

RESEARCH ARTICLE

Florentin Wörgötter · Eckart Nelle · Bing Li
Lei Wang · Yun-chen Diao

A possible basic cortical microcircuit called “cascaded inhibition”

Results from cortical network models and recording experiments from striate simple cells

Received: 21 April 1997 / Accepted: 28 April 1998

Abstract The robust behavior, the degree of response linearity, and the aspect of contrast gain control in visual cortical simple cells are (amongst others) the result of the interplay between excitatory and inhibitory afferent and intracortical connections. The goal of this study was to suggest a simple intracortical connection pattern, which could also play a role in other cortical substructures, in order to generically obtain these desired effects within large physiological parameter ranges. To this end we explored the degree of linearity of spatial summation in visual simple cells experimentally and in different models based on half-wave rectifying cells (“push-pull models”). Visual cortical push-pull connection schemes originated from antagonistic motor-control models. Thus, this model class is widely applicable but normally requires a rather specific design. On the other hand we showed that a more generic version of a push-pull model, the so-called cascaded inhibitory intracortical connection scheme, which we implemented in a biologically realistic simulation, naturally explains much of the experimental data. We investigated the influence of the afferent and intracortical connection structure on the measured linearity of spatial summation in simple cells. The analysis made use of the relative modulation measure, which is easy to apply but is limited to moving sinusoidal grating stimuli. We introduced two *basic push-pull models*, where the order of threshold non-linearity and linear summation is reversed. Very little difference is observed with the relative modulation measure for these models. Alternative models, like half-wave squaring models, were also briefly discussed. Of all model parameters, the ratio of excitation to inhibition in the simple cell exerts the most crucial influence on the relative modulation. Linearity deteriorates as soon as excitatory and inhibitory inputs are imbalanced and the relative modula-

tion drops. This prediction was tested experimentally by extracellular recordings from cat area 17 simple cells and we found that about 62% showed a significant deviation from linear behavior. The problem that individual basic push-pull models are hard to distinguish experimentally led us to suggest a different solution. In order to generically account for the observed behavior (e.g., imbalance of excitation versus inhibition), we suggested a rather generic version of a push-pull model where it no longer mattered about (the hard-to-distinguish) fine differences in connectivity. Thus, we introduced a new class of biophysically realistic models (“cascaded inhibition”). This model class requires very little connection specificity and is therefore highly robust against parameter variations. Up to 25 cells are connected to each target cell. Thereby a highly interconnected network is generated, which also leads to disinhibition at some parts of an individual receptive field. We showed that the performance of these models simulates the degree of linearity and its variability in real simple cells with comparatively high accuracy. This behavior can be explained by the self-regulating properties of a cascaded inhibitory connection scheme by which the balance between excitation and inhibition at a given cell is improved by the joint network effects. The virtues and the generic design of this connection pattern, therefore, allow to speculate that it is used also in other parts of the cortex.

Key words Cortical microcircuits · Relative modulation · Push-pull mechanism · Simple cell · Cascaded inhibition · Cat

Introduction

The temporal and spatial (quasi-) linearity¹ of neuronal responses in the visual system is a key issue of visual cortical physiology because linear behavior is highly desirable from the viewpoint of information processing in

¹ The notion of linearity for nerve cell responses must remain incomplete, because negative impulse rates cannot be represented. Thus, neurons can “at best” act as half-wave rectifiers. In order to emphasize this, we used the term “quasi-linear” throughout this study.

F. Wörgötter (✉) · E. Nelle
Institute of Physiology, Department of Neurophysiology,
Ruhr-Universität Bochum, D-44780 Bochum, Germany
e-mail: worgott@neurop.ruhr-uni-bochum.de,
Fax: +49-234-709-4192

B. Li · L. Wang · Y.-C. Diao
Institute of Biophysics, Academia Sinica, Beijing, China

any neuronal network. Thus, this aspect has been addressed in a multitude of experimental studies (Enroth-Cugell and Robson 1966; Hochstein and Shapley 1976; Movshon et al. 1978a; Kaplan and Shapley 1982; DeValois and Tootell 1983; Enroth-Cugell et al. 1983; Tolhurst and Dean 1987; Palmer et al. 1991; Reid et al. 1991; DeAngelis et al. 1993). It forms an essential trait of simple cell models, and it is also a prerequisite for many models of image processing (Adelson and Bergen 1985; Emerson and Citron 1988; Emerson et al. 1989; McLean and Palmer 1989; Heeger 1993).

As a measure for the linearity of cortical simple cells we focus on the so-called relative modulation (RM; Movshon et al. 1978a) of the spike density (i.e., the peristimulus time histograms, PSTHs) of a simple cell response that is stimulated with a drifting sinusoidal grating. It is precisely defined in Eq. 2 and gives an estimate of the amount of stimulus-induced periodic response compared with the mean nonspecific activity of a cell. This measure is often used to distinguish simple from complex cells (Skottun et al. 1991) and previously has been applied to draw conclusions about the (quasi-) linearity of simple cell models (Tolhurst and Dean 1987). In addition, it can be determined readily in experimental situations, which is important for the experimental part of this study.

A very successful class of models of simple cells, the so-called push-pull models were introduced more than a decade ago (Glezer et al. 1980; Palmer and Davies 1981; Pollen and Ronner 1982; Ferster 1988; Tolhurst and Dean 1990); they include an inhibitory component and account much better for the properties of real nerve cells than previously used purely excitatory models. Push-pull models assume that an excitatory (e.g., on) input is counterbalanced by an antagonistic inhibitory (e.g., off) input that covers the same part of the receptive field. As has been already pointed out by Tolhurst and Dean (1990), there are several subclasses of push-pull models, and it is not clear how far it is possible to easily distinguish between them experimentally. The original push-pull model, as well as many of its "subspecies," concentrated on the spatial domain, which we also do here. (For a discussion of spatiotemporal aspects, see Jones and Palmer 1987; McLean and Palmer 1989; Baker and Cynader 1988; DeAngelis et al. 1993). For all push-pull models, the amount of excitatory and inhibitory input is a crucial issue, as has been experimentally shown by Ferster (1988). We systematically studied the sensitivity to an imbalance of excitation and inhibition in this paper. In order to test this experimentally, the theoretical studies were complemented by a series of *in vivo* recordings from visual simple cells in cat. The restrictive design of the push-pull models and the experimentally observed, rather large variability in the linearity of cortical simple cells (Movshon et al. 1978a) raises the question of how to improve the models such that they can account for the observed behavior in a more generic way.

This report is organized in the following way: In the first part we try to show that many simple cells are rather nonlinear (see also Movshon et al. 1978a) and that this

feature could arise from an imbalance between excitation and inhibition at the respective target cell. We show that (as opposed to other parameters) such an imbalance is most strongly reflected by the behavior of the relative modulation measure regardless of the cell model used. On the other hand, we demonstrate that different versions of push-pull models cannot be distinguished by the use of the relative modulation measure. Thus, it seems that more complex methods to distinguish between the different model subclasses are required. Opposite to this view, we try to advocate the idea that the high sensitivity of cortical simple cells to some basic parameters (e.g., balance of excitation and inhibition) may thwart all attempts to design a unique push-pull model description of them. Instead in the second part of the study we propose that most of the experimentally observed effects can be embedded in a generic design of the push-pull models called "cascaded inhibition." Push-pull models are in general not restricted to visual information processing. Instead they probably play a role in all cortical (and subcortical) circuits in which two antagonistic nonlinear inputs (e.g., in the motor system) have to cooperate in order to achieve different degrees of linear behavior. The rather parsimonious design of our model, which nevertheless covers the different degrees of linearity in a rather robust way, points to the possibility that this behavior of simple cells is an emergent property of the cortical network and does not require specific wiring. This emergent network property together with the generic applicability of the push-pull principle could thus be of broader relevance within cortical microcircuits. Recent models of cortical circuitry strongly emphasize that also other cortical cell characteristics (e.g., direction and orientation specificity) could be emergent network properties (Douglas et al. 1995; Somers et al. 1995). The discussion about a rather specific design of the connection structure in order to achieve the different degrees of linear behavior that are experimentally observed might, thus, be questioned by our more generic approach.

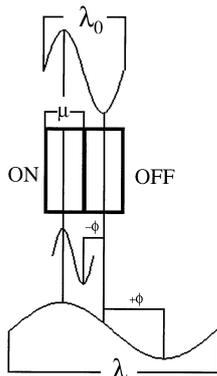
Materials and methods

Definitions for experiments and models

Response phase

Consider an idealized simple cell, with two subfields, responding to a moving sinusoidal grating stimulus of optimal orientation. We were interested in the (non)linearity of spatial summation between the subfields and hence the contributions of the individual subfields to the overall cell response. Such a range of problems could in principle be addressed by nonlinear systems theory. In order to make our results more easily comparable with those from previous studies, however, we took a simpler approach. Depending on the characteristics of the external stimulus, the responses of the individual subfields show different strengths and timing, leading to a temporal phase difference between the subfield responses: In an idealized simple cell, the on- and off-subfields are optimally separated (non-overlapping and without a "gap" between them) and of equal width, μ . In addition, we assume that there are no differences in the response latencies between the subfields. Therefore, there is only

Fig. 1 Schematic showing the definition of the temporal phase ϕ between the center of the off-subfield and the minimum luminance peak of the grating, when the maximum luminance peak is exactly located in the center of the on-subfield. λ is the spatial period for the grating with λ_0 corresponding to the optimal spatial period and, thus, to zero temporal phase. μ is the subfield width, with $\mu=\lambda_0/2$. See also Eq. 1



one optimal spatial frequency of the drifting grating, which we call $1/\lambda_0$, for which the on- and off-subfields are simultaneously stimulated with optimal illumination, depicted in Fig. 1 (top).

Let us associate t_0 with the moment when the positive luminance peak of the grating is exactly located in the center of the on-subfield regardless of the spatial frequency. This allows us to visualize the temporal phase ϕ between the time t_0 and the time corresponding to optimal coverage of the off-subfield by the negative luminance (darker than background) peak in Fig. 1 (bottom) given by the equation:

$$\phi = \frac{2\pi}{\lambda} \left(\frac{\lambda}{2} - \mu \right) = \pi - \frac{2\pi\mu}{\lambda} \quad (1)$$

where λ is the spatial period of the grating (inverse of the spatial frequency), μ the subfield width, and $\lambda \geq \mu$.

Some interesting cases are:

1. A very wide grating, $\lambda \gg \mu$, with $\phi \rightarrow +\pi$
2. The optimal grating, $\lambda = \lambda_0 = 2\mu$, with $\phi = 0$
3. A narrow grating, $\lambda = \lambda_0/2 = \mu$, with $\phi = -\pi$

Note that ϕ is ill-defined for $\lambda < \mu$, i.e., when the spatial frequency is so high that one (or more) full cycles cover a single subfield. This is directly related to the physiological finding of the so called null-response in the case of an ideally linear and space-time separable operator, which reflects the vanishing of the response if exactly two (opposing) half-cycles cover each subfield (Shapley and Hochstein 1975; Movshon et al. 1978a). The use of the ‘‘response phase’’ allows us to treat all real and model cells independently of their receptive field width.²

Relative modulation

To quantify the cell behavior, we make use of RM measure (see Movshon et al. 1978a), defined for inputs that vary sinusoidally in time. The RM measure can be applied on the spike-density function (PSTH) because the PSTH (as opposed to many other measurements of the cell activity) will directly reflect the spatial summation behavior of the cells. If the input-output transformation is linear, we expect the output to contain only the same frequency spectrum as the input. Hence, deviations from the ideal output frequency spectrum can be used to characterize the degree of nonlinearity of the transformation performed by the simple cell. Given a sinusoidal input of a specific temporal frequency, we can assign the input-output-transformation a family of RMs r_1, r_2, \dots , where the RM of order n is defined as

$$r_n = \frac{n^{\text{th}} \text{ harmonic component of amplitude spectrum}}{\text{d.c. component of amplitude spectrum}} \quad (2)$$

² Commonly, spatial frequency tuning curves are plotted on a log scale, which leads to shifts of the curves with changing receptive field widths. For our purposes the phase is much better suited because it eliminates the receptive field width dependence altogether.

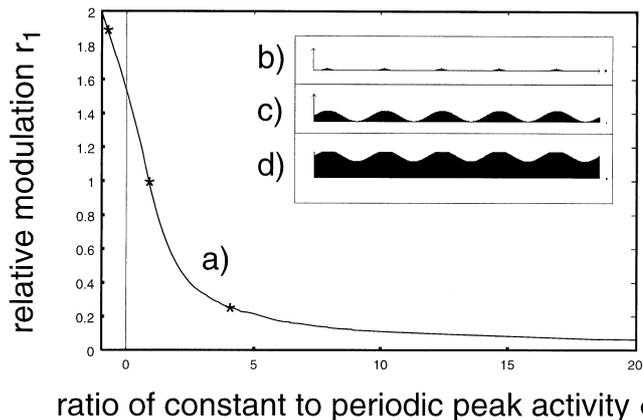


Fig. 2 Relative modulation for a simple model cell with sinusoidal input. The cell has a constant spontaneous activity S , which is added to the sine wave input to give the total activity before rectification. This signal is rectified at a given threshold to produce the output activity from which the relative modulation is determined. *a)* shows the relative modulation as the amplitude A of the sine wave input is held constant and the constant portion $\sigma=S/A$ of the activity is varied. The output is rectified at a threshold of zero activity to remain nonnegative. Relative modulation varies from $r_1=2.0$ for a negative spontaneous activity to 0 for a very high positive constant activity. A similar effect would be obtained for a fixed spontaneous activity and variation of the threshold $\tau=T/A$ from high to low values. *b)–d)* The insets show the corresponding output at intermediate stages marked by asterisks on the curve *b)* $S=-0.75$; relative modulation $r_1=1.9$. Only the crests of the sine wave are contained in the output. Even though the absolute output level is low, the output is timed very well around times of peak sinusoidal input. *c)* $S=1$; relative modulation $r_1=1.0$. Amplitude of sine wave and constant activity are equal. The variations of the input are reflected over the full output range. *d)* $S=4$; relative modulation $r_1=0.25$. Constant activity is 4 times higher than the amplitude of the sinusoidal input, leading to a low relative modulation of the output

The fundamental frequency or first harmonic is determined by the input, and the amplitude spectrum is calculated from the output signal. RM compares the higher order Fourier coefficients to the constant term of the Fourier decomposition $f(t) = a_0/2 + \sum_{i=1}^{\infty} [a_n \cos(nt) + b_n \sin(nt)]$. RM of order n measures how strongly the n th harmonic of the input frequency is represented in the output as compared to the overall activity of the output signal. Whenever we use ‘‘RM’’ without a specific order in the remainder of the paper, we refer to RM of order 1, r_1 . Defining the RM as a ratio leads to a normalization of r_i with respect to the overall output. Hence, the RM cannot directly be interpreted in terms of the absolute cell response. Whereas this does not present a problem for numerical treatment in the theoretical studies, experimentally the modulation frequency and its harmonics have to be detectable and separable from the noisy background in order to determine the corresponding RMs r_i .

Figure 2 shows the behavior of the RM for an input-output relation given by:

$$y^{\sigma, \tau, \psi}(t) = \begin{cases} A \sin(t + \psi) + \sigma A - \tau A, & A \sin(t + \psi) + \sigma A \geq \tau A \\ 0, & A \sin(t + \psi) + \sigma A < \tau A \end{cases} \quad (3)$$

This equation describes a possible PSTH of a simple cell, where the cell output $y(t)$ at time t is determined by the input amplitude A and the level of spontaneous activity $S=\sigma A$ and output threshold $T=\tau A$ of the cell. The parameters σ and τ reflect the state of the cell, as well as the contrast of the external stimulus. Because of their physiological interpretation, we treat τ and σ as separate variables instead of combining them into a single parameter. The input amplitude A in

the model is finite as the consequence of the response saturation already in the retina.

In spite of the threshold nonlinearity, one usually refers to the ideal half-wave rectifying case ($S=0$, $T=0$) as linear, since this is the closest that a half-wave rectifying cell without spontaneous discharge can follow the sinusoidal input. To avoid confusion with the case $r_1=1.0$, where the modulation of the output follows the full sine wave, we call this behavior quasi-linear. In this case we obtain a relative modulation of $r_1 = \frac{\pi}{2}$.

Experimental methods

Anesthesia and general procedures

Six adult cats (2–3.5 kg), inhousebred were used in the study. Following the international guidelines for animal research, the animals were initially anesthetized with sodium thiopental (50 mg/kg). Anesthesia was then maintained with urethane (20 mg/kg per hour), gallamine triethiodide (Flaxedil, 10 mg/kg per hour), glucose (200 mg/kg per hour) in Ringer solution (i.v.). The animals were artificially respired with room air and the end-expired CO_2 was kept at 3.8–4.0%. Atropine was used to dilate the pupils, phenylephrine for retraction of the nictitating membranes. We used contact lenses to keep the eyes wet and adjust the refraction to a viewing distance of 28 cm. Body temperature was kept at 37.5° C using a feedback-controlled heating pad. The electroencephalographic activity (EEG) above the frontal cortex and the electrocardiographic activity (EKG) were monitored to ensure the necessary depth of anesthesia. A local anesthetic, procaine (0.4%) was applied to wound and pressure points. An opening of the skull, was made above the projection area of the central ≈ 0 –30° of the visual field in area 17 (see also Wörgötter and Eysel 1991). Especial care was taken in order to obtain a stable experimental situation. After all, the measured linearity of a simple cell could in principle also be affected by changing the state of anesthesia, for example.

Recording, visual stimulation, and data collection

Extracellular recordings were made with single-barrel glass pipettes, with an outer tip diameter of 1.5–3 μm , filled with 4 M NaCl solution or glass-coated tungsten electrodes with a tip diameter of 2–4 μm at retinal eccentricities between 0° and 25°. Stimulus presentation and data acquisition were controlled by an 80386 PC. Visual stimuli were generated by a Picasso image generator (Picasso; Innisfree, Cambridge, Mass.) and presented on an oscilloscope screen (Tektronics 608) 28 cm in front of the cat's eyes. The stimulus was presented monocularly to the dominant eye.

Drifting sine wave gratings were used as stimuli. The spatial frequency was adjusted to different values mostly in 5–15 steps to cover a wide range of temporal phases, maximally $\pm 180^\circ$. Alternatively, recordings were ended with those two (high and low) spatial frequencies that did not elicit reliable responses anymore. The spatial frequency range we used in the total cell sample was between 0.05 cycles/deg and 3.0 cycles/deg. Contrast was set to a low value (40%) to avoid saturation effects and kept identical in all experiments. This procedure is generally used when assessing the linearity of simple cells, because at high-contrast values nonlinear effects are naturally induced partly due to saturation effects. Orientation, drift direction, and temporal frequency of the grating were adjusted individually for each cell to elicit optimal responses and remained unchanged during recording of the same cell.

The total recording time at an initial temporal resolution of 1 ms was 120–500 s for each spatial frequency tested, depending on the response strength of the cell and on the temporal frequency. This resulted in up to 900 grating cycles. Pauses were interspersed between every (approx.) 50 cycles to avoid adaptive effects. Spontaneous activity was measured before presenting the grating stimuli.

As a control, narrow light and dark bars ($0.5^\circ \times 10^\circ$) were moved at optimal orientation and optimal speed across the receptive field, which also provides an additional test for the cell class. Since re-

cordings were from area 17, slow velocities had to be used, between 0.2 and 10°/s. Simple cells usually required slow stimuli, while complex cell could often be stimulated with up to 10°/s. Between 10 and 50 sweeps of bar motion were made and responses were recorded also at an initial resolution of 1 ms.

Classification of the cells

We followed the suggestion of Tolhurst and Dean (Tolhurst and Dean 1987, 1990; Skottun et al. 1991) and based our cell classification mainly on the value of the relative modulation. Cells with a mean RM below 1.0 were classified as complex cells, those with a mean value above as simple cells. Averaging was performed for the three or four spatial frequencies closest to zero temporal phase. In addition we assessed the subfield overlap and found in almost all cases consistently that cells with low relative modulation values had highly overlapping subfields, while the relative modulation grew bigger with decreasing overlap.

Data analysis

The recorded spike trains were transformed into standard PSTHs showing either the response to the bar (20 ms/bin) or to the grating (2 ms/bin) stimuli. When using a grating, PSTHs representing 5 subsequent cycles were displayed for graphical reasons, for Fourier analysis the five sweeps were combined into one. Curves of the total mean response per cycle versus the spatial frequency were also computed. The PSTHs for the light and dark bars were used to measure the receptive field width and the overlap between subfields. This was done in a straightforward way by approximating the response peaks with triangles and subsequently measuring them.

In order to compute the response phase (Eq. 1), we followed the general convention and assumed that the optimal spatial frequency $1/\lambda_0$ was the one that elicited the strongest response to a grating. This corresponds to the peak of the spatial frequency tuning curve (Ikeda and Wright 1975; Movshon et al. 1978b; Andrews and Pollen 1979). The subfield width can be deduced from the optimal spatial frequency, for which the subfields should be optimally covered, with $\mu = \lambda/2$, such that $\phi=0$ as soon as $\lambda = \mu$.

Cell counts

A total of 84 nondirection-selective simple cells were recorded in layers II to (mostly) IV of area 17 as judged by the recording depth. In addition, 23 complex cells were recorded for comparison. For 71 simple and 18 complex cells, curves of the relative modulation versus temporal phase were obtained. Sixty simple cells of those could be used for analysis, because in the other 11 the curves were either incomplete or those cells had a significant spontaneous activity. We excluded all cells that were spontaneously active with mean of more than 1 Hz, as well as cells with a directional tuning of more than $\text{DI}=0.5$ (nondirection selective cells; Orban 1984) from this study, because this also affects the relation between RM and temporal phase (see Discussion). The spontaneous rates of the 71 simple cells were distributed as follows: less than 0.3 Hz, 16 cells; between 0.3 and 0.6 Hz, 34 cells; between 0.6 and 1.0 Hz, 10 cells; more than 1.0 Hz, 11 cells (excluded from the analysis). Responses to moving dark and light bars were recorded in 46 of the 71 simple and in 14 complex cells.

Models

Setup of the basic push-pull models

Consider a simple cell in area 17 with a receptive field divided into several on and off subfields (e.g., two subfields in Fig. 3). To model such a receptive field structure, we combine half-wave rectifying cells as given by Eq. 3 to a push-pull model of a simple cell in area 17.

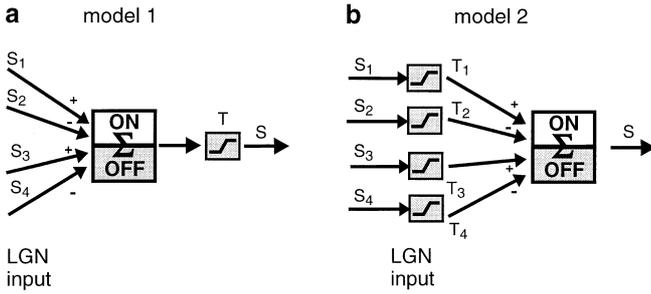


Fig. 3a, b Push-pull models of Tolhurst and Dean. Threshold parameters are labeled T , those indicating the spontaneous activity level, S . **a** Model 1: first summation, then submission to threshold. **b** Model 2: first threshold process, then summation. Only nonzero thresholds are shown (*LGN* lateral geniculate nucleus)

The output from a model push-pull cell is then given as:

$$z^{\sigma, \tau}(t) = \begin{cases} u(t) + \sigma A - \tau A, & u(t) + \sigma A \geq \tau A \\ 0, & u(t) + \sigma A < \tau A \end{cases} \quad (4)$$

where we define:

$$u(t) := \sum_{i=1}^4 w_i y_i^{\sigma_i, \tau_i, \psi_i}(t) \quad (5)$$

$y_i^{\sigma_i, \tau_i, \psi_i}$ is defined according to Eq. 3, and w_i are positive for excitatory and negative for inhibitory input. Two subfields are excitatory, two inhibitory, leaving us with a summation of four fields in total. For model 1 in Fig. 3 (first summation, then threshold process), we have:

$$\sigma_i = 0, \quad \tau_i = 0 \quad (i = 1 \dots 4) \quad \text{and} \quad \sigma, \tau \geq 0.$$

Model 2 is characterized by:

$$\sigma = 0, \tau = 0 \quad \text{and} \quad \sigma_i, \tau_i \geq 0 \quad (i = 1 \dots 4).$$

In addition, we will briefly discuss push-pull models that include a half-squaring mechanism (Heeger 1992a, b, 1993). For these models, the output is squared (after adding spontaneous activity and application of the threshold), i.e., y_i has to be replaced by $y_i \times y_i$. For the push-pull model, this implies that the parameters of the simple cell have to be scaled accordingly.

Biophysiologicaly realistic model

In the *biophysiologicaly realistic model* we implement the push-pull models in a large-scale simulator, which is built in a modular fashion. Only a sketch of the model and the properties relevant for the receptive field structure of the simple cells can be given here. (Details can be found in Wörgötter and Koch 1991 and Brettle and Niebur 1994.) The model setup is as follows: We simulated the membrane potential of about 16000 nerve cells in the primary visual on- and off-pathway of X cells. Cells were modeled as a single compartment with a leakage current, membrane capacitance, afterhyperpolarization, and a conductance change for each excitatory or inhibitory input described by an alpha function. The pathway modeled led from retinal ganglion cells via (*LGN*) relay cells to the primary visual cortex V1.

Since we wanted to make use of the relative modulation in our studies, we stimulated the model cells with a moving sine wave grating with a velocity of 1.5°/s, such that each point in the receptive field saw at least five periods of the stimulus.

The grating stimulus was temporally and spatially filtered in the simulator to mimic the processing of the retina. The output was converted into ganglion cell spikes, which served as input to the upper stages of the simulator. Each of the ganglion cells innervated four neighboring geniculate cells such that retinal topography was preserved. Projections of geniculate cells were arranged in the fashion

proposed by Hubel and Wiesel (1962) such that a nearly rectangular array of neighboring *LGN* cells projected excitatorily onto a single simple cell in area 17. One such array of receptive field centers of *LGN* cells corresponded to a single receptive subfield of a simple cell in area 17. A mean of 3 subfields with the same orientation, each consisting of the about 3×23 presynaptic *LGN* cells, was chosen for each simple cell. Because of the high *LGN*RF overlap, this arrangement resulted in an aspect ratio of about 1:3.2 of a cortical subfield. In order to account for variations in the projections, we selected only about 30% of the *LGN* cells at the right location to form excitatory projections to a cortical cell. The preferred orientation of the cortical simple cells varied continuously from 0° to 180° over the simulated cortex population. A total of 2.5 hypercolumns was simulated in our setup. Details that directly affect the results of the push-pull setup in the realistic model are described in Biophysical models.

Results

Basic push-pull models

We first studied simple, so-called basic push-pull models, in order to determine the fundamental and generic properties of push-pull models for simple cells in response to moving sinusoidal gratings. The study of the basic models led to predictions for the experiment and formed the basis for the much more elaborate realistic models. One can distinguish two versions of the basic push-pull model (Tolhurst and Dean 1990), differing in the order in which summation and static threshold were applied to the input of the simple cell. In model 1, the individual contributions to the input were first added and then submitted to a threshold. In model 2, first each contribution was submitted to an individual threshold and then the remaining signals were summed (see Fig 3). Both linear-nonlinear (*L-N*) and *N-L* type models are part of a more general *N-L-N* cascade model (Emerson et al. 1989; Jacobsen et al. 1993). However, we did not mix both versions, in order to evaluate their individual properties. We determined the linearity of spatial summation in the model as expressed in the *RM* when the spatial frequency was systematically varied. For simplicity, we set the level of spontaneous activity and output threshold to zero (hence, model 1 and model 2 became identical). The inhibitory inputs were assumed to be 180° out of temporal phase to the response of their antagonistic excitatory partners (corresponding to a perfect match of the antagonistic receptive fields). In the case of *equal amplitude* of excitatory and inhibitory input to the simple cell, the *RM* was *independent* of the temporal phase difference of the subfield responses, corresponding to quasi-linear spatial summation over the whole range of spatial frequencies (Fig. 4a, solid line). However, for real cells, there was no a priori reason why the amplitudes of excitatory and inhibitory input should be exactly balanced. When the amplitudes of excitatory and inhibitory inputs were different, leading to a *weighted push-pull model*, the relative modulations became *phase-dependent* (Fig. 4a). Figure 4b shows the dependence of the first four relative modulations on the ratio of the inhibitory and excitatory amplitudes for a fixed phase difference of 90° between the subfield responses.

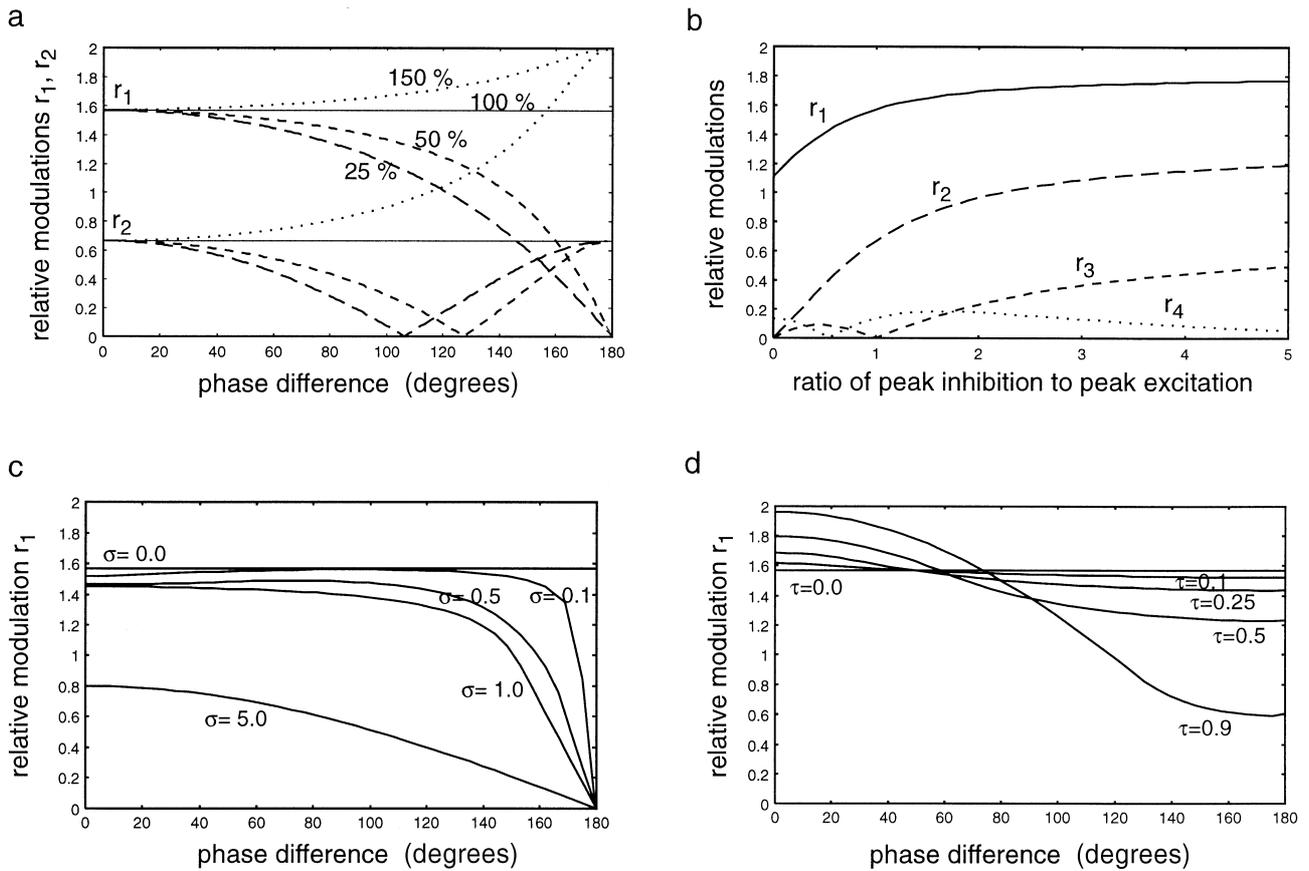


Fig. 4a–d Dependence of the first few relative modulations on temporal phase difference and ratio of inhibition to excitation in the basic push-pull models. The setting in **a** and **b** is $T_i=T=0$; $S_i=S=0$. **a** Inhibition fixed at 150%, 100%, 50%, and 25% of the excitation. Relative modulations r_1 and r_2 vary with temporal phase difference, except for the 100% case. For 180° the output vanishes and the relative modulations are set to their limiting values from below. Phase dependence of relative modulation is a generic feature of the basic push-pull models. For $S \neq 0$, $T \neq 0$ we obtain a phase dependence also in the 100% case. **b** Phase difference fixed at 90° . Relative modulations r_1, \dots, r_4 vary with ratio of inhibition to excitation. A high relative level of inhibition leads to a high relative modulation r_1 . In **c** and **d**, $S_{(ij)} \neq 0$ and $T_{(ij)} \neq 0$. **c** Now constant activity is added, $s = \frac{1}{2}\sigma_i > 0$; $T_i = T = 0$. The diagram contains the phase dependence of the relative modulation r_1 for various settings of σ . **d** In a similar fashion, here we have $\tau_i = \tau > 0$; $\sigma_i = \sigma = 0$, again showing the phase dependence of r_1

When we relaxed the restrictions on the spontaneous activity and threshold, and chose parameter settings $S, T > 0$, even for equal excitatory and inhibitory input amplitude we observed a phase dependence of the relative modulations. This is shown in Fig. 4c, d. Hence, phase dependence of relative modulation is a generic property of the basic push-pull models. However, the phase dependence observed for threshold and/or spontaneous activity changes are relatively moderate as compared to those from unequal excitation and inhibition. Only for unphysiological values of threshold and spontaneous activity was a strong phase dependence observed.

We then evaluated the two kinds of push-pull models (Fig. 3) under ideal conditions, i.e., same strength of excitatory and inhibitory input over a wide range of param-

eters σ and τ . To limit the number of free parameters, we arbitrarily chose a temporal phase of 90° between on and off responses. Since we could not study all possible different parameter combinations, we scaled all threshold base values by a common threshold factor τ , all base levels of spontaneous activity by a common factor of spontaneous activity σ . Thus, we evaluated two-dimensional sections of the full parameter space. Figure 5 shows the resulting variation of the relative modulation r_1 in the two push-pull models.

In both models the relative modulation changes slowly with varying parameter values in physiological ranges, and many different parameter combinations lead to the same relative modulation. In particular, for small ($T_i \lesssim A/5$, $S_i \lesssim A/5$), physiologically relevant values of constant activity and threshold there is little difference between the two models, making it impossible to gain conclusive evidence from the RM values in favor of one kind of the basic push-pull models. This shows that the threshold operation (beyond half-wave rectification), irrespective of the order in which it is applied, is a much less sensitive parameter for the observed non-linearity than the ratio of excitation to inhibition.

Finally, we studied the effects of using half-wave squaring cells (Heeger 1992a, b, 1993) instead of half-wave rectifying cells in the basic push-pull models. For a single cell, the relative modulation r_1 at zero constant activity level with zero thresholds has a value of about 1.7, as compared to about 1.57 in the case of a single, purely half-wave rectifying cell. Generally, the relative

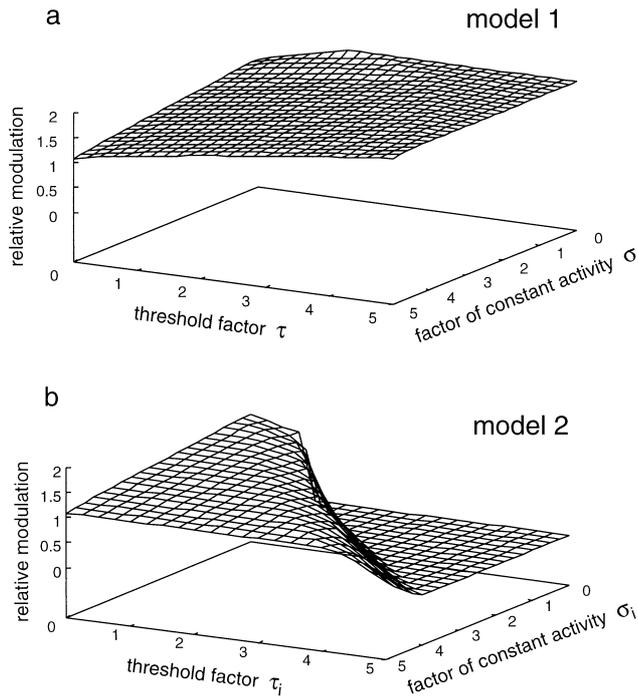


Fig. 5a, b Effect of parameter variation on the relative modulation r_1 in the basic push-pull models. Parameters are given in units of the common amplitude A . The phase difference is fixed at 90° . **a** Model 1: Relative modulation varies slowly and smoothly with level of constant activity and firing threshold. For large threshold values relative modulations close to 2.0 can be obtained, as in the single-cell model, which is strongly reflected in this model. **b** Model 2: Generally the relative modulation is smaller than in model 1, even for a low threshold. For high threshold values of the input cells the weak input to the simple cell results in a very low relative modulation. Where the threshold exceeds the activity, the relative modulation has been set to zero. Note that relative modulations for responses for which the threshold exceeds the total input (and, hence, there would be no output) have been set in **a** to 2.0 and in **b** to zero to obtain a smooth diagram. The level of constant activity has been set in the simple cell to half the value of presynaptic cells

modulation r_1 is slightly larger than for half-wave rectifying cells. This is due to the squaring, which emphasizes the high levels of the periodic activity. A principal difference to the half-wave rectifying model lies in the fact that, because of the reduced temporal overlap of the subfield responses, the relative modulation r_1 remains close to 1.4 in model 2 as the phase difference approaches 180° . Also, already for equal amplitudes of excitatory and inhibitory input we find a phase dependence of the relative modulation (see Fig. 6). Again the relative modulations also depend on the ratio of excitatory to inhibitory input for a fixed phase difference (data not shown). Whereas in principle the change of relative modulation with phase difference is a way to distinguish between purely rectifying cells and cells that combine a rectification with a squaring operation, under realistic conditions, the deviations between the results remain too limited to gain conclusive evidence.

In summary, it can be noted that the only parameter that has a strong influence on the RM is the balance be-

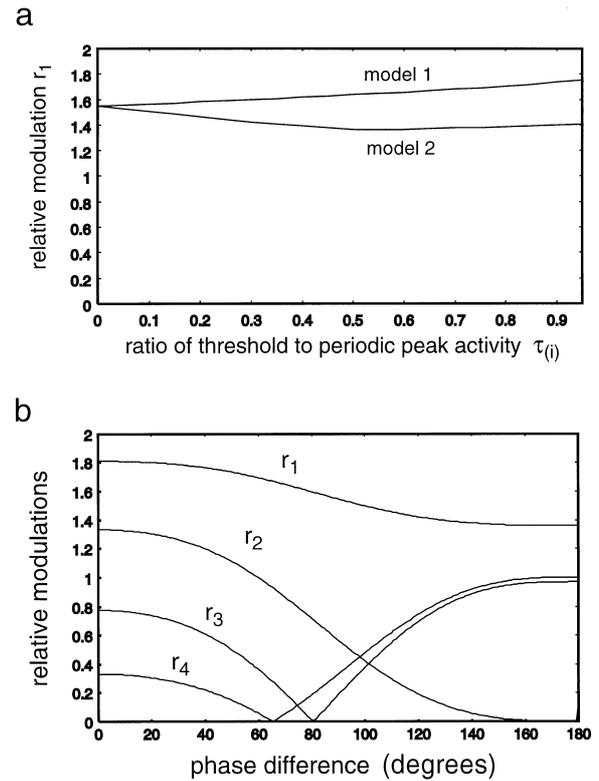


Fig. 6a, b Model with half-squaring cells. **a** Relative modulation r_1 as function of threshold in models 1 and 2 for a fixed phase difference of 90° , equal excitatory and inhibitory amplitudes, and zero spontaneous activity. As for the half-wave rectifying case, the relative modulation in model 2 is less than in model 1. Because of the reduced temporal overlap of the excitatory inputs, in both models the maximum total input does not exceed A , i.e., $\tau \leq 1$. **b** The first four relative modulations as the temporal phase difference is varied. Spontaneous activity and thresholds are set to zero; hence model 1 and 2 are equal. Excitatory and inhibitory inputs possess the same amplitude. The dependence of the relative modulations on phase difference is a distinctive feature of half-wave rectifying and half-wave squaring cells as r_1 remains close to 1.4 when the phase difference approaches 180°

tween excitation and inhibition, which will be used in the experimental part of this study to draw conclusions about the convergence pattern onto cortical simple cells.

Experimental results

Dependence of the relative modulation on the temporal phase of the grating

Figure 7 shows the behavior of three simple cells stimulated with drifting gratings of different spatial frequencies. The cell in Fig. 7c reflects a borderline case between simple and complex cell behavior. A selection of the recorded PSTHs to gratings of different spatial frequencies is shown in plots 1–5 of Fig. 7a–c. The time axis has been normalized in these plots to always show five subsequent cycles. The diagrams marked with short bars show the response to a moving light or dark bar at optimal direction and orientation. To obtain the curves of relative modula-

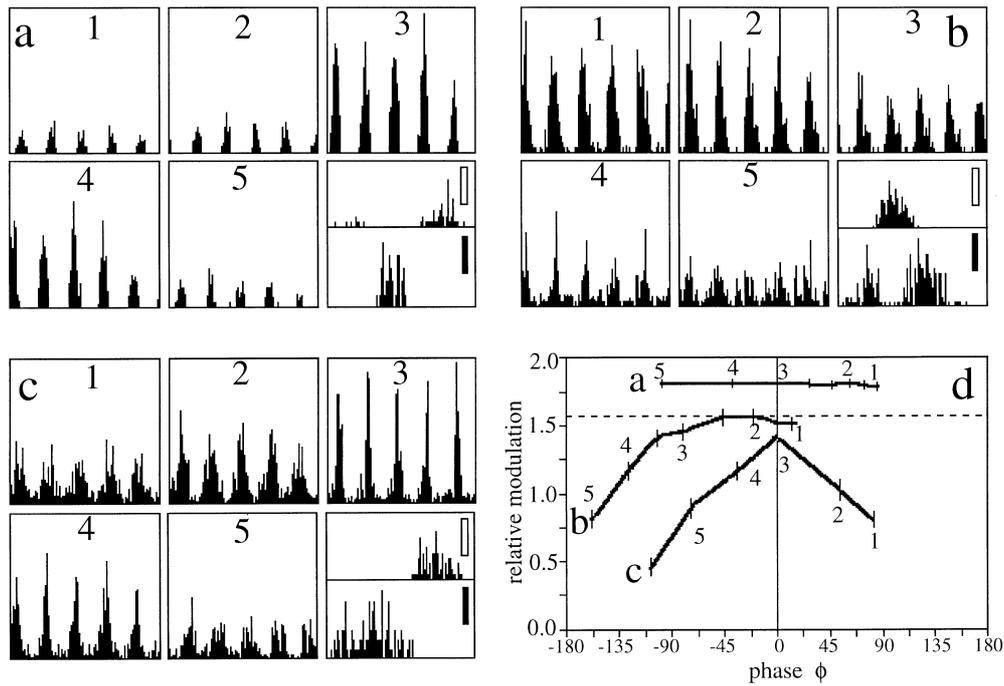


Fig. 7 PSTHs (a–c) and relative modulation versus temporal phase curves (d) of three simple cells. In a–c, the PSTHs (bin width 8 ms) of the responses to gratings of five different spatial frequencies (in order, from low to high) are displayed in panels 1–5, and in the last panel the responses to moving light and dark bars (marked with short bars) at optimal orientation and direction are shown (bin width 20 ms). Spontaneous activity was less than 0.3 Hz for these three cells. **a** The optimal spatial frequency located at $\phi=0$ is $SF=0.40$ cycles/deg.; drift frequency of the grating $DF=0.45$ Hz; preferred orientation $PO=88^\circ$, velocity of the moving bar $V=0.70^\circ/s$; total receptive field width $RF=2.24^\circ$. Vertical scale for the grating responses $SC=45I/s$. **b** $SF=0.29$ cycles/deg.; $DF=1.55$ Hz; $PO=24^\circ$, $V=1.7^\circ/s$; $RF=3.57^\circ$, $SC=43I/s$. **c** $SF=0.25$ cycles/deg.; $DF=1.68$ Hz; $PO=10^\circ$, $V=1.6^\circ/s$; $RF=5.80^\circ$, $SC=25I/s$. **d** Relative modulation versus temporal phase curves. More spatial frequency values than shown in a–c were used to plot the curves, and the frequency values have been converted into temporal phase values with Eq. 1

tion versus temporal phase, we defined that particular spatial frequency as optimal that produces the highest mean impulse rate per sweep. Thus, we assumed that this spatial frequency reflected twice the subfield width μ and was therefore associated with the temporal phase $\phi=0$ (see Eq. 1). The curves show three distinctive shapes.³ The cell in Fig. 7d, a, has a high mean RM and the curve is flat; in Fig. 7d, b, the curve drops to low RM values for negative values of ϕ . In Fig. 7d, c, the curve has a peak at $\phi=0$. Note, that the examples were selected to show the shape of the curves and not to give maximal relative modulation values. Therefore, they were selected to render nonoverlapping curves in Fig. 7d. From Eq. 1 it is clear that ϕ only reaches the value of $+180^\circ$ asymptotically for zero spatial frequency. Therefore, in almost all cases recordings had to be stopped at values of ϕ much smaller for two reasons: Either the cell would stop to respond pe-

³ In Figs. 5, 6, and 11, we show the curve only for positive values of ϕ because for the models discussed there the curve is symmetrical.

riodically to such a wide grating as in Fig. 7c, or – more often observed – the cell would start to show nonlinear biphasic responses so that the second harmonic of the spectrum dominates. This case was observed for the cell in Fig. 7b.

The examples in Fig. 7 were chosen to represent the typical behavior of the complete cell sample, which is displayed in Fig. 8. Two basic types of behavior were observed in the curves of the relative modulation versus temporal phase. In 23 (38%) cells, the curves were flat. Judging from the recording depth, we found these cells almost exclusively in the middle of layer IV. In another 37 (62%) cells, they showed a drop of the relative modulation to either both sides of the origin or in cases such as in Fig. 7b for negative values of ϕ . To distinguish between flat and not-flat curves, the mean RM was computed for all data points on the curve. If the data points were scattered randomly around the mean, the curve was regarded as flat. On the other hand, if the data points close to $\phi=0$ were larger than the mean and those for $|\phi| \gg 0$ were smaller than the mean, the curve was regarded as not flat.

In Fig. 8c the curves for eight complex cells are shown for comparison. These cells were selected to show at least a small degree of modulation (modulated C cell; see Pollen et al. 1978). The curves for the complex cell are almost always below 1.0. In Fig. 8d the curves in Fig. 8a–c were averaged. To achieve this, the original curves, which are plotted by linear interpolation between the data points, were sampled with 10° steps and averaging was performed on these samples. Error bars are not drawn, because the variance can be assessed from the original set of curves in Fig. 8a–c. The difference between the three data sets is quite obvious. The mean RM changes in the interval of $\phi=\pm 90^\circ$ by approximately 30% for a rather large group of simple cells. In addition, the maximum mean RM modulation is smaller for these cells than for the

Fig. 8a–d Relative modulation versus temporal phase curves of the complete cell sample. **a** Twenty-three simple cells with flat curves; **b** 37 simple cells with non-flat curves; **c** 8 modulated complex cells. **d** Mean curves for the different samples shown in **a–c**

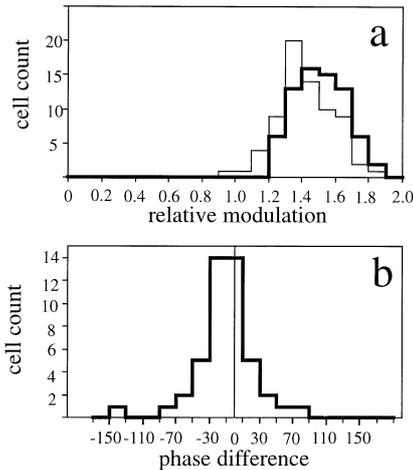
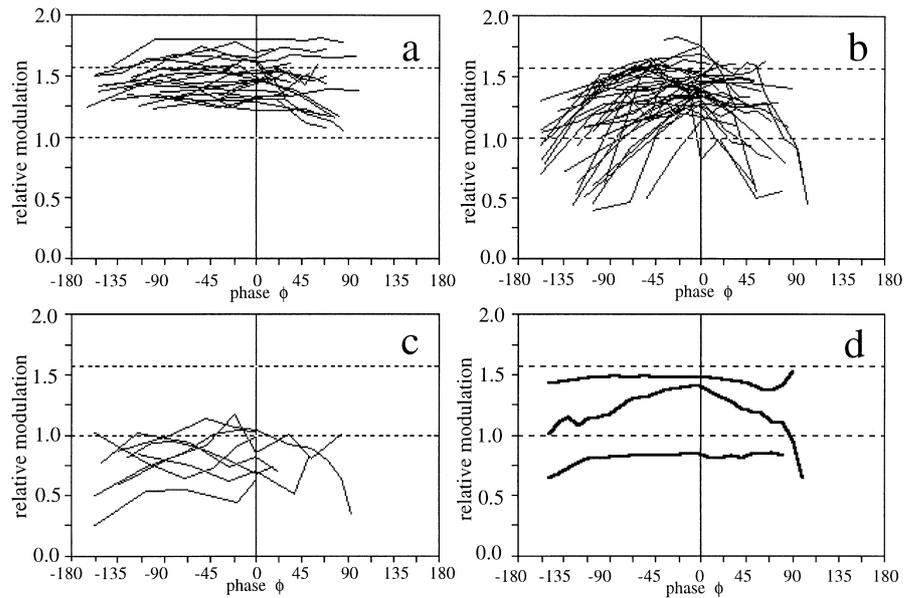


Fig. 9a, b Statistical results for the relative modulation versus temporal phase. **a** The distribution histograms of the relative modulation values at optimal spatial frequency (*thin line*) and the maximum relative modulation values (*thick line*) are shown. A small difference can be seen between the two histograms, and the mean values are 1.42 ± 0.17 and 1.51 ± 0.15 , respectively. A total of 71 simple cells is shown. **b** The distribution histogram of the temporal phase differences between $\phi=0$ and the temporal phase where the maximum relative modulation was found. A total of 46 simple cells is shown, because only the curves with a distinctive maximum are included. The mean value is $-11.37^\circ \pm 24.01^\circ$

group with a flat characteristic. This result might indicate different weights of the excitatory and inhibitory inputs.

Statistics

So far we have assumed that the optimal spatial frequency, which is associated with the value of $\phi=0$, is the one with the strongest response to the grating. The question arises whether this frequency is also associated with the

highest relative modulation value which would reflect the “most linear” (quasilinear) response. In Fig. 9a the distribution of the relative modulation values at $\phi=0$ (thin line) and the maximum relative modulations (thick line) are plotted as histograms. There is a small tendency for a difference between the two histograms, but it does not reach statistical significance at the 5% level. The means are: 1.42 for $\phi=0$ and 1.51 for the maxima. Both cases are below 1.57, which would represent the ideal situation given the cells act as half-wave rectifiers.

The histogram in Fig. 9b tries to answer the question of at what phase ϕ the maxima are found in relation to the chosen optimal spatial frequency (i.e., in relation to zero phase). For this histogram we could only use those curves that have a distinctive maximum, namely the curves in Fig. 8b and a few others excluded from Fig. 8 (curves too short). We plotted the distribution of the values of ϕ where the maxima were located and found small shift toward negative phases, i.e., higher spatial frequencies (mean $\phi = -11.4^\circ$). Therefore, the maximum RM is on average found for a spatial frequency higher than that which produces the maximal response. This difference essentially reflects the trend found in Fig. 9a, but it also does not reach statistical significance. Therefore, we conclude that the best quasi-linear responses for simple cells are indeed obtained at that spatial frequency for which they respond maximally.

Biophysical models

The basic push-pull models are highly abstract and do not directly incorporate biophysical parameters such as the membrane characteristics of the cells. In addition, their design cannot directly be compared with the actual wiring in a sub-cortical network. In particular, network effects that contribute to the receptive field properties of simple cells are not contained in the basic models. Furthermore,

an explanation of the experimentally observed cells with a flat RM curve would require a rather precise balance of excitation and inhibition over a wide range of stimulus conditions. Therefore, in this section we introduce a biophysically much more realistic model for cortical simple cells, and we concentrate on the possible role of the local cortical network in generating the experimentally observed quasi-linearity of spatial summation. In particular, we focus on three different connectivity patterns for antagonistic intracortical inhibition, corresponding to the push-pull setup described in the section Basic push-pull models. Firstly, the *strict, sparse inhibitory model*: at most two simple cells of precisely matching but antagonistic receptive field properties are connected to any simple cell. The target cell in turn inhibits these cortical source cells, resulting in an overall mutual inhibition (Fig. 10a). The specific type of inhibition (subtractive or shunting) is not crucial. Important for a linear operation is the fact that action potentials are eliminated at the “right” times, as prescribed by the input. We allow two cortical simple cells to project to a simple cell to give room for a small variability in order to make the model more robust (in Fig. 10a, b only one cell is shown).

Secondly, a more discretionary model, termed *weak sparse inhibition model*: the requirement of alignment of the preferred orientations and relative positions of the receptive fields are relaxed. But we still connect, at most, two simple cells to a given target cell (Fig. 10b).

Thirdly, a new introduced *cascaded intracortical inhibition model*: about 20–25 simple cells with loosely matching but antagonistic receptive fields are connected to a given target cell (Fig. 10c). The number of receptive subfields for each simple cell may vary between 2 and 4 for all three inhibition schemes. Models corresponding to the sparse mutual intracortical connection setup (i.e., push-pull models) have already been proposed (Glezer et al. 1980; Palmer and Davies 1981; Ferster 1988). In addition, there is experimental evidence for the existence of antagonistic, mutual intracortical inhibition that could subserve a push-pull mechanism (Liu et al. 1992).

The comparatively large number of intracortical connections in the cascaded connection scheme inevitably introduces statistical fluctuations of the contributing individual receptive fields, leading to a greater robustness of the model and resulting in more specific receptive fields of individual simple cells. In addition, the intracortical network allows even net disinhibition as a secondary-order effect. The cascaded inhibition scheme therefore tolerates a much larger variation of individual receptive field properties than the sparse mutual inhibition scheme, which requires a highly accurate matching of receptive field properties. It needs to be emphasized that we did not explicitly model inhibitory interneurons. Including those, however, would only slightly alter the observed behavior. Furthermore it should be noted that cascaded inhibition also often results in a “shared inhibition,” originating from one given source cell and terminating at different target cells – a situation that is often found in the real cortex.

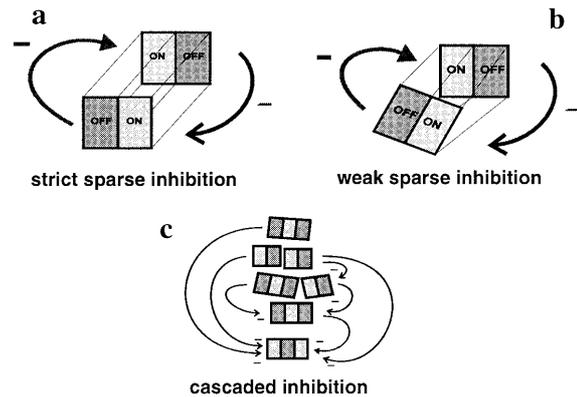


Fig. 10 Strict (a) and weak (b) sparse and cascaded (c) intracortical inhibition (schematic). Note, for simplicity, inhibitory interneurons that can mediate the inhibition are not drawn. In the biophysical model, inhibitory interneurons have not been modeled explicitly, but neurons were assumed to be able to provide inhibition and excitation at the same time for different targets. A similar approach has been adopted by others (Wörgötter and Koch 1991; Somers et al. 1995) without introducing significant differences in the cell behavior as opposed to an explicit implementation of inhibitory interneurons. For the sparse models, mutual inhibition is indicated by the *arrows*. For the cascaded model, reciprocal connections are not drawn to clearly show the sample wiring for a single cascade of seven cells. In this example, several “target cells” (at the *arrow end-points*) receive inhibition from several (different or same) “source cells.” In the model cascades can consist of up to 25 source cells for a single target. Cascaded inhibition allows higher order effects of cells not only inhibiting the target cell but also other presynaptic cells, possibly leading to a reduced inhibition or even net disinhibition of the target cell

The receptive field setup sketched so far presents an idealized case in which the receptive fields of all simple cells strongly resemble each other. To account for the variance found in nature, we have generally assigned model parameters a Gaussian-distributed jitter. In particular, a small jitter in the size (up to 0.1 deviation from an ideal grid position and up to 25% subfield overlap) and orientation ($\pm 10^\circ$) of receptive subfields has been introduced. In addition, the propagation delays are generally Gaussian distributed, with a standard deviation of 25% of the mean values.

We then investigated the behavior of the three models for stimulation with a moving sinusoidal grating of a spatial frequency, which corresponds to optimal stimulation of an idealized simple cell. Cells were arrayed in the model on a rectangular grid whose columns corresponded to a particular, preferred orientation of the simple cells. The resulting relative modulation r_1 in the cascaded model for all LGN and all cortical cells is shown in Fig. 11a. Whereas the relative modulation is about 1.57 (dark shading) in the LGN cells, the relative modulation in the cortex is only high for cells with a nearly optimal orientation. This is due to the fact that the overlap of the subfields with the stimulus halfwaves drops with the angle between preferred orientation and stimulus orientation. In Fig. 11b, the mean RMs for all thalamic cells in a column and all cortical cells in a column are plotted. The standard deviation of the RM is indicated for the cortical cells only.

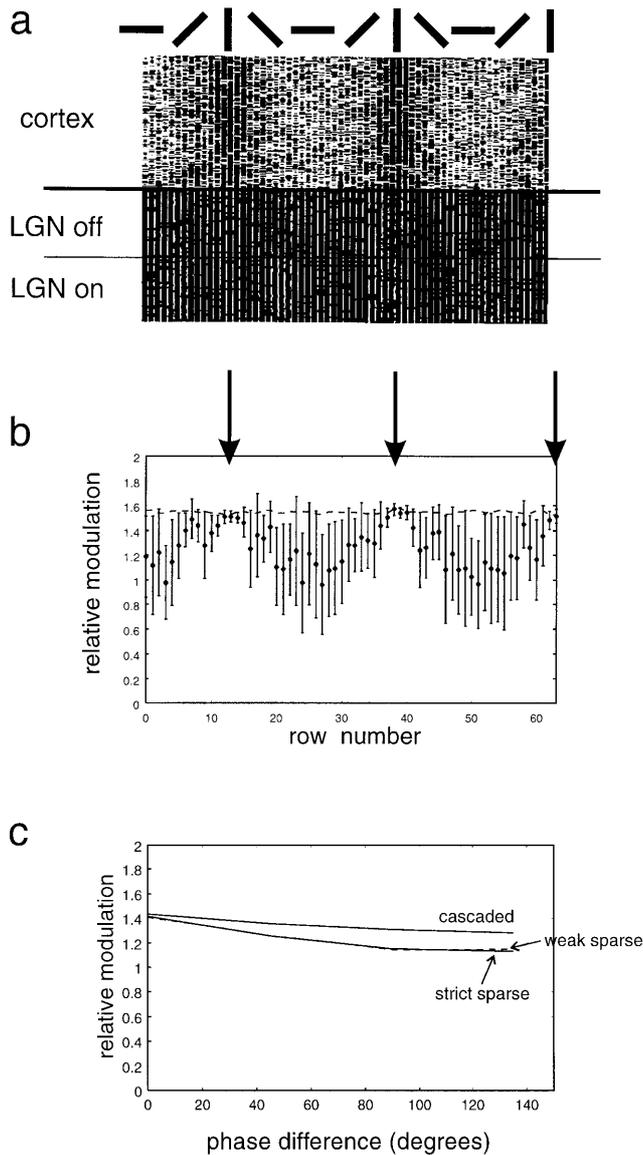


Fig. 11 **a** The RM is plotted in a population plot (for similar plots see Wörgötter and Koch 1991) for all 16384 cells in the simulated area of LGN and V1. Cells are arranged on a rectangular grid of 64×256 ($x \times y$) cells. The RM r_1 is plotted as a *black rectangle* for each cell. A *large rectangle* indicates high r_1 , a small rectangle, low RM. The stimulus size has been chosen to result in a phase difference of 0° . Cortical cells are connected according to the cascaded model. LGN cells possess a nearly constant RM of about 1.57. In the cortex, r_1 is only high for simple cells whose preferred orientation matches that of the stimulus. The *bars at the top* indicate the preferred orientation of the simple cells at the particular rows. **b** All RMs in a column are collapsed into a single value for LGN (*dotted line*) and cortex (*dots with error bars*). RM of the cortical cells coincides with a very small error with the relative modulation of the precortical cells only for the optimal orientation. Elsewhere we observe a sinusoidally shaped dependence of RM on the preferred orientation of the simple cells. **c** The phase dependence of RM in the three detailed models, based on the RM at 0° , 45° , 90° , and 135° phase difference. Only cells with optimal orientation preference have been selected, and the mean out of all cells with $r_1 \geq 1.0$ has been taken. Because of the skewness of the RM distribution, the mean RM is lower than the ideal value of 1.57. Whereas the relative modulation in the cascaded scheme only shows a weak phase dependence, the RM in the strict and weak sparse models falls off more strongly with increasing phase difference

Note, that the spread of relative modulations has a strong minimum for the optimally oriented cells. Some jitter is introduced in the orientation preference of the cortical model cells because of differing subcortical connections.

Out of the total of 8192 simulated cortical cells, we only used those cells whose preferred orientation matched that of the stimulus. Hence, we selected all cells in the columns corresponding to the stimulus orientation (marked with an arrow in Fig. 11b). To avoid pitfalls during the Fourier transformation, we excluded simple cells with an activity below $1/s$ from further analysis. The RM values for the remaining cells were sorted into bins of width 0.1 and the distribution for all cells with different relative modulations was drawn as a histogram for the three different models (Fig. 12).

Since all three models belong to the push-pull class, we did not expect drastically different behavior. Most notable was the difference between distributions corresponding to the sparse inhibitory and the cascaded inhibitory connection models. Whereas the former showed a bimodal distribution of relative modulations, in the latter we observed only a single, somewhat more pronounced peak around a relative modulation of 1.5. This means that the cascaded model allowed more simple cells to respond efficiently, i.e., with a substantial stimulus-induced modulation, to the input stimulus. If the simple cells in our model were true half-wave rectifiers, we would expect the peak number of cells to be clustered precisely around $r_1 = 1.57$. If one of the other parameters, such as the ratio of excitation and inhibition or the level of spontaneous activity, is *drastically* varied, the differences between the models deteriorate. This is due to the reduction in the RM, which is already present in the basic models. There is, however, a major difference in the robustness to such variations between the sparse and the cascaded models: The cascaded model is much less sensitive to local variations of these parameters, because these tend to cancel in the cascaded cortical network. Only a global variation over the whole cascaded network has a similar effect as for the sparse models. The peak at low relative modulations r_1 corresponds to simple cells that have not found an antagonistic partner that matches closely enough their own receptive field properties (corresponding to a pure push model). Accordingly, in the sparse models more simple cells (about 24% of the total for both types) have a relative modulation $r_1 \leq 1.0$ than in the cascaded model (about 8% of the total). Similarly, the broadening of the second peak in the strict versus the weak case corresponds to several target cells having found only a single matching partner instead of the maximum number of 2. In the area where the push-pull mechanism is fully effective ($r_1 > 1.0$), there is little difference between the models. The cascaded model can be implemented, on the computer or biologically, with much less rigor than its sparse companions. It is less sensitive to the specific properties of individual cells and more robust to changes in the pre-synaptic network. In addition, the distribution of the relative modulation values looks realistic. The degree of linearity obtained with cascaded inhibition nicely reflects

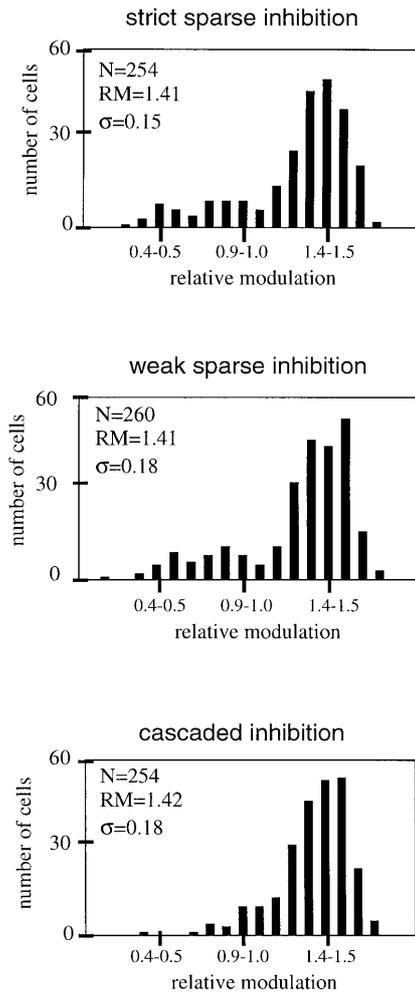


Fig. 12 Distribution of relative modulation r_1 in the detailed models. Using a bin size of 0.1, the distribution of relative modulation is shown for the strict sparse inhibitory model, the weak sparse inhibitory model, and the cascaded intracortical inhibition scheme for stimulation with a grating of optimal spatial period. Distributions in the *top two* cases are bimodal, whereas in the cascaded model we observe a single peak with a maximum around $r_1=1.57$. Indicated are the number of cells in the distribution N , the mean RM, calculated from all cells with $r_1 \geq 1.0$, and the corresponding standard deviation of the mean σ . Because of the skewness of the distribution, the mean RM is shifted toward lower r_1 -values when compared with the ideal value 1.57

the “linear” subpopulation of simple cells, which we found to be 38% of the total. It is, however, much easier to obtain nonlinear behavior by means of the same cascaded connection pattern just by using short, imbalanced cascades that lead to imbalanced excitation and inhibition.

It remains to be investigated how far the complicated receptive field structure introduced by the cascaded model could explain other properties of simple cells. This applies in particular to temporal effects not discussed here, e.g., phase differences leading to direction selectivity or facilitatory effects in other simple cells.

We now address the question of how sensitive the simple cells in the detailed model are to changes of the spatial frequency of the stimulus, i.e., we discuss the temporal

phase dependence of RM as in the previous section for the basic push-pull models and for real cells in the experimental part.

In order to determine the phase dependence in the detailed models, we simulated the behavior of the sparse and cascaded models for a temporal phase difference of 0° , 45° , 90° , and 135° in the response from the on and off subfields.

When we selected all cells with optimal orientation and varied the phase difference by using grating stimuli of different spatial period, we obtained the phase dependence of the RM shown in Fig. 11c. We averaged all RMs r_1 above 1.0 and plotted the mean against phase difference. Because of the skewness of the distributions, the mean was shifted toward lower RMs.

Both of the sparse models show a greater phase dependence of relative modulation than the cascaded model. This could be an indication that the phase-insensitive cells found in the experiment are connected according to the cascaded connection scheme, whereas a sparse connection scheme more readily results in a phase dependence of the RM. It also implies that quasi-linear behavior of simple cells over a relatively wide parameter range can be easily achieved in a cascaded setup.

Discussion

The goal of this study was threefold: (1) show the limitations of a widely used method (relative modulation); (2) provide evidence that specific push-pull models may be rather hard to distinguish experimentally; and (3) try to devise a parsimonious model for an antagonistically organized intracortical inhibition. In particular, our last goal was meant to contribute beyond visual cortical physiology to the discussion about general cortical design principles.

We first briefly discuss the basic push-pull models and the limits of the relative modulation measure before we try to argue in favor of the generic cascaded inhibition model, which might incorporate the most important aspects of cortical linearity.

The basic push-pull models

The relative modulation measure is relatively crude because it is based on a spatially and temporally extended stimulus. Hence, it is not the aim of this study to analyze the fine structure of the receptive field of a simple cell, for which much better methods exist such as the structural cascade models (Pollen and Ronner 1982; Adelson and Bergen 1985; Korenberg and Hunter 1986; Emerson and Citron 1988; Emerson et al. 1989) or the mapping techniques with or without reverse correlation between cells response and stimulus (DeAngelis et al. 1993; Jones and Palmer 1987; McLean and Palmer 1989; Baker and Cynader 1988). On a scale that takes into account the stimulus limitations, however, the spatial summation between subfields of a simple cell can already be studied in the basic

push-pull models that represent a very simple sketch of a neuron in the visual system. It turns out that the sensitivity of the relative modulation, being a measure of the linearity of spatial summation between the subfields, to changes of most parameters in the basic models is relatively weak. Various different model constellations may lead to the same relative modulation. The introduction of a half-squaring cell model does not qualitatively alter this situation. Ambiguity in the relative modulation remains, even though the squaring operation increases the modulation depth of this output and hence magnifies changes caused by different parameter settings. Thus, the studies of the *weighted* basic push-pull model (Fig. 4) show that the ratio of excitation to inhibition is the only parameter in the present models which – within physiological ranges – exerts a rather strong influence on the relative modulation. This relatively strong sensitivity to relative modulation to the balance of excitation and inhibition seems to be reflected in the experimental part of this study, where we find a large class of simple cells with a phase-sensitive relative modulation.

Experimental findings

In the theoretical part, we discussed several reasons that can lead to a curved relation between relative modulation and temporal phase. This effect occurs with nonzero spontaneous activity, but we have excluded such cells from the study. Furthermore, the experimentally observed bend sets in gradually starting from $\phi=0$, which is not similar to the bend of the simulated curves that was introduced by increasing the spontaneous activity. In this case a noticeable bend only occurs for large values of ϕ . In addition, firing thresholds unequal to zero lead to bended curves. To get a strong effect of this kind, the firing threshold, however, would have to be unphysiologically high. Another problem is posed by the question of whether a deviation from the true half-wave rectifier model by exponents unequal to 1.0 (e.g., a half-squarer has the exponent 2) could account for the bend in the curves. While higher exponents have been proposed by others (Albrecht and Geisler 1991; Heeger 1992a, b, 1993), in our hands there is no conclusive evidence in favor or against them. In case of higher exponents, we would expect larger RM values and a bend of the curves which occurs starting from these larger values, which we did not observe. A combination of these parameters could also account for the bend in the curves, but only if the exponent would be less or equal to one while spontaneous activity and threshold would have to deviate significantly from zero at the same time. While this would in principle be possible, the bend in the curves is more generically explained by an imbalance between excitation and inhibition. Palmer et al. (1991) discussed several models that include such an imbalance, one of which (their “model 3”) includes imbalanced lateral inhibition for which indirect support exists (Sillito 1975; Hata et al. 1988; Wörgötter and Eysel 1991).

In addition to such a later inhibition, an imbalance could also be introduced by a displacement between the excitatory and inhibitory subfields with respect to each other. In the detailed simulation (Figs. 10–12), we have proposed a cascaded intracortical inhibition scheme in which many (≈ 25) cells converge on and inhibit each other. This structure is easy to implement during development and excludes an exact balance between excitation and inhibition in almost all cases. Moreover, even completely imbalanced regions of pure excitation and/or pure inhibition can exist (see Fig. 10). Such effects most strongly contribute to the bent shape of the curves in the basic models (Fig. 4). The most robust behavior was obtained with the cascaded inhibition connection scheme, which is directly related to the “model 4” proposed by Palmer et al. (1991), in which displacements between the on and off (i.e., excitatory and inhibitory) regions exist. A similar scheme, which possibly accounts for cortical orientation selectivity by displaced receptive fields, was also proposed by Heggelund (1981).

The experimental and theoretical results discussed above make it seem likely that an imbalance between converging inputs could actually introduce the experimentally observed bend in the curves in Fig. 8b. Temporal effects, however, could add to this.

The imbalance between excitation and inhibition is spatially introduced in all the cases discussed so far. It is important to note, however, that temporal effects could also result in an *effective* imbalance between excitation and inhibition at any one point in time. Temporal delays between the different subfields could lead to a phase difference and, this, to a drop of the relative modulation at nonoptimal spatial frequencies. While we tried to avoid some of the possible temporally induced effects by removing all direction-selective cells from the study, this possibility cannot entirely be ruled out with the present set of experiments.

Regarding the high complexity of the cortical network and also the variability of the cell behavior, one would not necessarily expect that simple cells should have an exactly balanced excitatory and inhibitory input and, thus, a flat relation between relative modulation and temporal phase. Nevertheless, we found about 38% of the simple cells in Fig. 8a, b that did not show any significant change of the relative modulation, when we varied the spatial frequency of the grating. The mean relative modulation (1.49) for these cells was close to 1.57. Therefore, this sample seems to reflect rather ideal simple cells, with a behavior that closely represents quasi-linear summation. Judging from the recording depth, we found these cells almost exclusively in the middle of layer IV. This cell group probably corresponds to the “linear simple cells” described by Movshon et al. (1978a). The range and the mean of the relative modulations shown in Fig. 8a corresponds to that found by others (Dean and Tolhurst 1983; for a review see Skottun et al. 1991). After the previous discussion, the question must arise of how far these quasi-linear cells might principally differ from the cells that are more sensitive to the changes in the stimulus, if one

wants to avoid the awkward assumption that excitation and inhibition have been well balanced by chance over the whole set of stimulus conditions and recording time.

Biophysically realistic models

From the sections above, it becomes clear that basic push-pull models are designed in a specific way, which might not correspond to the actual connectivity pattern found in the visual cortex. In addition, some of the fine distinctions between the different flavors of basic push-pull models could be difficult to assess experimentally. In general, however, all these models are very sensitive with respect to an imbalance between their excitatory and inhibitory inputs. From these observations the question arose of whether there is a generic way to design a cortical network such that it will produce an (almost) quasilinear behavior, while exceptions from linearity will also be explained within this wiring framework.

Massive inhibitory intracortical connections as in the cascaded inhibition scheme easily leads in a robust way to some degree of linear spatial summation in simple cells. In addition, the variability of the cells' linearity can also be accounted for by this connection scheme, because many times an incomplete balance between excitation and inhibition arises as the consequence of this wiring pattern. One should in particular note that the effect of cascaded inhibition is more than a mere effect of increased inhibition. This can be seen firstly by increasing the level of inhibition in the sparse connection models, which does not lead to the same result as in the cascaded model. Secondly, in the cascaded model the relative modulation of a simple cell's output is less sensitive to changes in the phase difference of the individual responses from the receptive subfields. It is the range of spatial parameters (position and orientation) and temporal parameters (delays) of the whole network, that aids the simple cell in the cascaded model to respond more uniformly (as measured with relative modulation) to a variety of spatial frequencies of the stimulus. Size and orientation of subfields of real nerve cells vary slightly even for subfields of the same simple cell. Such a variability has larger consequences for the detailed sparse models than the cascaded model, because the latter involves an averaging process over a larger number of cortical cells, leading to a greater robustness with respect to variations in cell parameters and also in stimulus properties. In particular, this extends to the ratio of excitation to inhibition, because the local inhibitory feedback network tends to normalize the cell response to a certain extent. Additional properties such as the presence of disinhibitory secondary order effects in the cascaded network are predictions from the cascaded push-pull model, which might eventually also be tested experimentally. A recent study of Ferster et al. (1996) questions the role of the intracortical network in generating the orientation tuning of cortical simple cells. Even though it is not the aim of the cascaded inhibition model to explain orientation tuning in simple cells, a cer-

tain amount of "pull action," i.e., inhibition from other cortical simple cells with spatially opponent receptive field structure, is necessary in the cascaded inhibition model to overcome the flaws of a pure push model. If this condition is fulfilled, the cascaded setup presents a generic way to account for quasilinear as well as nonlinear behavior within a single model framework.

Analytically tractable models are elegant and lead to a more intuitive understanding of the behavior of a system. One of them, the fundamental push-pull model is largely accepted for the explanation of many aspects of cortical simple cell behavior. The specific design of this model, however, represents a tremendous oversimplification of real cortical wiring. Recent models of cortical direction and orientation selectivity make use of the intrinsic properties of a highly interconnected network such as the visual cortex and the cell specificities arise in these models as an "emergent" network property (Douglas et al. 1995; Somers et al. 1995). In light of this changing viewpoint, the approach of shaping the cells' linearity through the action of the intracortical network (as by means of cascaded inhibition) seems preferable to an approach of solely designing increasingly sophisticated models of single cells to explain the experimental data.

Acknowledgements We thank Dr. Klaus Funke for his valuable comments on the manuscript. We gratefully acknowledge the support of DFG grant Wo 388/4, and of the exchange program of the Max Planck Gesellschaft with the Academia Sinica (F.W. and B.L.).

References

- Adelson EH, Bergen JR (1985) Spatiotemporal energy models for the perception of motion. *J Opt Soc Am A* 2:284–299
- Albrecht DG, Geisler WS (1991) Motion sensitivity and the contrast response function of simple cells in the visual cortex. *Vis Neurosci* 7:531–546
- Andrews BW, Pollen DA (1979) Relationship between spatial frequency selectivity and receptive field profile of simple cells. *J Physiol (Lond)* 287:163–176
- Baker CL, Cynader MS (1988) Space-time separability of direction selectivity in cat striate cortex neurons. *Vision Res* 2:239–246
- Brette D, Niebur E (1994) A detailed parallel simulation of a biological neural network. *IEEE Comp Sci Eng* 1:31–43
- DeAngelis GC, Ohzawa I, Freeman RD (1993) Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. II. Linearity of temporal and spatial summation. *J Neurophysiol* 69:1118–1135
- Dean AF, Tolhurst DJ (1983) On the distinctness of simple and complex cells in the visual cortex of the cat. *J Physiol (Lond)* 344:305–325
- DeValois K, Tootell R (1983) Spatial-frequency-specific inhibition in cat striate cortex cells. *J Physiol* 336:359–376
- Douglas RJ, Koch C, Mahowald M, Martin KAC, Suarez HH (1995) Recurrent excitation in neocortical circuits. *Science* 269:981–985
- Emerson RC, Citron MC (1988) How linear and nonlinear mechanisms contribute to directional selectivity in simple cells of cat striate cortex (abstract). *Ophthalmol Vis Sci* 29:23
- Emerson RC, Korenberg J, Citron MC (1989) Identification of intensive non-linearities in cascade models of visual cortex and its relation to cell classification. In: Marmarelis V (ed) *Advanced methods of physiological system modelling*. Plenum Press, New York, pp 97–111

- Enroth-Cugell C, Robson JG (1966) The contrast sensitivity of retinal ganglion cells of the cat. *J Physiol* 247:551–578
- Enroth-Cugell C, Robson JG, Schweitzer-Tong DE, Watson AB (1983) Spatio-temporal interactions in cat retinal ganglion cells showing linear spatial summation. *J Physiol* 341:279–307
- Ferster D (1988) Spatially opponent excitation and inhibition in simple cells of the cat visual cortex. *J Neurosci* 8:1172–1180
- Ferster D, Sooyoung C, Wheat H (1996) Orientation selectivity of thalamic input to simple cells of cat visual cortex. *Nature* 380:249–252
- Glezer VD, Tsherbach TA, Gauselman VE, Bondarko VM (1980) Linear and non-linear properties of simple and complex receptive fields in area 17 of the cat visual cortex. *Biol Cybern* 37:195–208
- Hata Y, Tsumoto T, Sato H, Hagihara K, Tamura H (1988) Inhibition contributes to orientation selectivity in visual cortex of cat. *Nature* 335:815–817
- Heeger DJ (1992a) Normalization of cell responses in cat striate cortex. *Vis Neurosci* 9:181–197
- Heeger DJ (1992b) Half-squaring in responses in cat striate cortex. *Vis Neurosci* 9:427–443
- Heeger DJ (1993) Modeling simple-cell direction selectivity with normalized, half-squared, linear operators. *J Neurophysiol* 70:1885–1898
- Heggelund P (1981) Receptive field organization of simple cells in cat striate cortex. *Exp Brain Res* 42:89–98
- Hochstein S, Shapley RM (1976) Linear and nonlinear spatial subunits in Y cat retinal ganglion cells. *J Physiol* 262:265–284
- Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol* 160:106–156
- Ikeda H, Wright MJ (1975) Spatial and temporal properties of "sustained" and "transient" neurones in area 17 of the cat's visual cortex. *Exp Brain Res* 22:363–383
- Jacobson LD, Gaska JP, Chen H-W, Pollen DA (1993) Structural testing of multi-input linear-nonlinear cascade models for cells in macaque striate cortex. *Vision Res* 33:609–626
- Jones JP, Palmer LA (1987) The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *J Neurophysiol* 58:1187–1211
- Kaplan E, Shapley RM (1982) X and Y cells in the lateral geniculate nucleus of macaque monkeys. *J Physiol* 330:125–143
- Korenberg MJ, Hunter IW (1986) The identification of nonlinear biological systems: LNL cascade models. *Biol Cybern* 55:125–134
- Liu Z, Gaska JP, Jacobson LD, Pollen DA (1992) Interneuronal interaction between members of quadrature phase and anti-phase pairs in the cat's visual cortex. *Vision Res* 32:1193–1198
- McLean J, Palmer L (1989) Contribution of linear spatiotemporal receptive field structure to velocity selectivity of simple cells in area 17 of cat. *Vision Res* 29:675–679
- Movshon JA, Thompson ID, Tolhurst DJ (1978a) Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *J Physiol* 283:53–77
- Movshon JA, Thompson ID, Tolhurst DJ (1978b) Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat's visual cortex. *J Physiol* 283:101–120
- Orban GA (1984) Neuronal operations in the visual cortex. In: Barlow HB, Bullock TH, Florey E, Grüsser OJ, Peters A (eds) *Studies of brain function*, vol 11. Springer, Berlin
- Palmer LA, Davis TL (1981) Receptive-field structure in cat striate cortex. *J Neurophysiol* 46:260–276
- Palmer LA, Jones JP, Stepnowski RA (1991) Striate receptive fields as linear filters: characterization in two dimensions of space. In: Cronly-Dillon J, Leventhal AG (eds) *The neural basis of visual function. (Vision and visual dysfunction, vol 4)* Macmillan, London
- Pollen DA, Ronner SF (1982) Spatial computation performed by simple and complex cells in the visual cortex of the cat. *Vision Res* 22:101–118
- Pollen DA, Andrews BW, Feldon SE (1978) Spatial frequency selectivity of periodic complex cells in the visual cortex of the cat. *Vision Res* 18:665–682
- Reid RC, Soodak RE, Shapley RM (1991) Directional selectivity and spatiotemporal structure of receptive fields of simple cells in cat visual cortex. *J Neurophysiol* 66:505–529
- Shapley RM, Hochstein S (1975) Visual spatial summation in two classes of geniculate cells. *Nature* 256:411–413
- Sillito AM (1975) The contribution of inhibitory mechanisms to the receptive field properties of neurones in the striate cortex of the cat. *J Physiol* 250:305–329
- Skottun BC, De Valois RL, Grosf DH, Movshon JA, Albrecht DG, Bonds AB (1991) Classifying simple and complex cells on the basis of response modulation. *Vision Res* 31:1079–1086
- Somers DC, Nelson SB, Sur M (1995) An emergent model of orientation selectivity in cat visual cortex. *J Neurosci* 15:5448–5465
- Tolhurst DJ, Dean AF (1987) Spatial summation by simple cells in the striate cortex of the cat. *Exp Brain Res* 66:607–620
- Tolhurst DJ, Dean AF (1990) The effects of contrast on the linearity of spatial summation of simple cells in the cat's striate cortex. *Exp Brain Res* 79:582–588
- Wörgötter F, Eysel UT (1991) Topographical aspects of intracortical excitation and inhibition contributing to orientation specificity in area 17 of the cat visual cortex. *Eur J Neurosci* 3:1232–1244
- Wörgötter F, Koch C (1991) A detailed model of the primary visual pathway in the cat: comparison of afferent excitatory and intracortical inhibitory connection schemes for orientation selectivity. *J Neurosci* 11:1959–1979