Analysis of Neural Excitability and Oscillations

C. Gielen
Dept. of Biophysics
University of Nijmegen
Netherlands

September 6, 2009

Contents

1 Analysis of Neural Excitability and Oscillations

1.1 Introduction

Qualitative features of excitable or oscillatory dynamics are shared by broad classes of neuronal models. Expressed in models for single-cell behavior as well as for ensemble activity, these features include excitability and threshold behavior; beating and bursting oscillations and phase locking; and bistability and hystersis. Our goal here is to illustrate, by exploiting a specific model of the excitable membrane, some of the concepts and techniques that can be used to understand, predict, and interpret these dynamic phenomena biophysically. Our mathematical methods include numerical integration of the model equations, graphical or geometric representation of the dynamics (phase plane analysis), and analytic formulae for characterizing thresholds and stability conditions. The concepts are from the qualitative theory of nonlinear differential equations and nonlinear oscillations, and from perturbation and bifurcation theory. In this chapter, we will not consider the spatiotemporal aspects of distributed systems. Thus our methods apply directly only to a membrane patch, to a spatially uniform, equipotential cell, or to a network with each cell type perfectly synchronized.

Even seemingly simple models, that exhibit one or two of the different dynamic behaviors, such as generation of individual or repetitive action potentials, may display a great variety of response characteristics when a broad range of parameters is considered. This means that a given cell or ensemble may behave in many different modes, for example, as a generator of single pulses, as a bursting pacemaker, as a bistable "plateauing" cell, or as a beating oscillator, depending upon the physiological conditions (neuromodulator or ionic concentrations) or stimulus presentations (applied currents or synaptic inputs) (see 1). The nonlinear nature of the models provides the substrate for this broad repertoire.

In this chapter, we show that a simple, biophysically reasonable, two-current excitable membrane model is sufficiently robust to exhibit such behavioral richness, when parameters are systematically varied. The underlying qualitative structure for these behaviors will be revealed with graphical phase plane analysis, complemented by a few analytic formulas. The concepts we will cover include steady states, trajectories, limit cycles, stability, domains of attraction, and bifurcation of solutions. Phase plane characteristics and system dynamics will be interpreted biophysically in terms of activation curves, current-voltage relations, and the like. The concepts apply to higher-order systems, for which appropriate projections of phase space, motivated by differences in time scales for certain variables, can lead to similar insights.

1.2 Models for Excitable Cells and Networks

Most models for excitable membrane retain the general Hodgkin-Huxley (HH) format (Hodgkin and Huxley 1952), and can be written in the form

$$C\frac{dV}{dt} \; + \; I_{ion}(V, W_1, ..W_n) \; = \; I(t) \tag{1}$$

$$\frac{dW_i}{dt} \; = \; \phi\frac{W_{i,\infty} - W_i}{\tau_i(V)} \tag{2}$$

where V denotes the membrane potential (say, deviation from a reference, or "rest" level), C is the membrane capacity, and I_{ion} is the sum of V- and t-dependent currents through the various ionic channel types. I(t) is the applied current. The $W_i(t)$ variables describe the fraction of channels of a given type that are in various conducting states (e.g., open or closed) at time t. The first-order kinetics for W_i typically involve V dependence in the time constant τ_i; ϕ is a temperature-like time scale factor that may depend on i. If the current, I_j, for channel type j may be suitably modeled as ohmic, then it might be expressed as

$$I_j \;=\; \bar{g}_j \sigma_j(V, W_1, ..., W_n)(V - V_j) \tag{3}$$

where $\bar{g}_j$ is the total conductance with all j-type channels open (product of single-channel conductance with the total number of j channels), σ_j is the fraction of j channels that are open (it may depend on several of the W_i variables), and V_j is the reversal potential (usually Nernst potential) for this ion. For some channel types the current-voltage relation may be more appropriately represented by the Goldman-Hodgkin-Katz equation and the gating kinetics might involve a multistate Markov description. In the classical HH model for the squid giant axon, there are three variables W_i, denoted as m, h, and n, to describe the fractions m^3h and n^4 of open Na^+ channels and K^+ channels, respectively.

1.3 The Fitzhugh-Nagumo model

In the Hodgkin-Huxley (HH)-model the membrane potential V(t) and the sodium activation m(t) evolve on similar (fast) time scales during an action potential, while sodium inactivation h(t) and potassium activation n(t) change on a slower time scale. Given the great similarity of V(t) and m(t), it makes sense to simplify the model in terms of the number of parameters by lumping V(t) and m(t) into a single "activation" variable V. By the same argument, we can combine the parameters n(t) and 1-h(t) into a new single variable W, characterizing the degree of "accomodation" or "refractoriness" of the system. This provides the basis for the Fitzhugh-Nagumo (FN) model, which has only 2 parameters.

The equations underlying the FN model have their origin in the work by van der Pol, who formulated a nonlinear oscillator model to describe the cardiac pacemaker dynamics. The van der Pol oscillator is defined by the following set of differential equations:

$$\frac{d^2x}{dt^2} \;+\; c(x^2 - 1)\frac{dx}{dt} \;+\; x \;=\; 0 \tag{4}$$

This equation can be cast into the form

$$\frac{dx}{dt} \;=\; c\left(x + y - \frac{x^3}{3}\right)$$

$$\frac{dy}{dt} \;=\; -\;\frac{x}{c}$$

where $y = \frac{1}{c}\frac{dx}{dt} \;+\; \frac{x^3}{3} \;-\; x$.

Independently, Fitzhugh and Nagumo derived the following equations for the 2-parameter membrane dynamics for an excitable neuron:

$$\dot{V} \;=\; V \;-\; \frac{V^3}{3} \;-\; W \;+\; I \tag{5}$$

$$\dot{W} \;=\; \phi(V + a - bW) \tag{6}$$

where ϕ is the inverse of the time constant, which determines how fast the variable W changes relative to V. Typical values for the constants are ϕ=0.08, a=0.7, and b=0.8.

Due to the nonlinearities, closed form analytical solutions for equations 5 and 6 cannot be obtained. The only alternatives are numerical computer simulations, and local linearization. The latter is only useful in regions, where linearization makes sense, which is near the stable states of the system. A useful way to deduce the topological properties of the dynamic behavior is to define a vector $\mathbf{r}(t) = (V(t), W(t))^T$.

In order to understand how the system evolves in time, we will consider the so-called isoclines. An isocline is a curve in the (V,W) plane along which one of the ($\dot{V}$ or $\dot{W}$) is zero. The null-cline associated with the fast variable V, is the cubic function $W = V - V^3/3 + I$ (see figure 2). If the system is located on the V nullcline, its imminent future trajectory must be vertical, pointing either upward (for $\dot{W} > 0$) or downward (for $\dot{W} < 0$). Furthermore, for all points in the plane above this cubic polynomial $\dot{V} < 0$, with the converse for the points below the cubic polynomial. The nullcline associated with the slow variable W is specified by the linear equation $W = (V + a)/b$. Thus, if the evolution of the system brings it onto the W nullcline, its trajectory in the immediate future must be horizontal, for only V, not W, can change.

1.4 Stability of the equilibrium points.

The critical or singular points $\mathbf{r}^*$ of the system are the points (V_i, W_i) at which both derivatives are zero (i.e. : $(\dot{V}(V_i, W_i), \dot{W}(V_i, W_i)$=0). These singular points can be stable or unstable. In the absence of noise, the system would stay in a singular point forever. However, for a stable point, any perturbation by noise will bring the system back to the singular point, whereas for an unstable point, any noise will bring the system out of the neighbourhood of the singular point and will move it away from the singular point.

The stability of singular points can be evaluated by linearizing the system around the singular point. The linearization procedure corresponds to moving the origin of the system to the singular point and considering the fate of points in the immediate neigbourhood of the singular point. We can write for any perturbation $\delta\mathbf{r}$ around the fixed point $\mathbf{r}^*$

$$\dot{\bar{V}} + \delta\dot{V} = (\bar{V} + \delta V) - (\bar{V} + \delta V)^3/3 - (\bar{W} + \delta W) + I$$

$$\dot{\bar{W}} + \delta\dot{W} = \phi((\bar{V} + \delta V) + a - b(\bar{W} + \delta W))$$

Remembering that $(\bar{V} - \bar{V}^3/3 - \bar{W} + I) = 0$ and $\phi(\bar{V} + a - b\bar{W}) = 0$ and that $\dot{\bar{V}} = \dot{\bar{W}} = 0$ and neglecting higher order terms in δV and δW, we arrive at

$$\delta\dot{V} = (1 - \bar{V}^2)\delta V - \delta W$$

$$\delta\dot{W} = \phi(\delta V - b\delta W)$$

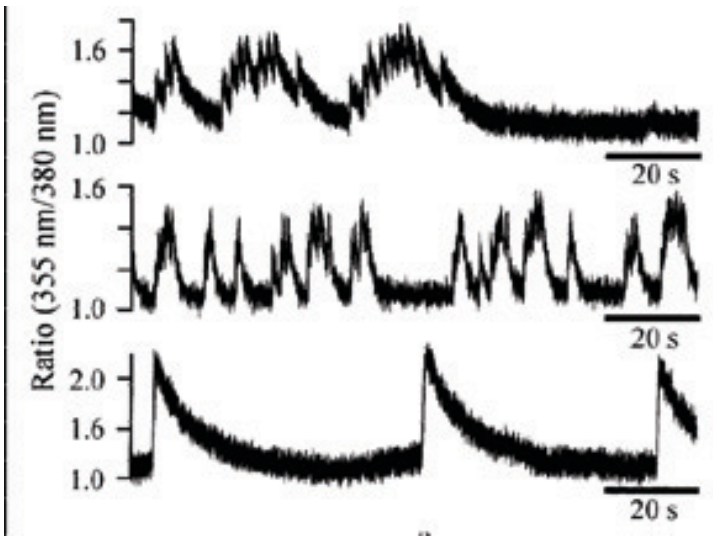


Figure 1: Various types of oscillations of intracellular $[Ca^{2+}]$ concentration in 3 different cells of Xaenopus Laevis (klauwpad): Deterministic, repetitive oscillations (bottom panel), chaotic oscillations (middle panel), and bursting behavior (top panel).

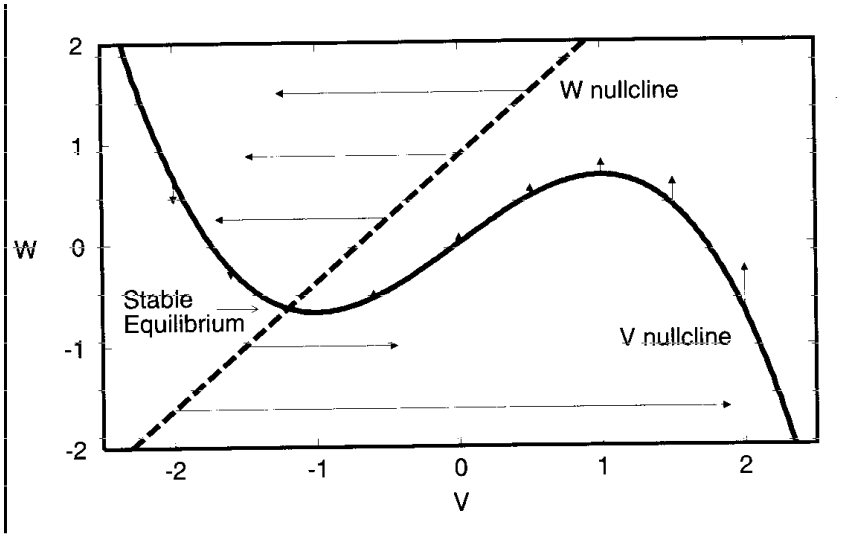


Figure 2: Phase plane associated with the FN model. The fast variable V corresponds to the membrane excitability, while the slower variable W can be visualized as the state of the membrane accomodation. The arrows are proportional to $(\dot{V}, \dot{W})$ and indicate the direction and rate of change of the system.

or in matrix notation

$$\delta\dot{\mathbf{r}} \;=\; M\;\delta\mathbf{r} \tag{7}$$

with the matrix M given by

$$M = \begin{pmatrix} (1-\bar{V}^2) & -1 \\ \phi & -b\phi \end{pmatrix} \tag{8}$$

We can characterize the behavior around the singular point by finding the eigenvalues and corresponding eigenvalues of M. The associated characteristic equation is

$$\lambda^2 + (\bar{V}^2 - 1 + b\phi)\lambda + (\bar{V}^2 - 1)b\phi + \phi = 0$$

The eigenvalues are given by

$$\lambda_{1,2} \;=\; 1/2[-(\bar{V}^2 - 1 + b\phi) \pm \sqrt{(\bar{V}^2 - 1 - b\phi)^2 - 4\phi}]$$

The evolution of the system takes the following form

$$\delta\mathbf{r} \;=\; c_1\mathbf{r}_1 e^{\lambda_1 t} \;+\; c_2\mathbf{r}_2 e^{\lambda_2 t} \tag{9}$$

with $\mathbf{r}_1$ and $\mathbf{r}_2$ the two eigenvectors. Obviously, if both eigenvalues are real and negative, the system is stable at the singular point and it is often called a sink. If one of the eigenvalues is greater than zero, the system is unstable. If both eigenvalues are real and positive, the singular point is called a source. If the two eigenvalues have opposite signs, the singular point is called a saddle point.

In this spirit, FitzHugh (1960) considered reductions of the HH and then introduced and idealized, an analytically tractable two-variable model widely studied as a qualitative prototype for excitable systems in many biological and chemical contexts. A FitzHugh-Nagumo/Hodgkin-Huxley hybrid was formulated and studied by Morris and Lecar (1981). The model incorporates a Voltage-gated Ca^{2+} channel and a Voltage-gated, delayed-rectifier K^+ channel; neither current inactivates. A simple version of this model is represented by the equations

$$C\frac{dV}{dt} \;=\; -\,I_{ion}(V,w) \;+\; I \tag{10}$$

$$\frac{dw}{dt} \;=\; \phi\frac{w_\infty(V) \;-\; w}{\tau_w(V)} \tag{11}$$

where

$$I_{ion}(V,w) \;= \bar{g}_{Ca} m_\infty(V)(V - V_{Ca}) + \;\bar{g}_K w(V - V_K) \;+\; \bar{g}_L(V - V_L). \tag{12}$$

In eqs. 10-12, w is the fraction of K^+ channels open, and the Ca^{2+} channels respond to V so rapidly that we assume instantaneous activation. One might introduce dimensionless variables in order (1) to reduce the number of free parameters and identify equivalent groups of parameters, and (2) identify and group "fast" and "slow" processes together. However, in the interest of clarity, we will keep all equations in their original form. In eq. 11, τ_w has been scaled so its maximum is now one, and ϕ equals the temperature factor divided by the prescaled maximum.

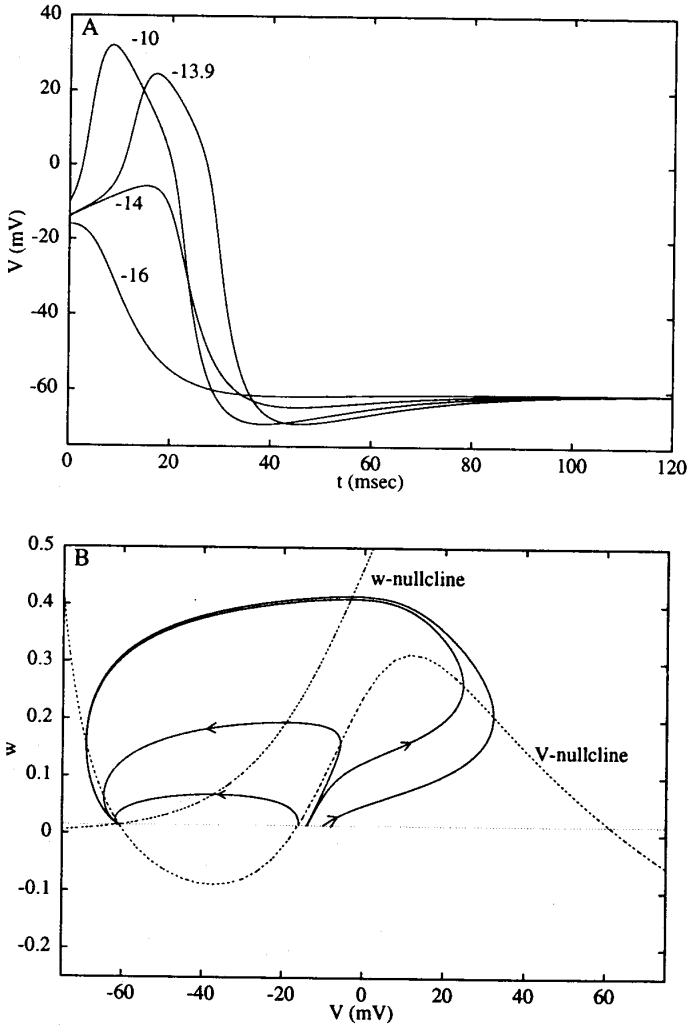


Figure 7.1
Response of the Morris-Lecar excitable system, eqs. 7.4–7.6, to a brief current pulse. For these parameters (see appendix A), the system has a unique stable rest state, $\bar{V} = -61\,\text{mV}$, $\bar{w} = .015$. The line $w = \bar{w}$ is shown lightly dashed. Four different stimuli lead to an instantaneous displacement of V from $\bar{V}$ to V_0 (values of V_0 are shown alongside the curves in panel A). Panel A shows the time course of the voltage. Notice that intermediate responses are possible with some stimuli: the threshold is graded; firing occurs with finite latency. Panel B shows trajectories in the V-w phase plane; nullclines are shown dashed and intersect only once. The effect of a stimulus is to displace the initial condition horizontally from rest.

Figure 3:

1.5 The Geometry of Excitability

We begin by considering the Morris-Lecar model , in the case that there is a unique rest state and a threshold-like behavior for action potential generation. The Morris-Lecar model is defined by the following differential equations and V-dependent functions:

$$C\frac{dV}{dt} \;=\; -\,\bar{g}_{Ca}m_{\infty}(V)(V-V_{Ca}) \;-\; \bar{g}_{K}w(V-V_{K}) \;-\; \bar{g}_{L}(V-V_{L}) \;+\; I \tag{13}$$

$$\frac{dw}{dt} \;=\; \phi\frac{w_{\infty}(V)-w}{\tau_{w}(V)} \tag{14}$$

where

$$m_{\infty}(V) \;=\; 0.5*[1+tanh\{(V-V_{1})/V_{2}\}] \tag{15}$$

$$w_{\infty}(V) \;=\; 0.5*[1+tanh\{(V-V_{3})/V_{4}\}] \tag{16}$$

and

$$\tau_{w}(V) \;=\; 1/cosh\{(V-V_{3})/(2V_{4})\} \tag{17}$$

For Figs. 3-5 we used the parameters $V_1 = -1.2$ mV, $V_2 = 18$ mV, $V_3 = 2$ mV, $V_4 = 30$ mV, $\bar{g}_{Ca} = 4.4\ mS/cm^2$, $\bar{g}_K = 8.0\ mS/cm^2$, $\bar{g}_L = 2\ mS/cm^2$, $V_K = -84$ mV, $V_L = -60$ mV, $V_{Ca} = 120$ mV, C=20 $\mu F/cm^2$, and The same parameters are used for figures 6-8 with the exceptions $V_3 = 12$ mV, $V_4 = 17.4$ mV, $\bar{g}_{Ca}$=4.0 mS/cm^2, and $\phi = 1/5$. In Figs. 9- 10 the parameters are as in Figs. 6-8 but $\phi = 0.23$. The current I in $\mu A/cm^2$ is generally the only free parameter.

Figure 3A shows the V responses to brief current pulses of different amplitudes. The peak V is graded, but the variation occurs over a very narrow range of stimuli; in this case, as in the standard HH model, the threshold phenomenon is not discrete, but rather, steeply graded. In figure 3B, these same responses are represented in the V-w plane. The solution path in the space of dependent variables is called a "trajectory," and direction of motion along a trajectory is often indicated by an arrowhead. In figure 3B, the flow is generally counterclockwise. All the trajectories shown here ultimately lead to the rest point: $V = \bar{V}, w = \bar{w} = w_{\infty}(\bar{V})$. The rest state is said to be "globally attracting." Each trajectory has a unique initial point, a horizontal displacement from the rest point corresponding to instantaneous depolarization by a brief current pulse. A trajectory's slope conveys the relative speed of w to V; thus a shallow slope means V is changing faster. The trajectory of an action potential shows the following features: an upstroke with rapid increase in V (trajectory is moving rightward with little vertical component) and then the transient depolarized plateau with the delayed major increase in w, corresponding to the slower opening of K+ channels. When w is large enough, the abrupt downstroke in V occurs-the trajectory moves leftward, nearly horizontal, as V tends toward V_K. Finally, as w decreases (the potassium channels close), the state point returns to rest with a slow recovery from hyperpolarization.

In the phase plane, the slope of a trajectory at a given point is dw/dV, which is simply the ratio of dw/dt to dV/dt, and these quantities are evaluated from the right-hand sides of the differential equations (eqs. 10-11). Thus a trajectory must be vertical or horizontal where dV/dt = 0 or dw/dt = 0, respectively. The conditions

$$0 \;=\; -\,\bar{g}_{Ca} m_\infty(V)(V - V_{Ca}) \;-\; \bar{g}_K w(V - V_K) \;-\; \bar{g}_L(V - V_L) \;+\; I \tag{18}$$

$$0 \;=\; \phi \frac{w_\infty(V) \;-\; w}{\tau_w(V)} \tag{19}$$

define curves, the V and w nullclines, which are shown dashed in figure 3B. This provides a geometrical realization for where V and w can reach their maximum and minimum values along a trajectory in the V-w plane (notice how the trajectories cross the nullclines either vertically or horizontally in figure 3B). The w nullcline is simply the w activation curve, $w \;=\; w_\infty(V)$. The V nullcline, from eq. 18, corresponds to V and w values at which the instantaneous ionic current plus applied current is zero; below the V nullcline, V is increasing and above it, V is decreasing. The cubic-like shape seen here reflects the N-shaped instantaneous I- V relation, $I_{ion}(V, w)$ versus V with w fixed (Eq. 12), typical of excitable membrane models in which the V-gated channels carrying inward current activate rapidly. From another viewpoint, motivated by the slower time scale of w, suppose we fix w, say, at a moderate value. Then the three points on the V nullcline at this w correspond to three pseudo-steady states; at the low- V state, small outward and inward currents cancel while at the high V state, both currents are larger but are again in balance. These states are transiently visited during the plateau phase and the return-to-rest phase of an action potential. Notice how the trajectory is near the right and left branches of the V nullcline during these phases.

If ϕ were smaller still, then the phase plane trajectories (except when near the V nullcline) would be nearly horizontal (because dw/dV would be small); the action potential trajectory during the plateau and recovery phases would essentially cling to, and move slowly along, either the right or left branch of the V nullcline. The downstroke would occur at the knee of the V nullcline. Also, in the case of smaller ϕ, the threshold phenomenon would be extremely steep; the middle branch of the V nullcline would act as an approximate separatrix between sub- and superthreshold initial conditions. In contrast, for larger ϕ, the response amplitude is more graded.

1.6 Oscillations Emerging with Nonzero Frequency

In the phase plane treatment, the rest state of the model is realized as the intersection of the two nullclines; such steady-state solutions are also referred to as **singular** or **equilibrium** points. From the geometrical viewpoint, one sees how different parameter values could easily lead to multiple singular points by changing the shapes and positions of the nullclines. In figure 3, the unique singular point is attracting. Technically, we say it is asymptotically stable, that is, for any nearby initial point the solution tends to the singular point as $t \to \infty$. In general, the local stability of a singular point can be determined by a simple algebraic criterion. The procedure is to linearize the differential equations near the singular point, evaluate the partial derivatives at the singular point (this matrix of partial derivatives is called the Jacobian), and to determine whether the eigenvalues of the Jacobian are positive or negative. If they are positive, the singular point is unstable; if all eigenvalues are negative, it is stable. For eqs. 10-12, the linearized equations that describe the behavior of small disturbances, $V \approx \bar{V} + x$, $w \approx \bar{w} + y$, from the singular

point are

$$\frac{dx}{dt} = ax + by \tag{20}$$

$$\frac{dy}{dt} = cx + dy \tag{21}$$

where

$$a = -\frac{\partial I_{ion}(V,w)}{\partial V}\frac{1}{C} \tag{22}$$

$$b = -\frac{\partial I_{ion}(V,w)}{\partial w}\frac{1}{C} \tag{23}$$

$$c = \frac{\phi}{\tau_w}\frac{dw_\infty}{dV} \tag{24}$$

$$d = -\frac{\phi}{\tau_w} \tag{25}$$

Solutions are of the form $exp(\lambda_1 t)$, $exp(\lambda_2 t)$, where $\lambda_{1,2}$ are the eigenvalues of the Jacobian matrix in eqs. 20-21. They are roots of the quadratic

$$\lambda^2 - (a+d)\lambda + (ad - bc) = 0 \tag{26}$$

For the parameters of figure 3, the two eigenvalues are both real and negative. As parameters are varied, the singular point may lose stability. In our example, the rest state could then no longer be maintained and the behavior of the system would change. It may fire repetitively or tend to a different steady state (if a stable one exists). Let us consider the effect of a steady applied current and ask how repetitive firing arises in this model. We will apply linear stability theory to find values of I for which the steady state is unstable. First, we note that for eqs. 10-12 a steady-state solution $\bar{V}$ for a given I must satisfy $I = I_{ss}(\bar{V})$, where $I_{ss}(V)$ is the steady-state I-V relation of the model given by

$$I_{ss} = I_{ion}(V, w_\infty(V)) \tag{27}$$

If I_{ss} is N-shaped, there will be three steady states for some range of I. If however, I_{ss} is monotonically increasing with V, as in the case of figure 3, then there is a unique $\bar{V}$ for each I. Moreover, $(\bar{V}, \bar{w})$ cannot lose stability by having a single real eigenvalue pass through zero. Destabilization can only occur by a complex conjugate pair of eigenvalues crossing the axis $Re\lambda = 0$ as I is varied through a critical value I_1. At such a transition, a periodic solution to eqs. 10-12 is born and we have the onset of repetitive activity. This solution, for I close to I_1, is of small amplitude and frequency proportional to $Im\lambda$. Emergence of a periodic solution in this way is called a Hopf bifurcation. From eqs. 20-21, or eq. 26, we know that $\lambda_1 + \lambda_2 = a + d$. Thus loss of stability occurs for the I whose corresponding V satisfies

$$\frac{1}{C}\frac{\partial I_{ion}(V,w)}{\partial V} + \frac{\phi}{\tau_w} = 0 \tag{28}$$

The first term here is the slope of the instantaneous I-V relation and the second is the rate of the recovery process; this condition also applies approximately to the HH model. From eq. 28 we conclude that loss of stability occurs:

1. only if the instantaneous I-V relation has negative slope at V;

2. when the destabilizing growth rate of V from this negative resistance just balances the recovery rate; and

3. only if recovery is sufficiently slow, i.e. if ϕ is small (low "temperature").

In figure 4A, V is plotted versus I (this is the steady-state I-V relation, but shown as V against I) and the region of instability is shown dashed. Figure 4A also shows the maximum and minimum values of V for the oscillatory response. Just as a singular point can be unstable, so, too, can a periodic solution; unstable periodics are indicated by open circles. Here we see that the small amplitude periodic solution born at $I = I_1 = 93.85\mu A/cm^2$ from the loss in stability of $\bar{V}$ is itself unstable; it would not be directly observable. Note that solutions along this branch depend continuously on parameters and they gain stability at the turning point or knee at $I = I_\nu = 88.3\mu A/cm^2$. A stable periodic solution is called a "limit cycle." The upper branch (solid) corresponds to the limit cycle of observed repetitive firing. The frequency increases with I over most of this branch (figure 4B). At sufficiently large I, repetitive firing ceases (depolarization block) as $\bar{V}$ regains stability at $I = I_2 = 212\mu A/cm^2$. This figure is referred to as a "bifurcation diagram"; it depicts steady-state and periodic solutions, and their stability, as functions of a parameter and it shows where one branch bifurcates from another. Bifurcation theory allows one to characterize the solution behavior analytically in the neighborhood of bifurcation points; for example, the frequency of the emergent oscillation at the Hopf point is proportional to $|Im\ \lambda_{1,2}|$. When the Hopf bifurcation leads to unstable periodic solutions, i.e., when the emergent branch bends back into the parameter region where the steady state is stable, then the bifurcation is subcritical (i.e., a hard oscillation); if the opposite occurs, it is supercritical.

Intermezzo about Hopf bifurcations

A simple Hopf bifurcation generates a limit cycle starting from a fixed point. For example, consider the following differential equation in polar coordinates:

$$\frac{dr}{dt} = -(\Gamma r + r^3) \quad ; \quad \Gamma = a - a_c$$

$$\frac{d\theta}{dt} = \omega$$

Their solutions are

$$r^2(t) = \frac{\Gamma r_0^2 e^{-2\Gamma t}}{r_0^2(1 - e^{-2\Gamma t}) + \Gamma}$$

with $r_0 = r(t = 0)$ and $\theta(t) = \omega t$ with $\theta(t = 0) = 0$. For $\Gamma \geq 0$ the trajectory approaches the origin (fixed point), whereas for $\Gamma < 0$ it spirals towards a limit cycle with radius $r_\infty = |(a - a_c)|^{1/2}$. Transformation of the differential equation in polar coordinates into cartesian coordinates gives

$$\frac{dx}{dt} = -[\Gamma + (x^2 + y^2)]x - y\omega$$

$$\frac{dy}{dt} = -[\Gamma + (x^2 + y^2)]y - + x\omega$$

Linearization about the origin gives

$$\frac{d\mathbf{f}}{dt} = A\mathbf{f}$$

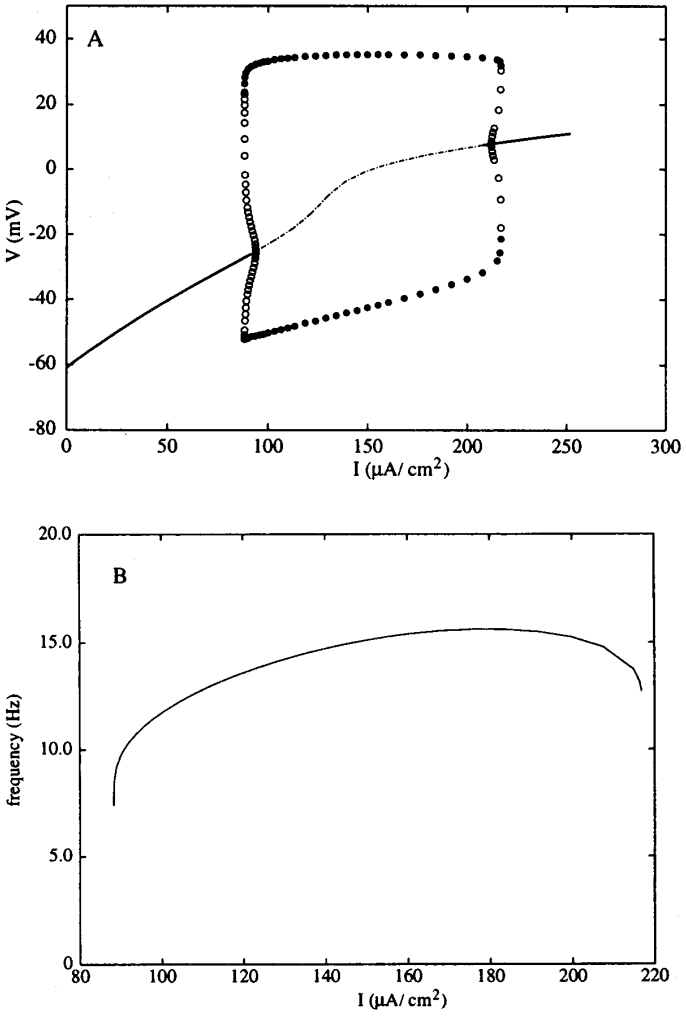


Figure 7.2
Repetitive firing in the Morris-Lecar model for steady current. Bifurcation diagram in panel A shows the steady-state voltage $\bar{V}$ versus I (thin lines; stable are solid, unstable are dashed) and the maximum and minimum voltage for periodic solutions shown as filled (stable) and unfilled (unstable) circles. The unstable branch of periodic solutions meets the branch of steady-state oscillations at $I = I_1 = 94\,\mu A/cm^2$ and $I = I_2 = 212\,\mu A/cm^2$ (Hopf bifurcation points). The unstable branch of periodic solutions coalesces with the stable branch of periodic solutions at $I = I_v = 88\,\mu A/cm^2$. A similar coalescence occurs near $I = 215\,\mu A/cm^2$. For these parameters, the steady-state I-V curve is monotonic. Furthermore, panel B shows that the frequency (plotted in Hz, and only for the stable limit cycles) as a function of current is always bounded away from zero. Parameters are as figure 7.1.

Figure 4:

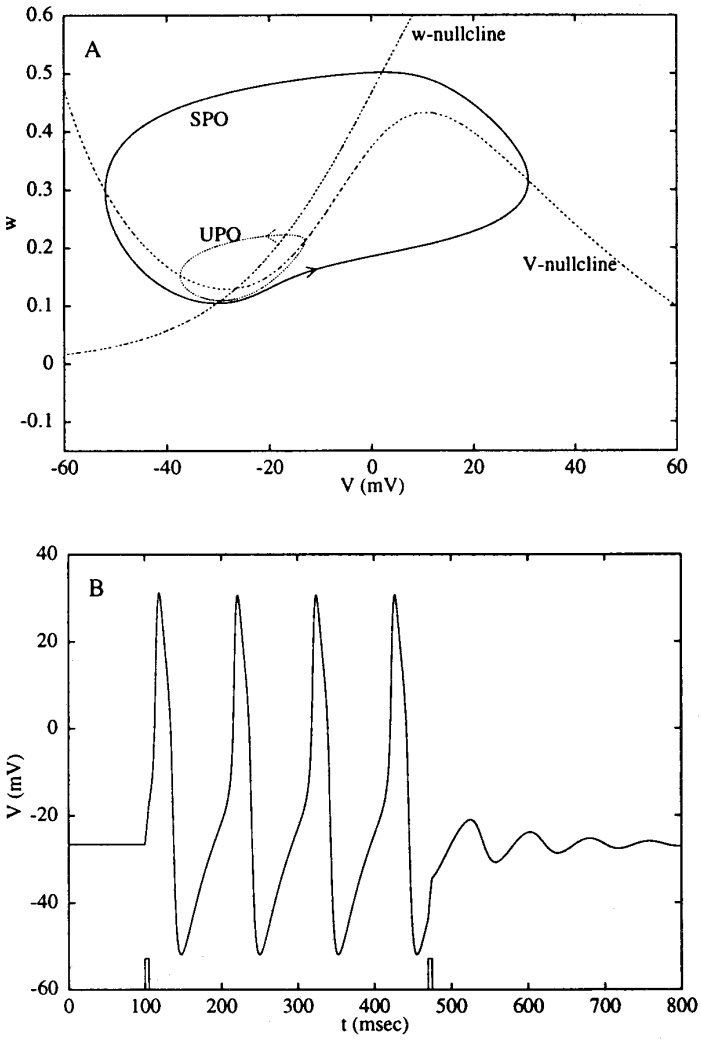


Figure 7.3
Bistability for steady current near the threshold for repetitive firing for the Morris-Lecar model with parameters as in figure 7.1 and $I = 90\,\mu A/cm^2$. In this region, where I is between the first Hopf bifurcation point, I_1, and the "knee," I_v, there are two stable states (cf figure 7.2): a rest state (the intersection of the nullclines) and a stable oscillation (SPO) separated by an unstable periodic solution (UPO). This is shown in panel A. Panel B demonstrate switching from rest to oscillation and then back to rest for two brief appropriately timed depolarizing current pulses.

Figure 5:

where $\mathbf{f} = (\Delta x, \Delta y)$ and A is the matrix

$$A = \begin{pmatrix} -\Gamma & -\omega \\ \omega & -\Gamma \end{pmatrix} \tag{29}$$

with eigenvalues $\lambda_{\pm} = -\Gamma \pm i\omega$. This means that at a Hopf bifurcation a pair of conjugate eigenvalues crosses the imaginary axis. These eigenvalues indicate, in another way, that the origin is a stable attractor for $\Gamma > 0$. If $\Gamma = 0$ the origin is still a stable attractor (verify !). However, when $\Gamma < 0$, the real part of the eigenvalue becomes positive at the origin, making the origin an unstable attractor. For $\Gamma < 0$, the system starts approaches oscillations with a radius $\sqrt{-\Gamma}$.

For a range of I values (between the knee, I_ν and the Hopf bifurcation, I_1), our model exhibits bistability: a stable steady state and a stable oscillation coexist. Figure 5A illustrates the phase plane profile in such a case; a periodic response here appears as a closed orbit. There is a stable fixed point shown as the intersection of the two nullclines and a stable periodic orbit (labeled SPO). The two attractors are separated by an unstable periodic orbit (UPO). Initial values inside the unstable orbit tend to the attracting steady state, while initial conditions outside of it will lead to the limit cycle of repetitive firing. A brief current pulse, whose phase and amplitude are in an appropriate range, can switch the system out of the oscillatory response back to the rest state. In figure 5B, two 30 $\mu A/cm^2$ current pulses with 5 ms duration are given, at t = 100 ms and then at t = 470 ms. The first pulse switches the membrane from rest to repetitive firing, while the second pushes the membrane back to rest. This bistable behavior is critical for the occurence of bursting oscillations when a very slow conductance is added to the model.

1.7 Oscillations Emerging with Zero Frequency

The Hopf bifurcation is one of a few generic mechanisms for the onset of oscillations in nonlinear differential equation models. In that case, the frequency at onset of repetitive activity has a well-defined, nonzero minimum. In contrast, some membranes and models exhibit zero (i.e., arbitrarily low) frequency as they enter the oscillatory regime of behavior. A basic feature in such systems is that I_{ss} versus V is N-shaped rather than monotonic, as in the previous section. For eqs. 10-12, this occurs if the V dependence of K^+ activation is translated rightward, so that the inward component of I_{ss} dominates over an intermediate V range. Thus, for some values of I, below the repetitive firing range, there are three singular points in the phase plane and the system is excitable.

We discuss this case first. In figure 6B, we see the nullclines intersecting three times. As determined by linear stability theory, the singular points are the stable rest state (R), and unstable saddle point threshold (T), and an unstable spiral (U). The system is excitable, with the lower state being a globally attracting rest state: initial conditions near R lead to a prompt decay to rest, while larger stimuli lead to an action potential-a long trajectory about the phase plane. The phase plane portrait moreover reveals that this case of excitability indeed has a distinct threshold which is due to the presence of the saddle point, T. To understand this, we note that associated with the saddle are a unique pair of incoming trajectories (bold dashed lines) corresponding to the negative eigenvalue of the Jacobian matrix; together, these represent the stable manifold. Corresponding to the positive eigenvalue are a pair of trajectories (bold lines) that enter the saddle as $t \to \infty$; these are the unstable manifold. The stable manifold defines a separatrix curve

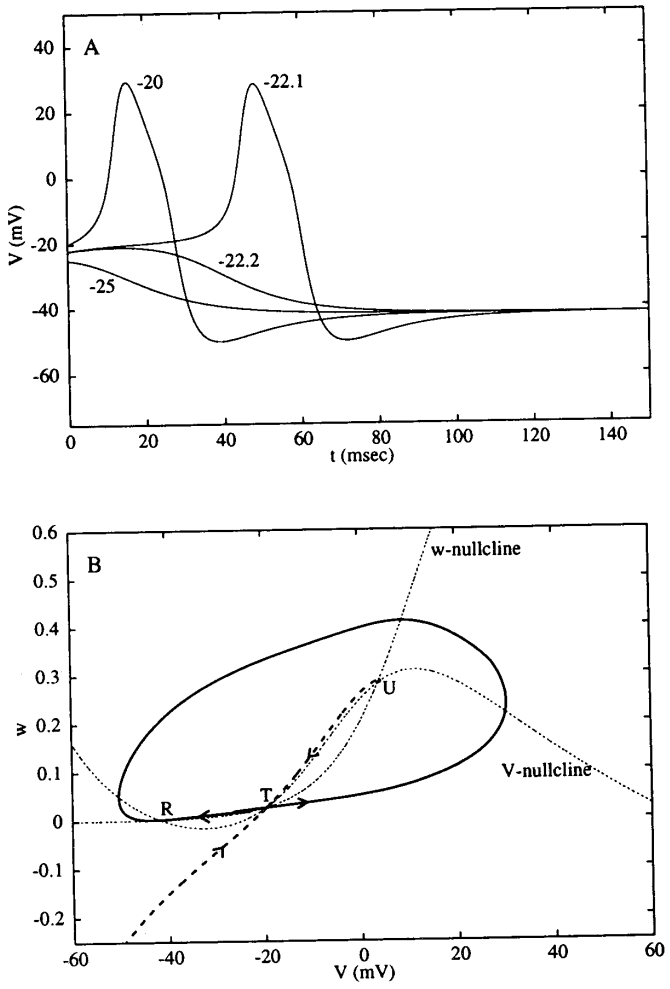


Figure 7.4
Excitability with three steady states and a distinct threshold; the response of the membrane to a brief current pulse from the stable rest state. Four different stimuli result in a displacement of V from $\bar{V}$ to V_0 (values of V_0 are given alongside the curves in panel A). (A) Time course of the voltage for $I = 30\,\mu A/cm^2$. (B) Phase plane for the dynamics illustrated in panel A. Nullclines intersect at three places: (1) R a stable rest state, (2) T, a saddle point threshold, and (3) U an unstable node. The thick solid line shows the unstable manifold for the saddle point; here, unstable refers to movement in opposing directions away from T (indicated by arrowheads). The manifold's two branches lead to the stable rest state and form a smooth loop in phase space. The heavy dashed line shows the stable manifold for the saddle point (arrowheads pointing toward T). Any initial conditions to the left of this manifold decay to rest. Initial conditions to the right lead to an action potential before returning to rest. Parameters are as in figure 7.1, except $\bar{g}_{Ca} = 4\,mS/cm^2$, $V_3 = 12\,mV$, $V_4 = 17.4\,mV$, $\phi = 1/15$.

Figure 6:

in the phase plane that sharply distinguishes sub- from superthreshold initial conditions. For initial conditions near the threshold separatrix, there is a long latency before a firing or decaying sub-threshold response (see figure 6A). This is because the trajectory starts close to (but not exactly on) the stable manifold and thus the solution comes very near the saddle singular point (where it moves very slowly) before taking off. If w is started at rest, WR, then there is a unique value of $V = V_T$ (between 22.1 and -22.2mV in the present example) called the "voltage threshold," where the stable manifold intersects the line $w = W_R$.

The action potential trajectory follows along the unstable manifold (bold lines), which passes around the unstable spiral and eventually tends to the rest point. Such a trajectory joining two singular points is called a "heteroclinic orbit." The other branch of the unstable manifold is also a heteroclinic orbit from the saddle to the rest point. This heteroclinic pair forces any trajectory that begins outside it to remain outside it-thus preserving the amplitude of the action potential. In this case we do not find graded responses for any brief current pulses from the rest state.

Next, we tune up I and ask when repetitive firing occurs. Because I_{ss} is N-shaped, we know that the lower and middle values of V move toward each other as I increases, and there is a critical value I_1 where they meet. In the phase plane, this means that the rest point and the saddle coalesce and then disappear; this is called a "saddle node bifurcation." Moreover, the heteroclinic pair becomes a single closed loop, a limit cycle, which for I just above I_1 has a very long period (figure 7). Thus, in this parameter regime, the transition to repetitive firing is marked by arbitrarily low frequency (figure 8). When $I = I_1$, the limit cycle has infinite period; it is called a "saddle node loop". Generally, an infinite period limit cycle is called a "homoclinic orbit," one that begins and ends at a singular point. The saddle node loop is one type of homoclinic orbit; we will encounter another type in the next section. This type of zero-frequency onset is generic and occurs over a range of parameters. Changing another parameter will typically lead to a smooth change in I_1. We emphasize that this mechanism allows arbitrarily low firing rates without relying on channel gating kinetics, which are necessarily slow. The value I_1 is determined by evaluating I_{SS} at the value of V for which $\partial I_{ss}/\partial V = 0$, and this latter condition is equivalent to having the determinant $ad - bc$ of the Jacobian matrix equal to zero.

The global picture of repetitive firing is shown in the bifurcation diagram of figure 8A, with frequency versus I in figure 8B. The branch of steady states (unstable shown dashed) form the S-shaped curve, and the oscillatory solutions are represented by the forked curve whose open end begins at $I = I_1$. As I increases beyond $I = I_1$ the peak-to-peak amplitude on the stable (repetitive firing) branch decreases and the frequency increases. The family of periodic solutions terminates at $I = I_2$ via a subcritical Hopf bifurcation. Except for I in a small interval of this upper range, this system is monostable. Annihilation of repetitive firing, as in figure 5, cannot be carried out for I near I_1 in this case (although at the high-current end where there is bistability, annihilation can occur).

1.8 More Bistability

It is important to realize that the solution behavior we have described in our bifurcation diagrams depends on other parameters in the model. The temperature parameter ϕ is particularly convenient, with useful interpretative value for additional parametric tuning: it plays no role in I_{ss} and thus does not affect the values along the S-shaped curve of

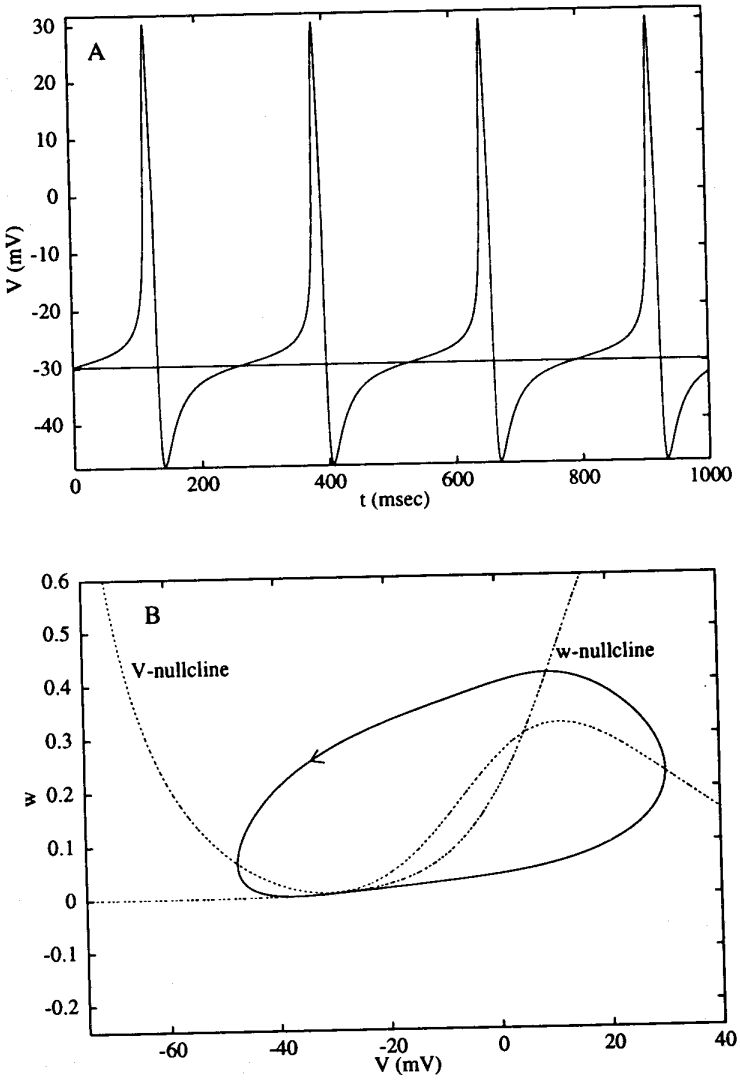


Figure 7.5
Onset of repetitive firing with arbitrarily low frequency for a constant current, $I = 40.76\,\mu A/cm^2$ shows an oscillation with a period of about 220 msec. Panel A shows the voltage time course and panel B shows the phase plane. Note the "narrow channel" between the two nullclines near $-30\,mV$, which accounts for most of the oscillation period (see Rinzel and Ermentrout 1989). Parameters are as in figure 7.4.

Figure 7:

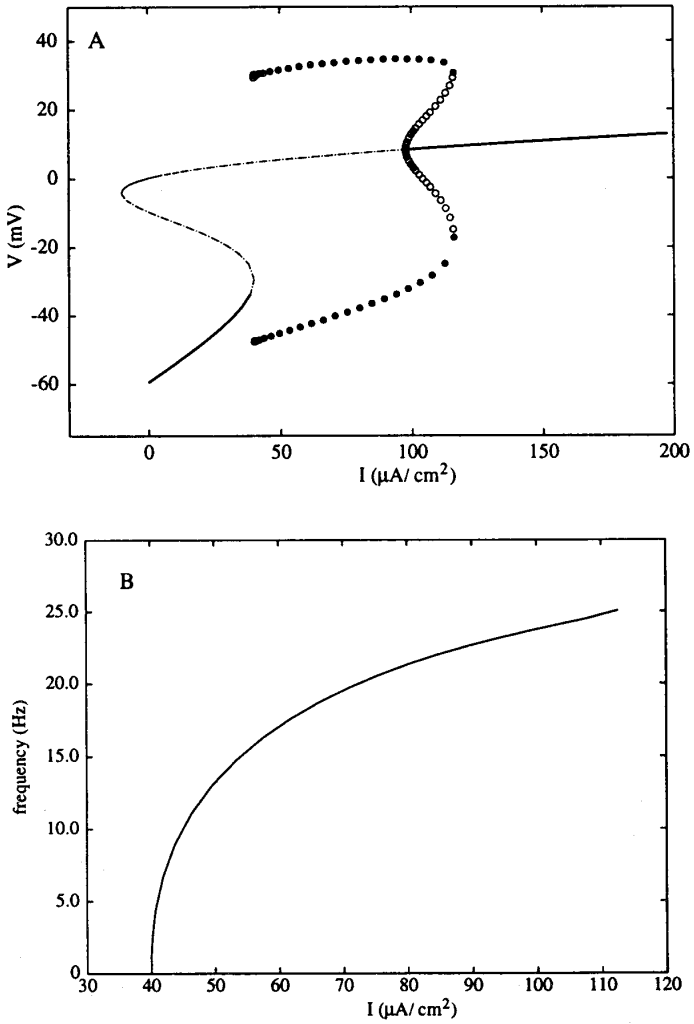


Figure 7.6
Multiple steady states and periodic orbits for a steady current when the I_{ss}–V relation is N-shaped. (A) Bifurcation diagram (line types as in figure 7.2A; parameters are as in figures 7.4–7.5). In spite of the coexistent states, the system is monostable for I between $I_1 = 40$, the turning point of the steady states, and $I_2 = 98$ where there is a Hopf bifurcation. Onset of repetitive firing at zero frequency occurs at $I = I_1$ where two fixed points coalesce. This corresponds to figure 7.4B when the unstable manifolds of the saddle point form a closed loop. The branch of periodic orbits has a turning point at $I = 116$ before terminating at the Hopf bifurcation point, $I = I_2$. All current values in $\mu A/cm^2$. (B) Frequency (in Hz) of stable branch of periodic orbits.

Figure 8:

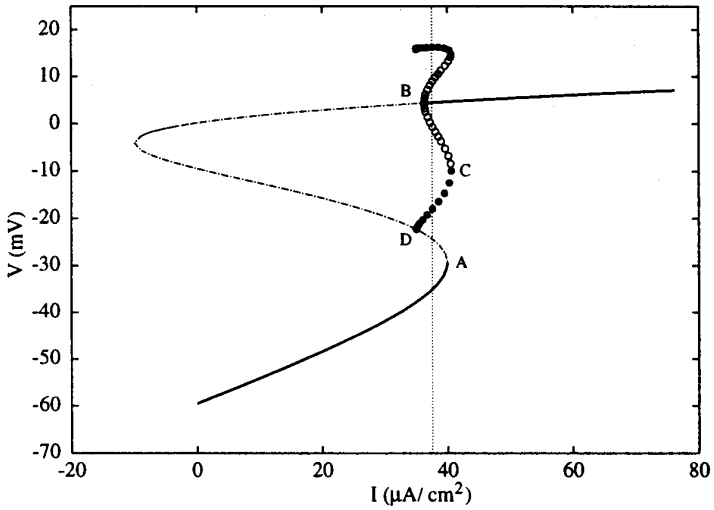


Figure 7.7
Bifurcation diagram (as in figure 7.6 but for $\phi = 0.23$). Point A shows where the two lower steady states coalesce, point B shows the Hopf bifurcation for the upper steady state, point C shows the coalescence of the stable and unstable periodic branches, and point D shows where the branch of stable oscillatory solutions terminates on the branch of saddle points (not on the knee, as in figure 7.6) at a saddle loop homoclinic. For currents between points B and A, there are three stable states: (1) a low-voltage rest state, (2) a high-voltage rest state, and (3) an oscillatory state. Note that the steady-state branch is identical to that of figure 7.6; ϕ only affects the stability of the steady states and the behavior of the periodic orbits. Vertical line at $I = 37.5\,\mu A/cm^2$ shows a current for which there are three stable states (cf. figure 7.8).

Figure 9:

steady states in figure 8, or the corresponding curve in figure 4. The stability of a steady state does, however, depend on ϕ. As is seen from eq. 28, when ϕ is large, oscillatory destablization is precluded; a Hopf bifurcation from a steady state only occurs when the time scale of w is slow compared to that of V. Thus, for large ϕ, both the upper and lower branches of the S-curve are stable; the middle branch is of course unstable. This system is bistable. In this large-ϕ limit, the kinetics of the K^+ system are so fast (essentially instantaneous, with $w = w_\infty(V)$ that the model reduces to one dynamic variable, V. Then stability is determined only by the slope of I_{ss} with the two "outer" states being stable and the "middle" unstable. This simple example also shows that sometimes a model can be conveniently reduced to a lower dimension when there are significant time scale differences between variables.

For intermediate values of ϕ, the dynamics of both V and w influence stability, and the upper branch is unstable for a certain range of I. Figure 9 shows a bifurcation diagram analogous to that in figure 8A, in which the branch of steady states is S-shaped and the stable rest state disappears at a turning point (point A). The high voltage equilibrium is stable for large currents but, as the current is reduced, loses stability at a subcritical Hopf bifurcation (point B). An unstable branch of periodic solutions emanates from the Hopf bifurcation point and then becomes stable at a turning point (C). Unlike figure 8A, however, this branch of stable periodic orbits (solid circles) does not terminate on the knee (point A) but instead on the unstable middle branch (point D on the diagram) as the current decreases to a critical value, 1D Again the frequency of the limit cycle tends to zero for this branch. At the critical value of current, 1D, the closed orbit has infinite period; it is called a "saddle loop homoclinic orbit." Recall that the middle branch of

solutions is a saddle point. For certain values of the current, this system has three stable states. If I is chosen to lie between the I values for points B and C, then the lower branch still exists and is stable, the upper branch of equilibria is stable, and there is a stable periodic orbit. Figure 10A shows the phase plane for this case. The stable manifold for the saddle point (bold dashed trajectory) acts to separate the stable periodic orbit from the lower rest state. The small unstable periodic orbit separates the upper rest state from the stable periodic solution. As in figure 5B, we can use brief current pulses to switch between states. Figure 10B shows the effect of three 5 ms current pulses switching from the periodic orbit to the lower rest state, back to the periodic orbit, and then to the upper rest state. (Note that perturbations from the upper rest state decay very slowly.) The HH model, adjusted for higher than normal external potassium, exhibits similar multistable behavior.

This example of coexistence between a depolarized limit cycle and a lower resting state is important because it also forms the basis for a general class of bursting phenomena.

1.9 Phase-Resetting and Phase-Locking of oscillators

We now turn our attention to a brief description of periodically forced and coupled neural oscillators. The behaviors generally involve issues that are difficult to analyze and we will only touch on them briefly. Before treating a specific example, it is useful to discuss certain important aspects of oscillators. We say that a periodic solution to an autonomous differential equation is (orbitally) "asymptotically stable" if perturbations from the oscillation return to the oscillation as $t \to \infty$. The difference between asymptotic stability of an oscillation and that of a steady-state solution is that, for the oscillation, the time course may exhibit a shift (see figure 11A) due to the time translation invariance of the periodic solution. Indeed, in phase space, the periodic trajectory is unchanged by translation in time. The shift that accompanies the perturbation of the limit cycle can be exploited in order to understand the behavior of the oscillator under external forcing. Suppose that an oscillator has a period, say T. We may let t = 0 correspond to the time of peak value of one of the oscillating variables, so that at t = T we are back to the peak. Given that we are on the periodic solution, if some I is specified, then we know precisely the state of each oscillating variable. This allows us to introduce the notion of phase of the periodic solution. Let $\theta = t/T$ define the phase of the periodic solution so that θ =0,1,2,... all define the same point on the periodic solution. For example, if $\theta = 8.5$, then we are halfway through the oscillator's nineth cycle.

1.9.1 Phase Response Curves

With the notion of phase defined, we now examine how a perturbation shifts the phase of the oscillator. In figure 11A, we show the voltage time course for the Morris-Lecar system in the oscillating regime. At a fixed time, say t, after the voltage peak, we apply a brief depolarizing current pulse. This shifts the time of the next peak (figure 11A) and this shift remains for all time (the solid curve is the perturbed oscillation and the dashed curve is the unperturbed - in this case the time for the next peak is shortened). If the time of the next peak is shortened from the natural time, we say that the stimulus has "advanced the phase"; if the time of the next peak is lengthened, we say that we have "delayed the phase." Let T_1 denote the time of the next peak. The phase shift is (T - T_1) / T, and T_1

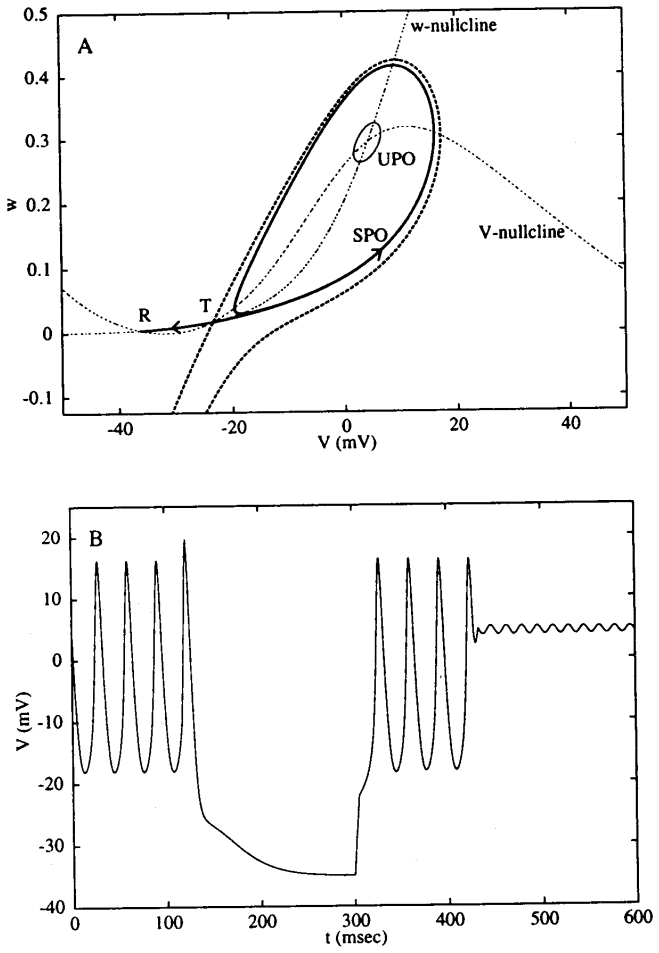


Figure 7.8
Multistability for a current between points B and A in figure 7.7 ($I = 37.5\,\mu A/cm^2$; other parameters are as in figure 7.7). Panel A depicts the V-w phase plane. The nullclines intersect at three places representing steady states: (1) a lower stable rest state (R), (2) an unstable saddle point (T), and (3) an upper stable rest state (unlabeled). The left branch of the unstable manifold of the saddle point (bold line) connects to the lower steady state. The right branch wraps around the stable periodic orbit (SPO). The branches of the stable manifold of the saddle point (bold dashed line) form a separatrix between the lower stable rest state and the stable periodic orbit. The unstable periodic orbit (UPO) separates the stable upper steady state from the stable periodic orbit. Panel B shows the effects of three successive depolarizing current pulses. Starting on the stable oscillation, the membrane is switched to the lower stable steady state. Another brief pulse pushes it back to the stable oscillation and a third pulse switches it to the upper steady state. No single brief current pulse can switch it from the lower steady state directly to the upper steady state, although the opposite transition is possible.

Figure 10:

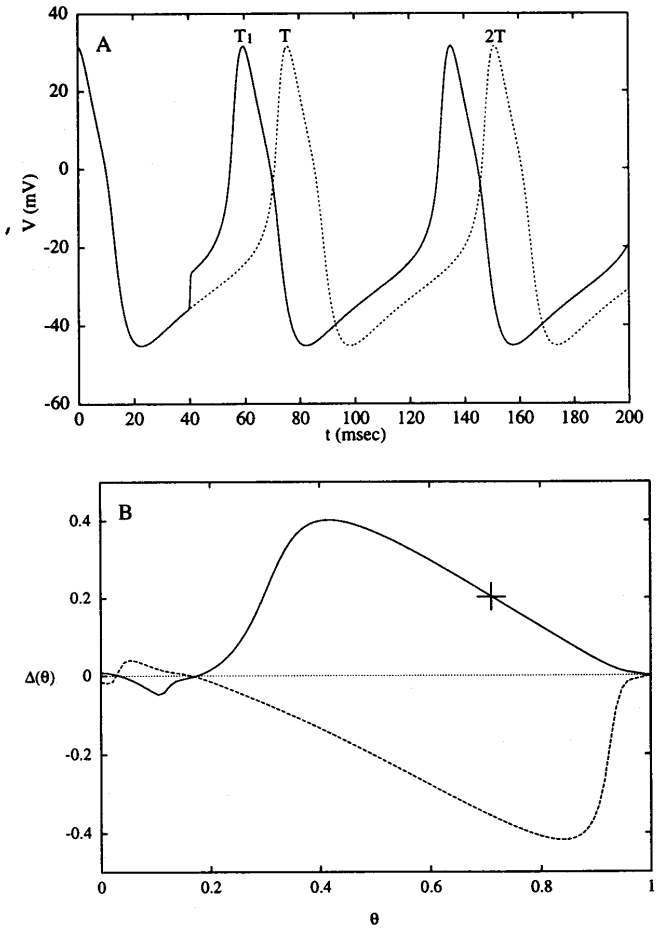


Figure 7.12
Phase-resetting of a Morris-Lecar oscillator. (A) A brief depolarizing stimulus can shorten the onset of the next spike and thus advance the phase. (B) Phase response curve (PRC) for the Morris-Lecar equations with parameters as in figures 7.4–7.6 and an applied current $I = 50\,\mu A/cm^2$. The stimulus is a 0.5 msec pulse with amplitude of $\pm 480\,\mu A/cm^2$ delivered at time $t = 40$ msec. The solid line shows the PRC for a depolarizing stimulus and the dashed for a hyperpolarizing pulse. The cross on the depolarizing PRC corresponds to the experiment in figure 7.13.

Figure 11:

depends on the time t or the phase $\theta = t/T$ at which the stimulus is applied. Thus we can define a phase shift $\Delta(\theta) \equiv (T - T_1(\theta))/T$. The graph of this function is called the "phase response curve" for the oscillator. If $\Delta(\theta)$ is positive, the perturbation advances the phase and the peak will occur sooner. On the other hand, if $\Delta(\theta)$ is negative, the phase is delayed and the next peak will occur later. We can easily compute this function numerically, and the same idea can be used to analyze an experimental system. Moreover, this curve can be used as a rough approximation of how the oscillator will be affected by repeated perturbation (periodic forcing) with the same current pulse. In figure 11B, we show a typical PRC for the Morris-Lecar model computed for both a depolarizing stimulus (solid line) and a hyperpolarizing stimulus (dashed line). The stimulus consists of a current pulse of magnitude 480 $\mu A/cm^2$ applied for 0.5 msec at different times after the voltage peak. The time of the next spike is determined, which yields the PRC, as above. The figure agrees with our intuition; if the depolarizing stimulus comes while V(t) is increasing (i.e., during the upstroke or slow depolarization of recovery), the peak will occur earlier and we will see a phase advance. If the stimulus occurs while V(t) is decreasing (i.e., during the downstroke), there will be a delay. The opposite occurs for hyperpolarizing stimuli. The curves show that it is difficult to delay the onset of an action potential with a depolarizing stimulus or advance it with a hyperpolarizing one.

We now show how this function can be used to analyze a periodically forced oscillator. Suppose that every P time units a current pulse is applied to the cell. Let θ_n denote the phase right before the time of the nth stimulus. This stimulus will either advance or delay the onset of the next peak depending on the phase at which the stimulus occurs. In any case, the new phase after time P and just before the next stimulus will be $\theta_n + \Delta(\theta_n) + P/T$. To understand this, first consider the case where there is no stimulus: after time P the oscillator will advance P/T in phase, but because the stimulus advances or delays the phase by an amount $\Delta(\theta_n)$, this amount is just added to the unperturbed phase, resulting in an equation for the new phase just before the next stimulus:

$$\theta_{n+1} = \theta_n + \Delta(\theta_n) + P/T. \tag{30}$$

This difference equation can be solved numerically. Here we consider the question of whether the periodic stimulus can entrain the voltage oscillation. That is, we ask whether there is a periodic solution to this forced neural oscillation. In general, a periodic solution is one for which there are M voltage spikes for N stimuli, where M and N are positive integers. When such a solution exists, we have what is known as "M : N phase-locking." Finding M:1 phase-locked solutions is quite easy. We require the oscillator to undergo M oscillations per stimulus period. In terms of eq. 30, this means we seek a solution that satisfies

$$\theta + M = \theta + \Delta(\theta) + P/T \tag{31}$$

for some value of θ. If such a solution exists and is stable (to be defined below), then, starting near θ, we can iterate eq. 30 and end up back at θ. This θ is the locking phase just before the next stimulus and because it does not change from stimulus to stimulus, the resulting solution must be periodic. Obviously, a necessary condition for a solution to eq. 31 is that M - P/T lie between the maximum and minimum of $\Delta(\theta)$, that is, we must solve

$$M - P/T = \Delta(\theta) \tag{32}$$

Having solved eq. 32, we need to determine the stability of the solution. For equations of the form of eq. 30, a necessary and sufficient condition for θ to be a stable solution

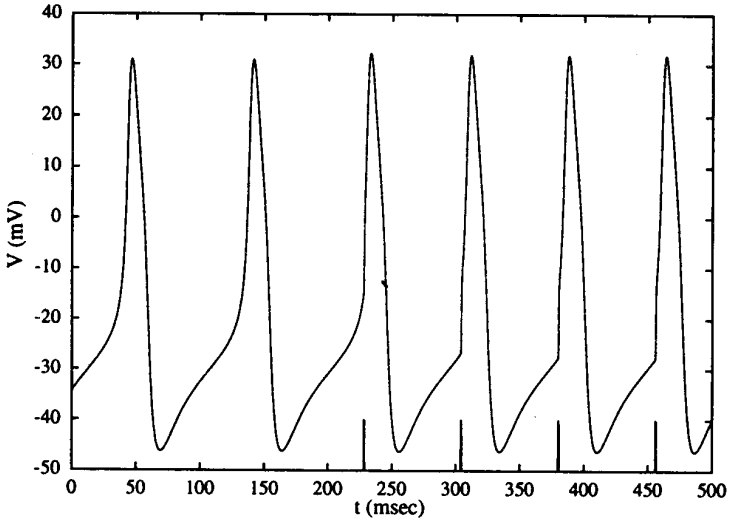


Figure 7.13
Phase-locking (1:1) of the Morris-Lecar model to a series of four current pulses with interpulse period of 76 msec. Intrinsic period of the membrane oscillator is 95 msec. Parameters are as in figure 7.12B. After phase-locking is achieved, the stimulus is seen to occur about 67 msec after the action potential's peak, just as predicted by the PRC.

Figure 12:

is that $-2 < \Delta'(\theta) < 0$. Because $\Delta(\theta)$ is periodic and continuous, there will in general be two solutions to eq. 32 (see figure 11B). But because only one of them will occur where $\Delta(\theta)$ has a negative slope, there will be a unique stable solution. We must also worry about whether the negative slope is too steep (i.e., more negative than -2); for small stimuli, this will never be the case - stability is assured. When $\Delta'(\theta) < -2$ (instability), very complex behavior can occur such as chaos. The case of M : N phase-locking where $N > 1$ is more difficult to explain and will not be considered here. It is clear that if the stimulus is weak, the magnitude of $\Delta(\theta)$ will also be small so that M - P/T must be small in order to achieve M:1 locking. On the other hand, if the stimulus is too strong, then we must be concerned with the stability of the locked solution. We note that, in a sense, eq. 30 is only valid for stimuli that are weak compared to the strength of attraction of the limit cycle; for stronger stimuli, it will take the solution more than a single oscillation to return to points close to the original cycle. The PRC in figure 11B shows that, when the stimulus is depolarizing, it is easier to advance the Morris-Lecar oscillator and thus force it at a higher frequency ($0 < P/T < 1$) than it is to force the oscillator at a lower frequency ($P/T > 1$). For hyperpolarizing stimuli, we can more easily drive the oscillator at frequencies below the natural frequency.

To illustrate these concepts, we have periodically stimulated the Morris-Lecar model (natural period of 95 msec) with the same brief depolarizing current pulse repeated every 76 msec. Figure 12 shows that the oscillation is quickly entrained to the new higher frequency. Equation 32 allows us to predict the time after the voltage peak that the stimulus will occur for 1:1 phase-locking. From the PRC we can see that $\Delta(\theta) = 1 - 76/95 = 0.2$ corresponds to two values of θ, one stable (cross in figure 11B) $\theta = 0.702$

and the other unstable. Thus the locking time after the voltage peak, that is, when the stimulus occurs, is predicted from the PRC to be $t = T\theta = 67$ ms. This is exactly the shift observed in figure 12.

The technique illustrated here is useful for analyzing the behavior of a single oscillator when forced with a short pulsatile stimulus. For more continuous types of forcing, such as an applied sinusoidal current, other techniques must be used. One such technique is the method of averaging, applicable when the forcing is weak. Periodic forcing is a special case of coupling, which we will now describe.

An example of forced synchronization: Dynamics of cardiac cells

In this section we will elaborate a detailed example of forced synchronization. This section is partly taken from Schuster (1988). Please note, that contrary to the previous section, which dealt with phase leads $\Delta(\theta)$ as positive and phase leads (delays) as negative, this section defines phase delays the other way around. As a consequence, the equation, which defines the time of the new phase θ_{n+1} has a negative sign for $\Delta(\theta)$ in this chapter, contrary to equation 30, which has a positive sign for the same term.

It has been found by M. R. Guevara, L. Glass, and A. Shrier (1981) that circle maps are also relevant for explaining the dynamics of cardiac cells. Fig. 13 shows the temporal behavior of the transmembrane electric potential from an aggregate of embryonic chick heart cells, which beat spontaneously. If the system is periodically stimulated via a current pulse through a microelectrode, the nature of the response depends on the interstimulus interval. The main idea is to reduce this response to a single stimulus by constructing an appropriate circle map.

Fig. 14 shows that the influence of a single pulse changes the period of the spontaneous beats from τ to T. The assumption is now that their ratio T/τ depends only on the phase shift $\theta = \delta/\tau$ of the stimulus with respect to the natural signal, that is,

$$T/\tau = g(\theta) \tag{33}$$

This assumption is supported by the experimentally determined function g(θ) displayed in Fig. 15.

Next we consider a train of stimuli separated by a uniform time interval t_s. Consultation of Fig. 16 leads to the relation

$$\delta_{i+1} + T_i = \delta_i + t_s \tag{34}$$

Division by τ, and assuming that the influence of a single stimulus decays sufficiently fast such that eq. 33 holds for every i, yields the phase relationship:

$$\theta_{i+1} = \theta_i + \Omega - g(\theta_i) \; ; \; \Omega = t_s/\tau \tag{35}$$

which has the form of a circle map (see Fig. 17) where the rate of rotation $\Omega = t_s/\tau$-is set by the interstimulus distance t_s.

Using g(θ) from Fig. 15, eq. (35) has been used to successfully predict the -esponse to a train of stimuli as a function of ty (see Fig. 18). The so-called Wenckebach phenomenon

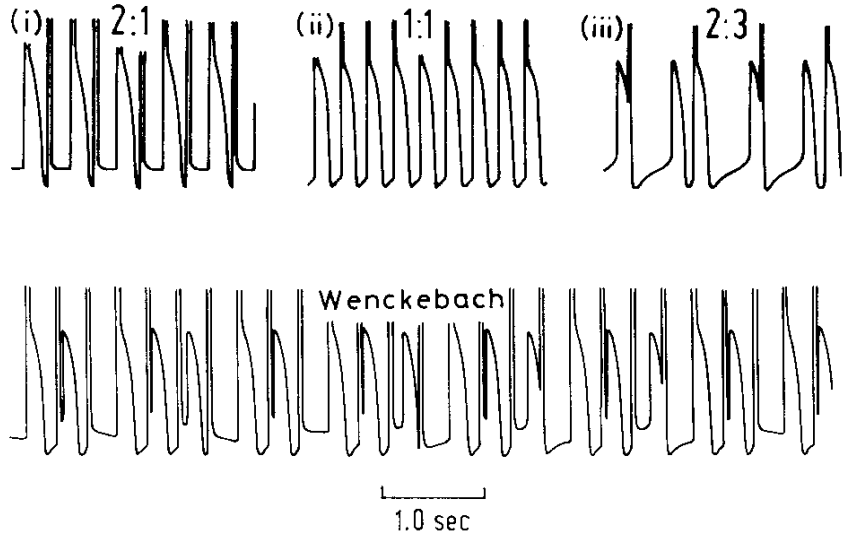


Figure 13: Influence of periodic stimulation as a function of the interstimulus interval t_s,: a) Stable phase locked pattern (i) 2:1 $t_s = 210$ msec; (ii) 1:1, $t_s = 240$ msec; (iii) 2:3 $t_s = 600$ msec. b) Irregular dynamics displaying the Wenckebach phenomenon, $t_s = 280$ msec. (After Guevara et al., 1981; copyright 1981 by the AAAS.)

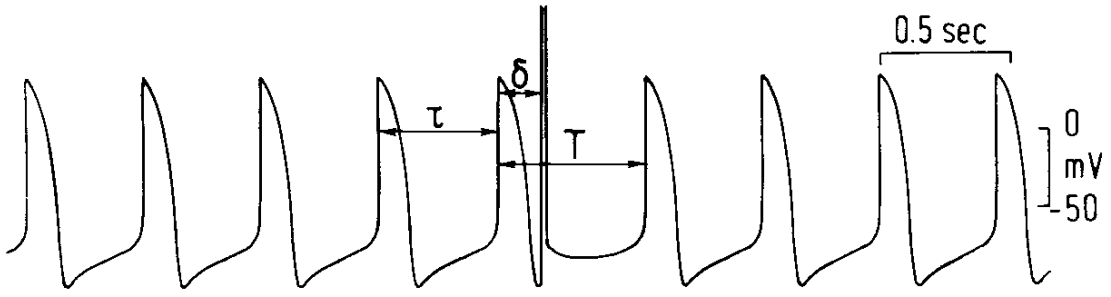


Figure 14: Time course of the transmembrane electrical potential from an aggregate of embryonic heart cells. Left: Spontaneous pulses. Right: After administration of a brief depolariz-ing stimulus (off-scale response) which occurs δ msec after the action potential upstroke. The graph sharply rises, and the spontaneous-state period τ is shifted to a new value T. (From Guevara et al., 1981; copyright 1981 by the AAAS.)

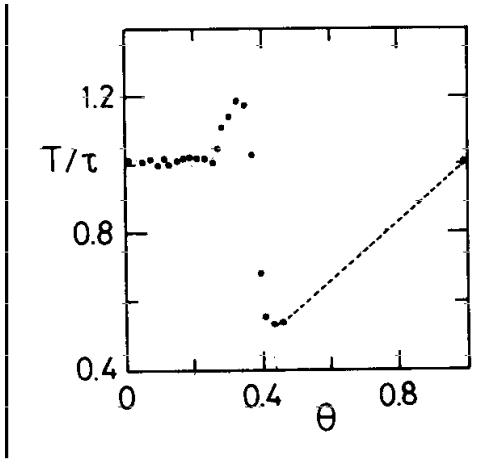


Figure 15: The function g(θ) defined in eq. (33), as experimentally determined for embryonic chick heart cell aggregates (from Guevara et al., 1981; copyright 1981 by the AAAS).

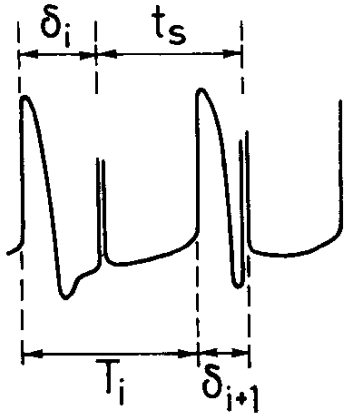


Figure 16: Graphical demonstration of the relation $T_i+\delta_{i+1} = \delta_i+t_s$ for $T_i < \delta_i+t_s < T_i+\tau$

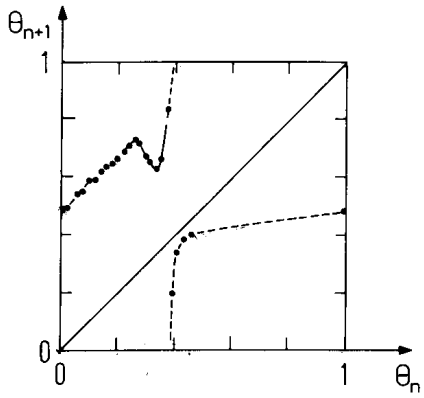


Figure 17: Experimentally determined circle map that describes the dynamics of beating chicken heart cell aggregates. This graph is obtained by using g(θ) from Fig. 15 in eq. (35). (From Guevara et al., 1981; copyright 1981 by the AAAS.)

in Fig. 13c (i. e., the gradual prolongation of the time between a stimulus and the subsequent action potential until an active potential is skipped either irregularly or in a phase locked pattern) occurs also in human electrocardiograms. There the external stimulus is replaced by the stimulus provided by the sinoatrial node.

1.9.2 Averaging and Weak Coupling

Although the general behavior of coupled neural oscillators is very difficult to analyze, limiting cases can be treated. We will describe one method, the method of averaging, used successfully to study the dynamics of two or more neural oscillators that are weakly coupled. In this limit, the coupling is sufficiently weak that each oscillator's trajectory remains close to its intrinsic limit cycle. The primary effect of the coupling is to perturb the relative phase between the oscillators, much as we described above. Because the perturbation per cycle is small (with weak coupling), however, the net effect occurs only over many cycles, and the per cycle effect is seen as averaged. For illustration, we summarize the use of averaging to describe the phaselocking properties of two identical Morris-Lecar oscillators when coupled with identical mutually excitatory synapses.

We assume that motion of each oscillator along its limit cycle can be rewritten in terms of a phase variable. Thus an oscillator's membrane potential is periodic with period T and follows the function $V(\theta_j)$, where θ_j is the phase of the j-th oscillator, j = 1,2, and V is the voltage component of the limit cycle trajectory. In the absence of coupling, the dynamics are given simply as $\theta_j = t + C_j$, where C_j is an arbitrary phase shift. Now consider the effect of small coupling. A brief, weak synaptic current I_{syn} to cell i from activity in cell j will cause a phase shift in cell i:

$$\Delta\theta_i = -\Delta^*(\theta_i)(t) I_{syn}(\theta_i(t), \theta_j(t)), \tag{36}$$

where $\Delta^*(t)$ is the infinitesimal phase response function, the minus sign converts excitatory current to positive phase shift. The synaptic current is given by

$$I_{syn}(\theta_i, \theta_j) = g_c \alpha(\theta_j(t))(V(\theta_i(t)) - V_{syn}) \tag{37}$$

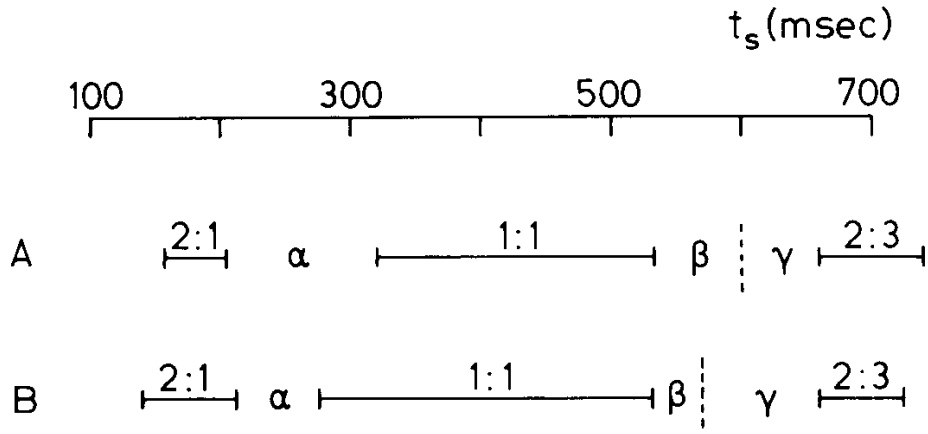


Figure 18: Experimentally determined and theoretically computed responses to periodic stimulation of period t_s with the same pulse durations and amplitudes as in Fig. 13 a), a) Experimentally determined dynamics: 2:1,1:1,2:3 mode locking regions and three zones α, β, γ of complicated dynamics, b) Theoretically predicted dynamics obtained via eq. (35). (After Guevara 3.L. 1981; copyright 1981 AAAS.)

where the postsynaptic gating variable $\alpha(t)$ in cell i is activated by the presynaptic voltage $V(\theta_j)$, V_{syn} is the reversal potential for the synapse, and g_c is the strength of the synaptic coupling. The gating variable $\alpha(t)$ could be represented by a so-called (event-triggered) alpha function, which looks like the impulse response of a second order low-pass filter, e.g. $te^{-t/\tau}$. Alternatively, it could obey a voltage-gated differential equation. In the method of averaging we simply "add up" all the phase shifts due to the synaptic perturbations and average them over one cycle of the oscillation. Thus, after averaging, the coupled system is found to satisfy

$$\frac{d\theta_1}{dt} \;=\; 1 \;+\; g_c H(\theta_2 - \theta_1) \;+ Order(g_c^2) \tag{38}$$

$$\frac{d\theta_2}{dt} \;=\; 1 \;+\; g_c H(\theta_1 - \theta_2) \;+ Order(g_c^2) \tag{39}$$

where H is a T-periodic "averaged" interaction function, given by

$$H(\phi) \;=\; \frac{1}{T}\int_0^T \Delta^*(t)\alpha(t+\phi)(V_{syn} - V(t))dt \tag{40}$$

The key to these models is the computation of H.

In figure 19A, we show the function $V^*(t)$ along with the synaptic gating variable $\alpha(t)$ over one cycle for exactly the same parameters as in figure 11B. Here $\alpha(t) \;=\; 0.04te^{\,t/5}$ is an alpha-function with a 5 ms time constant. Note the similarity (except for scale) of the excitatory PRC and the infinitesimal PRC, $V^*(t)$. As with the PRC, $V^*(t)$ is mainly positive, showing that the predominant effect of depolarizing perturbations is to advance the phase or, equivalently, to speed up the oscillator. In only a very small interval of time can the phase be delayed, and this is a general property of membranes that become oscillatory through a saddle node bifurcation.

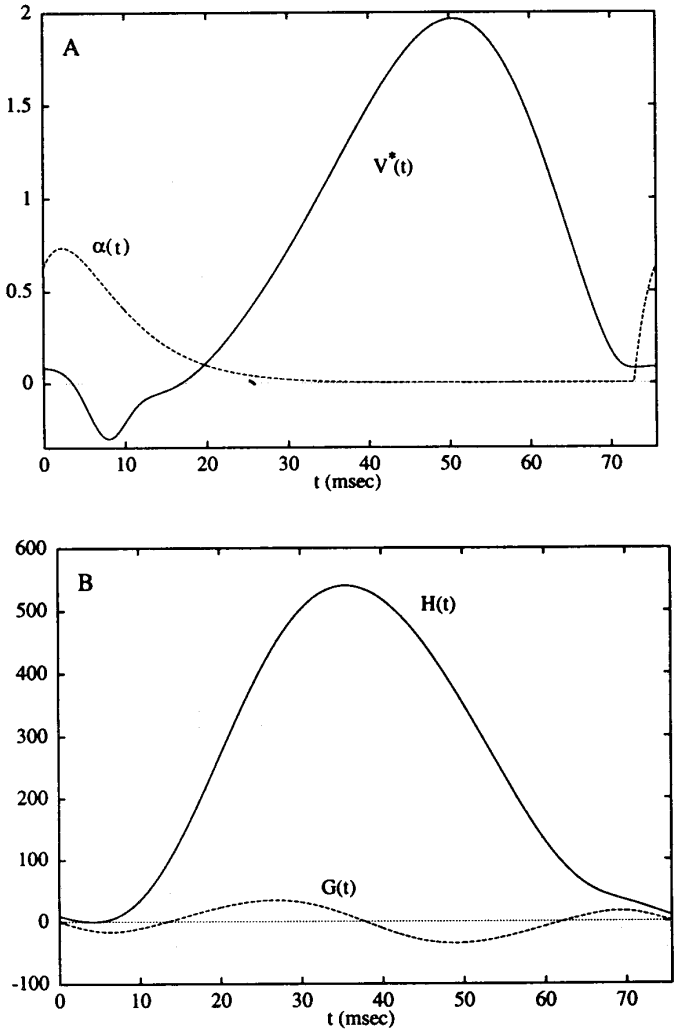


Figure 7.14
The method of averaging for two weakly coupled identical Morris-Lecar oscillators. Parameters for the oscillators are as in figures 7.4–7.6 (and figure 7.12A) and an applied current $I = 50\,\mu A/cm^2$. (A) Solid line shows the infinitesimal PRC, $V^*(t)$ and the dashed line shows the time course of the excitatory synaptic conductance (for plotting convenience, here multiplied by ten), modeled as an alpha function with a 5 msec time constant and peak of about $0.075\,mS/cm^2$. The alpha function "turns on" when V crosses 20 mV. (B) Interaction function $H(t)$ and its odd part $G(t)$ for the synaptic dynamics shown in panel A and a synaptic reversal potential of 0 mV. Zeros of the function $G(t)$ correspond to phase-locked solutions to the weakly coupled system; stable solutions have positive slopes and unstable have negative slopes.

Figure 19:

Figure 19B shows the function H(t) defined in eq. 40 for that figure's alpha function and for $V_{syn} = 0$ mV. We can use this function along with eqs. 38-39 to determine the stable phase-locked patterns for this coupled system. Let $\Theta = \theta_2 - \theta_1$ denote the phase difference between the two oscillators. From eqs. 38-39 we see that Θ satisfies

$$\frac{d\Theta}{dt} = g_c(H(-\Theta) - H(\Theta)) + Order(G_c^2) \equiv -2g_cG(\Theta) + Order(g_c^2) \qquad (41)$$

Here $G(\Theta)$ is just the odd part of the function H. Because the coupling is weak, the higher-order terms, $Order(g_c^2)$ are ignored. Equation 41 is just a first-order equation. Phase-locked states are those for which Θ does not change, that is, they are roots of the function $G(\Theta)$ and they are stable fixed points if $G'(\Theta) > 0$. Because any odd periodic function has at least two zeros, $\Theta = 0$ and $\Theta = T/2$, there will always exist phase-locked states, although, these may not be stable. Synchronous solutions ($\Theta = 0$) imply that both membranes fire together. Antiphase solutions ($\Theta = T/2$) are exactly one-half cycle apart. Figure 19B shows the function $G(\Theta)$, from which we see, that there are four distinct fixed points: the synchronous (precisely in-phase) solution; the antiphase solution; and a pair of phase-shifted solutions at $\Theta \approx +/-15ms$. Both the synchronous and antiphase solutions are unstable but the phase-shifted solution is stable. Thus, if two of these oscillators are coupled with weak excitatory coupling and the parameters chosen as above, they will phase-lock with a phase shift of about 20% of the period. Although the classical view is that mutual excitation leads to perfect synchrony, computations with a variety of neuronal models suggest that this is not generally the case. This type of analysis is easily extended to systems where the oscillators are not exactly identical, coupling is not symmetric, and there are many more oscillators. The behavior of such phase models and the forms of the interaction functions, H, are the topics of current research.

Summary

We have introduced and used some of the basic concepts of the qualititative theory of differential equations to describe the dynamic repertoire of a representative model of excitability. We believe that a geometrical treatment, as in the phase plane, gives one an opportunity to see more clearly and to appreciate the underlying qualitative structure of models. One can see which initial conditions, for example, those resulting from a brief perturbing stimulus, will lie in the domain of attraction of any particular stable steady state or limit cycle. This is especially helpful for the design of experiments to switch a multistable system from one mode to another. Analytic methods are also important for determining and interpreting the stability of solutions (e.g., eq. 28 for the Hopf bifurcation) and for approximating aspects of the solution behavior. Another useful conceptual device is the bifurcation diagram by which we have provided compact descriptions of the system attractors. Although in several of our illustrations, the bifurcation parameter was I, and the steady-state I-V relation appeared explicitly in the diagram, channel density, synaptic weight, or any other parameter can be used.

We have shown how a minimal but biophysically reasonable membrane model can be massaged to exhibit robustly a variety of physiologically identifiable firing behaviors. For the simplest two-variable Morris-Lecar model, we illustrated some qualitative differences in threshold behavior. When the steady-state current-voltage relation is monotonic, action potential size may be graded, although generally quite steeply with stimulus strength,

and latency for firing is finite; when it is N-shaped, there is a true (saddle point) threshold for action potentials, latency may be arbitrarily long, and intermediate-sized responses are not possible. Correspondingly, for a steady stimulus, the monotonic case leads to onset of oscillations with a well-defined, nonzero frequency (Hopf bifurcation), and with possibly small amplitude (super-critical). In contrast, in the N-shaped case repetitive firing first appears with zero frequency (homoclinic bifurcation). These features are consistent with some of those used to distinguish axons with different repetitive firing properties. Additionally, we have provided a geometric interpretation of some common forms of bursting neurons. Many bursters can be dissected into fast dynamics coupled to one or more slow processes that move the fast dynamics between resting and oscillatory states. Coupled and forced oscillators can often be reduced to maps or to continuous low-dimensional systems of phase equations, especially when the interactions are weak.

References

Good references to look for additional material on the topics discussed in this section are

- Methods in Neuronal Modeling. From Ions to Networks. C. Koch and I. Segev, MIT Press, Cambridge, Massachusetts.
- Biophysics of Computation. Information processing in single neurons. C. Koch, Oxford University Press.
- Handbook of Biological Physics, Vol. IV. Moss and Gielen (Eds.), Elsevier.
- Deterministic Chaos: An Introduction. H.G. Schuster VCH Verlagsgesellschaft mbH. Weinberg, Germany (1988)

Excercises

Excercises 1 to 4 require the use of the software APSIM. Ask your assistent for this program !

Problem 1. Parameters of the action potential

Set the Mode to Active. Set the Stimulus Strength to 30 nAmps and Duration to 0.5 msec. Click on Run to elicit an action potential. Use the Measure Window to measure the voltage at the peak of the action potential and the corresponding time at which the peak occurs by centering the cross hair on the peak.

How close is this voltage to E_{Na} (normally +55 mV)? Why doesn't it equal E_{Na} ?

Opgave 2. Membrane currents and conductances

Open up the Membrane Conductances and Membrane Currents windows from the Plots Menu on the menu bar.

Click on the view Vm box in the Membrane Conductances window.

Click Run to elicit an action potential.

Note the relationship between membrane voltage, membrane conductances and membrane currents.

- Why do you think the sodium current shows an initial brief peak during the rising phase of the action potential?
- Why does the sodium conductance rapidly decline after the action potential reaches its peak?
- Why does the sodium current continue to increase even as the sodium conductance declines?
- Why does the K current decline more rapidly than the K conductance?

Opgave 3. Channel gates

Close the Membrane Currents window by clicking on the small box in the upper left hand corner.

Now open up the Channel Gates window by choosing this option under the Plots Menu. Click on **Clear** and then click **Run** to start a simulation. Note how the Na channel inactivation gates (h, plotted in green), activation gates (m, plotted in red), and channel activation gates (n, plotted in blue) open and close during the action potential.

- Why does the Na conductance fall more rapidly than the rate at which its activation (m) gates close?
- What fraction of Na channel inactivation gates are shut at the resting potential?
- Rank the m, h. and n gates in order of speed, from slowest to fastest, during the rising and falling phases of the action potential.

opgave 4. Effect of changing sodium concentration on the action potential

Now close the **Channel Gates** and **Membrane Conductance** windows. Click on **Clear** and then run the simulation. Next choose the **Ionic Concentrations** option from the Edit menu.

Alter the external Na concentration from its normal value of 460 mM to 310 mM (approximately 2/3 of normal) and re-run the simulation.

How does this change the peak value of the action potential, action potential duration, after- hyperpolarization, resting potential, maximal rate of rise of the action potential, and maximal rate of repolarization?

Now re-determine the value of the threshold potential and the amount of stimulating current required to reach threshold (using a 0.5 msec long stimulus pulse).

Explain the observed changes (Hint: try opening up the **Membrane Conductances** and/or **Channel Gates** windows to gain insight into what is happening.)

Problem 5.

a. Consider the second-order equation

$$\ddot{y} \; + \; b\dot{y} \; + \; y \; = \; 0$$

What is the stability type of the equilibrium point at the origin for various values of b ?

b. Consider the second-order equation

$$\ddot{y} \; + \; b\dot{y}^3 \; + \; y \; = \; 0$$

What is the stability type of the equilibrium point at the origin for various values of b ?

Problem 6

Use MATLAB to simulate and to discuss the stability and instability of the equilibrium point of the system

$$\begin{aligned} \dot{x}_1 \; &= \; -\, x_2 \; + \; x_2^3 \\ \dot{x}_2 \; &= \; -\, x_1 \; + \; x_1^3 \end{aligned}$$

Note especially the equilibrium point (1,0).

Problem 7

Use MATLAB to simulate and to show that the system

$$\begin{aligned} \dot{x}_1 \; &= \; 2x_1x_2 \\ \dot{x}_2 \; &= \; \frac{1}{4} - x_1^2 \; + \; x_2^2 \end{aligned}$$

has two centers. Hint: use the change of variables $x_1 \rightarrow \frac{1}{2} + x_1$ to find one of the centers.

Problem 8

Discuss the stability properties of the origin for the system

$$\begin{aligned} \dot{x}_1 \; &= \; x_2 \; + \; x_1x_2 \; + ax_1x_2^2 \\ \dot{x}_2 \; &= \; -\, x_1 \; - \; x_1^2 \; + \; x_2^2 \end{aligned}$$

for various values of a.

Problem 9.

Use MATLAB to simulate and to analyze the van-der-Pol oscillator, given by

$$\ddot{x} + C(x^2 - 1)\dot{x} + \omega^2 x = 0$$

- Which of the parameters c and ω determine the frequency of the oscillations along the limit cycle ?
- Does the amplitude of the oscillations along the limit cycle depend on c and ω?
- Which of the parameters c and ω determine the rate of convergence towards the limit cycle ?

Problem 10

Consider the Rayleigh oscillator:

$$\ddot{y} \; + \; \dot{y}^3 \; - \; 2\lambda\dot{y} \; + \; y \; = \; 0$$

where λ is a small positive scalar.

Convert this into a set of two first-order differential equations and investigate the Hopf-bifurcation at $\lambda = 0$. Determine the stability of the periodic orbit.

Hint: The approximate bifurcation cruve is $\lambda \; = a^2/8 \; + \; O(a^3)$.

Problem 11.

Discuss the Hopf bifurcation near the periodic orbits of the system

$$\dot{x}_1 \; = \; \lambda x_1 \; + \; x_2 \; + \; x_1 x_2 \; + x_1 x_2^2$$

$$\dot{x}_2 \; = \; \; - \, x_1 \; + \; \lambda x_2 \; - \; x_1^2 \; + \; x_2^2$$

near the origin for $|\lambda|$ small.

Problem 12.

The simple Hopf-bifurcation is obtained by a system with the following dynamics:

$$\frac{dr}{dt} \; = \; r(c - r^2)$$

$$\frac{d\phi}{dt} \; = \; 2\pi$$

where the system is in polar coordinates with radius r and phase ϕ.

a. Plot the bifurcation diagram for this simple case. (i.e. plot the value of the steady-state solution of the radius r as a function of the value of parameter c).

b. Consider now the system

$$\frac{dr}{dt} \; = \; r(c + 2r^2 - r^4)$$

$$\frac{d\phi}{dt} \; = \; 2\pi$$

Plot the bifurcation diagram for this system. What happens to r when c starts from -2 to larger values up to +2 ? What happens if the value of c changes from +2 to -2 ?

Problem 12. Assume that we can ignore any stochastic input to neuron firing and that neurons tend to fire regularly at a constant firing rate. Assume that the firing of a neuron corresponds to sinusoidal oscilations of the phase where the action potential corresponds to one particular phase in the cycle. Also assume that neuron i in a network of coupled oscillators has the (constant) "firing rate" ω_i. In that case the only relevant parameter is the phase of neuron. The dynamics is given by

$$\frac{d\theta}{dt} \; = \; \omega_i \; + \; \frac{K}{N} \sum_{j=1}^{N} sin(\theta_j \; - \; \theta_i)$$

where

- ω_i is "firing rate" of neuron i

- θ_i is phase of neuron i
- K is coupling strength between neurons
- N is number of neurons.

a. Suppose that the frequencies ω_i are symmetrically distributed around a mean frequency ω_0. This allows us to introduce new variables $\psi_i = \theta_i - \omega_0 t$. By replacing ω_i with $\omega_i - \omega_0$, we obtain

$$\frac{d\psi_i}{dt} = \omega_i - \frac{K}{N}\sum_{j=1}^{N} sin(\psi_i - \psi_j) \tag{42}$$

b. Show that the mean-field approximation for dolving the behavior of the network of neurons gives rise to the differential equation

$$\frac{d\psi_i}{dt} = \omega_i - Krsin(\psi_i - \Theta) \tag{43}$$

where

- $rexp(i\Theta) = \frac{1}{N}\sum_{j=1}^{N} exp(i\psi_j)$
- r provides a measure for the amount of synchrony, with r=0 indicating independent uncoupled behavior of all oscillators and r=1 indicating perfect in-phase locking.
- Θ indicating the average phase of the oscillators.

c. show that Θ is an order parameter. Find the solutions for the differential equation as a function of K; indicate which solutions are stable modes for the system and which are unstable.

Problem 13. The circle map is a standard procedure to investigate whether a mapping of a process has stable states and whether or when it converges to the stable state(s). Consider the mapping

$$x_{n+1} = ax_n(1 - x_n)$$

for the interval $0/leqx/leq1$. a. Show that x=0 is a stable attractor for 1 ¡ a 1.
b. Show that there is another attractor on the interval $0 \leq x \leq 1$ for 1 ¡ a ¡3.

Problem 14.

Assume two neurons, each with a different firing rate. Assume that the neuron with the lower firing rate receives spike input from the neuron with the higher firing rate, such that the next action potential of the neuron with the lower firing rate arrives earlier in time (see Fig. 16 in lecture notes) according to the relation

$$\Theta_{n+1} = \Theta_n + \Delta(\Theta_n) + P/T$$

Show that a necessary and sufficient condition for convergence to a fixed phase relation Θ^* with one-to-one firing of the neurons with $\Delta(\Theta^*) = -P/T$ requires that $-2 < \frac{\partial\Delta(\Theta)}{\partial\Theta} < 0$.

The response of a population of classical Hodgkin-Huxley neurons to an inhibitory input pulse

Christoph Börgers[1], Martin Krupa[2], and Stan Gielen[2]

[1]*Department of Mathematics, Tufts University,*

Medford, Massachusetts 02155, USA

[2]*Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen,*

Dept. of Biophysics, Geert Grooteplein 21, 6525 EZ Nijmegen, Netherlands

Abstract:

corresponding author: Christoph Börgers, cborgers@tufts.edu, phone: 617-627-2366, fax: 617-627-3966

keywords: Hodgkin-Huxley equations, synchronization, type II neurons, gamma oscillations, shunting inhibition

1 Introduction

Transient synchronization of masses of neurons is believed to be common in the brain and important for brain function. In a seminal paper, van Vreeswijk et al. (1994) showed that often synaptic inhibition, not excitation, leads to synchronized activity. Fast-spiking inhibitory interneurons are believed to play the central role, in particular, in the generation of gamma (30–90 Hz) rhythms (Traub et al., 1997; Whittington et al., 2000; Traub et al., 2003; Csicsvari et al., 2003; Hájos et al., 2004; Mann et al., 2005; Compte et al., 2008).

Arguably the simplest example of synchronizing inhibition is that of a population of uncoupled neurons receiving a common strong inhibitory input pulse. Such a pulse can transiently drive all neurons of its target population towards a common quasi-steady state, thereby driving them (that is, the quantities characterizing their states, such as membrane potentials and ionic gating variables) towards each other. This is the foundation of the "PING" (Pyramidal-Interneuronal Network Gamma) mechanism (Whittington et al., 2000; Börgers and Kopell, 2003; Börgers and Kopell, 2005), in which gamma rhythms arise from the interaction between excitatory pyramidal cells (E-cells) and inhibitory fast-spiking interneurons (I-cells): Population spike volleys of the I-cells synchronize the E-cells, and population volleys of the E-cells trigger synchronous spike volleys of the I-cells. By "PING" we mean "strong PING" here in the terminology of Börgers et al. (2005), *i.e.*, PING with deterministic drive to the E-cells. "Weak PING", in which each E-cell is driven stochastically and spikes irregularly and infrequently (Börgers et al., 2005), will be addressed briefly in the Discussion.

In this paper, we take another look at the approximate synchronization of a population of neurons by a single inhibitory pulse, and find that it often fails for classical Hodgkin-Huxley neurons. The reason lies in the nature of the transition from excitability to spiking. For the classical Hodgkin-Huxley neuron, this transition involves a subcritical Hopf bifurcation. For many other neuronal models, on the other hand, it involves a saddle-node bifurcation on an invariant cycle; the simplest model of this sort is the theta neuron (Ermentrout and Kopell, 1986; Hoppensteadt and Izhikevich, 1997; Gutkin and Ermentrout, 1998). Neuronal models are often called of type I if the transition from rest to spiking involves a saddle-node bifurcation on an invariant cycle, and of type II if it involves a Hopf bifurcation (Rinzel and Ermentrout, 1998; Gutkin and Ermentrout, 1998; Ermentrout, 1996).

For both type I and type II neurons, a sufficiently strong inhibitory pulse introduces an attracting quasi-steady state. For the classical Hodgkin-Huxley neuron, this quasi-steady state is a focus, *i.e.*, the center of a spiral; as the inhibition decays, it turns from attracting to (weakly) repelling. For the theta neuron, on the other hand, and for other model neurons of type I, the attracting quasi-steady state is a node, which is annihilated altogether in a saddle-node collision as the inhibition decays. As we will show, this difference gives rise to crucial differences in synchronization behavior. Theta neurons are easily synchronized by a pulse of inhibition, provided only that the pulse is strong and long-lasting enough. On the other hand, for classical Hodgkin-Huxley neurons, synchronization by a pulse of inhibition is fragile. It often fails when the inhibition is shunting (that is, when the reversal potential is near the resting potential). Surprisingly, it is more likely to fail, even for hyperpolarizing inhibition, for stronger and longer-lasting inhibitory pulses.

We expect these results to have significance for the question which neurons in the brain participate in or are entrained by gamma oscillations. We briefly outline two possible implications here; for details and references, see the Discussion section. First, our results suggest that robust strong PING may require type I pyramidal cells. Several studies indicate that pyramidal cells in superficial layers of the cortex may in fact be of type I with cholinergic modulation, but of type II without it. This raises the possibility that cholinergic modulation may be required for robust strong PING. Second, gamma oscillations generated by purely inhibitory networks, often called "ING" (Interneuronal Network Gamma) (Whittington et al., 1995; Whittington et al., 2000), may be fragile when the neurons are of type II. There are reports in the literature suggesting that the fast-spiking inhibitory intereneurons believed to underlie gamma oscillations may in fact be of type II.

2 Methods

We describe here the models used in our computational study.

2.1 The theta neuron

The theta model (Ermentrout and Kopell, 1986; Hoppensteadt and Izhikevich, 1997; Gutkin and Ermentrout, 1998) represents a neuron by a point $P = (\cos\theta, \sin\theta)$ moving on the unit circle in the plane. This is analogous to Hodgkin-Huxley-like, conductance-based models, which represent a periodically spiking space-clamped neuron by a point

moving on a limit cycle in a higher-dimensional phase space. In the absence of synaptic input, the differential equation defining the theta neuron is

$$\frac{d\theta}{dt} = 1 - \cos\theta + I\,(1+\cos\theta)\ . \tag{1}$$

Here t should be thought of as time measured in milliseconds (see Börgers and Kopell, 2003, Section 2) and I as the analogue of an external input current. For $I < 0$, Eq. (1) has exactly two fixed points in $(-\pi,\pi)$, namely

$$\theta_0^{\pm} = \pm\arccos\frac{1+I}{1-I} = \pm 2\arccos\frac{1}{\sqrt{1-I}}\ . \tag{2}$$

(The second equality in Eq. (2) is a consequence of the angle doubling formula for the cosine function.) The fixed point $\theta_0^- \in (-\pi,0)$ is stable, and $\theta_0^+ \in (0,\pi)$ is unstable. As I increases, the fixed points approach each other. As I crosses 0 from below, a saddle-node bifurcation occurs: The fixed points collide at $\theta_0^- = \theta_0^+ = 0$, and there are no fixed points for $I > 0$. For a theta neuron, to "spike" means, by definition, to reach $\theta = \pi$ (modulo 2π). For $I > 0$, the theta neuron spikes with period

$$T = \frac{\pi}{\sqrt{I}}\ .$$

The transition from $I < 0$ to $I > 0$ is the analogue of the transition from excitability to spiking in a neuron.

The theta neuron is equivalent, up to a change of variable, to a quadratic integrate-and-fire neuron with threshold potential $V_T = +\infty$ and reset potential $V_{reset} = -\infty$; see Börgers and Kopell (2005) for a detailed discussion of this connection.

2.2 The theta neuron with inhibitory input

In Results, we will review the effect of adding an exponentially decaying inhibitory term to Eq. (1), discussed in detail in Börgers and Kopell (2003, 2005). Following Börgers and Kopell (2003), we model the inhibition as follows:

$$\frac{d\theta}{dt} = 1 - \cos\theta + \left(I - \left\{\begin{array}{cc} ge^{-(t-t^*)/\tau_i} & \text{if } t \geq t^* \\ 0 & \text{if } t < t^* \end{array}\right\}\right)(1+\cos\theta)\ , \tag{3}$$

where $g > 0$ is the strength of the pulse of inhibition, t^* is its arrival time, and τ_i its decay time constant. We primarily focus on $\tau_i = 10$, since the decay time constant of $GABA_A$-receptor mediated inhibition is on the order of 10 ms. However, we will also discuss the effects of varying τ_i.

Eq. (3) can be derived from the quadratic integrate-and-fire neuron with an exponentially decaying inhibitory *current* input term added to the right-hand side. A variation on this equation is obtained when starting with the quadratic integrate-and-fire neuron with an exponentially decaying inhibitory *synaptic* input term added to the right hand side, in the form $ge^{-(t-t^*)/\tau_i}(V_{syn} - V)$, where V denotes the membrane potential, g the maximal synaptic conductance, and V_{syn} the synaptic reversal potential (see Börgers and Kopell, 2005, for a detailed derivation). However, the difference between current inputs (Börgers and Kopell, 2003) and synaptic inputs (Börgers and Kopell, 2005) is not crucial in the current context. Here we will use current inputs, *i.e.*, Eq. (3), for simplicity.

2.3 E/I networks of theta neurons

For illustration, we show in Fig. 6 a spike rastergram resulting from a simulation of a network of 400 excitatory theta neurons (E-cells) and 100 inhibitory theta neurons (I-cells). All details of this simulation were as in Börgers and Kopell, 2003, Fig. 1B, with the following three exceptions. (1) Connectivity was all-to-all here, whereas it was sparse and random in Börgers and Kopell, 2003, Fig. 1B. (2) The (random) initializations were not the same. (3) We simulated only 100 ms here, whereas 200 ms were simulated in Börgers and Kopell, 2003. As in Börgers and Kopell, 2003, Fig. 1B, $\tau_i = 10$ in Fig. 6.

2.4 The classical Hodgkin-Huxley neuron

The classical Hodgkin-Huxley equations (Hodgkin and Huxley, 1952) for the space-clamped squid giant axon are

$$C\frac{dV}{dt} = g_{Na}m^3h(V_{Na}-V)+g_Kn^4(V_K-V)+g_L(V_L-V)+I\,, \tag{4}$$

$$\frac{dm}{dt} = \alpha_m(V)(1-m)-\beta_m(V)m\,, \tag{5}$$

$$\frac{dh}{dt} = \alpha_h(V)(1-h)-\beta_h(V)h\,, \tag{6}$$

$$\frac{dn}{dt} = \alpha_n(V)(1-n)-\beta_n(V)n\,. \tag{7}$$

The letters C, V, t, g, and I denote capacitance density, voltage, time, conductance density, and current density, respectively. The units used for these quantities are μF/cm^2, mV, ms, mS/cm^2, and μA/cm^2, respectively. For brevity, units will often be omitted from here on. Up to a change in notation, the parameter values that Hodgkin and Huxley chose are $V_{Na}=45$, $V_K=-82$, $V_L=-59.387$, $g_{Na}=120$, $g_K=36$, $g_L=0.3$, and $C=1$. The letters m, h, and n denote the gating variables, which are dimensionless real numbers between 0 and 1. The rate functions α_x and β_x, $x=m,h,n$, are given by

$$\alpha_m(V) = \frac{(V+45)/10}{1-\exp\left(-(V+45)/10\right)}\,, \tag{8}$$

$$\beta_m(V) = 4\exp\left(-(V+70)/18\right)\,, \tag{9}$$

$$\alpha_h(V) = 0.07\exp(-(V+70)/20)\,, \tag{10}$$

$$\beta_h(V) = \frac{1}{1+\exp(-(V+40)/10)}\,, \tag{11}$$

$$\alpha_n(V) = \frac{(V+60)/100}{1-\exp(-(V+60)/10)}\,, \tag{12}$$

$$\beta_n(V) = 0.125\exp(-(V+70)/80)\,. \tag{13}$$

Although of course a Hodgkin-Huxley model neuron has spikes of positive width, we say that there is a spike "at time t_0" if

$$V(t_0)=0 \quad \text{and} \quad \frac{dV}{dt}(t_0)>0. \tag{14}$$

2.5 The classical Hodgkin-Huxley neuron with inhibitory input

To model synaptic inhibition, we modify Eq. (4) by adding a term of the form

$$I_{syn} = gs(t)(V_{syn}-V) \tag{15}$$

to the right-hand side. *Constant* inhibitory input corresponds to

$$s(t)=1$$

for all t. A *decaying pulse* of inhibition is modeled by

$$\begin{cases} ge^{-(t-t^*)/\tau_i}(V_{syn}-V) & \text{if } t\geq t^*\,, \\ 0 & \text{if } t<t^*\,, \end{cases} \tag{16}$$

where t^* denotes the arrival time of the inhibitory pulse. The parameter V_{syn} is the synaptic reversal potential, and g is the maximum conductance associated with the synaptic input. The reversal potential of GABAergic synapses can be below the resting potential (*hyperpolarizing* inhibition), at the resting potential (*shunting* inhibition), or even above the resting potential (Isaev et al., 2007; Jeong and Gutkin, 2007; Lu and Trussell, 2001). We will therefore experiment with the values $V_{syn}=-80$ (hyperpolarizing inhibition) and $V_{syn}=-65$ (shunting inhibition) in this paper.

2.6 E/I networks in which the E-cells are classical Hodgkin-Huxley neurons

We will present simulations of networks in which the E-cells are classical Hodgkin-Huxley neurons (Eqs. (4)–(13)). The external drive I in Eq. (4) will in this context be denoted by I_e. Some of our simulations include a heterogeneous drive to the E-cells. By this we always mean that the external drive to the j-th E-cell is

$$I_{e,j} = (1+0.2X_j)I_e, \tag{17}$$

where the X_j are independent Gaussian random variables with mean 0 and variance 1. The drives $I_{e,j}$ are time independent. However, in the simulations in which drive to the E-cells is heterogeneous, we also add time-dependent noisy drive. This is done by adding a term to the right-hand side of the equation describing the time evolution of the membrane potential of the j-th E-cell in the form

$$-0.05 s_{stoch,j}(t) V_j(t), \tag{18}$$

where the functions $s_{stoch,j}$ jump to 1 at random times, and decay exponentially with time constant 3 ms between jumps. The jumps occur on Poisson schedules with mean frequency 10 Hz. The noisy inputs to different E-cells are independent of each other. Eq. (18) is intended to mimic the effect of excitatory synaptic input pulses arriving at random times. The decay time constant of 3 ms is motivated by the fact that the decay time constant of AMPA-receptor mediated glutamatergic synapses is on the order of 3 ms.

In the model networks in which the E-cells are classical Hodgkin-Huxley neurons, the I-cells are Wang-Buzsáki model neurons (Wang and Buzsáki, 1996). The Wang-Buzsáki neuron has the same general form as the classical Hodgkin-Huxley neuron (Eqs. (4)–(7)). However, the differential equation for the gating variable m, Eq. (5), is replaced by

$$m = m_\infty(V) = \frac{\alpha_m(V)}{\alpha_m(V)+\beta_m(V)},$$

and the parameters are $V_{Na} = 55$, $V_K = -90$, $V_L = -65$, $g_{Na} = 35$, $g_K = 9$, $g_L = 0.1$. As for the classical Hodgkin-Huxle neuron, $C = 1$. The rate functions α_x and β_x, $x = m,h,n$, are given by

$$\alpha_m(V) = \frac{0.1(V+35)}{1-\exp(-(V+35)/10))}, \tag{19}$$

$$\beta_m(V) = 4\exp(-(V+60)/18), \tag{20}$$

$$\alpha_h(V) = 0.07\exp(-(V+58)/20), \tag{21}$$

$$\beta_h(V) = \frac{1}{\exp(-0.1(V+28))+1}, \tag{22}$$

$$\alpha_n(V) = \frac{0.01(V+34)}{1-\exp(-0.1(V+34))}, \tag{23}$$

$$\beta_n(V) = 0.125\exp(-(V+44)/80). \tag{24}$$

The external drive to the I-cells is denoted by I_i. Heterogeneity and noise in external inputs are modeled for the I-cells in precisely the same way as for the E-cells.

We adopt the synaptic model of Ermentrout and Kopell (1998). Each synapse is characterized by a synaptic gating variable s associated with the presynaptic neuron, with $0 \le s \le 1$, which evolves according to the equation

$$\frac{ds}{dt} = \rho(V)\frac{1-s}{\tau_R} - \frac{s}{\tau_D},$$

where ρ denotes a smoothed Heaviside function:

$$\rho(V) = \frac{1+\tanh(V/4)}{2},$$

and τ_R and τ_D are the rise and decay time constants, respectively. To model the synaptic input from neuron j to neuron k, we add to the right-hand side of the equation governing the membrane potential V_k of neuron k a term of the form

$$g_{jk} s_j(t)(V_{syn} - V_k),$$

where g_{jk} denotes the maximal conductance associated with the synapse, s_j denotes the gating variable associated with neuron j, and V_{syn} denotes the synaptic reversal potential.

We use the notation $V_{syn,e}$ and $V_{syn,i}$ for the reversal potentials associated with excitatory and inhibitory synapses, respectively, $\tau_{R,e}$ and $\tau_{D,e}$ for the rise and decay time constants of excitatory synapses, and $\tau_{R,i}$ and $\tau_{D,i}$ for the rise and decay time constants of inhibitory synapses. We always use $\tau_{R,e} = 0.1$, $\tau_{D,e} = 3$, and $V_{syn,e} = 0$, values reminiscent of AMPA-receptor-mediated glutamatergic synapses. We use $\tau_{R,i} = 0.3$, but vary $\tau_{D,i}$ and $V_{syn,i}$. In most simulations, we use $\tau_{D,i} = 10$, reminiscent of $GABA_A$-receptor-mediated synapses.

Our model networks of Hodgkin-Huxley and Wang-Buzsáki neurons include 40 E-cells and 10 I-cells. The connectivity is all-to-all. We take the maximal conductance of the synaptic connection from the j-th I-cell to the k-th E-cell to be $g_{ie}/10$, where g_{ie} is independent of j and k. Thus the maximum possible value of the sum of all inhibitory conductances affecting an E-cell is g_{ie}. Similarly, the maximal conductance of the connection from an E-cell to an I-cell is $g_{ei}/40$, and the maximal conductance of the connection from an I-cell to an I-cell is $g_{ii}/10$. We do not include E→E synapses in these simulations.

We always use $g_{ei} = 0.2$. This parameter is chosen so that a population spike volley of the E-cells promptly triggers a population spike volley of the I-cells, but does not cause I-cells to spike multiple times. Unless otherwise indicated, we use $g_{ii} = 0.1$. PING does not require the presence of I→I synapses, but they significantly stabilize the rhythm (Börgers and Kopell, 2005). We experiment with various values of g_{ie}.

2.7 Visualizing the synchronizing effect of inhibition

To illustrate the synchronizing effect of a pulse of inhibition, we consider a model neuron (either a theta neuron, or a classical Hodgkin-Huxley neuron), with constant drive I above the spiking threshold. We denote by T the natural period of the neuron, that is, the period that would be seen without any additional input.

We assume that at time $t = 0$, the neuron spikes. For a theta neuron, this means $\theta = -\pi \bmod 2\pi$. For a classical Hodgkin-Huxley neuron, it means $V = 0$ and $dV/dt > 0$ (compare Eq. (14)). For the classical Hodgkin-Huxley neuron, we also assume, in addition to $V(0) = 0$ and $dV/dt(0) > 0$, that the point $(V(0), m(0), h(0), n(0))$ lies on the limit cycle. (These conditions uniquely determine $m(0)$, $h(0)$, and $n(0)$.) We now consider an inhibitory pulse arriving at some time t^* with $0 < t^* < T$. We denote by t_1 and t_2 the times of the first and second spikes following time t^*, respectively, and define

$$T_i = t_i - t^*, \quad i = 1, 2.$$

Plots of T_1 and T_2 as functions of t^*, as for instance in Fig. 1, help visualize the synchronizing effect of inhibition. Note for instance that immediate and perfect synchronization would correspond to T_1 and T_2 being independent of t^*.

2.8 Numerics

All differential equations were solved using the midpoint method, with the fixed time step $\Delta t = 0.02$. All codes are available from the first author upon request.

3 Results

3.1 Synchronization of theta neurons by a pulse of inhibition

The effect of adding an exponentially decaying inhibitory term to Eq. (1), as shown in Eq. (3), was discussed in detail in Börgers and Kopell (2003,2005). We review some of the material from Börgers and Kopell (2003,2005) here, and add a discussion of the parameter regime in which an inhibitory pulse synchronizes effectively.

First, to illustrate the synchronization resulting from an inhibitory pulse, we show in Fig. 1 the delays T_1 and T_2 between the arrival time t^* of a pulse of inhibition and the next two spikes (see Methods). T_1 and T_2 depend on I, g, and τ_i. Fig. 1 shows the example $I = 0.1$ (thus $T = \pi/\sqrt{I} \approx 9.935$), $g = 0.25$, and $\tau_i = 10$. In this example, T_1 is approximately independent of t^* as long as t^* is not too close to T. Thus most neurons in an asynchronous population will be brought to approximate synchrony by a single pulse of inhibition. Only those neurons that are quite close to spiking when the inhibition arrives ($t^* \approx T$) will escape. These neurons will spike soon after the arrival of the inhibitory pulse. However, Fig. 1 also shows that nearly all of those neurons will spike a second time at approximately the same time at which the others spike first. Thus a single pulse of inhibition comes very close to synchronizing the entire population.

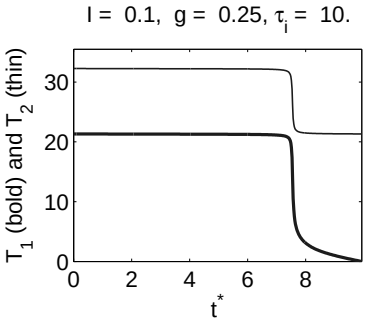


Fig. 1 *Graphs of T_1 and T_2, the time delays between the arrival time t^* of the inhibitory pulse and the first and second spikes following it, respectively, for the theta neuron. Here and in later figures, relevant parameter values are shown at the top.*

The mathematics underlying this synchronization effect were discussed in Börgers and Kopell (2003). There, the synchronization was interpreted as the effect of an attracting "river" (Diener, 1985a; Diener, 1985b)[1] in a phase plane parameterized by θ and the variable $J = I - ge^{-(t-t^*)/\tau_i}$. We will review and discuss the "river" picture below. Briefly, and without making reference to the geometry, the mechanism can be described as follows. If $g > I$, the inhibitory pulse transiently creates an attracting quasi-steady state, which is initially located at

$$\theta_g^- = -\arccos\frac{1+(I-g)}{1-(I-g)}$$

(compare Eq. (2)). While the inhibition decays, trajectories approach this quasi-steady state, and thereby approach each other.

3.2 The river picture for theta neurons

We review here the "river" picture described in Börgers and Kopell (2003,2005), and add some observations relevant to the present work. We consider a theta neuron with an inhibitory pulse arriving at time $t^* = 0$:

$$\frac{d\theta}{dt} = 1-\cos\theta + \left(I - ge^{-t/\tau_I}\right)(1+\cos\theta) \quad \text{for } t \geq 0. \tag{25}$$

Following Börgers and Kopell (2003), we make Eq. (25) autonomous by introducing the variable

$$J = I - ge^{-t/\tau_i}.$$

Eq. (25) then becomes

$$\frac{d\theta}{dt} = 1-\cos\theta + J(1+\cos\theta), \tag{26}$$

$$\frac{dJ}{dt} = \frac{I-J}{\tau_i}. \tag{27}$$

Fig. 2 shows the phase plane for Eqs. (26) and (27) for $I = 0.1$ and $\tau_i = 10$. (Very similar figures were presented in Börgers and Kopell, 2003 and 2005.) The flow in Fig. 2 is upwards, in the direction of increasing J. Trajectories are attracted (exponentially in forward time) to a single "stable river", which starts at $\theta = -\pi$, $J = -\infty$, and reaches $\theta = \pi$ at some value $J^* < I$. As trajectories are exponentially attracted to the stable river, they reach $\theta = \pi$ (which means spiking for the theta neuron) at a time T when $J \approx J^*$, which implies

$$I - ge^{-T/\tau_i} \approx J^*\ ,$$

or

$$T \approx \tau_i \ln\frac{I-J^*}{g}\ . \tag{28}$$

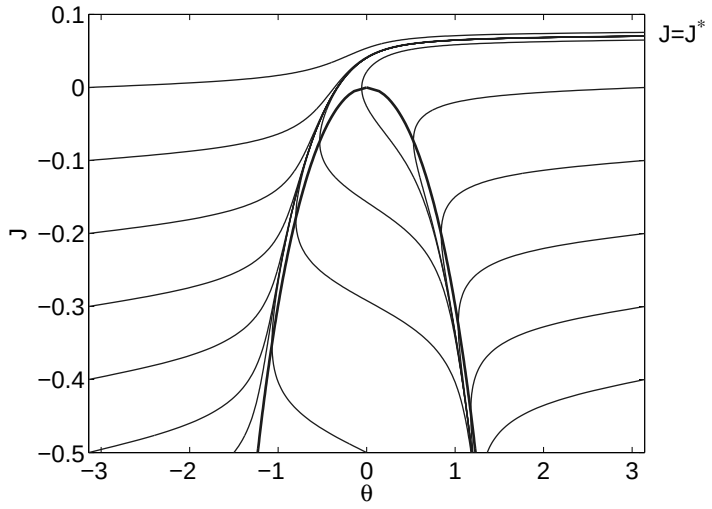


Fig. 2 *Trajectories of Eqs. (26) and (27), and the nullcline* $d\theta/dt = 0$ *(bold line), for* $I = 0.1$ *and* $\tau_i = 10$*. The flow is upwards, in the direction of increasing J.*

Since T is independent of $\theta(0)$, this implies synchronization.

We note that trajectories with different initial conditions typically come far closer to each other (and to the stable river) than to the quasi-steady state. This point, which will play a role later on, is demonstrated by Fig. 3, which shows θ as a function of t for several different initial conditions, together with the quasi-steady state (indicated in bold), for $I = 0.1$, $g = 0.3$, and $\tau_i = 10$.

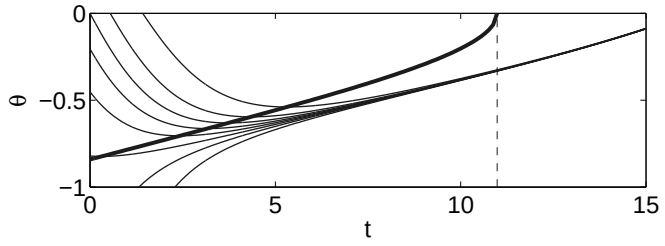


Fig. 3 *Graph of* $\theta(t)$ *for several different initial values* $\theta(0)$*, and the quasi-steady state (bold), for* $I = 0.1$*,* $g = 0.3$*, and* $\tau_i = 10$*. The dashed line indicates the time at which the quasi-steady state ceases to exist.*

3.3 On the parameter range in which a pulse of inhibition synchronizes a population of theta neurons

It is not easy to analyze rigorously for which values of I, g, and τ_i tight synchrony will be obtained by a single pulse of inhibition. However, the following argument does come close to answering this question. At time t^*, the time constant associated with the approach to the stable quasi-steady state is the reciprocal of

$$-\frac{d}{d\theta}\left(1-\cos\theta+(I-g)(1+\cos\theta)\right)\Big|_{\theta=\theta_g^-} = -\left(1+(g-I)\right)\sin\theta_g^- =$$

$$\left(1+(g-I)\right)\left(1-\cos^2\theta_g^-\right)^{1/2} = \left(1+(g-I)\right)\left(1-\left(\frac{1-(g-I)}{1+(g-I)}\right)^2\right)^{1/2} =$$

[1] Stable rivers correspond to stable Fenichel slow manifolds (Fenichel, 1979). We prefer to use the more intuitive term 'river' in this paper.

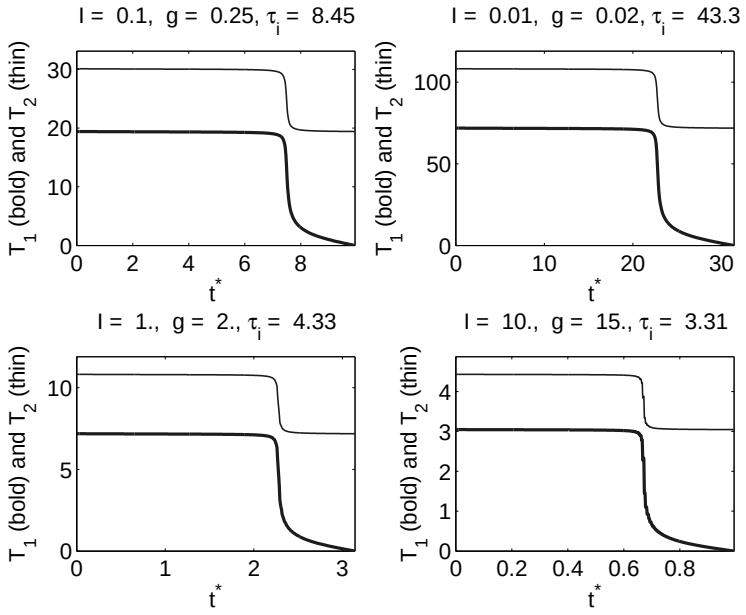


Fig. 4 *Graphs of T_1 and T_2 for various values of I and g with $\tau_i = 3\tau_0$.*

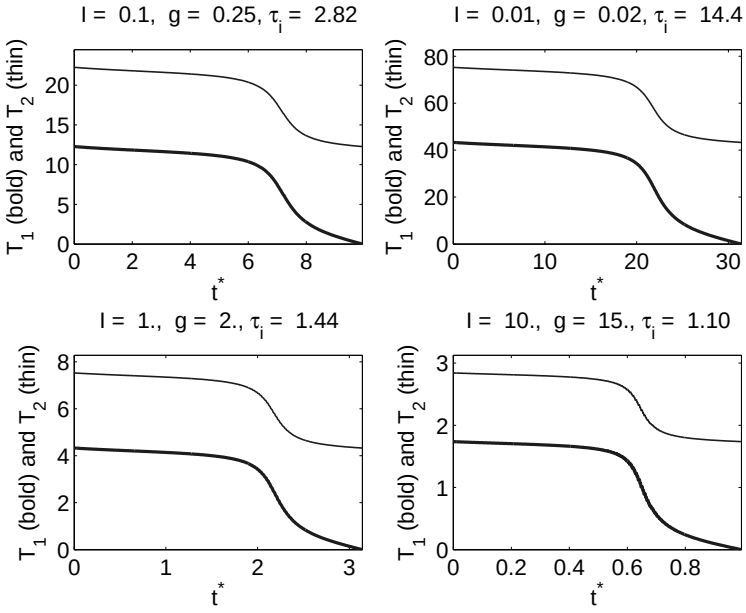


Fig. 5 *Graphs of T_1 and T_2 for various values of I and g with $\tau_i = \tau_0$.*

$$\left((1+(g-I))^2-(1-(g-I))^2\right)^{1/2} = 2(g-I)^{1/2}\,.$$

The stable quasi-steady state exists as long as

$$ge^{-(t-t^*)/\tau_i} > I\,,$$

that is, as long as

$$t - t^* < \tau_i \ln\frac{g}{I}\,.$$

This reasoning suggests (disregarding the fact that the quasi-steady state becomes less strongly attracting as the inhibition decays) that good synchronization should be expected as long as

$$\tau_i \ln \frac{g}{I} \gg \frac{1}{2(g-I)^{1/2}} \ .$$

Numerical experiments indicate, in good agreement with this reasoning, that for a very wide range of values of I, g, and τ_i, synchronization by a single pulse of inhibition is quite tight if

$$\tau_i \geq 3\tau_0, \tag{29}$$

and fairly loose if

$$\tau_i \leq \tau_0, \tag{30}$$

with

$$\tau_0 = \frac{1}{(g-I)^{1/2} \ln (g/I)} \ . \tag{31}$$

This is illustrated by Fig. 4 and 5, which show plots of T_1 and T_2 for $\tau_i = 3\tau_0$ and $\tau_i = \tau_0$, respectively, with widely varying values of I and $g > I$.

The conclusion is that synchronization will be good if τ_i is large enough in comparison with τ_0. Eq. (31) shows that τ_0 is small if g is sufficiently large in comparison with I. Note that both $g-I$ and g/I matter.

3.4 PING in E/I networks of theta neurons

Oscillations are common in networks of synaptically coupled excitatory and inhibitory theta neurons. For illustration, Fig. 6 shows a spike rastergram representing the results of a simulation of an E/I network of theta neurons. This simulation is very similar to one presented in Börgers and Kopell (2003); see Methods for the details.

At the start of the simulation, the E-cells spike asynchronously, and as a result the I-cells are gradually driven away from rest. Eventually, a population spike volley of the I-cells is triggered. Soon after this volley, the activity in the E-cells halts, resuming in near-synchrony approximately 25 ms later. The time between the spike volley of the I-cells and the resumption of spiking in the E-cells depends, in general, on the decay time constant τ_i of inhibition and, to a lesser extent, on the inhibitory conductances and external drives; see Eq. (28). The second spike volley of the I-cells, following the second spike volley of the E-cells, makes the synchrony perfect within plotting accuracy.

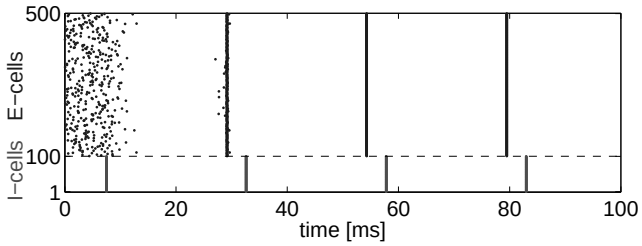


Fig. 6 *Gamma oscillation in a network of 400 excitatory and 100 inhibitory theta neurons.*

As discussed in Börgers and Kopell (2003,2005), similar rhythms, called PING (Pyramidal-Interneuronal Network Gamma) rhythms (Whittington et al., 2000), generally occur in networks of excitatory and inhibitory theta neurons whenever the E-cells spike spontaneously at a sufficiently high rate, the I-cells spike only in response to the E-cells but not on their own, and the E→I and I→E synaptic connections are sufficiently strong.

Population coding : efficiency and interpretation of neuronal activity.

C. Gielen
Dept. of Biophysics
University of Nijmegen
Netherlands

September 6, 2009

Contents

1 Introduction

A fundamental question in neuroscience concerns the understanding of the neural code. The standard doctrine is, that information is transmitted by action potentials. Since the shape and size of action potentials is almost uniform, the common believe is that the information is stored in the timing of the sequence of action potentials. In order to determine how information is represented by the nervous system, we need to understand two steps of neuronal information processing. First, we have to understand the transformation of a sensory signal into the sequence of action potentials of a single neuron. This aspect of neuronal coding has been an object of study for several decades and has resulted in a good insight into response properties of neurons in sensory pathways (e.g. the visual pathways from retina to cortex and from retina to superior colliculus, the auditory, the vestibular, and somatosensory pathways). With regard to motor control, we have to understand how neuronal activity in the final motor pathways is related to movement-related parameters, such as movement direction, movement velocity, and force. Second, there is the neurophysiological finding, that a single stimulus or movement is encoded in the activity of a large number of neurons. This has raised the question how neural activity in a population of cells can be interpreted in terms of external stimuli and actions, i.e. in terms of sensory input and motor output. This problem has been recognized since many years, but it is only in the last decade that considerable theoretical progress has been made to deal with this problem. The two problems mentioned above are crucial first steps before more complex issues like information processing and information storage in the brain can be addressed satisfactorily.

Neural encoding of information can be studied by measuring neural responses to external stimuli and during movements. Our understanding of the neural code can be tested by solving the inverse problem, inferring sensory input or motor output from a given set of neuronal activity. Solving this problem involves many other problems. For example, the evaluation of firing rate of a single cell brings several problems. In order to extract the continuous probability density of neuronal activity and its variance from the discrete spike data of a single cell, we must count the number of spikes that occur within some fixed time interval. Since a rapidly and regularly firing cell might fire some 100 spikes per second, we would need to count over at least 1 second in order to have an estimated error less than 1 percent if we had to deal with a single cell only. If the cell's firing is described by a Poisson process with an average rate of 100 spikes per second, we will need to count over a significantly longer interval to make an accurate estimate. This situation becomes even worse for chaotic firing at low firing rates.

Since the generation of actionpotentials is a stochastic process, the same sensory stimulus will never generate precisely the same neuronal activity pattern. This raises the question how the activity in a population of neurons should be interpreted. For a long time, the traditional view held that information is coded in recruitment and firing rate of neurons. However, firing rate is a continuous signal which can be obtained only by averaging over time. Obviously, averaging over time reduces the temporal resolution, which would be detremental for time-critical processes as required for sound localisation. Another solution might be averaging over many neurons. Instead of averaging over time, averaging of the activity of an ensemble of responding neurons has been proposed both to obtain accurate estimates by averaging noise in the population activity, and to combine information from many cells with different receptive fields and response properties. Ob-

viously, this approach allows to use the precise timing of each individual action potential, which has advantages not only from the point of view of temporal resolution as mentioned above, but also from a theoretical point of view since it avoids the problem of selecting the appropriate time window to determine firing rate accurately.

Theoretical considerations and an increasing body of experimental findings suggest that information is encoded not only in the recruitment and firing rate of activated neurons, but also in the temporal relations between their discharges. This becomes evident as a synchronization of firing of many neurons on a time scale of milliseconds. Therefore, synchrony of firing is thought to provide a major contribution in addition to recruitment and firing rate to code sensory and motor events.

Population codes, where information is represented in the activities of whole populations of neurons, are ubiquitous in the brain. There have been a number of theoretical analyses of population decoding in a variety of contexts. The theoretical studies often used methods that are optimal in some statistical sense usually based on probability distributions of the neuronal firing rates. However, these methods are sometimes highly implausible from a neurobiological point of view. At the other hand, experimental studies typically employed simple methods that may not be optimal from a statistical point of view, but were intuitively more clear in providing insight into the underlying mechanisms of neuronal coding and information processing.

2 Mathematics of (un)biased estimators and their variance

Before dealing with the interpretation of neuronal activity, we will first discuss various mathematical techniques to "measure" the information content of a neuronal signal. The problem, that we have to face is, how to interpret the action potentials of a large number of cells in a neuronal population as a function of time. Since the generation of an action potential is a stochastic process, we will have to rely on statistical techniques and we will have to develop probabilistic estimators. Estimators can be distinguished in *biased* and *unbiased* estimators. An unbiased estimator has the property that the output of the estimator approximates the true value of the parameter for a large number of data. One could then wonder, why people sometimes rely on unbiased estimators ? The answer is, that it may take quite some effort to obtain an unbiased estimate of some quantity and that a biased estimator may be easier to obtain.

2.1 Basic concepts from Information Theory

Figure 1 shows a schematic view of a communication channel. It consists of an Information Source, which emits a signal, which after corruption by noise, is received by a receiver, who has to estimate, as good as possible, the original message. A first question is: what measure should we use to "measure" the amount of information ? Suppose we have two independent messages x_1 and x_2 with information $f(x_1)$ and $f(x_2)$, respectively. The probability of two independent messages is the product of the probabilities of each of these two independent messages. Yet, the information adds linearly. Therefore, we require for the information function f $f(x_1 x_2) = f(x_1) + f(x_2)$. The only function, which satisfies

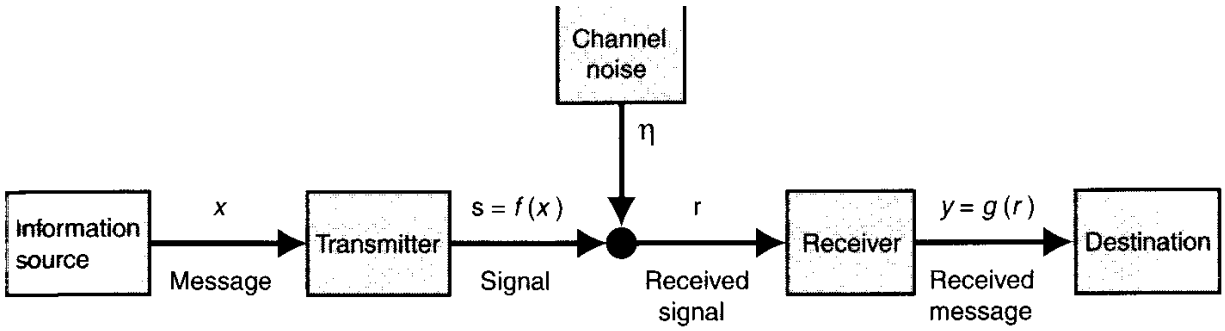


Figure 1:

this relation is the log-function. Therefore, we define the information in a signal x_i as

$$I(x_i) \;=\; -\,log_2\; p(x_i)$$

The minus sign makes the information positive as probabilities are always between zero and 1. The units of information are "bits" and the definition above also defines the units of a bit. For example, if we have a coin, with equal probabilities "up"and "down", then p("up") = p("down")=0.5. If we throw the coin, we will obtain the information $I = -log_2(0.5) = 1$, irrespective on whether the coin falls face "up" or "down". Therefore, the coin gives one bit of information.

2.1.1 Entropy

Suppose the coin is corrupted, such that the probabilities for $p("up")$ and $p("down")$ are not equal. Then the average amount of information, which we obtain after throwing the coin once is

$$S \;=\; -\sum_i p_i log_2(p_i)$$

This quantity is well known as the Entropy. The entropy is a quantity of the message set and is not defined for an individual message. The entropy of a message with N possible signals, all equally likely, is $-\sum_{i=1}^{N} \frac{1}{N} log_2 \frac{1}{N} \;=\; log_2(N)$ The entropy is hence equivalent to the logarithm of the number of possible states for equally likely states. Notice, that for N possible signals, the entropy is maximal when all signals have equal probability ! This is compatible with the physical interpretation of entropy as a measure of disorder: entropy is maximal for maximal disorder, which obviously is the case when all N signals are independent and all have equal probability.

For continuous distributions, the discrete summation is replaced by an integration:

$$S(X) \;=\; -\int_X p(x)\; log_2\; p(x)dx$$

For example, the entropy of a gaussian-distributed set with mean μ and variance σ^2, is given by

$$S \;=\; -\int_{x=-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} log_2 \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \right) \;=\; \frac{1}{2} log_2(2\pi e\sigma^2) \qquad (1)$$

Notice that the entropy, for obvious reasons, does not depend on the mean. However, it does depend on the variance σ. If σ approaches zero, the gaussian distribution approaches a δ-function and uncertainty decreases. This explains why the entropy decreases as the variance decreases. When σ is becomes too small ($\sigma < \sqrt{2\pi e}$), the entropy becomes negative, which is due to "pathologies" in the behavior of the gaussian distribution (p(x) > 1 for x close to the mean μ).

2.1.2 Mutual information

Since information coding and information transmission usually are partially corrupted by noise (see fig. 1), we have to distinguish between the information in the original signal x and in the final signal y. In fact, we want signal y to tell us as much as possible about the original signal x, such that we could reconstruct the original signal as good as possible. So except for the entropy of the signal y itself, we might wonder what the average information gain is from what a signal y tells us about the set of events that could happen.

The Entropy of a response y to a given stimulus x is

$$H_s \;=\; -\sum_y p(y|x) log_2 p(y|x)$$

Averaging over all stimuli gives the "noise-entropy"

$$H_{noise} \;=\; \sum_x p(x) H_s \;=\; -\sum_{x,y} p(x) p(y|x) log_2 p(y|x)$$

This noise-entropy relates to the response variability that is not due to the changes in the stimulus. The mutual information is obtained by subtracting H_{noise} from the response entropy:

$$I^{mutual} \;=\; H_y \;-\; H_{noise} \;=\; -\sum_y p(y) log_2 p(y) \;+\; \sum_{x,y} p(x) p(y|x) log_2 p(y|x)$$

With $p(y) = \sum_x p(x) p(y|x)$, we obtain

$$I^{mutual} \;=\; \sum_{x,y} p(x) p(y|x) log_2 \left(\frac{p(y|x)}{p(y)} \right)$$

and using $p(y,x) = p(x)p(y|x) = p(y)p(x|y)$ gives

$$I^{mutual} = \sum_{x,y} p(x,y) log_2 \left(\frac{p(x,y)}{p(y)p(x)} \right)$$

For continuous values of stimulus and response we can rewrite the mutual information as

$$I^{mutual} \;=\; \int_X \int_Y dx dy p(x,y) log_2 \frac{p(x,y)}{p(x)p(y)} \tag{2}$$

The "Mutual Information" (sometimes also called "cross-entropy") is symmetric in its arguments and describes the average amount of information that can be gained by receiving a message y when a signal x is sent (or the other way around). Notice that we can rewrite the mutual information in terms of entropies:

$$I^{mutual} \;=\; S(X) + S(Y) - S(X,Y)$$

If the two messages are completely independent, the mutual information is zero. Mutual information unequal to zero reflects a correlation between x and y. The mutual information in a communication channel describes the average amount of information we can expect to flow between input and output events.

The mutual information is related to a measure, which is used frequently in statistics and which is called the Kullback-Leibler (KL) divergence. The KL divergence provides a measure for the "distance" between two probability density functions $P(x)$ and $Q(x)$. It is defined as

$$D_{KL}(P,Q) \;=\; \int_X dx\; P(x)\; log_2\left(\frac{P(x)}{Q(x)}\right) \tag{3}$$

Notice that $D_{KL}(P,Q) \geq 0$, and $D_{KL}(P,Q) = 0$ only if the two distributions are equal: $P(x) = Q(x)$ for all x.However, unlike distance, it is not symmetric with respect to the distributions P and Q. Therefore, the KL-divergence is not a true "metric".

2.2 Maximum Likelihood and Maximum A Posteriori Estimators

A prototypical statistical problem is to estimate the value of some parameter θ from a finite set $\mathcal{X} = \{X_i\}$ of data. In the context of sensory coding, θ is a stimulus in the stimulus domain Θ, and the information about this stimulus is contained in the activities $\{r_i, i = 1, ..., N\}$ of a population of a large number of N neurons. Since θ is described as a parameter, this implies the existence of a family of probability densities $p(\mathbf{r};\theta)$ for $\theta \in \Theta$. When we make the assumption that the observations r_i are independent samples from an unknown density, then the likelihood is a product of set of conditional probability density functions of θ defined by

$$\mathcal{L}(\theta;\mathbf{r}) \;=\; p(\mathbf{r}|\theta) \;=\; \prod_i p(r_i|\theta)$$

where $p(r_i|\theta)$ represents the probability to measure neuronal activity r_i given the stimulus θ. The maximum likelihood estimator (MLE) associates to each set of data a value of $\hat{\theta}$, which maximizes $\mathcal{L}(\theta;\mathbf{r})$:

$$\hat{\theta}(\mathbf{r}) \;=\; \overset{argmax}{\theta}\; \mathcal{L}(\theta;\mathbf{r})$$

Instead of maximizing the likelyhood, it is easier to find the maximum of the logarithm of the likelyhood since the logarithm maps the product of conditional probabilities into a sum of logarithms of conditional probabilities. The maximum likelyhood is then found by solving the equation

$$\nabla_\theta\; log\; \mathcal{L}(\theta;\mathbf{r}) \;=\; \nabla_\theta \sum_{i=1}^{N} log\; p(r_i|\theta) \;=\; 0$$

Therefore, the MLE is the value of θ that maximizes the likelyhood $p(\mathbf{r}|\theta)$.

According to Bayes' relation the posterior density of θ can be found by

$$p(\theta|\mathbf{r}) \;\propto\; p(\mathbf{r}|\theta)p(\theta) = \mathcal{L}(\theta;\mathbf{r})p(\theta)$$

The maximum a posteriori (MAP) estimator of θ maximizes $p(\theta|\mathbf{r})$, or equivalently, $\mathcal{L}(\theta;\mathbf{r})p(\theta)$. Thus MLE is a MAP estimator for the "flat" prior over $p(\theta)$.

2.3 Fisher Information

A natural framework to study how neurons communicate, or transmit information, in the nervous system is information theory. Suppose we are collecting data and we know that each data sample comes from either of two distributions $f_1(x)$ and $f_2(x)$. (The number of 2 can easily be extended to any arbitrary number of distributions.) Define $I(1:2)$ as the average information per observation or sample from distribution $f_1(x)$ in favor of discrimination for the hypothesis that the sample is from f_1, against the hypothesis that the sample is from f_2. By this definition, $I(1:2)$ is a measure for the average information which is available to decide in favour of class 1 relative to class 2, given a sample from class 1. This definition gives

$$I(1:2) \;=\; \int f_1(x) log \frac{f_1(x)}{f_2(x)} dx$$

It is easy to see, that $I(1:2)$ approaches infinity when the two distributions are disjunct, and that $I(1:2)$ is zero, when $f_1(x) = f_2(x)$. Also note, that $I(1:2) \geq 0$ at all times. There are some conditions, which may seem pathological. Take for example the distributions f_1, which is defined as $f_1(x) = 1/x_1$ for $0 \leq x \leq x_1$ and zero otherwise, and f_2, which is defined as $f_2(x) = 1/x_2$ for $0 \leq x \leq x_2$ and zero otherwise, with $x_2 > x_1$. In that case f_1, $I(1:2) = log \frac{x_2}{x_1}$, but $I(2:1)$ does not exist. This can be remedied easily by setting some requirements, for example by requiring that a distribution $f(x) \neq 0$, which is true for a gaussian distribution.

We now define the Divergence

$$J(1,2) \;=\; I(1:2) \;+\; I(2:1) \;=\; \int (f_1(x) - f_2(x)) log \frac{f_1(x)}{f_2(x)} dx$$

Notice that

- $J(1,2) = J(2,1) \geq 0$
- $J(1,1) = 0$.
- triangle inequality does not hold.

$J(1,2)$ is a measure for the divergence between the hypotheses H_1 and H_2 and it is a measure for the difficulty to discriminate between H_1 and H_2.

Now suppose that we receive data from a neuron and we have to decide whether the neural response codes stimulus θ or stimulus $\theta + \Delta\theta$. Then

$$\begin{aligned}
J(\theta, \theta+\Delta\theta) &= \int (f(\theta) - f(\theta+\Delta\theta)) log \frac{f(\theta)}{f(\theta+\Delta\theta)} d\theta && (4)\\
&= \int (f(\theta+\Delta\theta) - f(\theta)) log \frac{f(\theta+\Delta\theta)}{f(\theta)} d\theta && (5)\\
&\approx \int \frac{\partial f}{\partial \theta} \Delta\theta log \frac{f(\theta) + \Delta\theta \frac{\partial f}{\partial \theta}}{f(\theta)} d\theta && (6)\\
&= \int \frac{\partial f}{\partial \theta} \Delta\theta log (1 + \frac{\frac{\partial f}{\partial \theta}}{f(\theta)} \Delta\theta) d\theta && (7)
\end{aligned}$$

$$\approx \int f(\theta) \left(\frac{\frac{\partial f}{\partial \theta}}{f(\theta)} \right)^2 (\Delta\theta)^2 d\theta \quad (8)$$

$$= \int f(\theta) \left(\frac{\partial log f(\theta)}{\partial \theta} \right)^2 (\Delta\theta)^2 d\theta \quad (9)$$

$$= -\int f(\theta) \frac{\partial^2 log f(\theta)}{\partial \theta^2} (\Delta\theta)^2 d\theta \quad (10)$$

The latter step assumes that the function $f(\theta)$ is symmetric (see below). This defines the Fisher Information, which is defined by $\int f(\theta) \left(\frac{\partial log f(\theta)}{\partial \theta} \right)^2 d\theta$ or, which is equivalent, by $-\int f(\theta) \frac{\partial^2 log f(\theta)}{\partial \theta^2} d\theta$ (see below).

Suppose we have a set of N neurons, whose activity is represented by the vector $\mathbf{r}$. This vector $\mathbf{r}$ then codes a specific signal θ. The Fisher information is a functional of $p(\mathbf{r}|\theta)$ and can be interpreted as the amount of information in $\mathbf{r}$ about the stimulus θ. The Fisher information is defined by

$$J[\mathbf{r}](\theta) = E[-\frac{\partial^2}{\partial \theta^2} log\ p(\mathbf{r}|\theta)] = \int d\mathbf{r} p(\mathbf{r}|\theta) \left(-\frac{\partial^2 log p(\mathbf{r}|\theta)}{\partial \theta^2} \right) \quad (11)$$

Note that, if the additional assumption is made that the probability function $p(\mathbf{r}|\theta)$ is symmmetric, like for a gaussian function, the Fisher Information can also be written as

$$J[\mathbf{r}](\theta) = \int d\mathbf{r} p(\mathbf{r}|\theta) \left(\frac{\partial log p(\mathbf{r}|\theta)}{\partial \theta} \right)^2$$

This follows easily from the following:

$$\int p(\mathbf{r}|\theta) \left(\frac{\partial log p(\mathbf{r}|\theta)}{\partial \theta} \right)^2 d\theta = \int p(\mathbf{r}|\theta) \frac{1}{p(\mathbf{r}|\theta)^2} \left(\frac{\partial p(\mathbf{r}|\theta)}{\partial \theta} \right)^2 d\theta$$

Starting from the other equation gives

$$\int d\mathbf{r} p(\mathbf{r}|\theta) \left(-\frac{\partial^2 log p(\mathbf{r}|\theta)}{\partial \theta^2} \right) = -\int d\mathbf{r} p(\mathbf{r}|\theta) \frac{\partial}{\partial \theta} \left(\frac{1}{p(\mathbf{r}|\theta)} \frac{\partial p(\mathbf{r}|\theta)}{\partial \theta} \right) \quad (12)$$

$$= -\int d\mathbf{r} p(\mathbf{r}|\theta) \left[-\frac{1}{p(\mathbf{r}|\theta)^2} \left(\frac{\partial p(\mathbf{r}|\theta)}{\partial \theta} \right)^2 + \frac{1}{p(\mathbf{r}|\theta)} \frac{\partial^2 p(\mathbf{r}|\theta)}{\partial^2 \theta} \right] \quad (13)$$

$$= \int d\mathbf{r} p(\mathbf{r}|\theta) \frac{1}{p(\mathbf{r}|\theta)^2} \left(\frac{\partial p(\mathbf{r}|\theta)}{\partial \theta} \right)^2 + \int d\mathbf{r} \frac{\partial^2 p(\mathbf{r}|\theta)}{\partial^2 \theta} \quad (14)$$

With the assumption that the probability function $p(\mathbf{r}|\theta)$ is a symmetric function of θ , the second term is zero, which proves that both definitions are identical for symmetric functions.

The Fisher information is a measure of the expected curvature of the log likelihood at the stimulus θ. Curvature is important because the likelihood is expected to be at a maximum near the true stimulus value θ that caused the responses. If the likelihood is

very curved, and thus the Fisher information is large, responses typical for the stimulus are much less likely for slightly different stimuli. If the log-likelihood is very flat, and thus the Fisher information is small, responses common for the stimulus θ are likely to occur for for different stimuli as well.

It is important to stress, that the Fisher information itself is not an information quantity. Rather, the Fisher information gives a measure for the accuracy to discriminate between different values of the stimulus near θ given the signals $\mathbf{r}$. The terminology comes from an intuitive interpretation of the bound: our knowledge ("information") about a stimulus θ is limited according to this bound.

Because the generation of actionpotentials by each neuron is an independent process, the responses r_i can be assumed to be independent. As a result $J[\mathbf{r}](\theta) = \sum_{i=1}^{N} J[r_i](\theta)$ so, that J is of the order of N, implying that the typical fluctuations of the ML estimate scale as $N^{-\frac{1}{2}}$. This is in contrast to the bias of the ML-estimate, which is of the order of N^{-1}. Hence, the variance is the dominant contribution to the error in the estimate in the limit for large N.

One of the reasons of the importance of the Fisher Information in neuronal information processing is found in the Cramer-Rao inequality, which states that the Fisher information $J(\theta)$ provides a lower bound for the mean squared error of any unbiased estimator:

$$\left\langle (\hat{\theta} - \theta)^2 \right\rangle \geq \frac{1}{J[\mathbf{r}](\theta)}.$$

This means that the Fisher information is a measure of how well one can estimate a parameter from an observation with a given probability distribution. Since the variance of the MLE approaches the inverse of the Fisher information, the MLE is asymptotically optimal.

2.4 Mutual information and Fisher Information

Consider an observable $\mathbf{r}$ and some stimulus θ. The information about the stimulus θ in the observable (response) $\mathbf{r}$ is given by

$$\int d^N\mathbf{r}\ p(\mathbf{r}|\theta) log \frac{p(\mathbf{r}|\theta)}{p(\mathbf{r})}$$

A frequently used measure to express the information contained in two parameters, is the mutual information, which is the only quantity (up to a multiplicative constant) satisfying a set of fundamental requirements. For an observable $\mathbf{r}$ and a stimulus θ, the mutual information is defined by

$$I(\theta, \mathbf{r}) = \int d\theta d^N\mathbf{r}\ p(\theta)p(\mathbf{r}|\theta)\ log\ \frac{p(\mathbf{r}|\theta)}{p(\mathbf{r})} \tag{15}$$

and can also be defined as the average information in $\mathbf{r}$ over all stimuli θ.

The mutual information is closely related to the concept of entropy. Entropy is a measure for the information required to code a variable with a certain probability distribution by characterizing how many states it can assume and the probability of each. The entropy $H(\theta) = -\int d\theta\ p(\theta)\ log\ p(\theta)$ corresponds to the number of bits required to specify all

stimuli. Similarly, the entropy $H(\mathbf{r}) = -\int d^N \mathbf{r}\ p(\mathbf{r})\ log\ p(\mathbf{r})$ corresponds to the number of bits required to specify all possible neuronal responses. The entropy in the neural response $\mathbf{r}$ given the stimulus θ is defined by $H(\mathbf{r}|\theta) = \int d\theta d^N \mathbf{r}\ p(\theta)p(\mathbf{r}|\theta)\ log\ p(\mathbf{r}|\theta)$. The mutual information, which is the information about the stimulus preserved in the neural response, is given by $H(\theta) - H(\theta|\mathbf{r}) = H(\mathbf{r}) - H(\mathbf{r}|\theta)$, which is equivalent to Eq. 15 .

Suppose there exists an unbiased efficient estimator $\hat{\theta}$ with mean θ and minimal variance (according to Cramer-Rao !) equal to the inverse of the Fisher matrix $J(\theta)$. With the definitions given above, the mutual information (i.e. the amount of information gained about θ in the computation of the estimate $\hat{\theta}$) is

$$I(\theta, \hat{\theta}) = -\int d\hat{\theta} p(\hat{\theta})\ log\ p(\hat{\theta}) + \int d\theta \rho(\theta) \int d\hat{\theta} p(\hat{\theta}|\theta)\ log\ p(\hat{\theta}|\theta).$$

This is truly an information metric, since the first term represents the entropy of the estimator $\hat{\theta}$ and $I(\theta, \hat{\theta})$ represents the gain of information about θ in the computation of that estimator. The term $-\int d\hat{\theta} p(\hat{\theta})\ log\ p(\hat{\theta})$ is the entropy given $\hat{\theta}$, which for each θ is smaller than the entropy of a gaussian distribution with the same variance $J^{-1}(\theta)$. Since processing cannot increase information, the information $I(\theta, \mathbf{r})$ conveyed by $\mathbf{r}$ about θ is at least equal or greater than that conveyed by the estimator. This gives

$$I(\theta, \mathbf{r}) \geq I(\theta, \hat{\theta}) \geq -\int d\theta p(\theta)\ log\ p(\theta) - \int d\theta \rho(\theta) \frac{1}{2} \left(\frac{2\pi e}{J(\theta)} \right) \qquad (16)$$

where the last term at the right hand side follows straightforward for a gaussian distribution with variance $J^{-1}(\theta)$ (see also Eq. 1).

When the distribution of the estimator is sharply peaked around its mean value (which implies $J(\theta) \gg 1$) the entropy of the estimator becomes identical to the entropy of the stimulus. When the estimator has a non-gaussian distribution, the inequality will be strict.

3 Probabilistic interpretation of population codes.

The starting point for almost all work on neural population codes is the neurophysiological finding that many neurons respond to a particular variable underlying a stimulus (such as the sensitivity of neurons in visual cortex to the orientation of a luminous line; see fig 2) according to a unimodal tuning function. For neurons involved in sensory perception, the set of variables, which affect the response of a neuron, is usually referred to as the receptive field. However, for neurons involved in movements a better terminology would be "movement field". Cells in motor cortex have as "preferred movement direction": they show the largest firing rates for movements in a particular direction. The firing rate seems to decrease with the cosine of the angle between movement direction and the cell's preferred movement direction (see figure 3). In order to summarize both types of neurons, and especially neurons in the sensory-motor pathway where neural responses have both sensory and motor components, we will use the term "response field". The value or set of values of the variables underlying the response field, which produce a peak in the tuning function, will be called the "preferred value".

The response field plays an important role in interpreting neuronal population codes. For many brain structures, the response fields of neurons are not known. Only for neurons

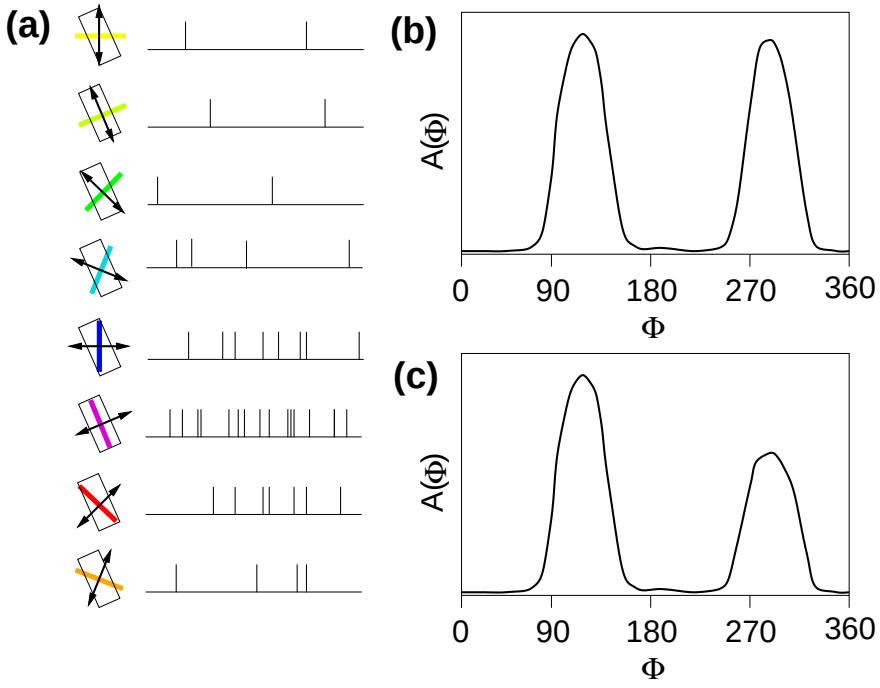


Figure 2: Orientation and direction preference of cortical cells. (a) Oriented bars moving across the receptive fields (black boxes) of neurons evoke response which are stronger when stimulating with the preferred orientation of the nerve cell (see examples of spike trains on the right). (b) The rate A of one neuron in dependence of the stimulus orientation Φ yields the tuning curve of the neuron. The response in (b) is only orientation selective, while the response in (c) displays direction selectivity.

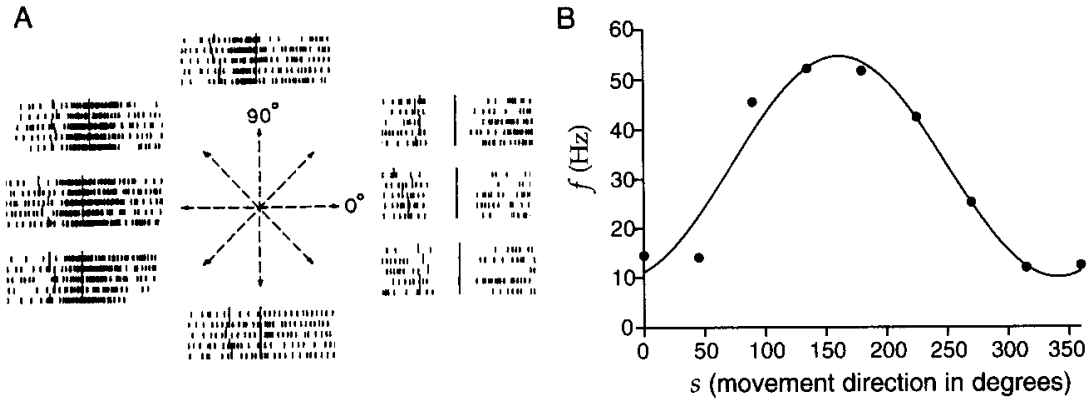


Figure 3: A. Recordings from cells in motor cortex in an arm reaching task. The hand of the monkey started from the middle target location and the monkey was instructed to make movements to each of the 8 surrounding targets, at 45 deg intervals. The rasters for each movement direction show action potentials fired on five trials. B. Average firing rate plotted as a function of movement direction.

in rather peripheral sensory pathways (such as retina, Lateral Geniculate Nucleus, area V1 in visual cortex) of motor pathways (for example motor cortex), it is possible to determine the response field. However, for neurons in more central brain structures, the relevant sensory and motor features, which underly the response field, may be very hard to discover.

Several authors have used gaussian white noise as stimulus. The reason for using white noise is that the characteristics of a dynamical system are hard to determine, because what happens now depends on what happened before. Thus all possible stimuli and neural responses have to be considered for a full characterization of the system. The use of Gaussian-White-Noise (GWN) stimuli is attractive, since a GWN-signal has the largest entropy given a particular variance and as such contains all possible combinations of stimulus values in space and time.

As a first order (linear) approximation, the response field $R_i(t)$ of neuron i can be defined by the crosscorrelation of the gaussian white noise stimulus $\mathbf{x}(t)$ and the neuronal response $r_i(t)$. This crosscorrelation can be shown to be equal to the averaged stimulus preceding an action potential or to the averaged response following a spike. We will refer to this as the averaged peri-spike-event:

$$
\begin{aligned}
R_{PSE}(\tau) &= \frac{1}{2T}\int_{-T}^{T}\mathbf{x}(t-\tau)r_i(t)dt && (17)\\
&= \frac{1}{2T}\int_{-T}^{T}\mathbf{x}(t-\tau)\sum_n \delta(t-t_n)dt && (18)\\
&= \sum_n \frac{1}{2T}\mathbf{x}(t_n-\tau) && (19)
\end{aligned}
$$

where the response $r_i(t)$ of neuron i is represented by a sequence of δ-pulses and where t_n is the time of occurrence of action potential n. As we will see later, this crosscorrelation technique can provide a first step to characterize the conditional probability $p(\mathbf{r}|\theta)$. However, for neurons with complex properties, the complexity of the GWN stimulus increases exponentially with the number of dimensions of the stimulus. Therefore, this approach to characterize the response field is only useful for neurons with simple, low dimensional response fields.

The characteristic properties of the response field can provide information to answer the question "How is an external event $\mathbf{x}(t)$ in the world encoded in the neuronal activity $\mathbf{r}(t)$ of the cells". A full characterization of the response field of a neuron (both spatial and temporal properties !) implies that the density function $p(\mathbf{r}|\theta)$ is known. The response fields are also indispensible for answering the question about the sensory or motor interpretation of neural activity. The response fields allow the mapping from the set of activities in a neural population $\mathbf{r}(t)$, with $r_i(t)$ representing the activity of neuron i at time t, to the events in the external world by Bayes' relation: $p(\theta|r_i) = \frac{p(\theta)p(r_i|\theta)}{p(r_i)}$.

Since the generation of action potentials is a stochastic process, the problems described above have to be addressed in a probabilistic way. We will define $p(\mathbf{r}|\mathbf{x})$ as the probability for the neuronal activity $\mathbf{r}$ given the stimulus $\mathbf{x}$. The simplest models assume that neuronal responses are independent, which gives $p(\mathbf{r}|\mathbf{x}) = \prod_i p(r_i|\mathbf{x})$. For the time being, we will

assume independence of firing. The case of correlated firing between neurons will be discussed later. A Bayesian decoding model specifies the information in $\mathbf{r}$ about $\mathbf{x}$ by

$$p(\mathbf{x}|\mathbf{r}) \;\propto\; p(\mathbf{r}|\mathbf{x})p(\mathbf{x}) \tag{20}$$

where $p(\mathbf{x})$ gives the prior distribution about $\mathbf{x}$. Note that starting with a specific stimulus $\mathbf{x}$, encoding it in the neural activity $\mathbf{r}$, and decoding it results in a probability distribution over $\mathbf{x}$. This uncertainty arises from the stochasticity of the spike generating mechanism of neurons and from the probability distribution $p(\mathbf{x})$.

4 Models for population codes

4.1 Simple version of Population Coding

The most simple and straightforward interpretation of neuronal population activity is obtained by simple summation of the response fields $\mathbf{R}_i$ of all neurons i, weighted by the firing rate r_i of each neuron:

$$\mathbf{x}_{est} \;=\; \frac{\sum_{i=1}^{N} r_i(\mathbf{x})\mathbf{R}_i}{\sum_{i=1}^{N} r_i(\mathbf{x})} \tag{21}$$

This choice corresponds to the so-called center-of-gravity estimate. Center-of-gravity coding can be statistically optimal. This is the case for perfectly regular arrays of sensors with gaussian tuning profiles that have an output described by independent Poisson statistics, and for arrays of sensors with a sinusoidal tuning profile for the parameter estimated. However, there are many cases in which center-of-gravity decoding is highly inefficient. This includes the important case (which is observed at nearly all parts of the brain), where sensor positions or response fields are not regularly spaced. We will come back on this topic later. Moreover, the center-of-gravity approach assumes a homogeneous distribution of response fields in the event space and a homogeneous distribution of stimuli $\mathbf{x}$ for sensory neurons. Given these assumptions, any deviations between the center-of-gravity result and the true parameter value are small provided that the noise is small and that the neurons sample the parameter space sufficiently dense. Moreover, the question arises, whether the estimate of this population coding scheme is optimal in the sense that it is unbiased and that the variance in its estimate is small. A good estimator should be unbiased, which is the case when the estimator gives the (expectation value of the) true stimulus $\mathbf{x}$. The center-of-gravity method is virtually bias-free. However, this simplistic version of the population vector is inefficient in the sense that the variance of the estimate is much larger than the smallest possible variance.

One of the first experimental data demonstrating the importance of the concept of population coding were obtained from motor cortex. Neurons in the arm area of primate motor cortex in monkey are broadly tuned in the sense that they increase firing rate for a broad range of arm movement directions (see figure 3). Each neuron appears to have a preferred movement direction (i.e. the movement direction, which corresponds to the largest response modulation of the neuron) and preferred movement directions are approximately homogeneously distributed in 3-D space. In the literature the population

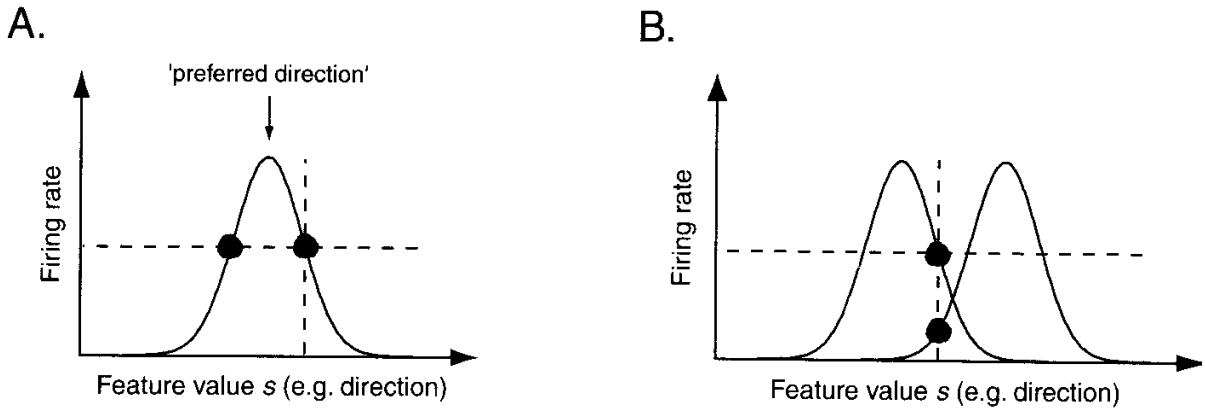


Figure 4: Gaussian tuning curves representing the firing rate of a neuron as a function of a stimulus feature. A. Single neuron cannot ambiguously decode the stimulus feature from the firing rate. B. A second neuron with shifted tuning curve can resolve the ambiguity.

activity has been interpreted as

$$\mathbf{M}(\mathbf{r}) \;=\; \sum_{i=1}^{N} r_i \mathbf{M}_i$$

where $\mathbf{M}_i$ is the preferred movement direction of neuron i and where $\mathbf{M}$ represents the estimated movement direction of the arm. Quite remarkably, the estimated movement direction by the population vector was very close to the actual measured movement direction of the monkey's arm.

For a simple array of N independent sensors with unit spacing between consecutive sensors and with a gaussian tuning function $f_n(\theta) \;=\; exp\left[-\frac{1}{2}\left(\frac{n-\theta}{\sigma}\right)^2\right]$ and with gaussian noise W_n superimposed on the response of neuron n ($R_n \;=\; f_n(\theta) \;+\; W_n$), the Fisher Information is given by $\frac{1}{N^2}\sum_n \left(\frac{\partial f_n(\theta)}{\partial \theta}\right)^2$. According to the Cramer-Rao bound, the minimal variance is given by $\frac{N^2}{\sum_n (f_n'(\theta))^2}$. When the summation is replaced by integration, which is a good approximation for large N and sufficiently large σ, the minimal variance reduces to $\frac{2\sigma N^2}{\sqrt{\pi}}$. Note, that the minimum attainable variance increases with the sensor tuning width σ, a result which is similar to the Maximum Likelihood result (see section 4.2 and Fig. 1). Also notice, that the Fisher Information for this neuron is proportional to the derivative of the receptive field. This can be understood from the following: when the slope is steep, a small change in e.g. stimulus orientation (for visual cortical cells) causes large changes in firing rate. Therefore, most of the information about the orientation of the visual stimulus is coded by neurons, which have the optimal tuning just neighboring to the orientation of the stimulus.

The results above show, that the minimal variance of the center-of-mass model is proportional to σ, i.e. the minimal variance increases as a function of the tuning width σ. Hence, it is advantageous to use narrowly tuned sensors. If we compare the variance

of the center-of-gravity model with that of the Cramer-Rao lower bound, we obtain

$$\frac{Var(\theta_{CR})}{Var(\theta_{CG})} \leq \frac{6\sqrt{\pi}\sigma^3}{\frac{N-1}{2}\frac{N+1}{2}\frac{N+3}{2}}$$

This illustrates that the efficiency of the center-of-gravity coding is low when the number of neurons is large. This is easily explained. When the number of neurons is large relative to the tuning width, many neurons do not respond to a stimulus, but do contribute to the population average by their noise, since sensor noise is independent of the response. Therefore, neurons, which do not respond to the stimulus, do contribute to the noise in the population average.

The analysis so far was for regular arrays of neurons. It can be shown that when the receptive fields of neurons are highly irregularly distributed, the largest contribution to errors in the center-of-gravity method originate from these irregularities, rather than from neuronal noise. As we will show below, the ML-estimate does not suffer from irregularities. Some linear estimators have been proposed which do not suffer from irregularities in the distribution of receptive fields either. However, these models do come at a price. The regular center-of-gravity estimator only needs to know the optimal stimulus parameter, whereas the models, that have been proposed to compensate for irregularities in distribution, also require knowledge of the distribution of neuronal tuning or overlap of tuning functions to invert a covariance matrix of neuronal activities .

4.2 Poisson model

Under the Poisson encoding model, the neuronal activities $r_i(t)$ are assumed to be independent with

$$p(r_i|\mathbf{x}) \;=\; e^{-f_i(\mathbf{x})}\frac{(f_i(\mathbf{x}))^{r_i}}{r_i!}$$

where $f_i(\mathbf{x})$ is the tuning function for neuron i and where $r_i(t)$ represents the firing rate or the number of action potentials in a particular time interval.

With regard to decoding, several studies have used Maximum Likelyhood (ML) for the Poisson encoding model. The ML estimate gives the stimulus $\mathbf{x}$, which maximizes the likelyhood $p(\mathbf{r}|\mathbf{x})$. It is defined as:

$$\mathbf{x}_{ML} \;=\; \overset{argmax}{\mathbf{x}}\; p(\mathbf{r}|\mathbf{x})$$

The ML estimate can be obtained by differentiating the logarithm of the response probability distribution

$$\frac{\partial\ log\ p(\mathbf{r}|x)}{\partial x} \;=\; \sum_n \frac{\partial\ log\ p(r_n|x)}{\partial x} \;=\; \sum_n \left[\frac{f_n'(x)}{f_n(x)}r_n \;-\; f_n'(x)\right] \tag{22}$$

For neurons with a gaussian tuning profile $f_n(\theta) \;=\; exp\left[-\frac{1}{2}\left(\frac{\theta_n-\theta}{\sigma}\right)^2\right]$ and with a regular, homogeneous distribution, the ratio $\frac{f_n'(\theta)}{f_n(\theta)}$ equals $(\theta_n-\theta)/\sigma^2$. For sufficiently dense neuron distributions, Eq. 22 reduces to $\frac{1}{\sigma^2}\sum_n(\theta_n-\theta)r_n$. The optimal estimate is obtained when the derivative in Eq. 22 is set to zero, which gives

$$\hat{\theta}_{ML} \;=\; \frac{\sum_n \theta_n r_n}{\sum_n r_n}$$

This result is identical to the center-of-gravity estimate for a regular homogeneous array of neurons. It illustrates that for a regular, homogeneous distribution of neurons with gaussian tuning functions and independent Poisson noise, the center-of-gravity method is optimal from a statistical point of view.

The full probability distribution over the quantity $\mathbf{x}$ from this Poisson model is

$$p(\mathbf{x}|\mathbf{r}) \;\propto\; p(\mathbf{x}) \prod_i e^{-f_i(\mathbf{x})} \frac{(f_i(\mathbf{x}))^{r_i}}{r_i!}$$

For independent noise between the neurons finding the ML estimate implies maximization of the likelyhood $p(\mathbf{r}|\mathbf{x})$. For a large number of neurons, the estimate is unbiased and the variance is given by $E[(\mathbf{x}_{est}-\mathbf{x})^2] \;=\; \frac{1}{J[\mathbf{r}](\mathbf{x})}$, where $J[\mathbf{r}](\mathbf{x})$ is the Fisher information as defined in Eq. 11. With the assumption of independent noise across units, the expression for the Fisher information becomes

$$J[\mathbf{r}](\mathbf{x}) \;=\; \sum_{i=1}^{N} E[-\frac{\partial^2}{\partial \mathbf{x}^2} log\ p(r_i|\mathbf{x})]$$

where $E[..]$ refers to the expectation value of the argument.

When the stochastic behaviour of neuronal firing is modeled by normally distributed noise on the response with variance σ^2 ($p(r_i|x) \propto exp\left(-\frac{(r_i-f_i(x))^2}{2\sigma^2}\right)$) (i.e., when $p(r_i|x) \propto exp\left(-\frac{(r_i-f_i(x))^2}{2\sigma^2}\right)$, then the Fisher information matrix is given by

$$J[\mathbf{r}](\mathbf{x}) \;=\; \frac{\sum_{i=1}^{N} f_i'(\mathbf{x})^2}{\sigma^2} \tag{23}$$

where $f_i'(\mathbf{x}) \;=\; \frac{\partial f_i(\mathbf{x})}{\partial \mathbf{x}}$. This follows from

$$\begin{aligned} E\left[-\frac{\partial^2}{\partial x^2} logp(r_i|x)\right] &= E\left[-\frac{\partial}{\partial x}\frac{(r_i - f_i(x))}{\sigma^2} f_i'(x)\right] \\ &= E\left[-\frac{r_i - f_i(x)}{\sigma^2} f_i''(x) + \frac{(f_i')^2(x)}{\sigma^2}\right] \\ &= \frac{(f_i')^2(x)}{\sigma^2} \end{aligned}$$

For Poisson distributed noise the Fisher information matrix for the MLE is given by

$$J[\mathbf{r}](\mathbf{x}) \;=\; \sum_{i=1}^{N} \frac{f_i'(\mathbf{x})^2}{f_i(\mathbf{x})} \tag{24}$$

This follows from

$$\begin{aligned} J[\mathbf{r}](\mathbf{x}) &= \sum_{n=1}^{N} E\left[-\frac{\partial^2}{\partial x^2} logp(r_n|x)\right] \\ &= \sum_{n=1}^{N} E\left[-\frac{\partial}{\partial x}\left(\frac{f_n'(x)}{f_n(x)} r_n - f_n'(x)\right)\right] \\ &= \sum_{n=1}^{N} E\left[-\frac{f_n''(x)}{f_n(x)} r_n + \frac{(f_n')^2(x)}{f_n^2(x)} r_n + f_n''(x)\right] = \sum_{n=1}^{N} \frac{f_n'(\mathbf{x})^2}{f_n(\mathbf{x})} \end{aligned}$$

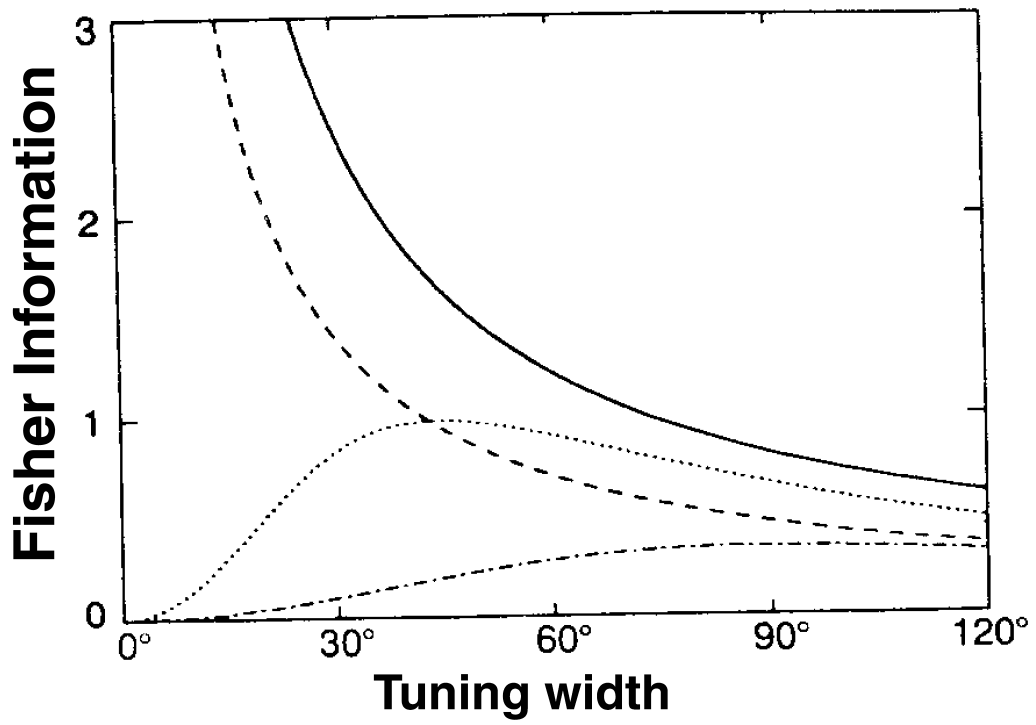


Figure 5: Fisher information (J/Nf_{max}) for the population of neurons with tuning functions according to Eq. 25 for the ML-estimate (solid line and broken line) and for the population vector (dotted line and dashed-dotted line) for ratio's of 0.1 (dashed line and dashed-dotted line) and 0.01 (solid line and dotted line) of f_{min} and $f_{min} + f_{max}$ as a function of tuning width a in degrees.

The Cramer-Rao inequality states that the average squared error for an unbiased estimator is greater than or equal to the inverse of the Fisher Information. Hence, the ML estimator is asymptotically optimal for the Poisson model, since its variance approximates the lower bound for a large number of neurons.

These ideas are illustrated in Figures 5 and 6, which show the Fisher information (the inverse of the variance in the ML estimate) for a hypothetical population of neurons in visual cortex. Each neuron is thought to have an optimal orientation sensitivity θ_i and the mean response of neuron i to a stimulus θ is given by

$$f(\theta - \theta_i) = \begin{cases} f_{min} + (f_{max} - f_{min})cos^2(\frac{\pi}{a}(\theta - \theta_i)) & \text{if } |\theta - \theta_i| < a/2 \\ f_{min} & \text{otherwise} \end{cases} \quad (25)$$

where a is the width of the receptive field of the neuron. When the stimulus θ is close to the preferred direction of the neuron, the probability of a large response is high. When the stimulus is outside the receptive field, the response is small with mean firing rate f_{min}. For the ML-estimator the Fisher information (Eq. 24) is proportional to $Nf_{max}a^{-1}$, which demonstrates that the Fisher information diverges when the width a approaches zero. The Fisher information for the ML-estimator decreases gradually for larger values of a, approaching the value zero (infinite variance !) for very large values of a.

For the population vector model with the tuning function according to Eq. 25, the

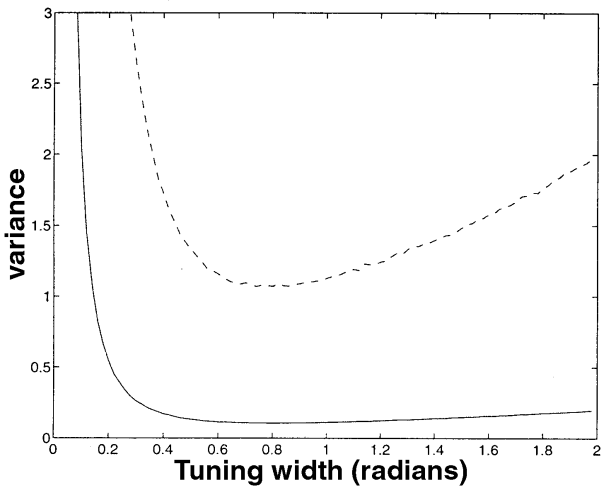


Figure 6: Figure 6 shows the variance for the population vector estimate for number of neurons N= 10^3 (dashed line) and for N=10^4 (solid line) as a function of the tuning width of the same set of neurons as in Fig. 5. Note that variance is related to the inverse of the Fisher Information.

variance (i.e. the inverse of the Fisher information) is given by $\frac{\bar{f}_o - \bar{f}_2}{2Nf_1^2}$, where $\bar{f}_n$ is the n-th Fourier component defined by $\bar{f}_n = \frac{1}{2\pi}\int_0^{2\pi} e^{in\theta} f(\theta) d\theta$. Obviously, the Fisher information for the population vector model mainly depends on the width a of the tuning function and on the background noise f_{min}. For small values of a, the Fisher information is zero, increases with a reaching a maximal value for finite values of a (see Fig. 5) after which the Fisher information decreases for larger values of a. It can be shown that the optimal width a_{max} is proportional to the ratio of background activity f_{min} to peak activity $(f_{min} + f_{max})$ to the power 1/3.

For the simple population vector (center of gravity vector), the Fisher information is zero for very small and very large values of a and therefore, the variance is infinity. This can be understood from the fact, that for small receptive fields a most neurons are below threshold and contribute noise with variance f_{min} without contributing to the signal. In contrast, the Fisher information increases for smaller values of a for ML, because ML is based on the gradient of the response (see Eq. 24), which approaches infinity for small a. As the tuning curve becomes more narrow, the increase in signal $|f'|$ more than offsets the decrease in the number of neurons above threshold. In addition, the number of neurons below threshold are completely ignored by the ML estimator. Both for ML and the population vector, the information decreases (and the variance increases) for large receptive fields, since for large receptive fields, a single stimulus will excite many neurons by the same amount, such that an accurate discrimination between responses of different neurons becomes impossible.

The ML model has several problems. First of all, the ML estimator assumes that there is one single stimulus $\mathbf{x}$ (for example one single visual bar at a given orientation for neurons in V1) which caused the neuronal activity. If multiple stimuli were present, the Poisson model will fail. Moreover, sometimes the estimation of the optimal decoding may

require the whole probability distribution $p(\mathbf{x}|w)$ over all values of the variable $\mathbf{x}$, where w represents all available information. The Poisson model will not be able to provide such a distribution in many cases. For example, when the tuning function $f_i(\mathbf{x})$ is gaussian with an optimal stimulus $\mathbf{x}_i$ for neuron i, then

$$\begin{aligned} log\ p(\mathbf{x}|\mathbf{r}) &\propto log\left[p(\mathbf{x})\prod_i e^{-f_i(\mathbf{x})}\frac{(f_i(\mathbf{x}))^{r_i}}{r_i!}\right] && (26)\\ &= C_1 - \sum_i f_i(\mathbf{x}) - \frac{1}{2\sigma^2}\sum_i r_i(\mathbf{x} - \mathbf{x}_i)^2 && (27)\\ &= C_2 - \frac{1}{2}\left(\frac{\sum_i r_i}{\sigma^2}\right)\left(\mathbf{x} - \frac{\sum_i r_i\mathbf{x}_i}{\sum_i r_i}\right)^2. && (28)\end{aligned}$$

This distribution has a mean $\mu = \frac{\sum_i r_i\mathbf{x}_i}{\sum_i r_i}$ and a variance $\frac{\sigma^2}{\sum_i r_i}$. Taking the mean of the distribution would give a single value, which is the same as that of the centre-of-gravity estimate, even in the case when the neuronal response was elicited by multiple stimuli. Therefore, the distribution of $p(\mathbf{x}|\mathbf{r})$ for the Poisson model for this model with gaussian tuning curves is unimodal. In addition, the variance will always be smaller than the variance of the gaussian tuning function, since $\sum_i r_i \geq 1$ for reasonably effective sets of stimuli. Thus the Poisson model is incapable of representing distributions that are broader than the tuning function, which points to a second problem for the Poisson model. Obviously, the proper way to find the true (set of) stimuli is to estimate the full conditional probability $p(\mathbf{x}|\mathbf{r})$.

4.3 Optimum Linear Estimator (OLE)

The simplest possible estimator is an estimator that is linear in the activities $\mathbf{r}$ of the neurons, which suggests a solution $\mathbf{x}_{est} = W^T\mathbf{r}$, where the problem is to find the optimal matrix W, which minimizes the mean square distance between the estimate $\mathbf{x}_{est}$ and the true stimulus $\mathbf{x}$:

$$\mathbf{w} = \overset{argmin}{W}\ E[(\mathbf{x}_{est} - \mathbf{x})^2]$$

One can think of the linear estimator as being the response of a two-layer Perceptron-like neural network with a set of output units, where output unit i has weights $\mathbf{w}_i$ to the input $\mathbf{r}$ and where W is the matrix with columns $\mathbf{w}_i$.

The OLE is known to be unbiased for a large number of units. Its variance given $\mathbf{x}$ is given by

$$E[(\hat{\mathbf{x}}_{OLE} - E\{\mathbf{x}\})^2] = \sum_{i=1}^{N} w_i^2\sigma_i^2$$

where $\sigma_i^2 = \sigma_n^2$ for normally distibuted noise with variance σ_n^2, and $\sigma_i^2 = f_i(\mathbf{x})$ for Poisson distributed noise.

Note, that the OLE model suffers from the same problem as the center-of-gravity estimate in the sense that many neurons contribute their noisy output to the population estimate, whereas only few neurons may respond to a stimulus. Therefore, a compromise has to be made between small tuning widths for a high resolution versus broad tuning

widths to elimate noise by averaging responses, thereby increasing the signal-to-noise ratio of the estimate.

5 Overlap of receptive fields and correlated noise in neural responses

In the analysis so far, we have made the assumption of independent noise in neighboring neurons. Also, we have demonstrated that the optimal tuning of neurons depends on the type of noise in the neural responses. In this section we will explore this in more detail, in particular in relation to optimal tuning width of neurons and to optimal information content of neuronal activity for various types of (correlated) noise.

5.1 Optimal receptive fields: broad or narrow ?

One of the central problems with population coding is how the neuronal code can be made as efficient and as accurate as possible. It is a common believe that sharper tuning in sensory or motor pathways improves the quality of the code, although only to a certain point; sharpening beyond that point is believed to be harmful. This was illustrated already in Figs. 5, which shows the Fisher information as a function of the receptive field width of model neurons, which have an orientation specificity, similar to that of neurons in visual cortex. Fig. 5 shows that sharp tuning (small receptive fields) is not efficient for the population coding model, since for very small receptive fields, the number of neurons, that respond to a narrow bar of light, is too small to reduce the noise in the neuronal responses. For broader tuning, more neurons will respond to the narrow bar, which allows noise reduction and improvement of the signal-to-noise ratio. Obviously, the optimal receptive field size depends on several parameters, such as the noise in the neuronal responses, the number of neurons, the distribution of receptive fields (homogeneous versus nonhomogeneous).

The best way to proceed is to start with the Fisher Information $J = E[-\frac{\partial^2}{\partial\theta^2} log\, p(\mathbf{r}|\theta)]$ where $p(\mathbf{r}|\theta)$ is the distribution of the activity conditioned on the encoded variable θ and $E[..]$ is the expected value over the distribution $p(\mathbf{r}|\theta)$. Instead of the Fisher information, one could also have chosen the Shannon information, which is simply and monotonically related to the Fisher information in the case of population coding with a large number of units.

Let us consider first the case, in which the noise distribution is fixed. For instance, for the population of neurons from the example in section 4.2, where we had a population with N neurons with bell-shaped tuning curves and independent gaussian white noise with variance σ^2, the Fisher information reduces to

$$J = \sum_{i=1}^{N} \frac{f_i'(\theta)^2}{\sigma^2} \tag{29}$$

where $f_i(\theta)$ is the mean activity of unit i in response to the stimulus with orientation θ, and $f_i'(\theta)$ is the derivative with respect to θ. Equation 29 illustrates, that as the width of the tuning curve decreases, the derivative $f_i'(\theta)$ will become steeper and thus the information increases up to infinity for infinitely small receptive fields. Clearly, this

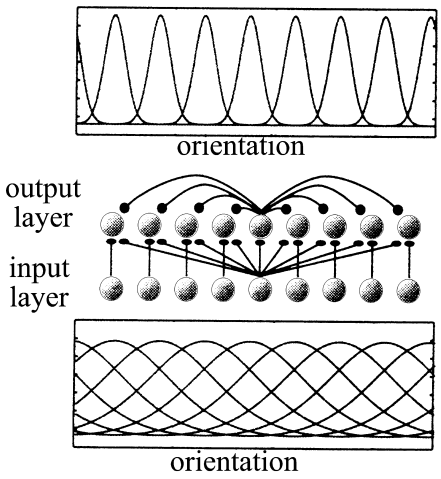


Figure 7: Two-layer neural network with feedforward excitatory connections between input layer and output layer and with lateral connections in the output layer. For visibility, only one representative set of connections is shown in each layer. The tuning of the input units was chosen broad, whereas lateral "mexican-hat"-like connections in the output layer create narrowly tuned neurons in the output layer. Since information cannot increase, this provides an example, where broad tuning in the input layer provides more information (or at least as much) as narrow tuning in the output layer.

corresponds to the ML estimate, discussed in section 4.2, where narrow tuning is better than broad tuning. Note, that for the same noise the minimal detectable change, which is inversely proportional to the square root of the Fisher information, reveals that narrow tuning may not be optimal for the population coding model (see Fig. 5).

When the noise distribution is not fixed, the results become different. Let us consider a two-layer network with an input layer and an output layer, with feedforward connections from input to output neurons and with lateral inhibitory connections in the output layer to sharpen the tuning curves (see Fig. 7). This case is particularly relevant for neurophysiologists. Since the output neurons can never contain more information than the input neurons, this model shows an example where broad tuning contains more information than narrow tuning. However, sharpening is done by lateral interactions, which induces correlated noise between neurons. The loss of information has to be attributed to this correlated noise.

The results above demonstrate that the answer to the question whether broad or narrow tuning is best, depends on the noise. In most neurophysiological experiments measuring single-unit activity it is impossible to detect correlated noise and in most cases it is not even possible at all to make a good estimate of the type of noise in the neuronal response. Therefore, usually independent noise is assumed. In the example above, this would lead to the erroneous conclusion that the output layer contains more information than the input layer. This simple example demonstrates that a proper characterization of the noise distribution is essential for a proper estimation and interpretation of the neuronal activity in a population. Multi-unit recording techniques may be an excellent tool for this purpose.

Many studies have convincingly demonstrated, that noise in a population of neurons is correlated. When the fluctuations of individual neurons about their mean firing rates would be uncorrelated, the variance of their average would decrease like 1/N for large N. In contrast, correlated fluctuations cause the variance of the average to approach a fixed limit as the number of neurons increases. The inverse of the Fisher information is the minimum averaged squared error for any unbiased estimator of an encoded variable. It thus sets a limit on the accuracy with which a population code can be read out by an unbiased decoding method.

5.2 The effect of correlated noise on the information content of neuronal activity

Let us consider a simple example of N neurons with firing rates r_i with mean values f_i, identical variances σ^2 and correlated variabilities so that

$$\langle (r_i - f_i)(r_j - f_j) \rangle \;=\; \sigma^2[\delta_{ij} \;+\; c(1 - \delta_{ij})] \tag{30}$$

with the correlation coefficient c satisfying $0 \leq c < 1$. In this case, the variance of the average of the rates

$$\bar{R} \;=\; \frac{1}{N} \sum_{i=1}^{N} r_i$$

is

$$\sigma^2_{\bar{R}} \;=\; \frac{\sigma^2}{N}[1 + c(N-1)].$$

This illustrates that the variance increases as a function of the correlation c for fixed N, and that for large N the variance approaches a fixed limit $c\sigma^2$. A typical correlation among activities of neurons in area MT (an area in the visual cortex, which is involved in the processing of moving visual scenes) has been estimated at about 0.1 to 0.2. This leads to the conclusion, that coding accuracy will not improve for populations of more than about 100 neurons.

In order to obtain a more basic insight in the effect of correlated noise, let us assume a population of N neurons, which respond to a stimulus with firing rates that depend on a variable x that parameterizes some stimulus attribute. When the average activity of neuron i to stimulus x is $f_i(x)$, its activity to a given trial is

$$r_i \;=\; f_i(x) \;+\; \eta_i$$

with η_i representing gaussian noise with zero mean and covariance matrix Q(x). We will consider three different types of variability: additive noise, multiplicative noise and correlation of noise for neurons within a limited range of each other. For additive noise, the covariance matrix is given by Eq. 30. For the limited-correlation model with an equidistant distribution of neurons the correlation matrix is given by

$$Q_{ij} \;=\; \sigma^2 \rho^{|i-j|}$$

where parameter ρ $(0 < \rho < 1)$ determines the range of correlations between neurones in the population. The parameter ρ can be expressed in terms of a correlation length L by writing

$$\rho \;=\; exp(-\Delta/L)$$

where Δ is the distance between peaks of adjacent tuning curves. For multiplicative noise, the covariance matrix is scaled by the average firing rates:

$$Q_{ij} \;=\; \sigma^2[\delta_{ij} \;+\; c(1-\delta_{ij})]f_i(x)f_j(x).$$

The Fisher information $J(\mathbf{x})$ is the best measure to estimate the effect of correlated noise on the population coding of stimulus $\mathbf{x}$, since the discriminability d, which quantifies how accurately discriminations can be made between two slightly different values $\mathbf{x}$ and $\mathbf{x} \;+\; \Delta\mathbf{x}$ based on the response $\mathbf{r}$, is related to the Fisher information by

$$d \;=\; \Delta\mathbf{x}\sqrt{J(\mathbf{x})}$$

The larger the Fisher information, the better the discriminability and the smaller the minimum unbiased decoding error.

When the random noise η is drawn from a gaussian probability distribution, the probability distribution $P[\mathbf{r}|\mathbf{x}]$, which determines the probability that a given response $\mathbf{r}$ is evoked by the stimulus $\mathbf{x}$, is given by

$$P[\mathbf{r}|\mathbf{x}] \;=\; \frac{1}{(2\pi)^N detQ(\mathbf{x})} exp\left[-\frac{1}{2}[\mathbf{r}-\mathbf{f}(\mathbf{x})]^T Q^{-1}(\mathbf{x})[\mathbf{r} \;-\; \mathbf{f}(\mathbf{x})]\right]$$

which results in the Fisher information

$$J(\mathbf{x}) \;=\; \mathbf{f}'(\mathbf{x})^T Q^{-1}(\mathbf{x})\mathbf{f}'(\mathbf{x}) \;+\; \frac{1}{2}Tr\left[Q'(\mathbf{x})Q^{-1}(\mathbf{x})Q'(\mathbf{x})Q^{-1}(\mathbf{x})\right] \tag{31}$$

where $Q'(\mathbf{x}) \;=\; \frac{dQ(\mathbf{x})}{d\mathbf{x}}$ and where $\mathbf{f}'(\mathbf{x}) \;=\; \frac{d\mathbf{f}(\mathbf{x})}{d\mathbf{x}}$. When Q is independent of $\mathbf{x}$, as it is for additive noise and limited range correlations, then only the first term in Eq. 31 survives.

This equation also illustrates that when the covariance matrix Q is independent of the stimulus (which includes that neural noise is the same for all neurons), the second term in Eq. 31 vanishes and the remaining variance is identical to that for Maximum Likelyhood (ML).

Additive noise

For the additive noise case and for large N, the Fisher information reduces to

$$J(\mathbf{x}) \;=\; \frac{N[F_1(\mathbf{x}) \;-\; F_2(\mathbf{x})]}{\sigma^2(1-c)}$$

where $F_1(\mathbf{x}) \;=\; \frac{1}{N}\sum_{i=1}^{N}(f_i'(\mathbf{x}))^2$ and $F_2(\mathbf{x}) \;=\; \left(\frac{1}{N}\sum_{i=1}^{N} f_i'(\mathbf{x})\right)^2$. This explains that the variance of the estimate (i.e. the inverse of the Fisher information) decreases with 1/N for large N, and also decreases as a function of the correlation c. The minimal error goes to zero as the correlation approaches one: any slight difference in the tuning curves can be exploited to calculate the noise exactly and to remove it.

The Fisher information will grow to infinity for correlations approaching one, only as long as $F_1(\mathbf{x}) \;-\; F_2(\mathbf{x})$ is not zero or does not approach zero for large N. When $F_1(\mathbf{x})$ differs from zero for any $\mathbf{x}$, this implies that always a fraction of the neurons will respond to any $\mathbf{x}$. This eliminates the case that $F_1(\mathbf{x}) \;-\; F_2(\mathbf{x})$ goes to zero for large N. The other case that $F_1(\mathbf{x}) \;-\; F_2(\mathbf{x}) \;=\; 0$ requires that $f_i'(\mathbf{x})$ is independent of i. $f_i'(\mathbf{x})$ independent

on i implies that all cells have the same tuning apart from a constant bias on the firing rate, which would be a pathological situation.

Multiplicative noise

For the multiplicative noise model, the Fisher information for large N is given by

$$J(\mathbf{x}) \;=\; \frac{N[G_1(\mathbf{x}) \;-\; G_2(\mathbf{x})]}{\sigma^2(1-c)} \;+\; \frac{N[(2-c)G_1(\mathbf{x}) \;-\; cG_2(\mathbf{x})]}{(1-c)} \tag{32}$$

where

$$G_1(\mathbf{x}) \;=\; \frac{1}{N}\sum_{i=1}^{N}\left(\frac{d\ log\ f_i(\mathbf{x})}{d\mathbf{x}}\right)^2$$

and

$$G_2(\mathbf{x}) \;=\; \left(\frac{1}{N}\sum_{i=1}^{N}\frac{d\ log\ f_i(\mathbf{x})}{d\mathbf{x}}\right)^2$$

The second term in the Fisher Information, which does not depend on the noise variance σ^2, arises because with multiplicative noise the encoded variable can be estimated from second order quantitaties, not merely from measurements of the firing rates themselves.

The Fisher information in Equation 32 is proportional to N (just as with the additive noise model) and is an increasing function of the correlation c, provided that $G_1(\mathbf{x}) > G_2(\mathbf{x})$. Since $G_1(\mathbf{x}) \geq G_2(\mathbf{x})$ by the Cauchy-Schwartz inequality, the only way that the Fisher information can become zero, is when $G_1(\mathbf{x}) = G_2(\mathbf{x})$, i.e. when $\frac{d\ log\ f_i(\mathbf{x})}{d\mathbf{x}}$ is independent of i. In other words, except for contrived artificial neuronal networks the Fisher information increases with correlation c and with the number of neurons N.

Limited range correlations

For limited-range correlations, the Fisher information is given by

$$J(\mathbf{x}) \;=\; \frac{N(1\;-\;\rho)F_1(\mathbf{x})}{\sigma^2(1\;+\;\rho)} \;+\; \frac{N^{1-2/D}\rho F_3(\mathbf{x})}{\sigma^2(1\;-\;\rho^2)} \tag{33}$$

where $F_1(\mathbf{x}) \;=\; \frac{1}{N}\sum_{i=1}^{N}(f_i'(\mathbf{x}))^2$, D the number of encoded variables and where

$$F_3(\mathbf{x}) \;=\; N^{2/D\;-\;1}\sum_{i=1}^{N}(f_{i+1}'(\mathbf{x}) - f_i'(\mathbf{x}))^2$$

(provided that the stimulus $\mathbf{x}$ is sufficiently far away from the boundaries of the stimulus domain). For fixed N the Fisher information is a non-monotonic function of the parameter ρ that determines the range and degree of the correlations. The first term in Eq. 33 is a decreasing function of ρ and hence of L, the correlation length. The second term has the opposite dependence. For fixed N, the first term dominates for small L, and the second term dominates for large L.

In the limit for large N, eq. 33 approaches

$$J(\mathbf{x}) \;=\; \frac{N(1\;-\;\rho)F_1(\mathbf{x})}{\sigma^2(1\;+\;\rho)}$$

which illustrates that, unlike the additive and multiplicative cases, increasing correlation decreases the Fisher information. However, the Fisher information still increases linearly with N for any $\rho < 1$.

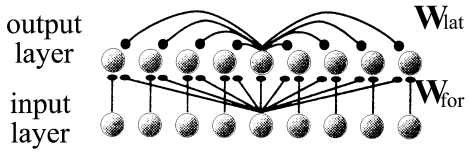


Figure 8: Example of a two-layer feedforward neural network with feedforward connections W_{for} from input layer to output layer and with lateral interactions W_{lat} in the output layer. For visibility, only one representative set of connections is shown in each layer.

This clearly illustrates that correlated noise can either lead to a decrease or an increase of the Fisher information, depending of the underlying model. The reader should be aware, that the models, discussed in this chapter are rather simple and all assume gaussian noise. For more complex models (as will be the case in biology) and for other types of noise, the results may not be valid.

6 Transformation of neural activity by the brain.

Most procedures discussed so far to estimate the neuronal activity are algorithmic approaches, suitable for off-line analysis by theoreticians. Some of the methods discussed require the complete sequence of stimuli and neural responses; others require complex analyses which do not seem to be biologically plausible. One might also wonder, how the brain is able to transform neuronal activity so as to make it easier to interpret. It might do so by mapping neuronal information in another format or to another frame of reference. Therefore, we will discuss possible neural architectures for this purpose.

Consider a two-layered feed-forward network, which is fully connected from the input to the output layer and with lateral connections in the output layer (Fig. 8). First we will assume a linear activation function in the output layer with dynamics governed by the following difference equation:

$$\mathbf{z}_t \;=\; ((1-\lambda)I + \lambda W_{lat})\mathbf{z}_{t-1}$$

where $\mathbf{z}_t$ represents the output of the network at time t, with λ a real-valued positive number in the interval [0,1], I the identity matrix, and W_{lat} a matrix for the lateral connections between the output units. At $t = 0$ the output $\mathbf{z}_0$ is initialized to $W_{for}\mathbf{r}$, where $\mathbf{r}$ is an input pattern and W_{for} is the feedforward matrix.

The dynamics of this network can be solved analytically. When the feedforward connections are equal to the lateral connections (i.e. $W_{for} = W_{lat}$), it converges to a state, corresponding to the eigenvector of the matrix $(1-\lambda)I+\lambda W_{lat}$ with the largest eigenvalue.

A slightly more complicated situation arises, when the output neurons have nonlinear activation functions. The weights W_{lat} can be set in such a way that a hill of activity arises centered around the optimal state $\mathbf{x}_{est}$. Sufficient conditions for this to occur are excitatory connections to neighbouring neurons and inhibitory connections to more distant units, such as in the well known "Mexican hat" profile of lateral connections (see the "winner-take-all" mechanism in the next chapter).

It can be shown that this recurrent network is able to provide a coarse code estimate of a stimulus $\mathbf{x}$, which is almost as efficient as the ML estimate for a large number of neurons. However, the method is in general suboptimal when the activity of input neurons is correlated.

These results show that it is possible to perform an efficient, unbiased estimation with coarse coding using a biologically plausible neural architecture like the two-layered recurrent neural network. The coarse coding and the lateral interactions serve to eliminate uncorrelated noise within a neural population and to obtain a more accurate estimate. In general, this recurrent network does not only preserve Fisher information. It can also change the format of information to make it more easily decodable. Whereas ML is a way to decode the input pattern efficiently, a complex estimator, or even a linear estimator, is sufficient to decode the stable hill while reaching the Cramer-Rao lower bound for the variance. One can therefore think of the relaxation of activity in the nonlinear recurrent network in two ways: as a clean-up mechanism of uncorrelated noise, or as a processing mechanism that makes information easier to decode.

7 Neurobiological data on neuronal population coding

Considering the many theoretical papers on efficiency of neuronal coding and about the interpretation of neuronal activity, the number of studies dealing with real experimental data is rather limited. This is certainly related to the fact, that application of the theoretical ideas on experimental data requires the simultaneous recording of many neurons in the same experimental conditions (i.e. to the same stimulus or during the same behavioral response). Simultaneous recording of actionpotentials from more than 10 neurons seems prohibitively difficult. An approximative solution to this problem is to make the best possible choice from various bad solutions: recording sequentially from single neurons in (as much as possible) the same experimental conditions. The assumption then is, that the neuronal system is time-invariant and that the response of a population of neurons can be obtained by sustituting the responses of all individual neurons in the individual recording conditions. Evidently, correlations in activity due to direct neuronal interactions in the population are lost.

7.1 Neuronal population coding in the auditory nerve

A theoretical framework for a probabilistic interpretation of neuronal activity should make a distinction between the most probable response given a stimulus (which is simply related

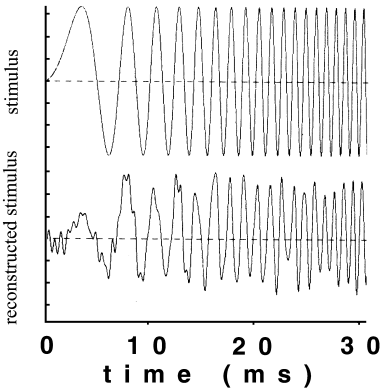


Figure 9: A frequency sweep starting at 100 Hz and rising to 1500 Hz within a time interval of 30 ms (upper panel) was presented to 64 model neurons. The tuning curve of each model neuron had a band-pass characteristic with slopes of 48 dB/octave. The central frequency ("characteristic frequency") of the neurons were distributed equidistantly on a logarithmic scale in the range between 200 and 2000 Hz. The output of the band-pass filter was supplemented by random gaussian white noise, one-sided rectified, and subsequently low-pass filtered by a low-pass filter with a 3-dB cut-off frequency at 1000 Hz and with a slope of 6 dB/octave. This signal was fed into a leaky-neural integrator (time-constant 10 ms) and a spike-generating mechanism, which generated an impulse whenever a threshold was exceeded in positive direction. The spike generating mechanism had an absolute refractory period of 1 ms. The reconstruction (lower panel) of the neuronal activity was obtained by substituting for each action potential the first order crosscorrelation between a gaussian-white-noise auditory signal and the neuronal response of each model neuron to this stimulus.

to the mean response $f_i(\mathbf{x})$ to stimulus $\mathbf{x}$), and the most plausible stimulus $\mathbf{x}$ given the neuronal response $\mathbf{r}$, which follows from Bayes relation $p(\mathbf{x}|\mathbf{r}) = \frac{p(\mathbf{x})}{p(\mathbf{r})} p(\mathbf{r}|\mathbf{x}) \approx \frac{p(\mathbf{x})}{p(\mathbf{r})} f_{\mathbf{r}}(\mathbf{x})$. As explained in sections 2.2 and 3, the simple population vector (Equation 21) is the maximum aposteriori estimator, given the measured neuronal activity $\mathbf{r}$ under the assumptions of a homogeneous distribution of independently firing neurons with independent noise and for a flat prior on the stimulus density space. Based on this result, the first order approximation implies the so-called population vector, which in this case implies that each action potential might be "substituted" by the most probable stimulus, which generated this action potential, and that the complete stimulus could be approximated by the summation of all most probable stimuli at the time of the action potentials of the individual neurons. Note that this substitution is fully equivalent to the construction of the population vector.

This theoretical framework was applied to provide a sensory interpretation of the activity in the auditory nerve. Neuronal activity in this study was obtained from a simulation of a set of 64 neurons with stochastic firing in the auditory nerve with the frequency selectivity equidistantly distributed on a logarithmic frequency scale in the

range between 200 Hz and 2000 Hz. For more details about the model neurons, see legend of Fig. 9).Figure 9 shows the stimulus in the upper panel (a frequency sweep from 100 Hz to 1500 Hz within 30 ms). The lower panel shows the reconstructed stimulus based on the neuronal population activity recorded in the 64 model-neurons. The reconstructed activity is noisy and small at the beginning of the sweep because of the low density of neurons with a characteristic frequency at low frequencies. Moreover, the sweep starts at 100 Hz, whereas the lower characteristic frequency of the neurons is at 200 Hz.

7.2 Neuronal population coding of movement direction.

The most influential study, which triggered experimental and theoretical research on population coding, was from neurons in primary motor cortex in monkey (and later also in parietal cortex and premotor cortex) for arm movements in various directions in 3-D space. Each neuron appeared to reveal the largest activity for movements in a neuron-specific particular direction, called the "preferred movement direction" (see figure 3). The preferred movement directions of all neurons appeared to be uniformly distributed in 3-D space. The directional tuning of the neurons was broad and bell-shaped. The kindness of nature, which led to uniformly distributed, unimodal, bell-shaped tuning curves, and the assumption of independent firing led to the use of the population vector, defined by Eq. 21: the summation of preferred direction vectors of cortical neurons, each weighted by the firing rate of that particular neuron. The estimated movement direction predicted by the population vector appeared to be similar to the actual movement direction within the confidence intervals.

Although the correlation with actual movement direction and predicted movement direction, based on the population vector, was quite high (typically above 0.95), this does not necessarily imply that motor cortex is explicitly and exclusively involved in the coding of movement direction. Later studies have reported that motor cortical cells also have a "preferred direction" for isometric force production in 3-D space. In these studies, monkeys were tested in a force task with an external load, such that three force variables could be dissociated: the force exerted by the subject, the net force exerted and the change in force. The directional tuning was invariant across different directions of a bias force. Cell activity appeared to be not related to the direction of force exerted by the subject, which changed drastically as the bias force changed. In contrast, the direction of net force, the direction of force change, and the visually instructed direction could all be the directional variables, alone or in combination, to which cell activity might be related. Obviously, this illustrates that the interpretation of population activity depends critically on the proper characterization of neuronal response characteristics. These observations do not violate the concept of a population vector, but indicate that an accurate and reliable interpretation of the population vector is possible, only when the response properties of single neurons are known in great detail for many experimental conditions (see Eq. 20).

8 Discussion

The aim of this chapter was to present an overview of theories about coding of sensory or motor events by neuronal activity and about the interpretation of neuronal activity in a population of neurons. For a broad range of neuronal properties, quantitative predictions can be made. The main hurdle for further progress is the experimental ability to make simultaneous recordings from many neurons. This will provide more information about important aspects related to correlation and/or independent neuronal activity due to common input, neuronal interactions and intrinsic noise in the membrane and spiking mechanism of cells.

In this context it is relevant to discuss recent observations about synchrony of firing. Synchrony of firing has been hypothesized as a way to solve the binding problem. Whatever the functional role of synchronous firing between neurons, synchronicity indicates a violation of independent firing, which poses some challenges for theoretical analyses to interpret neuronal activity.

The majority of studies on neuronal activity in sensory and motor pathways dealt with the problem how the activity of a neuron is related to a sensory stimulus or a motor response. This deals with the problem of encoding sensory and motor events into neuronal activity. Various studies, both theoretical and experimental studies, have shown that the variability of firing rate increases as a function of the number of excitatory and inhibitory inputs. A reasonable estimate is that the response variance is about 1.5 times the mean response and is fairly homogeneous throughout the cerebral cortex]. Because of this variability, an accurate estimate of firing rate of a single cell can only be obtained by averaging over time, which would eliminate fast temporal information transfer. Instead, it is thought that averaging takes place over an ensemble or population of cells, which suggests another role for synchronous firing and an alternative for the binding-hypothesis. Based on the studies mentioned above, it has been concluded that an ensemble of about 100 neurons might provide a reliable estimate of rate in just one spike interval (10-50 ms). Due to the fact that neurons share common input, resulting in a certain amount of common noise that ultimately limits the fidelity of signal transmission, little or no improvement is gained with larger pools.

9 References

Good references to look for additional material on the topics discussed in this section are

- Handbook of Biological Physics, Vol. IV. Moss and Gielen (Eds.), Elsevier.
- Kullback S. (1959) Information Theory and Statistics. John Wiley and Sons, New York.
- Shannon, S.E. and Weaver, W. (1949) The mathematical theory of communication. Urbana, Il: University of Illinois Press.
- Marmarelis, P.Z. and Marmarelis, V.Z. (1978) Analysis of Physiological Systems. Plenum Press, New York.
- T.P. Trappenberg. Fundamentals of Computational Neuroscience. Oxford University Press.
- P. Dayan and L.F. Abbott. Theoretical Neuroscience: Computational and mathematical modelling of neural systems. MIT Press.

10 Excercises

Problem 1.
In the visual system, many neurons have overlapping receptive fields, such that a single stimulus s can elicit neural responses in two neurons x and y. Demonstrate that $H(Y,Z|S) = H(Y|S) + H(Z|S)$ when the two neurons have no interactions, i.e. when $p(y_i, z_j|s_k) = p(y_i|s_k)p(z_j|s_k)$ for all i, j, k. (Note that $H(X) = -\sum_i p(x_i) log p(x_i)$).

Problem 2.
One of the major problems in brain research is to interpret neuronal activity: if an ensemble of cells generates a series of action potentials, what sensory or motor signal is coded in the action potentials. Information theory is a major tool for dealing with this problem. Suppose a signal from a sensor x is corrupted by noise n, such that the signal, which a neuron receives is given by $y = x + n$. Suppose that x is drawn from a gaussian distribution with mean μ and standard deviation σ_s. Suppose that the noise has mean zero and standard deviation σ_n. Calculate the following quantities:
a) Information in signal x
b) Information in signal y.
c) Note that the information in y is larger than the information in x. Is it correct to conclude that information has been added ?
d) Caluclate the mutual information between signal y and x. How does the noise effect the mutual information ?

Problem 3.
Suppose a neuron, which generates action potentials according to a Poisson process with mean firing rate $< r >$, such that the probability for an action potential at time τ after a previous actionpotential is given by

$$p(\tau) = < r > exp\{- < r > \tau\}$$

Also suppose that subsequent actionpotential intervalls are independent.

Calculate the mean information per actionpotential, which is obtained by observing the neural activity in the time interval between τ and $\tau + \delta\tau$.
NB: $log_2\ b = \frac{ln\ b}{ln\ 2}$

Problem 4

Assume 2 possible stimuli s^+ and s^- with $p(s^+) = p(s^-) = \frac{1}{2}$ and assume that a neuron can produce two responses: r^+ and r^-. However, encoding is not perfect. Assume that the probability of an incorrect response is p_x, such that $p(r^+|s^+) = p(r^-|s^-) = 1 - p_x$ and $p(r^+|s^-) = p(r^-|s^+) = p_x$.

a. Calculate the mutual information.
b. If the response is r^+, what is the probability that stimulus s^+ caused this response ?

Problem 5
Assume that a neuron can generate various firing rates between zero and r_{max} and that

the probability to generate a firing rate r is given by $p(r)$. Show that the optimal choice for the probability density function $p(r)$ (in terms of maximum information coding) is given by $p(r) = \frac{1}{r_{max}}$.
Hint: use $log\frac{q(r)}{p(r)} \leq \frac{q(r)}{p(r)} - 1$ with the equal sign only when $p(r) = q(r)$ and assume that $q(r) = \frac{1}{r_{max}}$. Show that $-\int p(r) log p(r) \leq log r_{max}$ with equal sign when $p(r) = \frac{1}{r_{max}}$.

Problem 6 What is the entropy of a Poisson spike train with duration T and mean firing rate r ?

Problem 7 Demonstrate, that the definition

$$I^{mutual} = \int_X \int_Y p(x,y) log_2 \frac{p(x,y)}{p(x)p(y)} dx\ dy$$

implies that

$$I^{mutual} = S(X) + S(Y) - S(X,Y)$$

Problem 8 Assume two distributions

$$f_1(x) = \frac{1}{\sqrt{2\pi}\sigma} exp(-x^2/2\sigma^2)$$

and

$$f_2(x) = \frac{1}{\sqrt{2\pi}\sigma} exp(-(x-\mu)^2/2\sigma^2)$$

Define the average information per observation from distribution 1 in favor of discrimination for the hypothesis that the sample is from population 1, against the hypothesis that the sample is from population 2 by

$$I(1:2) = \int f_1(x) log \frac{f_1(x)}{f_2(x)} dx$$

Show that $I(1:2)$ is a monotonically increasing function of μ, which is zero when $\mu = 0$ and which increasesto infinity when μ approaches infinity.

Problem 9 Calculate the Fisher Information $J(\Theta, \hat{\Theta})$ for the gaussian function $f_i(\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\theta-\theta_j)^2}{2\sigma^2}}$.

Problem 10 Assume that the response of a neuron j in visual cortex to a light bar in the receptive field is given by $p(r|\theta) = \frac{1}{Z} exp\{cos(\theta - \theta_j)\}$, where θ is the orientation of a light bar in the receptive field. Z is a normalization factor.
a. What is the optimal stimulus for this neuron ?
b. Assume that all orientations are equally likely, what is the most plausible stimulus for the neuron j given an action potential of the neuron j ?
c. Assume an array of N neurons which respond with firing rate $f_j(\theta) = Aexp\left(-\frac{(\theta-\theta_j)^2}{2\sigma^2}\right)$ to a stimulus orientation θ. The neurons are distributed equidistantly: $\theta_j = 2\pi j/N$.

Assume that the neuron fires with a poisson statistics: the probability for firing rate r_j to stimulus θ is $p(r_j|\theta) = e^{-f_j(\theta)} \frac{(f_j(\theta))^{r_j}}{r_j!}$. Assume that all stimulus orientations are equally likely ($p(\theta) = \frac{1}{2\pi}$). Show that the most plausible stimulus, which generated the firing rates r_j is proportional to $\sum_j r_j 2\pi j/N$.

Problem 11

Assume a set of N neurons with independent firing rates for a stimulus s, i.e. $p(\mathbf{r}|s) = \prod_i p(r_i|s)$. Assume that $p(r_i|s) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-(r_i - f_i(s))^2/2\sigma_i^2}$. Show that the maximum likelihood estimator for the stimulus s given a response $\mathbf{r}$ is given by the stimulus s, which meets the requirement that

$$\sum_i \frac{r_i - f(s)}{\sigma_i^2} \frac{\partial f_i(s)}{\partial s} = 0$$

Problem 12

Assume a set of N neurons in visual cortex, each of them having a preferred orientation tuning at orientation θ_n, and assume that the preferred orientations are equidistantly distributed between 0 and 2π: $\theta_n = 2\pi n/N$ and that the expected response to a stimulus θ is given by $f_n(\theta) = exp\left[-\left(\frac{(\theta_n - \theta)^2}{2\sigma^2}\right)\right]$. Also assume that each of the neurons generates actionpotentials according to a Poisson process.

a. Give an expression for the Fisher information for the Maximum Likelihood Estimator as a function of the "width" σ of the tuning.

b. Show that the largest contribution to the Fisher Information does not come from the neuron with the largest firing rate, but from neighbouring neurons !

c. Show that the Fisher Information approaches zero when σ approaches the value zero.

d. Show that there is an optimal value for σ, which gives the maximum Fisher Information.

Problem 13

We will address the question: what is the information entropy of a spike train with temporal coding. For this purpose, assume a neuron with firing rate r. If we discretise time in small bins Δt, such that the probability that the neuron generates mor than one action potential in a single time bin Δt is negligibly small. If there is an action potential in a time bin, we fix the value of the bin to the value '1' and it is '0' if there is no spike in the time bin.

a. What is the entropy of a spike train in a time interval T.

b. Show that the entropy can be approximated to

$$S = Tr\ log\left(\frac{e}{r\Delta t}\right)$$

for $r\Delta t \ll 1$.

c. Show that the entropy increases linearly with the recording time T and more than linearly with firing rate r.

d. Calculate the average entropy of an action potential when the neuron is firing at firing rate r.

Problem 14

Consider a one-sided rectifier with input x(t) and output y(t), such that $y(t) = 0$ if $x(t) < 0$ and $y(t) = x(t)$ if $x(t) \geq 0$. Assume that $x(t)$ has a gaussian distribution with $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp - \left(\frac{x^2}{2\sigma^2}\right)$ and $x \in IR$.

a. Calculate the entropy of the input signal x.

b. Calculate the entropy of the output signal y.

c. Calculate the mutual information in x and y.

Problem 15

Show, that the Fisher information for neurons with additive noise and with a response

$$r_i \;=\; f_i(x) \;+\; \eta_i$$

with η_i gaussian noise with mean zero and covariance matrix Q(x), is given by

$$J(x) \;=\; \frac{N\left[F_1(x) - F_2(x)\right]}{\sigma 2(1-c)}$$

with $F_1(x) = \frac{1}{N}\sum_{i=1}^{N}(f_i'(x))2$ and $F_2(x) = \left(\frac{1}{N}\sum_{i=1}^{N} f_i'(x)\right)^2$.

Topologically Ordered Neural Networks

C. Gielen
Dept. of Biophysics
University of Nijmegen
Netherlands

September 6, 2009

Contents

1 The Brain and Topologically Ordered Maps

It has long been known that the brain is ordered into many functionally specific areas. This is especially true for the cerebral cortex. Many cortical regions, such as the visual cortex and somatosensory areas of the cortex, are further divided into areas specific to different regions of the retina or the body. This ordering is not just a mapping all the sensory signals from one sensory organ into the same region of the cortex, but rather the topology of the sensory cells in the sensory organ is mapped onto the same topology of receiving neurons in the brain. For example in the visual cortex, the two dimensional retinal image is mapped to the visual cortex in such a way that spatial relationships present in the input stimulus are preserved when the image is transmitted to the visual cortex. Such a connection is commonly referred to as **topology preserving**. This topological mapping is the means by which information on the spatial relations between sensory cells is transmitted to, and decoded by the brain. How these topology preserving mappings can be formed in an unsupervised manner, and how they can be applied to processing information, form the basis of this chapter.

The first question is, how does the brain achieve such a high level of unsupervised ordering of all its various structures. Several theories have been put forward as to how the topographical order is maintained during growth and development. One hypothesis is that chemical markers play a role, which guide nerve fibers to specific areas in the brain during ontogenetic development. Although this may be partly true, it cannot provide a full explanation. The main argument is that the organization of neuronal areas depends on the input to these areas. For example, neuroimaging studies have shown, that a first piano lesson of about an hour, cause major changes in neuronal connectivity in somatosensory cortex (the area which receives sensory information from the fingers and from muscles controlling finger muscles) and in the finger region of motor cortex. Moreover, destruction of sensory organs, brain tissue or the deprivation of sensory stimulation at a young age, results in new neural connections which did not exist before and in the corresponding area being occupied by other projections. A good example is found in people, who were blind since birth. In these people, neurons, which would have participated in visual information processing in visual cortex, now participate in the processing of auditory or somatosensory information.

The most likely scenario seems to be that genetic coding defines coarse topological mappings of neurons in the brain, which is then fine tuned by neural plasticity in combination with neural activity. This idea leads to the idea of self-organization of topographic maps. In general, the notion of self-organization means that order can be created out of disorder without the use of a teacher or supervisor. The notion of emergent behavior is used when talking about the non random, non chaotic, complex behavior of very large spatio-temporal systems, comprised of many interconnected simple units. The formation of topographic maps can thus be considered as an emergent behavior in the brain.

2 Modeling the Self-Organizing Process

We will start with a very simple model of the neuron, where the neuron receives n inputs $\xi_j, j = 1, \ldots, n$, which are then weighted and summed to give the input activation I_i as,

$$I_i \;=\; \sum_{j=1}^{n} \mu_{ij}\, \xi_j, \tag{1}$$

where the μ_{ij} describe the synaptic efficiency or weights between neurons i and j. The most simple model for the output activity η_i is given by a static nonlinear function of the activity as,

$$\eta_i \;=\; const.\; f(I_i - \theta_i), \tag{2}$$

where $f(\bullet)$ could be the Heaviside function (i.e. $f(x) = 1,\ x > 0,\ f(x) = 0$, otherwise), and θ_i is a threshold. This computationally simple model of the neuron is used in many different Artificial Neural Network (ANN) algorithms (e.g. the perceptron). This model represents the very basic function of the neuron, and ignores its dynamical behavior. Usually, dynamical models of the neuron have been developed, many based on the so called additive model, where the change in neural activity in its most basic form is given by,

$$\frac{d\eta_i}{dt} \;=\; -\eta_i \;+\; \sum [\text{excitatory inputs}] \;-\; \sum [\text{inhibitory inputs}] \tag{3}$$

Note that the input may come from neurons or sensors in other parts of the brain, but can equally come by lateral interactions from neurons in the same brain area. A variation on this simple model treats the neuron dynamics as a form of integration of the neuron inputs, with nonlinear losses,

$$\frac{d\eta_i}{dt} \;=\; I_i - \gamma(\eta_i), \tag{4}$$

where $\gamma(\bullet)$ describes the sum of all nonlinear loss or leakage effects, and for large values of η_i it should be convex. It should be kept in mind that the activity η_i is in fact a frequency (firing rate) and as such cannot be negative. One interesting point to note about the formulation of equation (4) is in the stationary state (i.e. $d\eta_i/dt = 0$), that,

$$\eta_i \;=\; \gamma^{-1}(I_i). \tag{5}$$

Given the assumptions on the shape of γ, it is seen that γ^{-1} is compatible with the definition of the form of the thresholding function $f(\bullet)$ mentioned earlier. This last model will be used later on in discussing the physiological interpretation of the self-organizing map (SOM).

For the dynamics of neuronal maps, the learning rule is crucial. Most learning rules are based on the hypothesis of Hebb, which expresses how the neural synapses are modified,

> "When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process 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 hypothesis is written down in analytical form as,

$$\frac{d\mu_{ij}}{dt} \;=\; \alpha\, \eta_i\, \xi_j, \tag{6}$$

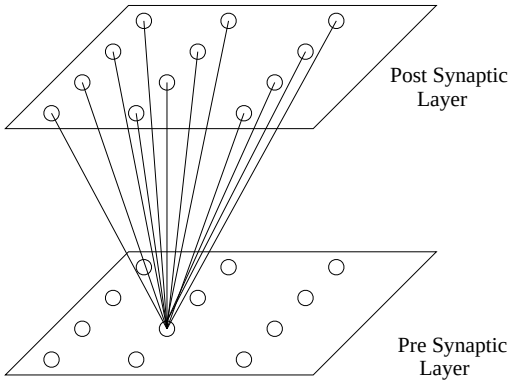


Figure 1: Illustration of Willshaw-von der Malsburg Model, showing two lattices of neurons, represented by circles, and the synaptic connections between one neuron in the presynaptic layer and all neurons in the postsynaptic layer are represented by the lines.

where μ_{ij} is the synaptic strength or neural weight between neuron j, which transmits a signal to neuron i, and α is a scalar parameter called the learning rate, η_i is the postsynaptic activity and ξ_j is the presynaptic activity. Since it was originally described, the Hebbian learning principle has been modified in many ways. One of the characteristics of Hebb's law as expressed in equation (6) is that the synaptic weight might increase to extreme values, and naturally there must be some saturation level. To overcome this problem a passive decay term was hypothesized:

$$\frac{d\mu_{ij}}{dt} \;=\; \alpha\ \eta_i\ \xi_j \;-\; \mu_{ij}. \tag{7}$$

We will illustrate these ideas with a simple model for self-organization in the visual system by assuming two layers of neurons, a presynaptic layer and a post synaptic layer, with each neuron in the presynaptic layer connected to each neuron in the post synaptic layer by a synaptic weight. Figure 1 shows the two layers, the presynaptic and postsynaptic layer. Each neuron is represented by a circle and the lines from one neuron in the presynaptic layer to the postsynaptic layer represent the synaptic connections. In the model every neuron in the presynaptic layer has a connection with every neuron in the postsynaptic layer. When subjected to an input signal the neuron weights are adapted using a Hebbian learning law, followed by renormalization. The activity of a neuron is described by equation (3). The exact form of the excitatory and inhibition interactions on the learning rate between neurons in the postsynaptic layer is such, that all neurons in the adjacent neighborhood of a neuron were excitatory (positive weight connection) while those further away were inhibitory (negative weight connection). Figure 2 illustrates a single function, which defines such an interaction, normally referred to as the Mexican hat function. The effect of this local excitation and global inhibition is to create competition between neurons for activity. The neurons, that respond best to an input, strengthen all their neighbor's responses while decreasing that of neurons further away. After repeatedly stimulating the neurons, clusters begin to form, and the neuron weights become organized. Organization occurs even when all the initial synaptic weight values have been set to approximately the same value. This organization is shown in figure 3, where the stimulation

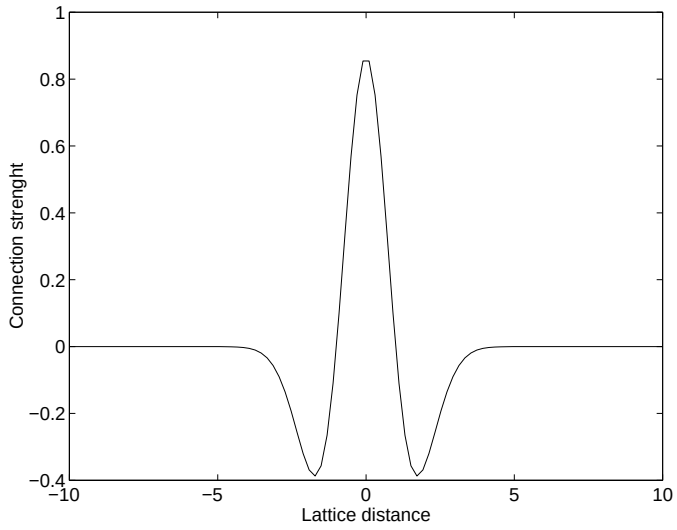


Figure 2: One dimensional Mexican hat function used to implement excitatory and inhibitory connections.

is two dimensional and is randomly chosen from the area bounded by the outer square. The synaptic weights, also two dimensional, are plotted as points in input space, and lines connect synaptic weights of immediately neighboring neurons in the postsynaptic layer. The fact that these lines do not cross except at the points of the synaptic weight vectors, is interpreted as the weights being organized. The lettered neurons were used as polarity markers, which break the symmetry of the map and ensure that the weights converge to the correct one of the eight possible orientations. This model explains clearly what is meant by self-organization and topology preserving mappings between neuron layers.

This simple model is limited in some sense. Apart from being computationally quite expensive during simulation, they are not very robust and self-organization is usually local. However, what these models do suggest is the mechanism which performs self-organization. The first requirement is some form of competition between the neurons. The winning neurons or the neurons, that respond maximally to an input, increase in a positive way the activity of their neighboring neurons in such a way, that their responsiveness to this type of input is increased, while at the same time decreasing the response of neurons further away. This competition and changing of the synaptic weights in the models discussed so far take place simultaneously. In the next section, we will provide a more detailed analysis of the concepts outlined in this introduction.

3 The Self-Organizing Map (SOM)

The SOM algorithm describes a mechanism, which allows for the formation of globally organized topology preserving maps. Originally presented as a simple numerical algorithm, it soon became clear that the algorithm could be stated in a much more general or abstract form and could be applied in many different settings. First consider a K-dimensional lattice of N neurons where the position of each neuron j in the lattice is given by a coordinate vector $\mathbf{i}_j = (i_{j1}, i_{j2}, \ldots, i_{jK})$. Define a metric space (I, d_a) and assume that $\mathbf{i}_j \in I,\ \forall\ j$.

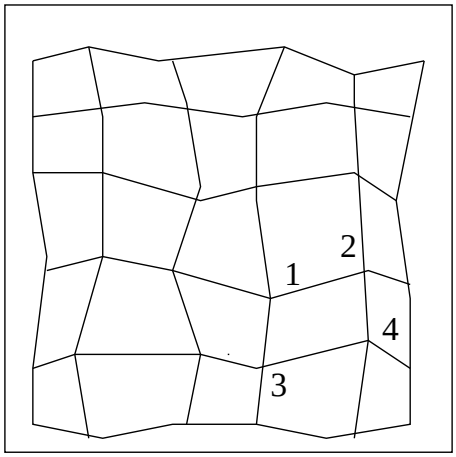


Figure 3: Typical result of Willshaw-von der Malsburg model for a 6×6 lattice. The outer box indicated the square which is the support from which the input samples were drawn. Each synaptic weight is plotted as a point and the synaptic weight for each neuron is connected to the synaptic weight of each of its immediate neighbors in the postsynaptic layer.

Associated with each neuron j is a D-dimensional weight vector $\mathbf{m}_j = (\mu_{j1}, \mu_{j2}, \ldots, \mu_{jD})^T$ where $D \geq K$. There is a D-dimensional input $\mathbf{x} = (\xi_1, \xi_2, \ldots, \xi_D)$ which is "presented" to each neuron. Figure 4 shows an illustration of this structure for $K = 2$, and the neurons are represented by the circles. Define a metric space (X, d) where $\mathbf{x} \in X$ and $\mathbf{m}_i \in X,\ 1 \leq i \leq N$, and d is a measure which satisfies the usual requirements of a distance metric. The SOM algorithm is carried out in a series of discrete time steps t, and at each time an input signal $\mathbf{x}(t)$ is taken and $d(\mathbf{x}, \mathbf{m}_i)$ is evaluated for each i. This step can be interpreted as a measure of the activity of a neuron in response to the input, the smaller the measure for the neuron the greater its activity. In the SOM algorithm the neuron with the greatest level of activity is called the winner $v(t)$, formally given by,

$$v(t) \ = \ \arg\ \min_{1 \leq i \leq N}\ d(\mathbf{x}(t), \mathbf{m}_i(t)). \tag{8}$$

This principle of one winning neuron is commonly referred to as the Winner Take All (WTA). The winner neuron is then used to define the change in the values of the neuron weights. The general principle is to change the weight values such that $d(\mathbf{x}(t), \mathbf{m}_i(t+1)) < d(\mathbf{x}(t), \mathbf{m}_i(t))$. If self-organization is to occur, then some weighting of these updates, dependent on the distance $d_a(\mathbf{i}_{v(t)}, \mathbf{i}_j)$ on the neuron lattice, between the winning neuron and the other neurons j must be used. Referring back to the excitation/inhibition Mexican hat function used in the models of the previous section, a similar type function h is used in the SOM. However, in this case h acts as a weighting function in the control of the synaptic plasticity during learning. Unlike the Mexican hat function, it does not describe the feedback activity of the signals. The function h is commonly referred to as the neighborhood function. A typical function h is shown in figure 5, where the strongest weighting is given to the update of the weights, whose neurons are closest to the winning neuron on the lattice. Note that unlike the Mexican hat function of figure 2, the function h is never negative. The update of the neuron weights is formally given as,

$$\mathbf{m}_i(t+1) \ = \ \mathbf{m}_i(t) \ + \ \alpha(t)\ h\Big(d_a(v(t), i)\Big)\ \mathbf{x}(t), \tag{9}$$

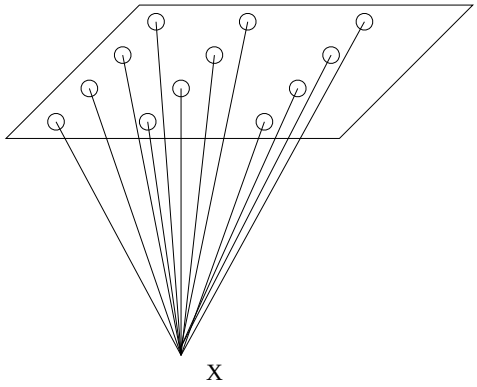


Figure 4: Illustration of the structure for the SOM, with a 2 dimensional lattice of neurons represented by circles. The input $\mathbf{x}$ is shown connected to each neuron by lines which represent the neural weights $\mathbf{m}_i$.

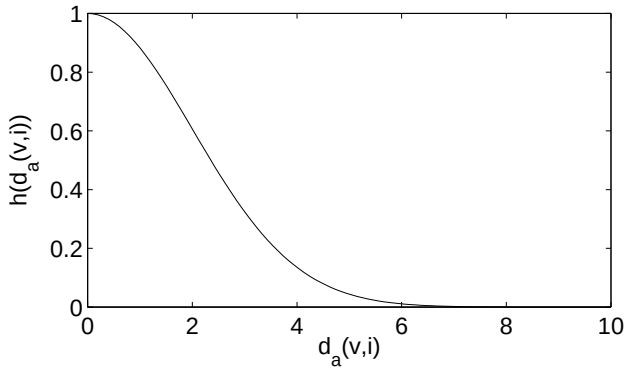


Figure 5: Plot of a typical neighborhood function $h(d_a(v,i))$.

where $\alpha(t) \to 0$ when $t \to \infty$ is a gain factor and,

$$\delta \mathbf{m}_i(t) \;=\; \epsilon \, \nabla_{\mathbf{m}_i} d(\mathbf{x}, \mathbf{m}_i), \tag{10}$$

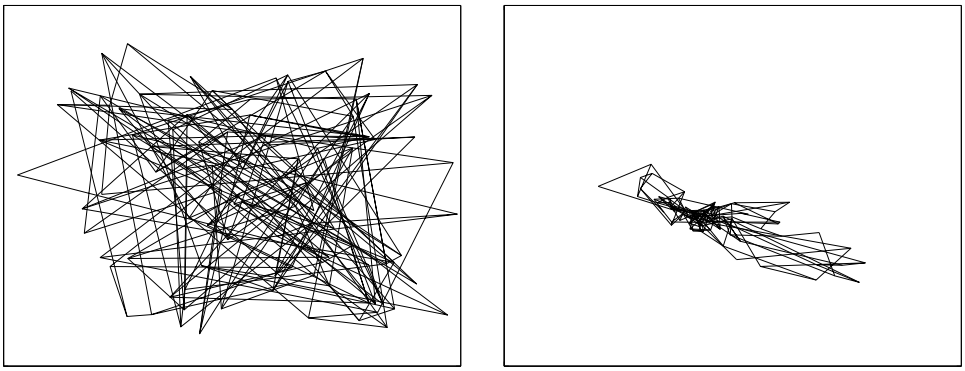
where ϵ is a constant, sufficiently small to ensure that $d(\delta \mathbf{m}_i(t), 0) \leq d(\mathbf{x}(t), \mathbf{m}_i(t)) \;\forall\; t$ and where $\nabla_{\mathbf{m}_i}$ represents the gradient relative to $\mathbf{m}_i$. Intuitively, it can be seen that the effect of the algorithm is to cluster the weights of neighboring neurons together, which eventually leads to a global self-organization of the weights.

To understand what is meant by the self-organized state of the neuron weights, consider a SOM with $K = D = 2$ and $\mathbf{x}$ uniformly distributed on the unit square. The metrics d, d_a in this case are taken to be the Euclidean distances. Figure 6 shows a series of plots of the neuron weight vectors, plotted on the unit square, the support of the input signal. The lines in the plot join the weight vectors of adjacent neurons on the neuron lattice. Figure 6 (a) shows a random initialization of the neuron weights, after 10 iterations figure 6 (b) shows the weights converging to the center of the support. In figure 6 (c) after 100 iterations the weights are already approaching an organized state. Finally after 100,000 iterations, figure 6 (d) shows the weights in an organized configuration spreading out over the support of the input. In figure 6 (d) the meaning of topographic order is quite clear with none of the lines joining the weight values intersecting. This means that the neuron weights are organized in a similar fashion to the way the neurons are ordered on the lattice. This example represents a very simple case of the SOM. Figure 7(a) shows a plot of the neuron weights after 50,000 training iterations for a SOM with a two dimensional input and a one dimensional neuron lattice. The curve shown shows the dimension reducing ability of the SOM, by mapping a two dimensional space onto a one dimensional lattice of neurons. Once again in this example the weights have reached an organized configuration, although in this case the definition of organized is more difficult to describe in general terms than it is to understand it intuitively. Figure 7 (b) shows an example of a topological defect for a two dimensional SOM, where the weights can be considered to be locally organized, but not globally organized, since there is a twist in the distribution of the weights.

These examples show a second characteristic of the SOM algorithm. Not only does it form an organized mapping but it tends to "spread" itself out or regresses onto the probability distribution of the input signal. The organization of the SOM require somewhat contradictory conditions. To arrive at a stage, where the SOM is globally organized and forms a good approximation of the input probability distribution, requires knowing a few "rules of thumb", gained from experience. These rules are quite robust but not perfect. The most important factor influencing the ability of the SOM to form globally organized mappings is the neighborhood function. A typically used form of the neighborhood function is Gaussian in nature,

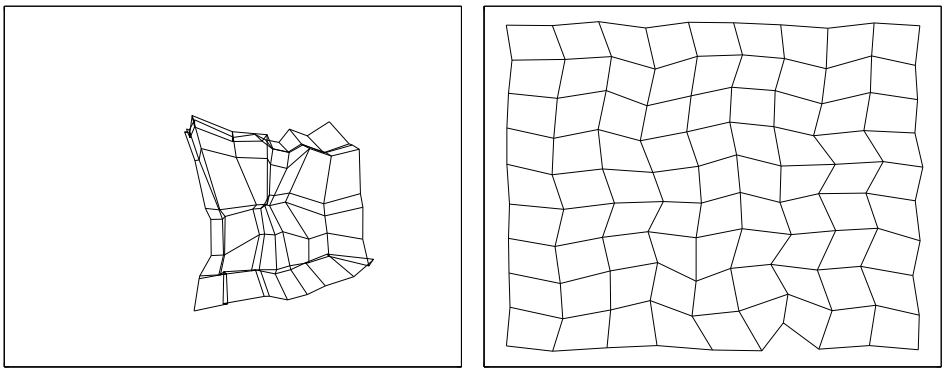
$$h(d_a(v,i)) \;=\; \begin{cases} exp\Big(-\frac{d_a^2(v,i)}{\sigma^2}\Big) & \text{if } d_a(v,i) < W \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

Some known effects of the neighborhood function will be discussed later on, but for now it is enough to say that if the width W of the neighborhood is too small, then the chances of global ordering happening are decreased. Another important factor in achieving an organized state is that the gain function $\alpha(t)$ does not decrease too rapidly. If its value



(a) Initial state.

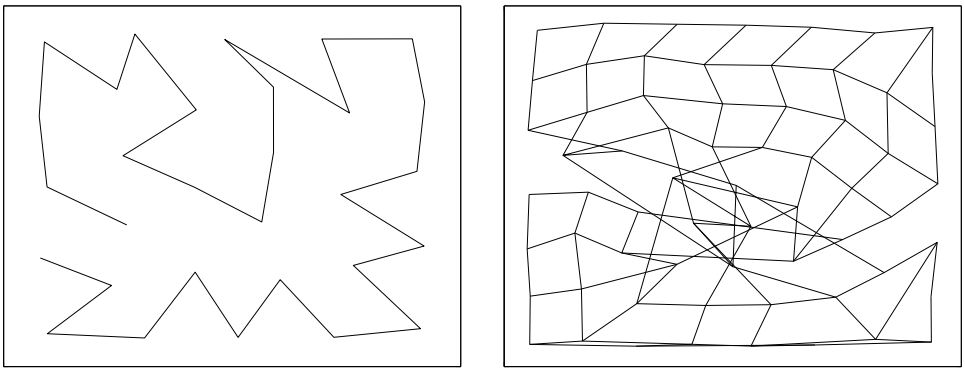
(b) After 10 iterations.



(c) After 100 iterations.

(d) After 100k iterations.

Figure 6: Plot of the neuron weights on the square support of the input signal, for a two dimensional SOM.



(a)

(b)

Figure 7: (a) Plot of the neuron weights on the square support of the input signal, for an SOM with a two dimensional input and one dimensional neuron lattice. (b) Two dimensional SOM with a topological fault.

becomes too small too quickly, then the SOM risks converging to a non-organized state. On the other hand, the wider the neighborhood function, the stronger its clustering effect, which tends to pull the neuron weights together. If the neuron weights are to spread out to form a true representation of the input space, then the influence of the neighborhood function must be decreased, that is $W \rightarrow 0,\ t \rightarrow \infty$. Similarly the gain function $\alpha(t)$ must reach small values to reduce the statistical variations in the value of the neuron weights, allowing them to converge to a state of optimal representation of the input distribution. Thus the two objectives, formation of a topologically ordered map and the representation of the probability distribution of the input require opposing conditions. Normally a compromise is reached, where the training of the SOM is divided into two phases : a) The ordering phase, where large W and α are used to allow for topological ordering of the weights. b) The convergence phase, where W is decreased towards 0, and α is small and decreases slowly to 0. This scheme works well in general, because once the neuron weights reach an organized state there is a strong tendency for them to remain in this organized state, even during phase b) of training. A semi-empirical rule for the average optimum gain $\alpha(t)$ is

$$\alpha(t) \ = \ \frac{A}{t+B}, \tag{12}$$

with A, B suitably chosen constants. The idea is that earlier and later input values are taken into account with approximately similar average weighting.

The SOM presented here was for a very general case, where the activity of a neuron in response to an input was calculated in terms of a distance between the input vector and the neuron weight vector. Another method of determining the neuron, which responds maximally to the input vector by measuring instead the correlation between the neuron weight vector and the input vector, is defined by,

$$v(t) \ = \ \arg\ \max_{1 \leq i \leq N}\ \mathbf{x}^T(t)\,.\mathbf{m}_i(t), \tag{13}$$

and the dot-product is used as a measure of the correlation. Using the dot-product however means that the input needs to be normalized before being used. Using the dot-product as a measure of activity also means, that to be compatible, the update rule for the neuron weights must be modified to,

$$\mathbf{m}_i(t+1) \ = \begin{cases} \frac{\mathbf{m}_i(t) \ + \ \alpha^*(t)\,\mathbf{x}(t)}{\|\mathbf{m}_i(t) \ + \ \alpha^*(t)\,\mathbf{x}(t)\|} & \text{if } |i - v(t)| < W \\ \\ m_i(t) & \text{otherwise,} \end{cases} \tag{14}$$

where now the gain function $0 < \alpha^*(t) < \infty$.

4 Physiological Interpretation of the SOM

Given that the SOM achieves its goal of global self-organization, the next logical question is ; given the SOM, what can this tell us about the physiology of the biological self-organizing mechanism ? It has already been stated that the reason for the global self-organizing ability of the SOM is the fact, that compared to other models, there is a single neuron chosen which has maximum response to an input and that the change of synaptic weights for each neuron depends only on this input and on the physical position of the

neurons with respect to this winner. Can this mechanism help in the understanding of the biological self-organizing mechanism, and can it suggest mechanisms of learning in the brain hitherto unknown ?

To model the WTA function the neurons must be allowed to interact with each other when an input signal is present. As before, there are two kinds of input to each neuron, an external input and lateral feedback between neurons. The activity due to inputs is written as,

$$I_i \;=\; I_i^e \;+\; I_i^l, \tag{15}$$

where I_i^e is due to the external inputs and can be simply described by,

$$I_i^e \;=\; \mathbf{x}^T.\mathbf{m}_i \;=\; \sum_{j=1}^{n} \mu_{ij}\,\xi_j, \tag{16}$$

and I_i^l is due to the laterally connected neurons and given by,

$$I_i^l \;=\; \sum_{j=1}^{n} g_{ij}\,\eta_j. \tag{17}$$

The coefficients $g_{ij} \in R$ are the effective lateral connection strengths of the cells. These were constrained such that $g_{ii} > 0$ and they have the same value $\forall\ i$. Also for all i, j with $i \neq j$ then $g_{ij} < 0$, $|g_{ij}| > |g_{ii}|$ and the g_{ij} are mutually equal. It is possible to implement these lateral interactions with interneurons whose dynamics are also described by equation (4). Starting from arbitrary initial positive values of the synaptic weight vector $\mathbf{m}_i(0)$ and zero initial activity of all the neurons, the output activity η_v of the neuron for which $\mathbf{x}^T.\mathbf{m}_i$ is maximum (i.e. the winner neuron), converges to an asymptotically high value, whereas the activity η_i, $i \neq v$ of all the other neurons converges to zero. This happens in a robust manner for a persistent input. Hence, a unique winning neuron is obtained and this model performs a WTA operation. In the case of a biological neural network however, the neurons must be able to respond to different inputs and there must be a reset of the neuron activities before the presentation of a new input. This reset is carried out by local, slow inhibitory interneurons with output variable ϕ_i. The dynamic model of equation (4) can be used for these interneurons but for simplification purposes it is written as,

$$\frac{d\phi_i}{dt} \;=\; a\eta_i \;-\; \theta, \tag{18}$$

where a, θ are scalar constants. This leads to a modification of the dynamic equation of the principal neurons in equation (4). Including the decay term ϕ_i it reads as,

$$\frac{d\eta_i}{dt} \;=\; I_i - a\phi_i - \gamma(\eta_i). \tag{19}$$

The result of this system of two coupled differential equations, which describes the dynamics of the neuron activity, is a WTA circuit with an automatic reset function.

Using a Taylor series expansion equation (14) can be shown to reduce to,

$$\mathbf{m}_i(t+1) \;\approx\; \mathbf{m}_i(t) \;+\; h(d_a(i, v(t)))[\mathbf{x}(t) - \mathbf{m}_i(t)\mathbf{m}_i^T(t)\mathbf{x}(t)]. \tag{20}$$

It should be noted that this equation has a tendency to normalize the weight values $\mathbf{m}_i$.

5 Dynamic neural fields

Dynamic neural fields have been proposed as models for the average of large ensembles of cortical neurons. The approach aims to give a macroscopic description of the neural system dynamics in terms of space- and time-continuous distribution of neural activity $u(x,t)$. The values of this function signify the average activity within a neural ensemble at location x of the neural tissue at time t.

In a simple form, the dynamics of the variable $u(x,t)$ is given by the nonlinear integro-differential equation

$$\tau \dot{u}(x,t) \;=\; -\,u(x,t) \;+\; h \;+\; \int w(x,x')\theta(u(x',t))dx' \;+\; s(x,t) \tag{21}$$

This equation describes the neuron as a leaky integrator with time constant τ, where the neural activity approaches the value h without any external input $s(x,t)$. The connectivity between neurons in the layer is represented by $w(x,x')$. For a homogeneous field, the interaction kernel can be written as $w(x,x') \;=\; w(x-x')$, in which case the integration term becomes a convolution of the thresholded output signal $\theta(u(x,t))$ with the interaction kernel.

When the θ-function is a simple linear function, a solution for the integro-differential equation above can be find easily for most external inputs $s(x,t)$. If the threshold function θ is a step-function, an analytical solution is sometimes hard to find, but numerical solutions can be found easily. We will deal with these issues in more detail in the next section.

The analysis of equations, like (21), can be done for specific values of the various parameters. However, it is more elegant and it provides more insight, when equations like (21) can be solbed analytically as a function of the various parameters. A way to do so is to find a so-called Lyapunov function. A Lyapunov function can be interpreted as a generalized Energy function, which are particularly useful for solving high-dimensional nonlinear dynamical systems. Like the energy for physical systems, the Lyapunov function is minimal for stable states of the dynamical system.

To give an intuitive idea about the role of a Lyapunov function, consider the system

$$\tau \dot{u}(t) \;=\; -\,u(t) \;+\; s \;-\; h \tag{22}$$

This can be rewritten in the simple form

$$\dot{u}(t) \;=\; -\,\frac{dE(u)}{du} \tag{23}$$

with $E(u) = (-u+s-h)^2/(2\tau)$. The last equation means that, over time, the activation u changes always in the direction that leads to a reduction of the function E(u). The activation reaches a stable state for the minimum of $E(u)$, which is obtained when $u^* \;=\; s-h$.

Obviously, for a set of n uncoupled differential equations

$$\tau \dot{u}_n(t) \;=\; -\,u_n(t) \;+\; s_n \;-\; h_n \tag{24}$$

the corresponding Lyapunov function is

$$E(u_1, u_2, ..., u_n) \;=\; \frac{1}{2\tau}\sum_n(-u_n + s_n - h)^2 \tag{25}$$

For the nonlinear discrete neuron dynamics with the interactions

$$\tau\dot{u}_n(t) \;=\; -u_n(t) + \sum_{m=1}^{N} W_{mn}\theta(u_m(t)) \;+\; s_n \;-\; h \tag{26}$$

it is not possible to write an energy function, such that the changes of the activities of the n neurons are proprtional to the negative gradient. However, it is possible to construct a function E, such that equation 26 can be written in the form

$$\dot{u}_n(t) \;=\; -\alpha(u_n)\frac{\partial E(u_1, u_2, .., u_n)}{\partial u_n}$$

where the function $\alpha(u) > 0$ is always positive. When the threshold function θ is differentiable (for example $\theta(u) = tanh(\beta u)$) and for symmetric weights ($W_{mn} = W_{nm}$), the scalingfactor $\alpha(u) = \theta'(u)$ and the Lyapunov function has the form

$$E(u_1, u_2, ..u_n) \;=\; \frac{1}{2}\sum_n\sum_m W_{mn}\theta(u_n)\theta(u_m) \;+\; \sum_n \int_{\theta(0)}^{\theta(u_n)} \theta^{-1}(\eta)d\eta \;-\; (s_n - h)\theta(u_n) \tag{27}$$

For the continuous case

$$\tau\dot{u}(\mathbf{x}, t) \;=\; -u(\mathbf{x}, t) \;+\; \int w(\mathbf{x} - \mathbf{x}')\theta(u(\mathbf{x}', t))d\mathbf{x}' \;+\; s(\mathbf{x}, t) \;-\; h \tag{28}$$

the corresponding Lyapunov function is given by

$$E(u) \;=\; \frac{1}{2}\int\int w(\mathbf{x}-\mathbf{x}')\theta(u(\mathbf{x}, t))\theta(u(\mathbf{x}', t))d\mathbf{x}d\mathbf{x}' + \int\left[\int_0^{u(\mathbf{x},t)} \eta\theta'(\eta)d\eta \;-\; (s(\mathbf{x}) \;-\; h)\theta(u(\mathbf{x}, t))\right]d\mathbf{x} \tag{29}$$

It can be shown that this Lyapunov function is

- bounded from below for all continuous functions $u(\mathbf{x})$
- $\frac{dE(u(\mathbf{x},t))}{dt} \leq 0$ for all trajectories $u(\mathbf{x}, t)$
- $\frac{dE(u(\mathbf{x},t))}{dt} = 0$ when the trajectories $u(\mathbf{x}, t)$ is a stationary solution of the neural field dynamics (which means $\dot{u}(\mathbf{x}, t) = 0$)

Notice, that not all minima of the Lyapunov function have to correspond to global minima: It may well be that the Lyapunov function has local minima or spurious attractors. Therefore, just finding the minima may not give the proper result for the stationary trajectories. Advanced techniques (e.g. simulated annealing) are required to find the global minima.

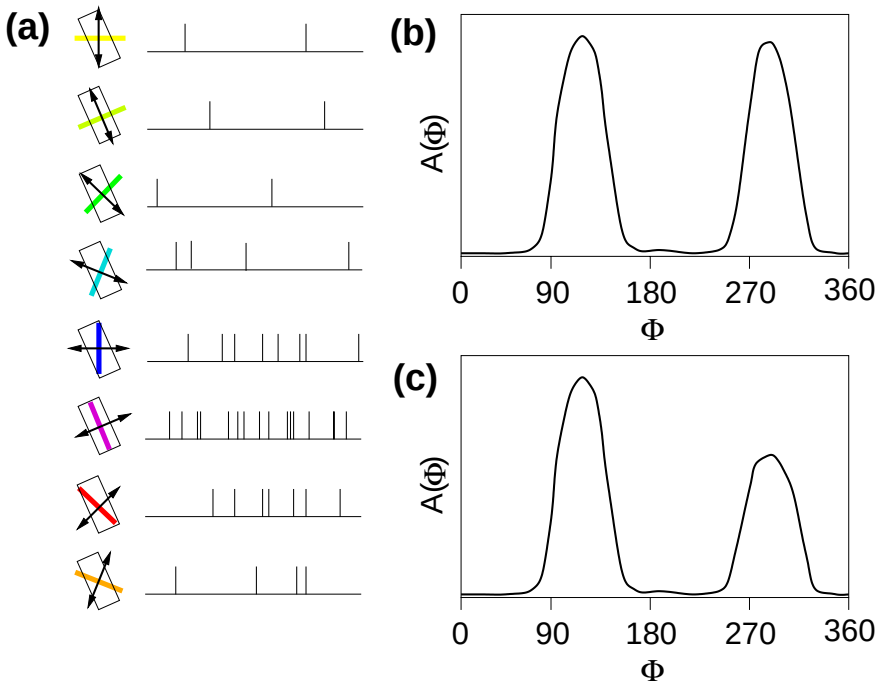


Figure 1: Orientation and direction preference of cortical cells. (a) Oriented bars moving across the receptive fields (black boxes) of neurons evoke response which are stronger when stimulating with the preferred orientation of the nerve cell (see examples of spike trains on the right). (b) The rate A of one neuron in dependence of the stimulus orientation Φ yields the tuning curve of the neuron. The response in (b) is only orientation selective, while the response in (c) displays direction selectivity.

1 Introduction

It was a major breakthrough for the investigation of neurobiological mechanisms underlying brain function, when David Hubel and Thorsten Wiesel (Nobel price for Medicine 1981) discovered that neurons in the primary visual cortex become most strongly activated by elongated visual stimuli moving across the visual field. 1(a) shows a typical recording of a simple cell of layer 4 in area 17 of the cat. A long bar of light moving across a screen leads to an increased number of action potentials only if

- the bar crosses a particular location, the so-called receptive field,
- the bar has a particular orientation, and
- a particular direction of motion.

The responses are weaker or even vanish if one or more of these conditions are changed. It appears as if the neuron was selective for a particular set of features of a stimulus and for this reason one speaks of feature selectivity of the response. The neuron is said to be tuned for the feature, which is quantified by plotting the dependency of the number of spikes on the features, the so-called tuning curves (1(b)-(c) show examples for orientation tuning and direction tuning curves, respectively).

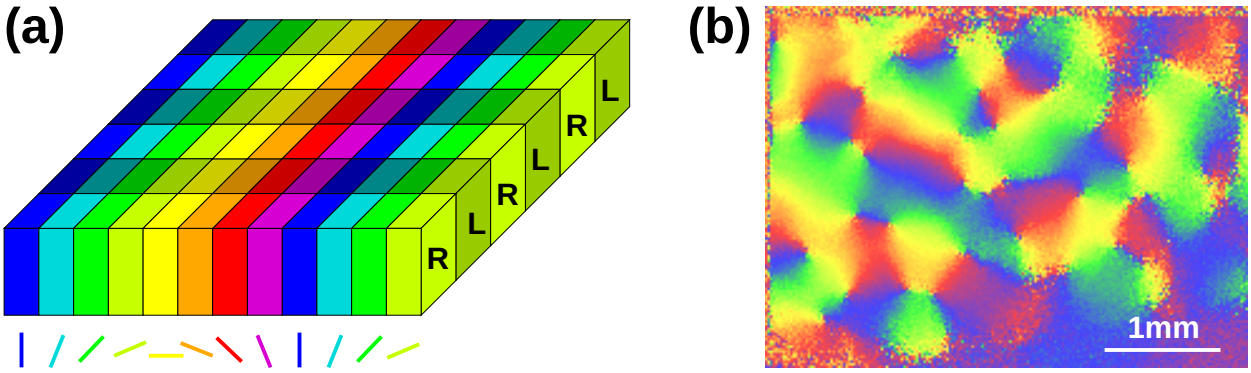


Figure 2: (a) Columnar organization of the visual cortex, as proposed in the ice-cube model by Hubel and Wiesel The cortical surface appears to be divided into columns sharing similar response properties, as e.g. the orientation (colored bars) or the ocular dominance (R=right eye, L=left eye) of a stimulus. (b) "Real" orientation maps, however, show a more complicated architecture with singularities (pinwheels) and fractures (optical imaging of area 17 of the cat, data from T. Bonhoeffer, scale bar 1mm).

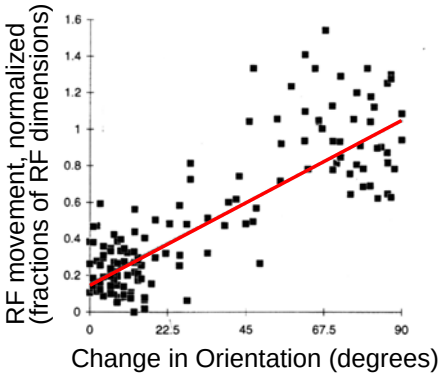


Figure 3: The distance of the centers of two neighboring receptive fields is a linear function of the difference of preferred orientations of the corresponding neurons. The slope of this function is approximately 1 receptive field per 90 degrees of orientation preference change, such that neurons with orthogonal orientation preferences have non-overlapping receptive fields.

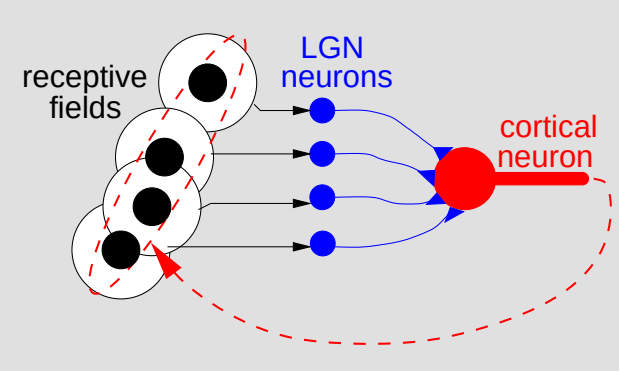


Figure 4: Classical model for orientation preference in a simple cell, adapted from Hubel and Wiesel. One neuron in the cortex (right) receives synaptic afferent input from many thalamic neurons (middle), whose receptive fields are aligned to match a specific stimulus orientation (left). Within this framework, inputs from simple cells with parallel receptive fields, converging onto a single postsynaptic cell, could explain the phase-independent response of complex cells to moving gratings.

Hubel and Wiesel also discovered that the selectivity for location and for orientation varies gradually when the recording site moves smoothly from one cortical location to the next parallel to the cortical surface ("horizontally", "tangentially", see 2). This observation is a reflection of the topographic organisation of visual cortex. In this way the responses realize mappings of retinal places and of orientations. Recently developed novel methods by optical imaging of intrinsic signals have uncovered the precise layout of these maps (2(b)). It turned out that the selectivities for particular orientations and also for directions of stimuli are arranged smoothly across the cortex except for some points and lines where they change abruptly. These discontinuities are called pinwheels and fractures, respectively, see 2(b)). While the electrophysiological experiments of Hubel and Wiesel and many others had already shown that the mapping of retinal location to cortex is on average also smooth ("retinotopy"), a very recent series of experiments revealed that the retinotopic mapping is correlated with the orientation map in a way that a movement in cortex which is associated with a change of 90 degrees in orientation preference entails on average a movement of the receptive field by one receptive field size (3).
When Hubel and Wiesel discovered the response properties of the neurons in primary visual cortex, they also offered a simple explanation for the selectivity for orientation (4). The idea was, that neurons selective for a particular orientation are connected to neurons in the lateral geniculate nucleus (LGN), that provide the input to the visual cortex, and whose (unoriented, circularly symmetric) receptive fields are arranged in an elongated manner.
An alternative picture assumes that intracortical mechanisms strongly influence the selectivity for orientation such that a small bias provided by the structure of input connections would suffice to generate the full orientation preference map. But also in this picture, the structure of the orientation maps is laid down in specific patterns of input connections. It is usually assumed that these patterns of input connections emerge by activity-dependent development.

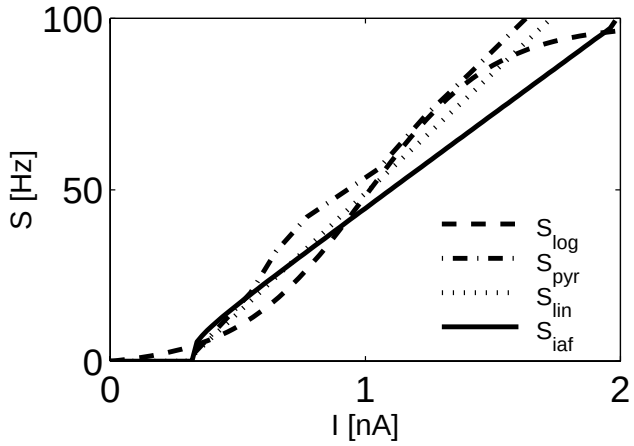


Figure 5: The response curve or gain function S models the excitability or rate of one neuron in dependence of its input current I. Above some threshold, S normally increases monotonically with I. $S_{\rm iaf}$, integrate-and-fire neuron; $S_{\rm log}$, logistic curve as used by Wilson and Cowan; $S_{\rm pyr}$, pyramidal cell in the cortex and $S_{\rm lin}$, threshold-linear neuron. Parameters are $s = 70$Hz/nA and $I_{\rm f} = 0.3$nA for $S_{\rm lin}$; $s = 140$Hz/nA and $I_{\rm f} = 1$nA for $S_{\rm log}$; $s = 1/(RCI_{\rm f}) = 60$Hz/nA and $I_{\rm f} = 0.3nA$ for an integrate-and-fire neuron obeying the differential equation $RC\ dV/dt = -V + RI$ for its membrane potential V (C=membrane capacitance, R=membrane resistance).

a threshold-linear and a sigmoidal gain function, we first consider only one population (excitatory or inhibitory) within a column, with the simplified dynamics of Eqs.(1-2) and $r = 0$

$$\tau \frac{dA}{dt} \quad = \quad -A + S(wA + I) \quad . \tag{3}$$

The fixed points can be found by solving $A_0 = S(wA_0 + I)$ for A_0; if one has to deal with two populations, one has to solve a system of two fixed point equations.

One population There are two different regimes depending on the gain parameter s and the coupling constant w; the weak coupling regime $sw < 1$, and the strong coupling regime $sw \geq 1$.
In the weak coupling regime, with both $S = S_{\rm lin}$ and $S = S_{\rm log}$, we find a stable fixed point A_0 whose absolute value that increases monotonically with increasing Input I (6(a) and (b)). The only difference between the two gain functions is that with $S_{\rm log}$, A_0 saturates at higher values of I.
In the strong coupling regime, with $S = S_{\rm lin}$, there is either one stable fixed point at $A_0 = 0$, or the activity increases beyond all limits. With $S = S_{\rm log}$, depending on I either one stable fixed point near 0, one stable fixed point near maximum activity, or both of these fixed points coexist. This behaviour results in hysteresis: with intermediate I, depending on the initial or previous activation level A, one of the fixed points either at the low activity or at the high activity level is reached (6(c) and (d)).

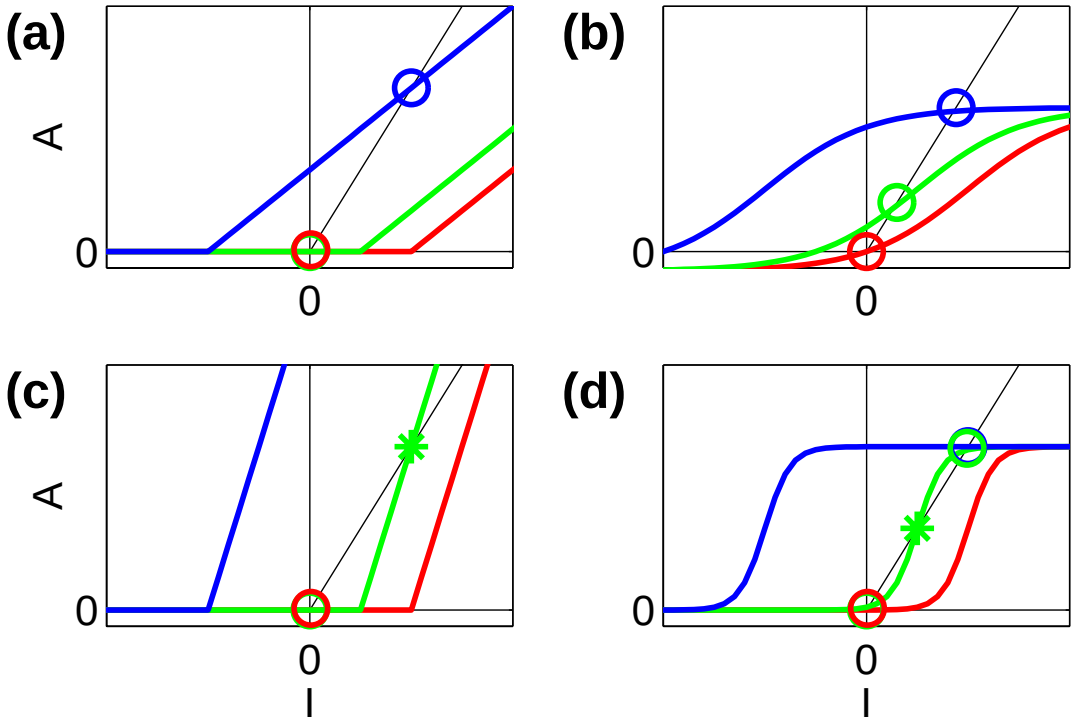


Figure 6: Fixed points of the dynamics of a single column. The figures show the fixed points A_0 (open circles, stable; stars, unstable) as revealed by the intersections of the gain functions S with the identity. For (a),(c) the threshold-linear gain function S_{lin}, and for (b),(d) the sigmoidal gain function S_{log} was used in (a),(b) the weak coupling regime, and (c),(d) in the strong coupling regime. The colors red, green, and blue mark increasing input levels I. In (a) and (b), the dynamics has one stable fixed point, while in (c), only the fixed point at $A = 0$ may be stable - otherwise, the activity diverges. In (d) the dynamics can have up to two fixed points with medium input levels; here, the systems undergoes hysteresis and the activity is limited by the saturating gain function.

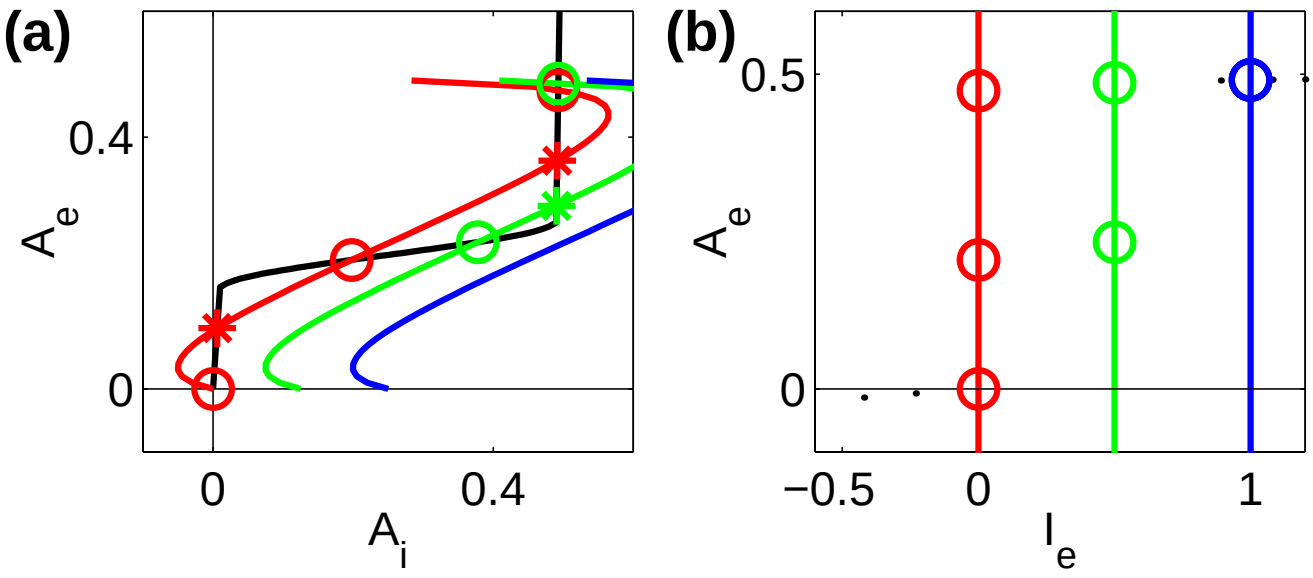


Figure 7: Hysteresis in the Wilson-and-Cowan model. (a) shows intersecting isoclines for three different excitatory input currents I_e (red, $I_e = 0$; green, $I_e = 0.5$; blue, $I_e = 1$). Fixed points are marked as in 6. (b) Depending on the excitatory input current I_e, either one (blue), two (green), or three fixed points (red) are stable, and the initial conditions determine which one is selected. Parameters were $w_{ee} = 13$, $w_{ie} = 4$, $w_{ei} = 22$, $w_{ii} = 2$, $s_e = 1.5$, $s_i = 6$, $I_{f,e} = 2.5$, $I_{f,i} = 4.3$, $r_e = r_i = 1$, $\tau_e = 10$, $\tau_i = 5$, and $I_i = 0$.

Two populations With two populations, Eqs.(1-2) yield two isoclines intersecting at the fixed points of the activation dynamics. Their stability can then be derived by linearization of Eqs.(1-2) around these fixed points and solving the characteristic equation. Using S_{lin}, the activation dynamics is very simple. There is no hysteresis in the system, and either a stable fixed point exists at $A_e \geq 0$ and $A_i \geq 0$, or the activation diverges because the interaction is too strong.
With $S_{\log}$, there is the possibility of multiple hysteresis phenomena. Increasing the constant input, one finds either one, two, or three stable fixed points existing simultaneously (7(a) and (b)). The existence of hysteresis is very important, because it can implement a form of short-term memory: brief pulses of external input can excite a column, which remains activated after the input has decayed, due to the dynamics of the internal couplings.
Additionally, there is a parameter range where the model can exhibit (damped) oscillations in the population activity. These solutions of the differential equations correspond to the existence of limit cycles in phase space. Limit cycles occur if there is only a single unstable fixed point of the dynamics, and if the input is sufficiently high. It can be shown that limit cycles occur naturally in coupled neuronal populations.

2.3 Coupled columns

To simulate more than a local cortical column, Wilson and Cowan extended their model and examined a chain of coupled neuronal populations. The activation A is now a function

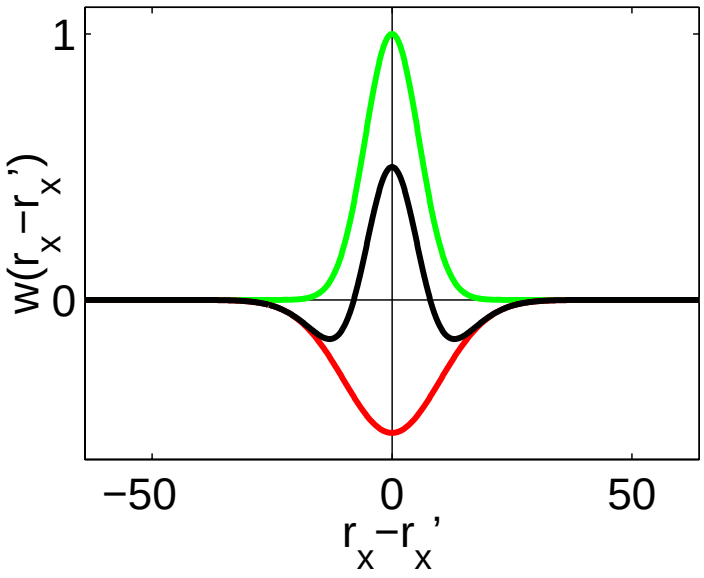


Figure 8: Excitatory couplings $W_{\mathrm{e}}(r_{\mathrm{x}} - r_{\mathrm{x}}')$ (green) having a shorter length scale than inhibitory couplings $W_{\mathrm{i}}(r_{\mathrm{x}} - r_{\mathrm{x}}')$ (red) lead to a coupling function $W(r_{\mathrm{x}} - r_{\mathrm{x}}') = W_{\mathrm{e}} - W_{\mathrm{i}}$ (black) having the shape of a mexican hat. Parameters of the coupling functions chosen (see Eqs.(7-8)) are $w_{\mathrm{e}} = 14$, $w_{\mathrm{i}} = 12.5$, $\sigma_{\mathrm{e}} = 5.6$, $\sigma_{\mathrm{i}} = 10$, and $d = 1$.

of time **and** space, $A(t) \to A(\mathbf{r}, t)$, and the synaptic input now depends not only on the activities of the populations in the same column, but also on the activities in all other columns. The products in Eqs.(1-2) therefore have to be replaced by the convolution of the activities with the corresponding coupling kernels $W_{\mathrm{ee}}(\mathbf{r} - \mathbf{r}')$, $W_{\mathrm{ei}}(\mathbf{r} - \mathbf{r}')$, $W_{\mathrm{ie}}(\mathbf{r} - \mathbf{r}')$, and $W_{\mathrm{ii}}(\mathbf{r} - \mathbf{r}')$, where $[W * A](\mathbf{r}, t) := \int_{CTX} W(\mathbf{r} - \mathbf{r}')A(\mathbf{r}')d\mathbf{r}'$

$$\begin{aligned} \tau_{\mathrm{e}} \frac{\partial A_{\mathrm{e}}(\mathbf{r}, t)}{\partial t} &= -A_{\mathrm{e}}(\mathbf{r}, t) \qquad (4) \\ &+ (k_{\mathrm{e}} - r_{\mathrm{e}} A_{\mathrm{e}}(\mathbf{r}, t)) S_{\mathrm{e}}([W_{\mathrm{ee}} * A_{\mathrm{e}}](\mathbf{r}, t) - [W_{\mathrm{ie}} * A_{\mathrm{i}}](\mathbf{r}, t) + I_{\mathrm{e}}(\mathbf{r}, t)) \\ \tau_{\mathrm{i}} \frac{\partial A_{\mathrm{i}}(\mathbf{r}, t)}{\partial t} &= -A_{\mathrm{i}}(\mathbf{r}, t) \qquad (5) \\ &+ (k_{\mathrm{i}} - r_{\mathrm{i}} A_{\mathrm{i}}(\mathbf{r}, t)) S_{\mathrm{i}}([W_{\mathrm{ei}} * A_{\mathrm{e}}](\mathbf{r}, t) - [W_{\mathrm{ii}} * A_{\mathrm{i}}](\mathbf{r}, t) + I_{\mathrm{i}}(\mathbf{r}, t)) \quad . \end{aligned}$$

The delay of synaptic transmission from $\mathbf{r}$ to $\mathbf{r}'$ has hereby been neglected.
The choice of the coupling functions W_{xx} is crucial for the dynamics of the system. A common assumption is that excitatory couplings prevail on short distances $\|\mathbf{r} - \mathbf{r}'\|$, while inhibitory interactions dominate on larger distances. This leads to a coupling function having the shape of a mexican hat (8). It is questionable if this assumption is really fulfilled in the visual cortex. It has been shown that long-range horizontal connections spanning several hypercolumns exist, while inhibitory interactions have a limiting range of about one hypercolumn. These long-ranging axons, however, are not distributed homogeneously but form dense clusters in columns having a similar orientation preference as the neuron from which they originate. Due to the typical structure of an orientation map in the visual cortex, it may still be possible that the interaction profile has indeed the shape of

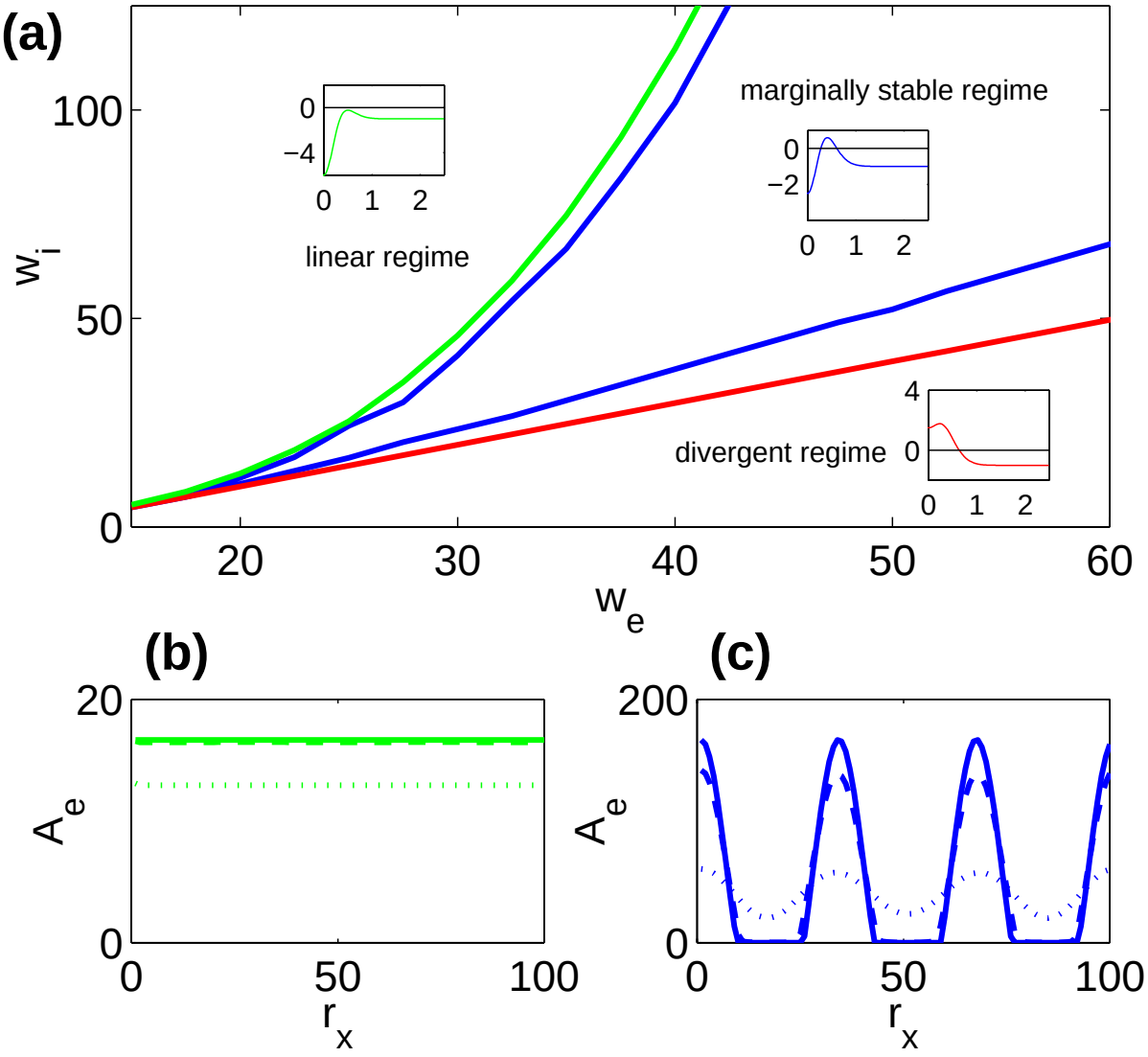


Figure 9: (a) Phase diagram in dependence of the excitatory and inhibitory coupling strengths w_{e} and w_{i}. The conditions B1-B3 partition the phase space in three regions: in the upper region, the homogeneous fixed point is stable (b), and in the lower region, no fixed point exists and the activtiy diverges exponentially. The region in between shows a different behaviour. Here, the homogeneous fixed point is unstable, so each minimal local perturbation of an otherwise constant synaptic input leads to pattern formation, which is stable (c) or unstable, depending on the actual strength of the inhibitory coupling. The green line separates the linear from the marginally stable, and the red line marks a lower boundary of the marginally stable regime. The blue lines are numerical estimates of the phase boundaries. (b) and (c) show successive activity profiles $A(r_{\mathrm{x}}, t)$ after the system has been stimulated with a homogeneous input with a small perturbation, at times $t = 1.25, 45$ (dotted), $t = 3.75, 50$ (dashed), and $t = 50, 55$ (solid), respectively. Parameters for the simulation were $w_{\mathrm{e}} = 30, 45$, $w_{\mathrm{e}} = 80, 32$, $l_{\mathrm{x}} = 100$, $\sigma_{\mathrm{e}} = 5.6$, $\sigma_{\mathrm{i}} = 10$, $\tau = 5$, $s = 100$, $I = 1$, $\Delta t = 0.25$. The insets in (a) display typical eigenspectra $\lambda(k)$ for the three cases.

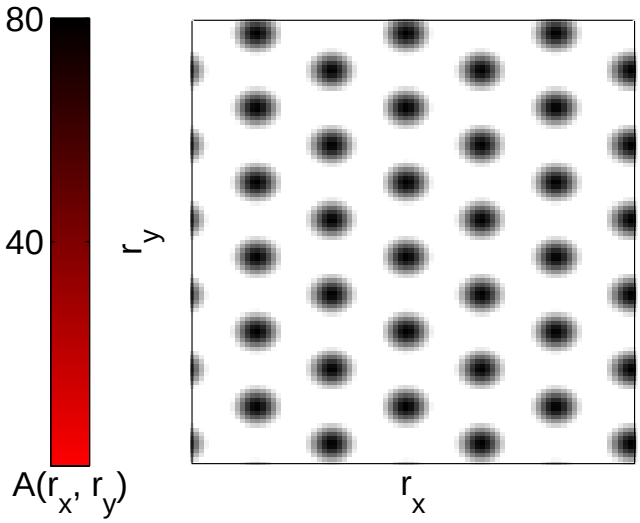


Figure 10: Stationary activation pattern $A(r_\mathrm{x}, r_\mathrm{y})$ in a two-dimensional homogeneous model cortex, obtained with a uniform stimulus. The blobs arrange in a regular hexagonal pattern. The activity A is coded in shades of grey, see color bar. Simulation parameters are $l_\mathrm{x} = 88$, $l_\mathrm{y} = 105$, $w_\mathrm{e} = 45$, $w_\mathrm{i} = 60$, $\sigma_\mathrm{e} = 2.8$, $\sigma_\mathrm{i} = 5$, $\tau = 5$, $s = 100$, $I_0 = 1$, and $\Delta t = 1$.

This leads to a localized activation blob centered around $l_\mathrm{x}/2$. If the afferent input is suprathreshold everywhere, other blobs appear in a specific distance which is determined by the length scales of the excitatory and inhibitory interactions. In 9(c), this distance is about half of the size of the chain, such that two activation clusters appear. This picture does not change significantly in higher dimensions: in a two-dimensional cortex, the activation clusters typically arrange in a hexagonal pattern (10). If the inhibitory interaction extends over larger distances, or even does not decay significantly, then the network implements some sort of a winner-takes-all network with global inhibition. Only one blob will appear at the location with strongest feedback and afferent input.
There are two other interesting dynamical states in this model leading to propagating waves or blobs of velocity Ω_b. In the first case, a movement of a periodic stimulus with velocity Ω_s as modeled e.g. by

$$I = I_0(1 + \epsilon \cos(2\pi(0.5 + \Omega_\mathrm{s} t + r_\mathrm{x}/l_\mathrm{x})) \tag{11}$$

drags the blobs into the direction of movement. Depending on the time scale of the lateral dynamics and the modulation amplitude ϵ of the stimulus, the activation either follows the stimulus perfectly with $\Omega_\mathrm{b} = \Omega_\mathrm{s}$, or misses some cycles ($\Omega_\mathrm{b} < \Omega_\mathrm{s}$). In the second case, a small asymmetry in the input leads to a self-propagating wave. Here, a necessary condition is $\epsilon << 0$, because otherwise the blob becomes pinned at the position of maximal input. Travelling waves typically occur when the inhibitory input is large.

3 A simple model of visual cortex

In the previous section, we discussed that a simple model of coupled neuronal populations can exhibit interesting dynamical properties: from simple fixed points to pattern

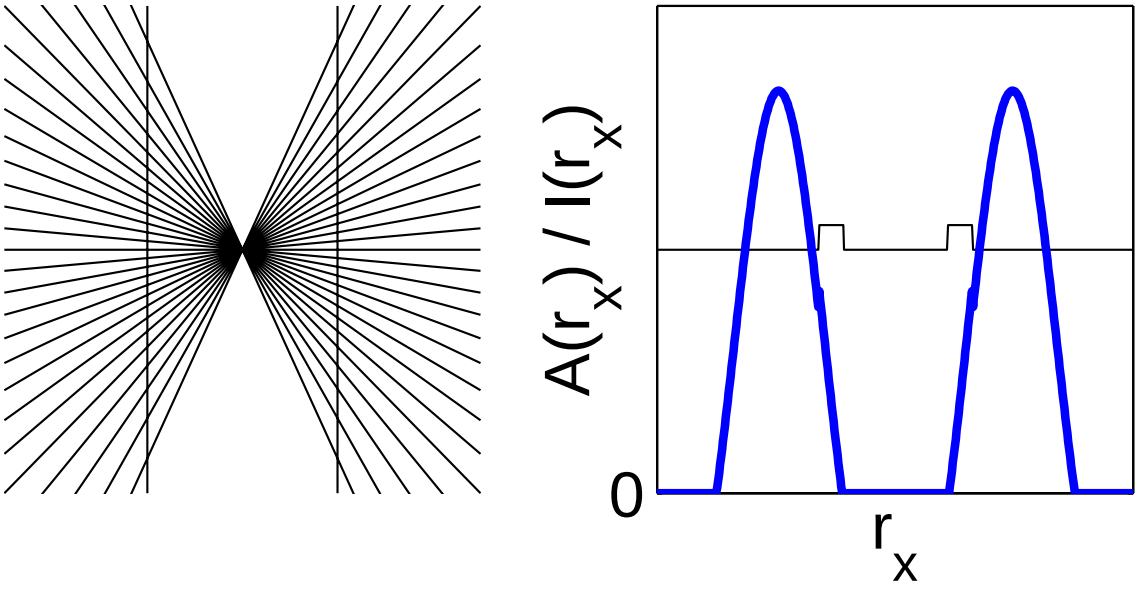


Figure 11: (a) Small angles between line segments are overestimated by our visual system; we perceive a tilt in the vertical lines despite they have been laid out in parallel. (b) Possible explanation for the phenomenon shown in (a). If difference in orientation $\Delta\Phi$ is identified with difference in cortical position Δr_{x} (compare with 2(a)), two line segments crossing at small angles lead to a bimodal input distribution $I(r_{\mathrm{x}})$ (thin line). The distance between the two maxima in this distribution is smaller than the minimal distance between two blobs, which appear shifted to the left or to the right, relative to the input. Thus, the position of their maxima in $A(r_{\mathrm{x}})$ may suggest for higher cortical areas that the stimulus displays a larger angle than present.

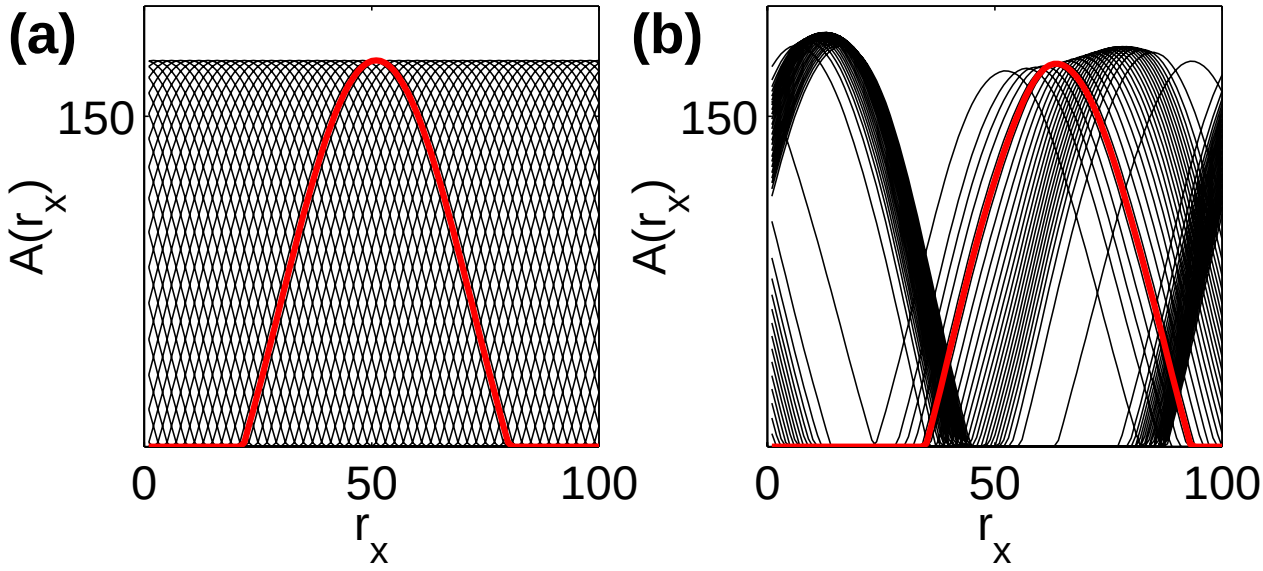


Figure 12: Superpositions of activation profiles $A(r_{\mathrm{x}})$ of a chain of coupled cortical columns to different afferent inputs (Eq.(10)). Each of the afferent inputs has a maximum at one specific location $r_{\mathrm{x}}{}'$. These locations have been chosen to be distributed equidistantly (in the simulations, the input has been shifted by equal distances, with periodic boundary conditions). While in (a), the network was homogeneous, in (b), random disorder has been introduced by applying a random displacement of $\eta(r_{\mathrm{x}}) = \mathrm{rand}(1)$ on the columnar positions. In (a), the positions of the clusters are exclusively determined by the input maximum, in (b), the marginally stable continuum of attractors has been broken up into a finite set of attractors located at positions with maximal cortical feedback. Here, system disorder and input perturbation determines the neuronal response. The response to the stimulus centered in the middle of the chain is marked in red. Parameters were $w_{\mathrm{e}} = 45$, $w_{\mathrm{i}} = 60$, $\sigma_{\mathrm{e}} = 20$, $\sigma_{\mathrm{i}} = 40$, $\tau = 5$, $s = 100$, $I_0 = 0.5$, $\epsilon = 0.05$, $\Delta t = 1$, $\sigma_{\mathrm{l}} = 5$.

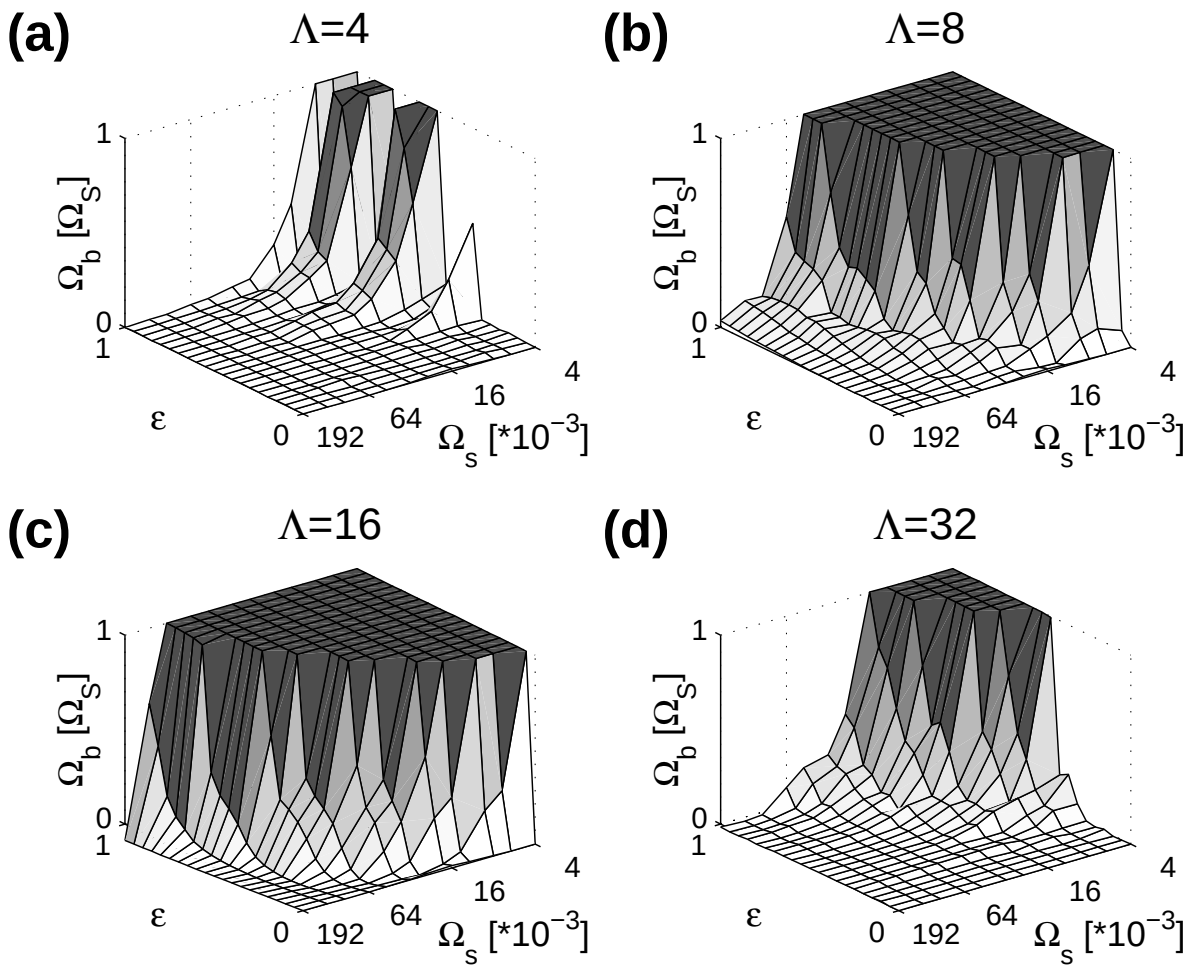


Figure 13: Movement Ω_b of the blobs, normalized to the movement of the stimulus, Ω_s, for several parameter combinations: (a)-(d) for $\Lambda = 4, 8, 16, 32$, respectively. In dependence of the effective stimulus modulation amplitude ϵ and the stimulus velocity Ω_s, the blobs either remain more or less stationary ($\Omega_b \approx 0$), or are being dragged by the stimulus ($\Omega_b > 0$). If $\Omega_b = 1$, the blobs move with the same velocity as the stimulus. Parameters as in 10.

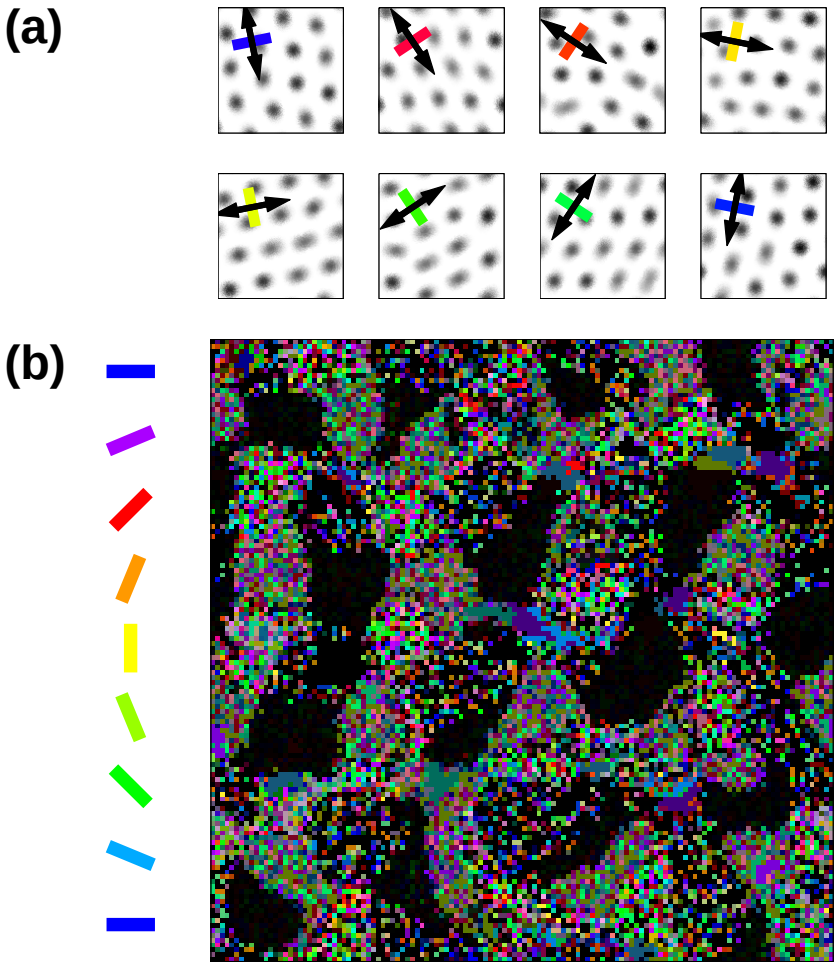


Figure 14: (a) Single conditon maps A_{n} for gratings drifting in $n = 8$ different orientations (see colored bars). (b) Orientation preference map $\Phi(\mathbf{r})$ obtained by vectorially summing up the single condition maps shown in (a). Parameters were $l_{\mathrm{x}} = l_{\mathrm{y}} = 128$, $\sigma_{\mathrm{e}} = 5.6$, $\sigma_{\mathrm{i}} = 10$, $w_{\mathrm{e}} = 45$, $w_{\mathrm{i}} = 60$, $\sigma_{\mathrm{aff}} = 4$, $< \eta >= 1$ (white noise), $I_0 = 10$, $s = 100$, $\tau = 5$, $\Delta t = 1$, $\Omega_{\mathrm{s}} = 0.2$, and $\Lambda \approx 18$, which is approximately the size of a blob.

In terms of the symmetry breaking introduced by small inhomogeneities in the distributions of the neurons, each stimulus selects a different subset of all possible attractors. The feedback of the lateral couplings both attenuates the irregularities in the system, and applies a spatial bandpass filter on the cortical response.
This bandpass filtering by the mexican-hat type of interaction is responsible for the emergence of orientation preference maps. The neurons not only achieve a preference for one stimulus orientation, but orientation preference also changes smoothly across the cortical surface. The excitatory interactions are responsible for the smoothness of the maps (cooperation), while inhibitory interactions realize a sort of competition. Typically for the mapping of a periodical quantity onto a manifold, the orientation preference map $\Phi(\mathbf{r})$ shows characteristic singularities like pinwheels (places where you can find cells of all orientation preferences nearby) and fractures (elongated curves where cells change abruptly their orientation preference), 14. In the models described in these pages, the filter here is

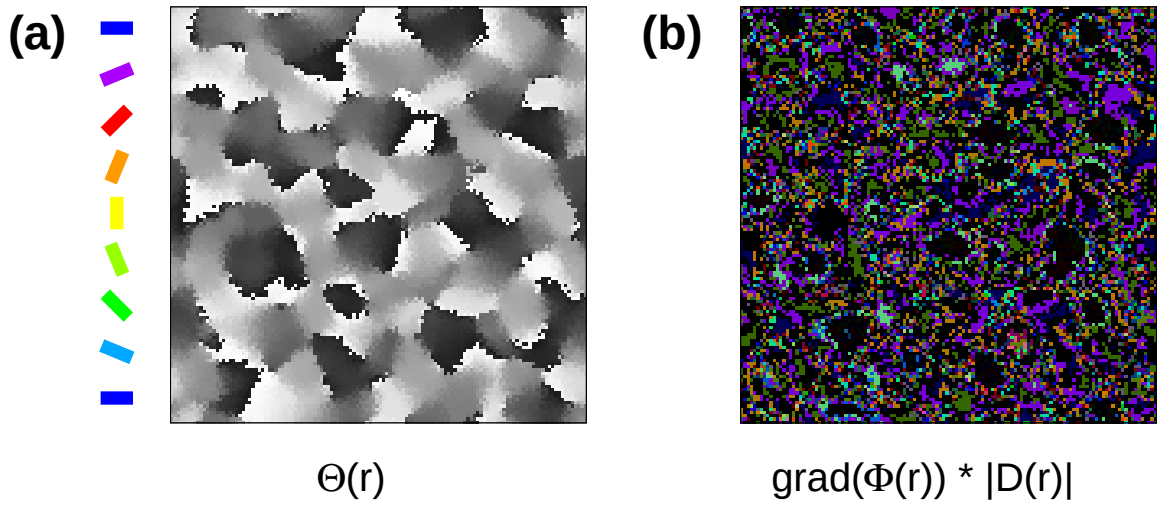


Figure 15: (a) Direction preference map $\Theta(\mathbf{r})$. The preferred direction is shown color-coded according to the color bars on the left. (b) Relation between direction and orientation map. Discontinuities in the orientation preference map, as obtained from a gradient transform $\mathrm{grad}(\Phi) = \sqrt{(\partial\Phi/\partial r_\mathrm{x})^2 + (\partial\Phi/\partial r_\mathrm{y})^2}$ with periodic boundary conditons and periodic argument $\Phi(\mathbf{r})$, are coded in shades of yellow superimposed on the map of the directional selectivity $\mid D(\mathbf{r}) \mid$ coded in grey shades. Dark colors represent low, and bright colors represent high absolute values (normalized color table). Same parameters as in 14.

related to the dynamics of the biologically plausible lateral interactions which also have much wider capabilities to account for experimental evidence like contrast invariance of orientation tuning and other neuronal properties described in the previous sections. Most importantly, the positions of active patches were robust against random initial conditions due to the stabilizing effect of the inhomogeneities in the lateral connections.

3.2 Direction preference maps

Having a closer look at the single condition maps A_n, we see that the neurons also exhibit a preference for a certain direction of the stimulus movement (17(a)). The corresponding direction map (15) closely resembles experimental data (16). Where does this preference comes from?
As we already know from the preceeding section, moving the stimulus implies a periodic movement of the activation clusters. The existence of the inhomogeneities in the neuronal density now introduces barriers like hills in a potential landscape, as in the example with the particle in a potential well discussed above. If the force of the stimulus movement does not suffice to drag the blob over the potential barrier into the basin of the neighboring attractor, the stimulus movement will only induce an oscillation of the blob around a position which is identified relative to its stationary location.
In 13, there exist regions where the blob moved, $\Omega_\mathrm{b} > 0$, or remained stationary, $\Omega_\mathrm{b} = 0$. The conditions supporting a fast movement are

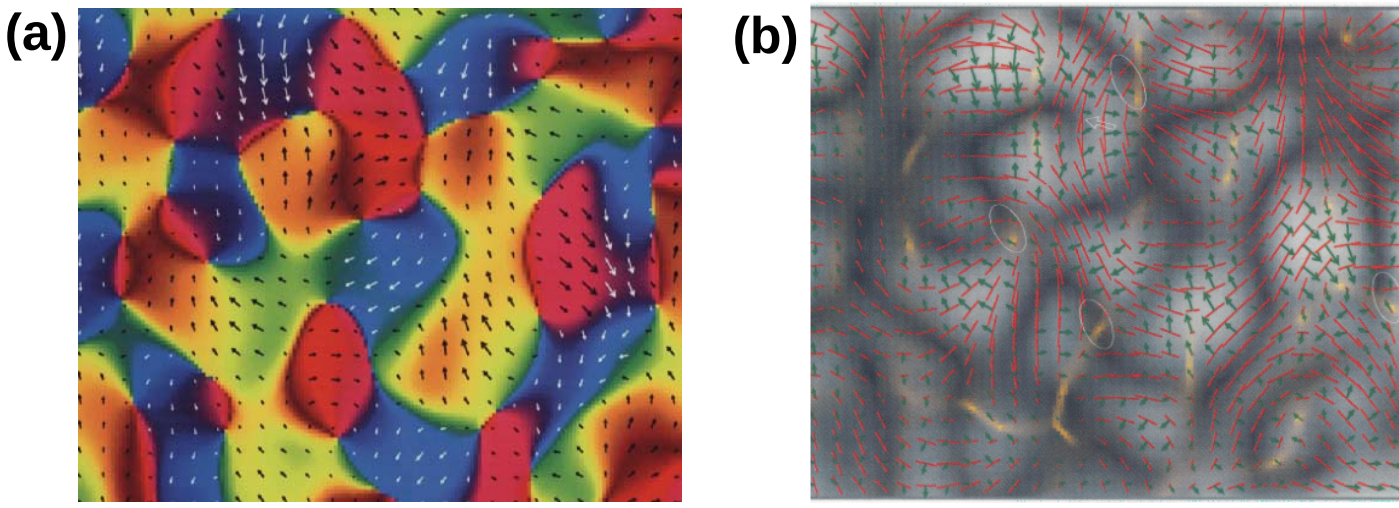


Figure 16: (a) Direction preference map and (b) relation between direction and orientation map found in experimental studies. Same representation as in 15, the sizes of the rectangles are approximately 3mm by 2.5mm.

a strong modulation ϵ of the sinusoidal stimulus, because of the competition between the stimulus movement and the localisation strength of dominant lateral interactions,

a stimulus velocity Ω_s with a time constant similar to the rate dynamics, and

a convenient grating period Λ in the same range or larger than the typical interblob distance.

Because all of these conditions have to be fulfilled to move the blobs over an inhomogeneous cortex, it is very improbable that the blobs will move being in a regime where lateral interactions are strong. Instead, most likely the shifts from the basin of attraction will lead to direction preference: moving the stimulus to the left, the average activation is shifted to the left, and moving the stimulus to the right, the average activation also shifts to the right. This effect provides a novel explanation for the direction selectivity of the neurons, not relying on special types of neurons, or asymmetries in the afferent connections, either leading to delayed summation of action potentials, or to a sharpening of these asymmetries by cortical feedback – here, the direction selectivity is caused by intracortical inhomogeneities.
From these observations, it is possible to explain another experimental result. It has been observed that patches having similar orientation preferences split into two subpopulations having opposite direction preferences (17). As a patch of similar orientation preferences corresponds to an activation cluster evoked by a specific stimulus orientation, it is obvious that the shift of the cluster in both stimulus directions divides this patch into two subregions being more strongly activated by the one or by the opposite direction of movement. Necessarily, the deformation being roughly symmetrical, there has to be a narrow region within the activated area where the activities for stimuli moving in either of the two directions are the same.

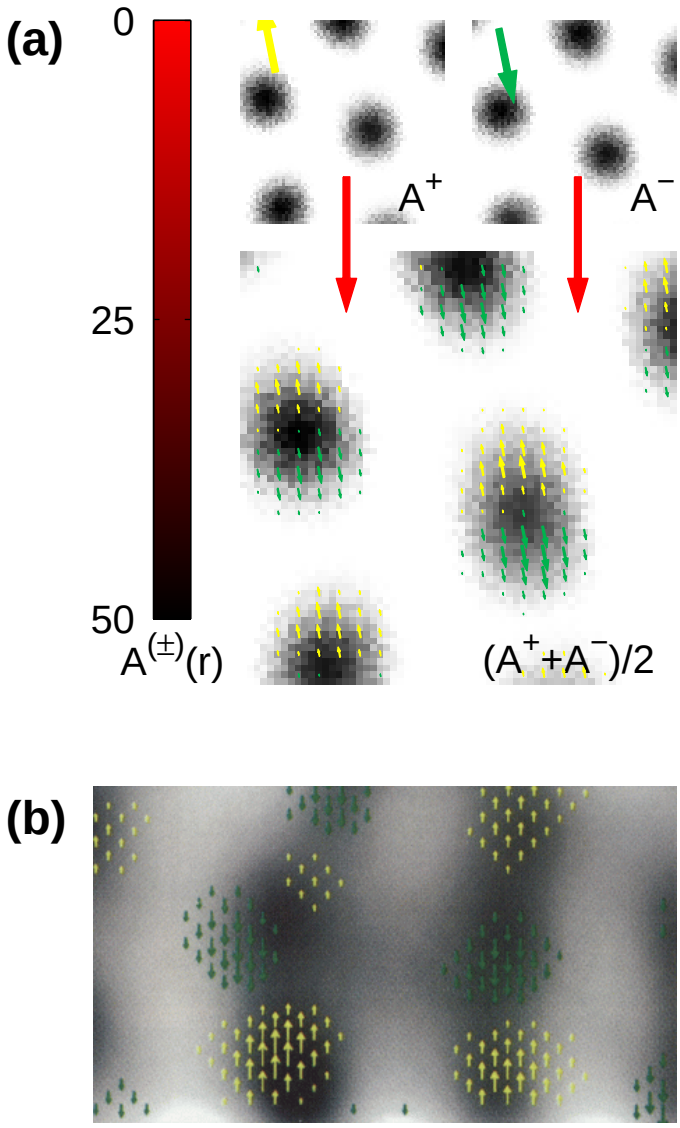


Figure 17: (a)Two activity maps $A^{\pm}(\mathbf{r})$ obtained by stimulation with a grating moving up (top left) or down (top right) yield a the mean activation $A(\mathbf{r}) = (A^{+} + A^{-})/2$ stimulating with a specific orientation (bottom). The single differential directional map is shown by the yellow and green arrows, their lengths coding the selectivity $A^{+} - A^{-}$ for one of the two conditions. (b) shows the equivalent differential map. Same parameters as in 14.

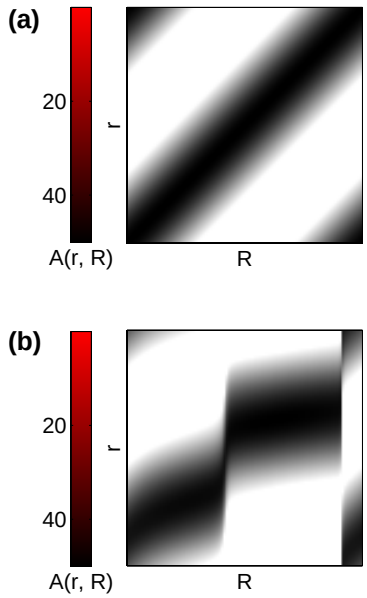


Figure 18: (a) Activations $A(r, R)$ in a simulation with a homogeneous distribution of neurons, and (b) in a simulation with randomly distributed neurons (data and parameters from 12), due to input distributions with maxima at position R. While in (a), receptive fields $A(r = r_{const.}, R)$ are of equal size and shape at all cortical positions r, in (b), receptive fields vary tremendously in size and shape, especially at positions where neuronal densities are small or, equivalently, cortical feedback is weak.

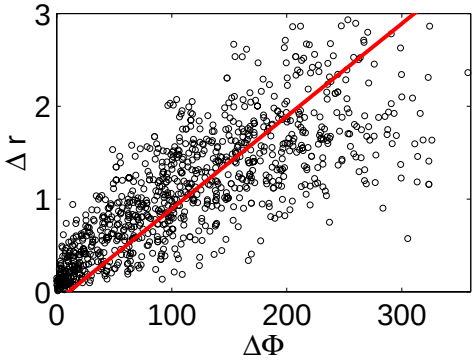


Figure 19: The distance, ΔR, of the centers of two neighboring receptive fields is an approximately linear function of the difference of preferred orientations, $\Delta\Phi$, of the corresponding neurons. Compare with 3. Parameters as in 14, stimulated with non-oriented stimuli G_1 with $\epsilon = 1$ and $\sigma_1 = 5$.