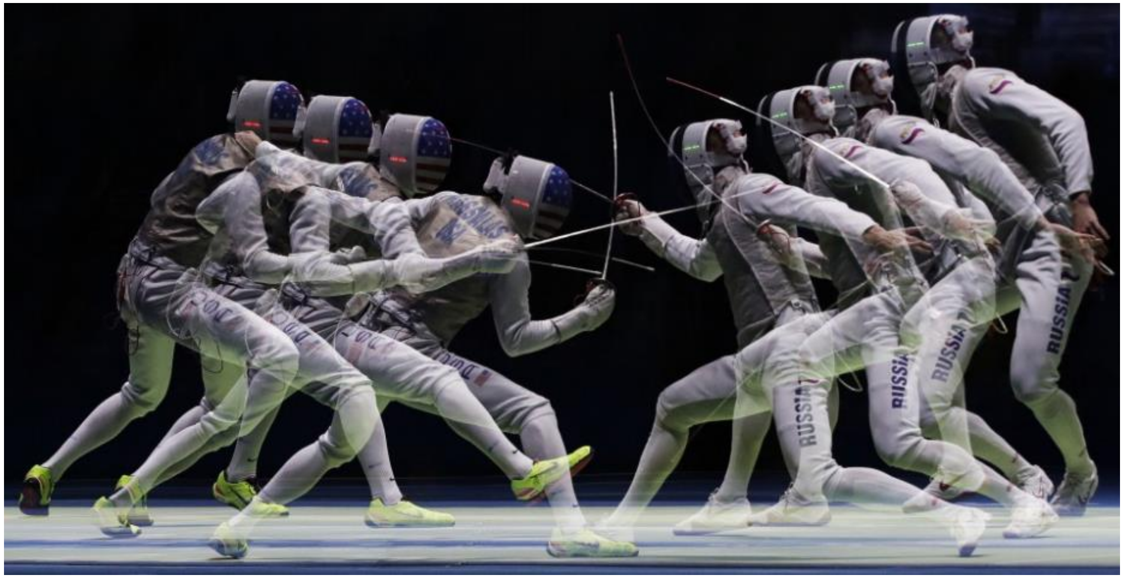
MOTION PERCEPTION. III

Modelling

Computational Strategies

1. **Feature Tracking**
2. **Spatiotemporal Correlation (Reichardt)**
3. **Motion Energy (Adelson and Bergen)**
4. Sign of brightness gradients (Marr and Ullman)
5. Ratio of brightness gradients (Johnston et al)
6. Flow (Johnston et al)

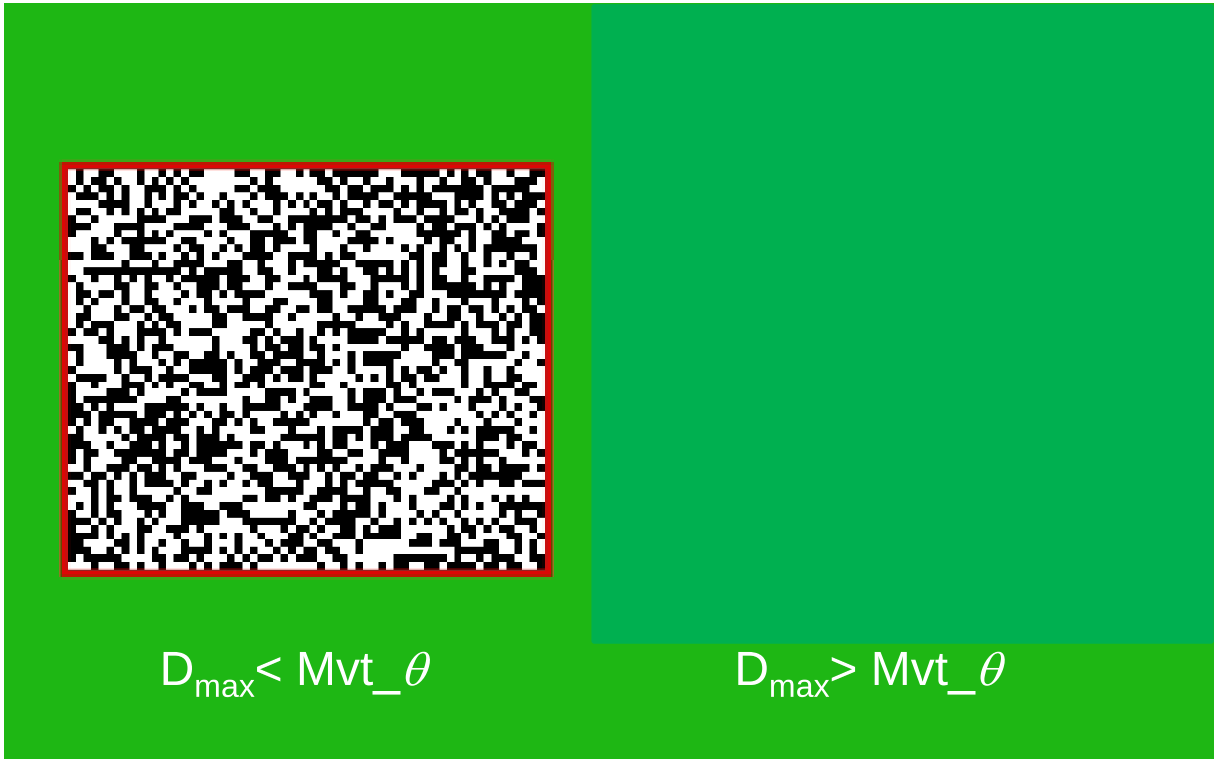
Feature tracking



Motion as a sequence
of frames (2 images/s)

Motion “per se”
(24 images/s)

Feature tracking



$D_{max} < Mvt\_\theta$ $D_{max} > Mvt\_\theta$

D_{max}: Braddick (1974)

Feature tracking

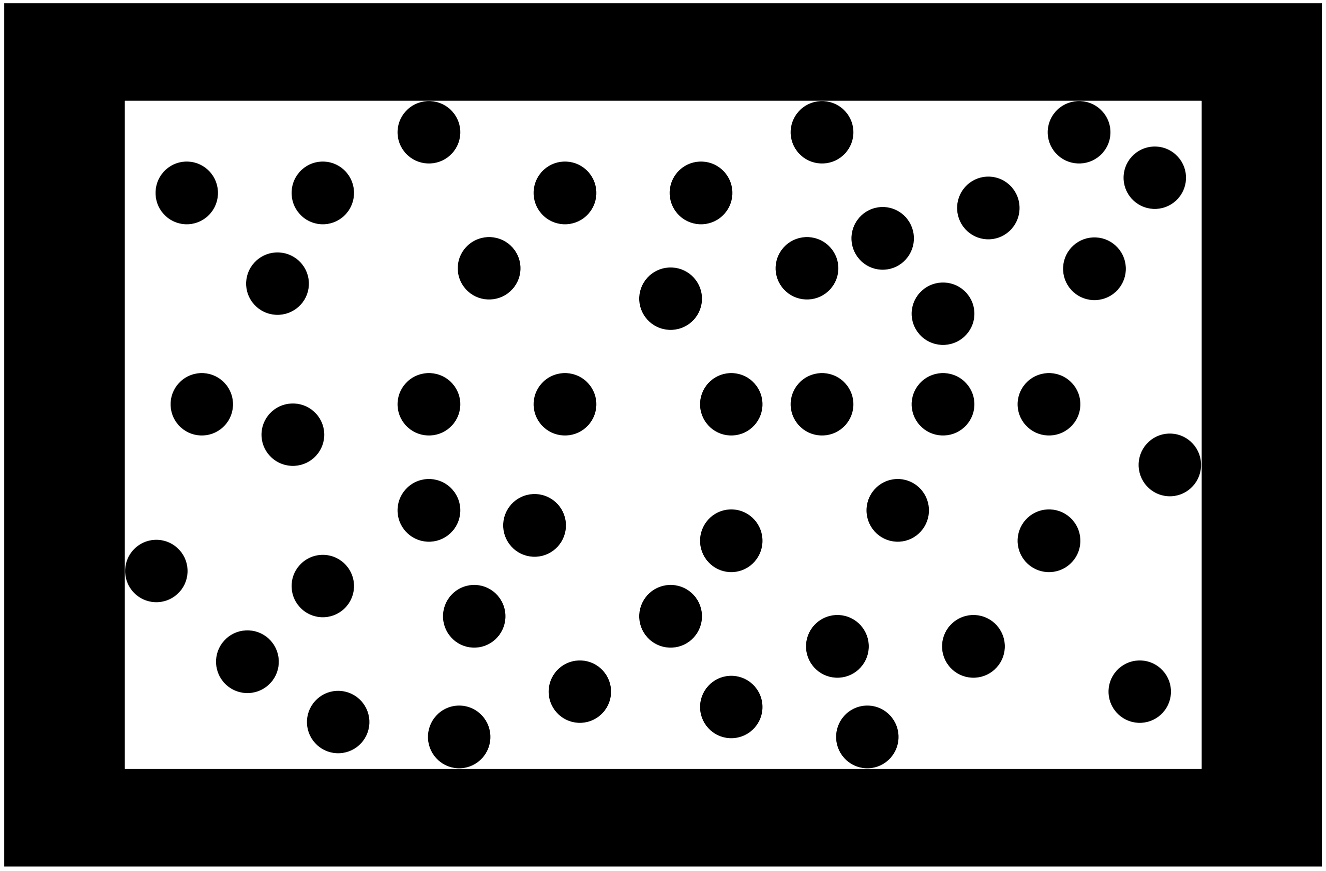
Advantages / Disadvantages

- Intuitive but..

- Problems
  - You measure position not motion
  - You only get the position of the features you track
  - You have to match corresponding features across frames
  - You have to convert spatial displacement into image speed

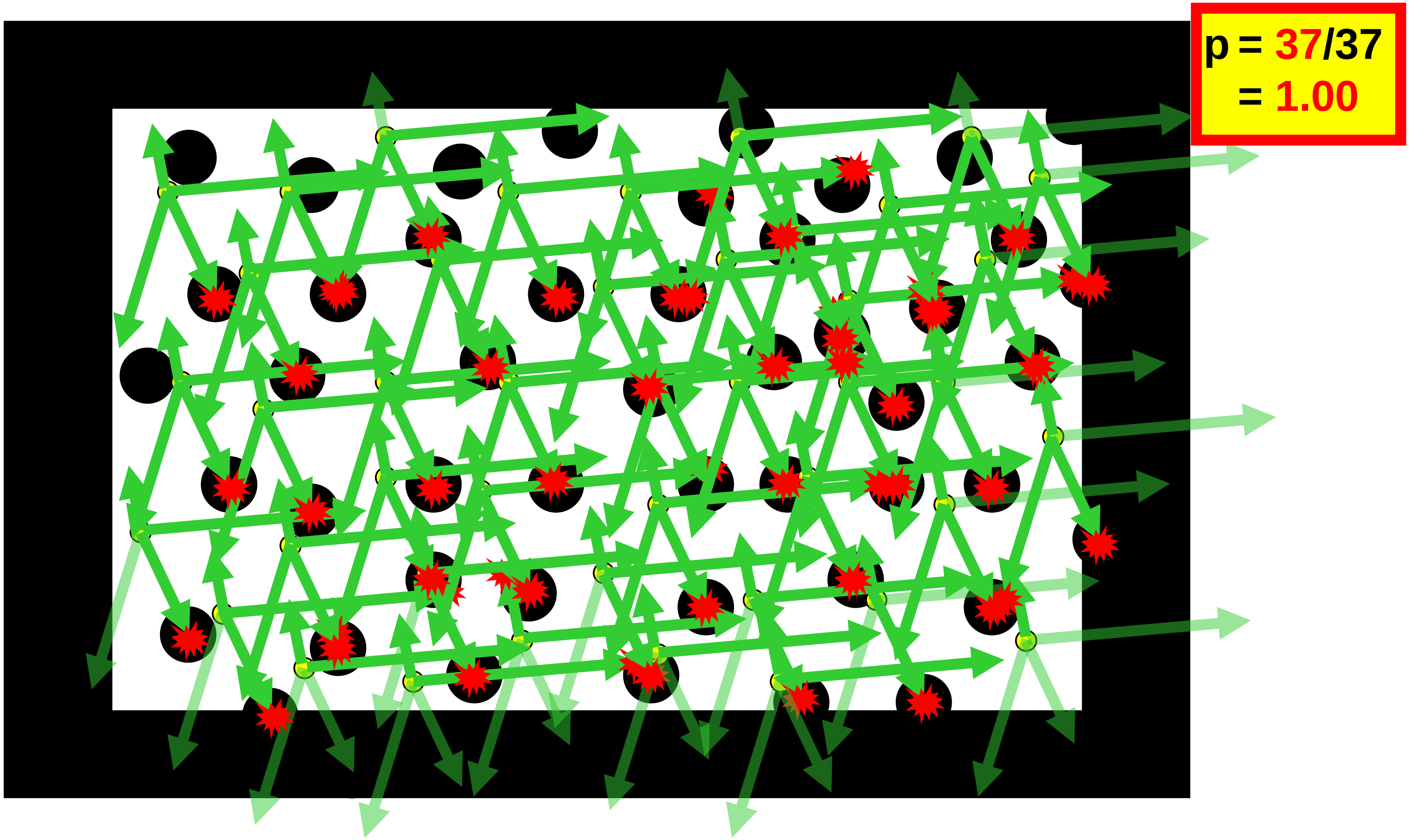
MOTION AS CORRELATION

Spatio-temporal correlation

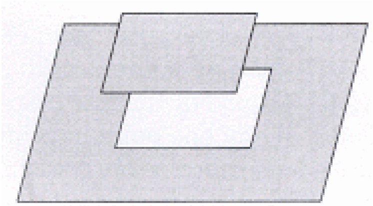


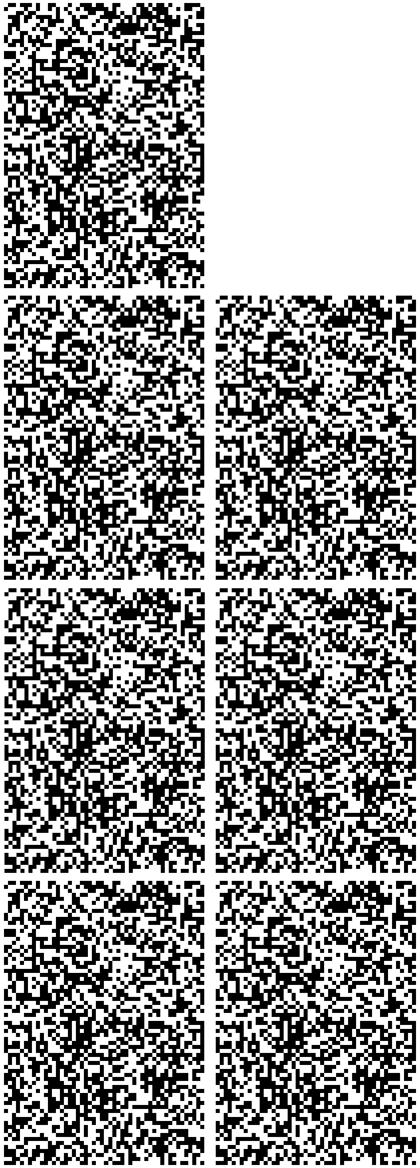
N = 43

Spatio-temporal correlation



Cross-correlation in Stereopsis (Julesz, 1961)





I. Create a random dot image.

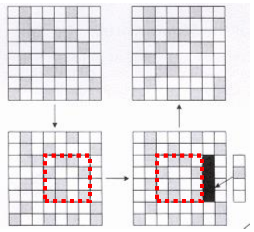
II. Copy image side by side.

III. Select a region of one image.

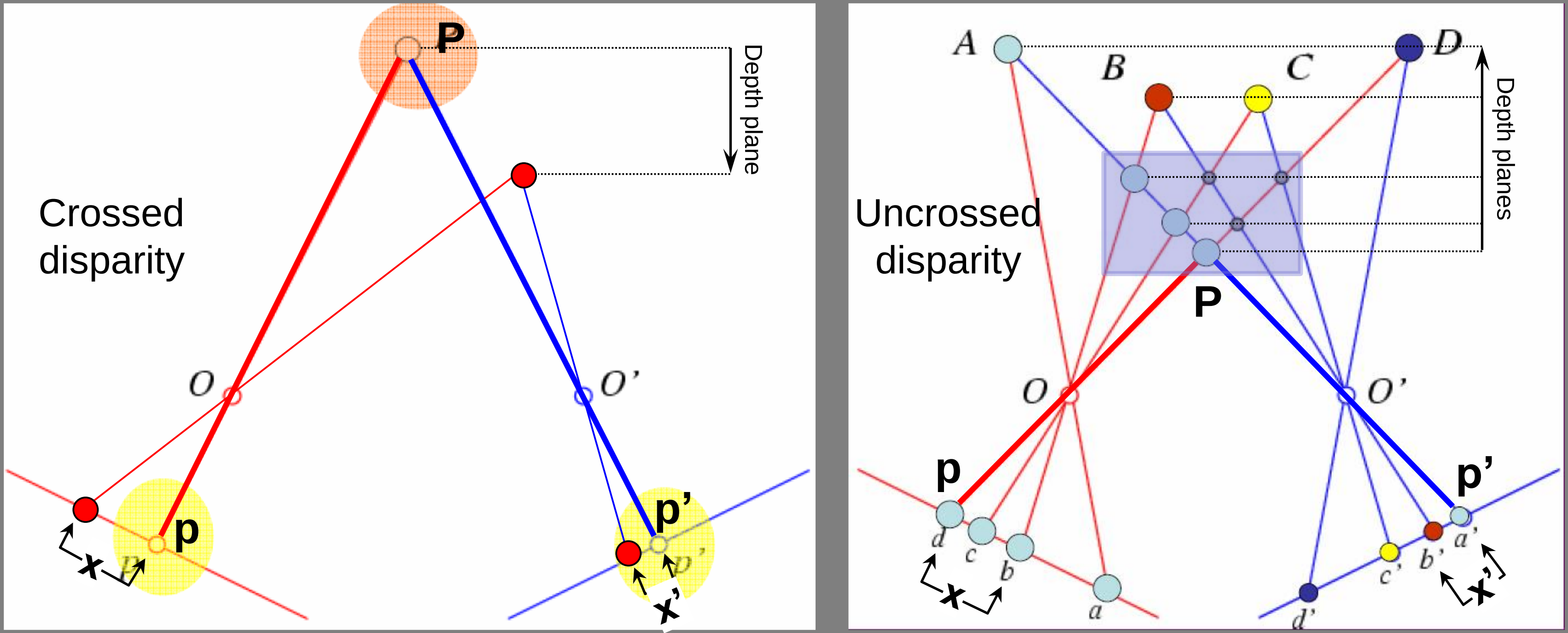
IV. Shift (horizontally) this region
and fill in the blank space left
behind with the random dots
to be replaced ahead.

The Random Dot Stereogram is ready.

To “reveal” the “hidden” square the brain presumably computes the cross-correlation between the 2 images.



Binocular disparity



Binocular disparity ➔ x – x' [deg]

Figure 1. The binocular fusion problem: in the simple case of the diagram shown on the left, there is no ambiguity and stereo reconstruction is a simple matter. In the more usual case shown on the right, any of the four points in the left picture may, a priori, match any of the four points in the right one. Only four of these correspondences are correct, the other ones yielding the incorrect reconstructions are shown in the purple shaded square.

Cross-correlation in Stereopsis

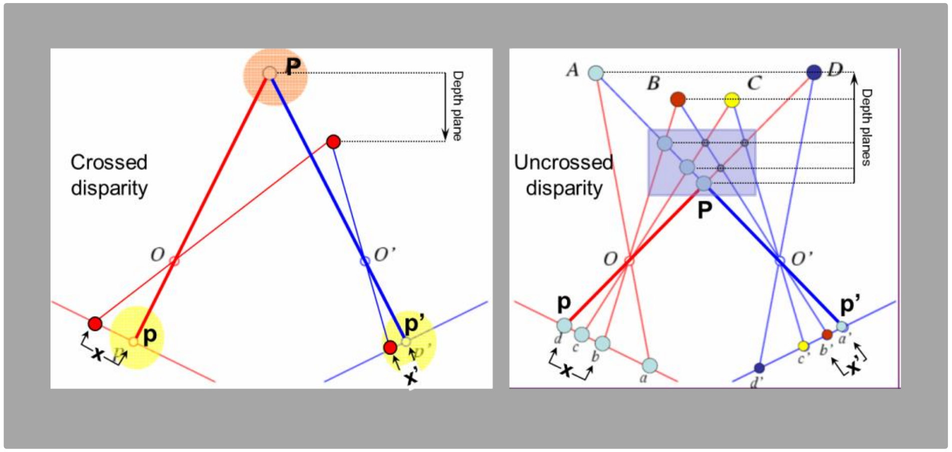


Figure 1. The binocular fusion problem: in the simple case of the diagram shown on the left, there is no ambiguity and stereo reconstruction is a simple matter. In the more usual case shown on the right, any of the four points in the left picture may, a priori, match any of the four points in the right one. Only four of these correspondences are correct, the other ones yielding the incorrect reconstructions shown as small grey discs

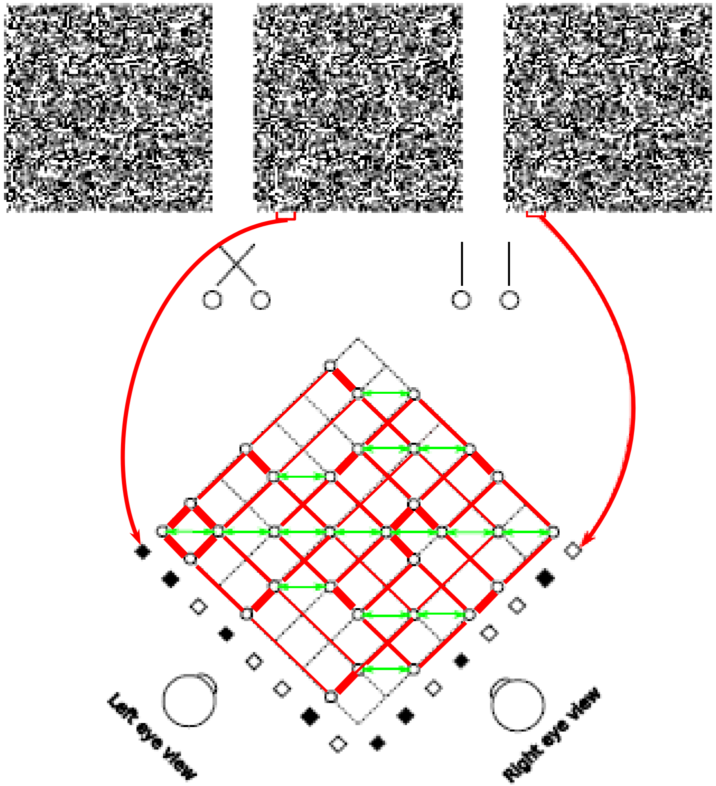


Figure 2. Eliminating 'false matches' in the stereo correspondence problem. A random dot stereogram at the top shows left and right eyes' images for crossed or uncrossed fusion (pair on the left or right respectively). Marr and Poggio's [10] proposal for establishing correct correspondences between dots in the two eyes' images is illustrated below, using only the dots highlighted in red (and dots from the same region of the left eye's image). The algorithm requires matches to be made between dots of the same colour, which gives rise to possible correspondences at all the nodes in the network marked by an open circle. Neighbouring matches with the same disparity support one another in the network, illustrated schematically by the green arrows (in their paper, the support extended farther). At the same time, matches along any line of sight (dotted lines) inhibit each other (red lines with thickness proportional to inhibition strength) since a ray reaching the eye must have come from only one surface. These constraints are sufficient to eliminate all but the correct matches, shown here along the main diagonal.

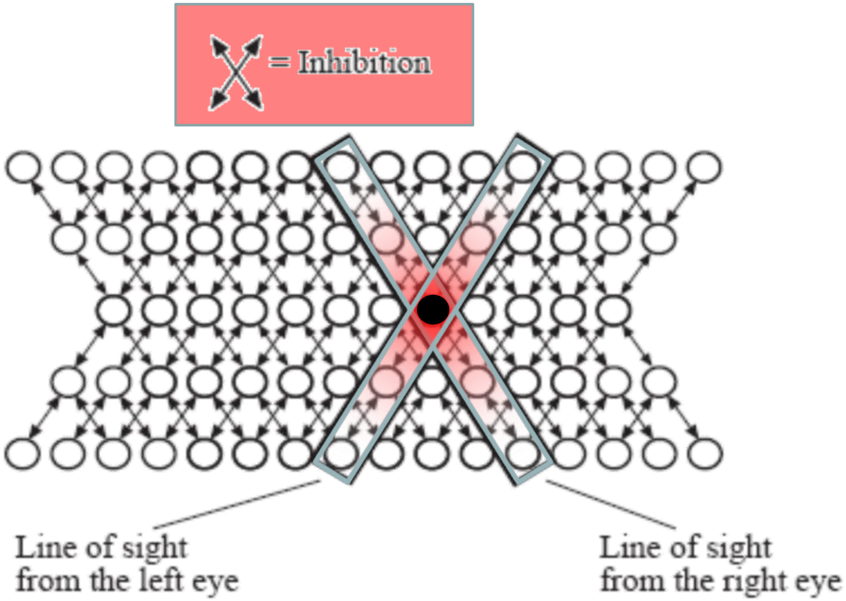


Figure 3. *Horizontal cross-section* of a disparity space. The **constraint of uniqueness** is implemented by letting all cells, along the two lines of sight, inhibit each other.

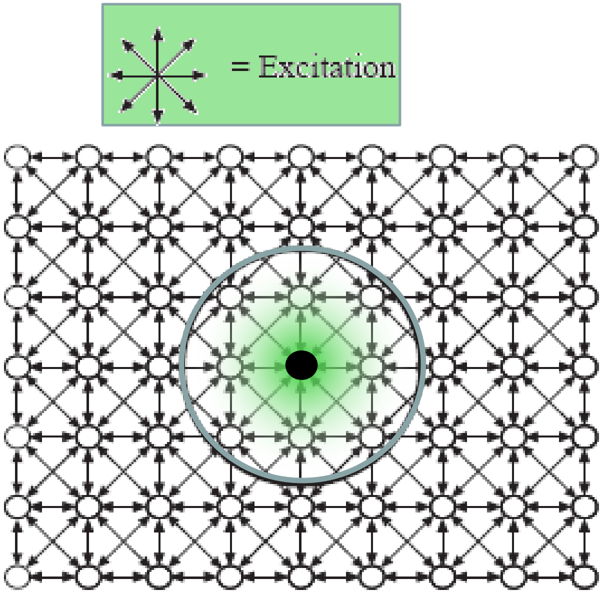
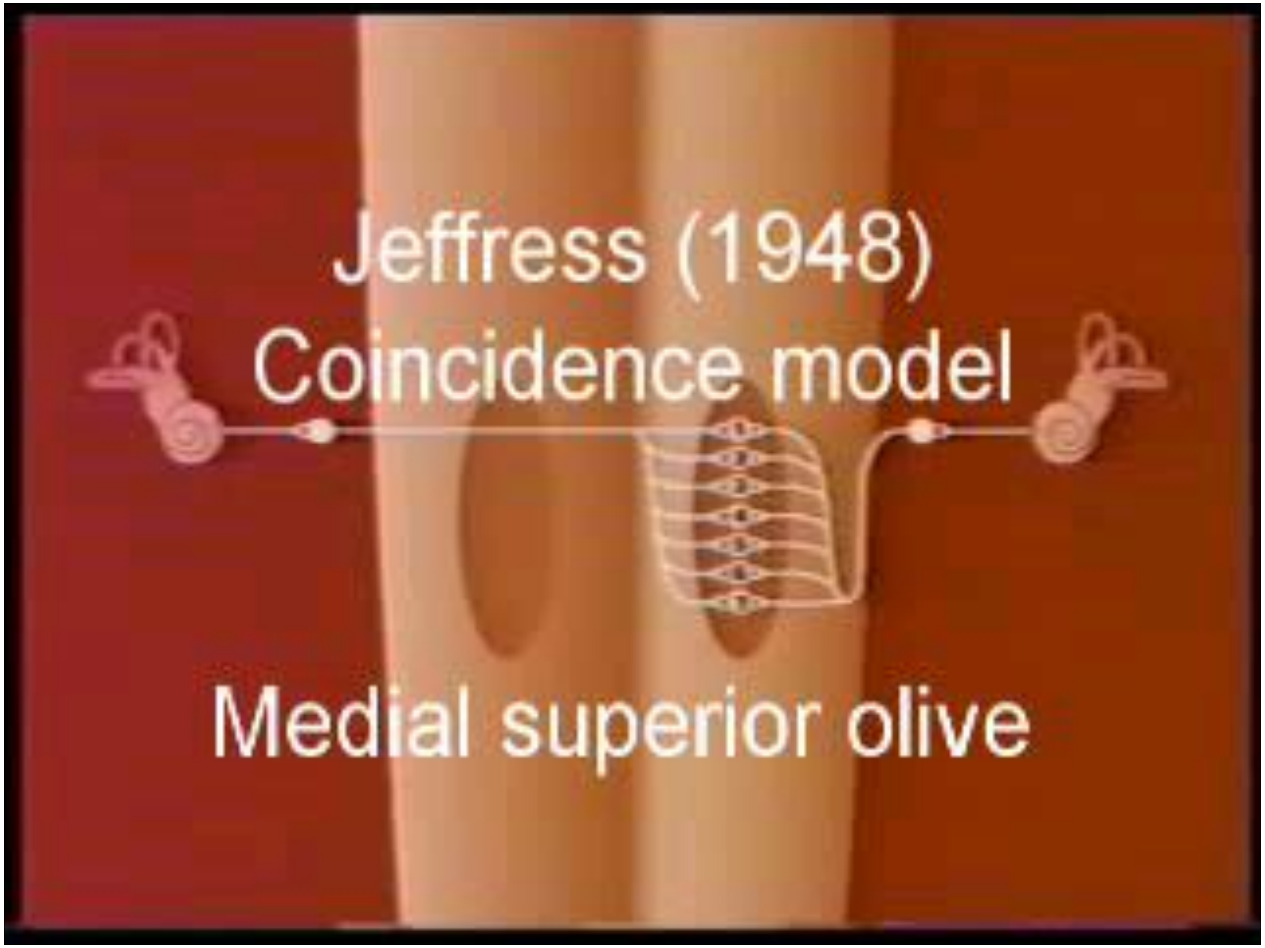


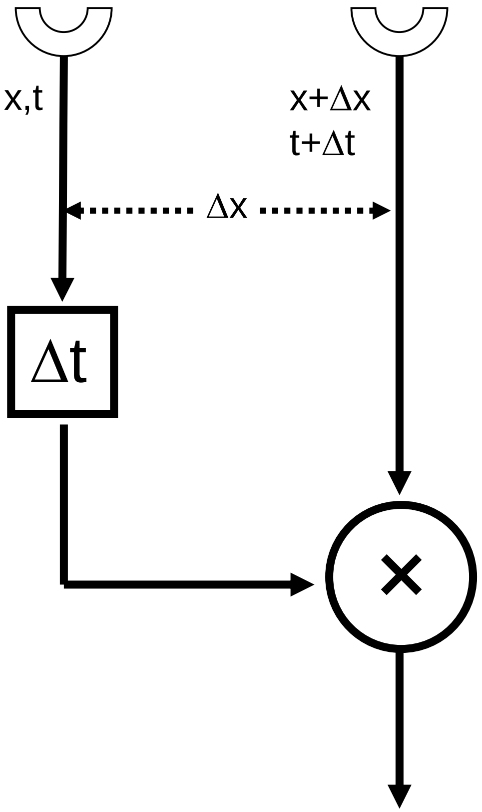
Figure 4. *Vertical cross-section* of a disparity-space. The **constraint of continuity** is implemented by letting all active cells excite the cells, in neighboring columns, that represent similar binocular disparity.

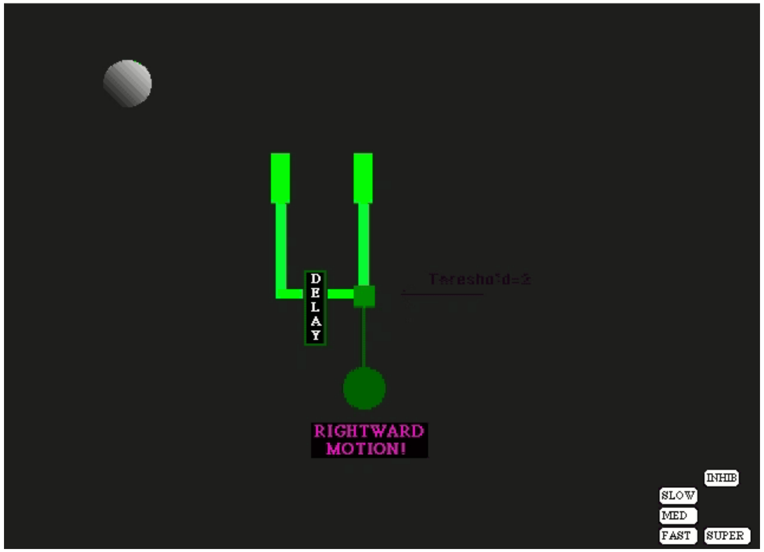
The coincidence model (Jeffress, 1948)

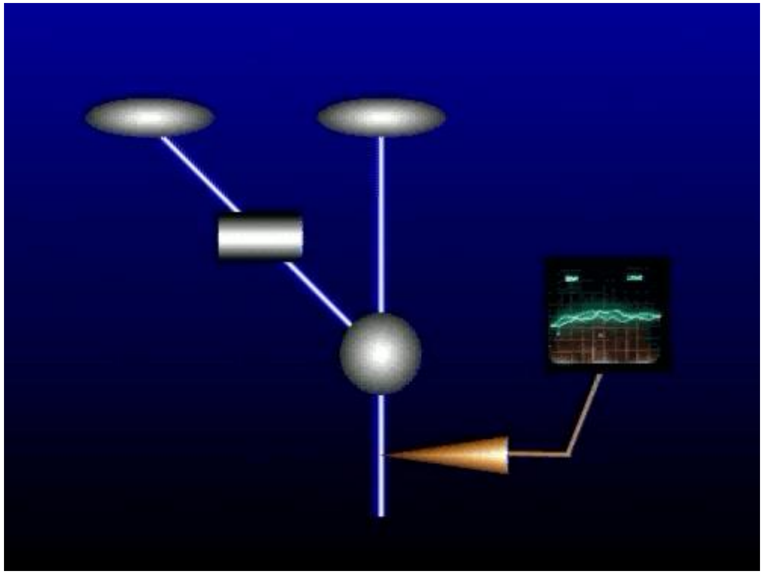


Neurocomputational model (derived from Boring’s “neural place-theory”) explaining how the auditory system might register and analyze small differences in the arrival time of sounds at the two ears in order to estimate the direction of sound sources in the azimuthal plane. Jeffress L. A. (1948). A place theory of sound localization," *J Comp Physiol Psychol 41*, 35-39

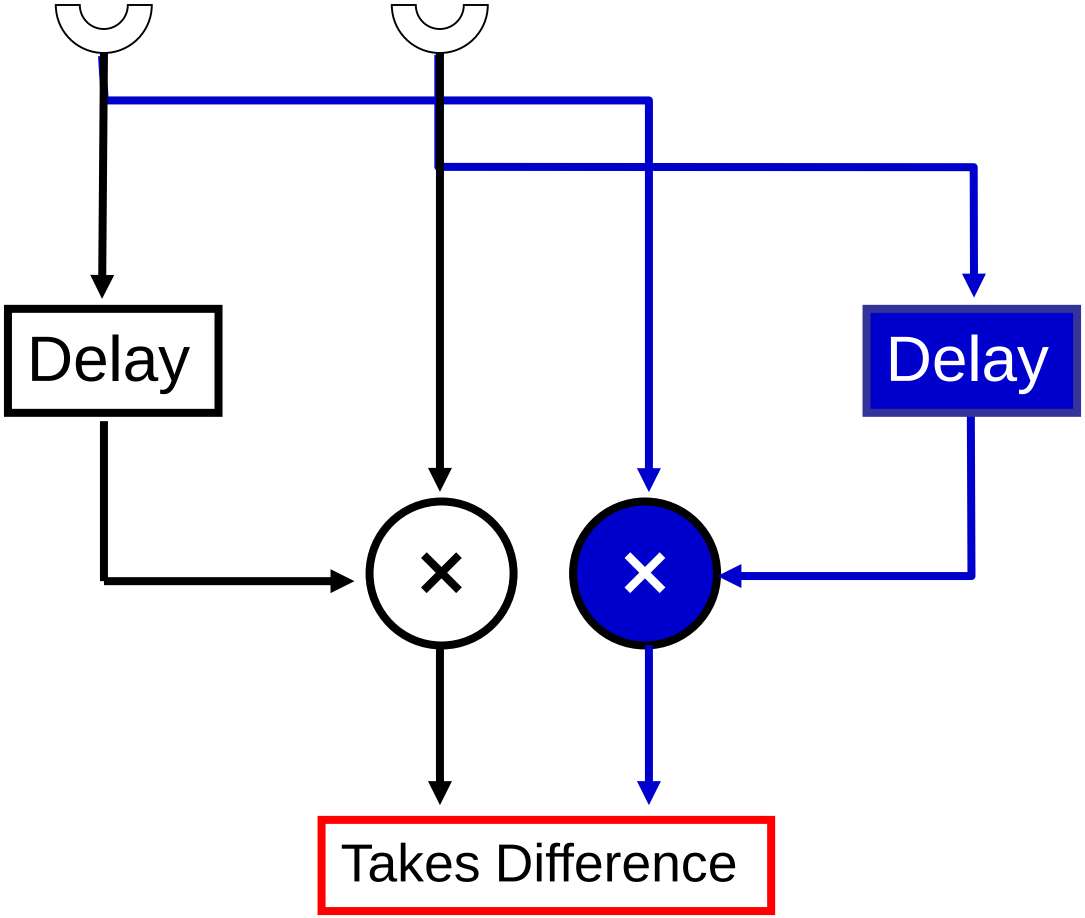
Spatio-temporal correlation (Reichardt, 1961)







Spatio-temporal correlation
+
Motion opponency



- Opponency enhances direction specificity
- Pairing “opposite” detectors can eliminate spurious activity

Spatio-temporal correlation

Elaborated Reichardt detector (Van Santen & Sperling, 1985)

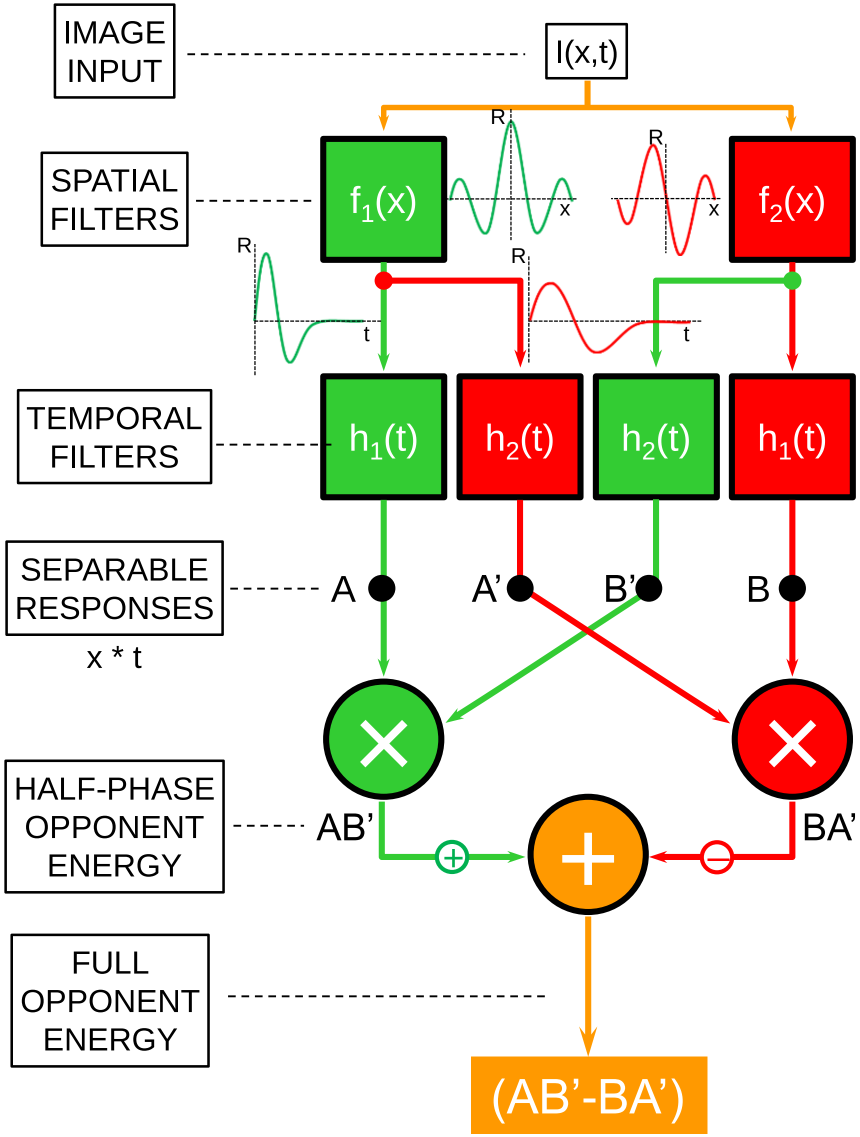


IMAGE INPUT

SPATIAL FILTERS

TEMPORAL FILTERS

SEPARABLE RESPONSES

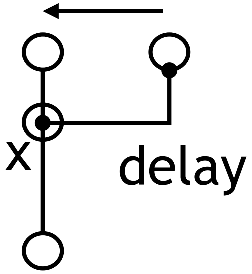
x * t

HALF-PHASE OPPONENT ENERGY

FULL OPPONENT ENERGY

Spatio-temporal correlation

Advantages / Disadvantages

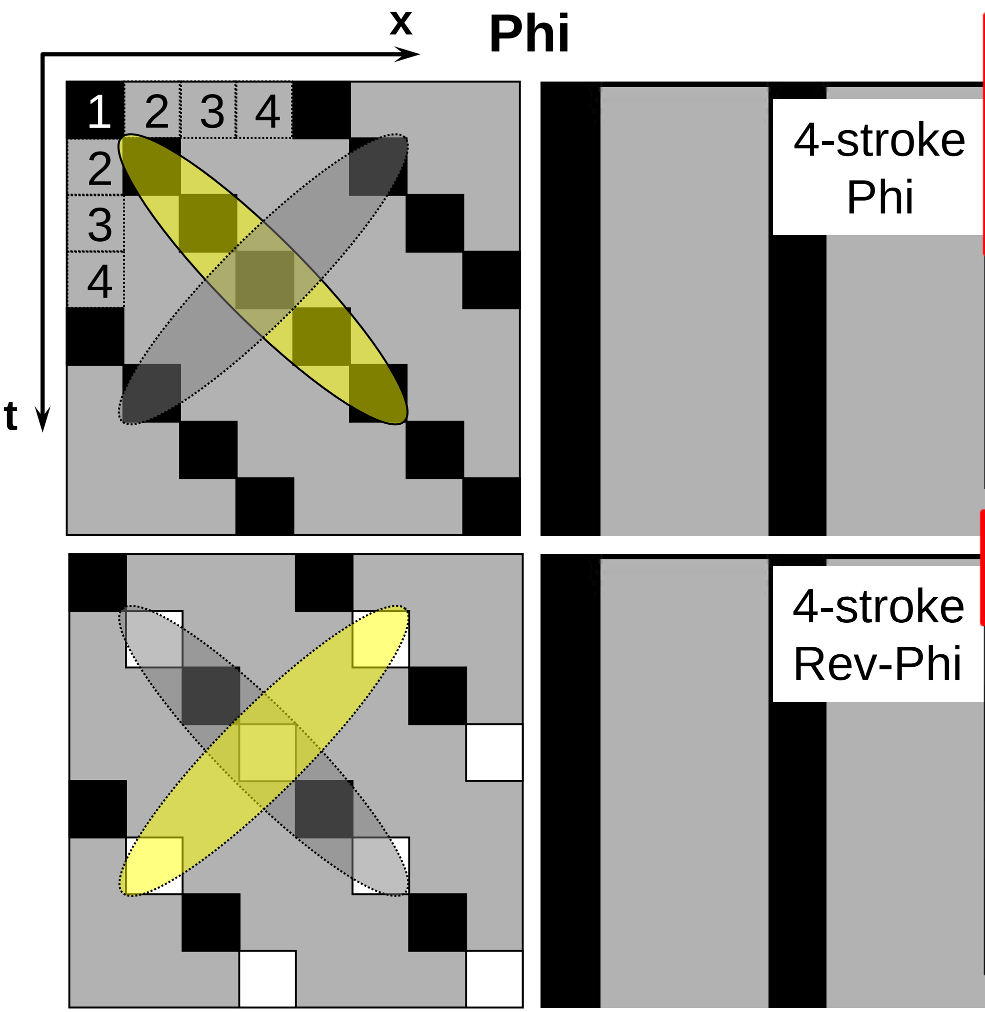


- Advantage – intuitive but..

- Disadvantages
  - Tuned to temporal frequency (rather than speed)
  - Not immediate – you have to wait for the traverse
  - Subject to aliasing/correspondence problems
  - Phase dependent response

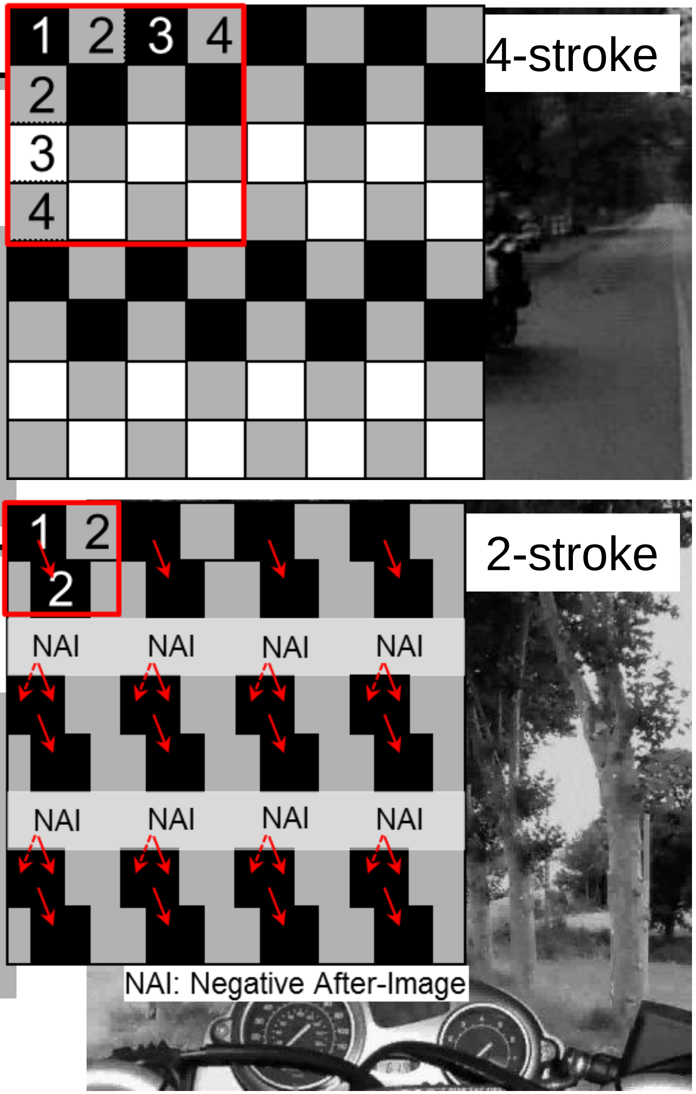
MOTION AS SPACE-TIME ORIENTATION

Motion as Orientation in space-time



Reversed-Phi

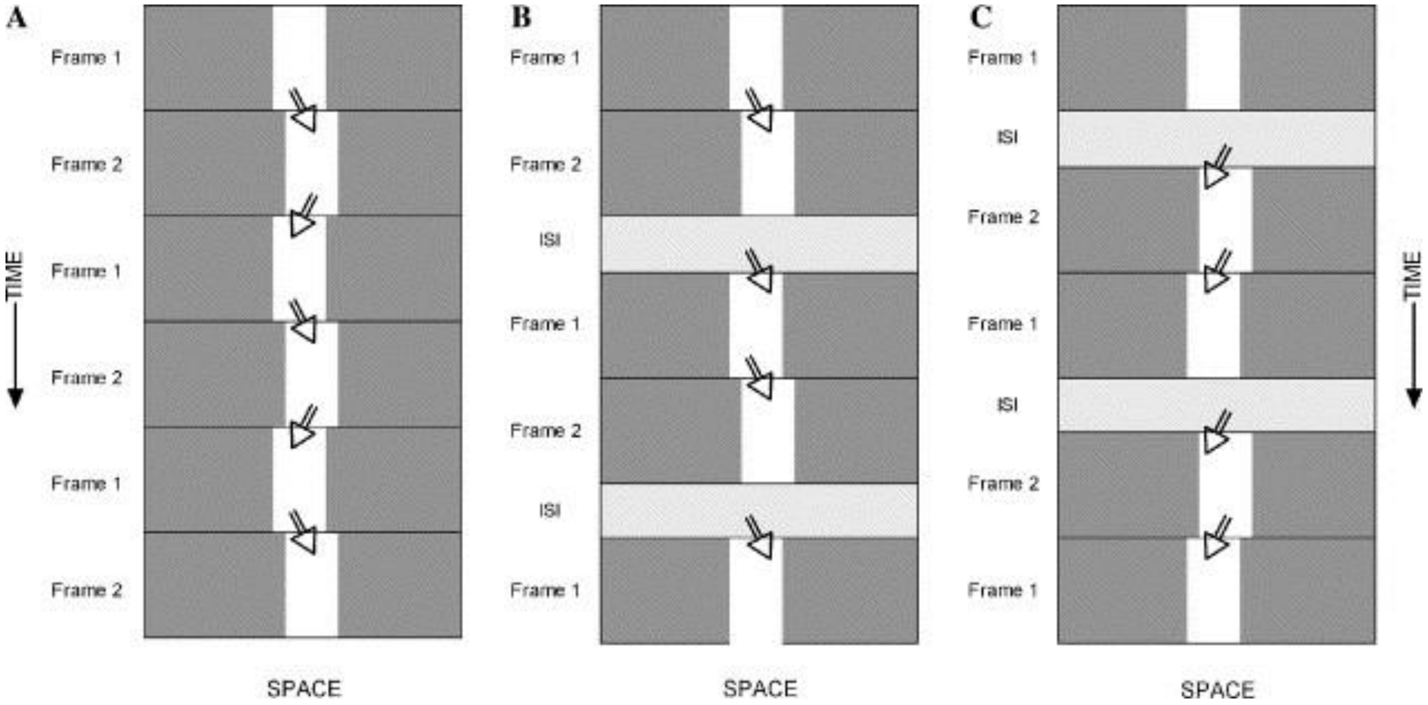
(Anstis, 1970; Anstis & Rogers, 1975)



NAI: Negative After-Image

http://www.lifesci.sussex.ac.uk/home/George_Mather/

http://www.lifesci.sussex.ac.uk/home/George_Mather/Motion/
http://www.michaelbach.de/ot/mot_reverse-phi/index.html



Two-stroke motion. (A) Illustrates standard apparent motion; (B and C) illustrate two-stroke motion. (A) A white rectangle shifts rightward from frame 1 to frame 2. When the two frames are presented repeatedly in alternation, they create an impression of to-and-fro movement (arrows). (B) If a brief blank inter-stimulus interval (ISI) is inserted at the transition between frame 2 and frame 1, the direction of apparent motion during the transition reverses, so that presentation of the whole sequence creates an impression of continuous forward motion (arrows). (C) If the blank ISI is inserted between frame 1 and frame 2, presentation of the whole sequence creates an impression of continuous backward motion (arrows)

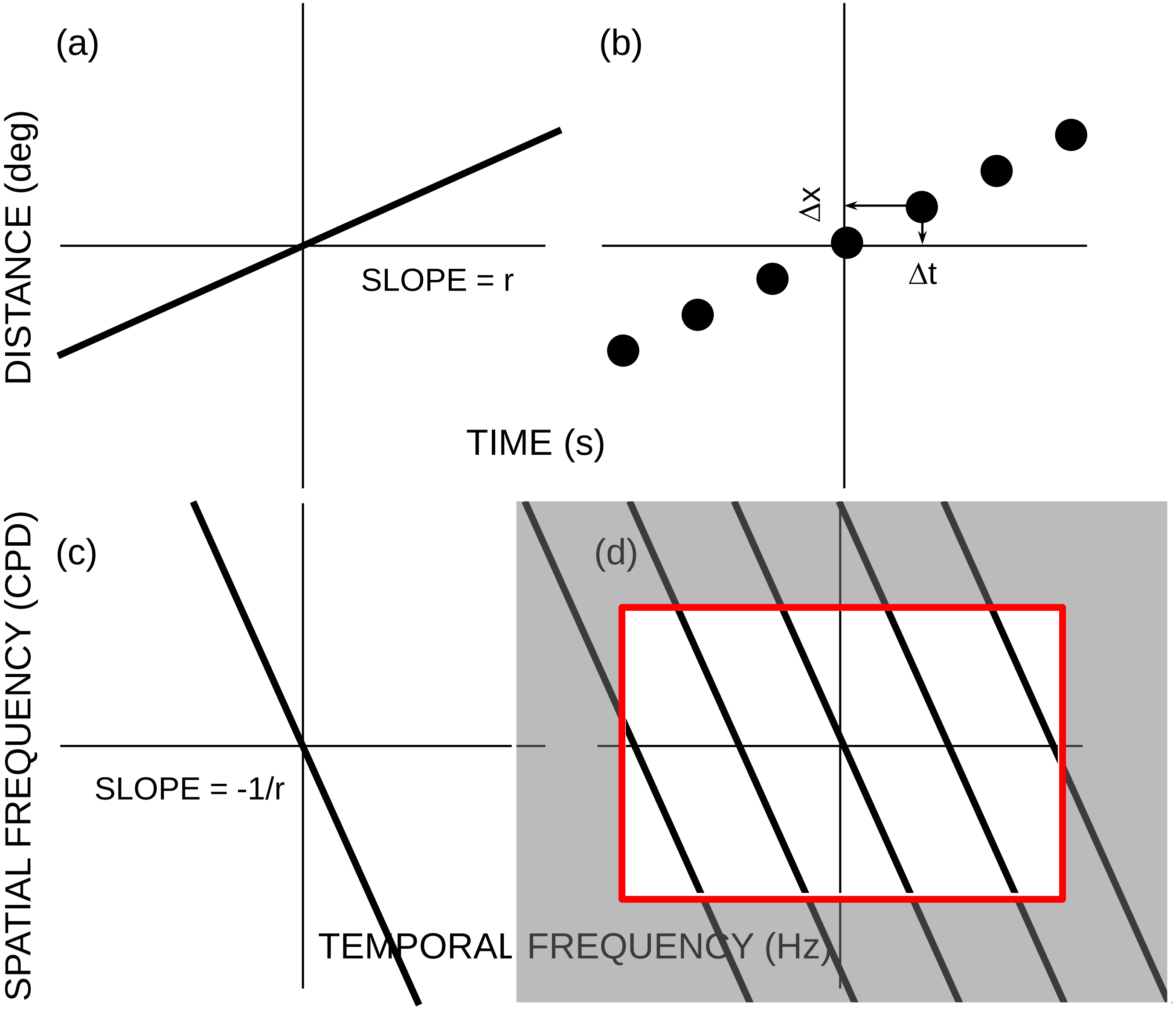
George Mather

Two-stroke: A new illusion of visual motion based on the time course of neural responses in the human visual system

Vision Research, Volume 46, Issue 13, 2006, 2015–2018

http://dx.doi.org/10.1016/j.visres.2005.12.022

THE WINDOW OF VISIBILITY



Watson, Ahumada & Farrell (1986). *J. Opt. Soc. Am. 3*(3) 300-307.

Motion as spatio-temporal texture

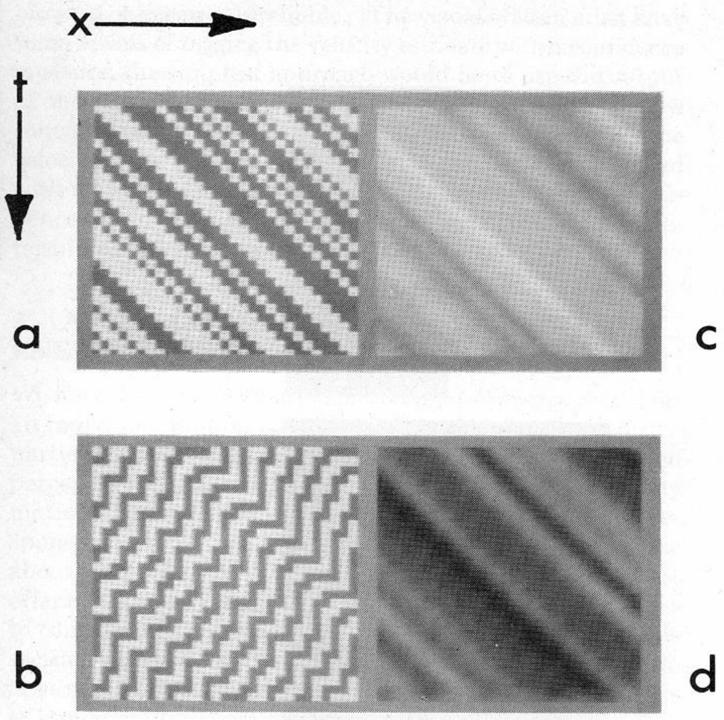


Fig. 16. a, An (x, t) plot of a random bar pattern, moving to the right in steps. b, The reverse-phi version: The pattern moves to the right and the bars reverse polarity on each step. c, The response of an opponent-energy channel to normal motion. The response is mainly positive, signaling rightward motion. d, The response of the channel to the reverse-phi display. Now the response is mainly negative, signaling leftward motion.

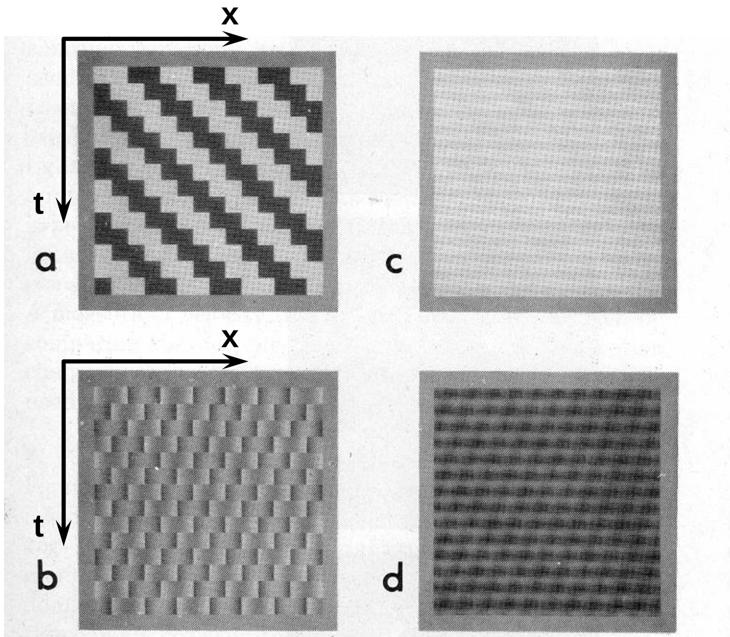
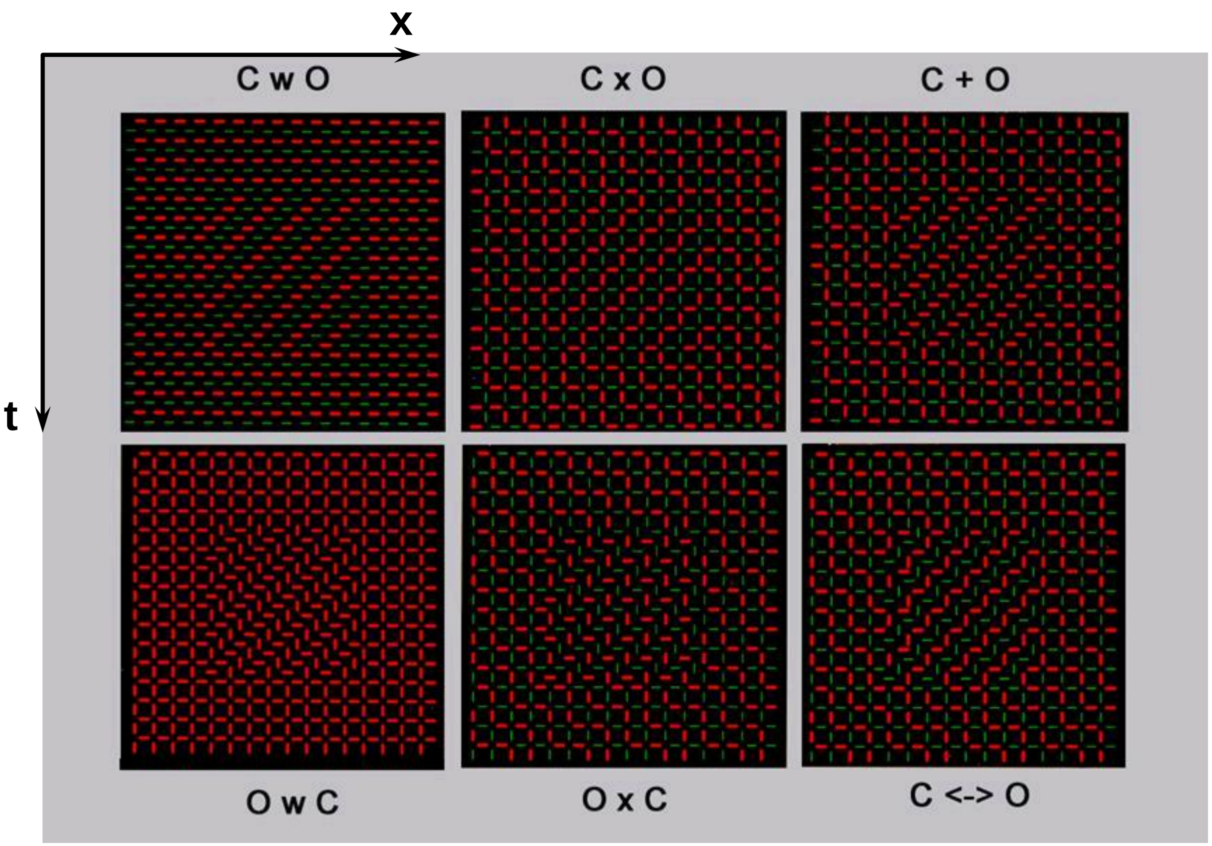


Fig. 17. a, An (x, t) plot of a square wave's motion. b, A (x, t) plot of a fluted square wave's motion. c, The response of a medium spatial-frequency opponent-motion channel when stimulated by the square wave. Rightward motion (bright) is signaled. d, The response of the same channel when stimulated by the fluted square wave. Leftward motion (dark) is signaled.

Adelson, E. H. & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A 2*, 284-299.

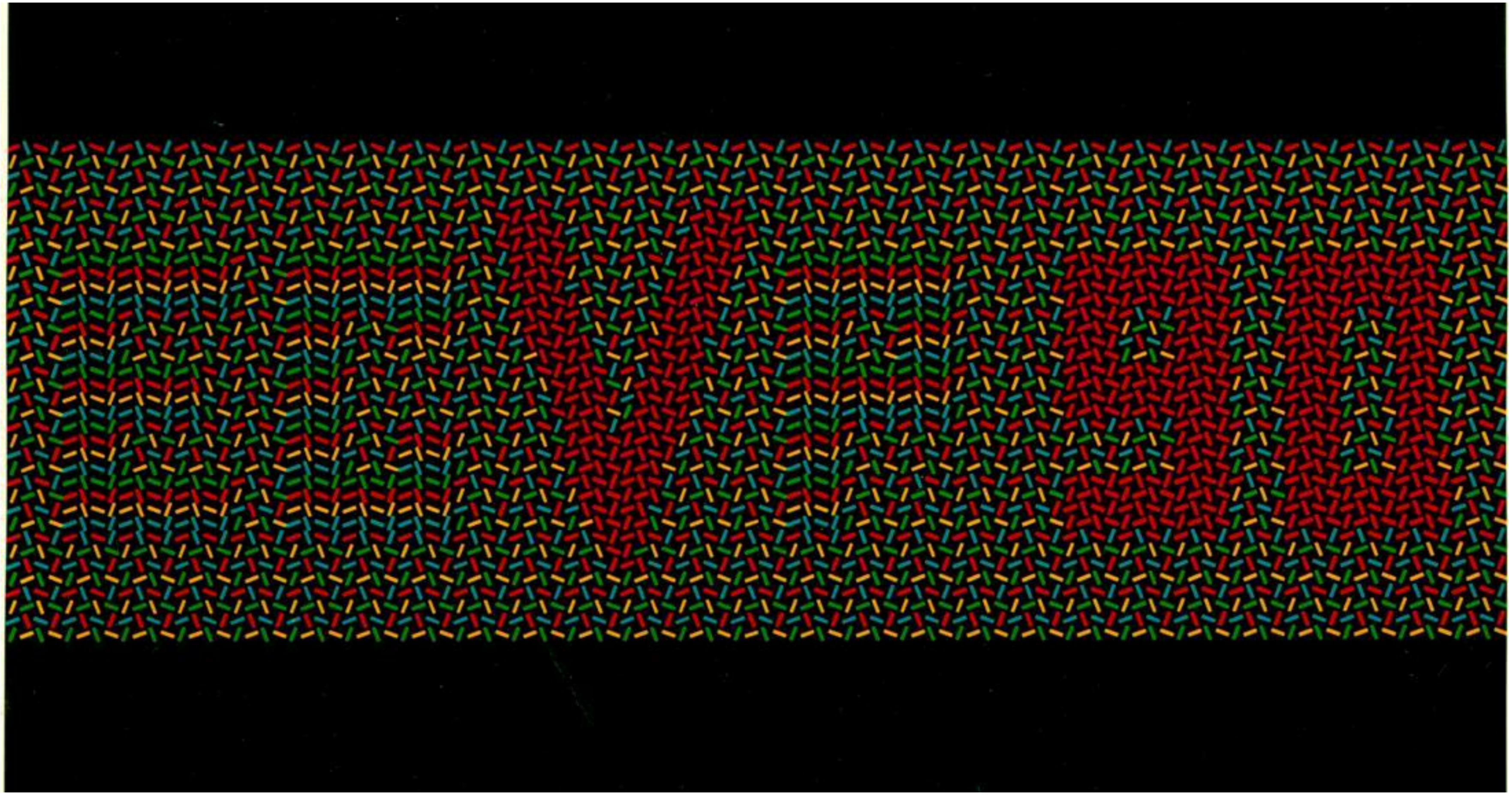
Motion & Texture

(Gorea & Papathomas, 1991)



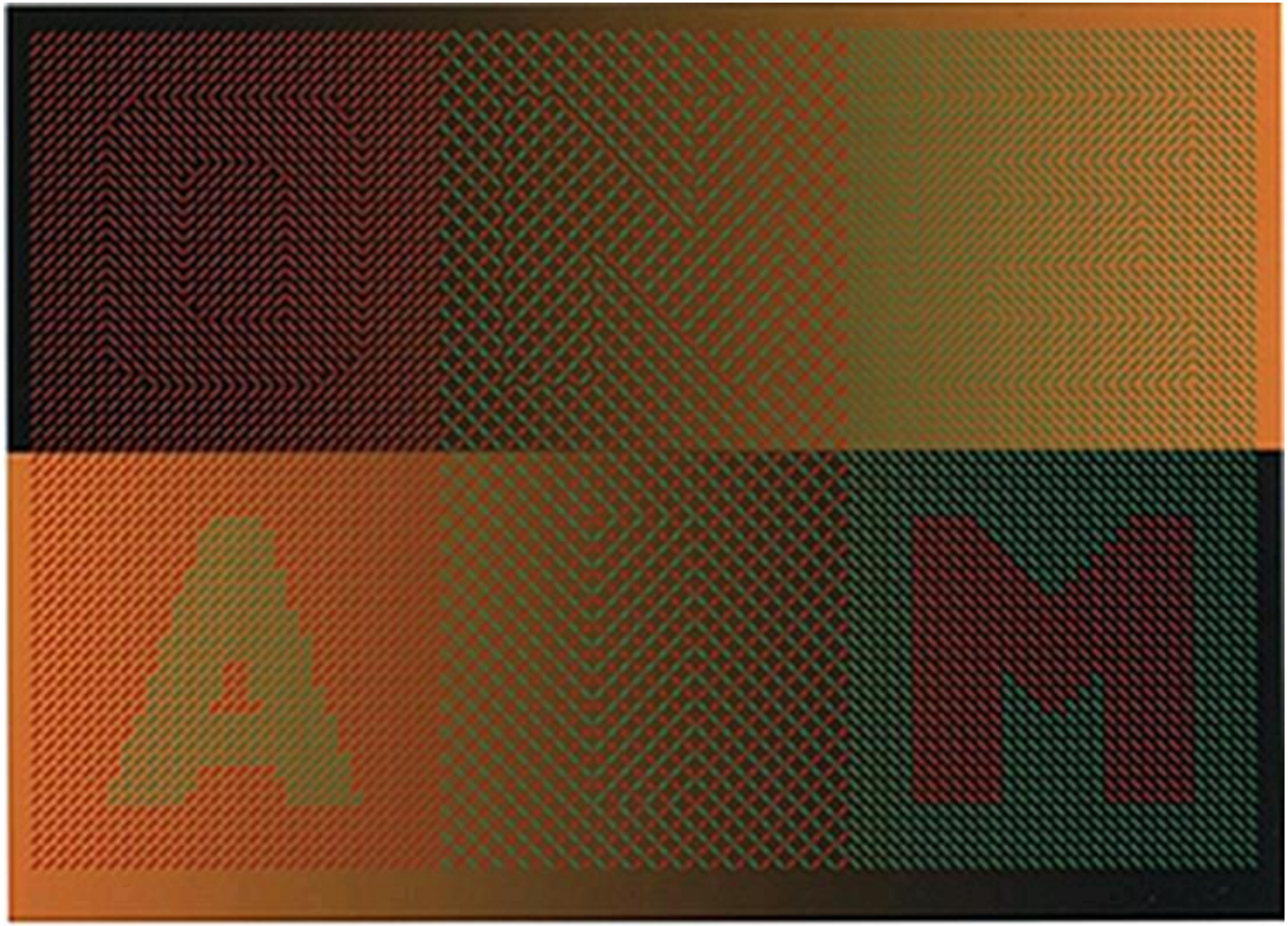
Motion & Texture

(Gorea & Papathomas, 1993)



Motion & Texture

(Gorea & Papathomas, 1997)



Motion “Energy”

Motion is orientation in space-time

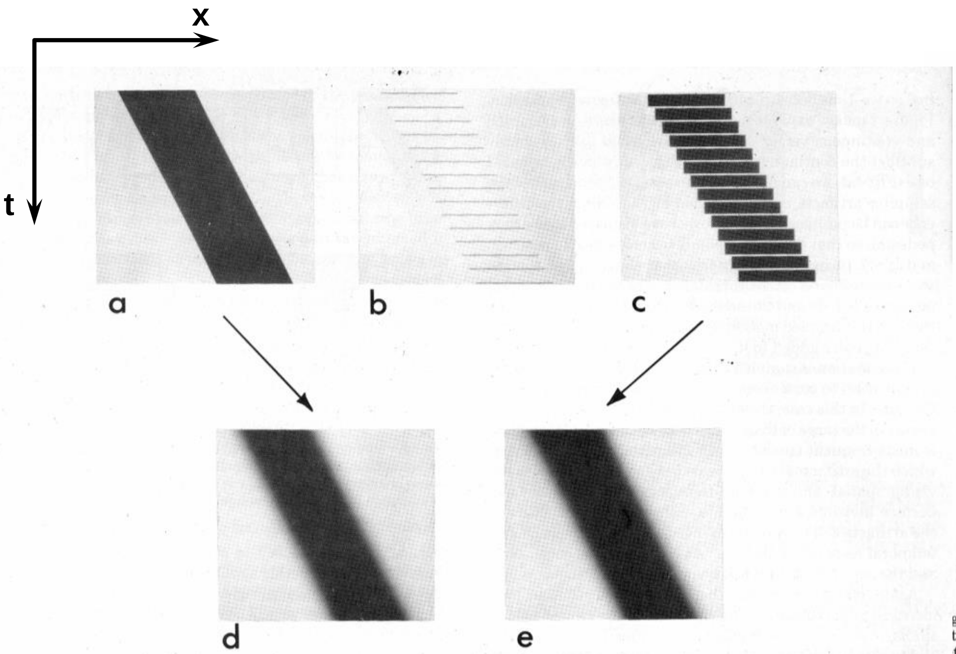
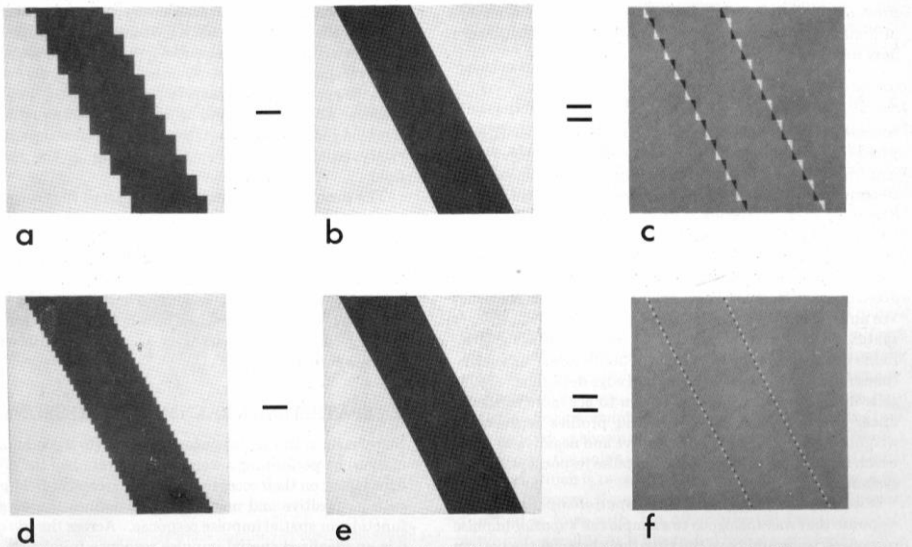


Fig. 3. a, An (x, t) plot of a bar moving to the right over time. Time proceeds downward. The vertical dimension is not shown. b, An (x, t) plot of the same bar, sampled in time. c, The sampled motion as displayed in a movie in which each frame remains on until the next one appears. d, Continuous motion after spatiotemporal blurring. e, Sampled motion after spatiotemporal blurring. The middle- and low-frequency information is almost the same for the two stimuli.

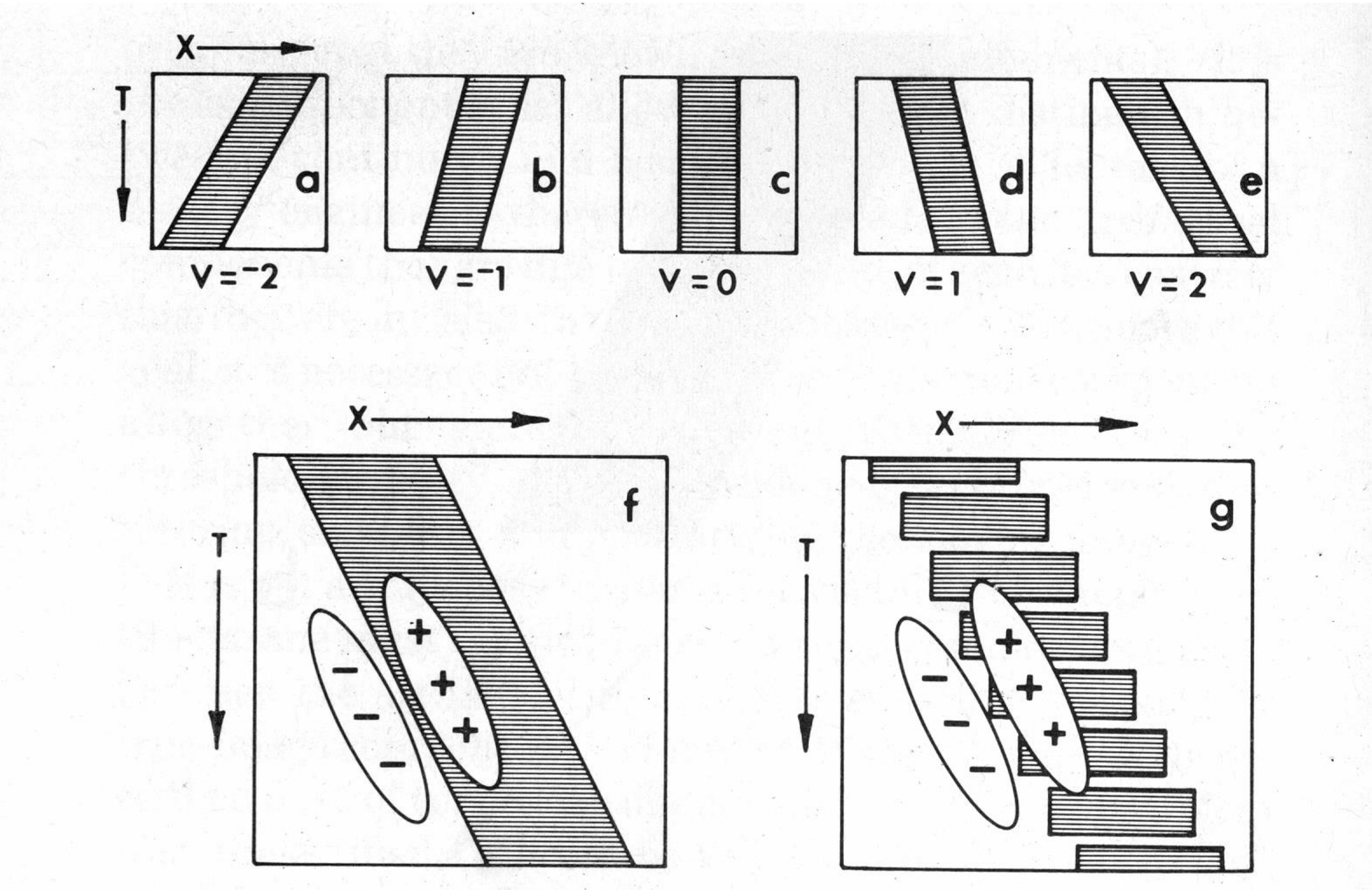


g. 4. (x, t) plots of moving bars. a, A movie of a bar moving to the right. b, A bar moving to the right continuously. c, The difference (sampling tifacts) between the sampled and continuous motions. d, A movie sampled at a high frame rate. e, Continuous motion. f, The difference tween the finely sampled and continuous motion. When the sampling rate is high, the sampling artifacts become difficult or impossible to e.

Adelson, E. H. & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A 2*, 284-299.

Motion "Energy"

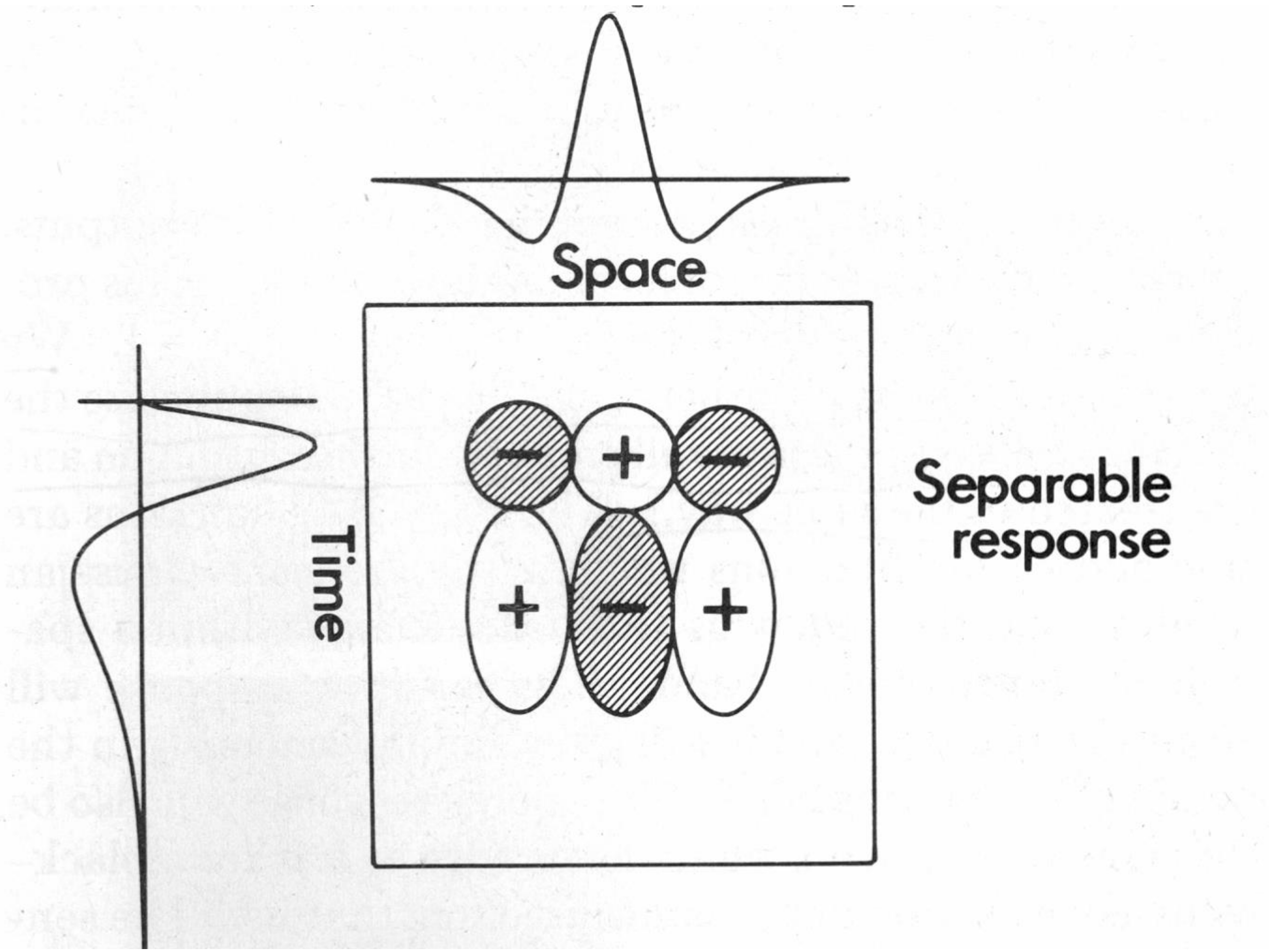
Motion is orientation in space-time



Adelson, E. H. & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A 2*, 284-299.

Motion “Energy”

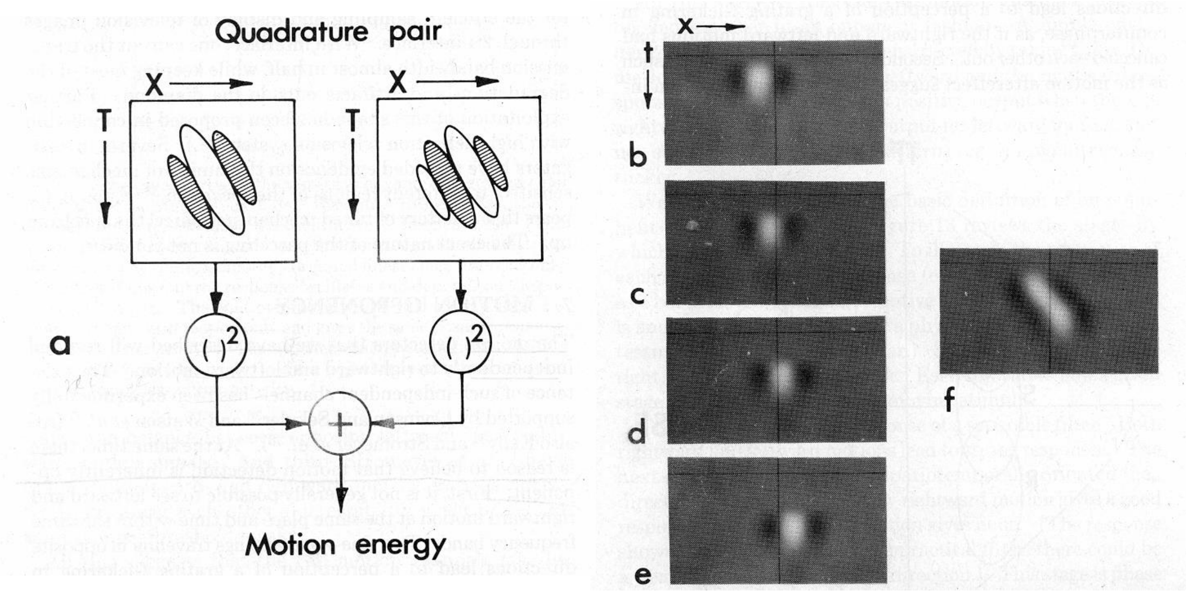
Non-oriented Spatio-Temporal RF

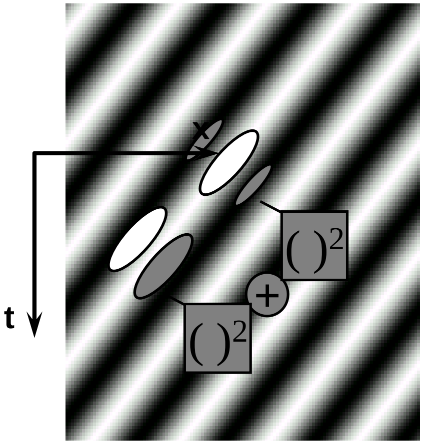


Adelson, E. H. & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A 2*, 284-299.

Motion "Energy"

Oriented, Phase independent Spatio-Temporal RF: "energy detectors"





Adelson, E. H. & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A 2*, 284-299.

Motion “Energy”

Building up an Oriented Spatio-Temporal RF

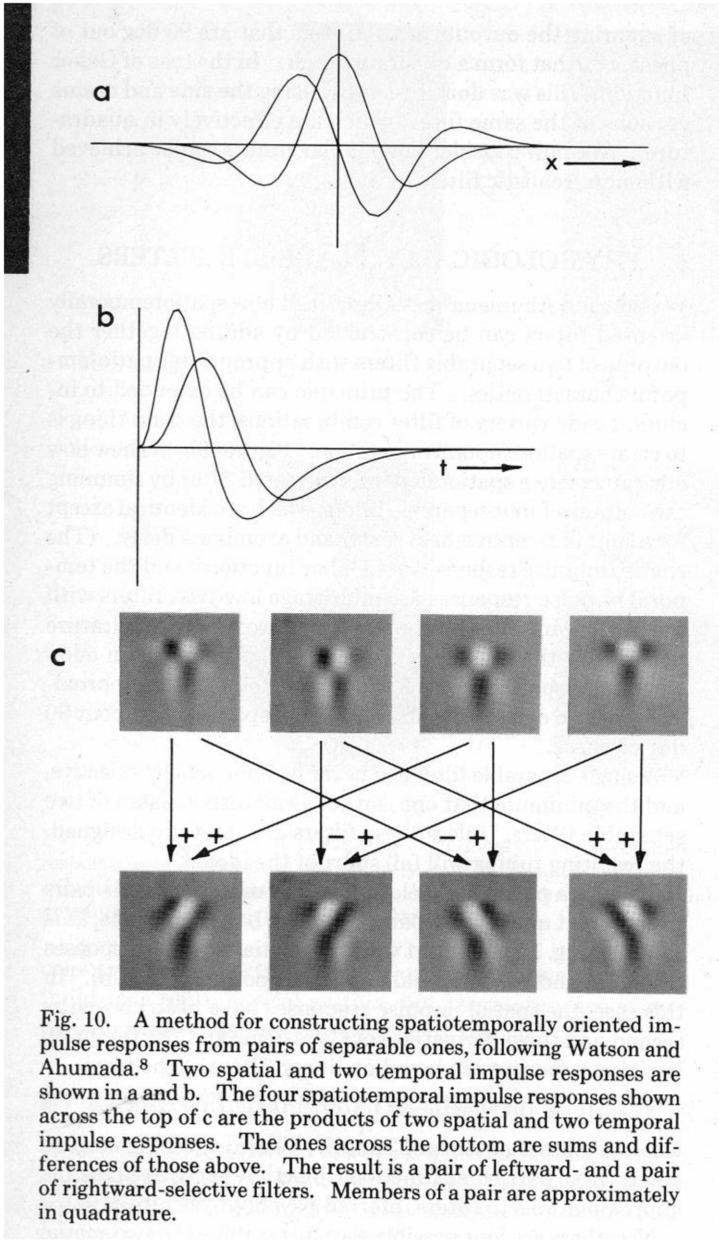
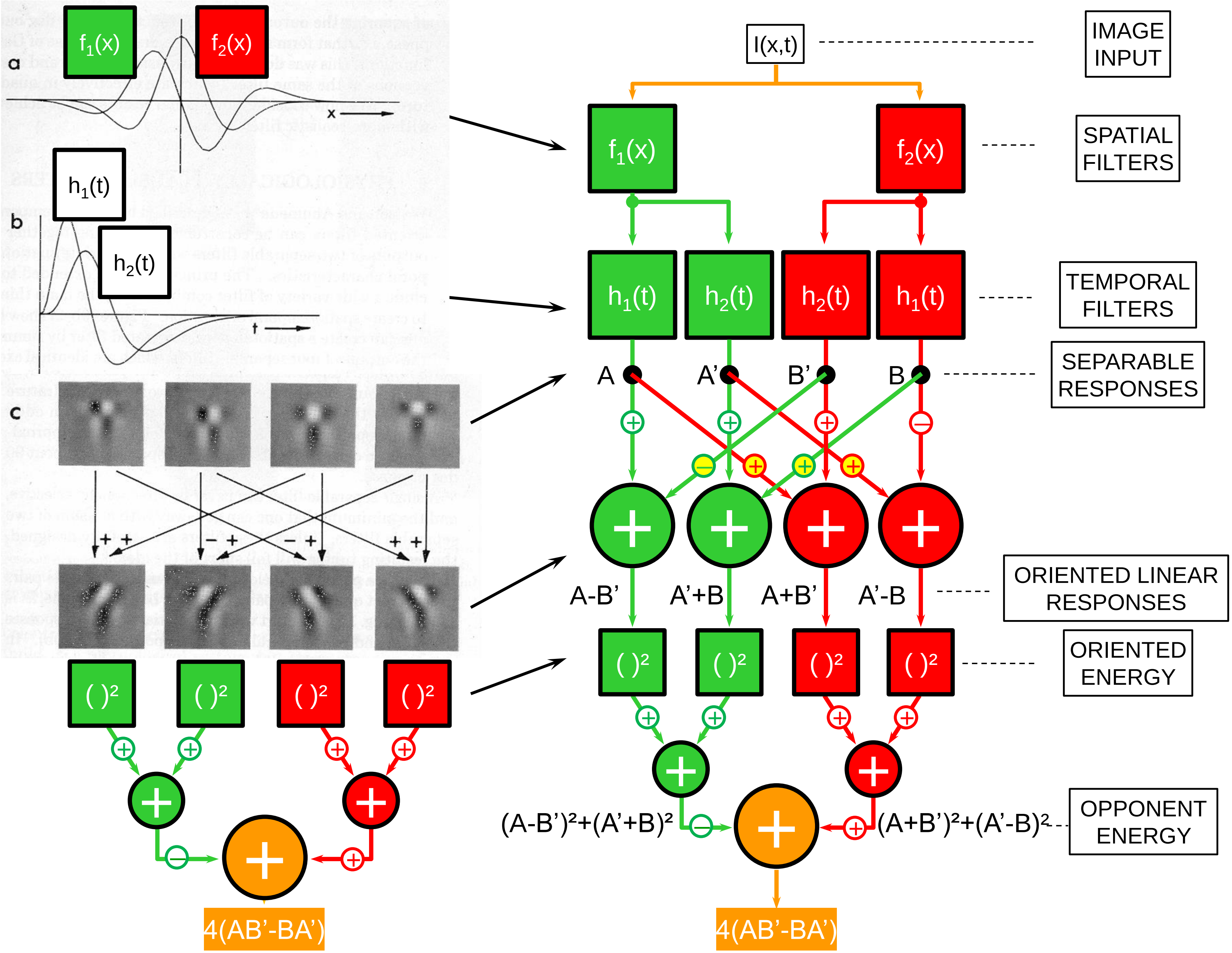


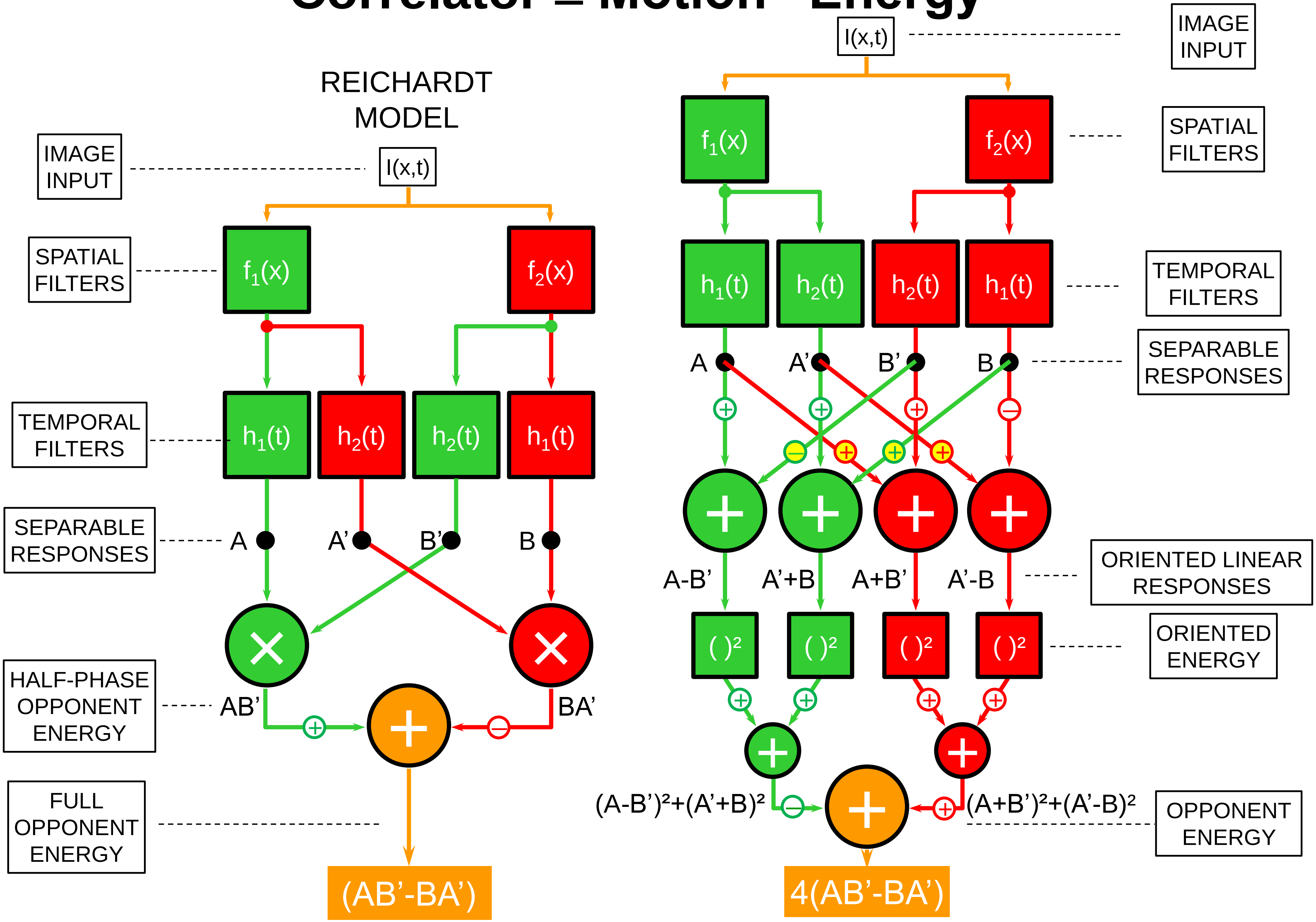
Fig. 10. A method for constructing spatiotemporally oriented impulse responses from pairs of separable ones, following Watson and Ahumada.[8] Two spatial and two temporal impulse responses are shown in a and b. The four spatiotemporal impulse responses shown across the top of c are the products of two spatial and two temporal impulse responses. The ones across the bottom are sums and differences of those above. The result is a pair of leftward- and a pair of rightward-selective filters. Members of a pair are approximately in quadrature.

Adelson, E. H. & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A 2*, 284-299.

ENERGY MODEL



Correlator ≡ Motion "Energy"



REICHARDT MODEL

IMAGE INPUT — $I(x,t)$

SPATIAL FILTERS — $f_1(x)$, $f_2(x)$

TEMPORAL FILTERS — $h_1(t)$, $h_2(t)$, $h_2(t)$, $h_1(t)$

SEPARABLE RESPONSES — A, A', B', B

HALF-PHASE OPPONENT ENERGY — AB', BA'

FULL OPPONENT ENERGY

(AB'-BA')

IMAGE INPUT — $I(x,t)$

SPATIAL FILTERS — $f_1(x)$, $f_2(x)$

TEMPORAL FILTERS — $h_1(t)$, $h_2(t)$, $h_2(t)$, $h_1(t)$

SEPARABLE RESPONSES — A, A', B', B

ORIENTED LINEAR RESPONSES — A-B', A'+B, A+B', A'-B

ORIENTED ENERGY — $(\ )^2$

OPPONENT ENERGY — $(A-B')^2+(A'+B)^2$, $(A+B')^2+(A'-B)^2$

4(AB'-BA')

Adelson, E. H. & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A 2*, 284-299.

Motion “Energy”

Advantages / Disadvantages

- Advantages – phase/polarity invariant

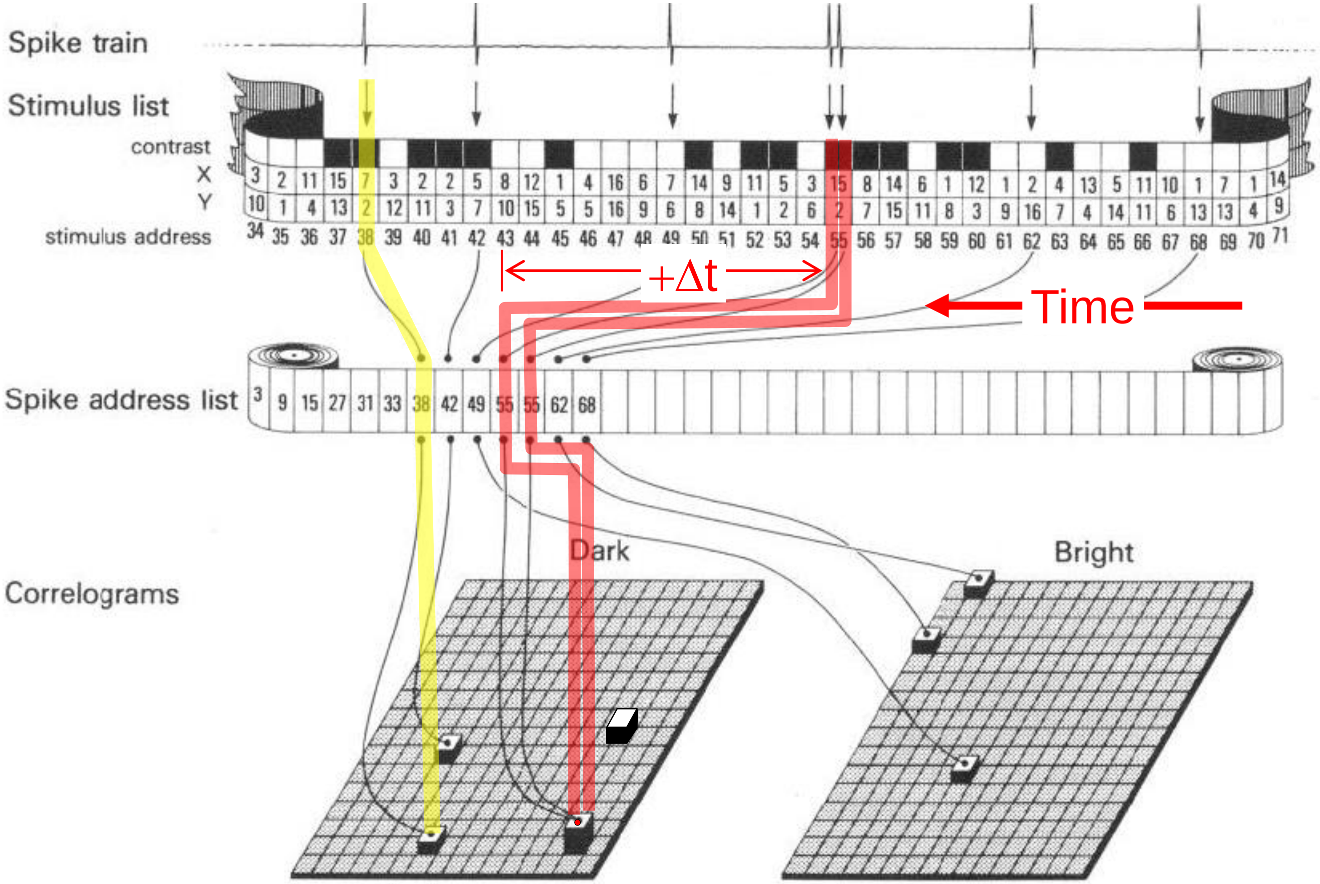
- Problems
  - Individual filters measure the degree of match not the speed:
  - For speed computation needs to interpret a population code

Jump to
THE APERTURE PROBLEM

▶

Reverse Correlation - Physiology

In the white-noise approach to RF mapping, a rapid, pseudo-random stimulus sequence that consists of patterns of spots or bars is presented, and the neuronal spike train is correlated to the stimulus sequence (that is, cross- or reverse-correlation). The aim of this correlation procedure is to characterize the transformation that occurs between the visual stimulus and the response of a neuron (that is, the neuron's 'transfer function'). Because stimuli are presented in rapid succession, without pausing to collect the response to each pattern, this technique is fast.

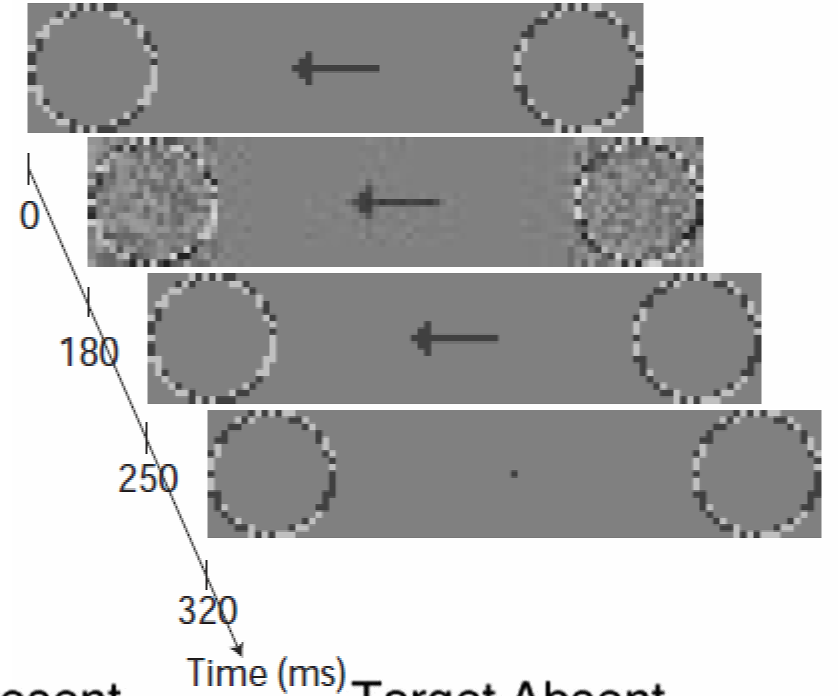


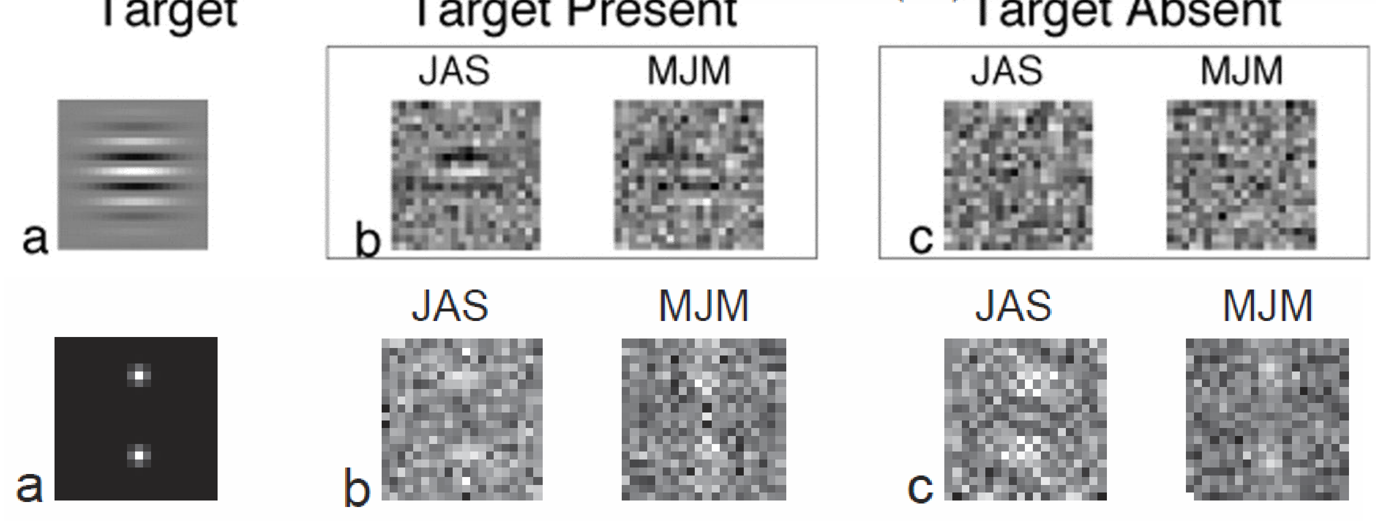
Jones & Palmer (1987) J. Neurophysiol., 58, 1187-1211.

Reverse Correlation - Psychophysics

Solomon (2002) *J. Vision.*, *2*, 105-120.

		STIMULUS	
		Noise	Signal
RÉPONSE	Yes	**FA**	**Hit**
	No	CR	Miss

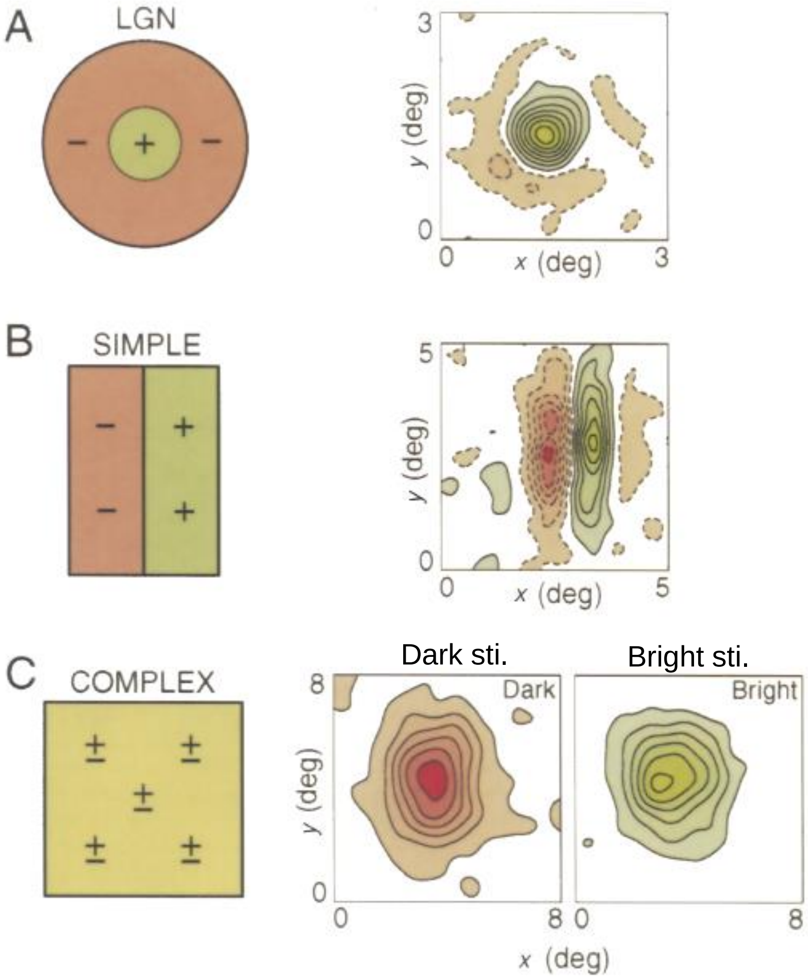




Receptive field (RF) *spatial* structure of the major classes of neurons in the geniculostriate pathway of the cat

Schematic RF

Measured RF
(Rev. Correl.)



Dark sti.

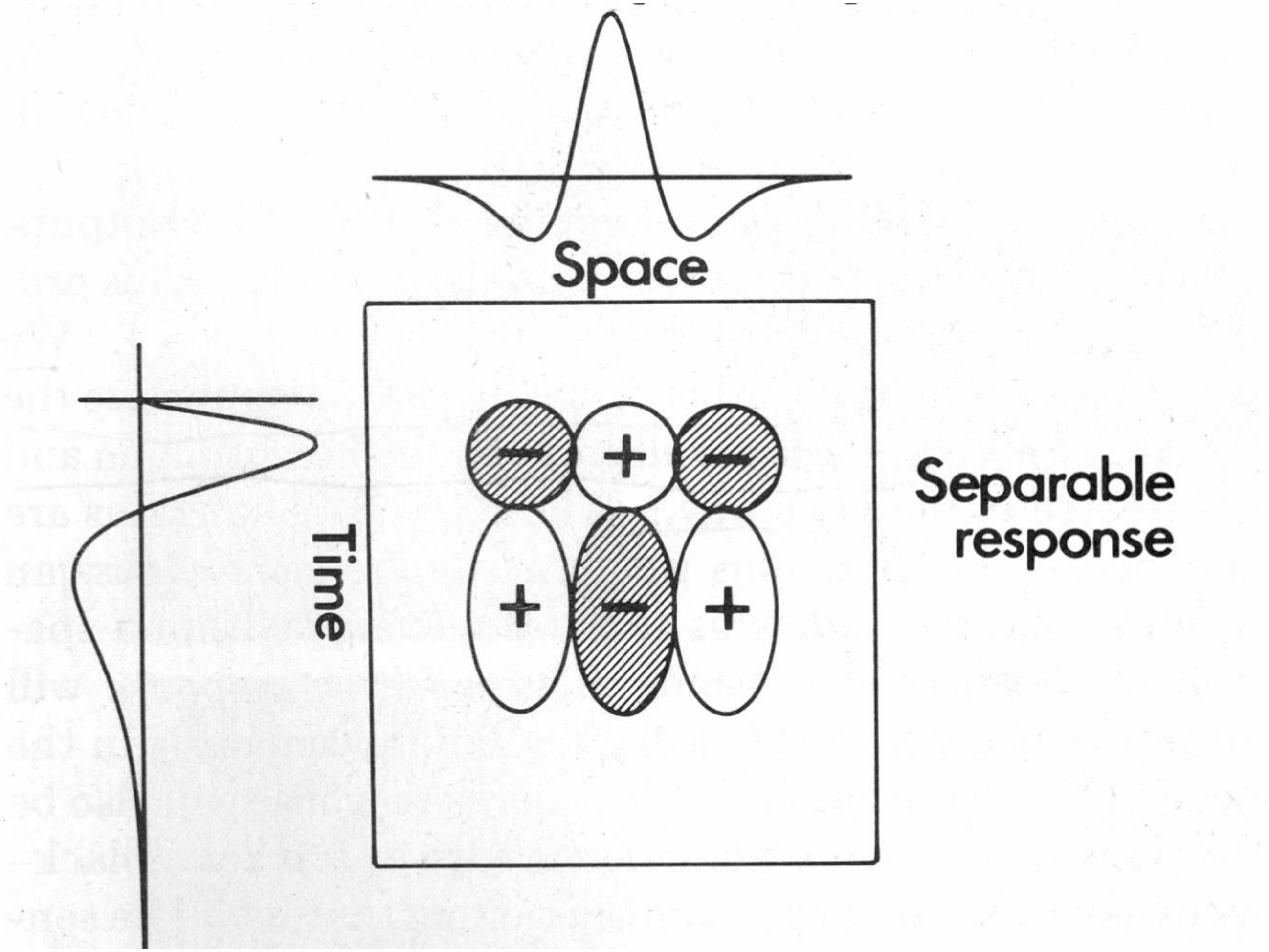
Bright sti.

Area 17

DeAngelis, Ohzawa & Freeman (1995). *TINS, 18*(10), 451-458.

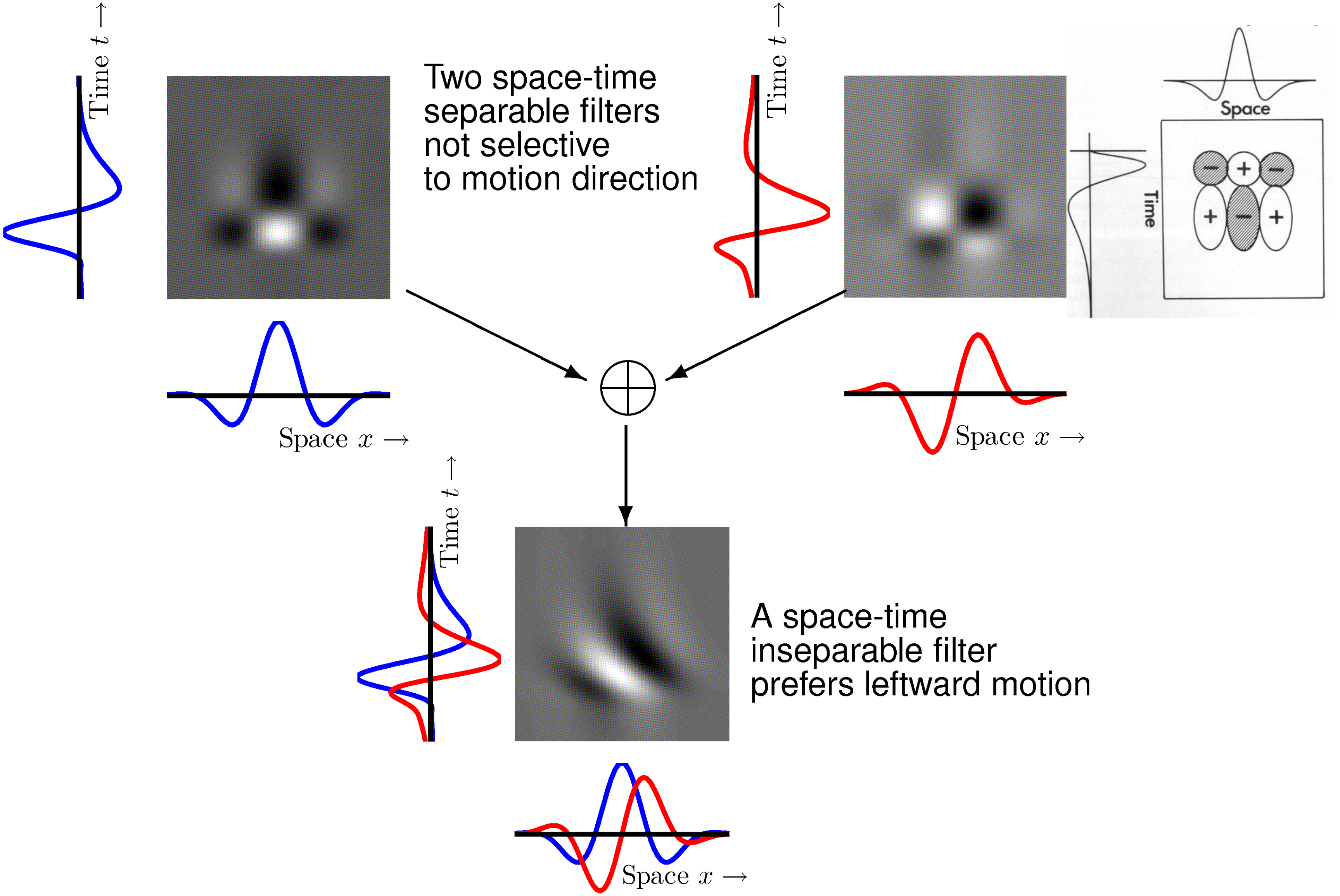
Motion “Energy”

Non-oriented Spatio-Temporal RF



Adelson, E. H. & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A 2*, 284-299.

Quadrature model of motion direction selectivity

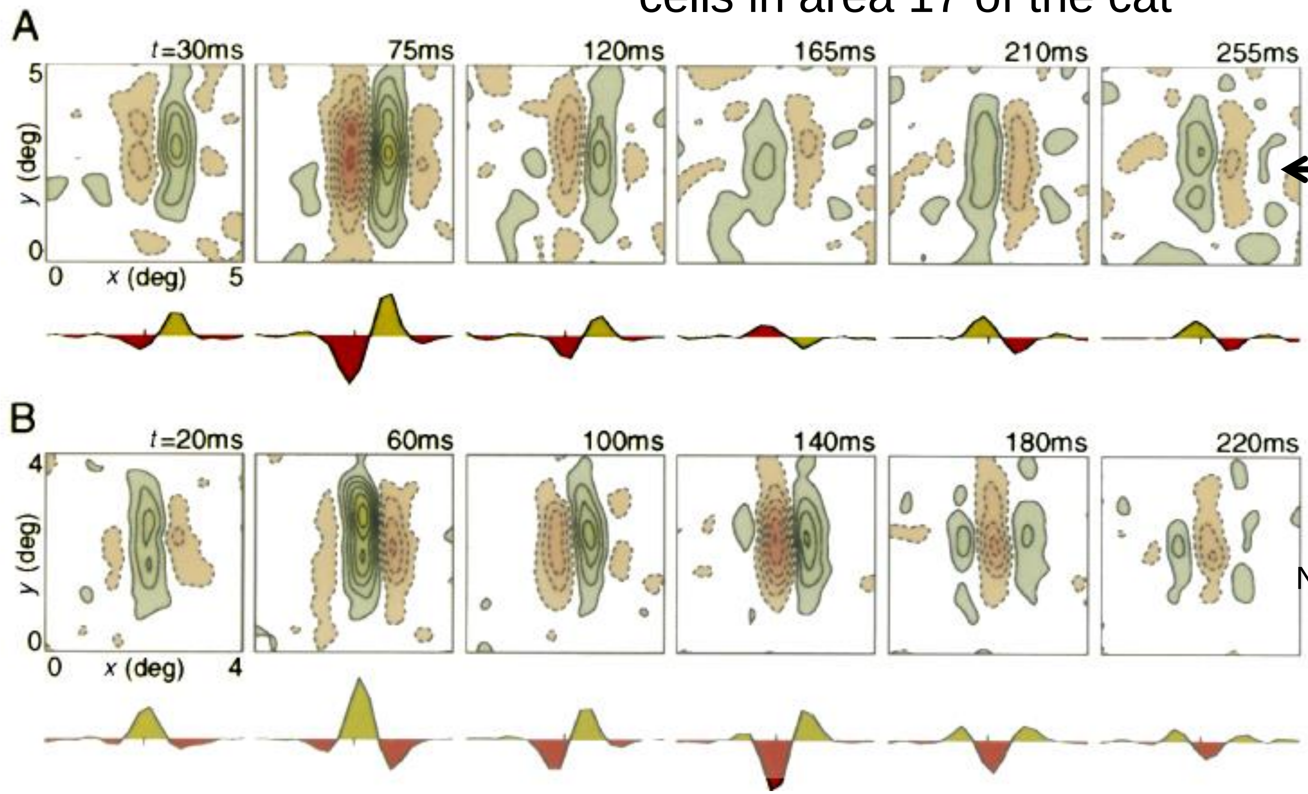


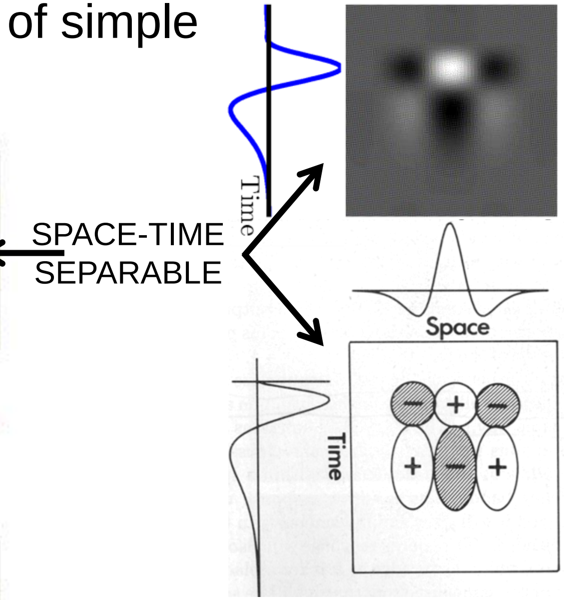
Two space-time
separable filters
not selective
to motion direction

A space-time
inseparable filter
prefers leftward motion

Figure from "Understanding Vision: theory, models, and data", by Li Zhaoping, Oxford University Press, 2014

Receptive field (RF) *spatial* structure *over time* of simple cells in area 17 of the cat





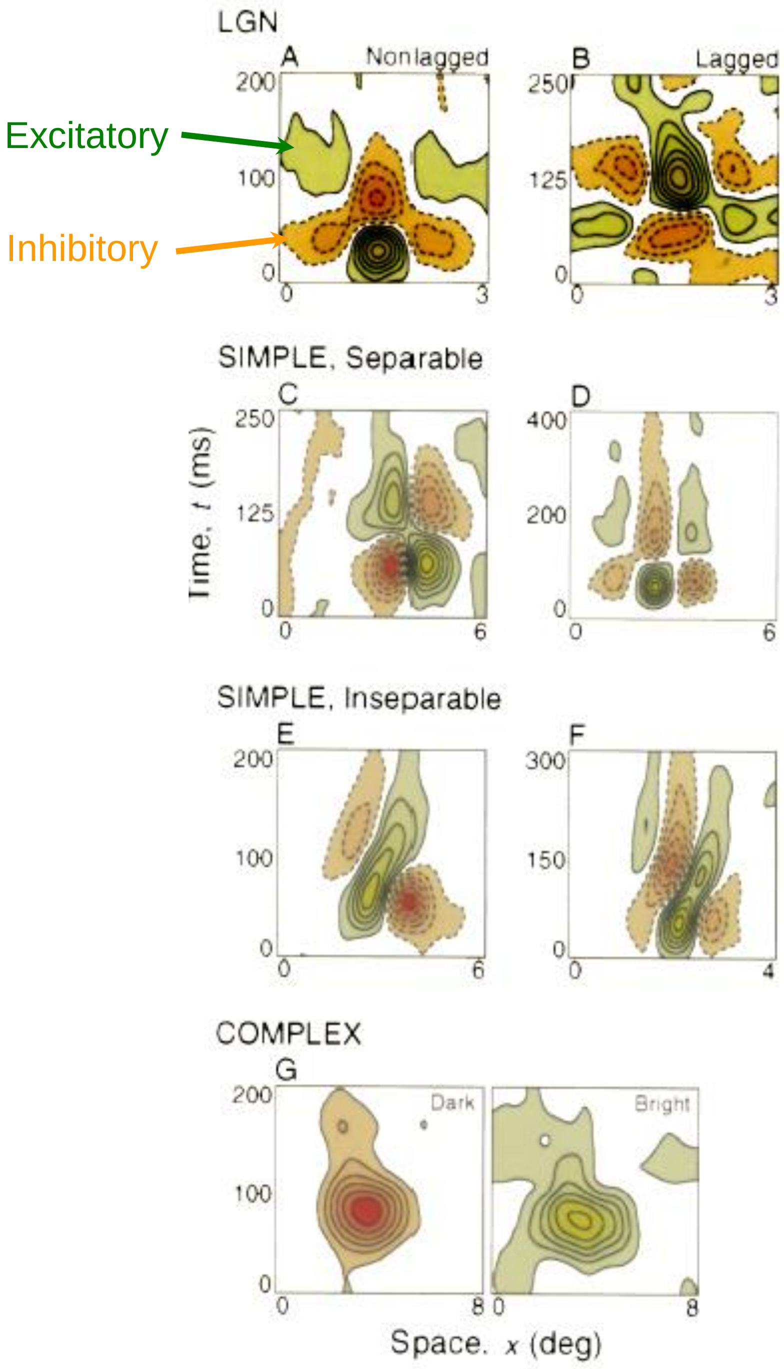
SPACE-TIME SEPARABLE

SPACE-TIME NON-SEPARABLE

Dynamics of receptive field (RF) structure of simple cells from striate cortex of the cat. By varying the correlation delay, t, in the RF mapping algorithm, 'snapshots' of the RF can be obtained at different times relative to stimulus onset. These data were obtained using a reverse correlation technique. For each cell, two-dimensional (2D) spatial (x-y) RF profiles are shown, as isoamplitude contour maps (conventions as in Fig. 1), for six values of t. Below each contour plot is a 1D RF profile that is obtained by integrating the 2D profile along the y axis, which is parallel to the cell's preferred orientation. Positive deflections (shaded green) in these 1D profiles indicate bright-excitatory subregions; negative deflections (shaded red) correspond to dark-excitatory subregions. (**A**) The RF of this simple cell is approximately space-time separable. From t = 30 ms to t = 120 ms, the RF profile has two dominant subregions, which are arranged with the dark-excitatory subregion on the left. These subregions are strongest at t = 75 ms. Between t = 120 ms and t = 165 ms, the RF reverses polarity, so that the bright-excitatory subregion is now on the left. This arrangement then persists over the remainder of the cell's response duration. Note that, at all values of t, the 1D RF profile is approximately odd symmetric (sine phase). (**B**) A fundamentally different type of spatiotemporal behavior is illustrated here. For this cell, the RF is space-time inseparable - the spatial organization of the RF changes over time. At t = 20 ms, the 1D profile is approximately even symmetric (cosine phase) whereas, at t = 100 ms, the RF profile is odd symmetric. Later, at t = 180 ms, the RF becomes even symmetric again but the profile is inverted relative to that at t = 20 ms.

DeAngelis, Ohzawa & Freeman (1995). *TINS, 18*(10), 451-458.

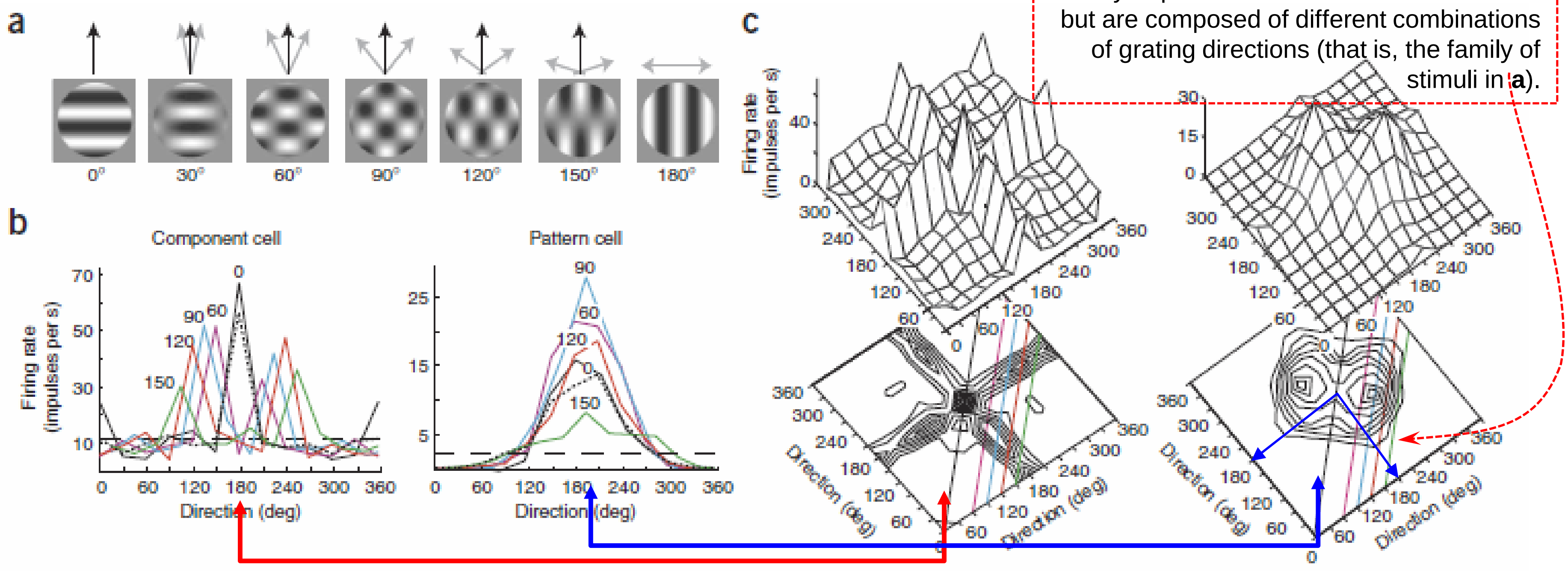
Spatiotemporal receptive field (RF) profiles (x-t plots) for neurons in the lateral geniculate nucleus (LGN) and striate cortex of the cat.



In each panel, the horizontal axis represents space (x), and the vertical axis represents time (t). For panels A-F, solid contours (with green shading) delimit bright-excitatory regions, whereas broken contours (with red shading) indicate dark-excitatory regions. To construct these x-t plots, 1D RF profiles are obtained, at finely spaced time intervals (5-10ms), over a range of values of t. These 1D profiles are then 'stacked up' to form a surface, which is smoothed and plotted as a contour map. (**A**) An x-t profile is shown here for a typical ON-center, non-lagged X-cell from the LGN. For t < 50ms, the RF has a bright-excitatory center and a dark-excitatory surround. However, for t > 50ms, the RF center becomes dark-~~excita~~inhibitory, and the surround becomes bright-excitatory. (**B**) An x-t plot of an ON-center, lagged X-cell. Note that the second temporal phase of the profile is strongest. (**C**) An x-t profile for a simple cell with a space-time separable RF. For t < 100 ms, the RF has a dark-excitatory subregion to the left of a bright-excitatory subregion. For t > 100ms, each subregion reverses polarity, so that the bright-excitatory region is now on the left. (**D**) Data for another simple cell with an approximately separable x-t profile. (**E**) Data are shown for a simple cell with a clearly inseparable x-t profile. Note how the spatial arrangement of bright- and dark-excitatory subregions (that is, the spatial phase of the RF) changes gradually with time. (**F**) An inseparable x-t profile is shown here for the same simple cell for which 2D spatial profiles are shown in Fig. 2B. Note that the subregions are tilted to the right in the space-time domain. (**G**) x-t profiles are shown for the same complex cell as in Fig. 1C. Responses to bright and dark stimuli are shown separately because these regions overlap extensively.

DeAngelis, Ohzawa & Freeman (1995). *TINS, 18*(10), 451-458.

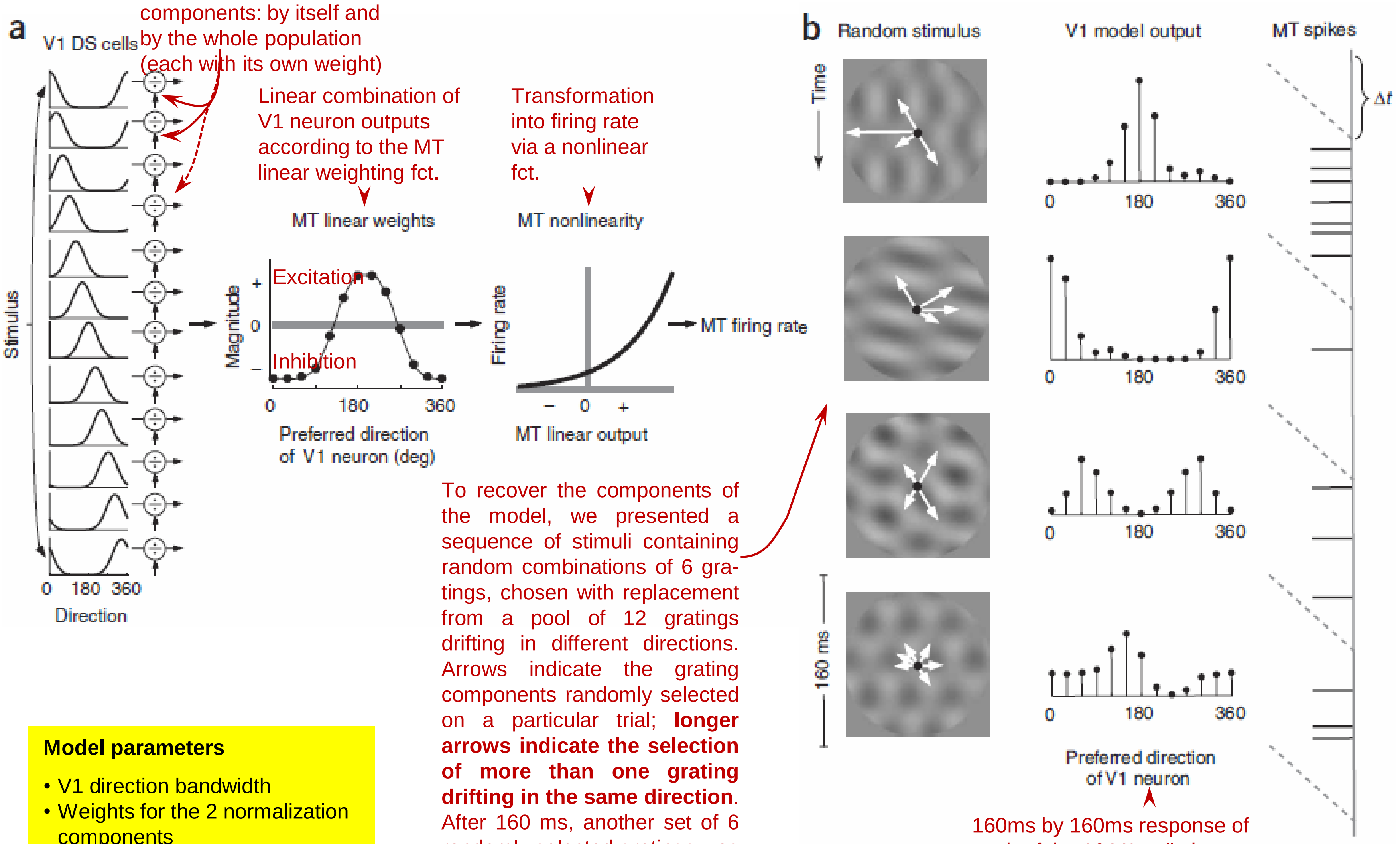
MODELING AND (A BIT OF) PHYSIOLOGY



Family of plaids that share an axis of motion but are composed of different combinations of grating directions (that is, the family of stimuli in **a**).

Component and pattern MT cell responses to plaid stimuli. (**a**) Example plaid stimuli. All plaids were constructed by superimposing two sinusoidal gratings of equal contrast moving at the same spatial and temporal frequency. Gray arrows show the motion directions of the two gratings; black arrows show the motion direction of the plaid. The angular separation between the two gratings (the plaid angle) is given below each stimulus. When the plaid angle is zero, the gratings form a single grating with twice the contrast. (**b**) Direction tuning curves for example component (left) and pattern (right) cells, collected for plaids with different plaid angles. Each colored tuning curve represents the response to a particular plaid. For the component cell on the left, the functions have two peaks, displaced from one another by the angle of the plaid; for the pattern cell on the right, all have similar shapes and preferred directions. Also shown is the half-contrast grating tuning curve (black dotted line). (**c**) Surface and contour plots of response of **a neuron preferring a 180° direction** as a function of the direction of the two grating components. The colored lines in c indicate the loci of the particular 2 component angular differences whose responses are shown in the same colors in b. Each direction interaction plot is symmetrical about the main diagonal (a plaid constructed from gratings 0° and 120° gratings is equivalent to a plaid constructed from 120° and 0° gratings). In the contour plots, contours begin at 20% of the maximum firing rate and subsequent contours indicate 10% increments.

Rust, Mante, Simoncelli & Movshon (2006). How MT cells analyze the motion of visual patterns. *Nat. Neurosci., 9*(11), 1420-31.



Divisive normalization by 2 components: by itself and by the whole population (each with its own weight)

Linear combination of V1 neuron outputs according to the MT linear weighting fct.

Transformation into firing rate via a nonlinear fct.

Excitation

Inhibition

To recover the components of the model, we presented a sequence of stimuli containing random combinations of 6 gratings, chosen with replacement from a pool of 12 gratings drifting in different directions. Arrows indicate the grating components randomly selected on a particular trial; **longer arrows indicate the selection of more than one grating drifting in the same direction**. After 160 ms, another set of 6 randomly selected gratings was immediately presented.

Model's parameters (bandwidth, 2 weights for normalization) were optimized optimized for each cell to maximize the likelihood of the observed responses.

160ms by 160ms response of each of the 12 V1 cells in **a**.

Hypothetical spike train of an MT cell in response to the random grating stimulus. The spikes were shifted by the latency of the cell's response (Δt) and the number of spikes occurring in a 160-ms bin counted. The cascade model is then fit to these spike count data.

Model parameters

- V1 direction bandwidth
- Weights for the 2 normalization components
- Exponent of the nonlinear transform

Rust, Mante, Simoncelli & Movshon (2006). How MT cells analyze the motion of visual patterns. *Nat. Neurosci., 9*(11), 1420-31.

RESPONSE NORMALIZATION

Normalization works (here) by

dividing each output

by the sum of all outputs.

Divisive inhibition

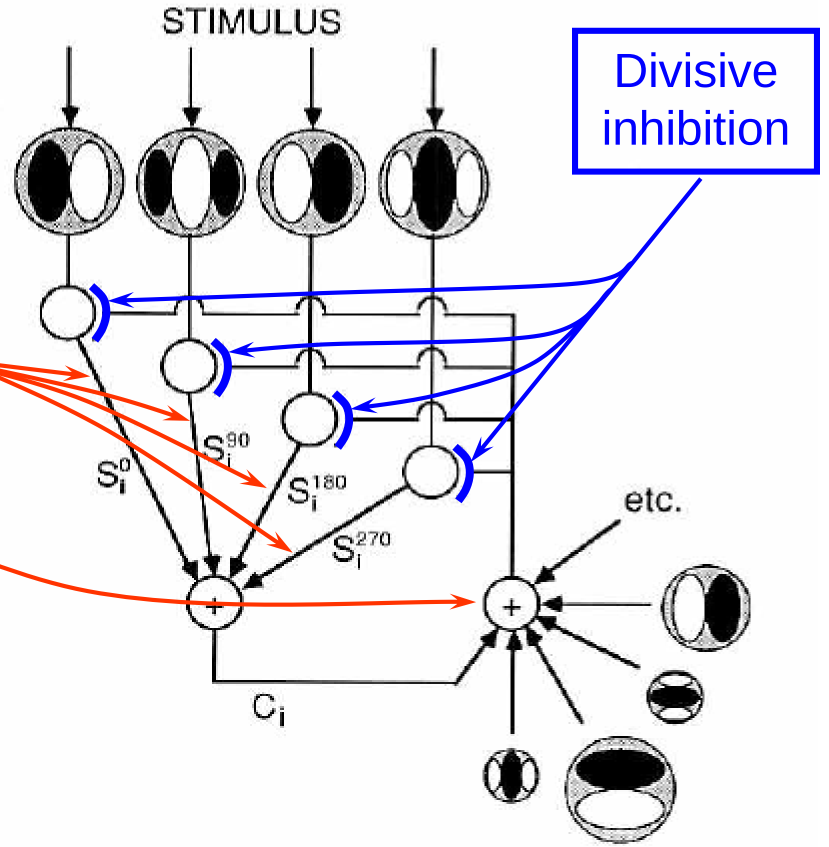
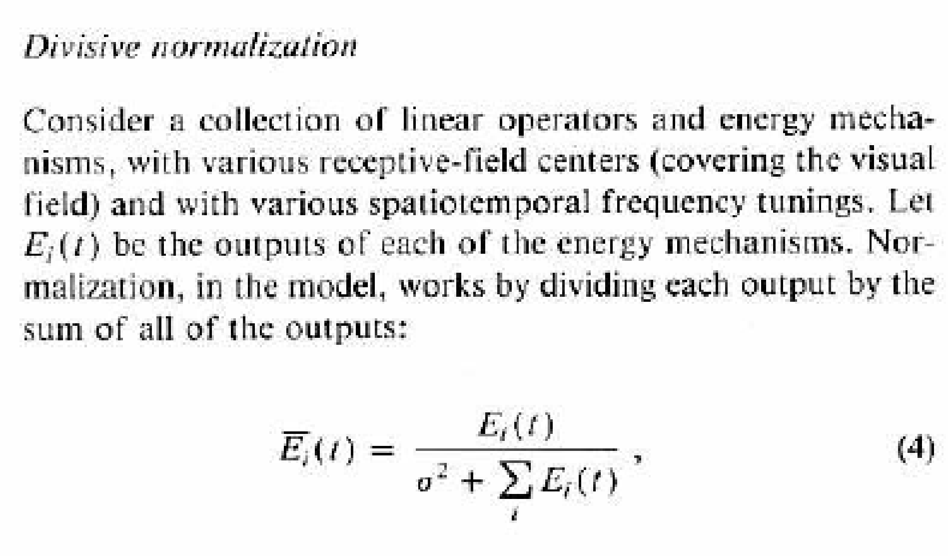


Fig. 1. Diagram of the various stages of the model. Linear weighting functions are depicted as circles, subdivided into excitatory (bright) and inhibitory (dark) subregions. The $S_i^?$ labels represent simple cell outputs, and the C_i label represents a complex cell output. The feedback signal is the combined energy at all orientations and nearby spatial frequencies, averaged over space and time. The feedback signal suppresses the simple cell responses by way of divisive suppression.

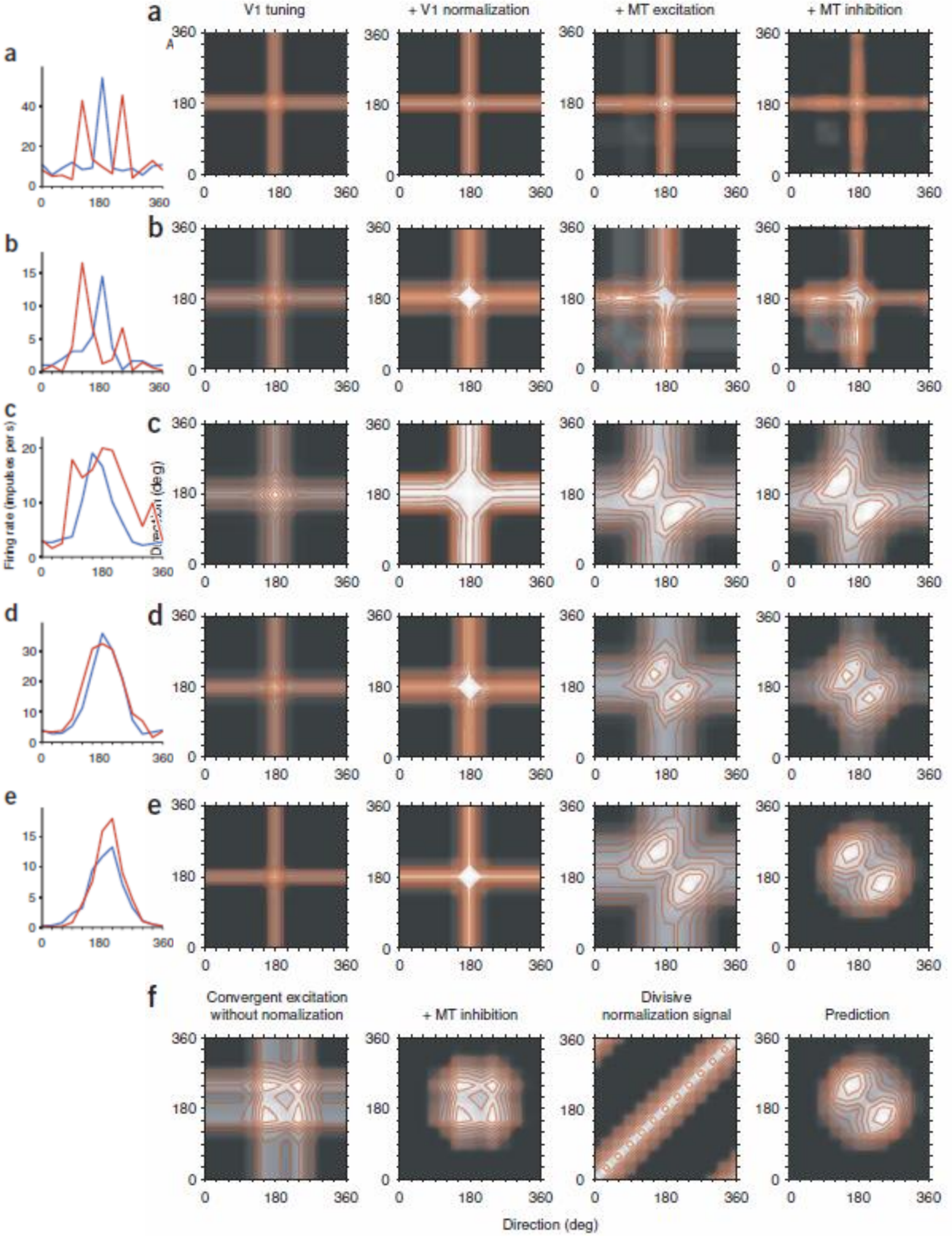


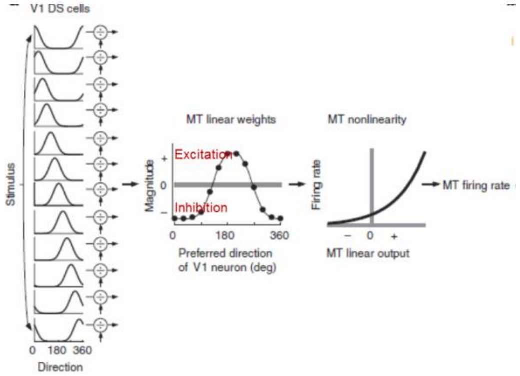
Divisive normalization

Consider a collection of linear operators and energy mechanisms, with various receptive-field centers (covering the visual field) and with various spatiotemporal frequency tunings. Let $E_i(t)$ be the outputs of each of the energy mechanisms. Normalization, in the model, works by dividing each output by the sum of all of the outputs:

$$\overline{E}_i(t) = \frac{E_i(t)}{\sigma^2 + \sum_i E_i(t)}, \qquad (4)$$

Heeger, D.J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, *9*, 181-197.





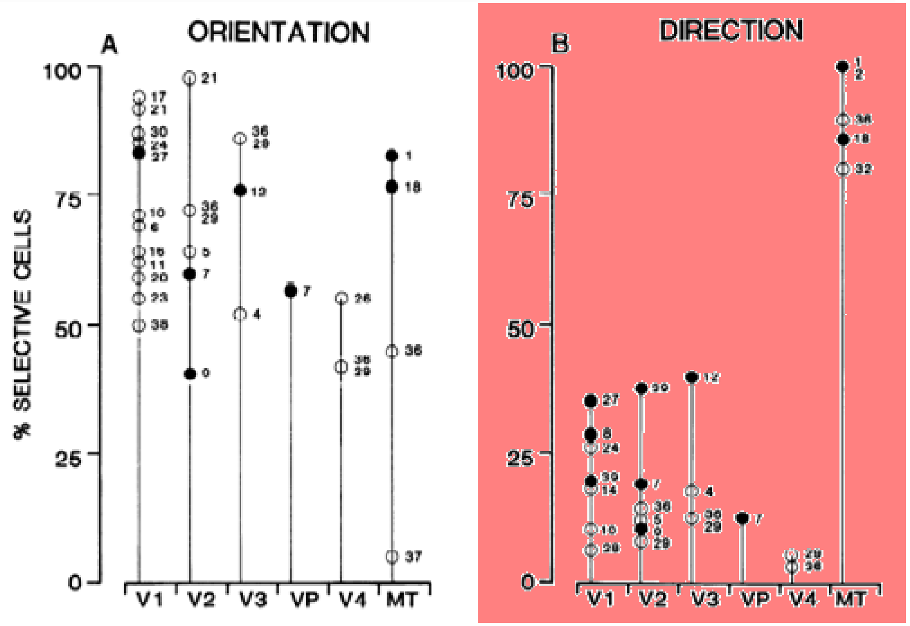
Dissection of the elements of the cascade model that create particular kinds of selectivity for the motion of plaids for five example cells ordered by pattern index (a–e). The first column shows the response of a version of the V1 stage of the model, without normalization, equivalent to the square root of the response of an energy model. The 2nd column shows the response of the V1 stage when normalization is included. 3rd column: if V1 stage projected to MT stage using only excitatory weights.

(f) Plots illustrating the role of V1 normalization in the computation of pattern motion for the fifth example cell. The first panel shows the combined effects of the V1 tuning bandwidth and the distribution of MT excitatory weights, without normalization. The second panel demonstrates the effect of including the MT inhibitory weights. The third panel illustrates the distribution of the V1 normalization signal across the direction-interaction surface. The final panel shows the effect of adding the normalization signal at the V1 stage. This is like dividing the response due to convergent excitation and inhibition (panel 2) by the normalization signal (panel 3).

Only in this final panel is direction-invariant pattern selectivity seen.

Rust, Mante, Simoncelli & Movshon (2006). How MT cells analyze the motion of visual patterns. *Nat. Neurosci., 9*(11), 2006

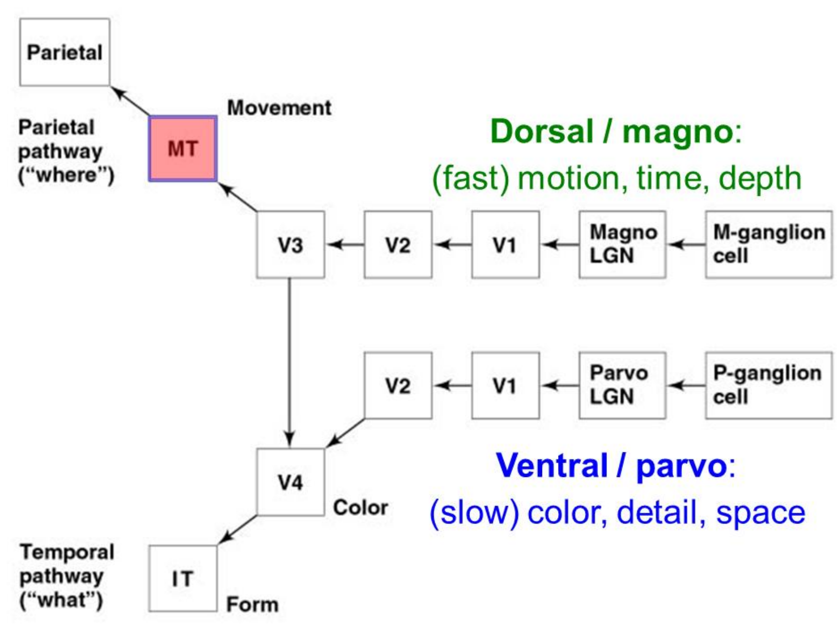
Directional preference across visual areas of the macaque



Comparison of neuronal selectivities across visual cortical areas. Estimates of the proportion of cells in different cortical areas selective for *A,* orientation; *B,* direction; *C,* binocular disparity; and *D,* color. Qualitative estimates indicated by *open symbols* and quantitative estimates by *solid symbols.* *Numbers* beside each *symbol* refer to individual published studies as indicated below.

I. Albright ('84)
2. Albright et al. ('84)
3. Anderson et al. ('82)
4. Baizer ('82)
5. Baizer et al. {'77)
6. Bullier and Henry ('80)
7. Burkhalter and Van Essen ('86)
R. De Valois et al. ('82)
9. DeYoe and Van Essen ('85)
10. Dow ('74)
II. Dow and Gouras ('73)
12. Felleman and Van Essen (present

20. Poggio ('72)
21. Poggio and Fischer ('77)
22. Poggio and Talbot ('81)
23. Poggio et al. ('75)
24. Poggio et al. ('77)
25. Poggio et al. ('85)
26. Schein et al. ('82)
27. Schiller et al. ('76)
28. Spinelli et al. ('70)
29. Van Essen and Zeki ('78)



Parietal

Movement

Parietal pathway ("where")

MT

V3 ← V2 ← V1 ← Magno LGN ← M-ganglion cell

V2 ← V1 ← Parvo LGN ← P-ganglion cell

V4

Color

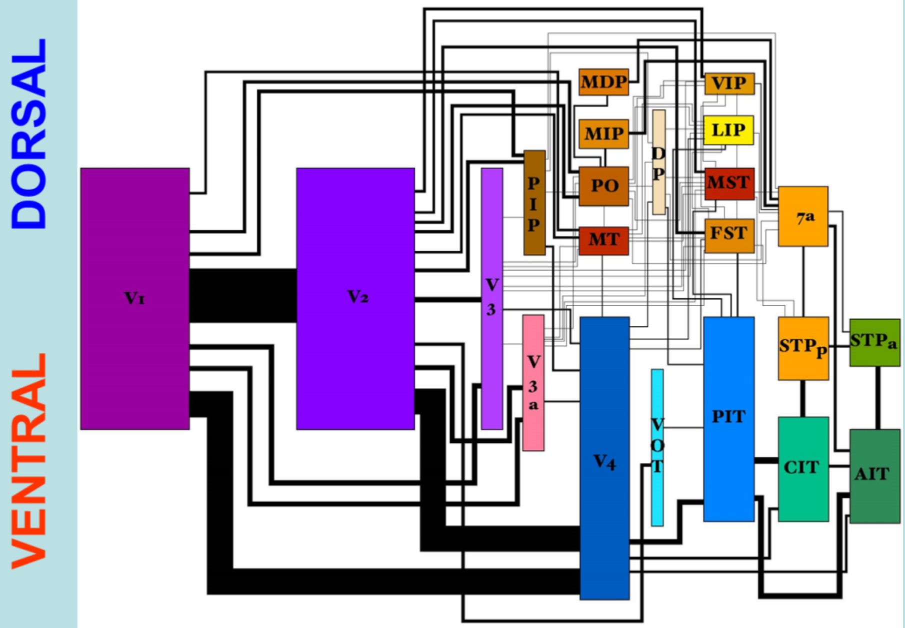
Temporal pathway ("what")

IT

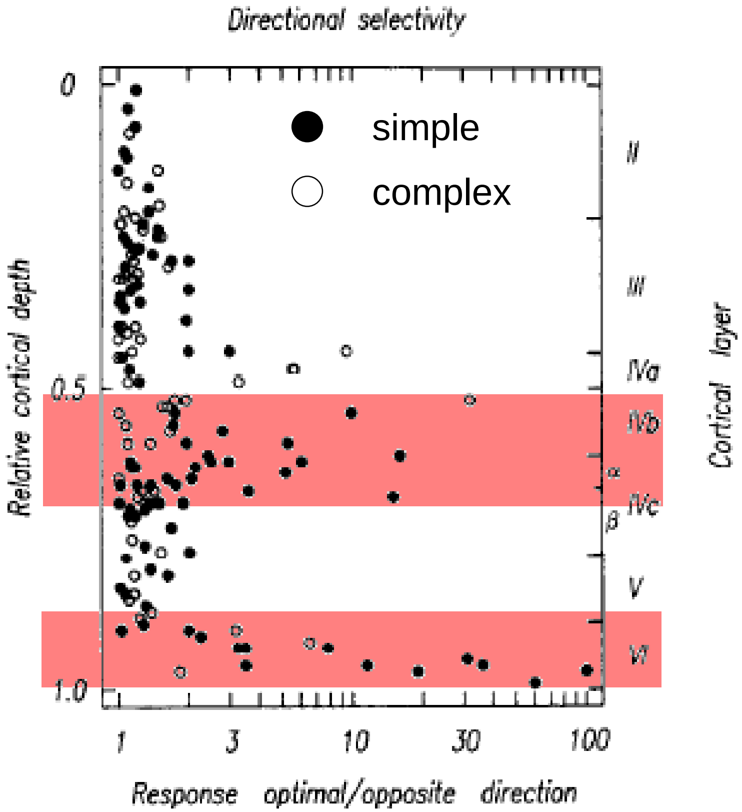
Form

Dorsal / magno:
(fast) motion, time, depth

Ventral / parvo:
(slow) color, detail, space



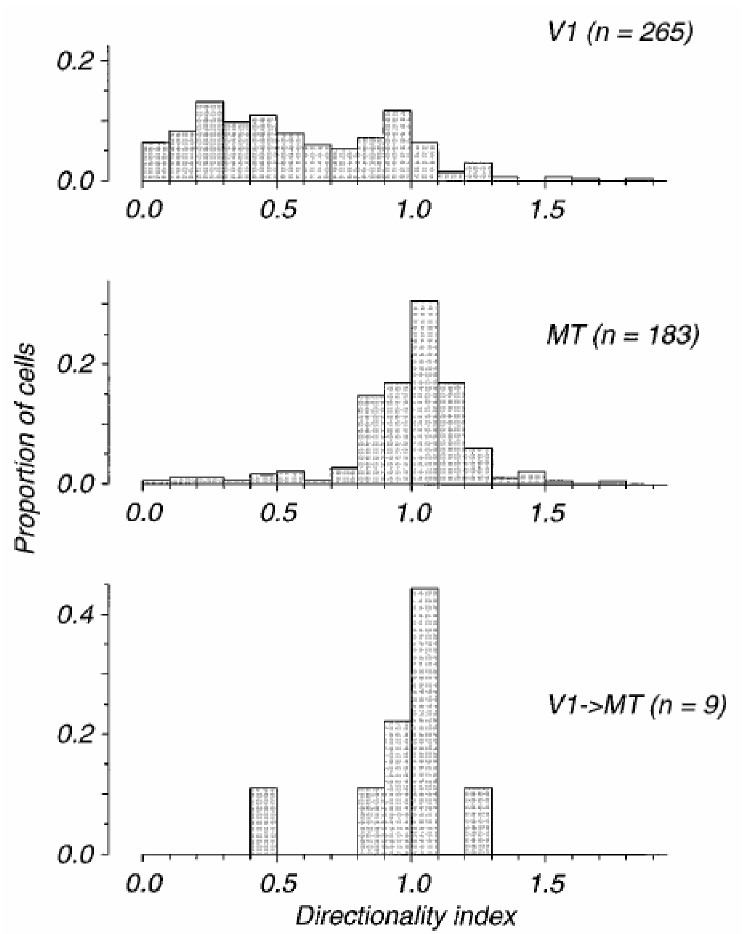
Directional preference of V1 macaque



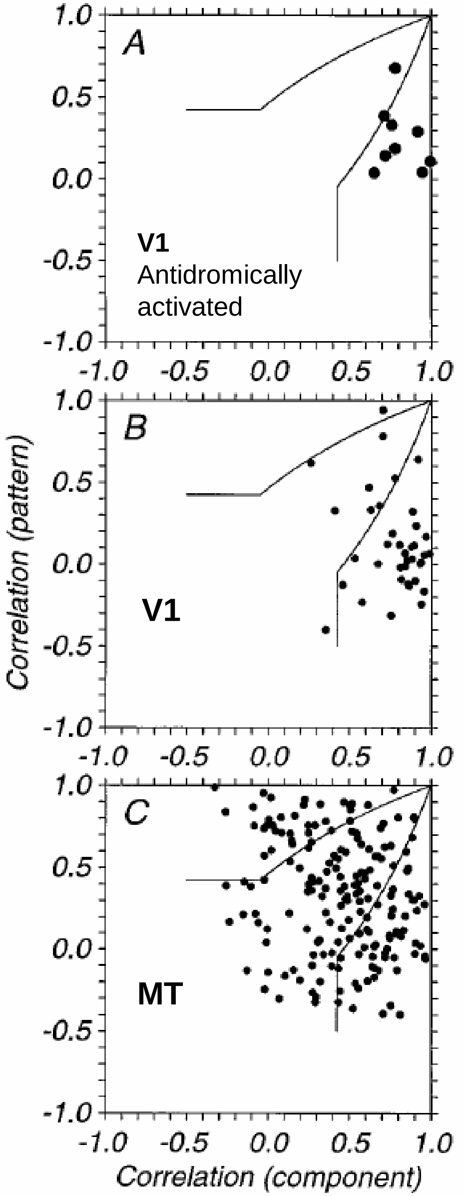
Directional preference of 147 V1 macaque cells is plotted as a function of laminar position. Directional preference calculated as ratio of peak response in the preferred direction to the response in the non-preferred direction. **Only cells in layer 6 and middle layer 4 show pronounced directional tuning**. Twelve of the cells included in this distribution were non-oriented. Open symbols, 54 complex cells plus 3 nonlinear, non-oriented cells; closed symbols, 81 simple cells plus 9 linear non-oriented cells.

Hawken, Parker & Lund (1988). *J. Neurosci.*, *8*(10), 3541-48.

Directional preference in V1 & MT areas of the macaque

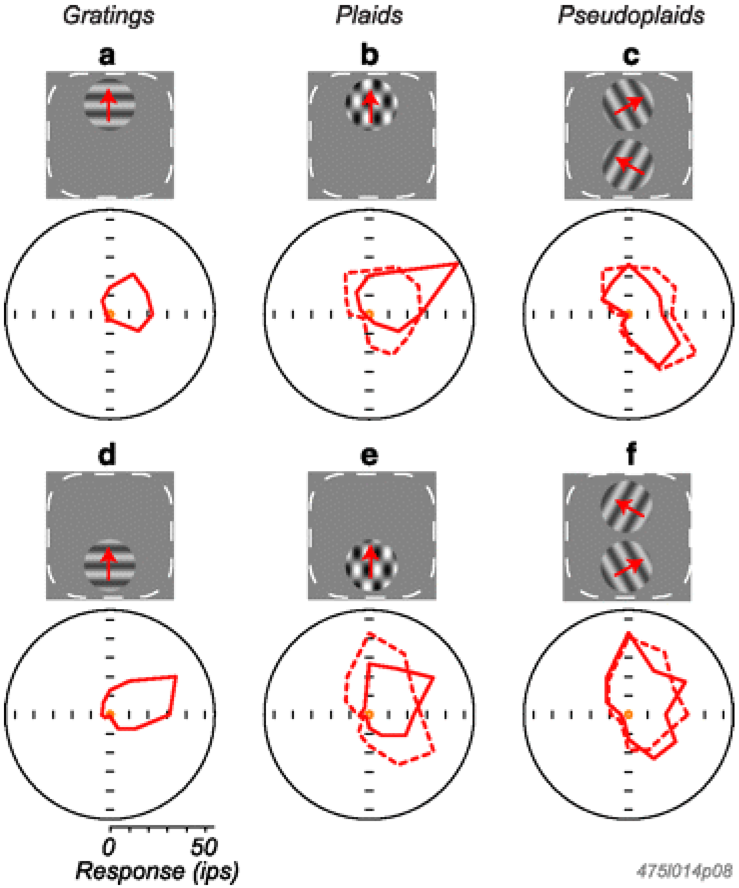


Frequency histograms of directionality indices for randomly sampled V1 neurons (*top*), randomly sampled MT neurons (*middle*), and V1 neurons antidromically activated from MT (*bottom*).

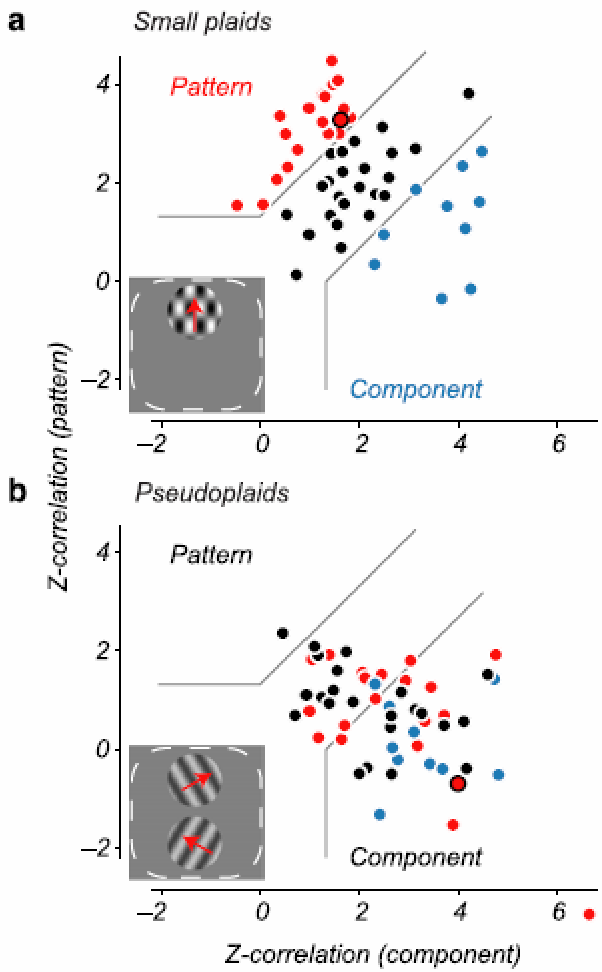


A. Scatterplot of the partial correlation coefficients calculated for each antidromically activated neuron. The abscissa shows the partial correlation between the data and the “component” prediction, whereas the ordinate shows the partial correlation between the data and the “pattern” prediction. The bullet-shaped contour divides this space into three regions of interest. Down and to the right is a region in which the correlation with the component prediction significantly exceeded the correlation with the pattern prediction or 0, whichever was larger. Neurons falling into this region most closely reflected the motion of the component gratings, and we consider such neurons to be component direction-selective. The converse relationship holds in the region up and to the left, and we consider neurons falling in this area to be pattern direction-selective. In between is a region in which cells cannot be classified as selective for either pattern or component motion. For cells in this region, neither correlation coefficient differed significantly from 0, or the two coefficients did not differ significantly from each other. For comparison, panels ***B*** and ***C***, shows similar scatterplots for populations of neurons randomly sampled from V1 and MT, respectively. Overall, the distribution from the antidromically activated neurons appeared to be indistinguishable from that observed for V1 neurons (Fig. 6*B*) and was obviously different from the much broader distribution observed for MT neurons (Fig. 6*C*).

Movshon J.A. & Newsome W.T. (**1996**). Visual Response Properties of Striate Cortical Neurons Projecting to Area MT in Macaque Monkeys, *J. Neurosci, 16*(23), 7733-41.



Responses of an MT cell to gratings, plaids, and pseudoplaids. Polar plots express cell response in spikes per second as distance from the origin, with the angle indicating the stimulus direction of motion. The small orange circles indicate spontaneous firing rate. ***a***, ***d***, Responses to a drifting grating covering one or the other of two patches within the receptive field, as indicated by the stimulus icons. ***b***, ***e***, Responses to the plaids created by summing two of the gratings tested in ***a***, with orientations differing by 120°. Dashed red curves indicate the predicted response of a component-direction-selective cell. ***c***, ***f***, Responses to pseudoplaids obtained with gratings in the two patches arranged with a direction difference of +120° (***c***) or –120° (***f***).



Summary of the results across the cell population. ***a***, The degree to which cells are selective for the direction of a whole pattern or of the individual components as determined with small plaid stimuli. The Z-transformed partial correlations between the data and the "component" and "pattern" models are plotted against one another. Gray lines separate regions within which cells are classified as pattern-direction selective or component-direction selective, according to a conservative statistical criterion. Cells classified as pattern selective are indicated in red, as component selective in blue– green, and as unclassed by this method in black. ***b***, A similar plot made from data taken from the same cells using pseudoplaids. The colors indicating the classification of the cells using small plaids are retained from ***a***.

Majaj, N. J., Carandini, M. a Movshon, J. A. (2007). Motion integration in Macaque MT is local, not global. *J. Neuroscience, 27*, 366-370.

THE APERTURE PROBLEM
and
VELOCITY COMPUTATION

The barber pole illusion

(Guilford, 1929)

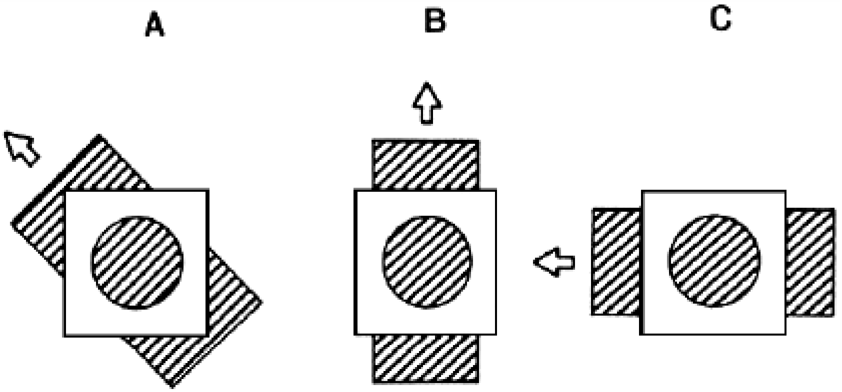


Fig. 1. Three different motions that produce the same physical stimulus.

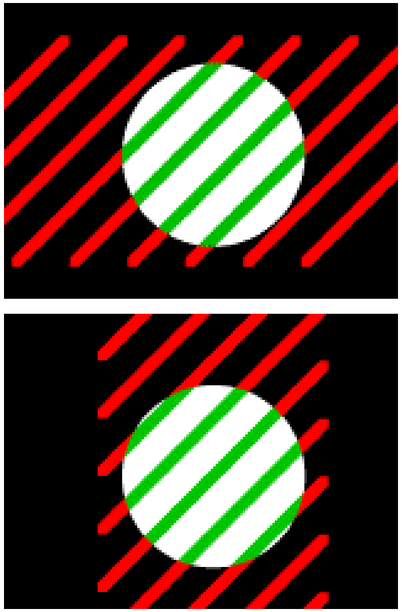
Movshon, Adelson, Gizzi & Newsome (1985).

http://en.wikipedia.org/wiki/Barberpole_illusion

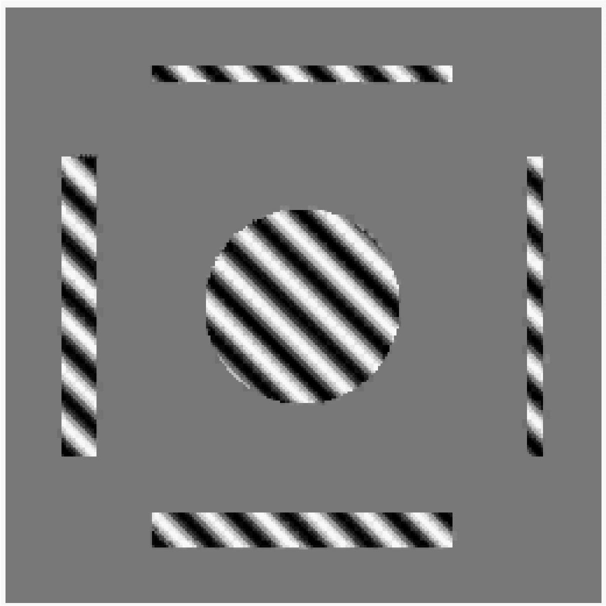
L

The aperture problem

(Wallach, 1935)

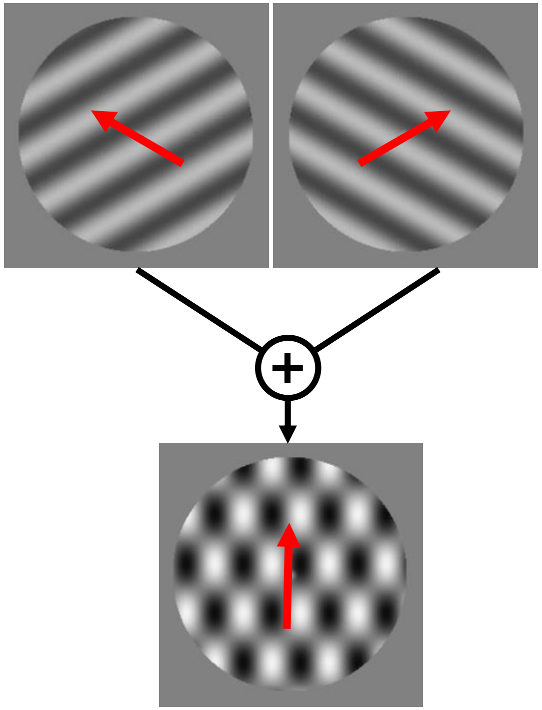


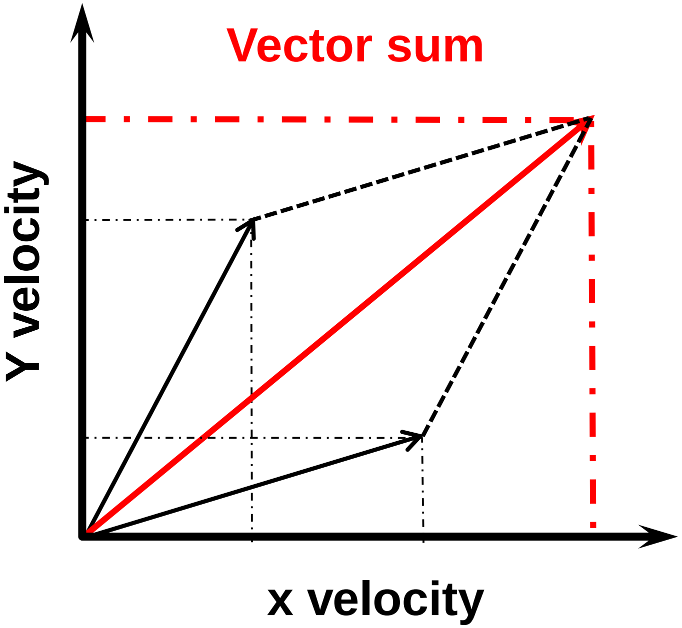
Barber Pole



The aperture problem

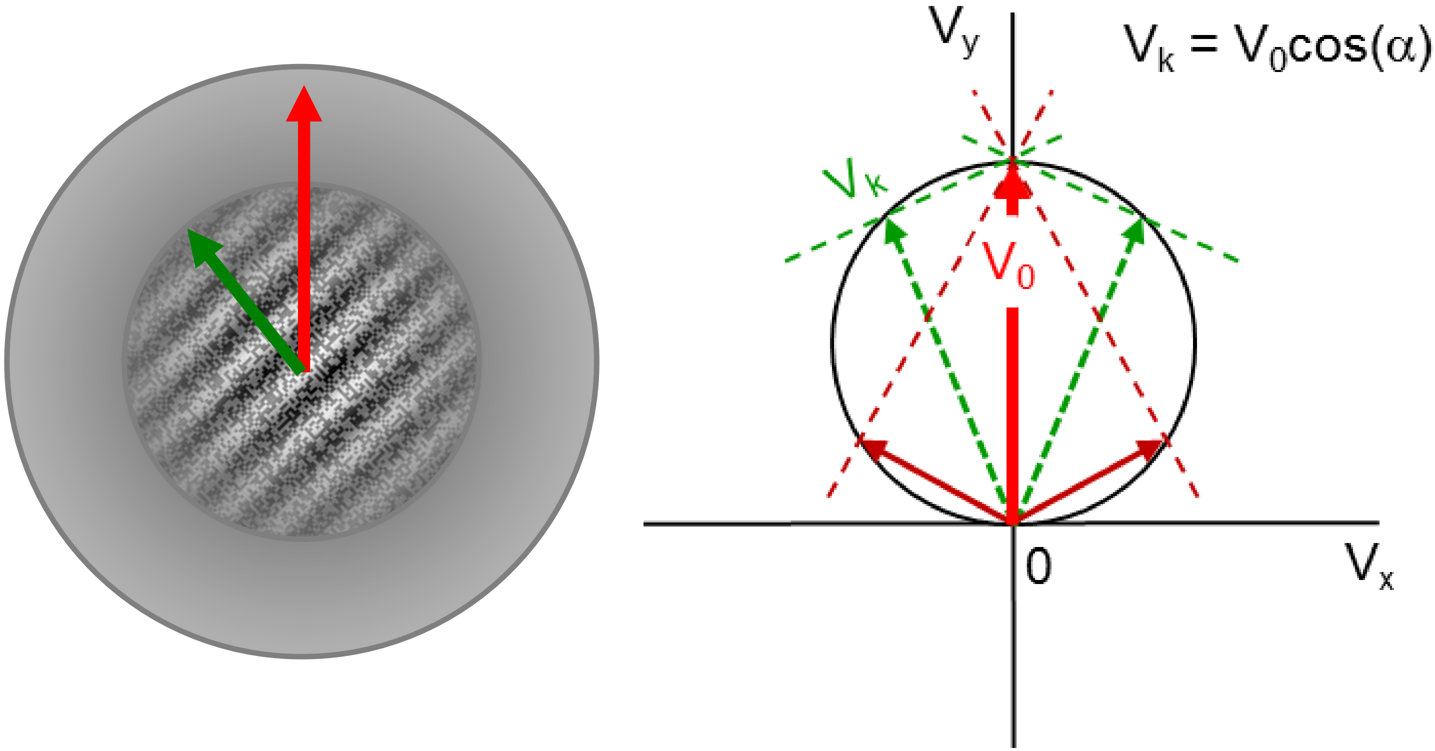
Vector sum





The aperture problem

Intersection of Constraints (***IOC***; Adelson & Movshon, 1982)



The velocity-space representation of a random-texture field moving to the right. The field contains components of all orientations, and the primary motion vector for each ends on the circle passing through the origin; the common motion 'implied' is the rightward motion given by the bold arrow This circular locus of primary motions represents all the motions that can exist in a single rightward-moving pattern. As such, the locus could represent the preferences of a family of primary motion analysers whose outputs are combined to signal coherent two-dimensional motion.

Adelson & Movshon, *Nature*, 1982

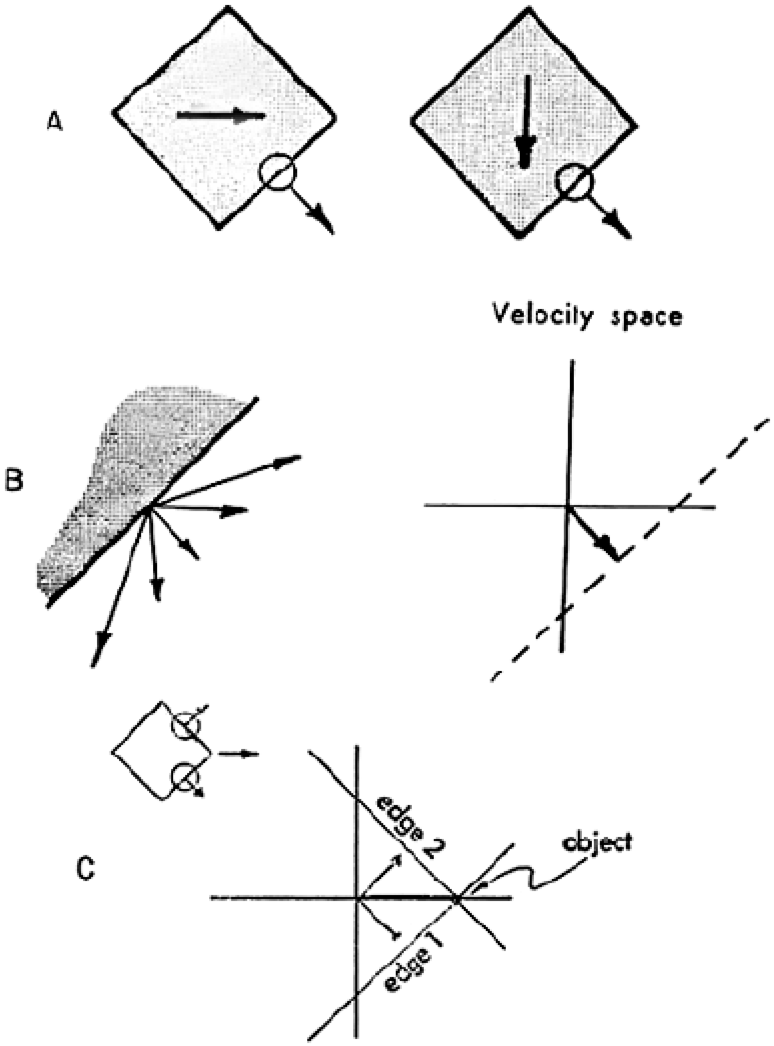


Fig. 2. A. Two moving diamonds. The local regions circled on each diamond's border have identical motions. B. A single moving contour, with the representation of its possible motions in a polar "velocity space", in which each vector represents a possible direction and speed. C. The solution to the ambiguity of one-dimensional motion based on an intersection of constraints. Each border's motion establishes a family of possible motions; the single intersection of these two families represents the only possible motion for a single object containing both contours.

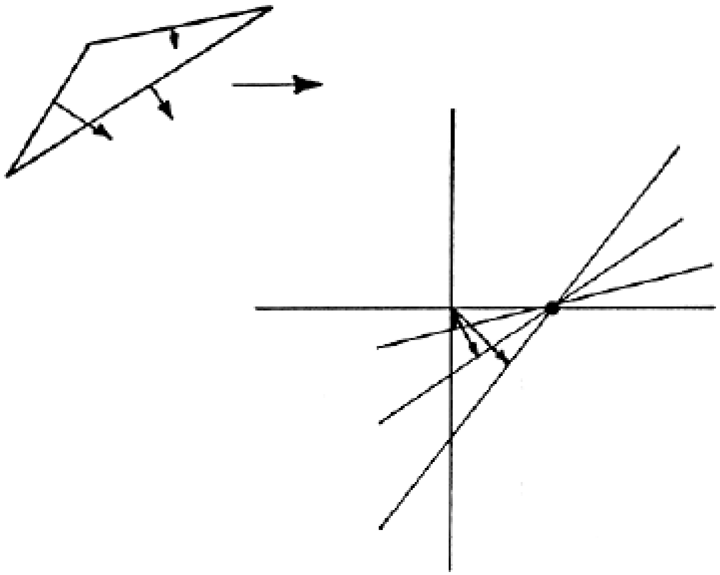
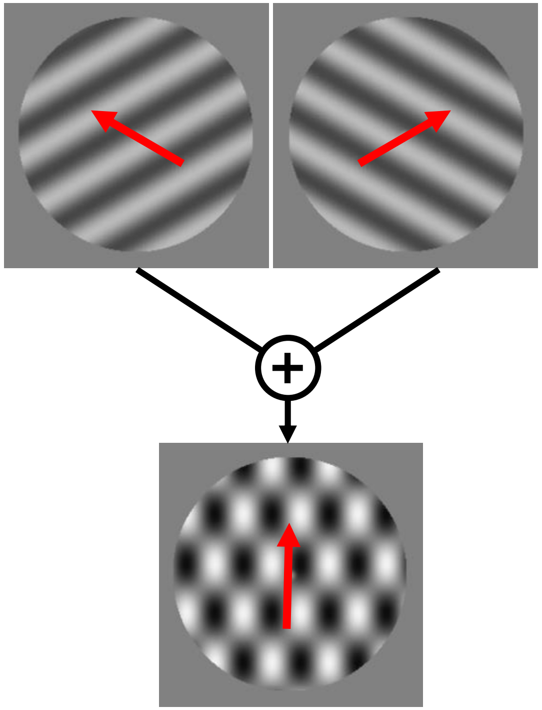


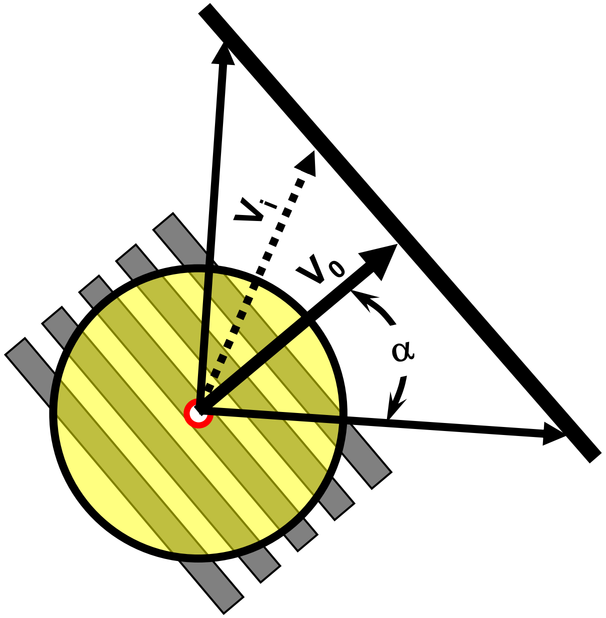
Fig. 3. An illustration of the inadequacy of vector summation as a solution to motion ambiguity. All three primary motion of the triangle's borders have a downward component but the true motion is directly to the right, as given by the intersection of constraints.

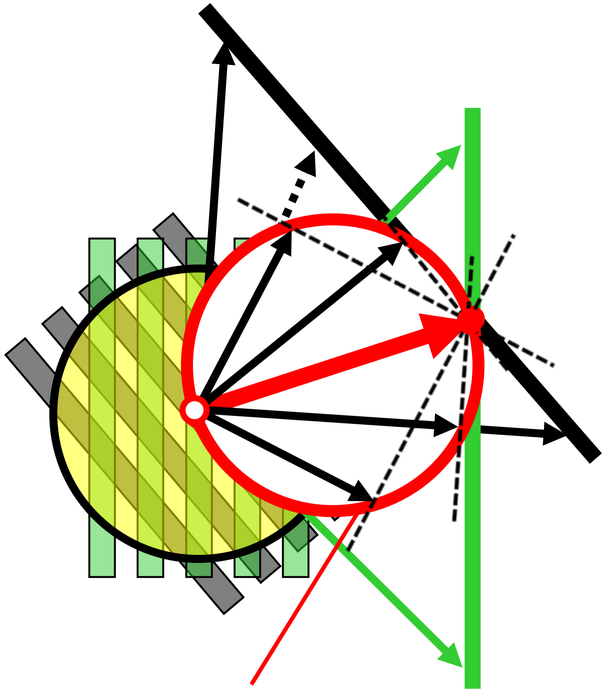
Movshon, Adelson, Gizzi & Newsome (1985).

The aperture problem

Intersection of Constraints (***IOC***; Adelson & Movshon, 1982)







Intersection of Constraints
IOC

VELOCITY COMPUTATION

Jump to MODEL

▶

Motion Energy

Velocity computation

Because of the *univariance principle* (the output of a velocity tuned neuron varies with both velocity **and contrast** – and also with SF and TF) and *the aperture problem* neither a Reichardt nor a motion energy unit will be able to code velocity by itself. Some kind of *population code* is required.

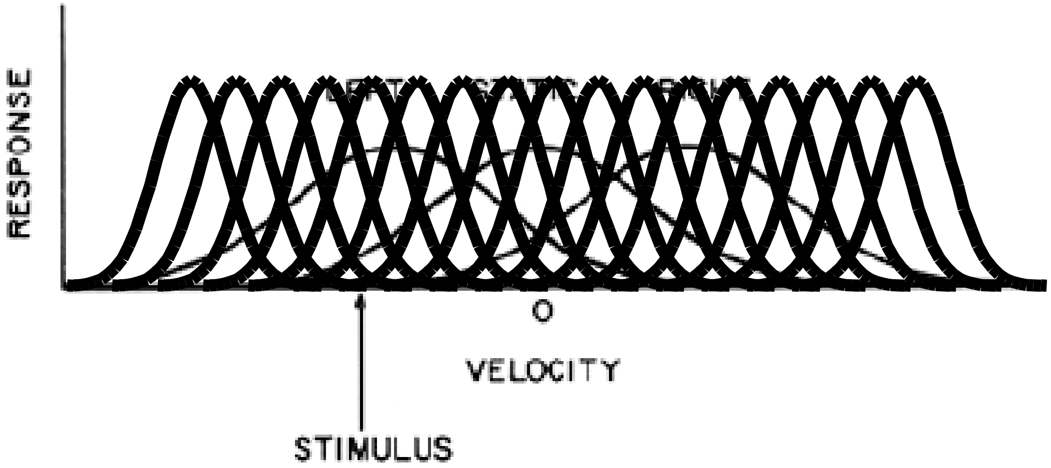
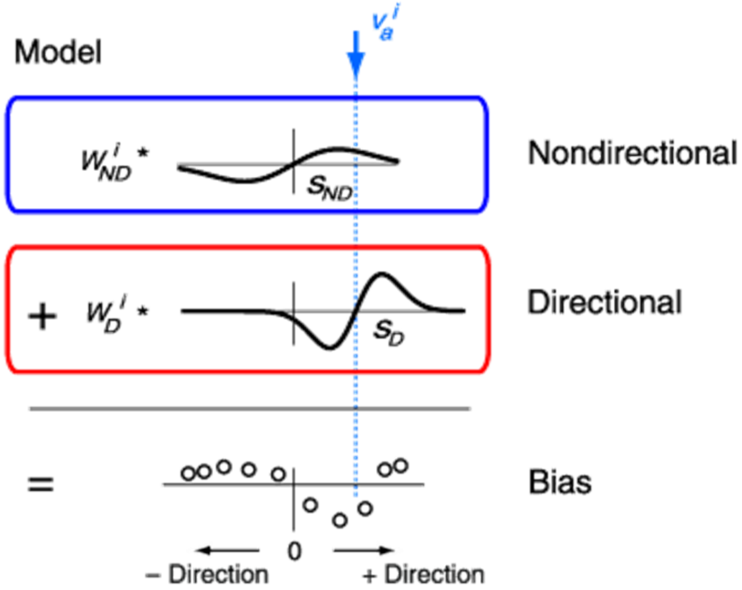
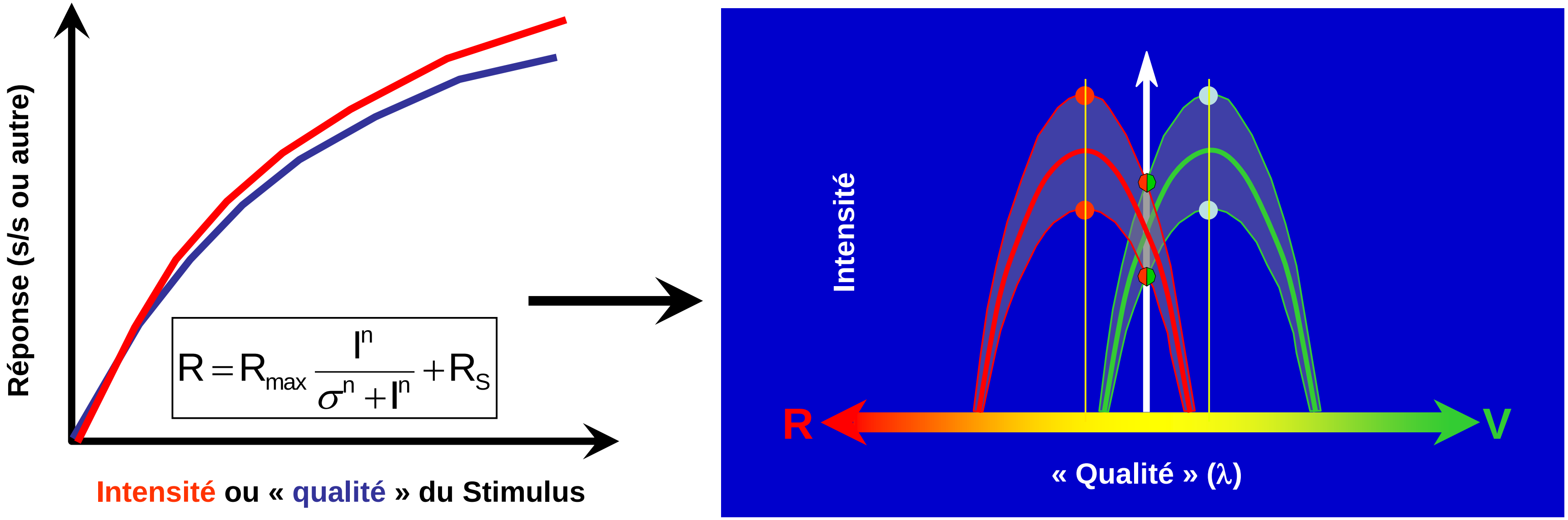


Fig 14. The overlapping response curves of three motion units plotted as a function of velocity. Any single unit's response is a function of both the velocity and the contrast of a stimulus. However, the relative responses of the various units can be used to compute a measure of velocity that is invariant with contrast.



Adelson, E. H. & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A* 2, 284-299.

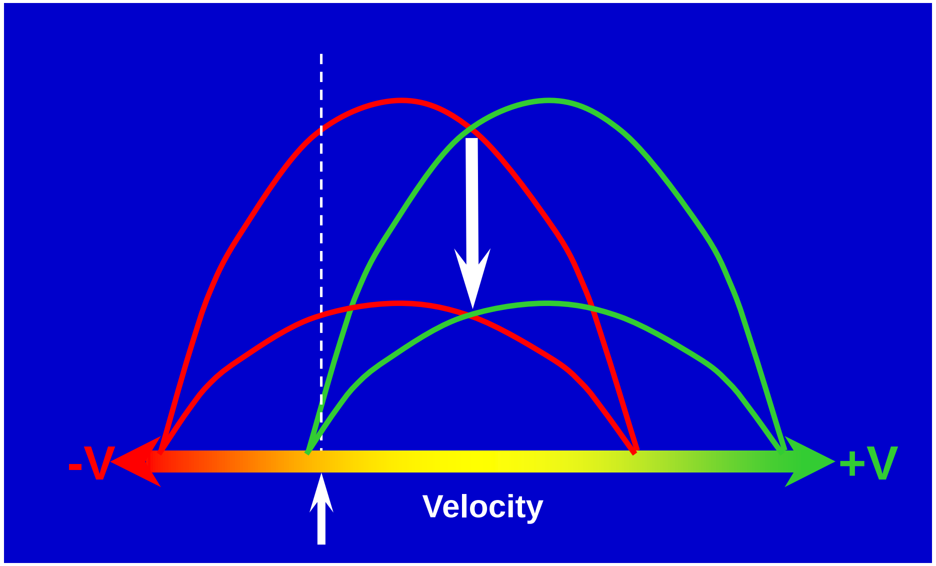
Le principe de l'UNIVARIANCE



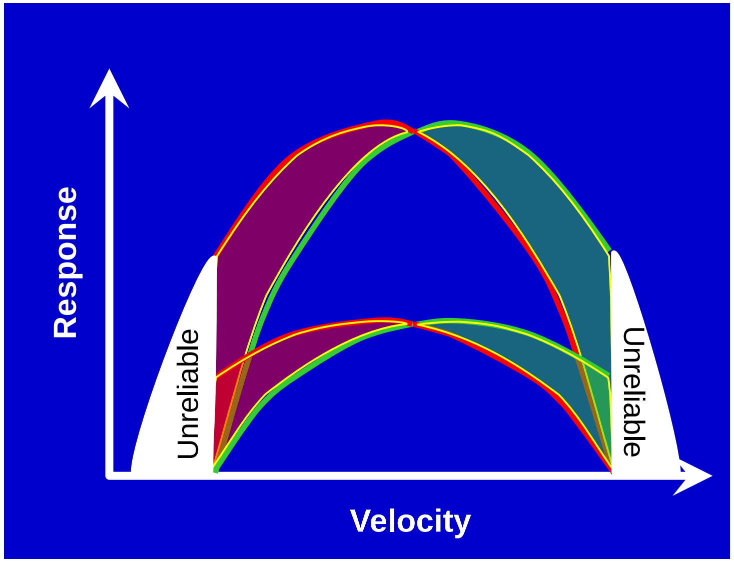
$$R = R_{max} \frac{I^n}{\sigma^n + I^n} + R_S$$

I = stimulus intensity
R = response
R_{max} = max resp.
N = max slope
S = semi-saturation cst.
R_s = spontaneous R

Tuning functions



Response functions & response ratios



At high velocities or low contrasts, the denominator in the ratio can become quite small, and so the velocity estimate will blow up or become unreliable. The visual system must have some means of tagging the velocity estimate with a confidence measure; the simplest approach would be to use the output of the static channel as it stands. High velocities or low contrasts would then lead to low-confidence measures for the velocity of the pattern. When this information is combined with motion information from other channels, its low confidence would cause it to receive a relatively low weight in determining the final motion percept. (Adelson & Bergen, 1985, p 293-294)

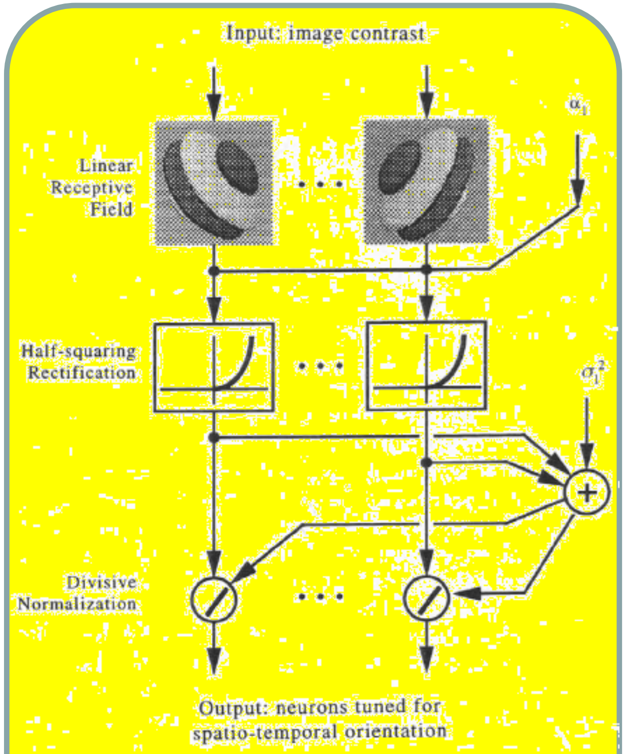


FIGURE 1. Model of V1 simple cells. Each neuron computes a weighted sum of its inputs followed by halfwave rectification, squaring, and response normalization. The underlying linear receptive fields (depicted as monochrome images) are localized in space and time, and are tuned for spatio-temporal orientation. An additive constant, α_1, is included in the summation, allowing for a spontaneous firing rate. The divisive normalization factor is computed as a sum of half-squared responses and a squared semi-saturation constant, σ_1.

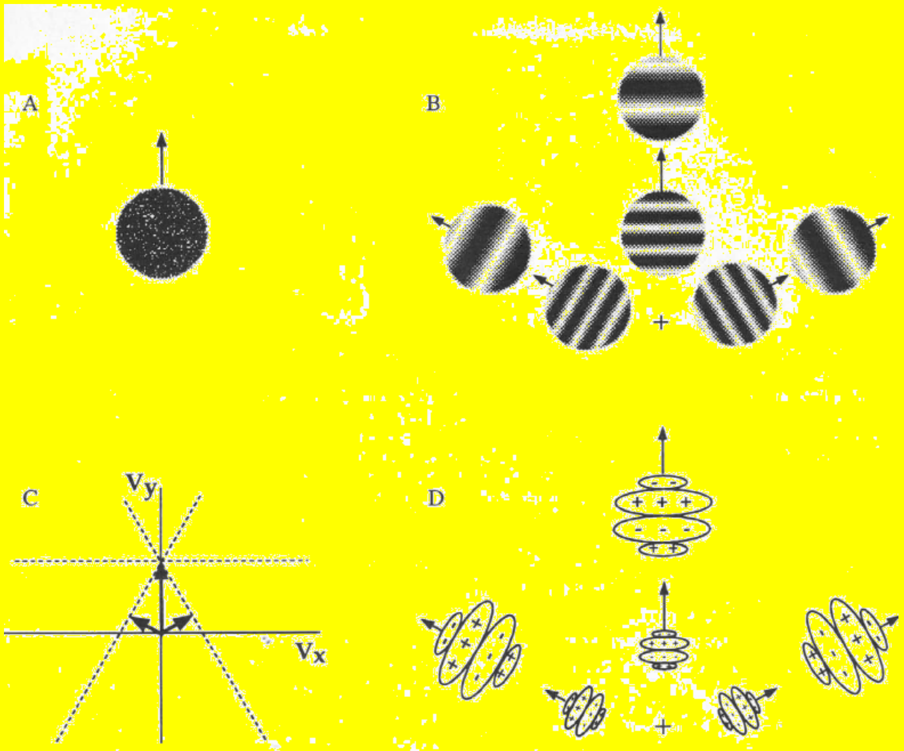
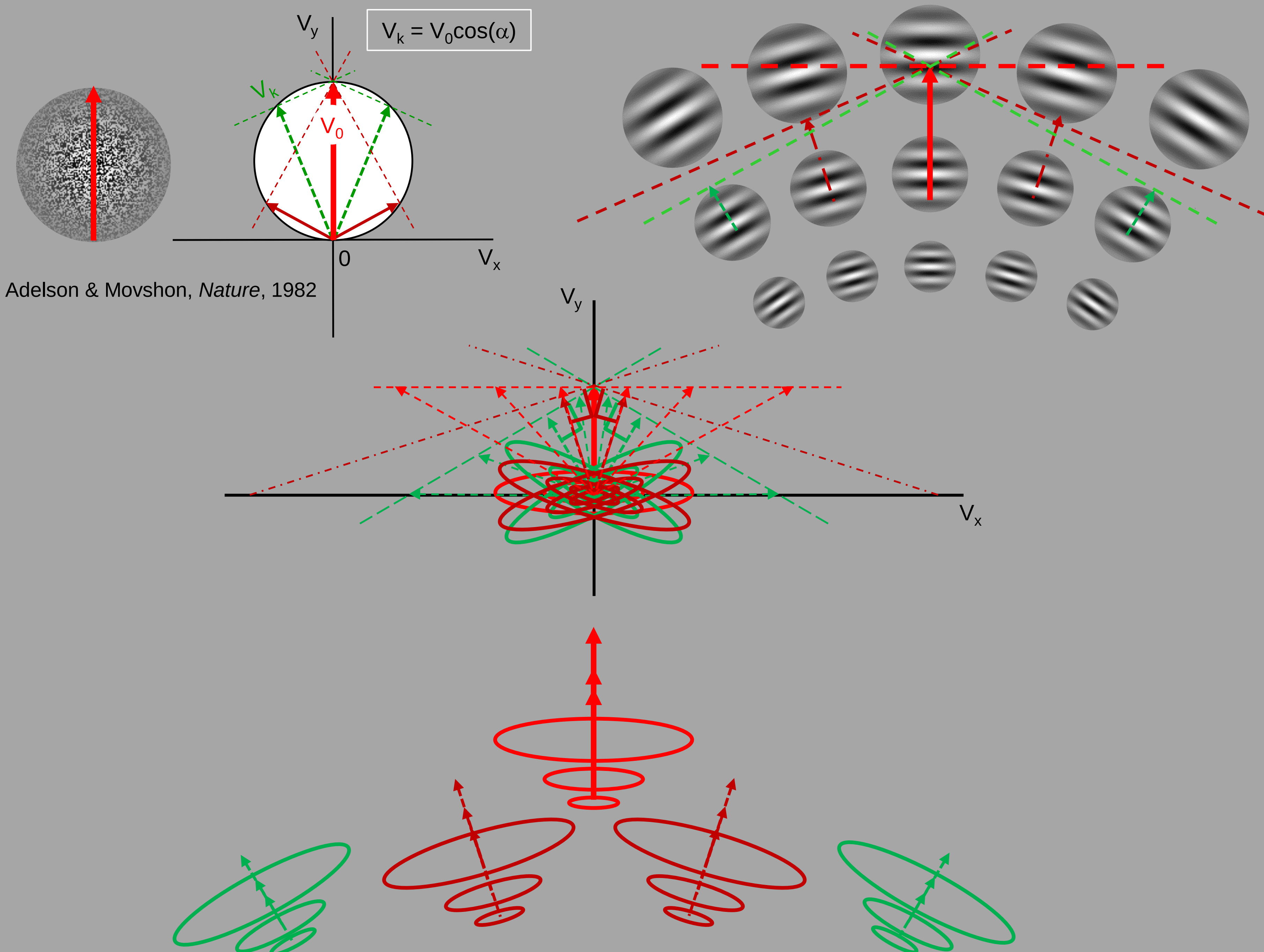


FIGURE 2. Construction of MT pattern cell velocity selectivity via combination of V1 complex cell afferents. (A) Random dot field stimulus, drifting upward. (B) Fourier decomposition of the dot stimulus. The stimulus is written as a sum of drifting sinusoidal components of appropriate orientation and normal velocity. A small subset of these are shown. (C) Intersection of constraints (IOC) construction. **The motion of a grating is ambiguous, since the component of velocity along the grating stripes produces no change in image intensity.** Each arrow corresponds to the normal component of velocity for two of the gratings shown in (B), and the dashed lines indicate the set of velocities consistent with the motions of those gratings. **The intersection point of these constraint lines is the only velocity consistent with the motion of *all* of the components, and corresponds to the velocity of the dot stimulus.** (D) Set of V1 complex receptive fields selective for each of the components shown in (B). The summed responses of such V1 neurons yield a pattern MT response that is selective for this stimulus velocity.

Spikes are positive

& cross-orientation inhibition

Simoncelli, E.P. & Heeger, D.J (1998) *Vision Res., 38*(5), 743-761.

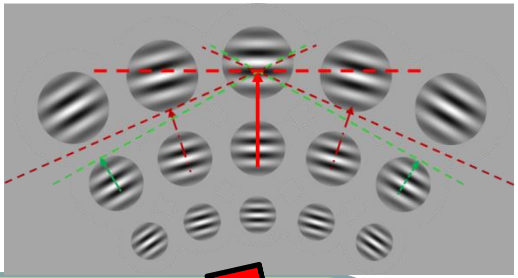


$V_k = V_0\cos(\alpha)$

Adelson & Movshon, *Nature*, 1982

Motion Energy

Velocity computation



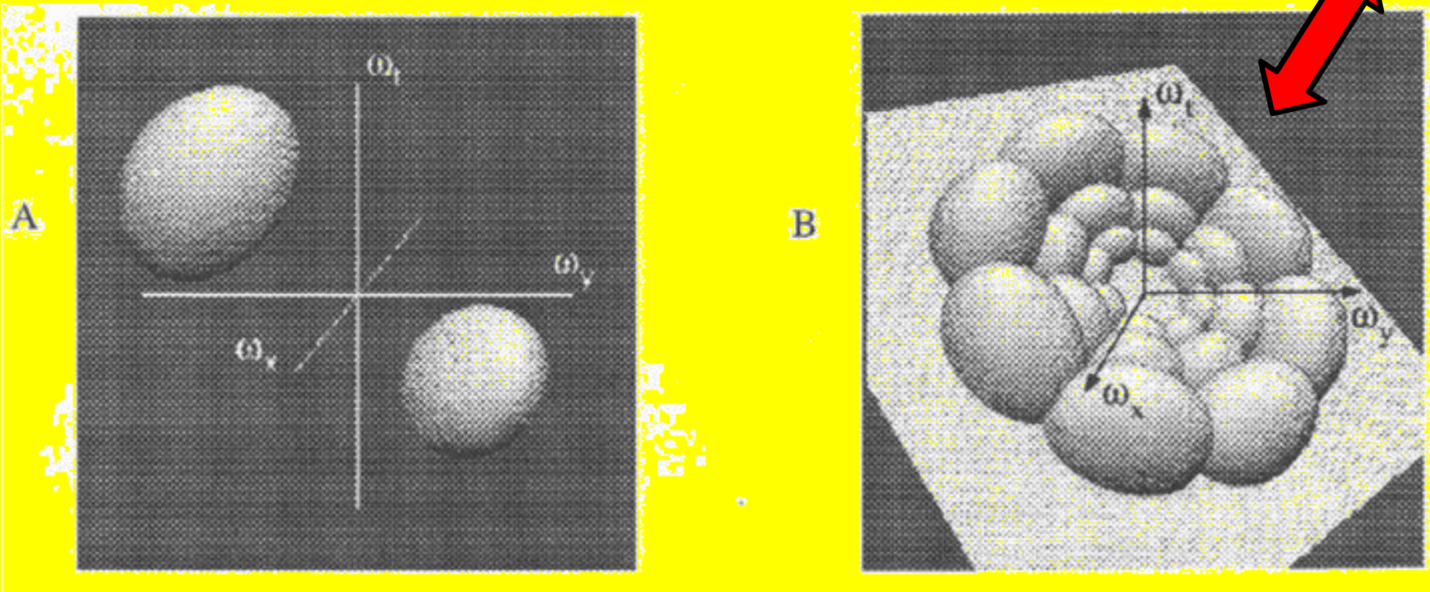


FIGURE 3. Construction of MT pattern cell velocity selectivity via combination of V1 complex cell afferents, shown in the Fourier domain. (A) Selectivity of a V1 neuron corresponds to a pair of localized spatio-temporal frequency bands, symmetrically arranged about the origin. (B) Selectivities of V1 neurons tuned for four orientations and three spatial scales, each consistent with a common velocity. The illustrated plane corresponds to the power spectrum of a stimulus moving at this common velocity. Responses of these V1 neurons are summed using positive (excitatory) weights to yield an MT response selective for this velocity. Not shown are a set of V1 neurons whose tuning bands lie off of the plane: these are combined using negative (inhibitory) weights. Also not indicated is the fact that the summation is performed over V1 neurons with receptive fields distributed over a local spatial region.

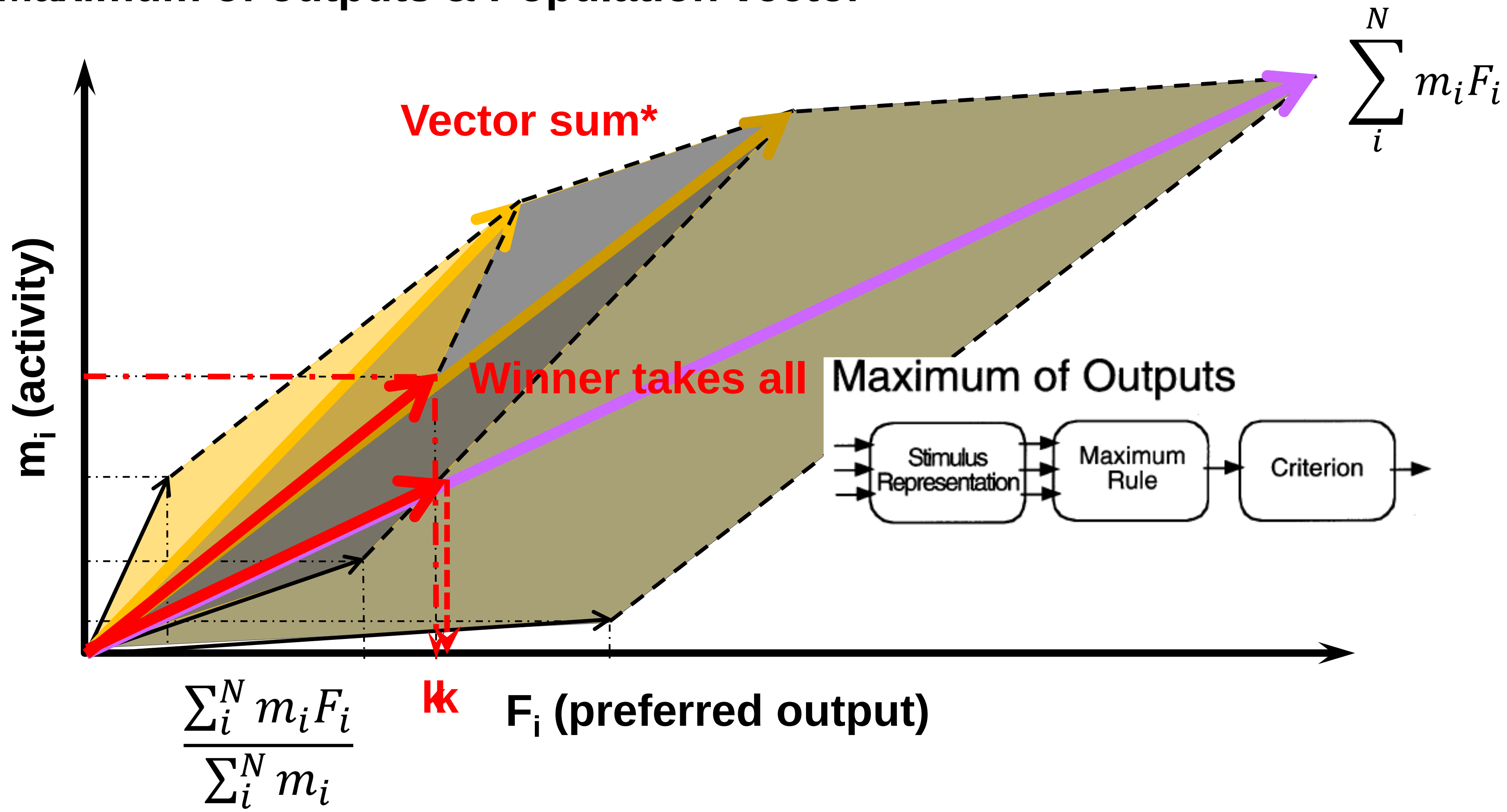
Simoncelli, E.P. & Heeger, D.J (1998) *Vision Res., 38*(5), 743-761.

Jump to the BINDING PROBLEM

▶

POPULATION CODING/DECODING

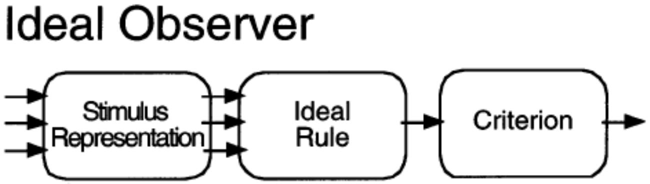
Maximum of outputs & Population vector



Drawbacks to population vector:

- It computes just *one single estimate* of the stimulus (rather than the full likelihood function);
- It ignores the contribution of the *full tuning properties* of the neurons

*Georgopoulos, Kalaska, Caminiti, & Massey (1982). *J. Neurosci. 2*, 1527–1537.



INTERPRETING POPULATION RESPONSE

State of affairs: *Neurons are broadly tuned; their response is noisy*

➔ A given stimulus evokes a noisy population response, i.e. no. of spikes for each neuron, n_i.

The task: infer the stimulus θ_0 given R.

Question

how likely it is that each possible stimulus θ elicited the observed response, n_i, i.e. what is $p(\theta \mid n_i)$?

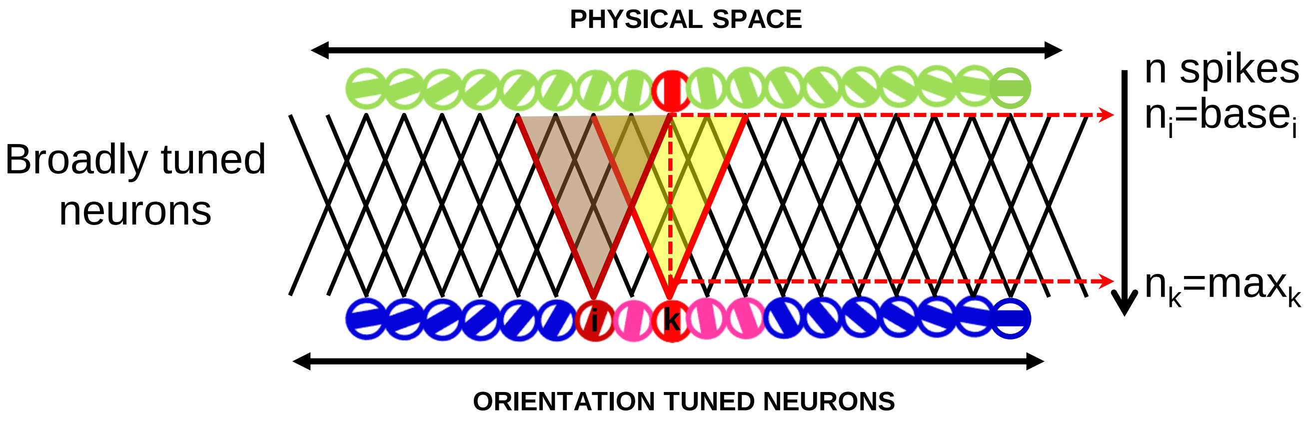
Strategy:

(1) determine for each neuron tuned to θ_i the likelihood of its observed response n_i given θ, i.e. $p(n_i \mid \theta)$

(2) Repeat the procedure for all θ for the observed population response,
➔ the likelihood function.

…Some mathematics…

Jazayeri, M. & Movshon, J.A. (2006). *Nat. Neurosci., 9*, 690-696.



Q: How likely it is that each stimulus (θ) elicited the observed population response (p($\theta_i \backslash n_i$)?

1. "Ask" each neuron i the *likelihood* that its response, n_i, was elicited by that stimulus, i.e. p($n_i \backslash \theta$) ➔ the likelihood function of each neuron, $L_i(\theta)$.

   1a. n_i is a stochastic variable and its mean, given a stimulus θ_0, is $f_i(\theta_0)$ where $f_i(\theta)$ is the tuning function of the neuron.

   1b. The likelihood of stimulus θ_0 is but the probability that this neuron would fire n_i spikes in response to that stimulus, given its tuning function, i.e. $L_i(\theta_0) = p(n_i \backslash \theta_0)$

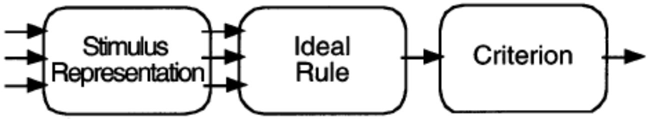
2. Combine the likelihoods over all neurons and all possible θ to determine the overall likelihood of each possible θ.

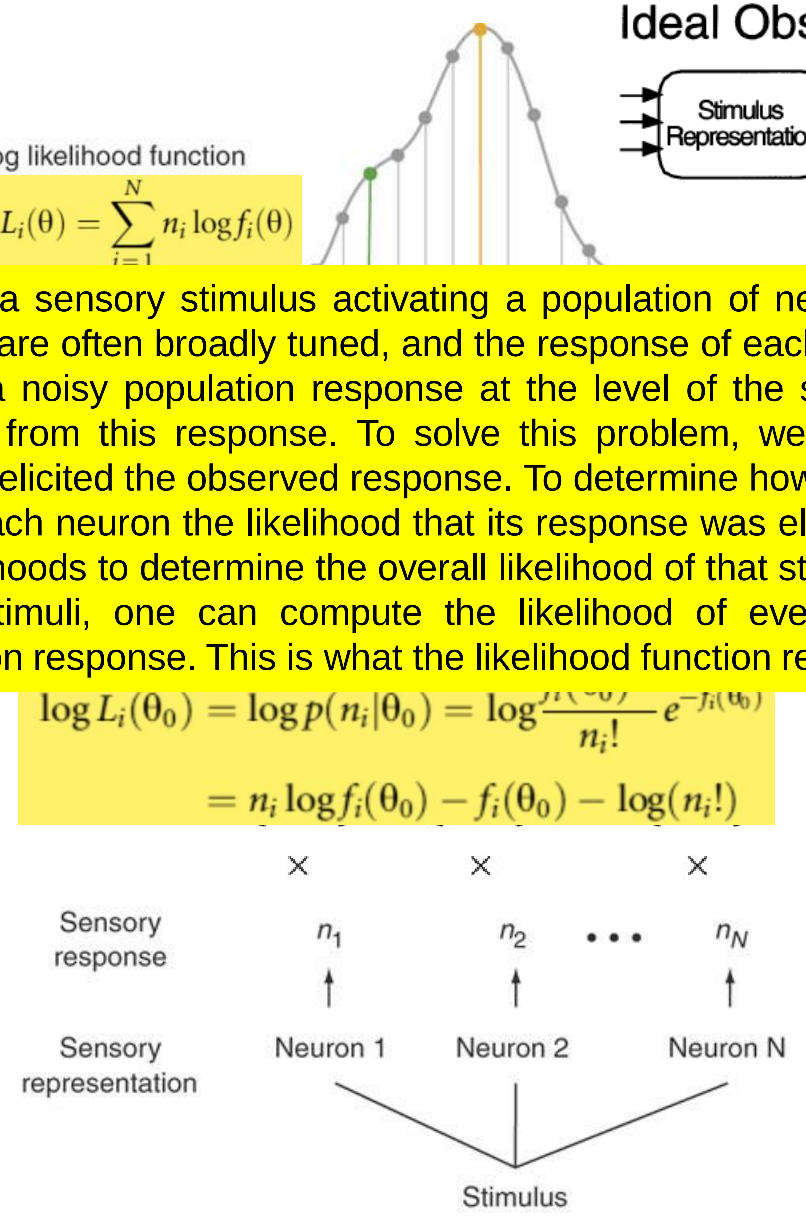
   2a. For each θ compute the joint probability that each neuron i responds according to its tuning function $f_i(\theta)$, i.e. $p(n_1 \backslash \theta) \times p(n_2 \backslash \theta) \times \ldots \times p(n_n \backslash \theta)$;

   2.b If the likelihood is computed in log space, then the multiplication transforms into an arithmetic sum,

INTERPRETING POPULATION RESPONSE

Ideal Observer





$$\log L(\theta) = \sum_{i=1}^{N} \log L_i(\theta) = \sum_{i=1}^{N} n_i \log f_i(\theta)$$

Imagine a sensory stimulus activating a population of neurons in a cortical sensory area. These neurons are often broadly tuned, and the response of each one is noisy. As a result, every stimulus evokes a noisy population response at the level of the sensory neurons; the task is to infer the stimulus from this response. To solve this problem, we ask how likely it is that each possible stimulus elicited the observed response. To determine how likely a given stimulus is, one strategy is to ask each neuron the likelihood that its response was elicited by that stimulus, and then combine the likelihoods to determine the overall likelihood of that stimulus. By repeating the same procedure for all stimuli, one can compute the likelihood of every stimulus for the particular observed population response. This is what the likelihood function represents.

log[p

Weights

×

n_i

$$\log L_i(\theta_0) = \log p(n_i|\theta_0) = \log\frac{f_i(\theta_0)^{n_i}}{n_i!}e^{-f_i(\theta_0)}$$

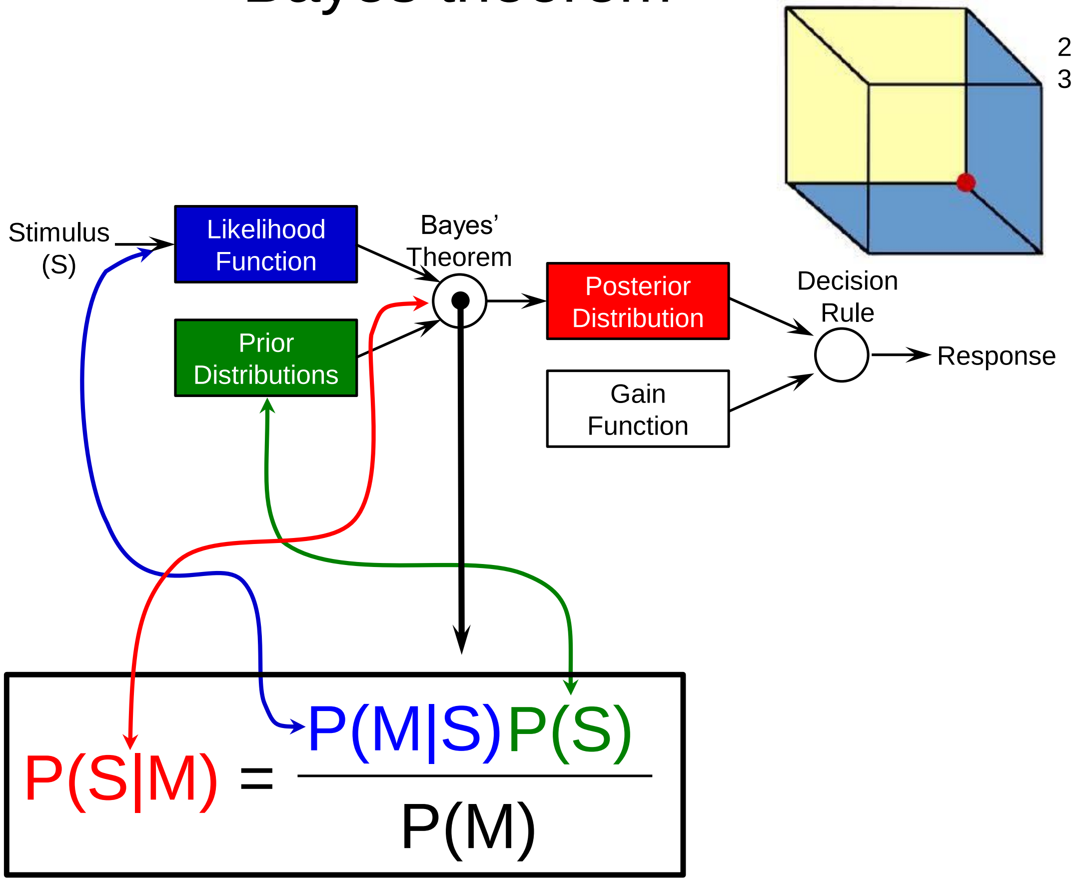
$$= n_i \log f_i(\theta_0) - f_i(\theta_0) - \log(n_i!)$$

Figure 1. Computing the log likelihood function in a *feedforward* function. Adding the contribution of individual neurons (shown for two example stimulus values in orange and green) gives the overall log likelihood function, log $L(\theta)$ for all values of θ that could have elicited this pattern of responses. Here, the orange point at the peak of the log likelihood function indicates the most likely stimulus.

Jazayeri & Movshon (2006). *Nat. Neurosci.*, *9*, 690-696.

BAYESIAN MOTION

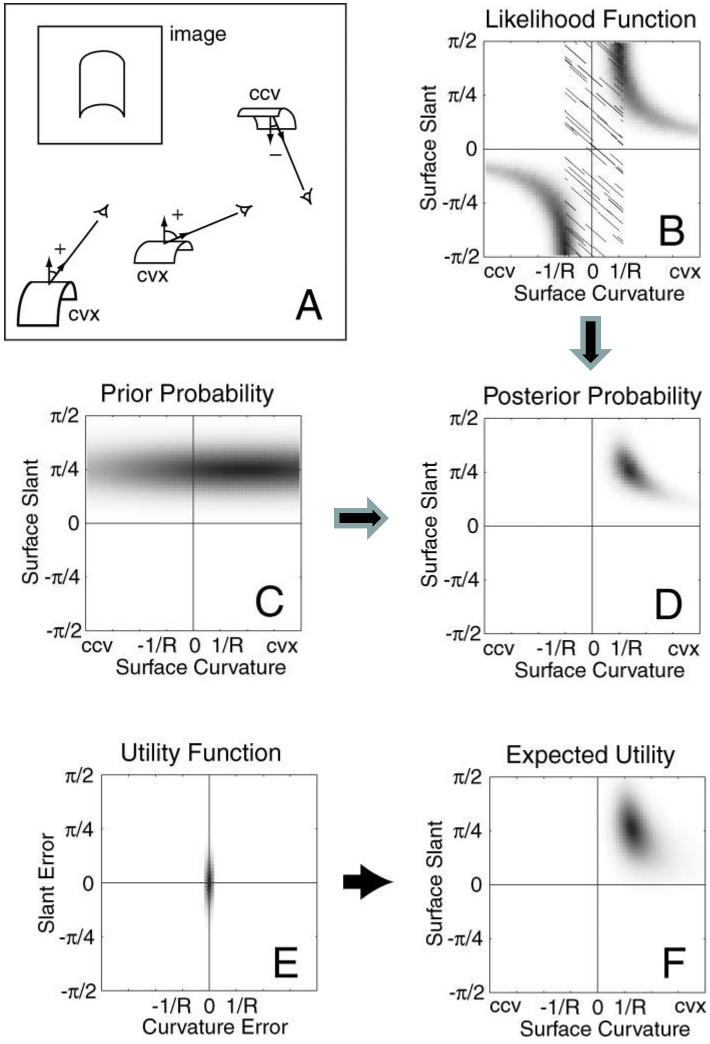
Bayes theorem



$$P(S|M) = \frac{P(M|S)P(S)}{P(M)}$$

Possible interpretations:

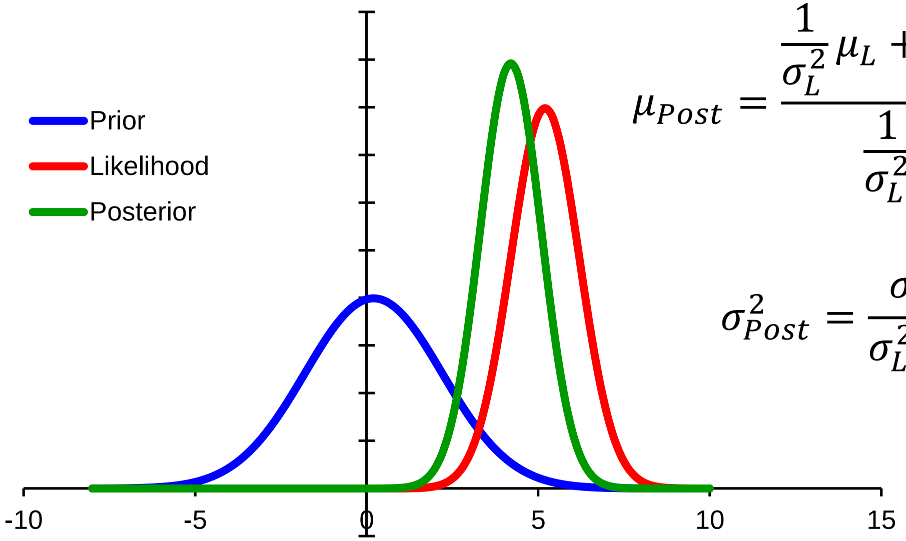
1. highly curved cylinder from a high angle of view,
2. a flatter cylinder from a low angle of view,
3. a concave cylinder from yet another viewpoint.



M→ measure or neural response

Kersten, Mamassian & Yuille, 2004

Bayesian combination rule of two random variables

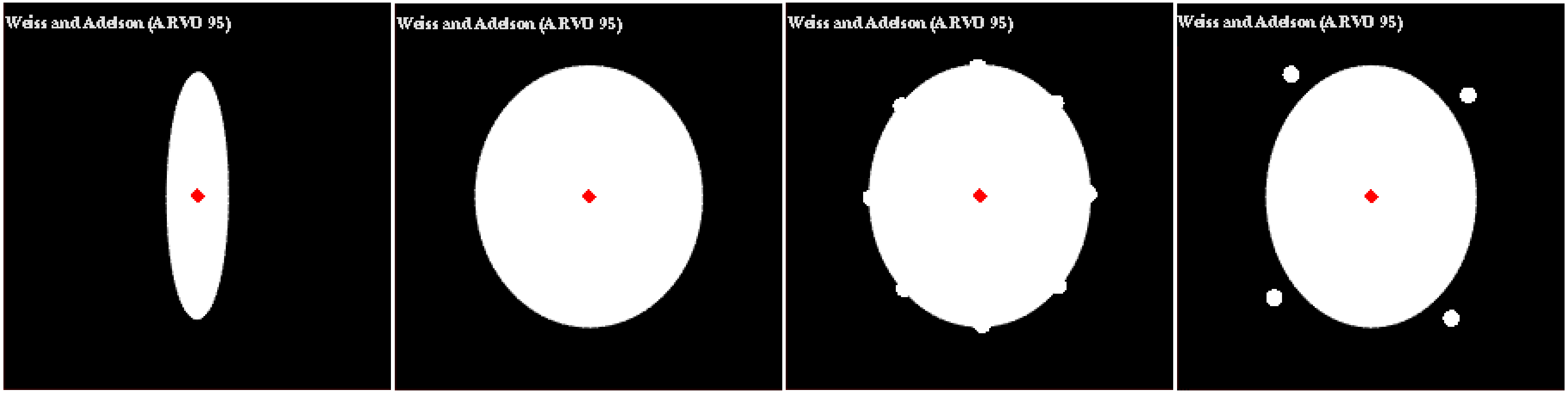


$$\mu_{Post} = \frac{\frac{1}{\sigma_L^2}\mu_L + \frac{1}{\sigma_{Prior}^2}\mu_{Prior}}{\frac{1}{\sigma_L^2} + \frac{1}{\sigma_{Prior}^2}}$$

$$\sigma_{Post}^2 = \frac{\sigma_L^2 \sigma_{Prior}^2}{\sigma_L^2 + \sigma_{Prior}^2}$$

Motion illusions as optimal percepts

(Weiss, Simoncelli & Adelson, *Nature Neuroscience* 2002)

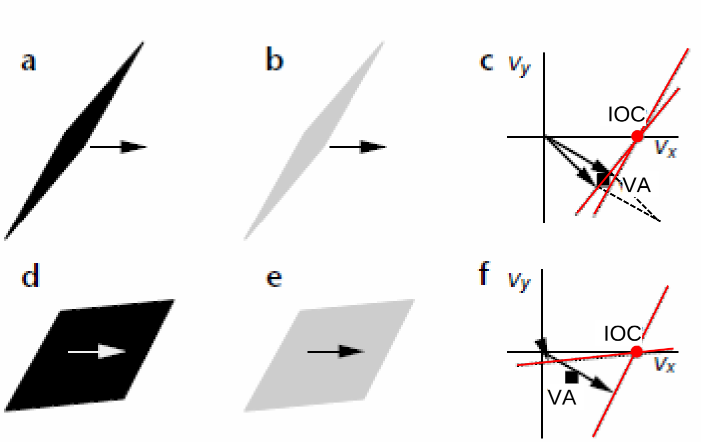


The pattern of local image velocities on the retina encodes important environmental information. Although humans are generally able to extract this information, they can easily be deceived into seeing incorrect velocities. We show that these ‘illusions’ arise naturally in a system that attempts to estimate local image velocity. We formulated a model of visual motion perception using standard estimation theory, under the assumptions that (i) there is noise in the initial measurements and (ii) slower motions are more likely to occur than faster ones.

http://www.cs.huji.ac.il/~yweiss/Rhombus/rhombus.html

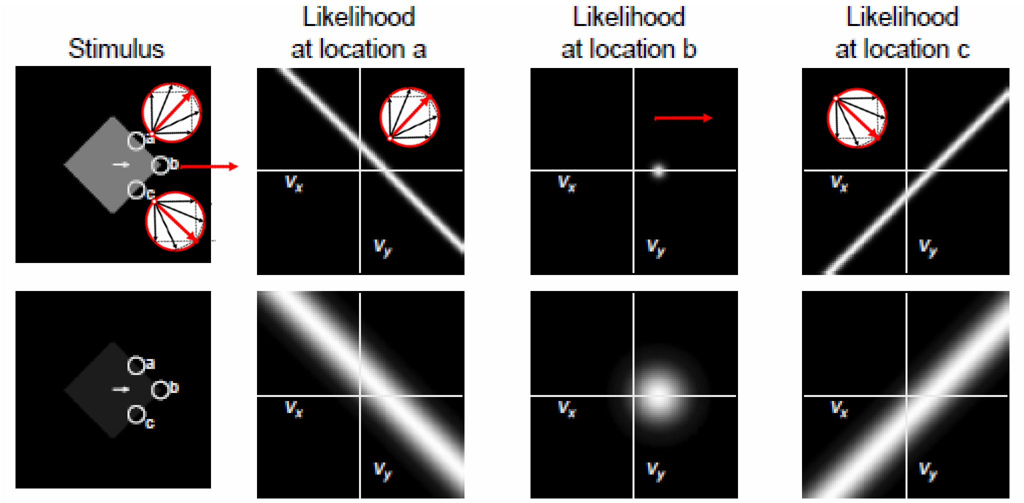
Motion illusions as optimal percepts

Intersection of Constraints & Priors (Weiss, Simoncelli & Adelson, 2002)



Insufficiency of either Vector Average, IOC or Feature Tracking rules as an explanation for human perception of a horizontally moving rhombus. (**a**) A ‘narrow’ rhombus at high contrast appears to move horizontally (consistent with IOC/FT). (**b**) A narrow rhombus at low contrast appears to move diagonally (consistent with VA). (**c**) Velocity space constraints for a narrow rhombus. (**d**,**e**) A ‘fat’ rhombus at low or high contrast appears to move horizontally (consistent with IOC/FT). (**f**) Velocity space constraints for a fat rhombus.

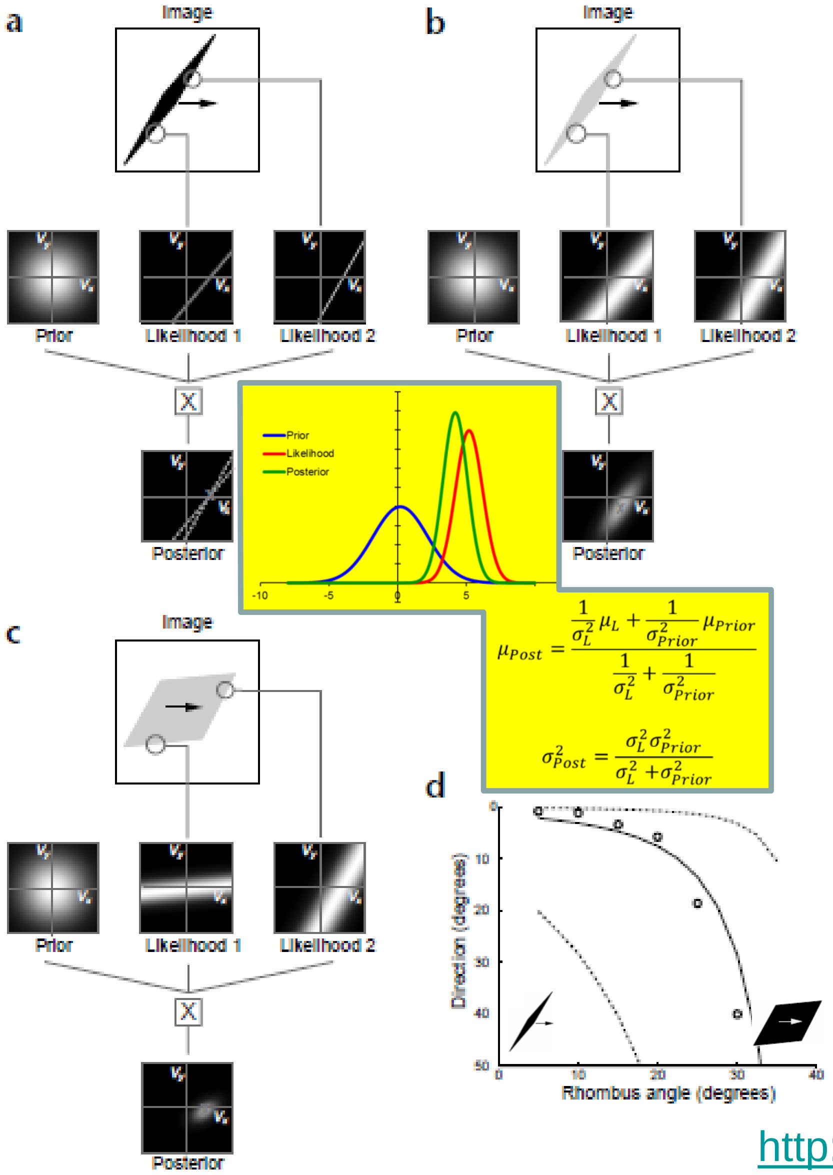
Likelihood functions for three local patches of a horizontally translating diamond stimulus, computed using equation (4). Intensity corresponds to probability. Top, high-contrast sequence. Bottom, low-contrast sequence, with the same parameter σ. At edges, the local likelihood is a ‘fuzzy’ constraint line; at corners, the local likelihood peaks around the veridical velocity. The sharpness of the likelihood decreases with decreasing contrast.



http://www.cs.huji.ac.il/~yweiss/Rhombus/rhombus.html

Motion illusions as optimal percepts

Intersection of Constraints & Priors (Weiss, Simoncelli & Adelson, 2002)



$$\mu_{Post} = \frac{\frac{1}{\sigma_L^2}\mu_L + \frac{1}{\sigma_{Prior}^2}\mu_{Prior}}{\frac{1}{\sigma_L^2} + \frac{1}{\sigma_{Prior}^2}}$$

$$\sigma_{Post}^2 = \frac{\sigma_L^2 \sigma_{Prior}^2}{\sigma_L^2 + \sigma_{Prior}^2}$$

Predictions of ideal observer for rhombus stimuli. (**a**–**c**) Construction of the posterior distribution for the rhombus stimuli. For clarity, likelihood functions for only two locations are shown; the estimator used in our study incorporated likelihoods from all locations. (**d**) Circles show perceived direction for a single human subject as rhombus angle was shifted gradually from thin to fat rhombuses (all three subjects showed a similar effect, and all gave informed consent to participate in the study). Each subject was given 100 presentations. Solid line shows the predictions of the Bayesian estimator computed using equation (1), where the free parameter was varied manually to fit the data. Dotted lines indicate the predictions when the free parameter was decreased by a factor of 10 (top dotted line) or increased by a factor of 10 (bottom line).

http://www.cs.huji.ac.il/~yweiss/Rhombus/rhombus.html

> The binding problem is when you look at what's happening in the brain, you find there's a division of labor. You have some parts of your brain that care about vision, some about hearing, some about touch. And even within a system, like vision, you have parts that care about colors, parts that care about orientations, parts that care about angles. And how this all comes together so that you have a unified perception of the world is one of the unsolved mysteries in neuroscience.
>
> David Eagleman

THE BINDING PROBLEM

Intersection of Constraints (for motion only)

Synchronization
Gestalt rules (e.g. common fate)
Feature integration (Treisman)

Feature Integration Theory revisited (Treisman, 1988)

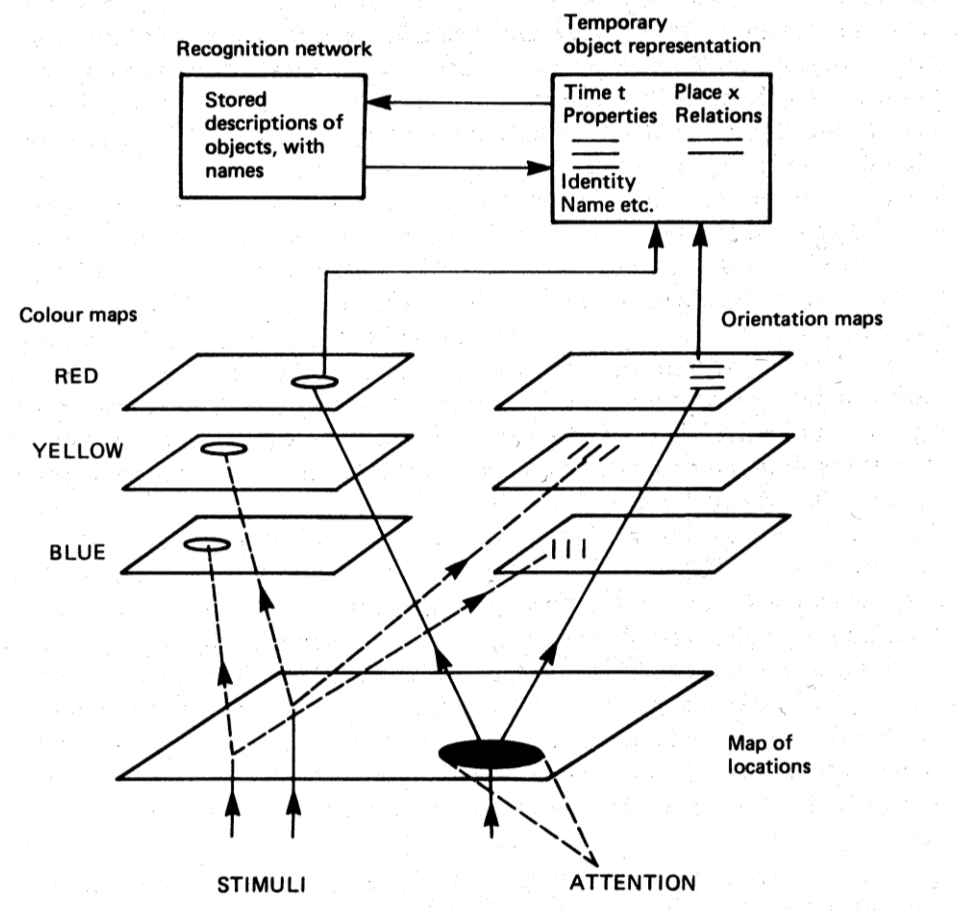
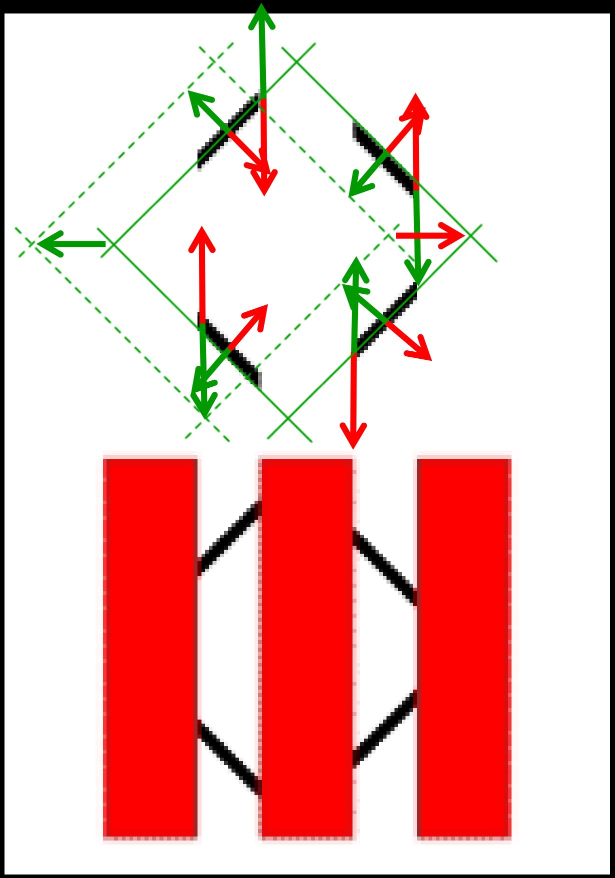
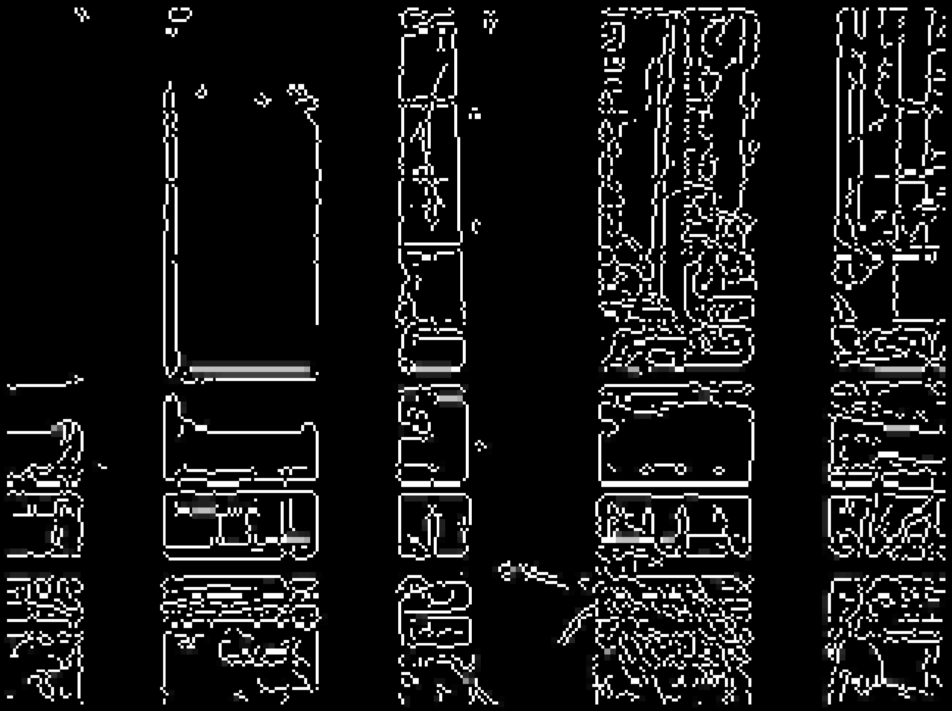


FIG. 5.18. A processing framework illustrating feature integration theory. (Reproduced from Treisman, 1988, with permission.)

Récupération des coordonnées x-y

The binding problem

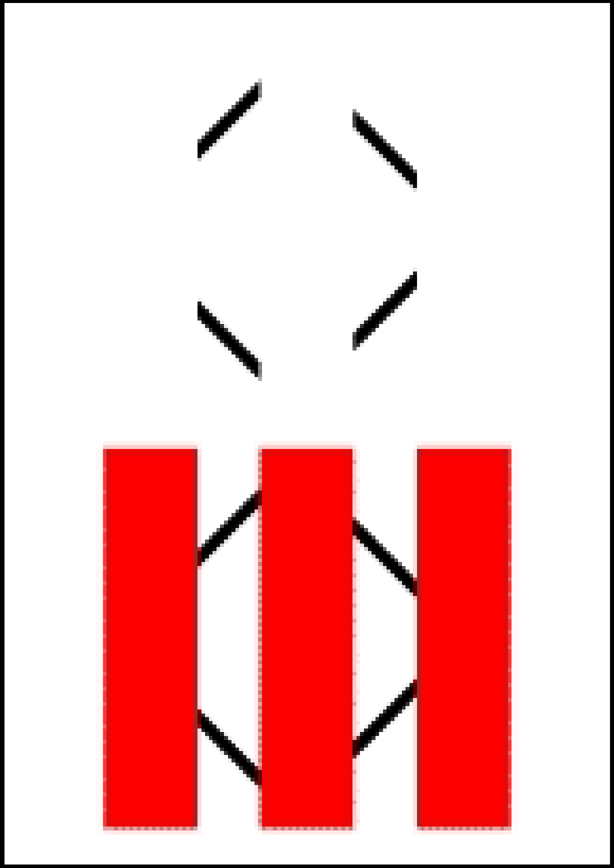




Lorenceau & Shiffrar (1992) *Vis. Res., 32(*2), 263-273.
Lorenceau & Shiffrar (1996) *Vis. Res., 36*(14), 2061-2067.

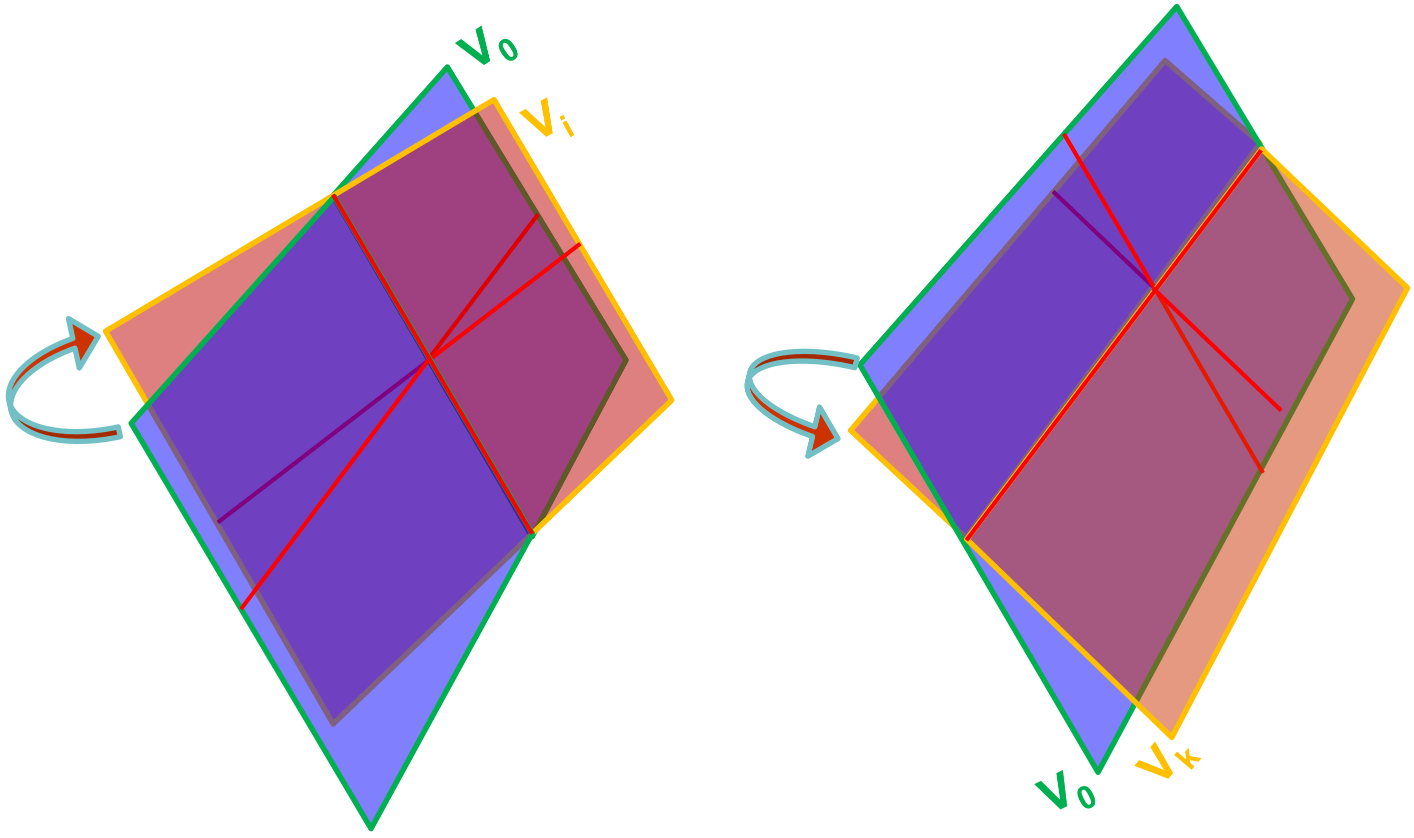
http://www.michaelbach.de/ot/mot-motionBinding/index.html

The binding problem



Lorenceau & Shiffrar (1992) *Vis. Res., 32(*2), 263-273.
Lorenceau & Shiffrar (1996) *Vis. Res., 36*(14), 2061-2067.

THE END


V_0
V_i
V_0
V_k