).

5.3 Simulating I&F neurons

To simulate I&F neurons we can write a straight-forward computer program to do so. Here we give example code of an integrate and fire neuron embedded in a network. The synapses are modelled

here as exponentially decaying conductances.

The starting point is the total current into the neuron. First we collect all input current. With $g(k,l)$ we denote the input conductances from neurons k to neuron l. For simplicity we consider one type of synapse between the neurons. We have

```
i_total = i_external_stimulus
forall inputs k {
  i_total += input_g(k,l)*(E_syn - v)
}
```

The voltage is given by $C_m\, dV/dt = -(V - V_{rest})/R_m + I$. Insert the definition of the derivative $dV(t)/dt = [V(t+\delta t) - V(t)]/\delta t$ or

$$V(t+\delta t) = V(t) + \delta t/\tau_m \left[-(V(t) - V_{rest}) + I.R_m\right]$$

where $\tau_m = R_m C_m$. In code:

```
v += dt/tau*(-v+vrest+i_total*rm)
```

Next, we have to detect spiking. A spike will reset the potential but will also provide input to other neurons, indexed with m, receiving input from it,

```
if (v > vthr){
  v=vreset
  forall m {
    input_g(l,m) += synaptic_weight(l,m) % a simple update model for input to other cells
  }
}
```

We have to update (decay) the synaptic conductances. When the synaptic conductances are modelled as exponential synapses, the corresponding differential equation for the conductance is $\tau_{syn} dg(t)/dt = -g(t)$.

```
forall synapses (k, l){
  input_g(k,l) -= dt/tau_syn*input_g(k,l)
}
```

Given that in a network there are usually many more synapses than neurons, the synaptic calculations will take the most time. If all synapses have the same time-constant (and assuming linear summation of the conductances), a simple but effective speed-up is to collect all incoming synaptic conductance into a single conductance. Now one only needs to decay the sum of the conductances instead of the individual ones.

A note about time-steps and integration.

In the current integration scheme spikes occur always on the time-step. A too large time-step will incorrectly prevent jitter in the spike times. Especially in large recurrent networks, this can lead to spurious synchronisation. This can be avoided by taking small enough time-steps. In addition, better algorithms calculate the precise firing time by interpolating between the time-steps.

The numerical integration used above is called the Euler scheme. Higher order integration routines are worthwhile, although the simplicity of the Euler scheme is attractive, for more details see Hansel et al. (1998).

Well known network simulation packages based on integrate-and-fire neurons are NEST and BRIAN.

Chapter 6

Single neuron firing statistics and noise

6.1 Neural variability

In the previous chapters we have developed models for firing of the neurons. Now one can measure the firing properties of neurons*in vivo*, and see how they match the models, or lead to refinement of the models. One of the most obvious aspects of neural activity not present in the models so far is the large variability, so here we look at computational considerations about that.

Suppose one has found a cell which reacts to certain visual feature, such as a white bar on a black background, or maybe a face. When we present the same stimulus over and over again to the neuron, its spike-train shows quite a bit of variability. This is shown for a visual neuron in Fig. 6.1. One common explanation is that this variability is due to noise. Another possibility is that for every repetition the nervous system perceives the stimulus differently, leading to a different response. Something like: "Hey, a new face", "Here it is again", "And yet again", and finally "It's boring". A more formal way to state this is to say that the *internal state* of the organism is important and the the system is non-stationary. After all, unlike a transistor that always responds similarly, the neuron is a living biological cell and in a constant state of flux. For instance, neural excitability might change (Goris, Movshon, and Simoncelli 2014). This could potentially be a serious concern. By checking whether firing rates and patterns remain similar throughout the experiment one can try to control for this. In early sensory cells, such as in the retina, one would not expect such effects. Anaesthesia might also partly relieve this problem, although that introduces other problems.

Alternatively, the neural code is perhaps not precisely stimulus-locked, but could use very precise timing between neurons. Also in that case fluctuations might not corresponds necessarily to noise. This could easily happen in neurons further removed from the sensory system.

Although these possibilities are worth more study both experimentally and theoretically, one usually interprets all variability as noise. A certain portion of the variability is certainly noise, for instance the random vesicle release and the noise of voltage gated channels.

A single trial response is noisy and typically only contains a few spikes. Therefore in order to measure the response characteristic of a given neuron, responses are commonly averaged over trials and by binning the spikes the Post Stimulus Time Histogram (PSTH) is produced. This PSTH gives the average temporal response profile to the stimulus. Note that although signals might be obvious in the PSTH, the nervous system itself does not have access to the many trials and has to rely on other information, such as combining the responses from multiple neurons.

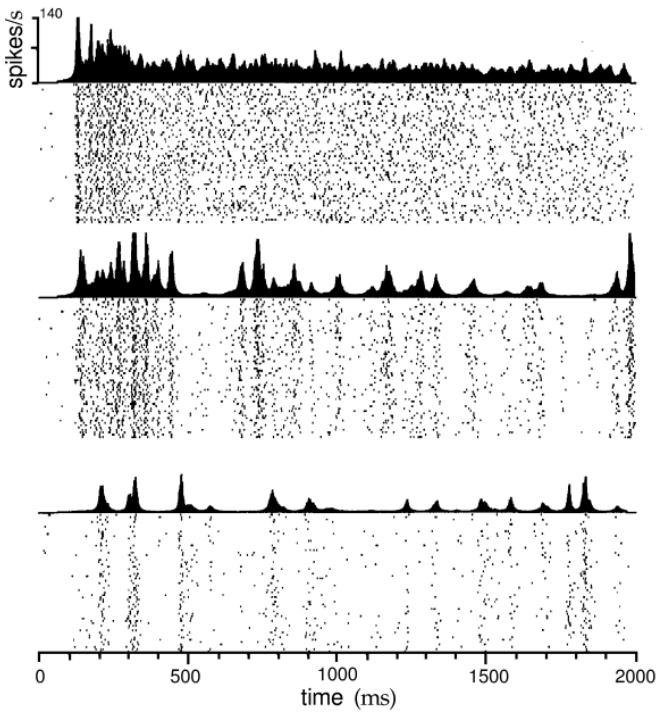


Figure 6.1: Example PSTH of a visual cortex neuron that is motion sensitive (area MT). The dots indicate the spikes from the neuron, each row corresponding to a different trial, firing frequency averaged over trials is shown above. The three panels correspond to different spatial temporal stimuli. The top stimulus is a constant motion signal, the lower two are fluctuating stimuli. These sort of firing statistics are seen throughout the cortex. The figure also shows that despite fluctuations, the firing can be more (bottom) or less (top) stimulus locked. From Dayan and Abbott 2002 after Bair and Koch 1996.

6.2 Interval statistics

We first consider the simple case when the stimulus is constant. Various statistical measures have been introduced to measure the variability of spike trains. One common measure is the inter-spike interval (ISI) distribution, $P_{ISI}(t)$. From this distribution we can determine the average and variance of the ISI. The following equations are valid for any probability distribution. The mean is given by

$$\langle t \rangle = \int dt\, P_{ISI}(t) t$$

The variance (the standard deviation squared) is[1]

$$\langle \delta t^2 \rangle = \sigma_t^2 = -\langle t \rangle^2 + \int dt\, P_{ISI}(t) t^2 \tag{6.1}$$

The coefficient of variation is defined as

$$CoV = \frac{\sigma_t}{\langle t \rangle}$$

[1]Especially for complicated problems, many mistakes are made in calculating the variance; use Eq. 6.1 and you will make no such mistakes.

The simplest category of spike interval models are the so-called RENEWAL processes. In renewal processes each spike time is chosen from 1) the same interval distribution and 2) independent of other intervals. Therefore the interval distribution completely determines the spike times. This rules out effects such as spike frequency adaptation known to be present in biology. Despite this lack of reality the description as a renewal process opens up many mathematical results and techniques. The spike time models in the next two sections are renewal models.

Intermezzo As we are talking statistics anyway The following information is useful when you are reading experimental papers. In experiments is often important to see if a certain treatment changes some mean property of something. Therefore it is important how reliable the estimate of the mean is. This quantity is called the STANDARD ERROR σ_M and is given by $\sigma_M = \sigma/\sqrt{N}$. By measuring enough data, the standard error can usually be made arbitrarily small, whereas σ usually tends to a fixed value when enough data are measured. The means the two conditions can be compared using a t-test. The resulting p value indicates how likely it was that the effect occurred due to a random coincidence (testing the null-hypothesis). The lower the standard errors, the more significant the result. Values below 5% ($p<0.05$) are considered significant and publishable. Note, that the lack of a significant result does not mean that there is no effect, it might be too small to see, or the data might be too noisy.

6.3 Poisson model

The presence of irregularity in neural activity has led to a number of noise models for neurons. The simplest model for spike generation is a Poisson process. (The prime example of a Poisson process is radio active decay). In a homogeneous Poisson process the probability for a spike is identical at every time. Suppose the rate of the Poisson process is $\lambda = 1/\tau$. What is the distribution of inter spike intervals? Divide the inter-spike interval in many small steps of duration Δt. The probability to spike after time t_{isi} is given by

$$\begin{aligned} P(t_{isi} = n\,\Delta t) &= (1 - \frac{\Delta t}{\tau})^{n-1}\frac{\Delta t}{\tau} = \frac{\Delta t}{\tau}\exp[(n-1)\log(1-\frac{\Delta t}{\tau})] \\ &\approx \frac{\Delta t}{\tau}\exp(-t_{isi}/\tau) \end{aligned}$$

We have to normalise this expression such that $\int_0^\infty dt_{isi} P(t_{isi}) = 1$, and we find

$$P_{Poisson}(t_{isi}) = \frac{1}{\tau}\exp(-t_{isi}/\tau)$$

In other words, the intervals are exponentially distributed. The mean interval is τ and the variance is τ^2. In the Poisson process the probability for an event is constant (there is no time-scale intrinsic to the Poisson process). Therefore the autocorrelation function is a constant plus a δ function at zero. As a result the power-spectrum is flat as well with a δ function at zero.

Of course, the spike rate of a neuron usually changes when a stimulus is presented/removed or the stimulus fluctuates, as in Fig. 6.1. A rate-modulated Poisson process can be used to model this. In that case at every instant the firing probability is calculated using a time-dependent rate $\lambda(t)$ (the plotted trial-averaged firing rate in Fig. 6.1). This will of course destroy the exponential distribution of inter-spike intervals. However, if we probe many trials, we would find a Poisson spike count distribution for every instant t, with rate $\lambda(t)$.[2]

The Poisson process is only a rough approximation for the firing of real neurons. Although reasonably correct for low firing rates, the Poisson approximation goes wrong for higher rates. When the instantaneous rate is high, a Poisson model can fire twice in rapid succession. However, a real neuron's membrane potential will have to climb up from reset to threshold voltage. In addition, refractory mechanisms prevent a rapid succession of spikes in real neurons. As a result real neurons fire less fast and also more reliable (Berry and Meister 1998). Nevertheless, the Poisson model is a reasonable and mathematically attractive model for neural firing *in vivo*.

[2] A good example of a Poisson process with a time-dependent rate is pop-corn in the microwave.

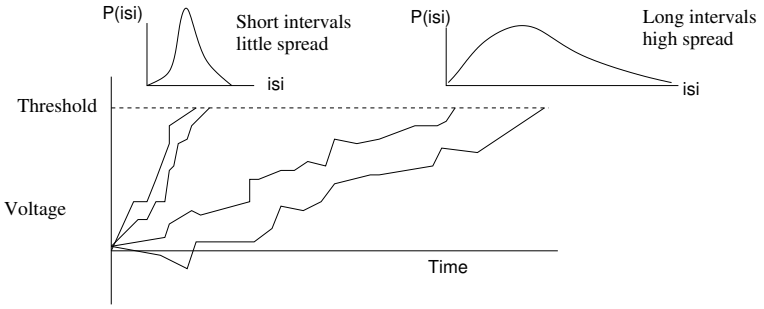


Figure 6.2: Noisy integrate and fire neuron. Different sample paths from rest to threshold. For strong stimuli (left 2 sample traces) the slope is steep and the variations in the spike timing is small. For weak stimuli (right 2 traces) the slope is shallow and noise accumulates. The shape of the inter-spike interval histograms sketched on top is like those observed in physiology.

6.4 Noisy integrate and fire neuron

In the Poisson process the intervals are exponentially distributed. This is a reasonable approximation for neurons with low firing rates (for instance background firing). The inter-spike interval distribution of real neurons with higher firing rates is usually a skewed Gaussian-like distribution (Note, that mathematically the ISI distribution can never be strictly Gaussian as negative intervals do not exist).

There are various noise models which can be inserted to give realistic variability in the firing. One can, for instance, make the threshold fluctuate, chose a random reset potential, or inject a noise current in the neuron. When we supply a constant noise term to the leak-less integrate and fire neuron, one obtains reasonably realistic firing statistics. Simplifying such that $V_{reset} = V_{rest} = 0$, and $C = 1$ the governing equation is

$$\frac{dV}{dt} = I_{stim} + I_{noise}(t) \tag{6.2}$$

and of course, as before the neuron spikes and resets when it reaches the threshold.

The problem is now to determine when the neuron will fire, given that $V(t = 0) = 0$. The noise term is easy to simulate, from trial to trial the neuron will take different paths to the threshold, Fig. 6.2. Under some simplifying assumptions analytical treatment is possible: The voltage obeys a diffusion equation as we will see shortly (Gerstein and Mandelbrot 1964). For mathematical simplicity we need to assume that the noise term has zero mean $\langle I_{noise} \rangle = 0$, and its correlation is $\langle I_{noise}(t) I_{noise}(t') \rangle = \sigma^2 \delta(t - t')$.[3]

The spike-timing can be calculated using a Fokker-Planck equation. Consider the distribution of possible voltages $P(V, t)$. The Fokker-Planck equation simply collects the contributions which change this distribution, as is shown schematically in Fig. 6.3.

$$P(V, t + \Delta t) = P(V, t) + \int dW T(V - W, V) P(V - W, t) - P(V, t) \int dV' T(V, V')$$

where $T(V', V) = T(V - W, V)$ is the transition probability that voltage jump from V' to become V. Consider a small time interval Δt, the voltage can change due to two effects: 1) due to the stimulus $V \to V + \Delta t I_{stim}$, and 2) due to the noise $V \to V +$ noise. In other words, $T(V', V)$ has a Gaussian distribution with mean $I_{stim}\Delta t$ and variance $\sigma^2 \Delta t$.

One now approximates $[T(V - W, V)P(V - W, t)]$ by a Taylor expansion around $[T(V, V + W)P(V, t)]$. This requires small steps and a smooth $P(V, t)$. After this one can pull $P(V, t)$ out of

[3]This is so called white noise; its power-spectrum is flat, i.e. it contains all possible frequencies. For a treatment with correlated noise see Brunel and Sergi (1998).

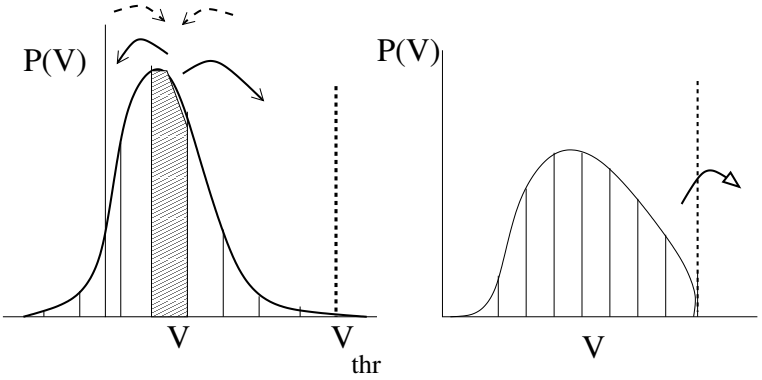


Figure 6.3: Cartoon of the Fokker-Planck approach. Left: Situation shortly after the neuron has fired. The Fokker-Planck equation collects all loss and gain term to $P(V)$ at each voltage (solid arrows loss terms, dashed arrows gain terms). Right: Later situation: the distribution is spread due to the noise, and its mean has shifted toward the threshold due to the stimulus current. When the probability reaches threshold, the neuron fires (open arrow).

the integrals over W and one obtains the Fokker-Planck equation

$$\begin{aligned}\frac{\partial P(V,t)}{\partial t} &= -\frac{\partial}{\partial V}[A(V)P(V,t)] + \frac{1}{2}\frac{\partial^2}{\partial V^2}[B(V)P(V,t)] \\ A(V) &= \int_{-\infty}^{\infty} dW\, WT(V, V+W) \\ B(V) &= \int_{-\infty}^{\infty} dW\, W^2 T(V, V+W)\end{aligned}$$

In this case, using Eq. 6.2 $W = \mathcal{N}(I_{stim}\Delta t, \sigma^2\Delta t)$ we simply have $A(V) = I_{stim}$ and $B(V) = \sigma^2 + O(\Delta t)$.In general Fokker Planck equations these so called jump-moments $A(V)$ and $B(V)$ could depend on V, but here they are constants. The distribution of voltages obeys a diffusion equation, just like the cable equation:

$$\frac{\partial P(V,t)}{\partial t} = -I_{stim}\frac{\partial P(V,t)}{\partial V} + \frac{1}{2}\sigma^2\frac{\partial^2 P(V,t)}{\partial V^2}$$

Suppose that at time 0, the neuron has just fired, thus at time $t = 0$ the potential is distributed as $P(V, 0) = \delta(V)$. Assume $I_{stim} = 0$, from the discussion of the cable equation, we know the solution in infinite space, namely[4]

$$P_{\infty,0}(V,t) = \frac{1}{\sqrt{2\pi\sigma^2 t}}\mathrm{e}^{-V^2/2\sigma^2 t}$$

The voltage moves around like a diffusing particle, Fig. 6.2.

When will the voltage reach the threshold? This is the meandering drunk with a ditch problem. There are two complications: First, there is a drift term, the term proportional to I_{stim} (the drunk walks on a down-hill slope). Secondly, the boundary conditions are important. Once the threshold is reached the neuron fires and this particular voltage trace needs to be discounted. This means that we have to impose an absorbing boundary (a sink) at the threshold: $P(V_T, t) = 0$. The neuron that spiked will re-enter the population at the rest voltage, but here we are interested only in the timing of the*first* spike, so we can consider neurons that spiked once to be gone. The probability turns out to be

$$P(V,t) = \frac{1}{\sqrt{2\pi\sigma^2 t}}\left(\mathrm{e}^{-V^2/2\sigma^2 t} - \mathrm{e}^{-(2V_T - V)^2/2\sigma^2 t}\right)\mathrm{e}^{I_{stim}V/\sigma^2 - I_{stim}^2 t/2\sigma^2}$$

[4]And for arbitrary I_{stim}, $P_{\infty,0}(V,t) = \frac{1}{\sqrt{2\pi\sigma^2 t}}\mathrm{e}^{-(V - I_{stim}t)^2/2\sigma^2 t}$

(Check that $P(V_T, t) = 0$; for the experts: the solution uses so called mirror charges to achieve this). Next, we need the firing rate of the neuron. This is given by the fraction of neurons that pass the threshold, Fig. 6.3. Fick's law relates the spatial gradient of the density to the probability current: $P_{Fire}(t) = -\frac{1}{2}\sigma^2 \frac{\partial P}{\partial V}|_{V=V_{thr}}$. Thus we get

$$P_{Fire}(t) = \frac{V_T}{t\sqrt{2\pi\sigma^2 t}} \mathrm{e}^{-(V_T - I_{stim}t)^2/2\sigma^2 t}$$

This is called an inverse Gaussian distribution. The mean interval is

$$\langle t \rangle = \frac{V_T}{I_{stim}}$$

Note that because the neuron has no leak, there is no threshold in the f/I curve. And the variance in the intervals is

$$\langle \delta t^2 \rangle = \frac{V_T \sigma^2}{I_{stim}^3}$$

Or the Coefficient of Variation (CoV) of the inter-spike interval is $CoV \equiv \frac{\sqrt{\langle \delta t^2 \rangle}}{\langle t \rangle} = \sigma / \sqrt{V_T I_{stim}}$. From this last equation we see that when the stimulus is strong, the firing frequency is high and fluctuation in the inter-spike interval are small. When the stimulus is weak, the inter-spike interval has a wider distribution as the noise has more time to influence the voltage, Fig. 6.2. This in reasonable agreement with cortical neurons (Gerstein and Mandelbrot 1964).

The noise typical has other effects as well. The strong threshold for firing seen *in vitro* is smeared out and the input-output relation becomes more or less threshold-linear (Anderson et al. 2000).

6.5 Stimulus locking

Now we turn to the situation when the stimulus is not constant. An interesting effect of noisy I&F neurons is that their timing precision seems to depend no only on the stimulus strength but also on the presence of transients in the stimulus (Bryant and Segundo 1976; Mainen and Sejnowski 1995). This is shown in Fig. 6.4. From the spiking pattern in the left panel, one would not expect that the neurons could spike so reliably as they do in the right panel. Transients in the stimulus can synchronise and lock the firing. This effect appears to be even stronger for real neurons probed with current injections. Note the similarity with in vivo neurons, Fig. 6.1.

6.6 Count statistics

The interval statistics is one way to measure spiking variability. Alternatively, one can count the number of spikes and calculate its statistical properties. Suppose one counts the spikes occurring in a window of duration T, one repeats the experiment and measures the mean and variance in the count. The Fano factor is defined as

$$F(T) = \frac{\sigma_N^2(T)}{\mu_N(T)}$$

For a Poisson process, see Eq. 4.1, the mean count is T/τ, and the variance is T/τ as well, so the Fano factor is 1 for all T. (It is important to distinguish the count statistics and interval statistics especially for the Poisson process). The behaviour for large T is revealing: most simple neuron models converge to fixed values for long T, however real neurons show often a steady increase in the Fano factor for long times (seconds), Fig. 6.5. This could either mean that slow variations in for instance excitability occur, or that the firing has a fractal structure (Lowen et al. 2001). (Exercise: Try to create a cell model with such a Fano factor as in Fig. 6.5).

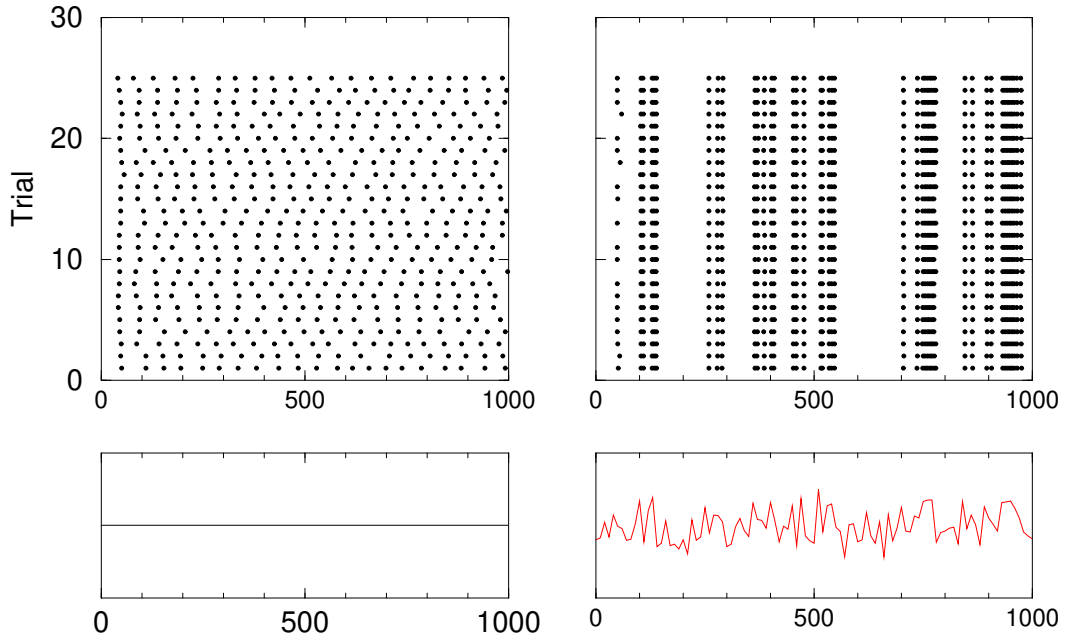


Figure 6.4: Left: Noisy integrate and fire neuron stimulated with a DC current (bottom panel), responses (top) are shown trial-to-trial. After a few spikes, spikes are not stimulus locked. Right: Stimulation with a fluctuating current leads to much more precise spiking; the firing locks to the stimulus. Time in ms on x-axis. Every dot indicates a spike. This is for a single neuron, but presumably the same occurs for a population of neurons, all measured at once (see Chapter 8).

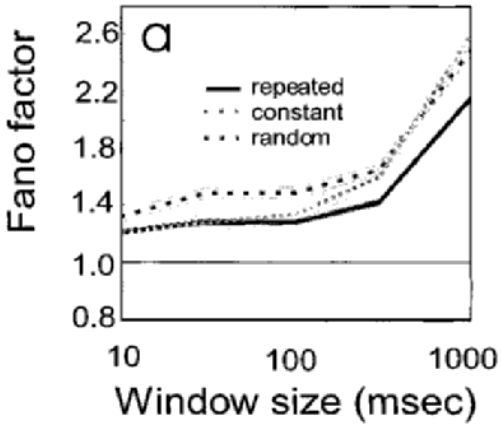


Figure 6.5: The Fano factor of the spike count for an experiment like in Fig. 6.1. The flat line corresponds to a Poisson process. From Buracas et al. 1998.

6.7 Neuronal activity and variability *in vivo*

As we have discussed the neural activity about as variable as a Poisson train. What underlies this high variability? A neuron in vivo continuously receives synaptic inputs. In a resting, awake state neurons fire at an average rate of about 3Hz. Cortical neurons have about 10,000 synapses. A connection between two neurons involves on the average about five synapses. Maybe each synapse is 50% reliable. This means that a neuron receives input at a frequency of $3 \times 10.000/5/2Hz = 3000Hz$. Now suppose for a moment that all these inputs are identical and excitatory. Two problems arise:

- Because there are so many synapses, the fluctuations in the input are averaged out. Therefore one would expect the post-synaptic neuron to fire very regular. Note, that this will be the case despite all synaptic and channel noise sources we have encountered.
- A single input depolarises the cell with about 0.1 mV. Maybe some 20..200 inputs are required to fire the cell. The total excitation will cause the cell to fire at very high rates, even though is just receives background input.

A number of factors has been identified to explain this discrepancy:

- The input is not accurately described by uncorrelated processes, as often a considerable amount of correlation is present. The presence of correlations effectively reduce the number of independent inputs, and as a result firing is more irregular (Stevens and Zador 1998).
- The excitatory input is balanced by inhibitory inputs. This will reduce the net drive, and at the same time increase the noise. In addition, because the inhibition is a conductance (and not just a current), the effective leak of the cell is increased (see Section 5.2). The typical background activity can reduce the input resistance by some 80%, reducing the time-constant of cells to about 5 ms, and the electrotonic length will become shorter as well (Paré et al. 1998). All reduce the integration of synaptic inputs in the cell. Completely balanced models have been proposed, in which the mean input remains constant; instead the fluctuations drive the activity (Vreeswijk and Sompolinsky 1996; Shadlen and Newsome 1998). It is even possible to create networks with excitatory and inhibitory neurons that exist in a chaotic state which is characterized by highly irregular activity in all neurons (see below).
- Geometry and active properties of real neurons might be important as well.
- The activity might be sparser than originally thought. The reason is that the extra-cellular recordings tend to be biased towards highly active cells. Unbiased sampling has revealed that many neurons are very quiet. So the actual input rate might be lower.
- The different inputs have different weights, so the number of effective inputs might be much smaller than the above argument assumes.

Two final remarks on this issue: 1) With some simple estimates the approximate behaviour of a neuron was calculated. The realisation that this is not consistent with data, is a nice example of neural computational thinking (Shadlen and Newsome 1998; Holt et al. 1996).

2) It shows that it is not only necessary to have an accurate model of the cell at rest, but that an effective model for the cell*in vivo* is as important.

More reading on statistical description of neurons: Tuckwell (1988)

Chapter 7

A visual processing task: Retina and V1

In the following lectures we discuss a typical laboratory task that a monkey might have to perform. We will follow the neural signal through the different brain areas and discuss the neural systems which are involved in performing this task. The monkey sits in a chair in front of a monitor. A set of pictures flashed on the monitor. One of these pictures contains a banana. The monitor goes blank and after a delay a 'Go' command is shown. Now the monkey has to move a joystick to the point where the banana was. If the response is correct and within a certain time limit, it will get a reward.

Such a simple task highlights the various tasks the nervous system has to do:

- convert the picture into a neural signal
- process and recognise the image
- encode the task
- remember the location of the banana
- perform the arm movement, relative to the body position.
- finally, the monkey needs to learn to perform the task, guided by the rewards it can earn.

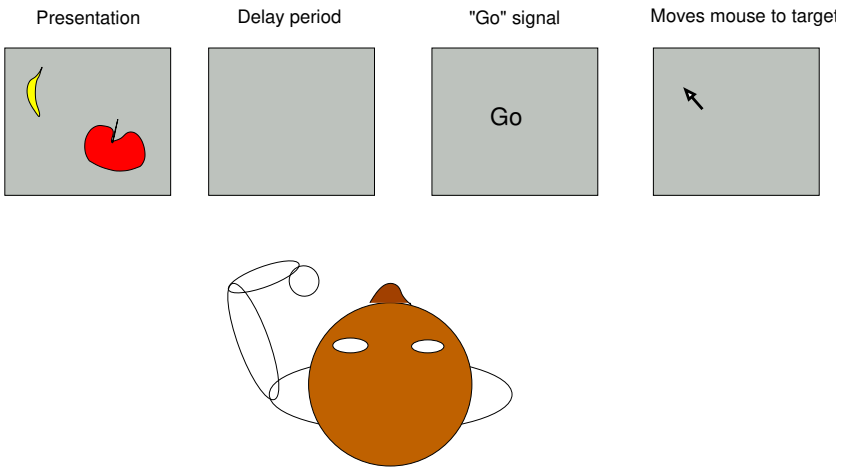


Figure 7.1: Experimental set-up for our thought experiment.

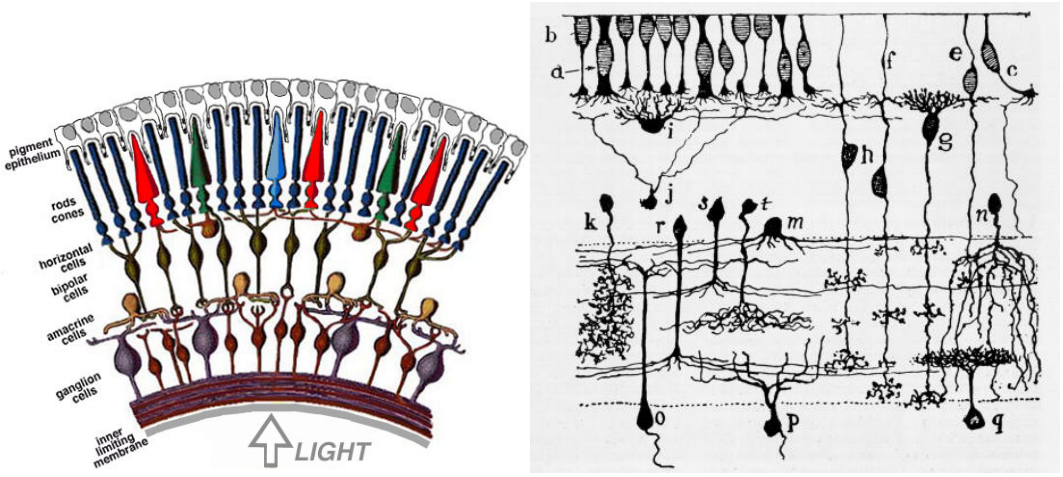


Figure 7.2: Left: Diagram of the retina and its different cell types. Note, that the light passes through all cells before reaching the photo-receptors rods and cones. The ganglion cells send the retinal output signal to the rest of the brain.
Right: Drawing by Ramon y Cahal the famous 19th century anatomist, who said of the retina:"I must not conceal the fact that in the study of this membrane I for the first time felt my faith in Darwinism weakened, being amazed and confounded by the supreme constructive ingenuity revealed not only in the retina and in the dioptric apparatus of the vertebrates but even in the meanest insect eye."

As we analyse this task in neural terms, we will encounter many properties of neurons and their computation, but also many unknown aspects.

We will start with description of the visual system, which has been studied most extensively of all sensory modalities. The few of the computational challenges for the visual system are: 1) the large amount of (partially redundant) information arriving every time, 2) the large dynamic range of the signals and the noise that dominates at low light levels, and, 3) the need to have robust classification and identification of, say, objects.

7.1 Retina

The description of the processing of our task starts with the retina. The retina is a thin layer of cells in the back of the eye-ball which convert the light into neural signals, Fig. 7.2. In the centre is the fovea, which is a high resolution area. The resolution gets less and less further away from the fovea. Except for the ganglion cells, which have as one of their tasks to send the signal to the thalamus, the neurons in the retina do not spike. Instead, the neurons use their graded membrane potential and neurotransmitter is released all the time with a rate that is a function of the membrane potential. Nevertheless most retinal neurons do have voltage-gated channels which help to adjust their characteristics which is useful in the adaptation process.

The ganglion cells come both in ON and OFF types. The ON responds maximally to a white spot on a black background, whereas the OFF cell likes a dark spot on a white background. (This is an interesting symmetry, that one does not encounter often in the brain.) The second distinction is between the cells in the M- and P- pathway. The P have a fine spatial resolution, are slow and some are colour selective. The M cells encompass larger receptive fields and therefore have a lower spatial resolution, but they are faster. The M and P pathway are processed more or less in parallel in the retina, LGN and cortex.

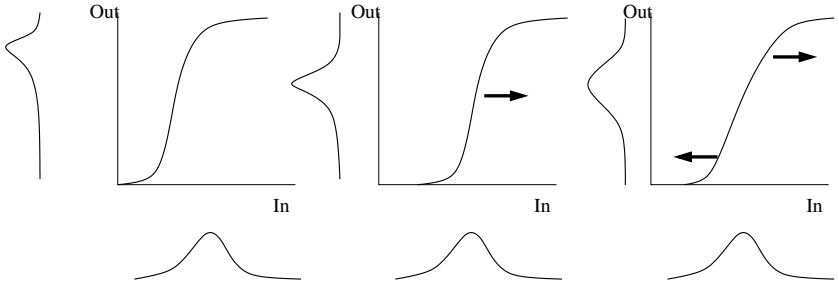


Figure 7.3: Adapting the neurons response to one feature of the input statistics, namely the probability distribution of the pixel intensities. A (mismatched) input-output relation (top) and the distribution of the input signal (bottom). Middle: Adaptation to the average signal. Right: Adaptation to the variance in the signal (contrast adaptation).

7.1.1 Adaptation

The retina works over a very large range of possible light intensities: a sunny day is about 10^{10} times brighter than a starlit night. The dynamic range of the output (which is defined the ratio the strongest and weakest perceptible signal) is about 100 (assuming rate coding with firing rates ranging from a few to 200 Hz). So adaptation is required. (Opening and closing the pupil only contributes a small bit to the total adaptation).

The principle of adaptation is shown in Fig. 7.3. Adaptation can be understood as changing the input/output relation of cell such that the dynamic range of the cell is optimally used. Like on your TV, both brightness, or luminance, and contrast are regulated, to map the mean light level, $\langle I(x) \rangle$ and its fluctuations $\langle I(x)^2 \rangle$ to the full output range of the retina without saturation[1]. Adaptation is implemented at many levels in the retina, from the photoreceptor to the ganglion circuit. There are both global mechanisms (mediated through neuro-modulators such as dopamine), and local mechanisms. The importance of these mechanisms becomes clear when suddenly the light turns on or off, and the adaptation state of the retina has not been adjusted yet - you won't see much.[2] In particular, adaptation to the starlight (or scotopic) regime, takes about 20 minutes.

7.1.2 Photon noise

In detecting the signal, the retina has to deal with the photon noise. Most light sources emit photons according to a Poisson process. The photon noise is most clear at night; the image that one sees has a snowy pattern superimposed on it. This is the photon noise. Indeed, the retina can almost see single photons and works close to the hard physical limits in obtaining this performance (Bialek 1987).

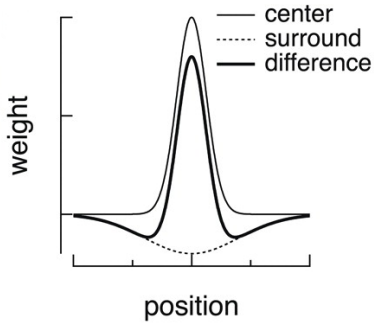
Already in dimly lit rooms, photon noise is discernible. This provides a good reason why the pupil can not do all adaptation. That would limit the number of photons impinging on the photo-receptors, increasing the relative noise. Rather the gain in the system is reduced at higher light levels. At low light levels one obvious way to reduce the photon noise is to average the signal either temporally or spatially, but of course this will come at the cost of reducing temporal/spatial resolution.

7.1.3 Spatial filtering

In addition to luminance and contrast gain control, the image is spatially and temporally filtered to improve performance. The question what is the best (linear) filter to process the image is can be

[1] The typical intensity distribution of natural images is log-normal

[2] In the visual sciences contrast is typically defined as $c = (L_{max} - L_{min})/(L_{max} + L_{min})$, where L_{max} (L_{min}) is the maximum (minimum) luminance in the image. Defining it as the relative standard deviation is mathematically more convenient.



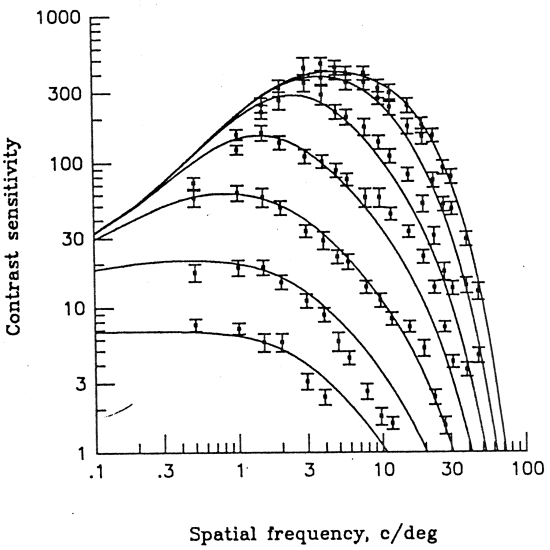


Figure 4: The family of solid curves are the predicted retinal filters (equation 2.7) at different I_0 separated by one log units, assuming that the dominant source of input noise is quantum noise ($N^2 \sim I_0$). No other parameters depend on I_0. The fixed parameters are $f_c = 22$ c/deg, $\alpha = 1.4$, $\rho = 2.7 \times 10^5$, $N' = 1.0$. The data are from human psychophysical contrast sensitivity measurements of Van Ness and Bouman (1967).

Figure 7.4: Left: Receptive fields of retinal ganglion cells can be modelled by a difference of Gaussians. This yields a bandpass spatial filtering. By reducing inhibition it becomes more low-pass.
Right:Spatial transfer function of the retina at different mean light intensities (lowest curve: lowest light level). At low light levels the transfer function is low-pass; at higher light levels it becomes band-pass. From Atick and Redlich 1992.

answered using information theory. Here we discuss this argument for the spatial filtering (Hateren 1992; Atick and Redlich 1992; Barlow and Levick 1965).

The argument is that one task of the retina is to remove redundancies in the signal. Redundancy means that the information sent by a given neuron was already (partly) present in the information that another neuron sent. Redundancy reduction is perhaps needed because the optic nerve has only limited capacity, or because the nervous systems can not process all incoming information. Redundancy can be removed by de-correlating the signal, such that its auto-correlation function is just a δ-function in space. That means there is no correlation (except at zero), which means that neighbouring pixels can not be predicted from the pixel under consideration. From our discussion of correlations and power-spectra, we know that this correlation function corresponds to a flat ('white') spectrum (cf Section 4.7.2). This means that the most informative coding strategy will whiten the spectrum of the output.

Remarkably, although visual input appears to have much randomness, the spectrum of visual input is not white. The spectrum of natural images behaves as $1/f_{spat}^2$, that means that natural image have a tendency to have much more low frequency components than high frequencies (you are encouraged to check this with Matlab, take a complex enough image and compute the power-spectrum). Whitening the signal would mean to multiply the signal with f_{spat}^2, i.e. high pass filtering.[3]

[3]There is a cut-off somewhere: it would not make sense to increase spatial frequencies which are beyond the eye maximal resolution. The resolution of the eye is determined by lens scatter and the number of ganglion cells in the

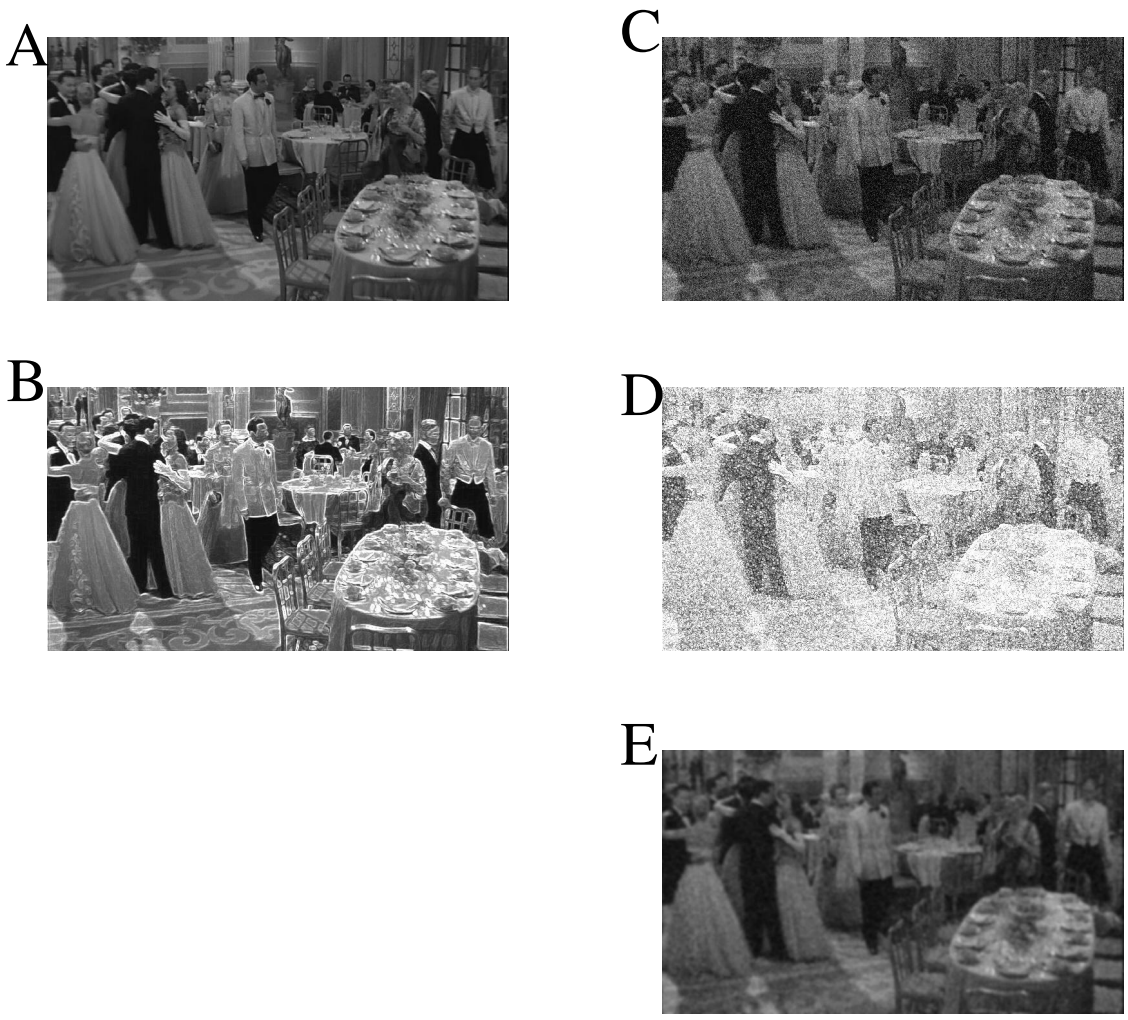


Figure 7.5: A) Original image. B) Edge enhanced version (high pass filtered). C). With noise added. D). The same edge enhancement does not gain anything. E). In this case low pass filtering works better. Note: made with GIMP for illustration purposes only, the filters were not exact.

However, because the signal is corrupted with photon noise, this filtering would be catastrophic, as the noise has a flat power-spectrum. Therefore multiplying the signal plus noise with f_{spat}^2 would cause most bandwidth would be taken up by noise. Instead there is a trade-off between whitening and getting too much noise. A bandpass filter is the result, Fig. 7.5.

Therefore (at normal light levels) the resulting spatial filter has a Mexican hat, or centre-surround profile. Its centre collects signals from neighbouring cells, thus doing a low-pass filter to reduce high frequency noise. The inhibition removes partly low-frequency components in the input. The result is a bandpass filter. In contrast, at low light levels the optimal filter is purely low-pass, Fig. 7.4. Similarly, at low light levels the temporal integration time increases ('longer shutter time').

The idea of removing correlation (redundancies) in the visual signal before sending it on the cortex dates back to Barlow (1958). The typical images are also redundant in time and colour-space (Buchsbaum and Gottschalk 1983). Redundancy reduction is a compression algorithm, a bit like JPEG compression. This sort of information reduction is probably going on at many stages in the brain. Our retina sends some 10^6 axons, with maybe some 10 bits/sec each. Yet, consciously we

retina.

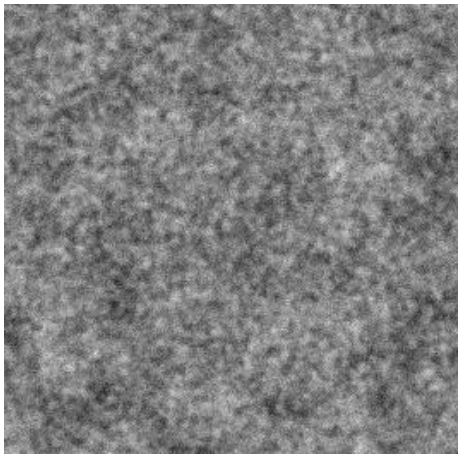


Figure 7.6: An artificially generated image with roughly $1/f^2_{spat}$ power spectrum. Despite the correct first and second order statistics it does not look very natural. Higher order statistics are important.

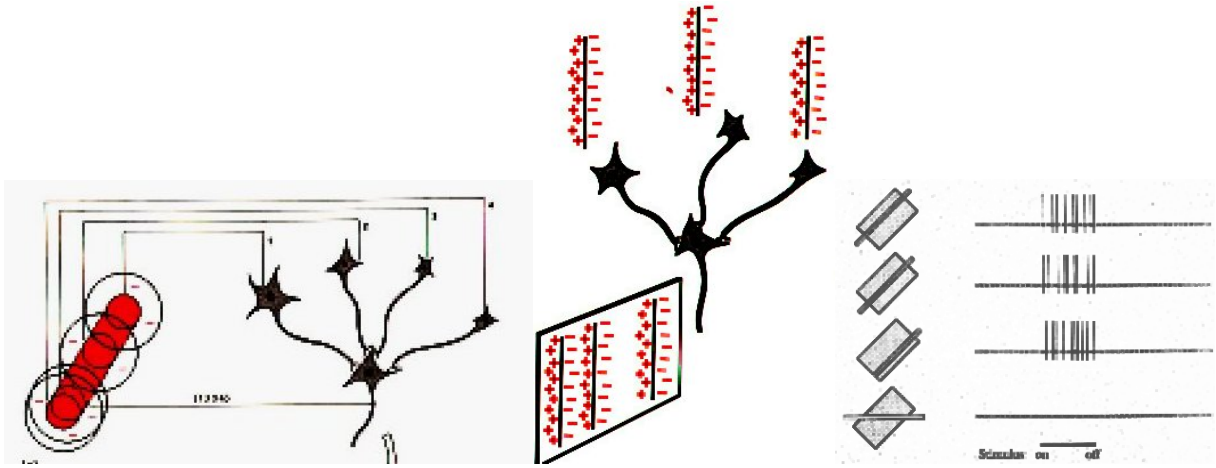


Figure 7.7: Left: Simple cell model according to Hubel and Wiesel. Middle: Corresponding feed-forward model of a complex cell. Right: Complex cell response.

process much less information (estimate are about 100 bits/sec).

These properties of the input are so called second order statistics, as the calculation of the correlation uses terms such as $\langle I(x)I(x+\Delta x)\rangle$. It is unclear yet if similar principles can be used for higher level (smarter) stages of processing. In principle, higher order statistics could yield additional compression tricks. Natural images might have a $1/f^2_{spat}$ power-spectrum, but have just having $1/f^2_{spat}$ statistics does not yield a natural looking image is shown in Fig. 7.6. This means that higher order statistics are important in describing natural images, and presumably are important to describe the visual cortex.

7.2 Primary visual cortex, V1

The ganglion cells project to the LGN (also called thalamus). The LGN is considered a relay station, blocking and allowing external input to enter the brain, its cells have similar receptive fields as the ganglion cells.

After passing through the LGN or thalamus, the visual signal enters the primary visual cortex,

V1. Most prominent is the existence of simple and complex cells in V1, which were discovered by Hubel and Wiesel. They were measuring cells in the visual cortex, they first used stimuli commonly used for the retina: circular spots. However, these stimuli caused only small responses in V1. But when they presented a light bar (allegedly accidentally when they replaced a slide) a much stronger response resulted. These cells are so called SIMPLE CELLS.They prefer a white bar next to a black bar and thus act like edge detectors.

A number of computational reason for these type of receptive fields has been proposed, such as sparse coding, wiring length minimization and ICA-like decomposition of the images (Hyvärinen, Hurri, and Hoyer 2009).

Another cell type is the complex cell, which still responds to edges, but does not care about the exact spatial position of the edge. This is most clear when a drifting grating is presented. The simple cell will only be active during a half period, whereas the complex cell will be active most of the time. Hubel and Wiesel came up with a simple hierarchical model to describe these responses. The simple cell gets its input from LGN cells, and the complex cell gets its input from the simple cells, Fig. 7.7.

Like the ganglion cells and LGN cells, the simple cells are reasonably described by linear filters, followed by a non-linearity. The firing rate of cell i can be written as

$$r_i(t) = f[\int d\mathbf{x} \int_{-\infty}^{t} dt'\, I(\mathbf{x}, t') K_i(\mathbf{x}, t - t')]$$

in which $I(\mathbf{x}, t)$ is the input image, K_i is the spatio-temporal filter (or 'kernel') of the cell which depends on the feed-forward input that this cell receives (and perhaps also lateral and feedback input), and f is a non-linearity, such as a rectification, to describe the F/I curve and impose that firing rates can't be negative. There are systematic methods to determine K and f experimentally for a given simple cell.

The complex cell rate can in this scheme for instance be written as,

$$c_k(t) = f(r_i(t) + r_j(t))$$

where i and j are opponent simple cells. The validity of these models can be quantified by the prediction quality of the firing rate on novel stimuli.

The primary visual cortex maps many features of the input: there is a systematic relation of the location of the cell to its function. Different simple cells have different directional tuning, but neighbouring cells will have similar tuning curves, Fig. 7.8B. But the visual cortex also has other structure: especially in the input layer (layer 4) there are OCULAR DOMINANCE COLUMNS, in which cell receive mainly input from either right or left eye, but not from both Fig. 7.8A+B. The other layers are less monocular. Finally, the primary visual cortex is also RETINO-TOPIC. Neighbouring cells have neighbouring receptive fields, Fig. 7.8C. Not surprisingly, more cortical area is dedicated to central or foveal vision. Moreover, the cortical representation is over-complete: there are many more V1 cells than there are retinal ganglion cells, their activity.

The mapping of the various features in the visual cortex (left/right eye, orientation selectivity, retinotopy, ...) has been a fruitful modelling subject, and both activity dependent models (among others by Jim Bednar) and growth factor dependent models (amongst others by David Willshaw) have been developed.

7.2.1 Reverse correlation

The Hubel and Wiesel experiments saw a clear response of the simple cells to a bar. But is this all that the cell does? How well does a bigger bar work, how about a 'T'-shape instead of a bar? This raises the general question: how to measure the receptive field of a cell without biasing the result, or, how to know what the cell codes for? A big problem in answering this question is the enormous dimension of the stimulus space; it has as many dimensions as there are pixels in the image. For a something simpler such as a stretch receptor, Fig. 9.1 , this is much less of a problem.

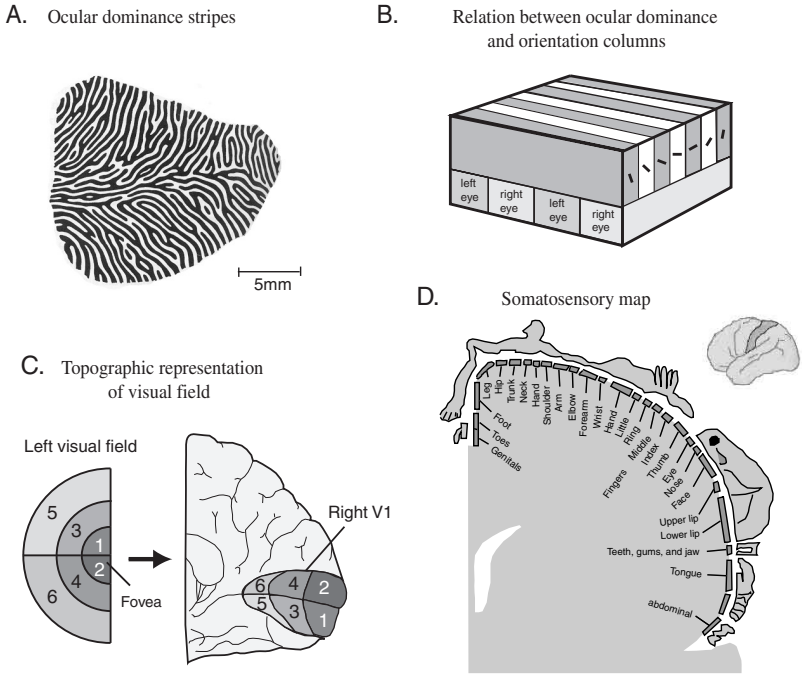


Figure 7.8: A) The ocular dominance columns in visual cortex. Left eye is dark, right eye is white (note the scale bar; comparable to your fingerprints !). B) The orientation selectivity has a columnar layout in the primary visual cortex. C) The primary visual cortex has a retinotopic map: nearby retinal cells project to nearby cortical cells. The fovea is over-represented compared to the periphery. D) Also somato-sensory cortex has a regular mapping.From Trappenberg 2002.

One way to measure the receptive fields of, for instance, simple cells is called reverse correlation or m-sequences. Here one stimulates the cells with a white noise pattern (uncorrelated across space and time). Every time the cell spikes, one records the preceding stimulus, Fig. 7.9. Apparently, the cell liked this stimulus. After measuring many spikes, one collects all the successful stimuli and averages them. This way one obtains the spatial (and temporal) receptive field.

A good check for the reverse correlation method is to present a somewhat smoothed stimulus, and then try to reconstruct the stimulus from the spike train, Fig. 7.10. The match between reconstructed stimulus and actual stimulus can be quite good (Rieke et al. 1996).

When applied to simple cells one stimulates with spatial noise pattern (like television with unplugged antenna) we find indeed a bar-like reverse correlate: the simple cells typically like a bright flank on one side, and a dark flank on the other, Fig. 7.9 right.

This version of the reverse correlation technique only works for (nearly) linear systems. For complex cells, for instance, the reverse correlation method already fails. This is simple to see: both a stimulus with dark left and bright right, and its inverse, dark right and bright left, will excite the complex cell. If both stimuli cause the same response, the reverse correlate averages out. A new development is to use 'higher order' techniques in this case (Touryan, Felsen, and Dan 2005). These methods can work well, but require much more data.

7.2.2 Role of lateral and feedback connections

Despite an enormous research effort, the precise circuitry of the V1 cortex is still not known. The responses of the (excitatory) cells are pretty well known. However, knowing the responses of the

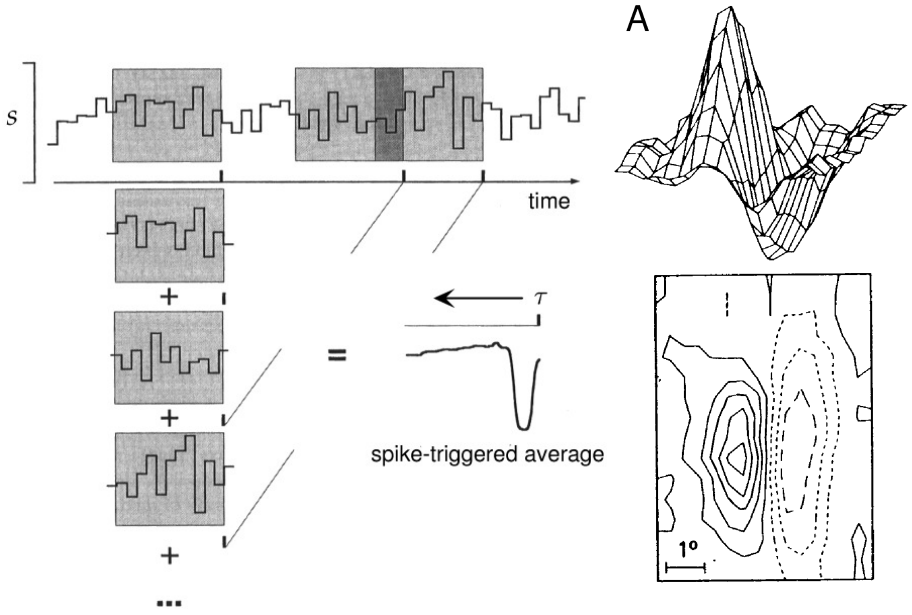


Figure 7.9: Left: Principle of reverse correlation. A random stimulus is given and every time spike is produced the preceding stimulus is recorded. The average gives the SPIKE TRIGGERED AVERAGE. Right: the spike triggered average thus constructed for simple cells in V1. From Rieke et al. 1996 and from Jones and Palmer 1987.

cells does not tell us what the underlying circuitry is. Where are the cell's inputs, excitatory and inhibitory, coming from? From LGN inputs, from neighbouring cells, feedback from higher layers? Various techniques have been developed to try to get at this question: cooling parts of the cortex to stop either feedback or lateral input, applying pharmacological blockers, and making lesions. However, neither of these techniques are without problems,

We have discussed that the original Hubel Wiesel feed-forward model, Fig. 7.7, is roughly correct at least when simple stimuli are presented. One thus wonders: what is the role of the many lateral and feedback connections? Most input to a cortical neuron is from lateral and feedback connections and only 20% is due to direct feed-forward input. One role of the feedback is thought to be attentional effects.With smart experiments one can change the attentional state of a neuron. One can ask the monkey to detect a small change in a visual stimulus, without allowing it to change its eye-position. Neurons show a stronger response (+20%) when attention is directed to their receptive field, Fig. 7.11. Attention is an interesting phenomenon because it*has* to be a feedback effect.

Also contextual effects might be mediated through lateral or feedback connections. Usually fairly simple stimuli are presented in studying the visual system, but when more complex stimuli are presented sometimes effects of parts of the image which fall outside the receptive field are observed (Heydt, Peterhans, and Baumgartner 1984). An example is shown in Fig. 7.12. We have seen that in lower visual areas there is usually a centre-surround layout of the receptive field, however for higher order stimuli the story is more complex. The response of the cells can be non-linear, as in the figure, so that the interpretation in terms of linear filters is fraud.

The precision of spike timing in visual cortex is still a matter of debate (see later as well). In Fig. 7.13 it is shown that response become more precise and sparse when natural stimuli are used. The cortex is probably not optimized to process typical laboratory stimuli, like bars and (drifting) gratings.

All these issues are difficult to study: The stimulus space is gigantic: think of all different 1000x1000 pixels colour images one can construct... Such a large stimuli space makes it difficult to see which stimulus is optimal for a neuron and how stimulus elements interact. Secondly, and

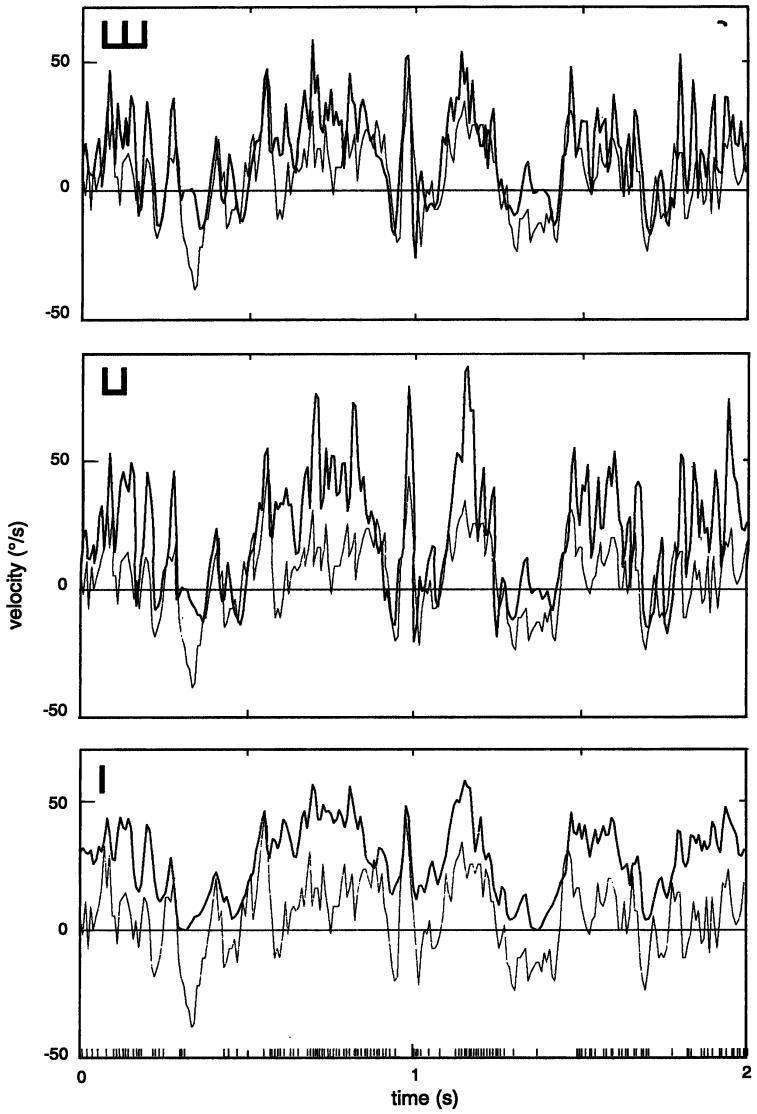


Figure 2.16
Reconstructions (dark lines) of angular velocity (thin line) using reconstruction depth of 1, 2, and 3 spike sequences (from bottom). Reconstructions using only single spike sequences (bottom) capture large fluctuations in the stimulus but miss many details. Including sequences of two spikes (middle) improves the reconstructions, but clearly the reconstructions systematically overestimate some aspects of the stimulus. These systematic errors are reduced in reconstructions based on triplets of spikes (top).

Figure 7.10: Stimulus reconstruction. The lower graph shows the measured spike train and the reconstruction. The other panels show higher order reconstructions. Fly motion sensitive neuron. From Rieke et al. 1996.

related to the first point, the shown contextual effect only appear for certain (low) contrasts. But these contextual effects might very well be important in explaining certain aspects of perception, Fig. 7.12, right.

More reading: NIP course, CNV course, Dayan and Abbott (2002) and Rolls and Deco (2002)

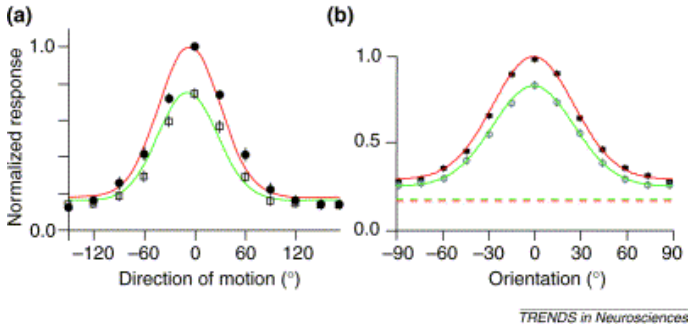


Figure 7.11: Attention increases the neuron's response. Upper curves are with attention (the neuron fires more), lower curves without attention. Left for area MST (a higher motion area) neurons, right for V4 (higher visual area).From Treue 2001.

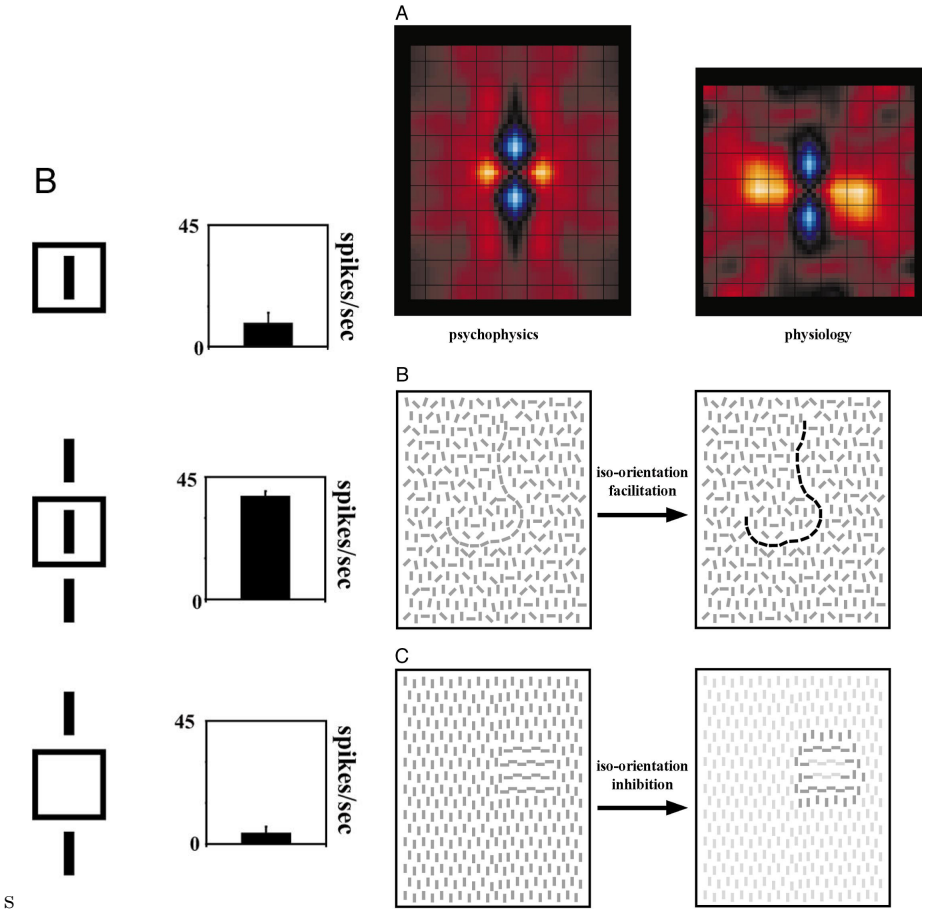


Figure 7.12: Contextual effects in V1. Left: non-linear interaction of centre stimuli and its context. The response to both centre and co-linear segment is much stronger than either parts alone. Right: The interactions in the vertical direction of the grey scale plots are excitatory, in the percept they lead to line completion. The interactions in the horizontal direction are inhibitory (parallel lines). Perceptually such interaction can cause pop-out of connected structures (B and C).From Kapadia, Westheimer, and Gilbert 2000.

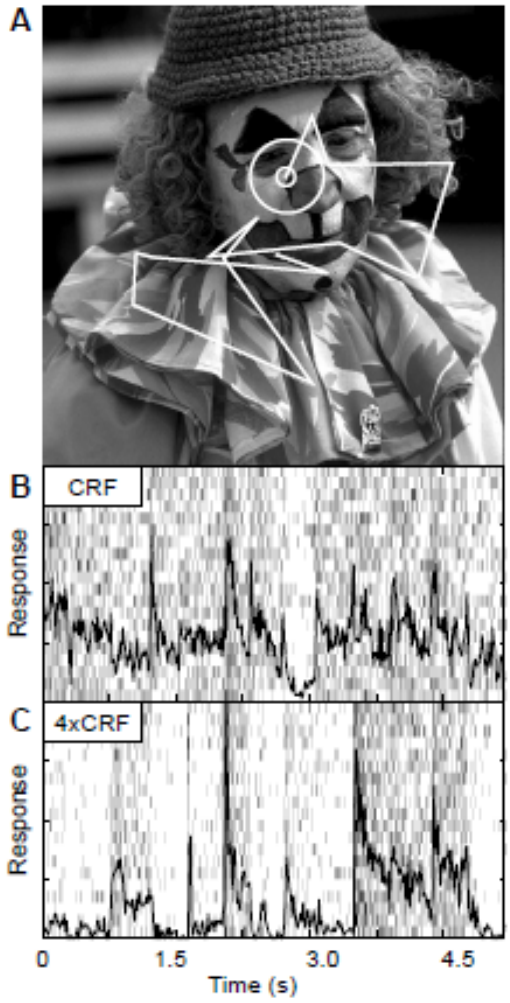


Figure 7.13: A) Diagram of monkey's eye movements on a picture. B) The firing rates of a neuron when the eye movements are replayed as a movie with a narrow aperture (small circle in A). C) Spiking is more precise and sparse when replayed with a wide aperture.From Vinje and Gallant 2000.

Chapter 8

Higher visual processing

In this chapter we discuss how the visual cortex further processes the image that is presented to the monkey. How does the monkey recognise the banana?

8.1 The importance of bars and edges

We have seen that retina and LGN cells respond to spots, simple cells in the primary visual cortex respond to edges, and complex cells respond to bars independent of their spatial phase. Edges of objects are indeed important in image processing and object recognition. Or from a more information theoretic point of view, edges contain a lot of information: Instead of describing an image using intensity, using just edges can be much more efficient.

There are interesting problems here: How well can one reconstruct an image from the responses of the simple cells? Why is making good line drawings so difficult, both for humans and computers? In a line rendering junctions of edges are of special importance. Junctions tell you which object is in front of which. In contrast, straight or slightly curved lines can be interpolated, and are therefore somewhat less informative.

8.2 Higher visual areas

From the primary visual cortex the visual signal goes to higher order areas. There seem to be two main pathways: one involved with the location of the object (the 'where' pathway or dorsal pathway). The other pathway is concerned with the identity of the object (the 'what' pathway or temporal pathway). In our task, both pathways will be important. The monkey has to recognise the object and know its location. How precisely this information is linked is not known (see chapter 9). Here we will look into the 'what' pathway.

In the early visual system the receptive fields are simple, and in the higher visual areas they become more complex. This suggests there is a hierarchical stack of processing layers (perhaps some 10 stages) in which the image is processed more and more until in the end objective selective cells pop up. The speed by which visual images are processed is very quick. Even complicated natural scenes are processed very fast: for example categorisation whether an image contains an animal or not, is done in some 150 ms of which 40ms is already taken by the retina (Thorpe, Fize, and Marlot 1996). Of course, we could have guessed this from daily experience, but this figure stresses the computational demands in visual processing.

8.2.1 Sparse coding and representation

How are objects represented in the higher visual areas? One approach taken is to pick a neuron and try to determine its receptive field. First see which stimulus from a large set causes a response. Next, one can try to see which stimulus is optimal for the neuron and determine the tuning. Such

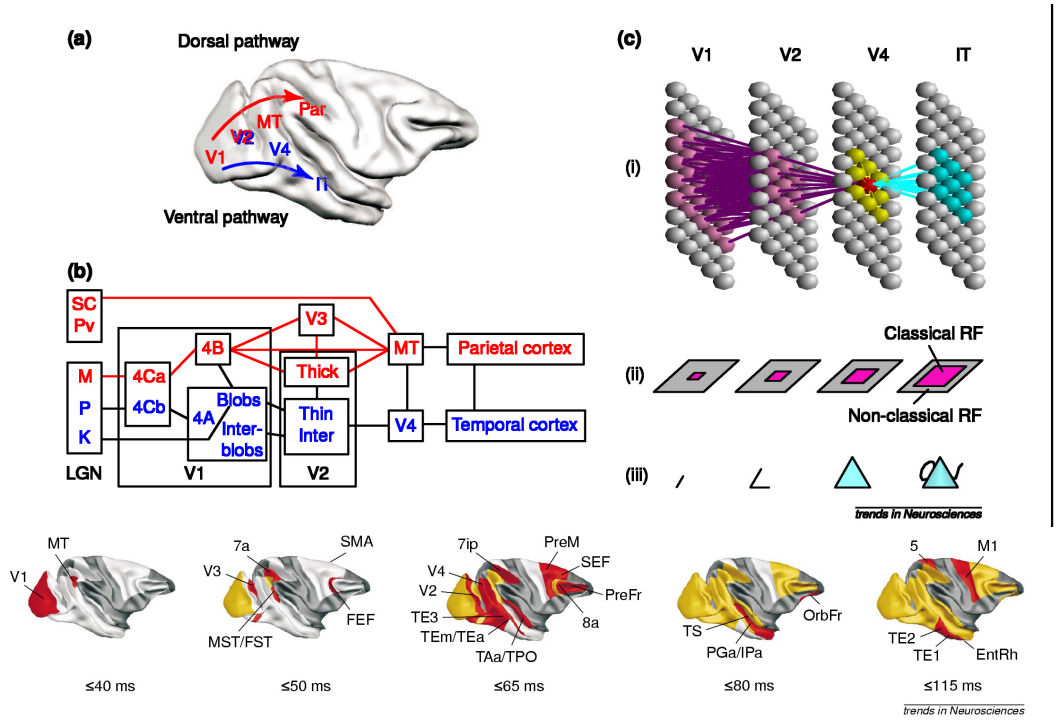


Figure 8.1: Top: Schematic of the different visual areas. Bottom: activity as it sweeps through the visual system. Both fromLamme and Roelfsema 2000.

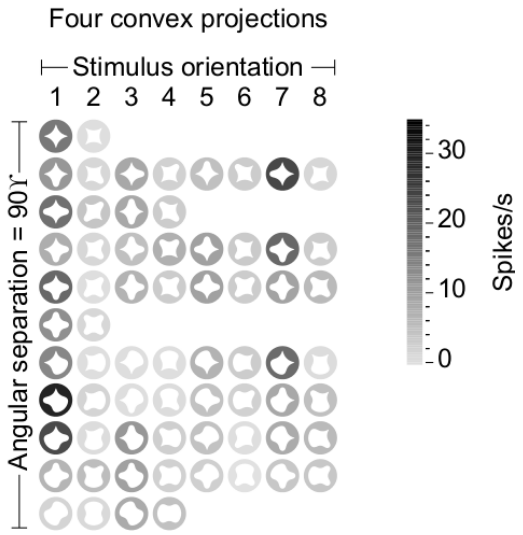


Figure 8.2: Coding in intermediate visual area V4, in this case analyzed in terms of square star-like objects. The darker backgrounds indicate higher response rates, they were not part of the stimulusPasupathy and Connor 2001.

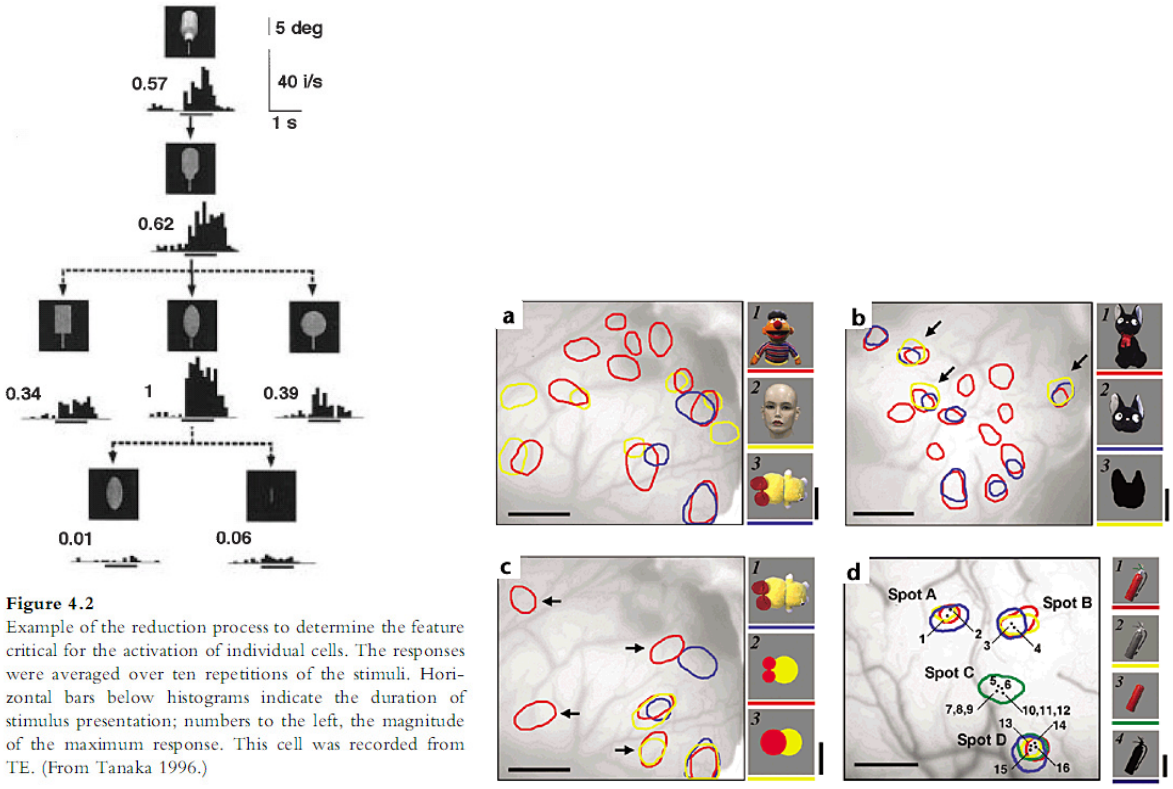


Figure 4.2
Example of the reduction process to determine the feature critical for the activation of individual cells. The responses were averaged over ten repetitions of the stimuli. Horizontal bars below histograms indicate the duration of stimulus presentation; numbers to the left, the magnitude of the maximum response. This cell was recorded from TE. (From Tanaka 1996.)

Figure 8.3: Neural responses to complex objects in higher visual areas. Left: single cell recordings, right: optical imaging (scale bar 1mm). Left from Tanaka 1996, right from Tsunoda et al. 2001. (The right picture is hard to see in grey-scale. Consult pdf-file or original paper).

an experiment is shown in Fig. 8.3. The result is that the neuron responds to limited set of similar stimuli.

Clearly to read out the code, the response of other neurons matters as well. Optical imaging allows to observe the response of many neurons at the same time. Objects appear to be arranged in a columnar structure, not unlike the regular layout in V1. Similar objects seem to evoke similar responses, but this is a tricky question: what makes images similar in content?

The precise structure of the representation in these areas is still under debate. From the above experiments it seems that also higher representations of objects are represented by smallish populations of neurons (COLUMNS), like the population code we encountered in V1 and will study further in the next chapter. This makes sense, as neurons and architecture are similar across the cortex.

In theory we can have a few possible coding schemes. Suppose we have N neurons of which k neurons are activated by the stimulus. The possibilities are

- $k = 1$: grandmother cell. This has a low coding capacity; there are only N possible outputs (for binary units)[1]. Furthermore, it would be hard to generalize between stimuli, as each stimulus gives a completely different response. However, as an advantage this code can represent N stimuli at once, and some computations are easy.
- $1 \ll k \ll N$: sparse code. This has a decent capacity $\frac{N!}{k!(N-k)!}$. It can represent a few stimuli at once, and it can help with generalization (depends on details of representation).

[1]For rate coding units (Ch.10) the number of possible outputs per cell depends on the signal to noise ratio: $(rate_{max} - rate_{min})/\sigma_{rate}$.

TABLE 19.1
A Binary Code Showing the Distinction between Sparseness and Selectivity, Demonstrating the Internal and Semantic Properties of a Binary Neural Code

		Neurons										
		n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9	n_{10}	
Stimuli	a_1	0	1	0	0	1	0	0	0	0	0	.2
	a_2	0	1	0	0	0	0	0	1	1	0	.3
	b_1	0	0	0	1	0	0	1	0	0	0	.2
	b_2	0	0	0	0	1	0	0	0	0	0	.1 ← local
	b_3	0	0	0	0	1	0	0	0	1	1	.3
	c_1	0	0	1	0	0	0	0	0	1	0	.2
	c_2	1	0	0	0	0	1	0	0	0	0	.2 ← sparse
	d_1	0	0	0	0	0	0	0	1	1	0	.2
	e_1	0	0	1	1	1	1	0	0	0	1	.5 ← dense
	e_2	0	0	1	0	0	0	0	0	1	0	.2
		.1	.2	.3	.2	.4	.2	.1	.2	.5	.2	.24
			^				^			^		
			grandm.				narrow			broad		

Note: The 10 columns correspond to 10 neurons, while the 10 rows correspond to 10 views of 5 objects (a, b, c, d, and e). Sparseness and codeword density are properties of the encoding of an item across a neural population (row), whereas selectivity and breadth of tuning are properties of a neuron across the encoded items (columns). Sparsity (examples in marked rows): local: one active neuron per item; sparse: few active neurons per item; dense: many active neurons per item. Codeword density is shown on the right margin. Selectivity (examples in marked columns): broad: many effective stimuli per neuron; narrow: few effective stimuli per neuron. Breadth of tuning is shown on the bottom margin. Note that density and breadth of tuning only have to be equal on average (0.24 in this example). "Grandmother-cell": a semantic property of n_2, which fires if and only if one of the views (a_1 or a_2) of object "a" is present. It codes the category "a" explicitly, and it is not the most narrowly tuned neuron in this example.

Figure 8.4: FromFöldiák 2013.

- $k \approx 0.5N$: dense code (e.g. binary computer). This has the highest capacity, namely 2^N (for binary units). However, it can represent only one stimulus at a time. Furthermore it can be hard to extract information as every bits counts. For the same reason it is hard to combine it with learning. Unless special coding schemes are used, a single bit error can give strange results. Finally, it consumes more energy than sparse codes.

Note we are mixing two quantities: how many stimuli activate a certain neuron, and, how many neurons does a single stimulus activate. While on the average both these sparseness measures are similar for homogeneous populations, it is important to recognize the difference, see Table 8.4. In experiments, one commonly studies a single neuron's response sampled across many images. In that case neurons commonly have an exponential distribution of firing rates, i.e. sometimes it is very active, but mostly it has no or low activity when stimulated by an arbitrary stimulus. This means that the coding is sparse. One way to define sparseness is with

$$a = \frac{\langle r \rangle^2}{\langle r^2 \rangle} = \frac{(\frac{1}{M}\sum_\mu r^\mu)^2}{\frac{1}{M}\sum_\mu (r^\mu)^2}$$

where the average is taken over a set of (M) relevant stimuli. When the code does not have any variation $a \approx 1$, while computers have $a = 1/2$. In monkey visual cortex $a \approx 0.3$ and the firing rates follow roughly an exponential distribution (Rolls and Deco 2002). An example of this is shown in Fig. 8.5, this figure also shows that blocking inhibition causes the neuron to be active in response to some of the previously non-exciting stimuli, i.e. inhibition from neighbouring cells determines some, but not all of the selectivity. Of course, the choice of stimulus ensemble is crucial here.

In these experiments the code was also independent: each neuron contributes a comparable number of bits (Abbott, Rolls, and Tovee 1996) to the readout. There was not much redundancy, i.e. each extra neuron helped to further identify the face.

This representation of sensory information can again be compared to a dictionary in our written language. There is a trade-off in our language between having a large dictionary of very precise words where a single word can say all, or a smaller dictionary with less precise words in which more words are needed. In this analogy, the receptive fields of the neurons are the words in the dictionary, the stimuli (sentences or stories) are represented by a combination of activity across neurons (words). (Also compare to our discussion of hyper-acuity and population coding.)

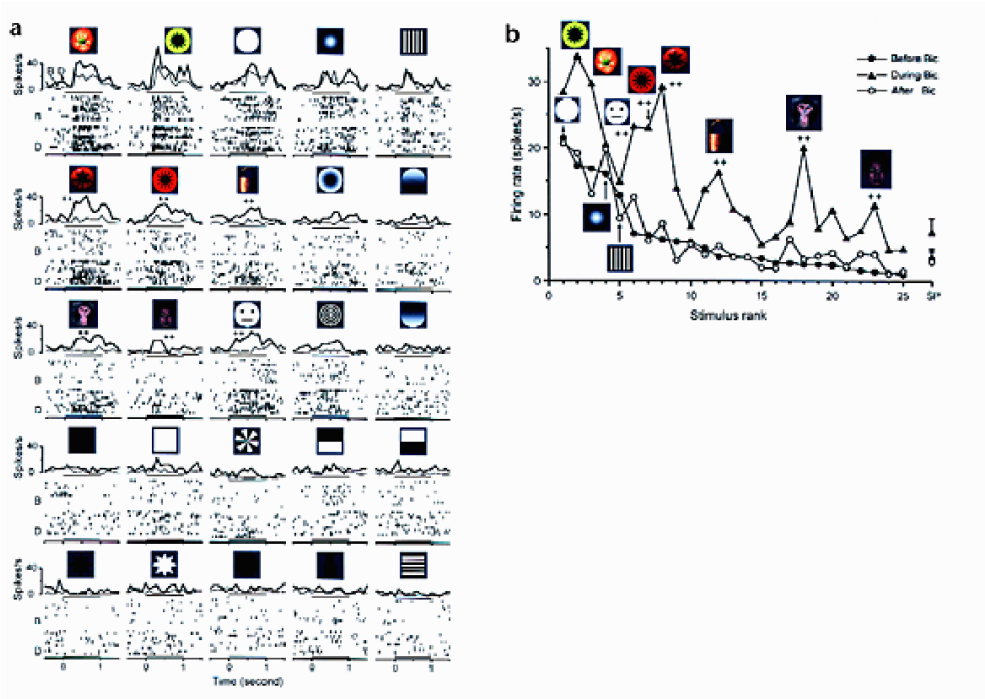


Figure 8.5: Stimulus selectivity for a single neuron in IT, both in normal condition and with inhibition blocked (Bic=Bicucilline, a GABA blocker). The selectivity of the neuron is determined by the mix of excitatory and inhibitory input it receives. FromWang, Fujita, and Murayama 2000 .

Sparse representations usually benefit machine learning algorithms. Another reason why sparse representations are used might be related to minimising energy consumption. The brain demands some 20% of the body's total oxygen and energy at rest (about 20 watts). Some models include this constraint to help to explain the observed representations (Stemmler and Koch 1999).

So the consensus is that the coding in the brain is SPARSE AND DISTRIBUTED.The fact that the code is sparse and distributed still leaves many possibilities (check for yourself). In analogy, saying that a computer works with ones and zeros, only tells you so much. It would be possible that the neurons have some random, yet sparse code. But there appears to be more order:
1) The neurons represent some stimulus feature consistently. (Although this is harder to verify in higher areas)
2) Nearby neurons have typically similar tuning properties.
3) The information in the neurons is not redundant (see above), because the responses are independent (yet partly overlapping).

8.2.2 Connectivity and computation

How do the activity patterns in the visual cortex help us in understanding the computations? Unfortunately, knowing the activity patterns in the cortex, does not tell us how the computations are implemented, or, in other words, how the cells are wired up. This is maybe the most challenging question in neurophysiology: how to extract computation from activity. (This problem is most clear with fMRI data, simply knowing which brain region is active during a task, is not very helpful.) A few techniques are currently used to measure the inputs to cells:

- Intracellular recording in vivo. This in principle allows to measure both excitatory and inhibitory inputs to a certain cell. However, technical problems, such as a bad space-clamp make this difficult.

- Multi-unit extracellular recording. In principle cross-correlations between neurons can be used to figure out connections. However, without more information, correlations are hard to interpret. Suppose two cells are found to have a strong correlation. Is this because they receive common input, or does one provide input to the other?
- Anatomy can be very useful as well, but it is very labour intensive.

8.3 Plasticity

The visual cortex is remarkably plastic and has an enormous storage capacity. After passively watching thousands of images, people can tell which one they have seen and which ones they didn't when tested later. All this information must be stored somewhere, most likely in (higher) visual cortex.

Indeed, the response of the neurons changes as monkeys becomes familiar with images, even when the monkeys have to do nothing but passively view them (Erickson, Jagadeesh, and Desimone 2000). In another experiment monkeys were trained to do image pairing tasks. It was found that cells pick up rather the pairing between the images rather than similarities in the shape of the images (Sakai and Miyashita 1991). These experiments show that the representation of objects is in a constant state of flux. It is surprising that the changes do not lead to the degradation of older memories.

Plasticity might be less in lower visual areas, but it is still there. In particular perceptual learning studies in humans and animals have been used to research how visual representation is altered when subjects learn to, say, detect faint low-contrast gratings in noise.

8.4 Computational models of object recognition

Building artificial object recognition systems is a long standing aim of computer vision and computational neuroscience. The challenge is that object recognition needs to be invariant towards viewpoint and lightning transformation (e.g. translation, rotation, scaling), but on the other hand it needs to be highly sensitive to minor changes in the input, for instance when distinguishing makes of cars or different faces. Invariances can be learned using the trace-rule, which links inputs that are presented subsequently (Földiák 1991). The idea is that visual input is not a random sequence but that the objects present in the scene do not change often, while the viewpoint might. There is some experimental evidence for such learning rules (Li and DiCarlo 2010). Recent biologically inspired systems of object recognition use combinatorially many neurons that calculate many features combinations (Fukushima 1980; Riesenhuber and Poggio 1999). Interestingly, in those models don't learn, except for the readout layer, which learns which neurons are active when a particular object category is presented. These models already show some invariance of the output with respect to translation or rotation of the image.

Anatomy indicates that there is no strict hierarchy in the processing stages, instead many cross and feedback connections exist. The above object recognition models often operate exclusively in a feed-forward mode. There is in principle no operation that a recurrent net could do that a feed-forward net could not do (given an arbitrary set of time delays). Therefore the question is perhaps: 'how do recurrent connections make the network more efficient?'

More reading: Rolls and Deco 2002

Chapter 9

Neural coding

In this chapter we look into the question how stimuli and actions are coded in single neurons and networks and formalize some of the interpretation of the data encountered in the previous chapter. Knowing the neural code is like having a dictionary that translates between the neural activity in the brain on the inside, to sensory stimuli (or motor output) on the outside. That is, if I know your neural activity and know the code, I know what you think.

Typically questions are: How should one decode? What is maximally attainable accuracy of a code? How does that relate to psychophysical accuracy?

We present a few coding schemes and methods to test hypotheses about coding experimentally.

9.1 Rate coding

When Edward Adrian in the 1920s measured the stretch receptor neuron in the muscle of a frog, he noted that with more force, the neuron fired with a higher frequency, Fig. 9.1. Rate coding is the

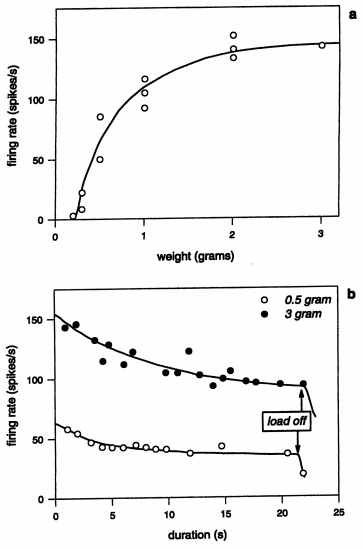


Figure 1.4
Rate coding and adaptation. (a) Average firing rate of a stretch receptor as a function of the weight applied to the muscle, in an experiment similar to that of Fig. 1.3. (b) Decrease in firing rate with time following the onset of a static stimulus at $t = 0$, adapted from Adrian (1926). This desensitization, or *adaptation*, is a general property of neural coding.

Figure 9.1: (Figure taken from Rieke's book)

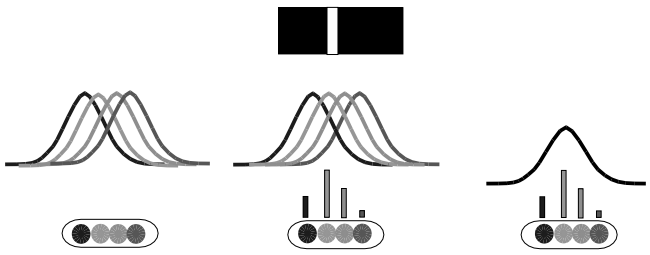


Figure 9.2: Population coding. Left: four cells with their tuning curves. The centres of the tuning curves are different, but they overlap. Middle: When a single bar is presented, all neurons respond, but with different amplitudes. Right: Hyper-acuity. The location of the bar can be estimated with a accuracy better than the distance between the neurons. The stimulus must have been located slightly to the right of the second neuron (because the third neuron is more active than the first neuron).

coding hypothesis that draws from this observation and states that all information is coded in the firing rate of the neuron. I.e. given the rate I can decode the force that was applied to the muscle. This can be compared to a coding scheme in which the spike timing relation between neurons matters. Rate coding is not very efficient in terms of information transmitted per spike, but it is quite robust against fluctuations in the precise spike timing. Methods like fMRI are silently based on rate-coding. Also many studies that record spikes, implicitly assume rate-coding in their analysis.

Rate coding requires a temporal integration window, which has been used to argue for rate coding being too slow for the fastest computations. However, when reading out from a population this window can be very short. Furthermore, there is no need for a down-stream neuron to wait for the end of the window, before it can start responding (see below).

9.2 Population code

Assuming rate coding for now, what happens when we consider multiple neurons? When a stimulus is presented to our monkey usually not just one, but a whole ensemble of neurons becomes active. We have already encountered an example of this effect, namely the simple cells in V1. Because the V1 neurons have a wide tuning curves (Fig. 10.10) many neurons will respond when a bar is presented. In analogy with the dictionary: a simple stimulus word in sensory language, leads to a long sentence with similar words in neural language. This distribution of activity over multiple neurons is called population coding (Paradiso 1988). This might at first seem a wasteful scheme: wouldn't it be more convenient if for a given stimulus only a few finely tuned neurons responds? But population coding has some interesting advantages. Here we consider coding of a continuous, angular quantity denoted with θ.

In Fig. 9.2 we show a example of population coding. Four neurons have overlapping tuning curves, denoted $f_i(\theta)$, where $i = 1..4$ indexes the neuron. When a stimulus is presented all neurons respond, albeit with different amplitudes. Although the neuron with the highest activity most clearly signals the location of the bar, the other neurons contribute as well. One advantage is almost immediately evident: a population code is robust against failure of neurons. In case a neuron fails the rough position of the stimulus can still be estimated; the information might be distorted but is not completely lost. Similarly, given that the response from a single neuron is noisy, using multiple neurons in parallel can speed up processing as it reduces the need for long averaging times. Because the neurons have slightly different tuning curves, this parallel processing is not redundant. If the tuning curves were narrow, only few neurons would respond to a given stimulus, making it more noisy.

Another interesting property of population codes is HYPER-ACUITY. Hyper-acuity means that the position of the stimulus can be estimated more finely than the distance between the neurons. Fig. 9.2(right) shows why hyper-acuity can occur. Because the third neuron reacts a bit stronger than the first neuron we know that the stimulus was slightly toward the third neuron (right of the second neuron). Hyper-acuity in our visual system was directly illustrated using Vernier lines: When one is presented with two co-linear line segments which are slightly displaced, one can discriminate very small displacements. Displacements that projected onto the retina are less than the distance between two photo-receptors. This was puzzling at first. However, it turn it turns out that the lens blurs the image just so much that strip of a few photoreceptor wide will be activated by even the thinnest line, so that the photoreceptor and subsequent retinal circuitry can employ a population code. This is quite different from computer pixels, although also there related methods are now being used (anti-aliasing, hyper-resolution)[1]It is important to note that the *resolution* of our vision is not enhanced by hyper-acuity. Hyper-acuity does not improve our ability to see a very fine pattern, only the position estimate of isolated stimuli is improved. Else blurring (incorrect glasses) would improve vision.

Population coding occurs in many, if not all, locations in the brain. For instance in primary visual cortex, in higher visual areas, in working memory, in hippocampal place cells, and in motor cortex, Fig. 9.3. A interesting question when using a population code is: how to estimate the encoded signal given the population response? A decent (but typically sub-optimal) read-out is given by the POPULATION VECTOR. In the population vector, every neuron votes with a weight proportional to its activity that the stimulus was their optimal stimulus. In the case of angles, the sum over all votes is best seen as a vector, Fig. 9.3.

$$\mathbf{p} = \sum_i r_i \mathbf{e}_i$$

where $\mathbf{e}_i$ is a vector with length 1, pointing in the preferred direction of neuron i. In experiments the thus constructed population vector gives a reasonable description of actual angle, see Fig. 9.3. The population vector will not be a very good estimate when the neurons have a high background rate and narrow tuning curves. In that case a lot of non-sense votes are included. These votes average out when many trials are included (the estimator is 'unbiased'), but worsen the trial to trial variations in the estimate, as we discuss below.

9.3 Fisher information

Using the so called Fisher information it is possible to make a more accurate statement about the information present in a population response (Seung and Sompolinsky 1993). How accurate can one estimate the encoded quantity? As one could expect from the example of Fig. 9.2, when the neurons are noise-free, the accuracy of the estimate is in principle infinite (the stimulus angle or location can be recovered with infinite precision). The question is therefore: given a certain noise level in the neurons, what the highest accuracy one can achieve.

Suppose we have a large population of N neurons that encode an angle θ. For convenience we suppose that they all have tuning curves which only differ in their centre $f_i(\theta) = A\exp[\cos(\theta - \theta_i)/w^2]$.

The error in the angle estimate can be calculated using the Fisher information[2], which is given by

$$I_F = -\int d\mathbf{r} P(\mathbf{r}|\theta) \frac{\partial^2 \log P(\mathbf{r}|\theta)}{\partial \theta^2} = \int d\mathbf{r} P(\mathbf{r}|\theta) \left(\frac{\partial \log P(\mathbf{r}|\theta)}{\partial \theta} \right)^2$$

[1]It is not a bad idea to blur the retinal image a bit anyway. Otherwise, one gets so called aliasing effects: stimuli with high spatial frequency (above the Nyquist frequency) can cause illusory low frequency percepts (like the Moire patterns sometimes seen on television).

[2]The Fisher information is not an information measure in the strict sense (bits). It is related though, for details see (Brunel and Nadal 1998; Yarrow, Challis, and Seri?s 2012).

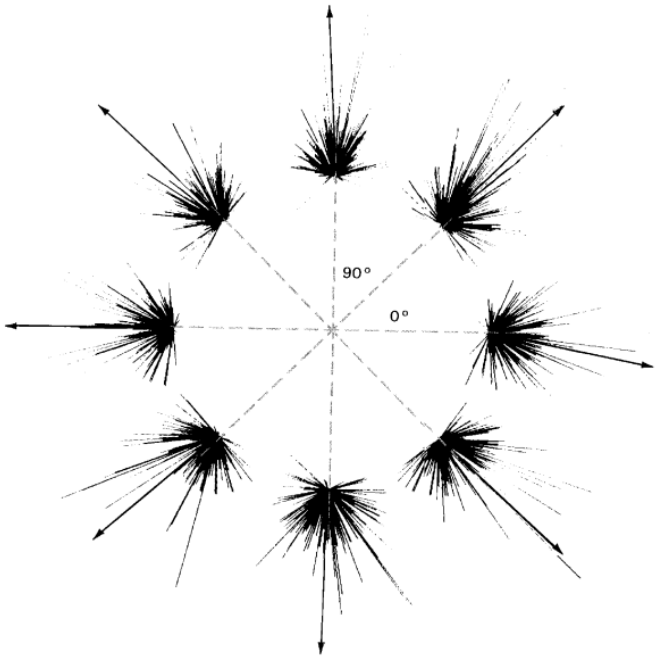


Figure 9.3: Population coding in the motor cortex. The neurons shown here code for a 2D arm movement. The actual movement directions of the arm are along the dashed lines. The population vector response is drawn with the thick arrows. There is a reasonable match between the actual response and the population vector. The line bundles give the response of the individual neurons, the sum of which gives the population vector.From Dayan and Abbott 2002 after Kalaska, Caminiti, and Georgopoulos 1983.

The Cramer-Rao bound says that it is not possible to estimate the original stimulus with a variance less than the inverse Fisher information (for a so-called unbiased estimator[3]), that is

$$\sigma_\theta^2 \geq 1/I_F$$

This inequality hold not only for angles, but for any encoded continuous quantity. Suppose the noise on the neurons is Gaussian, the probability for a certain response r_i from neuron i is $P(r_i|\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp[-(r_i - f_i(\theta))^2/2\sigma^2]$. When the neurons are independent, the probability factorizes: $P(\mathbf{r}|\theta) = \prod_{i=1}^N P(r_i|\theta)$. Hence

$$\begin{aligned}
I_F &= -\int dr_1 dr_2 ... dr_n P(r_1|\theta)P(r_2|\theta)..P(r_n|\theta) \sum_{i=1}^n \frac{\partial^2[-(r_i - f_i(\theta))^2/2\sigma^2]}{\partial\theta^2} \\
&= -\frac{1}{\sigma^2}\sum_{i=1}^n \int dr_i P(r_i|\theta)\left(r_i f_i''(\theta) - f_i(\theta)f_i''(\theta) - f_i'^2(\theta)\right) \\
&= \frac{1}{\sigma^2}\sum_i^n \left\{ f_i'^2(\theta)[\int dr_i P(r_i|\theta)] - f_i''(\theta)\int dr_i P(r_i|\theta)[r_i - f_i(\theta)]\right\} \\
&= \frac{1}{\sigma^2}\sum_{i=1}^n f_i'^2(\theta)
\end{aligned}$$

There are various things to note: 1) In general, the Fisher Information can depend on θ, but here we assumed a homogeneous system, so that all θ are coded equally well. 2) The more noise (

[3]An unbiased estimator estimates the angle correctly when given enough trials. A counter example is an estimator which always estimates the angle to be 30 °, of course the variance is in that case zero, but the estimate is useless.

σ^2), the less Fisher information; the more neurons, the more information. 3.) The information is proportional to the derivative on the tuning function squared. This means that there is no information when the derivative is zero, i.e. when the stimulus is right in the centre of the neuron's receptive field (!) or when the neuron fires hardly at all. Instead, most information is provided by neurons in the flanks. This makes sense, if we move the stimulus slightly, neurons which show a big change in response are most informative. This does not mean we can simply remove the un-informative neurons, because with another stimulus these neurons might become important.

The next question is how to read out the code, such that one reaches the Cramer-Rao bound. If the estimator is able to reach the Cramer-Rao bound, one has done the best possible. The population vector above, is one method, but often not optimal. Various methods have been developed such as Bayesian estimates and maximum likelihood methods, see (Dayan and Abbott 2002) for a overview.

However, it is good to realise that except at the motor output the nervous system itself does not need to 'read out' the population code. Instead, it computes with it. How that differs from regular computation is not fully clear. One proposed scheme is to use Radial Basis Functions (RBF) (Hertz, Krogh, and Palmer 1991; Pouget and Sejnowski 1997).One variant of this coding scheme is as follows: Suppose one wants to calculate a function $f(A, B)$ of the variables A and B (A and B might for instance represent the head direction and the eye direction). A typical transformation is to calculate the sum of these angles. When quantities are populated coded, already a simple sum is actually a non-linear operation. Using radial basis functions such a sum is calculated as follows: One first creates a 2-dimensional layer with an activity equal to the (outer) product of the population activity in A and B, i.e. $f(x, y) = f_A(x).f_B(y)$. Next, a projection from this layer to an output layer implements the output function. Under the quite general conditions this allows for the calculation of arbitrary functions.

9.4 Temporal codes and synchronisation

In our discussion of population coding we considered the firing of neurons as independent. However, this need not be the case in general.

Suppose one measures the spike count over a few hundred ms in response to some stimulus. Repeated over many trials one finds a certain mean count and some variation. However, this variation is not independent from cell to cell: if cell A is firing at a higher frequency than normal, cell B is likely to be more active as well. This is called noise correlation. It could simply be due to both neurons receiving some of the same noisy input, or A helping to excite B. A typical values for the correlation coefficient for neuron which have a similar receptive field is $r = 0.2$ (for 200ms time-bins) (Zohary, Shadlen, and Newsome 1994). If fluctuations were independent between neurons they could be averaged out by collecting signal from many neurons. However, the presence of correlations will limit the amount of averaging one can do, and thus worsens the stimulus estimate (the Fisher information becomes sublinear, rather than linear in N).[4]

One basically finds that correlations with a range comparable to the tuning curve are most deteriorating. Short range and infinite range correlations can be filtered out (Sompolinsky et al. 2002).

However, tight correlation could also signal something and in some cases precise correlated firing was found between neurons. This synchronisation does not need to be stimulus locked, so only if one measures two neurons simultaneously one has a chance to observe it. Singer and co-workers have proposed that synchronisation might be help for solving the binding problem.

The binding problem is the following: suppose our monkey observes a banana on the left on the screen, and an apple on the right. The neural processing streams for form and location are separated. How does the nervous system know that the banana is left and the apple is right? It seems that the brain could easily confuse the stimulus attributes. However, we seldom have such experiences in life. Similarly, when an moving object is partially occluded, how does the nervous system know that disjunct parts can correspond to the same object when shape and motion are

[4] Although more recent theoretical studies challenge this view...

processed in different brain regions? This is called the BINDING PROBLEM, it says that needs to be some way for the brain to keep 'banana' and 'left' bound together (a whole issue of Neuron was dedicated to it, good reading). If the binding problem is really a problem is not clear. Synchrony could provide a tag to the neural signal, labelling features which belong to the same object. In our example all banana attributes (left, yellow, banana) could spike synchronously and independently all apple attributes could spike synchronously, solving the binding problem.

The data in Fig. 9.4 show that under some stimulus conditions, peaks arise in the cross-correlation between neurons, even though the neurons are far apart. The cross-correlation is defined in analogy with the autocorrelation. The cross-correlation between spiketrain 1 and spiketrain 2 is

$$C_{12}(\tau) = \lim_{T\to\infty} \int_0^T dt\, r_1(t) r_2(t+\tau)$$

where $r_i(t)$ is the discretised spike-train from neuron i.

The peak in the cross-correlation is usually quite small, so that it can not not carry much information. On the other hand if many neurons synchronise, synchrony might provide a very powerful stimulus to a downstream neuron. Synchronised input to a neuron is more effective in firing it than uncorrelated input. You can easily see this with the Integrate-and-Fire neuron. Using synchronisation is an elegant way to solve the binding problem. However, evidence is conflicting and precise conditions for synchrony are still debated.

In principle information theory techniques can be used to estimate the amount of information carried by synchronous spikes, this can help to estimate the importance of synchronous spikes, e.g. (Oram et al. 2001).

9.5 Hippocampal sequences

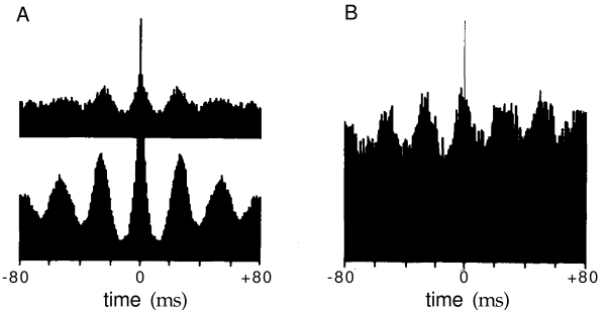
The hippocampus is a particularly interesting area for temporal patterns and sequences of spikes. In rodents, neurons in the hippocampus fire whenever the animal is in a certain location in space, so called place-cells. Hence, during movement cells will become sequentially active. When one looks at single neurons, one finds that neurons firing with a certain relation to the theta-rhythm (about 8Hz), so that first they fire with a high phase, and later they fire early wrt to the phase. This is called*phase precession*. On the population level this means that when the place fields overlap, experiences are also present on a much quicker time-scale than the experience itself, Fig. 9.5A. This temporal compression might help plasticity mechanism to bind together events that occur far apart from each other. See e.g. (Chadwick, Rossum, and Nolan 2015; Chadwick, Rossum, and Nolan 2016) for models.

Second, during sleep as well as quiet resting (Jadhav et al. 2012), sequences of activity are replayed by the hippocampus (Lee and Wilson 2002), Fig. 9.5B. This is generally believed to play an important role in memory consolidation. For instance, interrupting replay will disrupt memory.

9.6 Information theory

In section 9.3 we implicitly assumed that the neurons use a rate code: the response was measured in their firing rate. Although such rate coding is not in conflict with experimental data, various efforts have been made to see if the firing rate is the whole story, or if neurons also code information differently. How can we know if there is more information in the code?

Already a single neuron could in theory pack an enormous amount of information in its spiketrain when it would fire very precisely. A stupid example is a neuron fires at times $t = 1.0072$, 2.0069, 3.0076, 4.0080, 5.0033s. One might think the neuron just fires at a frequency of 1 Hz but it is somewhat noisy. But instead the neuron sends a secret message in the last digits ("HELP!" in ASCII). Without knowing this neuron's coding scheme it would have been difficult to decipher this spike train. When many neurons are active this issue is even more complicated. This problem is known as discovering the neural code.



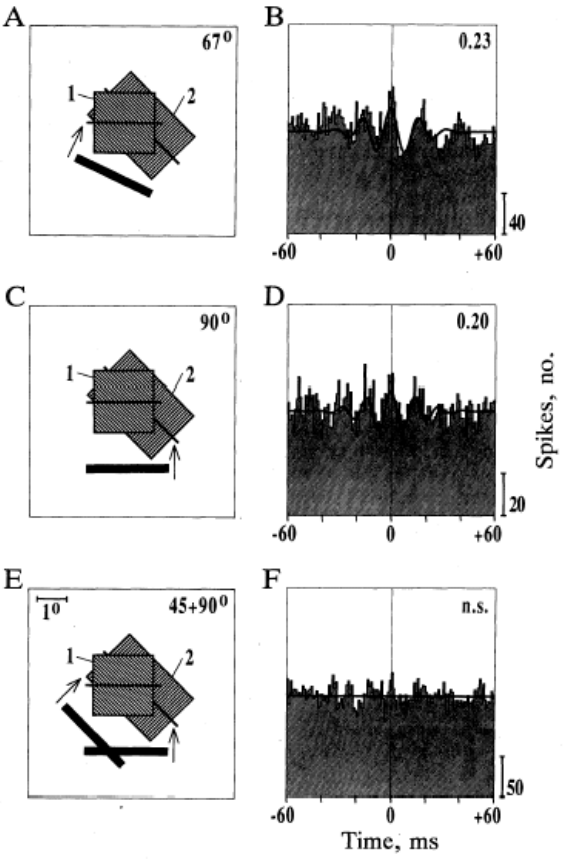


FIG. 2. Effect on conflicting stimuli on cell pairs with broadly overlapping orientation tuning. Recorded cells had a spatial separation of 400 μm and differed in their orientation preference by 45°. (*A* and *B*) Stimulation with a light bar of intermediate orientation elicited oscillatory responses at both sites that were synchronized to a significant degree. (*C* and *D*) Due to their broad orientation tuning, the cells at site 2 responded also to a horizontally oriented stimulus, which was optimal for the cells at site 1. With this stimulus, the two responses were also synchronized. (*E* and *F*) Adding the optimal stimulus for cell group 2 to the stimulus configuration eliminated the synchronization between the two responses (n.s.). Based on the evidence presented in Fig. 1, it can be assumed that cell groups 1 and 2 were still synchronized with other cells having the same respective orientation preference. Thus, these cell groups participated in two large assemblies desynchronized with respect to each other. In *B*, *D*, and *F*, the thick continuous line represents the Gabor function that was fitted to the correlogram. The number in the upper right corner indicates relative modulation amplitude. Vertical scale bars indicate numbers of spikes.

Figure 9.4: Left: A. Autocorrelation of two cells. B: The cross-correlation shows a peak, despite the neurons being millimetres apart. Right: Data from the same paper, showing that the correlation depends on the stimulus: disjunct stimuli do not cause synchrony.Left figure from Dayan and Abbott 2002 after Engel, Konig, and Singer 1991(?). Right from Engel, Konig, and Singer 1991

Using information theory it is in principle possible to measure the amount of information in a signal without knowing the coding it uses, or imposing a coding scheme. It is therefore a very clean way to answer questions about coding. Nevertheless, one should keep in mind that information theory does not imply that all information present is actually used. Practical use is limited to low dimensional measurements (i.e. only a few time bins, only few neurons in parallel), as the amount of data required rises very rapidly with the dimension.

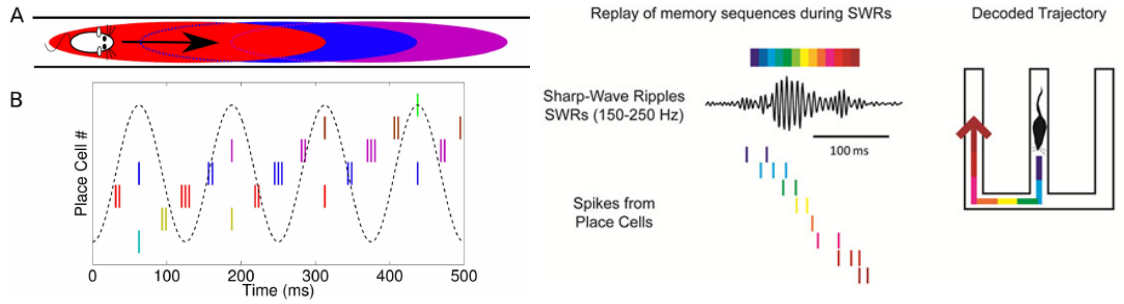


Figure 9.5: Left: Sequence compression of neural population activity by phase precessing neurons. Right: Replay of sequences while the rate is resting.Left figure from Chadwick, Rossum, and Nolan 2015. Right after Jadhav et al. 2012

Suppose we have a response-vector r that can take n values, value r has a probability $p(r)$. Entropy is defined as (Shannon and Weaver 1949)

$$H = -\sum_{r=0}^{n} p(r)\, \log_2 p(r)$$

When dealing with a spiketrain, the response will be a discretized spiketrain, such as $r = (0, 1, 0, 0, 1)$. When we split the total time T in bins of length Δt there are $n = 2^{T/\Delta t}$ possible responses. If the response pattern is always the same, one $p(r)$ is one while the others are zero. In that case the information is zero. Indeed such a response would not be very informative. Instead, when all probabilities are equal ($p(r) = 1/n$), information is maximal, $H = \log_2(n)$. In that case the response itself is very informative, we could not have predicted the response with much certainty. High entropy corresponds to having a rich response ensemble.

However, the richness in the response could just be noise. We want to know the information that the response provides about the stimulus, this is called the MUTUAL INFORMATION. We need to subtract the noise part. The noise part can be measured by repeating the stimulus over and over, and measuring the fluctuations in the neural response (see our discussion on spike statistics for problems with stimulus repetition). We write the stimulus as s, and the response as r. The noise part is for given s and its average are, respectively,[5]

$$\begin{aligned} H_{noise}(s) &= -\sum_{r} P(r|s) \log_2 P(r|s) \\ H_{noise} &= \sum_{s} H_{noise}(s) P(s) = -\sum_{r,s} P(s) P(r|s) \log_2 P(r|s) \end{aligned}$$

The mutual information is defined as the difference between total and noise entropy

$$\begin{aligned} I_m &= H - H_{noise} \\ &= -\sum_{r} P(r) \log_2 P(r) + \sum_{r,s} P(s) P(r|s) \log_2 P(r|s) \\ &= \sum_{r,s} P(s) P(r|s) \log_2 \frac{P(r|s)}{P(r)} \\ &= \sum_{r,s} P(r,s) \log_2 \frac{P(r,s)}{P(r)P(s)} \end{aligned}$$

where we used that $P(r) = \sum_s P(s)P(r|s)$. Note the symmetry between r and s.

[5]With $P(r|s)$ we denote the probability for response r given a stimulus s.

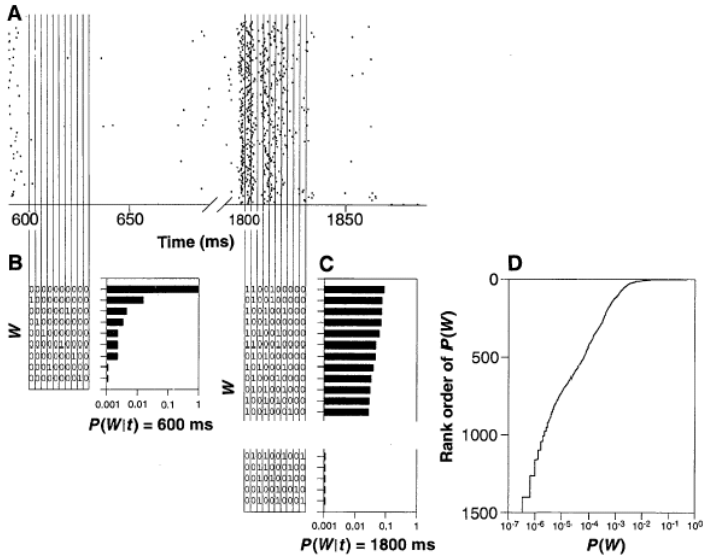


Fig. 3. Word frequency distributions and information transfer. (**A**) Two segments from 100 response traces of H1, starting at about 600 and 1800 ms, respectively, after onset of the repeated stimulus of Fig. 2. (**B**) Construction of local word frequencies. We start with a set of spike trains in response to a repeated random velocity sequence. Beginning at 600 ms these spike trains are divided in 10 contiguous 3-ms bins, as indicated by the array of vertical lines. For each trial, the spikes in each of the 10 bins are counted, and this set of 10 numbers forms a word, *W*. Here almost all words are binary strings, as two spikes occur only very rarely within 3 ms. This procedure gives us as many words as there are trials (here 900). From this set we compute the probability for each word, and the resulting distribution is depicted in the histogram, $P(W|t)$ = 600 ms, where the words are ordered according to their probability. (**C**) As in (B), but now starting at 1800 ms. (**D**) Distribution, $P(W)$, of all words throughout the experiment. Words are defined in the same way as in (B) and (C). However, here they are taken from the long (900 times 10 s) nonrepeated part of the stimulus sequence in order to obtain a large number of independent stimulus samples. Thus, stepping in 3-ms bins, $\sim 3 \times 10^6$ words are sampled, and the distribution shown here describes their ranked frequencies. In these windows, by far the most likely word is 0000000000, and roughly 1500 different words are observed.

Figure 9.6: Step in the calculation the information in a spiketrain: counting 'words'. The probability for words enters in the information calculation.From Steveninck et al. 1997.

In practise the information calculation goes as follows: pick a neuron and give it a stimulus from a chosen ensemble. Measure the response and repeat many times. The output spike-train is written a string of '1's and '0's. Response r is a possible 'word' in the response, such as $r = (1, 0, 0, 1, 0, 1)$. Next, one can calculate the information, Fig. 9.6. It is important to choose the right stimulus ensemble (a visual neuron will not give information about an auditory stimulus) and time-discretization. It is easy to understand that given the many possible words, an accurate information calculation requires a lot of data. In principle the method can be extended to measurements of many neuron at once, but this becomes numerically hard, requiring lots and lots of data.

Typical values for the mutual information are that a spike carries 1-2 bits of information. Note, that is not because a spike is binary. When temporal precision is high, information content can be much higher (see our example in the beginning of this section). When precision is low or the cell is not tuned to the stimulus, information content can be virtually zero.

More reading: NIP course, Rieke et al. (1996) for a good treatment of coding, reconstruction and information theory.

Chapter 10

Networks of neurons

After concentrating on single neurons, an obvious question is how a network consisting of many neurons will behave. This is where the fun starts. The final dynamics will of course to large extent depend on the connections between the neurons. Here we study very simple connectivity patterns, which in some cases give already very rich and interesting behaviour. We will model the neurons with I&F neurons or similar simplified models.

10.1 Rate approximation

We start with a description in which the individual spikes of the neurons are not explicitly taken into account, instead we describe neurons by their average firing rate. This abstraction is often used in connectionist and artificial neural networks.

$$\tau_{syn}\frac{dr(t)}{dt} = -r(t) + g(in) \tag{10.1}$$

A simple choice for the non-linearity $g()$ is

$$g(in) = [in]_+ \tag{10.2}$$

where $[x]_+ = x$ if $x > 0$, and $[x]_+ = 0$ otherwise, or, equivalently and useful in Matlab, $[x]_+ = \max(0, x)$. Eq.(10.2) is called the RATE APPROXIMATION(see Dayan and Abbott (2002) for a derivation and variants).

As an example, suppose we have a two-layer network. A vector denotes the activity of each neuron in the layer, the input layer activity we write as $\mathbf{s}$ and the output layer as $\mathbf{r}$. The connection weights between the two layers can be written as a matrix W, Fig. 10.1A. When enough noise is present such that the neurons will fire independently at different times, we can approximate the population response with

$$\tau_{syn}\frac{d\mathbf{r}(t)}{dt} = -\mathbf{r} + [W\mathbf{s}]_+ \tag{10.3}$$

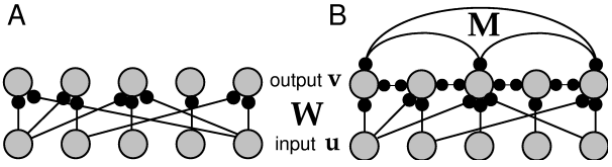


Figure 10.1: A) A simple two-layer feed-forward network, input layer $\mathbf{u}$ is connected with a weight matrix W to output layer $\mathbf{v}$. B) A two-layer network with lateral connections that feedback the output activity to other neurons in the layer.

Here the input is $\mathbf{in} = W\mathbf{s}$. More generally the input can contain recurrent connections and direct input.

- The firing rate sums the excitatory inputs, subtracts the inhibitory input and rectifies the result. This is not exact, but seems reasonable: above we have seen that inhibition can indeed act in a subtractive manner, and that the F/I curve can be more or less linear.
- The non-linearity can form the basis for doing computations. Without the rectification, even many stacks of layers could only do linear computations. In that case the transformation between two layers can be written as a matrix. The computation of three layers in series is the product of the two matrices, which is again a matrix. (This is the famous argument of Minsky and Papert against perceptrons). But with the non-linearity we can easily create a network that solve the XOR problem. (Try it!). Unfortunately, analysis is not so straightforward once non-linearities are included.
- Unlike more abstract neural nets, the dynamics of Eq. (10.2) are not saturating, but this can be built in easily, for instance by replacing $[.]_+$ with $\min(1, [.]_+)$ in Eq.(10.2). One does expect some mechanisms to be present in biology to prevent too high levels of activation.
- The dynamics of a isolated neuron are given by the synaptic time-constant τ_{syn}. (Assuming equal excitatory and inhibitory time-constants and no extra delays). This is fast. The synaptic time-constant will be the time-constant of the computation. This means our brain has some 300 clock cycles per second ($\tau_{AMPA} \approx 3ms$). When one simulates integrate-and-fire neurons, one finds that the dynamics indeed scale with the synaptic time-constant, but there are corrections which slow down the dynamics. These are usually small effects (Gerstner 2000; Brunel 2000; Dayan and Abbott 2002). The answer critically depends on the distribution of membrane potentials when the stimulus arrives. When the membrane potentials are near rest before the stimulus arrives, it takes longer for the neuron to spike, than when they are close to threshold.
- The dynamics are assumed very simple: there is no adaptation, or synaptic depression. Synapses have all the same time-constant and are linear.

The rate approximation is often used in artificial neural networks, perceptrons and cognitive models. In some aspects the rate approximation seems a decent description of cortical activity. Especially when the firing rate instead of a single neuron describes a group of neurons firing independently (without synchrony). This could describe a column. Of course, using a rate approximation, also implicitly implies rate coding (see Chapter on coding).

10.2 Spiking network

What happens when one uses spiking neurons? Suppose we have a population of unconnected I&F neurons. Now we present some external stimulus to all the neurons in the population. How does the firing change? This is not a trivial question, especially not when noise is present (Gerstner 2000). The response will depend on the different states in which the neurons in the population are, that is, the distribution of the membrane potentials. If the neurons have the same potential, they will all fire simultaneously. But right after that all neurons are in the refractory period and the population will have much less activity. This can cause damped oscillations in the firing rate, Fig. 10.2. Note that in this figure due to the noise the neurons have slightly different initial conditions, else they would all spike at *precisely* the same time. We can also see that the neurons can follow transients without any sluggishness, despite the presence of a slow membrane time-constant (here taken 20 ms). Compare to Fig. 6.4.

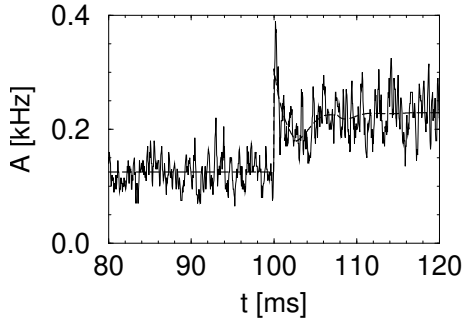


Figure 10.2: Response of a simulated population of spiking neurons to a step current (note, not synaptic input). Spiking neurons can follow transients rapidly.From Gerstner 2000

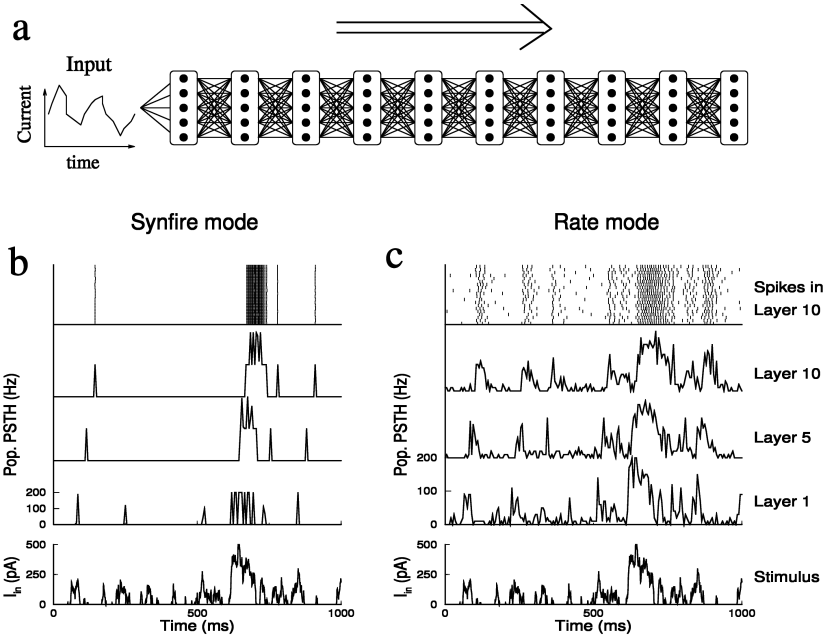


Figure 10.3: Propagation of a current stimulus through a layered feed-forward network. a) Architecture of the network: 10 layers with 20 integrate-and-fire neurons per layer (5 neurons shown), all-to-all connectivity. b) Syn-fire behaviour: without noise the neurons strongly synchronise. c) But if the neurons are noisy, firing rates can propagate rapidly through layered networks without much distortion. FromRossum, Turrigiano, and Nelson 2002.

10.3 Many layers of spiking neurons: syn-fire or not

Next, we extent this spiking network to include multiple layers. If, using a rate model, we would chain many layers after each other with uniform connections ($W_{ij} = w_0$) and ignore the non-linearity, the rate in each layer would be the low-passed. However, when multiple layers of spiking feed-forward networks are modelled, firing can start to synchronise. This behaviour is called the syn-fire chain (Abeles 1991), and is not captured with the rate description. In Fig. 10.3b the response of a syn-fire chain to a random stimulus is shown. The parallel architecture averages out temporal fluctuations in the spike timing. It allows for stable, and temporally very precise population spikes, as small amounts of noise will not prevent synchronisation. Although syn-fire chains are easily generated in computer simulations, experimental evidence is weak.

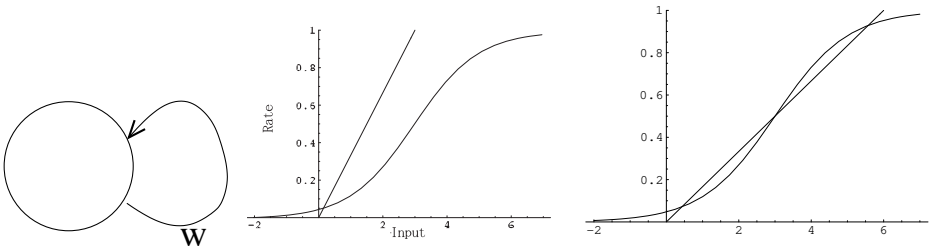


Figure 10.4: Left: A single neuron with a recurrent connection with strength w. Middle: rate as a function of the input, and the line $r = in/w$. The only solution is where the curves intersect. Right: For stronger w there are multiple solutions, the middle one is unstable, the two outer ones are stable.

When instead a 'decent amount' of noise is injected to all neurons, propagation can be described by the rate equation, Fig 10.3c. Note that the stimulus propagates rapidly through the network. The speed of the propagation helps to explain why one can do such fast processing of, for instance, visual stimuli. The tendency to synchronise is counteracted by the noise and by the synapses which act as a temporal low-pass filter between the layers, removing sharp transients.

It might be possible that in vivo a subset of neurons propagates information using a syn-fire mode, whereas the population as a whole appears to be asynchronous. This could have interesting computational effects. The existence of such situations would be very hard to verify as one has to measure the possibly very small subset of neurons.

10.4 Recurrent networks

Next we include recurrent connections in the network, initially using a rate model. Suppose we have just a single neuron. It does not receive any external input, but it does receive recurrent input with strength w, Fig. 10.4 left. The steady rate of neuron for a given input is here modelled as $0 = \tau dr/dt = -r + g(in)$, so $r(in) = 1/(1 + \exp(-in + 5/2)$. The recurrent input is given by $in(r) = wr$. In the steady state $dr/dt = 0$, so that $0 = -r + g(in)$. In other words, $r = g(wr)$, so that we need to find values for r that satisfy $r = 1/(1 + \exp(-wr + 5/2)$. This is impossible analytically. We can solve this equation graphically by plotting both $r(in) = 1/(1+\exp(-in+5/2)$ and $r(in) = in/w$ in one graph, Fig. 10.4.

Depending on the value w we can have one or two stable points. When w is small (or absent altogether) the rate is just small as well. When w is large enough, we have two stable points, one with low activity and one with high activity, in which the feedback keeps the neuron going. We have basically a flip-flop memory! Once the neuron is in one state, it will stay there, until we disturb it with external input.

When w is even larger, the only fixed point is near $r = 1$. You can see this by imagining an even shallower line in Fig. 10.4. Although this an unbiological example the same principle on a network level is thought to underlie working memory.

10.4.1 Two recurrent rate neurons

We now consider two connected recurrently connected neurons and are again interested in the dynamics and its fixed points. The definition for a fixed point is that at the fixed point the temporal derivatives are zero. Fixed points can be stable and unstable. Consider two recurrently connected neurons: $r_1 \rightleftarrows r_2$

$$\tau \frac{dr_1}{dt} = -r_1 + [w_{12} r_2 + in_1]_+$$

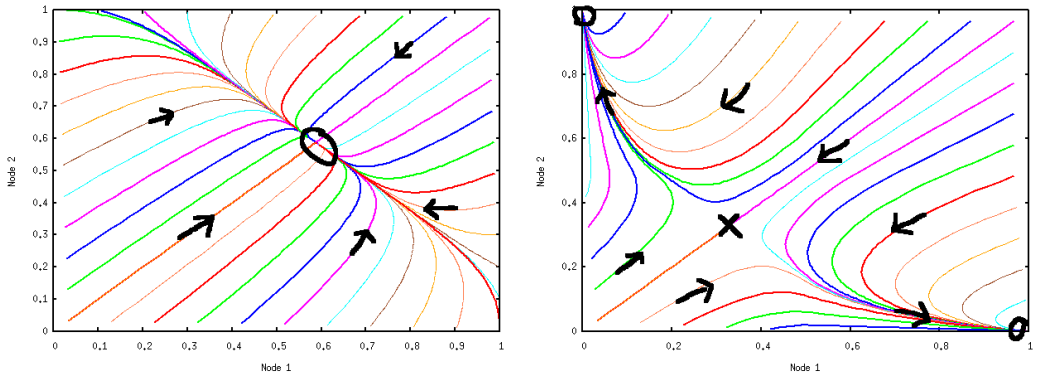


Figure 10.5: Stability of two connected nodes. Left ($w_{12} = w_{21} = -0.8$) Stable fixed point at $(0.59, 0.59)$. Right ($w_{12} = w_{21} = -2$). Unstable (saddle point) fixed point at $(1/3, 1/3)$. Note, changing the inputs changes the fixed points, but not the basic picture.

$$\tau \frac{dr_2}{dt} = -r_2 + [w_{21} r_1 + in_2]_+ \tag{10.4}$$

where $[x]_+$ is a rectification with $[x]_+ = x$ if $x > 0$ and $[x]_+ = 0$ otherwise. This describes two neurons that provide input to each other. If $w_{12} > 0$ ($w_{12} < 0$) then neuron 2 has an excitatory (inhibitory) influence on neuron 1. Apart from that they receive external input (in_1, in_2) which is assumed to be constant.

Let's assume (to be confirmed post hoc) that $w_{21}u_2 + in_1$ and $w_{12}u_1 + in_2$ are much larger than 0, so that we can ignore the rectification. In that case we have,

$$\begin{aligned} \tau \frac{d\mathbf{u}}{dt} &= \begin{pmatrix} -1 & w_{12} \\ w_{21} & -1 \end{pmatrix} \mathbf{u}(t) + \mathbf{in} \\ &= Q.\mathbf{u}(t) + \mathbf{in} \end{aligned}$$

The matrix $Q = W - I$ incorporates both the decay terms (on diagonal and negative) and interactions (off-diagonal). Let's find the fixed points, that is the $\mathbf{u}$ for which $\tau \frac{d\mathbf{u}}{dt} = 0$, i.e. $\mathbf{u}_\infty = -Q^{-1}.\mathbf{in}$. Let's assume $\mathbf{in} = (1, 1)$. In Fig. 10.5 these fixed points are right in the middle of the plot (indicated with a cross and a circle).

Next, we perform a stability analysis to see if these fixed point is stable. To this end we look at what happens if we perturb the system away from the fixed point, i.e. $\mathbf{u}(t) = \mathbf{u}_\infty + \mathbf{v}(t)$. Now $\tau \frac{d\mathbf{u}(t)}{dt} = Q.[\mathbf{u}_\infty + \mathbf{v}(t)] + \mathbf{in}$, so $\tau \frac{d\mathbf{v}(t)}{dt} = Q.\mathbf{v}(t)$ where $\mathbf{v}(t)$ is a small vector. The only thing we need to know if such a perturbation grows or shrinks over time. An easy way is to perturb in the direction of the eigenvectors W. An eigenvector $\mathbf{s}_i$ of Q will behave as $\tau \frac{d\mathbf{s_i}}{dt} = \lambda_i \mathbf{s}_i$, and the perturbation will therefore develop as $\mathbf{s_i}(t) = \mathbf{c} \exp(\lambda_i t/\tau)$. The sign of λ_i will determine whether the system runs away from the fixed point or returns to it. We can distinguish a few possibilities.

- $\lambda_{1,2} < 0$. The dynamics are stable, the system converges to fixed point. This is illustrated in Fig. 10.5 left. The figure illustrates the system's evolution. We simulated Eq. 10.4, and followed the system over time. Different sets of the initial conditions were taken, all along the edges of the graph.

- $\lambda_1 > 0$, $\lambda_2 < 0$. Saddle point. Although the dynamics are stable in one direction, in the other direction it is unstable. Therefore the fixed point as a whole is unstable. This is illustrated in Fig. 10.5 right. Along the diagonal the dynamics moves towards the fixed point $(\frac{1}{3}, \frac{1}{3})$, but then bends off towards to $(0, in_2)$ or $(in_1, 0)$. The intuition is that in this case, because the inhibition is stronger, the nodes strongly inhibit each other and there can only be one winner. Such WINNER-TAKE-ALL NETWORKS have been used to model decision making (see below).

- $\lambda_{1,2} > 0$. Dynamics are unstable. This means that a minuscule fluctuation will drive the solution further and further from the equilibrium. Like in the previous case, the solution will either grow to infinity or till the linear approximation breaks down.
- If the eigenvalues are complex the system will oscillate. Remember: $e^{x+iy} = e^x[\cos(y) + i\sin(y)]$. Stability determined by the real part of the eigenvalue $Re(\lambda)$. When the real part is < 0 the oscillations die out, otherwise they get stronger over time.

The above technique can also be applied to the case when the equations are non-linear. In that case the fixed points usually have to be determined numerically, but around the fixed point one can make a Taylor expansion, so that for small perturbations $\tau \frac{d\mathbf{u}}{dt} \approx Q.\delta\mathbf{u}$ and one can study the eigenvalues of Q again.[1]

In Fig. 10.5 left, the system will always go to the same fixed point. TheBASIN of attraction in this case encompasses all possible initial conditions. In Fig. 10.5 right we have two basins of attraction, starting above the line $in_1 = in_2$ the system will go to the upper left fixed point, starting below the line the system will go to the lower right fixed point.

10.5 Many recurrent neurons: Chaotic dynamics and Hopfield net

If we recurrently connect a decent number of nodes to each other with random weights, we can get chaotic dynamics. If a system is chaotic it will show usually wildly fluctuating dynamics. The system is still deterministic, but it is nevertheless very hard to predict its course. The reason is that small perturbations in the initial conditions can lead to very different outcomes and trajectories (The butterfly in China, that causes rain in Scotland). This contrasts with the non-chaotic case, where small perturbations in the cause small deviations in the final outcome. On the website there is a script which allows you to play with a chaotic system. There has been much speculation, but not much evidence for possible roles of chaotic dynamics in the brain.

10.6 Spiking recurrent networks

Above we dealt with a feed-forward network, the other extreme of modelling is a recurrent network receiving constant input (the real brain should be somewhere between these two approaches...) What happens to such a recurrent network, does any interesting activity emerge? When neurons are connected in an all-to-all fashion and little or no noise is present, the network easily settles into a state of global high frequency oscillation or complete quiescence. (Easy to simulate). This might relevant in the study of epilepsy or dead, respectively, but it does not reflect the normal state of the brain.

More realistically, the probability for connections between two arbitrarily chosen neurons is small. This is called a sparsely connected network. Secondly, inhibition should be incorporated. Interestingly, these networks have a whole range of possible stable states, including a randomly firing low-activity state (Vreeswijk and Sompolinsky 1996).

The possible states can be analysed by a Fokker-Planck equation for the membrane potential, much like we have seen for the noisy integrate and fire neuron. But here the firing rate of the network is fed back into the network: a self-consistent approach (Brunel 2000). This way one can find the conditions for an asynchronous state. The oscillation regimes can be found by perturbing the asynchronous state. These studies can then be confirmed in simulations.

In Fig. 10.6 the various possible states of a sparse, recurrent network are shown. When recurrent inhibition is weak, oscillatory high activity states develop. When inhibition is strong enough however (on the right side in the figure), chaotic firing occurs. The average activity, which

[1]The other two fixed points in the corners of Fig. 10.5 right can be found using this technique. The Taylor expansion of the transfer function we used is unfortunately ill-defined near $x = 0$. Replacing it with $g(x) = \log(1 + \exp(10x))/10$, which is very similar, will work better.

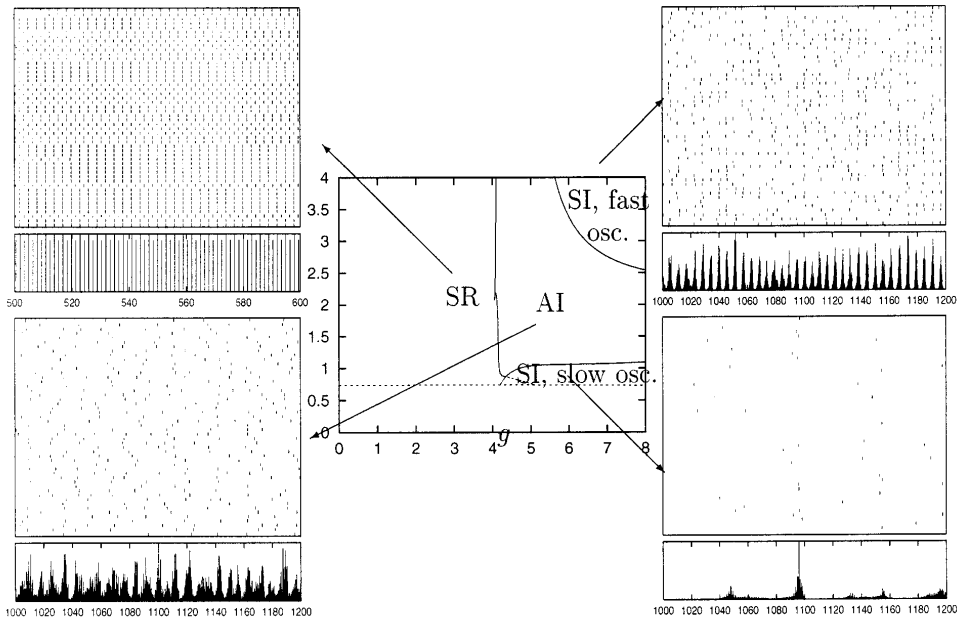


Figure 11. Simulations of a network of 10 000 pyramidal cells and 2 500 interneurones illustrate the different types of collective states, or 'phases' of the system. For each of the four examples are indicated the temporal evolution of the global activity of the system (instantaneous firing frequency computed in bins of 0.1 ms), together with the firing times (rasters) of fifty randomly chosen neurones. In the SR state, the network is almost fully synchronized and neurones fire regularly at high rates. In the fast oscillatory SI state, there is a fast oscillation of the global activity, and neurones fire irregularly at a rate which is lower than the global frequency. In the AI state, the global activity is stationary (fluctuations seen in the graph are a finite size effect, see Section 3.3.4), neurones fire irregularly. In the slow oscillatory SI state, there is a slow oscillation of the global activity, and neurones firing irregularly at very low rates.

Figure 10.6: Possible states of activity in a recurrent network of I&F neurons. The x-axis gives the ratio between recurrent inhibition and recurrent excitation, the y-axis represent the external input to the network.
From Brunel 2000.

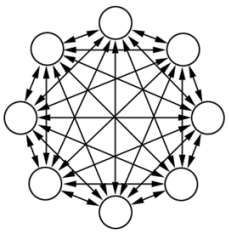


Figure 10.7: Hopfield net with 8 nodes. The arrows denote the (symmetric) weights between them.

corresponds approximately to the EEGs or field potentials measured experimentally, can still be oscillatory, but individual neurons don't fire every cycle (top right). This is much like what is seen in hippocampal rhythms. The state AI could correspond to cortical activity, with neurons showing low asynchronous background activity.

10.7 Hopfield net

Importantly, when the connections between neurons are symmetric, $w_{ij} = w_{ji}$, the dynamics is simple. The system always goes to one of its equilibrium states and stays there.

An even more simplified network is the Hopfield network. The neural activity can only take

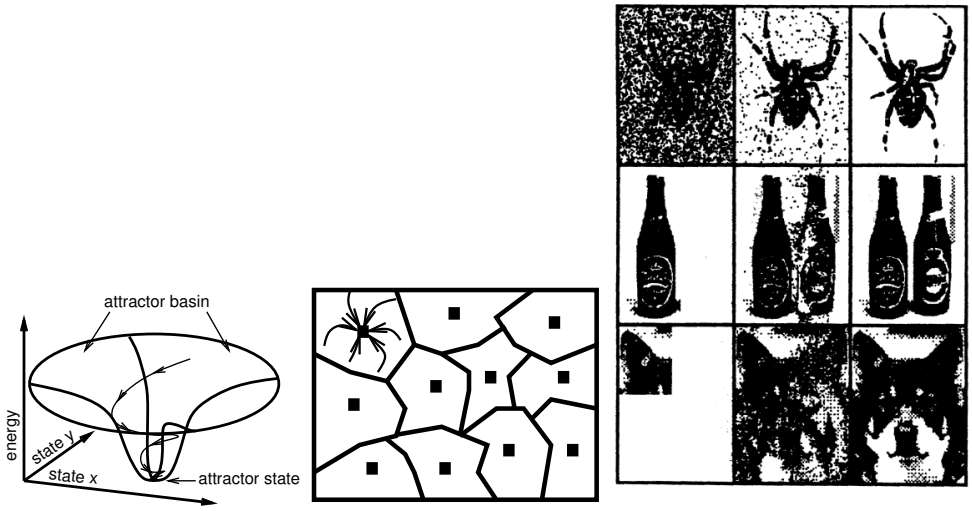


Figure 10.8:
Left: Whenever the network starts in a state close enough to an attractor, it will 'fall in the hole' and reach the attractor state.
Middle: Multiple attractor are present in the network, each with their own basin of attraction. Each corresponds to a different memory.
Right: Pattern completion in a Hopfield network. The network is trained on the rightmost images. Each time the leftmost (distorted) input is given, the network evolves via the intermediate state to the stored state. These different images can all be stored in the same network. (From the Hertz book).

values ± 1, and obeys the update rule

$$r_i(t+1) = \text{sign}[\sum w_{ij} r_j(t)]$$

The updating should take place with one (randomly chosen) unit at a time, that guarantees that the system will always reach an attractor. The Hopfield model is an influential memory model. It can store multiple binary patterns with a simple learning rule. The weight between node i and j should be set according to the rule $w_{ij} = \sum_\mu p_i^\mu p_j^\mu$, where p_i^μ is a binary bit (± 1) representing entry i for pattern μ.

The final state of the network is called an attractor state. Attractor states can be seen as memory states. Each stored pattern will correspond to a stable fixed point, or attractor state, see Fig. 10.8. However, the storage capacity of the network is limited. As we increase number of stored patterns, the memories states becomes unstable. Stable mixed states appear. At the critical amount of storage, performance decreases suddenly. The network will still equilibrate, but it will end up in spurious attractor states rather than in the memory states it was suppose to find. The numbers of patterns we can store is proportional to the number of nodes, $n_{stored} = \alpha n$ where α is called the capacity. Simulation and statistical physics gives $\alpha_c n = 0.138n$. Hence a network with 100 nodes can store about 14 patterns. Many more details can be found in Hertz, Krogh, and Palmer (1991).

The Hopfield network is called an AUTO-ASSOCIATIVE MEMORY: Presenting a partial stimulus leads to a recall of the full memory, see Fig. 10.8. Auto-associative memories are very different from computer memory (a bit like Google).

Under certain conditions the properties of the Hopfield network carry through for models with the rate equations introduced above, as well as spiking models. In parallel to the Hopfield net, attractors can be achieved by creating subpopulations in the network, which have stronger connections among themselves (w_+ in Fig. 10.9) than connections between subpopulations (w_-).

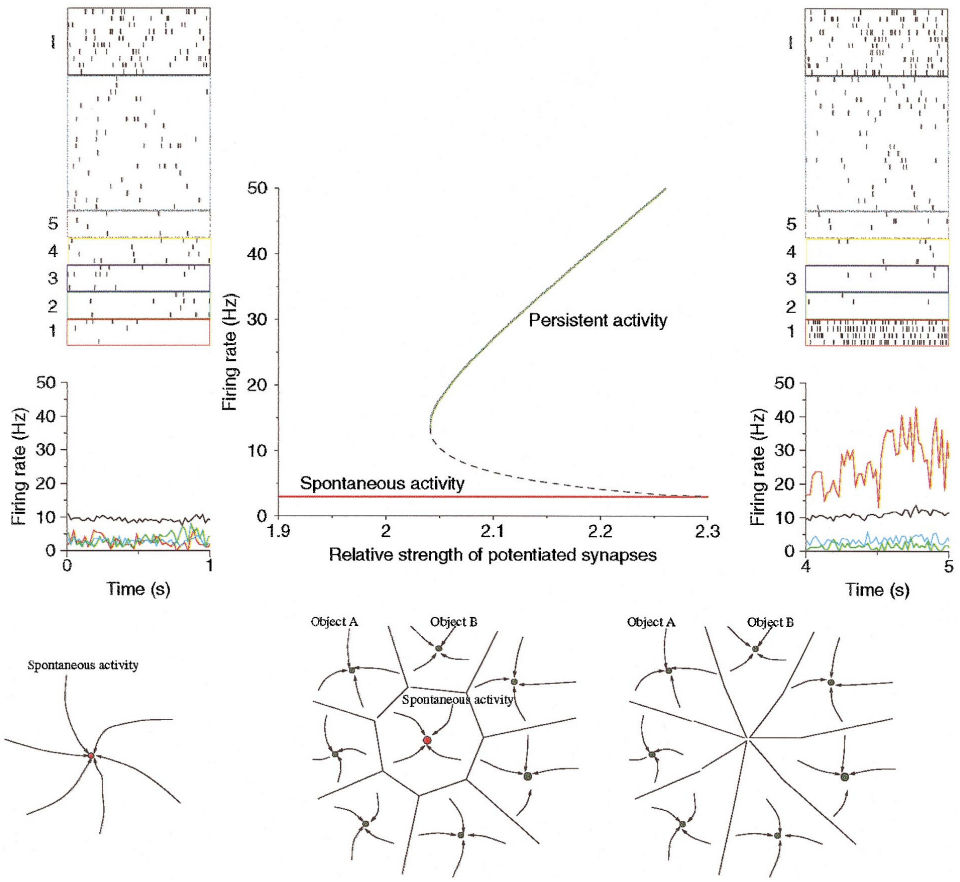


Figure 12. Top: at the centre, a graph shows how activity in a subpopulation coding for an object depends on the strength of potentiated synaptic connections w_+. It shows two stable branches, one corresponding to spontaneous activity, the other to persistent memory activity. An unstable branch, shown with a dashed line, indicates the boundary of the basins of attraction of spontaneous and persistent activity. To the left of the graph, a simulation of the spontaneous activity state in a network with 1 000 cells and $w_+ = 2.1$. To the right of the graph, a simulation of the persistent activity state in the same network. The network can switch from one state to another using transient selective external inputs to the subpopulations coding for a particular object. Bottom: schematic representation of the 'phase space' of the network. Left, below $w_+ \sim 2.05$, the network has a single fixed point attractor, the spontaneous activity state. Above this critical value, many additional attractors form, corresponding to the objects 'stored' in the synaptic matrix (centre). Last, a second bifurcation occurs at $w_+ \sim 2.3$ at which the spontaneous activity state becomes unstable (right).

Figure 10.9: Multiple stable states in a recurrent network.From Brunel 2000.

To represent 'unknown' or non-memorized input, one can create a point attractor with a large basin of attraction corresponding to spontaneous activity, Fig. 10.9. Once activity reaches a point attractor state, it will be very stable, assuming adaptation is absent. When the neurons or synapses adapt, the network will not stay in the state but will jump out of it after while, so called winner-less competition. The system will then visit attractor states in succession.

Networks can also be engineered to have particular asymmetric connections, so that particular sequences of attractors can be stored (sequence memory).

10.8 Single recurrent layer

We return to slightly simpler architectures. The next complication one can consider is to allow lateral or recurrent connections, Fig. 10.1. The equation for v becomes

$$\tau \frac{d\mathbf{v}(t)}{dt} = -\mathbf{v}(t) + [W.\mathbf{u}(t) + M.\mathbf{v}(t)]_+ \tag{10.5}$$

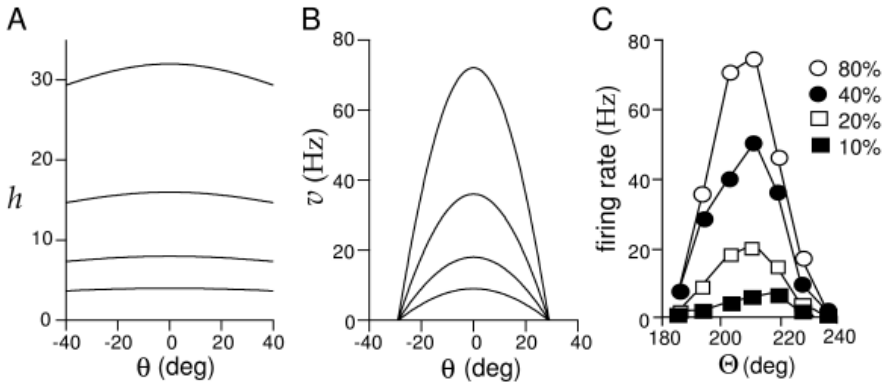


Figure 10.10: Recurrent model of simple cells in V1. Left: input to the network, which can be very weakly tuned. Middle: The network output is strongly tuned. Right: experimental data from V1.Figure from Dayan and Abbott 2002.

where M is the matrix describing the recurrent connections. An interesting application of this model has been to explain contrast invariant tuning curves (Ben-Yishai, Bar-Or, and Sompolinsky 1995). Although as mentioned above, there is now doubt that this is the correct model explaining contrast invariant tuning curves, it is still worth a study.

The model describes a population of simple cells with their preferred angle $\theta = -\pi/2...\pi/2$. Rather than specifying u and W, the network receives an input $\mathbf{h}$ (i.e. $\mathbf{h} = W.\mathbf{u}$), which is parameterised as

$$h(\theta) = Ac[1 + \epsilon(-1 + \cos(2\theta))]$$

where A describes the amplitude, c the stimulus contrast, and ϵ the amount of tuning of the input. We label the neurons v with the angular variable θ. The lateral interaction is described with two parameters λ_0, which describes uniform inhibition and λ_1 which gives the strength of tuned excitation. So the full equation is (the sum implicit in Eq. 10.5 can be written as an integral when there are many neurons):

$$\tau \frac{dv(\theta)}{dt} = -v(\theta) + \left[h(\theta) + \int_{-\pi/2}^{\pi/2} \frac{d\theta'}{\pi} \left\{ -\lambda_0 + \lambda_1 cos(2\theta - 2\theta') \right\} v(\theta') \right]_+$$

If the network were linear, the output of the network would smoothly depend on the input parameters. However, in this model, because of the rectification, excitation is not always counteracted by the uniform inhibition: a silent neuron can not inhibit... The effect is that when the input is weakly tuned, the output of the network can be much sharper, amplifying the differences in the input, Fig. 10.10.

Importantly, the width of the tuning curves becomes almost independent of the input 'sharpness'. This way the recurrent model explains the contrast invariance of tuning curves,[2] although nowadays this is a less popular model to explain contrast invariance.

10.8.1 Working memory

When the connections are made stronger, the network can maintain activity in the absence of input (cf. the networks described above). This is not useful for the early visual system, but it does allow it to store a memory in the location of the activity bump. These networks are called ring attractors (or sometimes line or continuous attractors). Unlike point attractors, the network state can without much effort move between the attractor state (like a ball in a Mexican hat). This

[2]Note, that the firing threshold in this model is zero, unlike in the discussion of the iceberg effect above, section 7.2.2. This does not change the behaviour of the model.

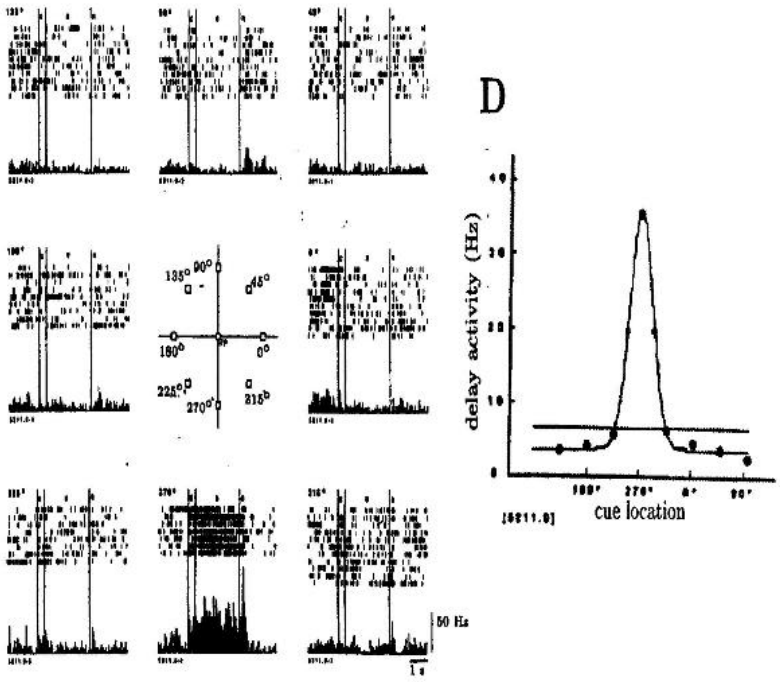


Figure 10.11: Working memory in pre-frontal cortex. A stimulus was presented (onset and offset marked by the two leftmost vertical bars). The monkey had to remember this location for two seconds. After this the monkey had to make an eye-movement (final bar). The neurons have (in this case) a spatial tuning function. FromFunahashi, Bruce, and Goldman-Rakic 1998.

means that the memory state will only be stable for few minutes. The memory will start to drift away, both in model and data.

There are cells in the pre-frontal cortex which fire selectively during such working memory tasks. An example of such a cell is shown in Fig. 10.11. Note that this cell is most active in the absence of the stimulus, this is presumably due to recurrent input. When the monkey has to remember the spatial location of the banana, the cells show a tuning related to the location. But given the flexible role of the pre-frontal cortex, the encoding is probably much more general than that. Look back at Chapter 1, Fig. 1.3 for examples of pre-frontal damage.

The architecture of the model is the ring model above with local recurrent excitation and long range inhibition. Networks with persistent activity need strong recurrent excitation. As one starts to model such networks (for instance with I&F neurons) one finds that is not easy to have a memory state which has relatively low activity, and no synchronisation. One can achieve such by applying the observations above: having large, sparsely connected networks with excitation and inhibition and including enough external noise. As the time-constant for the dynamics is typically given by the synaptic time-constant, using a lot of NMDA will help to stabilise the network. See (Compte et al. 2000; Wimmer et al. 2014) for a spiking model of a working memory network. Again, adaptation or synaptic fatigue will destabilize the systems and can lead to rotation of the state, e.g. (York and van Rossum 2009).

Another example of these ring attractors are the head direction cells in animals (and probably also humans) that keep track of the direction of the head with respect to the environment. It acts like an internal compass, except that it does not rely on a magnetic field but instead relies on updates from the visual and vestibular system. Continuous attractor networks are thought to be important tasks such as maintaining eye position and remembering body position as well. The best evidence for the above ring attractor dynamics has been found in the fruit-fly (Seelig and

Jayaraman 2015; Kim et al. 2017). Interestingly, also *anatomically* the cells appear to be arranged in a ring (something that the theory does not require).

10.8.2 General networks

It is important to realize that we have considered only a few networks with highly engineered connectivity (even if the connections were random). The real richness computational behaviour probably comes from the specific network connections. In fact, McCullough and Pitts showed that already with very simple threshold units, networks can be engineered that can carry out *any* computation (McCullogh and Pitts 1943). (A result that, as von Neumann realized, was powerful, but also ultimately sterile).

Chapter 11

Making decisions and Motor output

In many instances a task involves making a decision: an eye-movement, a bar press, or a report to the researcher. And also in real life decisions need to be made. Also our monkey has to decide where to move the joystick. It is therefore important to know how the decision is made and how errors in the decision relate to the underlying quantities.

One of the simplest cases for a decision with noise is a yes/no decision between Gaussian distributed quantities with an equal standard deviation. The errors in the decision are called FALSE POSITIVES (saying yes when the answer was no) and false negatives (saying no when the answer is yes). The error rates will depend on the width of the distribution and their distance. By changing the threshold, the trade-off between false positives and false negatives can be changed. This way one creates the so called RECEIVER OPERATOR CHARACTERISTIC, or ROC curve.

The variability in the quantities can either come from the stimulus (for instance when the stimulus is very dim and photon noise matters) or from the nervous system itself (for instance when you try to decide to go left or right on a forgotten road). Many stimuli have been developed in which the error can be systematically varied. An interesting stimulus is the random moving dots stimulus, Fig. 11.2. In area MT of the monkey there are neurons which respond to the (local)

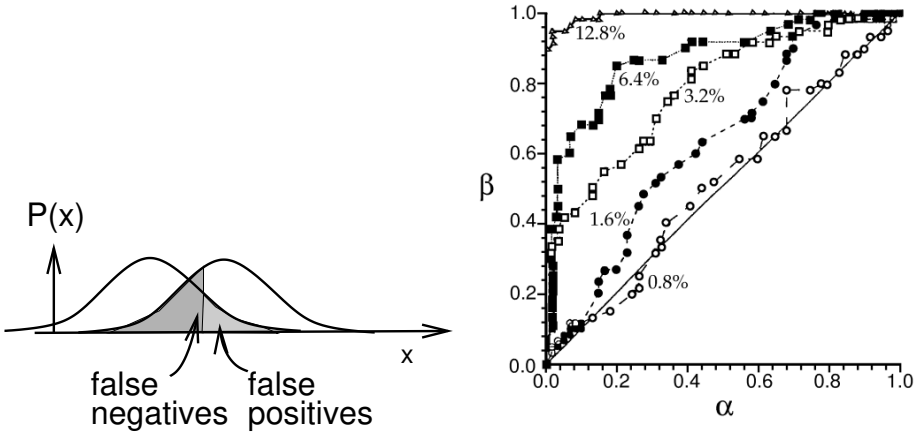


Figure 11.1: Left: principle of deciding between two quantities. The more the distributions overlap, the harder the task, the higher the error rate. Right: ROC curve of the random moving dots stimulus, described below (the numbers indicate the coherence of the stimulus). In the graph $1-\beta$ is the false negative rate, α is the false positive rate.

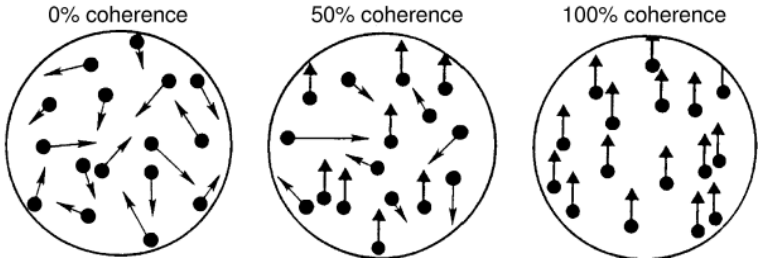


Figure 11.2: Moving random dot stimulus. Left: no net motion is present (but the observer might guess that there is motion). Middle: difficult but observable motion. Right: clear net motion.

motion in their receptive field. For instance, the stimulus in the right panel can be a very effective stimulus for such a neuron when the motion is in the direction of the neuron's preferred motion, whereas the stimulus in the left causes a much less vigorous response. By changing the coherence of the stimulus, the difficulty of the task can be changed and the neural responses varies accordingly. Recording the neuron while at the same time measuring actual performance, it was found that the performance of a single (!) neuron already matched the behaviour of the monkey (Newsome, Britten, and Movshon 1989; Britten et al. 1992). (Note, in these studies it is interesting to study both correct and false responses).

How does a certain stimulus and the corresponding firing of a neuron lead to a decision? The neurons in the higher brain regions seem to accumulate evidence until they reach a certain threshold, Fig. 11.3. Also in psychology there are models for the integration of evidence (Ratcliff and Rouder 1998). How do noise, error rate and reaction time relate? Integration of signal in the presence of noise is something we have encountered before: the noisy integrate and fire neuron. Indeed such models give reasonable fits to the distribution of response times. This section is one of the few cases where human psychology and neuroscience have become closely entangled.

Making decisions is also closely related to rewards. If the decision leads to punishment the monkey should better try something else next time. Recommended reading on the relation to reward (Schall, Stuphorn, and Brown 2002) and the rest of the special issue of Neuron, Oct. 2002.

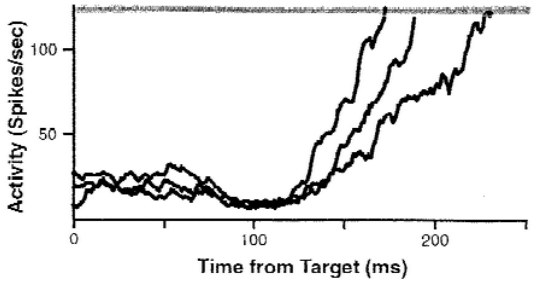
The activity of the decision neurons (if there are such neurons), reflect not only the integration of evidence but also the size and probability of the rewards (Platt and Glimcher 1999). This means that when multiple alternatives for action are available, the neurons can directly compete with each other and the one with highest (expected) reward will win, without the need for a deciding arbiter.

It seems reasonable to assume that decision making involves a close to optimal weighting of the evidence. How this is exactly done, is not clear. Many computational challenges remain: how to make good integrators from neurons, how to implement learning, how is the threshold implemented, are but a few.

11.1 Motor output

In the final stage of our task our monkey will have to produce motor output. The neurons in the motor cortex are a good example of population codes, e.g. Fig. 9.3. A check of the coding of the motor cortex is provided by experiments in which a computer cursor is directly controlled by neural activity (Wessberg et al. 2000). A multi-electrode array is implanted in the monkey's motor cortex. Suppose the cursor is controlled by a joystick. First the receptive field of the neurons are measured during a experimental phase where the arm moves a joystick freely. Next, link between the cursor and the joystick is broken (without the monkey being told), instead the spike rates of the neurons are used to move the cursor. With tens of neurons a bit unreliable, but usable, motor command can be constructed, Fig. 11.4. Interestingly, with training and visual feedback, the animal can improve performance, and thus steer the responses of the recorded neurons (Taylor, Helms Tillery, and Schwartz 2002). This means that learning can directly affect the coding in the motor cortex.

It is also remarkable that the simplest rate decoding mechanism gives good performance. So a rate coding scheme accounts for a major part of the information. Although one could argue that in motor cortex there is no place for more complicated coding schemes, as the output is sent fairly directly to the muscles. At least it seems we are out of the woods. These experiments have potentially important clinical applications: paralysed patients could potentially control artificial limbs with brain signals through chronically implanted electrodes.



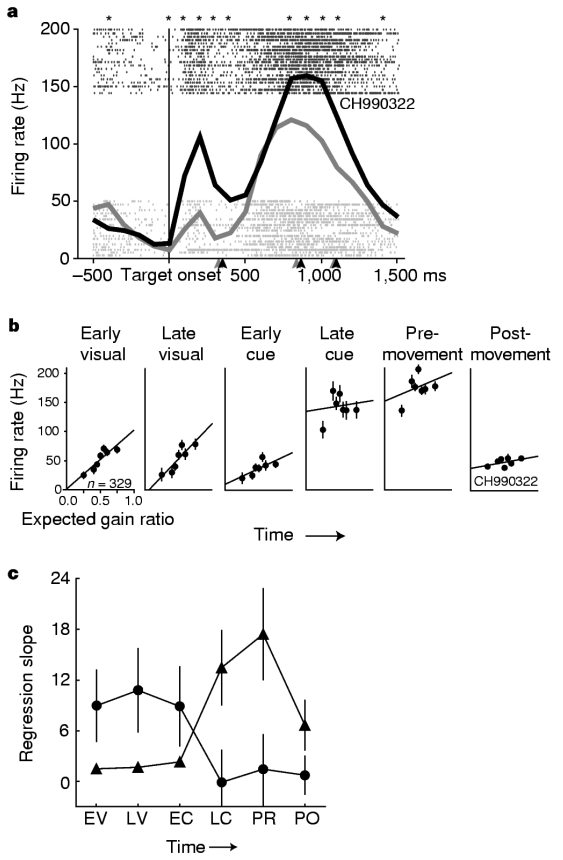


Figure 1 Modulation of neuronal activity by expected gain. The volume of juice delivered for each response was varied across blocks of cued saccade trials. **a**, Firing rate of an intraparietal neuron on trials instructing gaze shifts into the response field, averaged in 100 ms bins and synchronized, at vertical line, on target onset. Thick black line, expected gain ratio = 0.75 (n = 48); thick grey line, expected gain ratio = 0.25 (n = 37). Raster panels show spike times during the first 20 high-gain (dark grey) and first 20 low gain (light grey) trials. Panels show individual trials in sequential order from top to bottom. Arrows plot, successively, mean times of instruction cue onset, central fixation stimulus offset and saccade onset during high- (black arrow) and low- (grey arrow) gain blocks. Stars indicate 100-ms bins in which firing rate was significantly different in the high-gain block than in the low-gain block (t-test, $P < 0.05$). **b**, Mean firing rate (± s.e.) during each interval plotted against expected gain ratio, for all trials instructing gaze shifts into the neuronal response field. Lines indicate best-fit linear regressions through the raw data. **c**, Mean (± s.e.) regression slope for expected gain (circles) and the instructed movement (triangles), plotted against time (n = 40 neurons). Intervals: EV, early visual; LV, late visual; EC, early cue; LC, late cue; PR, pre-movement; PO, post-movement.

Figure 11.3: Left: Climbing of a neural response to threshold. This neuron (in the so-called frontal eye field) initiates an eye-movement. The time at which the neuron reaches threshold has a fixed latency to the eye-movement, and does not have a fixed latency to the stimulus onset. This shows that the reaching of threshold is strongly coupled to making the movement. From Schall, Stuphorn, and Brown 2002.
Right: From Platt and Glimcher 1999..

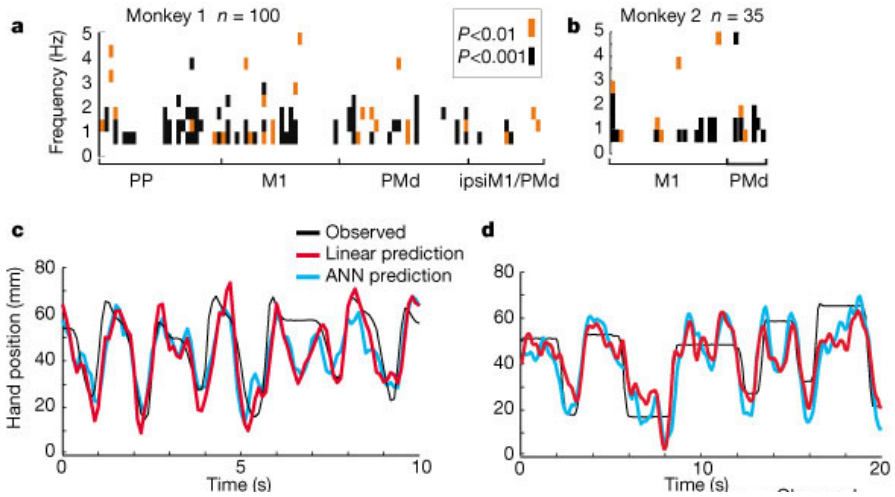


Figure 11.4: Predicting 1D hand movement from neural recordings in the motor cortex. Even a simple linear model predicts the movement fairly well. (The 'observed' curve in c,d is the one with the flat regions). From Wessberg et al. 2000.

Chapter 12

Synaptic plasticity

One of the characteristics of the nervous system is that it is not a fixed structure, but it is in a constant state of change. This plasticity takes places at many different levels and timescales:

1. Development: young brains develop, part by themselves, part helped by sensory signals. This development often goes one way, and damage can only partly be repaired.
2. Throughout life memories are stored, new procedures are invented and stored, new motor tasks are learnt. One assumes that such memories are stored in the synaptic weights and that the modification depend on the neural activity (and possible other factors such as reward).
3. Finally, on short time scales (milliseconds to minutes) neurons adapt, especially in the sensory systems. This is generally done by biophysical mechanisms and feedback loops.

All these three adaptation mechanisms are a hallmarks of neural computation. Here we focus on the second form of plasticity, long term synaptic plasticity. It is generally believed that there is not enough genetic information to specify all 10^{14} synaptic connections and their strengths, instead some limited set of rules might be at work. The most famous rule for learning is Hebbian learning. In 1949 Hebb, who was mainly thinking of reverberating loops, stated that:

"When an axon of cell A is near enough to excite cell B or repeatedly or consistently takes part in firing it, some growth or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased."

This remained an unproven theory for a long time, and changes in neural excitability and genes where also considered as possible memory storage sites. But evidence for Hebbian learning has now been found in many places in the brain. That is, synapses can undergo long lasting changes in strength consistent with this rule.

What remains less clear is whether Hebbian learning, as we shall shortly describe it, explains all forms of learning. Neuro-psychological case studies show that different memory systems exist. The different situations quickly mentioned above under 2) are all processed in different 'memory systems'. It is not clear if all use the same Hebbian mechanism. In human memory one distinguishes:

Implicit memory Includes procedural memory. Examples: motor learning, pattern completion. These tasks can be learned by amnesiacs. Their performance improves, but they will not remember learning it. Implicit memory has different modalities like explicit memory has (motor, visual, auditory, etc.).

Explicit memory Also known as declarative memory. This is the type of memory that requires concious processing. It can be divided in two sorts.
Semantic memory: Factual knowledge about the world.
Episodic memory: Personal events memory. This is invoked when one is asked questions like"Where were you when JFK was killed?", and"What did you have for dinner last night?" The medial temporal lobe (which includes the hippocampus) is thought to be the storage site for these episodic memories.

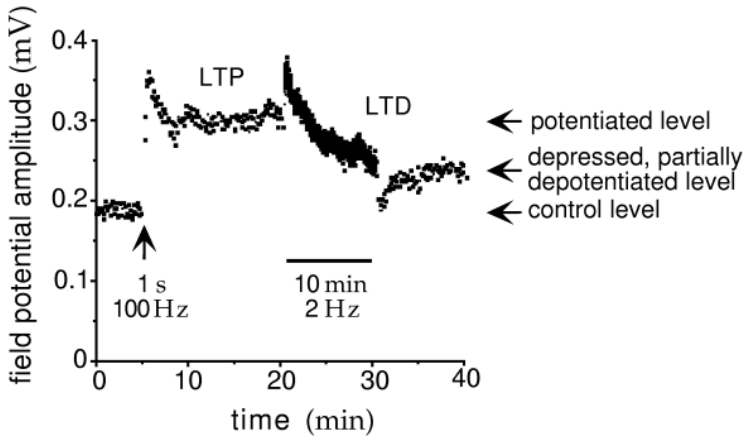


Figure 12.1: Hippocampal plasticity. First, LTP is induced by high frequency extracellular stimulation, and 10 minutes later it is partly reversed by low frequency stimulation.From Dayan and Abbott 2002.

In all learning networks we present input and look how the output of the network develops. We can distinguish*supervised* from*unsupervised* learning. In supervised learning we have a certain task that we want the network to perform, feedback is given to the network whether the performance was correct. A good example is a feed-forward network with back-propagation. These networks can in principle be taught any function between input and output. Essential is the presence of a hidden layer and some non-linearity in the transfer function. One can argue that supervised algorithms are nothing more but a fancy way of curve fitting an input-output relation. This is true, but the supervised algorithms remain an interesting and inspiring subject.

On an abstract level the brain is certainly able to do supervised learning. But which algorithm the brain uses for that is unclear. The backpropagation algorithm is hard to implement biologically. Here however, we will concentrate mainly on unsupervised learning, which is thought be important for receptive field learning, learning about regularities in the world, and associative memory. Hebbian learning is unsupervised when cell B does not receive a feedback signal from a teacher. Here we research variants of Hebb's learning rule and see how the network and synaptic weights develop under these rules.

12.1 LTP and LTD

Experimentally, LONG TERM POTENTIATION (LTP) was discovered in hippocampal slice (Bliss and Lomo 1973). An example is shown in Fig. 12.1. LTP can be induced in a number of ways: by brief high frequency extracellular stimulation, by providing input while depolarising the post-synaptic membrane potential, or by inducing pre- and postsynaptic spikes. This leads to strengthening of the synapses. In the figure one sees that LTP usually partly decreases after induction, but despite the stimulus only lasting a second, the remaining part is known to persist for months!

Although other brain regions show similar plasticity, hippocampus has remained the preferred system for studying LTP and its counterpart LTD (long term depression). The hippocampus is an associative shorter term storage in which information is thought to be stored for days or weeks, before it is transferred to the cortex (many details are unknown). In the hippocampus LTP is easy to induce. Furthermore, in the hippocampus the input path and synapses are clearly separated which makes extracellular experiments easier. Testing of hippocampal memory formation can be done in the Morris water-maze, in which the rat has to remember the location of a submerged platform in milky water. It can do that by remembering landmarks. In the hippocampus there is fairly good evidence that interfering with LTP, worsens the rat's performance, see e.g. (Riedel

et al. 1999).

The role of the hippocampus in humans was shown dramatically in patient H.M. who had both his hippocampi removed in an operation. As a result he could not store new memories anymore, and memories from shortly before the operation were lost. Memories from long before the operation were preserved, suggesting that these had over time been transferred to the cortex.

12.2 Biology of LTP and LTD

In studying the neurobiology of plasticity it is useful to consider the Induction, Expression and Maintainance of plasticity.

Induction

The induction of LTP is best understood: 1) synaptic input arrives, opening the AMPA receptors. 2) the EPSP propagates sub-threshold to the soma. 3) a spike is generated 4) the spike travels not only down the axon, but also back-propagates into the dendrites. The back-propagation spike arriving at the synapse releases the voltage-dependent Mg block of the NMDA receptors. 4) Ca flows into the synapse (both through NMDA channels and through voltage dependent Ca channels). 5) The Ca influx leads to the induction of LTP. (The buildup of calcium provides a possible reason why higher pairing frequencies give more LTP).

Expression

Via a complicated biochemical cascase the Ca influx leads to LTP. As we briefly mentioned in our discussion of synapses, there are multiple ways the synapse could potentially be modified to change its strength. 1) The release probablity of the synapse increases. 2) There is some evidence that right after LTP the AMPA receptors are put in a phosphorilated state which gives them a higher conductance than normal, un-phosphorilated AMPA receptors. 3) Also after induction, new receptors are inserted and the phosphorilated ones go back to an un-phosphorilated state. Perhaps now the synapse is in a naive state again but with a higher conductance as there are more receptors. However, the sequence of events causing LTP is not known yet, and different systems might use different LTP mechanisms. For a recent computational model see e.g. (Costa et al. 2015).

Maintainance

Memories can last a very long time. While this does not necessarily imply that the synapses are equally stable - the synapses that code the memory could change - many people believe they are.

There are two potential problems: the overwriting of old memories with new one (the stability-plasticity dilemma). This has to be solved on an algorithmic level. The second issue is to make synapses robust to random fluctuations common in biology. One way is to, like computers, make the synapses binary. This can dramatically increase the synaptic persistence.

More recently, a molecule called PKM ζ has been proposed as a molecular switch that, when set, maintains the synapse in a strengthened state. Interestingly, in animal experiments blocking this molecule with a drug leads to erasure of even quite old memories (Pastalkova et al. 2006).

12.3 Memory and LTP/LTD

Although the idea that alterations in the synaptic strength is appealing, there remained for a long time a healthy doubt whether synaptic plasticity really underlies learning. In (Martin and Morris 2002) 4 criteria are identified that must be fulfilled to make this link:

DETECTABILITY: If an animal displays memory of some previous experience, a change in synaptic efficacy should be detectable somewhere in its nervous system, Fig. 12.2.

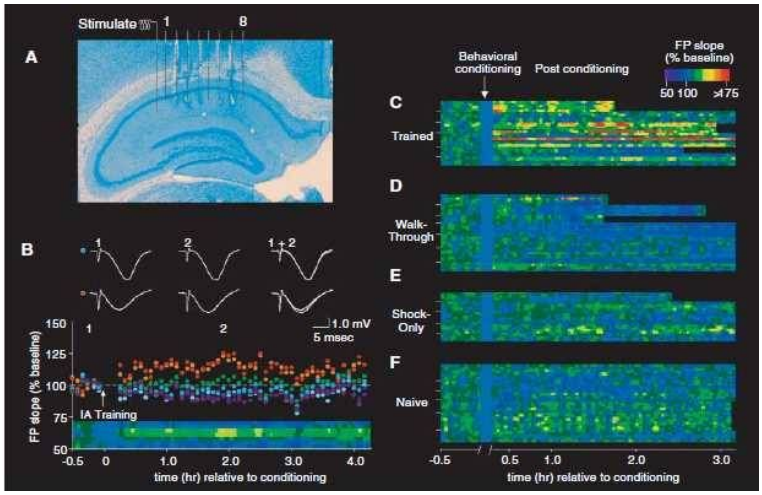


Figure 12.2: Demostration that learning in vivo leads to strengthening of synapses in the hippocampus.From Whitlock et al. 2006.

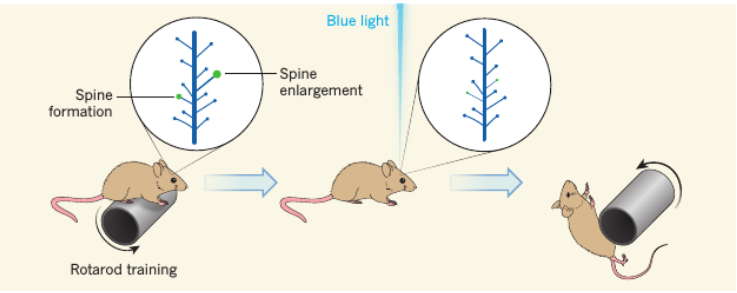


Figure 1 | Inducing forgetting. A neuron receives excitatory signals from other neurons through dendritic spines. When a mouse learns a new task, such as running on an accelerating rotating rod (a rotarod), spines involved in learning this task become potentiated (new spines form and existing spines increase in size). Hayashi-Takagi *et al.*[2] developed an 'optogenetic construct' based on a light-activatable form of the small signalling protein Rac1, which targets recently potentiated dendritic spines. Blue light activates the modified Rac1, which induces shrinkage of the spines. The authors found that spine shrinkage caused the mouse to forget the skill it had learnt, so it soon fell off the rotating rod.

Figure 12.3: Optically induced forgetting of memory. From comment onHayashi-Takagi et al. 2015.

MIMICRY: If it were possible to induce the appropriate pattern of synaptic weight changes artificially, the animal should display 'apparent' memory for some past experience which did not in practice occur.

ANTEROGRADE ALTERATION: Interventions that prevent the induction of synaptic weight changes during a learning experience should impair the memory of that experience.

RETROGRADE ALTERATION: Interventions that alter the spatial distribution of synaptic weight changes induced by a prior learning experience (see detectability) should alter the animal's memory of that experience.

In recent years the evidence has grown quite strong (although mimicry remains the toughest criterion to test). In one recent paper synapses were manupilated such that those that were potentiated could be optically reset (Hayashi-Takagi et al. 2015). This reset caused the animal to 'forget' the motor task it had learned, Fig. 12.3.

It also is now possible to track neurons and synapses optically over many days. This has shown that in the hippocampus the representation are not as stable as once thought (Ziv et al. 2013), and

that in hippocampusm but not in neo-cortex synapses turn over at a substantial rate (Attardo, Fitzgerald, and Schnitzer 2015).

Chapter 13

Computational models of plasticity

Mathematically we can describe Hebbian learning as follows. Suppose, there are N inputs neurons, which we label with $i = 1..N$. The pre-synaptic neurons all connect to one post-synaptic neuron. In the simplest rate approximation, the post-synaptic activity y equals the weighted sum of the inputs

$$y = \sum_{j=1}^{N} w_j x_j = \mathbf{w}.\mathbf{x}$$

where $\mathbf{w}$ are the weights of the inputs, which will be modified in the learning. Note that we have a fully linear neuron for which the activity and the weights can be negative.

The experiments seem to indicate that combining high pre- and post-synaptic activity leads to a strengthening of the synapse. One of the simplest ways to interpret the experimental results is to write

$$\begin{aligned} \Delta w_i &= \epsilon y x_i \\ &= \epsilon x_i \sum_{j=1}^{N} w_j . x_j \end{aligned}$$

where ϵ is a small constant which is the learning rate. It is usually chosen small so that a single presentation only gives a small change in the weights. This is more mathematical convenience more than an effort to describe biology; 'one-shot learning' seems to exists as we all know from daily life.

Suppose we have M patterns, labelled with $\mu = 1..M$. Every so many milliseconds we change the pattern. The total weight change after all patterns have occurred an equal amount of time is

$$\Delta w_i = \epsilon \sum_{\mu=1}^{M} x_i^\mu \sum_{j=1}^{N} w_j x_j^\mu$$

Define the correlation matrix of the inputs as $Q_{ij} = \sum_\mu x_i^\mu x_j^\mu$. The weights w_i can be written as a vector $\mathbf{w}$. We find that

$$\Delta \mathbf{w} = \epsilon \, Q.\mathbf{w}$$

As a last step one can write as a differential equation to describe to evolution of the weights

$$\boxed{\tau \frac{d\mathbf{w}}{dt} = Q.\mathbf{w}}$$

where τ describes the learning dynamics.

The above learning rule is the most plain Hebb rule. We have made the following assumptions, which all are somewhat un-biological and deserve further study:

- The output neuron is linear. It can have negative activity, and the activity does not saturate.
- The synaptic weights can have both positive and negative values.
- The update is linear in pre- and post-synaptic activity.
- Every event causes the same amount of synaptic change. This ignores effects such as saturation of the synapses and neural firing. In biology there are also some indications that the amount of change depends on the synaptic weight itself such that LTP is weaker for already strong synapses, naturally capping the maximal weight.
- Related to the last point, the update rule has no history. It also ignores modulators (e.g. dopamine) which can facilitate and prevent learning.
- The learning rule is 'symmetric' in LTP and LTD; as we flip x_i to $-x_i$ LTP and LTD are interchanged, but magnitude and speed of depression and potentiation are identical. As the two are biophysically quite distinct processes, this is a strong assumption.

It is a good exercise to try to investigate variants of the above learning rule for which these assumptions no longer hold. The effect of many assumptions is unclear.

The correlation matrix Q is a special matrix: It is symmetric and it is*positive semi-definite,* which means that its eigenvalues are larger or equal than zero. From our discussion of Markov-diagrams we know how the weights are going to evolve:

$$\mathbf{w}(t) = \sum_k c_k \mathbf{w}_k \mathrm{e}^{\lambda_k t/\tau}$$

where λ_k is the k-th eigenvalue, with eigenvector $\mathbf{w}_k$. The coefficients c_k are to be solved from the initial condition $\mathbf{w}(0) = \sum_k c_k \mathbf{w}_k$. Because the eigenvalues are larger than zero, the weights will run off to infinity. The reason for this divergence is a positive feedback: when a set of synapses becomes potentiated, their weights increase, which causes even more post-synaptic activity, which potentiates the synapses even further, etc. As a simple example, suppose that there are three independent patterns (0, 0, 10Hz), (0, 20Hz, 0) and (30Hz, 0, 0). The correlation matrix is proportional to $\begin{pmatrix} 900 & 0 & 0 \\ 0 & 400 & 0 \\ 0 & 0 & 100 \end{pmatrix}$ with eigenvalues 900, 400, and 100. When all the weights are non-zero to start with, the first synapse will eventually become the biggest. More generally, the learning will pick out the pattern which causes most post-synaptic activity, either because the activity at one input is high or because many inputs are simultaneously active (i.e. correlated).

13.1 Covariance rule

Realistic neurons can only have positive firing rates. If we therefore apply the above rule to rate based neurons, synapses can only get potentiated and never be depressed which is of course problematic as its leads to higher and higher activity. One can introduce a covariance learning rule (Sejnowski 1976)

$$\Delta w_i \quad = \quad \epsilon(x_i - \langle x_i \rangle)(y - \langle y \rangle)$$

where $\langle x_i \rangle = \frac{1}{M} \sum_\mu x_i^\mu$ is the activity averaged over the patterns. This is nothing more than the above rule, but now with Q the*covariance* matrix $Q_{ij} = \sum_\mu (x_i^\mu - \langle x \rangle)(x_j^\mu - \langle x \rangle)$. According to this rule, the synapse now becomes weaker, if either input or output are below average. Rather oddly, if both input and output are below average, the synapse becomes stronger as well. This latter condition can be removed without much consequences.

As stated, the learning picks out the inputs with the strongest correlation. These inputs are able to fire the post synaptic cell most and will therefore be enhanced. The unsupervised learning

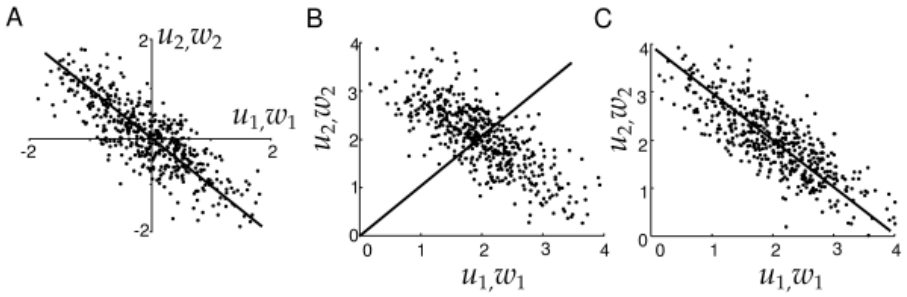


Figure 13.1: Hebbian learning of a neuron with two inputs. A) Two inputs which have strong negative correlation (each point represents a sample from the input distribution). The inputs are zero-mean. After learning with the plain Hebb rule the weight vector aligns with the data. B) When the data are not zero-mean, the weight vector aligns with the mean. C) In that case the covariance rule aligns with the data. From Dayan and Abbott 2002 and Hertz, Krogh, and Palmer 1991.

rules perform a PRINCIPAL COMPONENT ANALYSIS (PCA) on the data. (Below we discuss how to extract more than one principal component). Principal component analysis projects the data along the eigenvectors of the correlation matrix with the largest eigenvalues, Fig. 13.1. Describing the data this way is an efficient compression method, we can have many more input dimensions than output neurons and still have a decent description of the original input. This situation could occur in vision. An image has a very high dimension: each pixel is a separate dimensions (to describe an image you need a vector with as many elements as pixels). As we have seen in our discussion of visual processing there are many regularities in natural images which can be used to describe them more succinctly. For instance, neighbouring pixels will be positively correlated. The unsupervised learning rules pick up such regularities because they are driven by input correlations.

13.2 Normalisation

As one simulates the above rules, it soon becomes clear that the synaptic weights and the post-synaptic activity run off to un-biologically large values. One can stop learning after a certain time, or normalisation condition have to be imposed. The precise way one normalises the learning, however, turns out to be of crucial importance for the final weights, and hence the computation the neuron performs.

The easiest way is to incorporate a hard limit for each synaptic weight. For instance, $0 < w < w_{max}$. This can already have interesting effects on the dynamics, see Fig. 13.2.

Two other simple ways to constrain the synaptic weight are multiplicative and subtractive scaling. In the first we normalise with a second term which is proportional to the weight itself, in the second we subtract a constant factor:

$$\begin{aligned} \tau\frac{d\mathbf{w}}{dt} &= Q.\mathbf{w} - \gamma(w)\mathbf{w}(t) = Q.\mathbf{w} - [\frac{\mathbf{n}.Q.\mathbf{w}}{\mathbf{n}.\mathbf{w}}]\mathbf{w} \text{ Multiplicative} \\ &= Q.\mathbf{w} - \epsilon(w)\mathbf{n} = Q.\mathbf{w} - [\frac{\mathbf{n}.Q.\mathbf{w}}{\mathbf{n}.\mathbf{n}}]\mathbf{n} \text{ Subtractive} \end{aligned}$$

where $\mathbf{n}$ is the unit vector, $\mathbf{n} = (1, 1, 1, ...)$. You can easily check that in both cases $d(\mathbf{n}.\mathbf{w})/dt = 0$, that is the sum of weights $\sum_i w_i$ is constant.[1]Normalisation schemes automatically cause competition between the synaptic weights: one synapse can only win because the other gets reduced. The subtractive scheme is more competitive than the multiplicative scheme (see Practical). See (Miller and MacKay 1994) for an full analysis of these schemes.

[1]This way of writing the normalisation rules is useful for further analysis. If you just want to*simulate* learning, it is easier to determine the net weight change at each simulation step and adjust the weights so that their sum remains the same.

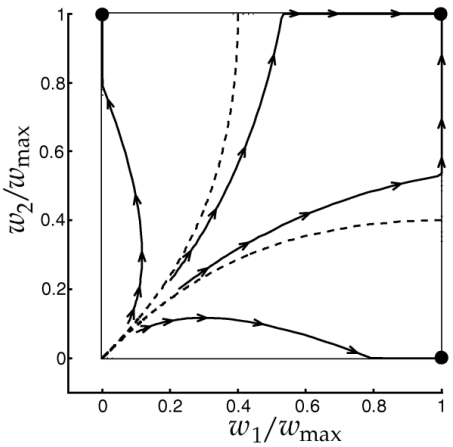


Figure 13.2: Weight development of two weights in which the weights have hard limits. Plotted in weight space, the arrows show how the weights develop under continuous stimulation. The inputs are anti-correlated with an correlation matrix $\begin{pmatrix} 1 & -0.4 \\ -0.4 & 1 \end{pmatrix}$, the largest eigenvector is $(1, -1)$. However, because of the limits on the growth, depending on the initial conditions different final weights are reached. There are three stable fix-points (filled circles) where the weights can end up. (With positively correlated inputs this does not happen and $(1, 1)$ is the only fixed points.)

In biology evidence for homeostasis of synaptic weights has been found (Turrigiano et al. 1998). On long time scales a neuron will try to make sure that it's activity is not too low or too high. It does this by both scaling its synaptic weight (in a multiplicative manner it seems) and by changing the densities of voltage gated channels.

One has introduced the terms HETERO-SYNAPTIC LTD and HOMO-SYNAPTIC LTD. In hetero-synaptic LTD, synapses depress in the absence of input to that synapse (maybe concurrently with the strengthening of a stimulated input). In homo-synaptic LTD, low pre-synaptic activity is required. With the inclusion of these normalization mechanisms the learning is no longerHOMO-SYNAPTIC (affecting a single synapse only), butHETERO-SYNAPTIC (a stimulus causes change in not only the associated synapse, but changes many synapses of the cell).

13.3 Oja's rule

Oja introduced a rule which automatically picks out the principal component in the inputs*and* normalises it. This rule is (Oja 1982)

$$\Delta w_i = \epsilon(x_i y - w_i y^2)$$

Note the quadratic term, which will constrain the weight. It is not difficult to show that the steady state needs to be an eigenvector of Q.

$$\begin{aligned} \Delta w_i &= \epsilon \sum_{\mu,j} w_j x_i^\mu x_j^\mu - \sum_{\mu,j,k} w_i w_j w_k x_j^\mu x_k^\mu \\ 0 &= Q.\mathbf{w} - (\mathbf{w}.Q.\mathbf{w})\mathbf{w} \end{aligned}$$

which can only be true if $\mathbf{w}$ is an eigenvector of Q, because for an eigenvector $Q\mathbf{w} = const.\mathbf{w}$. It is somewhat more effort to show that only the eigenvector with the largest eigenvalue is stable (Oja 1982; Hertz, Krogh, and Palmer 1991).

Oja's rule is a typical engineered rule, without direct biological relevance. There are two camps in this field, one trying to research the consequences of rules derived from biological data, and the other that tries to solve a task and derives the corresponding learning rule.

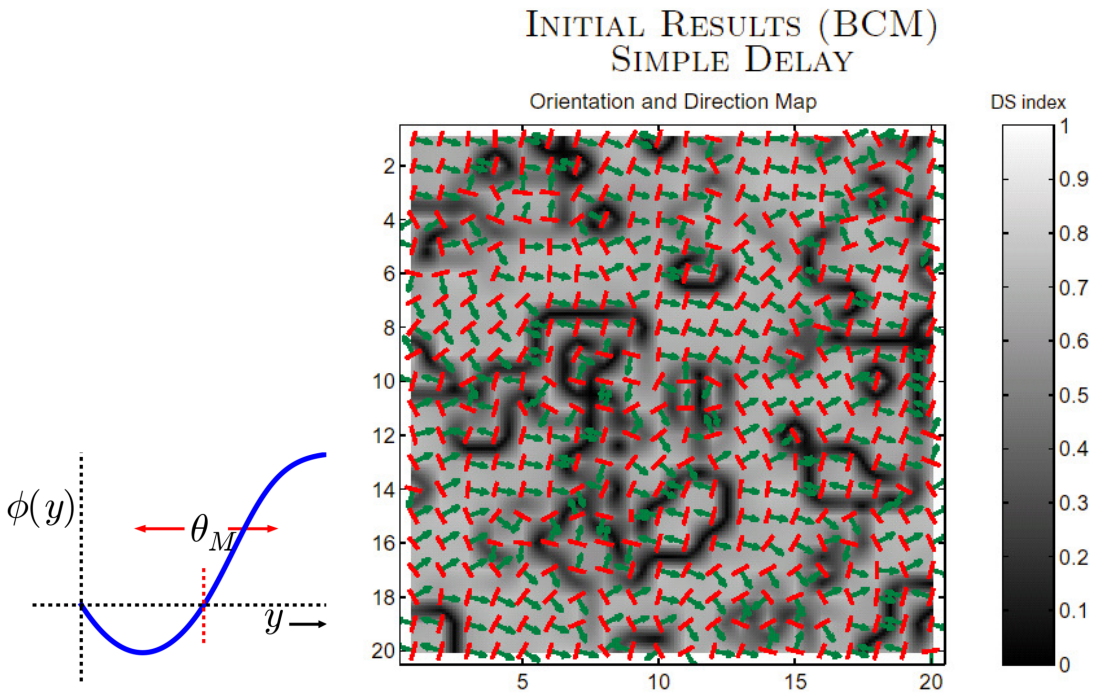


Figure 13.3: BCM rule. Left: the weight change as a function of the instantaneous post-synaptic activity (measured through the calcium concentration). In BCM, θ_+ shifts slowly with activity changes (θ_- is often set to zero).
Right: An application of the BCM rule. Development of orientation and direction selectivity in primary visual cortex modelled using the BCM rule.After Blais, Cooper, and Shouval 2000.

13.4 BCM rule

Bienenstock, Cooper and Munro have introduced a different way to deal with run-away synaptic weights (Bienenstock, Cooper, and Munro 1982). Their model is also known as BCM. In BCM, the learning rule is

$$\begin{aligned} \Delta w_i &= x_i \phi(y, \theta) \\ &= x_i [y(y - \theta)] \end{aligned}$$

Again, this learning rule would lead to run away excitation. But the trick is to continuously change the modification curve depending on the recent activity of the post-synaptic cell. We can do this by varying the threshold parameter θ. A good choice is to set $\theta = \langle y^2 \rangle$, that is, take a average of recent activity $\tau_\theta d\theta/dt = -\theta + y^2$. Now when, for instance, activity is high for a long time, the threshold is enhanced, leading to a depression for most patterns, and thus in turn lowering activity again. This way the learning rule has negative feedback build into it.

The shape of the function ϕ, see Fig. 13.3, is roughly what has been seen in experiments. BCM has homo-synaptic LTD; the modification is proportional to x_i, hence no modification without pre-synaptic input (in contrast to the covariance rule).

What really distinguished BCM from other learning rules, is the change in the threshold. Some evidence for that has been found (Kirkwood, Rioult, and Bear 1996). It seems a powerful learning model but its mathematical analysis is more tricky than the above models. Apparently, BCM can does PCA like decomposition like the above rules (Intrator and Cooper 1992). It also has been applied to the development of the visual system, e.g. Fig. 13.3.

13.5 Multiple neurons in the output layer

If we think for instance about the connection between LGN and V1, there is not a single neuron in the output, but there is a whole layer of output cells. Indeed there are many more V1 than LGN cells. If all the neurons in the output layer would select the same component in the input, there would be a lot of redundancy: the outputs would be fully correlated. The goal is therefore to on one hand to still pick out the interesting components, but at the same time to de-correlate.

For instance, using Oja's single unit rule above, all output neurons would select the same principal component of the data. We can add a term which is a bit like lateral inhibition as follows

$$\Delta w_{ij} = \epsilon(x_i y_j - y_j \sum_{k=1}^{M} w_{ik} y_k)$$

where M is the number of output neurons. Now the population of output neurons represents the first M principal components. Another method is Sanger's rule, which roughly does the same thing (Hertz, Krogh, and Palmer 1991).

Introducing lateral interaction between the neurons to induce competition is of much more general use. Inhibitory interactions reduce the redundancy between the neurons. This is called competitive learning.

In contrast excitatory interactions help to form populations of neurons coding the same stimulus properties. This decreases the noise associated to a single neuron spike train, increasing reliability and reducing integration times. Excitatory interactions do not seem necessary for noiseless, rate-based model neurons. However, when the neurons are noisy, such interactions help to average the signal. In addition they might perhaps be useful if you want to create population codes with rate based neurons.

13.6 ICA

PCA uses onlyFIRST AND SECOND ORDER STATISTICS which means it contains only terms such as $\langle x_i x_j \rangle$, $\langle x_i^2 \rangle$. This is because it is based on the correlation matrix. PCA analysis works therefore best on Gaussian distributed variables. Now the sum of many random numbers tends to a Gaussian, so it seems not a bad assumption. (see also the remarks about Fig. 7.6).

However, suppose that we are dealing with a signal in which many variables are mixed and we want to disentangle them. One way is to make linear combinations of inputs that are explicitly not Gaussian distributed, as these combinations are likely to be independent. This can be formulated in terms of information theory. This leads to so an extension of PCA, called independent component analysis (ICA), see Bell et al. (1997) (nice read).

The derivation (for noiseless neurons) goes as follows: suppose we have a set of neuron which have a sigmoid non-linearity. We can now maximise the mutual information of the output about the input by changing the synaptic weights. For a single neuron this leads to the weights (and an offset) to be such that the full output range of the neuron is used (much like the retinal adaptation we described). For multiple neurons the learning rule will de-correlate inputs. A learning rule which does this is

$$\Delta w_{ij} = \frac{cof(w_{ji})}{detW} + x_i(1 - 2y_j)$$

where $cof(w_{ij})$ is $(-1)^{i+j}$ times the determinant of the weight matrix W with row i and column j removed. However, it is not so clear how this would be implemented biologically, as it is a very non-local rule.

ICA is a more powerful technique than PCA, for instance, it allows the de-mixing of mixed signals. Suppose that three microphones each record in a room in which three persons are all speaking simultaneously. The signal of the microphones will be strongly correlated. How can we disentangle the voices. This is called the cocktail party problem. Applying the ICA algorithm will de-mix the signals and split the output into the 3 distinct sound sources. Whether this is like what

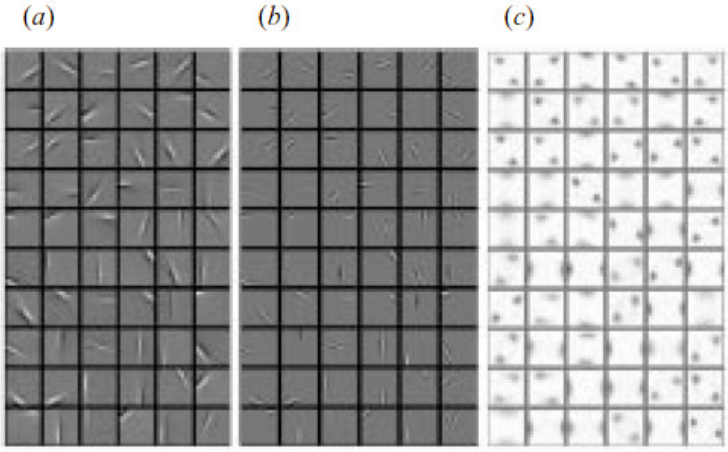


Figure 1. Independent component analysis on natural images (18×18 patches from 103 consecutive images, dimension reduced via PCA to 240 of 324). (*a*) IC basis vectors, and (*b*) corresponding IC filters (filtering an image with an IC filter yields the strength of the corresponding basis vector in the image). Signs of basis vectors and filters are arbitrary. (*c*) Amplitude spectra of the filters of (*b*), with darker grey values coding larger amplitudes. Zero spatial frequency is at the centre of each patch.

Figure 13.4: Receptive fields that result from doing ICA on natural images. FromHateren and van der Schaaf 1998.

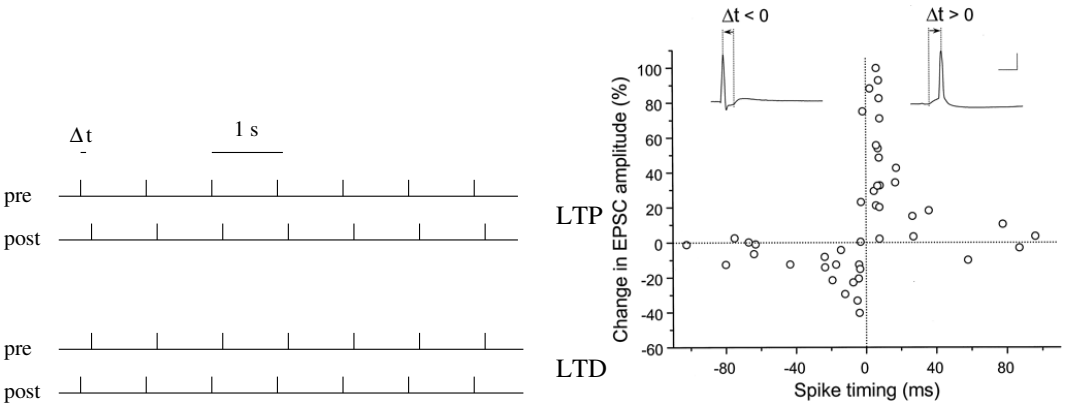


Figure 13.5: Left: Protocol for measuring spike time dependent plasticity. Right: Experimental data. Changes in synaptic strength after pairing pre- and post-synaptic spikes with 60 repetitions of pairings at 1Hz. FromBi and Poo 1998.

the brain does in auditory processing is not clear, but it is a useful algorithm for signal processing such as EEGs. It can also predict reasonably realistic V1 receptive fields, Fig. 13.4.

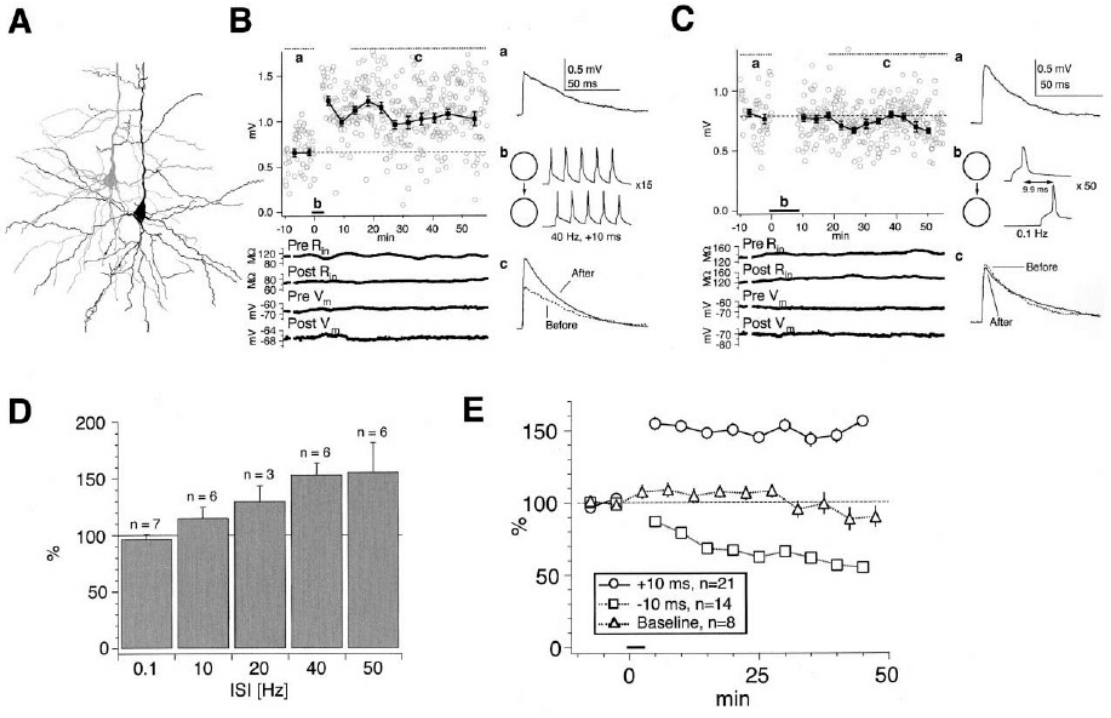


Figure 1. Spike Timing-Dependent LTP in Visual Cortical L5 Pairs Depends on Frequency

(A) Camera lucida drawing of a connected L5 pair. The left cell is presynaptic. Responses are shown in (B).

(B) LTP induction at 40 Hz. Top graph shows individual response amplitudes (open symbols) and five-min averages (filled symbols). Pre- and postsynaptic input resistance and membrane potential (lower traces) remained stable throughout the experiment. Right hand traces illustrate average unitary EPSPs before and after (Ba, Bc) induction (Bb), consisting of 15 bursts of five spikes at 40 Hz with presynaptic spikes preceding postsynaptic spikes by 10 ms (+10 ms). Average EPSP amplitude was potentiated by 163% (after/before, $p < 0.01$).

(C) Failure of LTP induction at low frequency. Responses from another pair in which 50 +10 ms pairings were delivered at 0.1 Hz (Cb) and no potentiation was observed.

(D) Pre-before-post firing (+10 ms) produced LTP at frequencies of 10 Hz and higher. The degree of potentiation increased with frequency.

(E) Pre-before-post firing (+10 ms) at frequencies between 10 and 50 Hz produced LTP (circles, n = 21), while post-before-pre firing (−10 ms) at frequencies between 0.1 and 20 Hz produced LTD (squares, n = 14). Low frequency pairing at large temporal offsets (+ 400 ms) produced little change in EPSP amplitude (triangles, n = 8).

Figure 13.6: Spike timing dependent plasticity in cortical cells.From Sjöström, Turrigiano, and Nelson 2001.

13.7 Spike Timing Dependent Plasticity

So far we looked at plasticity using average firing rates. Experiments show that the timing of the spikes is also of critical importance in the induction of LTP and LTD. Using patch clamp recording the pre- and post-synaptic event can be precisely timed. Such an experiment is shown in Fig. 13.5. If the pre-synaptic activity precedes the post-synaptic spike (and the input thus helps to induce the post-synaptic spike), the synapse is strengthened, but if the input lags behind the post-synaptic spike, the synapse is depressed. Thus the modification depends on the precise spike sequence. Note that this in remarkable agreement with Hebb's statement which implies that the pre-synaptic activity helps to*cause* the post-synaptic activity. The surprising part is the LTD condition. It is not clear yet whether all forms of the rate-based activity can be analysed in terms of this timing dependent plasticity.

It is know that again Calcium plays an important role, but the spatial- temporal profile of Ca concentration might very well matter as well (Nevian and Sakmann 2006).

Suppose we have one pre-synaptic from neuron i and one post-synaptic spike from neuron j. The simplest way to model this plasticity is to model the plasticity window as an exponential

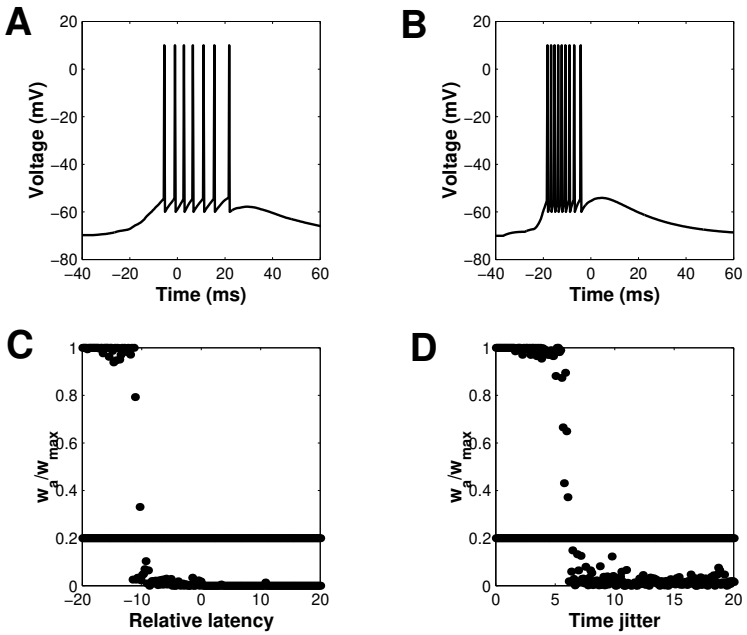


Figure 13.7: Spike timing dependent plasticity prefers early and reliable input. A) Post-synaptic spiketrain in response to periodic burst before training. B) After training the neuron responds earlier to the same input. C) Correspondingly, inputs with the shortest latencies are potentiated most. D) In another simulation different jitters were given to the inputs, the inputs with the least jitter were potentiated.

function, and to update the synaptic weight according to

$$\begin{aligned} \Delta w_{ij} &= A_+ \exp(-|t_i - t_j|/\tau_+) \quad (\text{if } t_i < t_j) \\ &= A_- \exp(-|t_i - t_j|/\tau_-) \quad (\text{if } t_j < t_i) \end{aligned}$$

where t_i is the pre-synaptic spike-time and t_j the post-synaptic spike-time, and A_+, A_-, τ_-, and τ_+ are constants to be extracted from the data. The τ_+ is usually tens of milliseconds, τ_- can be longer (100ms).

But how to cope with multiple spikes? Suppose pre- and post-synaptic activity are not well separated pairs as in Fig. 13.5, but contain trains of many spikes. Should the total strengthening be found by summing over all possible pairs, or perhaps should we just take closest pairs into account? In other words, interactions beyond simple pairing might play a role. Fig. 13.5 does not tell us how to deal with that situation (and is therefore a bit confusing). Such effects need to be measured experimentally, and are currently actively researched (Froemke and Dan 2002; Sjöström, Turrigiano, and Nelson 2001; Pfister and Gerstner 2006).

Very much related to that, the potentiation part is stronger for higher pairing frequencies, whereas the depression part does not depend much on pairing frequency, Fig. 13.6D, and see Markram et al. (1997).

13.8 Implications of spike timing dependent plasticity

What makes this learning rule different from the rate based learning? We first consider the case that the inputs have no particular temporal pattern. Again the post-synaptic neuron will pick out the correlated inputs. The rule will pick up correlations on the order of the LTP window, i.e. some 20 ms. Tightly correlated inputs will be potentiated. (Higher order interactions, such as dependence of post-synaptic activity might modify this picture).

As in the rate based case, precautions have to be taken against run away synaptic weights. Either hard bounds can be introduced (Song, Miller, and Abbott 2000), or the learning rules can be made converging, a bit like in Oja's rule (Rossum, Bi, and Turrigiano 2000).

Temporal structure of the inputs will be of course important in these learning rules. If we stimulate with repeating patterns, the post-synaptic neuron will potentiate the earliest inputs. The reason is that an early input will never come after the post-synaptic spike, and therefore will not undergo depression. The selection of fastest inputs can be used to shorten navigation routes on a map (Blum and Abbott 1996). In addition inputs with little jitter are selected above inputs with more jitter, Fig. 13.7D.

Also in a network setting these plasticity models are interesting. They predict an interesting effect in which the lateral connections at first teach the neighbouring neurons, Fig. 13.8. The lateral interactions disappear later, their output lags behind feed-forward input, and thus these connections are removed.

13.9 Concluding remarks

There is a deep link between information processing and mathematical models of unsupervised learning (Linsker 1992). It would be nice if there were more direct experimental evidence that showed that this is what biological Hebbian learning does. It would be interesting to know if the cortex's goal is to maximise mutual information and that Hebbian learning is just a way to do that.

Many models can reproduce some organisation structure in the primary visual cortex: monocular dominance columns, orientation columns or simple cell receptive fields. And curiously, from a biological point of view the structure in V1 is perhaps more related to development than to learning... It is not so clear how to proceed to construct learning models that do higher level vision. Not many models have reached beyond V1 and explained the emergence of, say, a face cell. In addition, most models have build in rules and constraints which make sense when formulating the model, but have not much experimental support from biology.

More reading: NIP-course, Hertz, Krogh, and Palmer (1991) and Hyvärinen, Hurri, and Hoyer (2009)

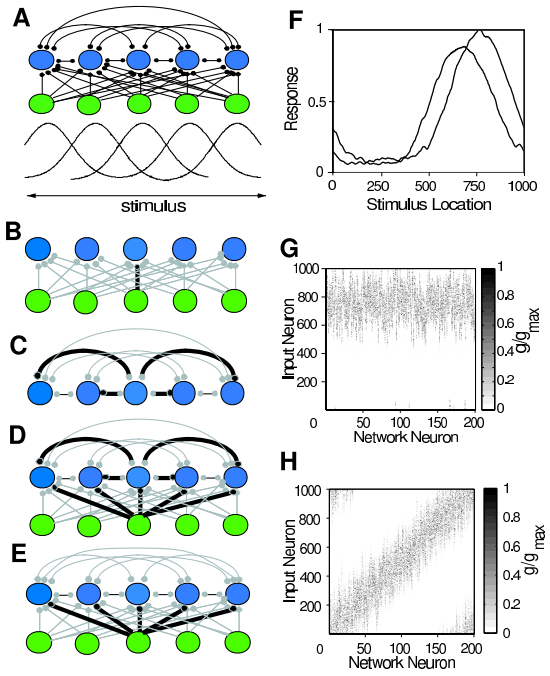


Figure 13.8: Formation of columns and maps under STDP. A) Network diagram. Two layers are simulated. The input layer contained neurons that fire Poisson trains at different rates and the network layer contained integrate-and-fire neurons. Every neuron in the input layer is randomly connected to one fifth of the neurons in the network layer and all the neurons in the network layer are recurrently connected. All synaptic connections are governed by STDP. Input correlations are introduced by moving a Gaussian hill of activity along the input layer. The Gaussian hill is centred on a random input neuron for a period of time and then shifted to another random location at the end of the period. The length of the period is chosen from an exponential distribution and the time constant is similar to the time window of STDP. B) First stage of column formation. Seed placed in the forward connections creates a correlated group of network neurons. C) Second stage of column formation. Correlated group of networks neurons send out connections to other neurons in the network. D) Third stage of column formation. Transfer of information from recurrent layer to feed-forward layer. E) Last stage of column formation. Recurrent connections weaken as feed-forward connections become well formed. F) Receptive fields of two network neurons. G. Feed-forward synaptic strengths define a column. Dark dots represent strong synapses. The horizontal stripe indicates that the network neurons have similar input connections, I.e.. receptive fields. H. When short range excitatory and global inhibitory recurrent connections are introduced in the network layer, a map forms in the feed-forward connection. The diagonal bar reflects the progression of receptive field centres as we move across the sheet of network neurons. AfterSong and Abbott 2001.

Bibliography

van Kampen, N. G. (1992). *Stochastic processes in physics and chemistry*. Second. Amsterdam: North-Holland.

Ben-Yishai, R., R. L. Bar-Or, and H. Sompolinsky (1995). "Theory of orientation tuning in visual cortex." In: *Proc. Natl. Acad. Sci.* 92, pp. 3844–3848.

Bi, G.-q. and M.-m. Poo (1998). "Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type." In: *J Neurosci* 18, pp. 10464–10472.

Liu, Y. H. and X. J. Wang (2001). "Spike-Frequency Adaptation of a Generalized Leaky Integrate-and-Fire Model Neuron." In: *J. of Comput. Neurosc.* 10, pp. 25–45.

Blum, K. I. and L. F. Abbott (1996). "A model of spatial map formation in the hippocampus of the rat." In: *Neural Comp.* 8, pp. 85–93.

Oram, M. W. et al. (2001). "Excess synchrony in the motor cortical neurons provides redundant direction information with thtat from coarse temporal measures." In: *J Neurophysiol* 86, pp. 1700–1716.

Bell, C. C. et al. (1997). "Synaptic plasticity in cerebellum-like structure depends on temporal order." In: *Nature* 387, pp. 278–281.

Holt, G. R. et al. (1996). "Comparison of discharge variability In vitro and In vivo in cat visual neurons." In: *J Neurophysiol* 75, pp. 1806–1814.

Blais, B. A., L. N. Cooper, and H. Shouval (2000). "Formation of direction selectivity in natural scene environments." In: *Neural Comp.* 12, pp. 1057–1066.

Atick, J. J. and A. N. Redlich (1992). "What does the retina know about natural scenes." In: *Neural Comp.* 4, pp. 196–210.

Vinje, W. E. and J. L. Gallant (2000). "Sparse coding and decorrelation in primary visual cortex during natural vision." eng. In: *Science* 287.5456, pp. 1273–1276.

Platt, M. L. and P. W. Glimcher (1999). "Neural correlates of decision variables in parietal cortex." In: *Science* 400, pp. 233–238.

Lamme, V. A. and P. R. Roelfsema (2000). "The distinct modes of vision offered by feedforward and recurrent processing." In: *Trends in Neurosci.* 23, pp. 571–9.

Desai, N. S., L. C. Rutherford, and G. G. Turrigiano (1999). "Plasticity in the intrinsic electrical properties of cortical pyramidal neurons." In: *Nat. Neuro.* 2, pp. 515–520.

Lowen, S. B. et al. (2001). "Fractal features of dark, maintained, and driven neural discharges in the cat visual system." In: *Methods* 24, pp. 377–94.

Seung, H. S. and H. Sompolinsky (1993). "Simple models for reading neuronal population codes." In: *Proc. Natl. Acad. Sci.* 90, pp. 10749–10753.

Rolls, E. T. and G. Deco (2002). *Computational neuroscience of vision*. Oxford.

Engel, A. K., O. Konig, and W. Singer (1991). "Direct physiological evidence for scene segmentation by temporal coding." In: *Proc. Natl. Acad. Sci.* 88, pp. 9136–9140.

Berry, M. J. and M. Meister (1998). "Refractoriness and Neural Precision." In: *J Neurosci* 18, pp. 2200–2211.

Hayashi-Takagi, Akiko et al. (2015). "Labelling and optical erasure of synaptic memory traces in the motor cortex." In: *Nature*.

Rossum, M. C. W. van, B. J. O'Brien, and R. G. Smith (2003). "The effects of noise on the timing precision of retinal ganglion cells." In: *J Neurophysiol* 89, pp. 2406–2419.

Rossum, M. C. W. van, G. G. Turrigiano, and S. B. Nelson (2002). "Fast propagation of firing rates through layered networks of noisy neurons." In: *J Neurosci* 22, pp. 1956–1966.

Rossum, M. C. W. van, G. q. Bi, and G. G. Turrigiano (2000). "Stable Hebbian learning from spike timing dependent plasticity." In: *J Neurosci* 20.23, pp. 8812–8821.

Taylor, D. M., S. I. Helms Tillery, and A. B. Schwartz (2002). "Direct cortical control of 3D neuroprosthetic devices." In: *Science* 296, p. 1829.

Miller, K. D. and D. J. C. MacKay (1994). "The role of constraints in Hebbian learning." In: *Neural Comp.* 6.1, pp. 100–126.

Huxley, A. F. (1959). "Ion movements during nerve activity." In: *Ann N Y Acad Sci* 81, pp. 221–246.

Martin, S. J. and R. G. Morris (2002). "New life in an old idea: the synaptic plasticity and memory hypothesis revisited." In: *Hippocampus* 12, pp. 609–36.

Barlow, H. B. and W. R. Levick (1965). "The mechanism of directionally selective units in rabbit's retina." In: *J. Physiol.* 178, pp. 477–504.

Battro, A. M. (2000). *Half a brain is enough. The story of Nico.* Cambridge: Cambridge University Press.

Bryant, H. L. and J. P. Segundo (1976). "Spike initiation by transmembrane current: a white-noise analysis." In: *J. Physiol.* 260, pp. 279–314.

Kandel, E. R., J. H. Schwartz, and T. M. Jessel (2000). *Principles of Neural science.* New York: McGraw Hill.

Varela, J. A. et al. (1997). "A quantitative description of short-term plasticity at excitatory synapses in layer 2/3 of rat primary visual cortex." In: *J Neurosci* 17, pp. 7926–7940.

Mainen, Z. F. and T. J. Sejnowski (1995). "Reliability of Spike timing in neocortical neurons." In: *Science* 268, pp. 1503–1506.

Barlow, H. B. (1958). "Temporal and spatial summation in human vision at different background intensities." In: *J. Physiol.* 141, pp. 337–350.

Kim, Sung Soo et al. (2017). "Ring attractor dynamics in the Drosophila central brain." In: *Science* 356.6340, pp. 849–853.

Abbott, L. F. et al. (1997). "Synaptic depression and cortical gain control." In: *Science* 275, pp. 220–224.

Schall, J. D., V. Stuphorn, and J. W. Brown (2002). "Monitoring and control of action by the frontal lobes." In: *Neuron* 36, pp. 309–322.

Hateren, J. H. van and A. van der Schaaf (1998). "Independent component filters of natural images compared with simple cells in primary visual cortex." In: *Proc. Roc. Soc. B* 265, pp. 359–366.

Stevens, C. F. and A. M. Zador (1998). "Input synchrony and the irregular firing of cortical neurons." In: *Nat. Neuro.* 1, pp. 210–217.

Hodgkin, A. L. and A. F. Huxley (1952). "A quantitative description of membrane current and its application to conduction and excitation in nerve." In: *J. Physiol.* 117, pp. 500–544.

Shadlen, M. N. and W. T. Newsome (1998). "The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding." In: *J Neurosci* 18, pp. 3870–3896.

Newsome, W. T., K. H. Britten, and J. A. Movshon (1989). "Neuronal correlates of a perceptual decision." In: *Nature* 341, pp. 52–54.

Hoffman, D. A. et al. (1997). "K^+ channel regulation of signal propagation in dendrites of hippocampal pyramidal neurons." In: *Nature* 387, pp. 869–875.

Hateren, J. H. van (1992). "Real and optimal neural images in early vision." In: *Nature* 360, pp. 68–70.

Britten, K. H. et al. (1992). "The analysis of visual motion: a comparison of neuronal and psychophysical performance." In: *J Neurosci* 12, pp. 4745–4765.

Johnson, D. M. et al. (2000). "New features of connectivity in piriform cortex visualized by intracellular injection of pyramidal cells suggest that "primary" olfactory cortex functions like "association" cortex in other sensory systems." In: *J Neurosci* 20.18, pp. 6974–6982.

Colbert, C. M. and E. Pan (2002). "Ion channel properties underlying axonal action potential initiation in pyramidal neurons." In: *Nat. Neuro.* 5, pp. 533–548.

Shannon, C. E. and W. Weaver (1949). *The mathematical theory of communication.* Illinois: Univeristy of Illinois Press.

Stevens, C. F. (1972). "Interferences about membrane properties from electrical noise measurements." In: *Biophys J* 12, pp. 1028–1047.

Tsodyks, M. V. and H. Markram (1997). "The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability." In: *Proc. Natl. Acad. Sci.* 94, pp. 719–723.

Buracas, G. T. et al. (1998). "Efficient discrimination of temporal patters by motion sensitive neurons in primate visual cortex." In: *Neuron* 20, pp. 959–969.

Herculano-Houzel, Suzana (2009). "The human brain in numbers: a linearly scaled-up primate brain." eng. In: *Front Hum Neurosci* 3, p. 31. DOI: 10.3389/neuro.09.031.2009.

Costa, Rui Ponte et al. (2015). "Unified pre-and postsynaptic long-term plasticity enables reliable and flexible learning." In: *eLife* 4, e09457.

Kapadia, M. K., G. Westheimer, and C. D. Gilbert (2000). "Spatial distribution of contextual interactions in primary visual cortex and in visual perception." In: *J Neurophysiol* 84, pp. 2048–2062.

Kalaska, J. F., R. Caminiti, and A. P. Georgopoulos (1983). "Cortical mechanisms related to the direction of two-dimensional arm movements: Relations in parietal area 5 and comparison with motor cortex." In: *Exp Brain Res* 51, pp. 247–260.

Paradiso, M. A. (1988). "A theory for the use of visual orientation information which exploits the columnar structure of striate cortex." In: *Biol Cybern* 58, pp. 35–49.

Mitchell, S. J. and R. A. Silver (2003). "Shunting inhibition modulates neuronal gain during synaptic excitation." In: *Neuron* 8, pp. 433–445.

Anderson, J. S. et al. (2000). "The contribution of noise to contrast invariance of orientation tuning in cat visual cortex." In: *Science* 290, pp. 1968–1972.

Lee, Albert K and Matthew A Wilson (2002). "Memory of sequential experience in the hippocampus during slow wave sleep." In: *Neuron* 36.6, pp. 1183–1194.

Tuckwell, H. C. (1988). *Introduction to theoretical neurobiology.* Cambridge: Cambridge University Press.

Shepherd, G. M. (1994). *Neurobiology.* New York: Oxford.

Gerstein, G. L. and B. Mandelbrot (1964). "Random walk models for the spike activity of a single neuron." In: *Biophys J* 4, pp. 41–68.

Erickson, C. A., B. Jagadeesh, and R. Desimone (2000). "Clustering of perirhinal neurons with similar properties following visual experience in adult monkeys." In: *Nat. Neuro.* 3, pp. 1143–8.

Goris, Robbe L T., J Anthony Movshon, and Eero P. Simoncelli (2014). "Partitioning neuronal variability." eng. In: *Nat Neurosci* 17.6, pp. 858–865. DOI: 10.1038/nn.3711. URL: http://dx.doi.org/10.1038/nn.3711.

Rossum, Mark C W van et al. (2008). "Adaptive integration in the visual cortex by depressing recurrent cortical circuits." In: *Neural Comput* 20.7, pp. 1847–1872. DOI: 10.1162/neco.2008.06-07-546.

McCullogh, W. S. and W. Pitts (1943). "A logical calculus of the ideas immanent in nervous activity." In: *Bull Math Biophys* 5, pp. 115–113.

Sejnowski, T. J. (1976). "Statistical constraints on synaptic plasticity." In: *J Theor Biol* 69, pp. 385–398.

Heydt, R. von der, E. Peterhans, and G. Baumgartner (1984). "Illusory contours and cortical neuron responses." In: *Science* 224, pp. 1260–2.

Steveninck, R. R. de Ruyter van et al. (1997). "Reproducibility and variability in neural spike trains." In: *Science* 275, pp. 1805–1809.

Oja, E. (1982). "A simplified neuron model as a principal component analyzer." In: *J Math Biol* 15, pp. 267–273.

Churchland, P. S. and T. J. Sejnowski (1994). *The Computational Brain.* MIT Press.

Turrigiano, G. G. et al. (1998). "Activity-dependent scaling of quantal amplitude in neocortical neurons." In: *Nature* 391, pp. 892–896.

Pfister, Jean-Pascal and Wulfram Gerstner (2006). "Triplets of spikes in a model of spike timing-dependent plasticity." In: *J Neurosci* 26.38, pp. 9673–9682. DOI: 10.1523/JNEUROSCI.1425-06.2006.

Holt, G. (1998). "A Critical Reexamination of Some Assumptions and Implications of Cable Theory in Neurobiology." PhD thesis. Caltech.

Bienenstock, E. L., L. N. Cooper, and P. W. Munro (1982). "Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex." In: *J Neurosci* 2, pp. 32–48.

York, Lawrence Christopher and Mark C W. van Rossum (2009). "Recurrent networks with short term synaptic depression." eng. In: *J Comput Neurosci* 27.3, pp. 607–620. DOI: 10.1007/s10827-009-0172-4.

Trappenberg, T. P. (2002). *Fundamentals of computational neuroscience*. Oxford.

Grubb, Matthew S. and Juan Burrone (2010). "Activity-dependent relocation of the axon initial segment fine-tunes neuronal excitability." eng. In: *Nature* 465.7301, pp. 1070–1074. DOI: 10.1038/nature09160.

Song, S., K. D. Miller, and L. F. Abbott (2000). "Competitive Hebbian Learning Through Spike-Timing-Dependent Synaptic Plasticity." In: *Nat. Neuro.* 3, pp. 919–926.

Jadhav, Shantanu P et al. (2012). "Awake hippocampal sharp-wave ripples support spatial memory." In: *Science* 336.6087, pp. 1454–1458.

Seelig, Johannes D and Vivek Jayaraman (2015). "Neural dynamics for landmark orientation and angular path integration." In: *Nature* 521.7551, pp. 186–191.

Sacks, O. (1985). *The man who mistook his wife for a hat and other clinical tales*. New York: Summit Books.

Bear, M., B. W. Connors, and M. A. Paradiso (2000). *Neuroscience: Exploring the Brain*. Lippincott, Williams and Wilkins.

Brunel, N. and J.-P. Nadal (1998). "Mutual information, Fisher information, and population coding." In: *Neural Comp.* 10, pp. 1731–1757.

Dayan, P. and L. F. Abbott (2002). *Theoretical Neuroscience*. Cambridge, MA: MIT press.

Abbott, L.F., E. T. Rolls, and M. J. Tovee (1996). "Representational Capacity of Face Coding in Monkeys." In: *Cereb Cortex* 6, pp. 498–505.

O'Donnell, Cian and Mark CW Van Rossum (2014). "Systematic analysis of the contributions of stochastic voltage gated channels to neuronal noise." In: *Front Comput Neurosci* 8, p. 105.

Whitlock, Jonathan R et al. (2006). "Learning induces long-term potentiation in the hippocampus." In: *Science* 313.5790, pp. 1093–1097. DOI: 10.1126/science.1128134.

Hille, B. (2001). *Ionic Channels of excitable membranes*. Sunderland, MA: Sinauer.

Zohary, E., M. N. Shadlen, and W. T. Newsome (1994). "Correlated neuronal discharge rate and its implications for psychophysical performance." In: *Nature* 370, pp. 140–144.

Luria, A. (1966). *Higher Cortical function in man*. Plenum.

Kirkwood, A., M. C. Rioult, and M. F. Bear (1996). "Experience-dependent modification of synaptic plasticity in visual cortex." In: *Nature* 381, pp. 526–528.

Funahashi, S., C. J. Bruce, and P. S. Goldman-Rakic (1998). "Mnemonic encoding of visual space in the monkey's dorsolateral prefrontal cortex." In: *J Neurophysiol* 61, pp. 1464–1483.

Treue, S. (2001). "Neural correlates of attention in primate visual cortex." In: *Trend in Neurosc.* 24, pp. 295–300.

Destexhe, A., Z. F. Mainen, and T. J. Sejnowski (1998). "Kinetic models of synaptic transmission." In: *Methods in neuronal modeling (2nd ed.)* Ed. by Koch and Segev. Cambridge: MIT Press.

Ratcliff, R. and J. N. Rouder (1998). "Modeling response times for two-choice decisions." In: *Psychological Science* 9. maybe other ratcliff papers are better, pp. 347–356.

Intrator, N. and L. N. Cooper (1992). "Objective function formulation of the BCM theory of visual cortical plasticity." In: *Neural Netw.* 5, pp. 3–17.

Blakemore, C. (1988). *The mind machine*. BBC books.

Linsker, R. (1992). "Local synaptic learning rules suffice to maximise mutual information in a linear network." In: *Neural Comp.* 4, pp. 691–704.

Pasupathy, A. and C. E. Connor (2001). "Shape representation in area V4: position-specific tuning for boundary conformation." eng. In: *J Neurophysiol* 86.5, pp. 2505–2519.

Pouget, A. and T. J. Sejnowski (1997). "Spatial transformations in the parietal cortex using basis functions." In: *J. Cogn. Neurosci.* 9, pp. 222–237.

Li, Nuo and James J. DiCarlo (2010). "Unsupervised natural visual experience rapidly reshapes size-invariant object representation in inferior temporal cortex." eng. In: *Neuron* 67.6, pp. 1062–1075. DOI: 10.1016/j.neuron.2010.08.029.

Hansel, D. et al. (1998). "On numerical simulations of integrate-and-fire neural networks." In: *Neural Comp.* 10, pp. 467–483.

Benda, Jan and Andreas V M Herz (2003). "A universal model for spike-frequency adaptation." In: *Neural Comput* 15.11, pp. 2523–2564. DOI: 10.1162/089976603322385063.

Brunel, N. (2000). "Dynamics of networks of randomly connected excitatory and inhibitory spiking neurons." In: *Journal of Physiology-Paris* 94, pp. 445–463.

Koch, C. (1999). *Biophysics of computation*. New York: Oxford University Press.

Bialek, W. (1987). "Physical limits to sensation and perception." In: *Annu Rev Biophys Biophys Chem* 16, pp. 455–78.

Arbib(editor), M. (1995). *The handbook of brain theory and neural networks*. Cambridge, MA: MIT press.

Paré, D. et al. (1998). "Impact of spontaneous synaptic activity on the resting properties of cat neocortical pyramidal neurons in vivo." In: *J Neurophysiol* 79, pp. 1450–1460.

Stricker, C. (2002). "Central Synaptic integration: Linear after all?" In: *News Physiol. Sci* 17, pp. 138–143.

Földiák, P. (1991). "Learning invariance from transformation sequences." In: *Neural Comp.* 3, pp. 194–200.

Ziv, Yaniv et al. (2013). "Long-term dynamics of CA1 hippocampal place codes." In: *Nat Neurosci* 16.3, pp. 264–266.

Rieke, F. et al. (1996). *Spikes: Exploring the neural code*. Cambridge: MIT Press.

Hertz, J., A. Krogh, and R. G. Palmer (1991). *Introduction to the theory of neural computation*. Reading, MA: Perseus.

Földiák, Peter (2013). "Sparse and explicit neural coding." In: ed. by Panzeri et al. CRC Press, pp. 379–390.

Jones, J. and L. Palmer (1987). "The two-dimensional spatial structure of simple receptive fields in cat striate cortex." In: *J Neurophysiol* 58, pp. 1187–1211.

Wessberg, J. et al. (2000). "Real-time prediction of hand trajectory by ensembles of cortical neurons in primates." In: *Nature* 408, pp. 361–368.

Abeles, M. (1991). *Corticonics: neural circuits of the cerebral cortex*. Cambridge: Cambridge University Press.

Br?maud, Antoine, David C West, and Alex M Thomson (2007). "Binomial parameters differ across neocortical layers and with different classes of connections in adult rat and cat neocortex." In: *Proc Natl Acad Sci U S A* 104.35, pp. 14134–14139. DOI: 10.1073/pnas.0705661104.

Brunel, Nicolas and Mark C W van Rossum (2007). "Lapicque's 1907 paper: from frogs to integrate-and-fire." In: *Biol Cybern* 97.5-6, pp. 337–339. DOI: 10.1007/s00422-007-0190-0.

Stemmler, M. and C. Koch (1999). "How voltage-dependent conductances can adapt to maximize the information encoded by neuronal firing rate." In: *Nat. Neuro.* 2, pp. 512–527.

Bliss, T.V.P. and T. Lomo (1973). "Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path," in: *J. Physiol.* 232, pp. 331–56.

Gerstner, W. (2000). "Population dynamics of spiking neurons: Fast transients, asynchronous state, and locking." In: *Neural Comp.* 12, pp. 43–89.

Lapicque, L. (1907). "Recherches quantitatives sur l'excitation electrique de nerfs traitee comme une polarization." In: *J. Physiol. Pathol. Gen.* 9.

Chadwick, Angus, Mark CW van Rossum, and Matthew F Nolan (2016). "Flexible theta sequence compression mediated via phase precessing interneurons." In: *eLife* 5, e20349.

Chadwick, Angus, Mark CW van Rossum, and Matthew F Nolan (2015). "Independent theta phase coding accounts for CA1 population sequences and enables flexible remapping." In: *Elife* 4, e03542.

Froemke, R.C. and Y. Dan (2002). "Spike-timing-dependent synaptic modification induced by natural spike trains." In: *Nature* 416.6879, pp. 433–8.

Tsunoda, K. et al. (2001). "Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns." In: *Science* 4, pp. 832–838.

Tam?s, G. et al. (2000). "Proximally targeted GABAergic synapses and gap junctions synchronize cortical interneurons." In: *Nat Neurosci* 3.4, pp. 366–371. DOI: 10.1038/73936.

Bair, W. and C. Koch (1996). "Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey." In: *Neural Comp.* 8, pp. 1185–1202.

Wimmer, Klaus et al. (2014). "Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory." In: *Nature neuroscience* 17.3, pp. 431–439.

Compte, A. et al. (2000). "Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model." In: *Cereb Cortex* 10, pp. 910–923.

Holt, G. and C. Koch (1997). "Shunting inhibition does not have a divisive effect on firing rates." In: *Neural Comp.* 9, pp. 1001–1013.

Song, S. and L.F. Abbott (2001). "Column and Map Development and Cortical Re-Mapping Through Spike-Timing Dependent Plasticity." In: *Neuron* 32, pp. 339–350.

Tanaka, K. (1996). "Inferotemporal cortex and object vision." In: *Annu Rev Neurosci* 19, pp. 109–139.

Brunel, N and S Sergi (1998). "Firing frequency of leaky intergrate-and-fire neurons with synaptic current dynamics." In: *J Theor Biol* 195.1, pp. 87–95.

Mack, V. et al. (2001). "Conditional restoration of hippocampal synaptic potentiation in Glur-A-deficient mice." In: *Science* 292, pp. 2501–4.

Fukushima, K. (1980). "Neocognitron: A self-organising multi-layered neural network." In: *Biol Cybern* 20, pp. 121–136.

Riedel, G. et al. (1999). "Reversible neural inactivation reveals hippocampal participation in several memory processes." In: *Nat. Neuro.* 2, pp. 898–907.

Watanabe, S. et al. (2002). "Dendritic K+ channels contribute to spike-timing dependent long-term potentiation in hippocampal pyramidal neurons." In: *Proc. Natl. Acad. Sci.* 99, pp. 8366–8371.

Markram, H. et al. (1997). "Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs." In: *Science* 275, pp. 213–215.

Andrewes, D. (2002). *Neuropsychology: From theory to practice*. Psychology Press.

Sakai, K. and Y. Miyashita (1991). "Neural organization for the long-term memory of paired associates." In: *Nature* 354, pp. 152–157.

Pastalkova, Eva et al. (2006). "Storage of spatial information by the maintenance mechanism of LTP." In: *Science* 313.5790, pp. 1141–1144.

Johnston, D. and S. Wu (1995). *Foundations of cellular Neurophysiology*. Cambridge, MA: MIT Press.

Attardo, Alessio, James E Fitzgerald, and Mark J Schnitzer (2015). "Impermanence of dendritic spines in live adult CA1 hippocampus." In: *Nature* 523 (7562), pp. 592–596. ISSN: 1476-4687. DOI: 10.1038/nature14467.

Alle, Henrik, Arnd Roth, and Jörg R P. Geiger (2009). "Energy-efficient action potentials in hippocampal mossy fibers." eng. In: *Science* 325.5946, pp. 1405–1408. DOI: 10.1126/science.1174331.

Roth, A and M van Rossum (2009). "Modeling synapses." In: *Computational Modeling Methods for Neuroscientists*. Ed. by E De Schutter. MIT Press.

Sompolinsky, H. et al. (2002). "Population coding in neuronal systems with correlated noise." In: *Phys. Rev E* 64, p. 51904.

Thorpe, S., D. Fize, and C. Marlot (1996). "Speed of processing in the human visual system." In: *Nature* 381, pp. 520–522.

Sterratt, David et al. (2011). *Principles of computational modelling in neuroscience*. Cambridge University Press.

Braitenberg, V. and A. Schüz (1998). *Statistics and geometry of neuronal connectivity*. Springer.

Nevian, Thomas and Bert Sakmann (2006). "Spine Ca2+ signaling in spike-timing-dependent plasticity." In: *J Neurosci* 26.43, pp. 11001–11013. DOI: 10.1523/JNEUROSCI.1749-06.2006.

Touryan, J., G. Felsen, and Y. Dan (2005). "Spatial Structure of Complex Cell Receptive Fields Measured with Natural Images." In: *Neuron* 45, pp. 781–791.

Markram, Henry et al. (2015). "Reconstruction and Simulation of Neocortical Microcircuitry." eng. In: *Cell* 163.2, pp. 456–492. DOI: 10.1016/j.cell.2015.09.029.

Schutter, E. de and Paul Smolen (1998). "Calcium dynamics in Large neuronal models." In: *Methods in neuronal modeling, 2^{nd} ed.* Ed. by C. Koch and I. Segev. Cambridge: MIT Press, pp. 211–250.

Vreeswijk, C. van and H. Sompolinsky (1996). "Chaos in neuronal networks with balanced excitatory and inhibitory activity." In: *Science* 274, pp. 1724–1726.

Riesenhuber, M. and T. Poggio (1999). "Hierarchical models of object recognition in cortex." In: *Nat. Neuro.* 2, pp. 1019–1025.

Wang, Y., I. Fujita, and Y. Murayama (2000). "Neuronal mechanisms of selectivity for object features revealed by blocking inhibition in inferotemporal cortex." In: *Nat. Neuro.* 3, pp. 807–812.

Hyvärinen, A., J. Hurri, and P.O. Hoyer (2009). *Natural Image Statistics*. Spinger.

Buchsbaum, G. and A. Gottschalk (1983). "Trichromacy, Opponent Colours Coding and Optimum Colour Information Transmission in the Retina." In: *Proc. R. Soc. B* 220, pp. 89–113.

Logothetis, N.K. et al. (2001). "Neurophysiological investigation of the basis of the fMRI signal." In: *Nature* 412, pp. 150–7.

Markram, H., Y. Wang, and M. Tsodyks (1998). "Differential signaling via the same axon of neocortical pyramidal neurons." In: *Proc Natl Acad Sci U S A* 95.9, pp. 5323–5328.

Sjöström, P.J., G.G. Turrigiano, and S.B. Nelson (2001). "Rate, timing, and cooperativity jointly determine cortical synaptic plasticity." In: *Neuron* 32, pp. 1149–1164.

Yarrow, Stuart, Edward Challis, and Peggy Seri?s (2012). "Fisher and Shannon information in finite neural populations." eng. In: *Neural Comput* 24.7, pp. 1740–1780. DOI: 10.1162/NECO_a_00292.

Feuillet, Lionel, Henry Dufour, and Jean Pelletier (2007). "Brain of a white-collar worker." In: *Lancet* 370.9583, p. 262. DOI: 10.1016/S0140-6736(07)61127-1.