SRGI conference: Sub-Riemannian Geometry and Interactions
Paris, September 7–11, 2020

The primary visual cortex as a sub-Riemannian engine

Jean Petitot
CAMS, EHESS, Paris
jean.petitot@ehess.fr

1. Introduction

Emmanuel asked me to explain in a not too technical way why and how sub-Riemannian geometry could be operational for models of visual perception.

This SRGI (I for "interaction") is easy to formulate but difficult to carry out.

2. Introduction (continued)

1. The links between lived (experienced) visual perception and geometry are obvious. Visual perception is highly geometrically structured, and conversely, geometry is largely an idealization and abstract generalization of perceptual intuitions.

2. We have known since antiquity a host of things about visual perception: phenomenology of visual lived experience, psychology of perception (Gestalt), psychophysics.

It all comes under psychology, that is, the study of *the mind*.

3. Introduction (continued)

3. But until very recently, absolutely *nothing* was known about the *neural correlates* of perception, which in turn come under the study of *the brain*.

The neural implementation was a black box. At most, we knew that such or such lesion produced such or such visual deficit.

It was not until the end of the 1950s that the activity of a few neurons activated by perceptual tasks began to be recorded by means of electrodes.

But the data remained very local, sparsed and scattered.

4. Introduction (continued)

4. It was only much later, at the turn of the 1980s and 1990s, that the imagery revolution made it possible to obtain *global* images of the activity of visual areas.

This break made it possible to begin to investigate the links between the psychology of perception (mind) and its neurophysiology (brain).

The implementation of perceptual software by neural hardware then became a cutting edge problem and a new research horizon.

5. Introduction (continued)

This is the challenge I would like to address in this presentation. It will necessarily be only a sketch. If you are interested in more precise experimental data, you can consult my book *Elements of Neurogeometry*. Volume 1: *Functional Architectures of Vision*, Springer 2017.

6. Introduction (continued)

A preliminary epistemological and methodological question is how to think about *implementation*.

Indeed, we cannot use the analogy with computers as universal Turing machines because brain hardwares are *dedicated*. It is their material hardware architecture that explains the abstract structures of the mind softwares that are implemented in.

7. Introduction (continued)

We then come up against a fundamental problem.

The geometry of visual perception involves many differential computations. But neurons *cannot* compute. They are simply active and connected, their connections allowing them to transmit activities.

It is their connectivity that can "compute" something.

It is therefore necessary to understand how a connectivity can perform differential computations. It must have a *very special* functional architecture.

I will try to show that this functional architecture can be modeled by jet spaces, contact structures and sub-Riemannian metrics.

8. Planets cannot compute

It is perhaps relevant to give here a completely different, but very well-known, example of a similar problem.

In the Newtonian theory of gravitation, we have the differential equations giving by integration the trajectories of the planets
(i) within the framework of a globally Euclidean space-time background structure (classical Galilean relativity), and
(ii) with the crazy hypothesis of long distance instantaneous forces.

But planets cannot "compute" anything: they can only follow *inertial* motions.

So, how then can simple inertial motions be equivalent to complicated observed and computed trajectories?

9. Riemannian geometry can compute Gravitation

How is it possible to explain gravitation purely locally and inertially?

It was general relativity that provided the solution by changing the globally Enclidean-Galilean background structure and using (semi-)Riemannian geometry.

Trajectories become geodesics and, in this new framework, can now be as complicated as needed to fit the empirical data.

10. Sub-Riemannian geometry can compute Vision

The same is true here.

We observe complicated geometric structures of visual perception and we compute models using the Euclidean plane $\mathbb{R}^2$ as background structure (2D image theory).

But neurons can't compute anything. They are simply active and connected.

Their connection systems must therefore be equivalent to the computations.

To "achieve this goal" (even if it is Darwinian and without "goal"), biological evolution has progressively implemented, in sophisticated modular connectivities, fiber structures of dimension $2 + n$ as jets, contact structures, and sub-Riemannian metrics ...

11. Outline

I will sketch the following propositions:

1. a class of cortical neurons detect contact elements (a, p);
2. the fibration $(a, p) \longrightarrow a$ is neurophysiologically implemented;
3. the contact structure of the 1-jets of plane curves is neurophysiologically implemented;
4. the sub-Riemannian geometry of this contact structure is neurophysiologically implemented.

Each of these propositions points to crucial neurophysiological discoveries which are all experimental "tours de force".

That biological evolution has been able to lead, from simple photosensitive sensors, to such genetically controlled neurophysiological structures is a real miracle.

12. A fun story and a long journey

Let us now begin a fun story and a long journey.

It begins, as often in science, with a revolution in the *means of observation*.

Extraordinary advances in recording and imaging techniques have transformed the brain from a "black box" to a more "transparent" box.

This led to a complete transformation of our understanding of mental activities.

13. Functional architectures

Since the early 1990s, methods of "in vivo optical imaging based on activity-dependent intrinsic signals" enabled to visualize the extremely special connectivity of the primary areas of the visual cortex, what is called their "functional architectures".

Single cortical visual neurons detect very *local* geometric cues (positions, contrasts, orientations, tangents, curvatures, inflection points, cusps, corners, crossings, etc.).

But it is the *intracortical architectures* that explain how these local cues can be *integrated* and generate the *global geometry* of the perceived images, with all the well-known phenomena studied since Gestalt theory, in particular *long range illusory contours*.

14. Neurogeometry

In the 1990s, I coined the expression "neurogeometry of vision" to refer to geometrical models of functional architectures.

The very particular connectivity of the functional neuroanatomy of these areas explains the geometry of percepts and must therefore be implemented in the synaptic weights of the neural nets used for modelling.

Neurogeometry is mainly based on the discovery that the functional architecture of $V1$ (the first primary visual area) implements the contact structure and the sub-Riemannian geometry of the 1-jet space of plane curves.

15. Illusory contours as geodesics

From there, illusory contours can be interpreted as *geodesics* of this contact structure.

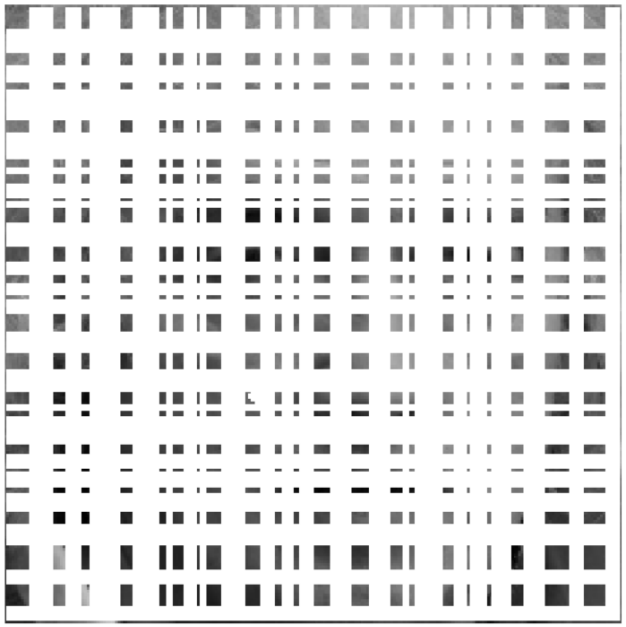
Such sub-Riemannian models generalize a previous model due to David Mumford and based on the theory of *elastica*.

They have many applications, in particular for *inpainting* algorithms.

16. An example of sub-Riemannian inpainting

The following picture shows how a highly corrupted image can be very well restored using sub-Riemannian diffusion.

The face of our friend Jean-Paul appears out of the blue.



17. Perceptual geometry. I

"Neurogeometry" concerns mathematical models for the neural algorithms processing perceptual geometry.

The relationships between visual perception and geometry are as old and diverse as geometry itself since antiquity.

The geometry of visual perception have been described, but not modeled, since the times of Goethe, Helmholtz, Hering, Brentano, Poincaré, Husserl and Gestalttheorie (von Ehrenfels, Wertheimer, Stumpf, Koffka, Köhler, Klüver, etc.), from Kanizsa to Marr in psychology, from Evans to Peacocke or McDowell in philosophy of mind.

18. Perceptual geometry. II

In modern times consider, for example, Helmholtz's answer "On the *facts* underlying geometry" [*Über die Tatsachen, die der Geometrie zum Grunde liegen*] (1868) to the founding text of Riemann (1854), "On the *hypotheses* which underlie geometry" [*Über die Hypothesen welche der Geometrie zu Grunde liegen*].

The so-called Riemann–Helmholtz "Raum Problem", concerning a system of axioms for perceptual space, was solved by Sophus Lie and Friedrich Engel in the third volume of their "Theory of transformation groups" [*Theorie der Transformationsgruppen*] (1888-1893).

19. Perceptual geometry. III

Regarding Henri Poincaré, consider Chapters IV and V of "La Science et l'Hypothèse" (1902) "L'espace et la géométrie" and "L'expérience et la géométrie".

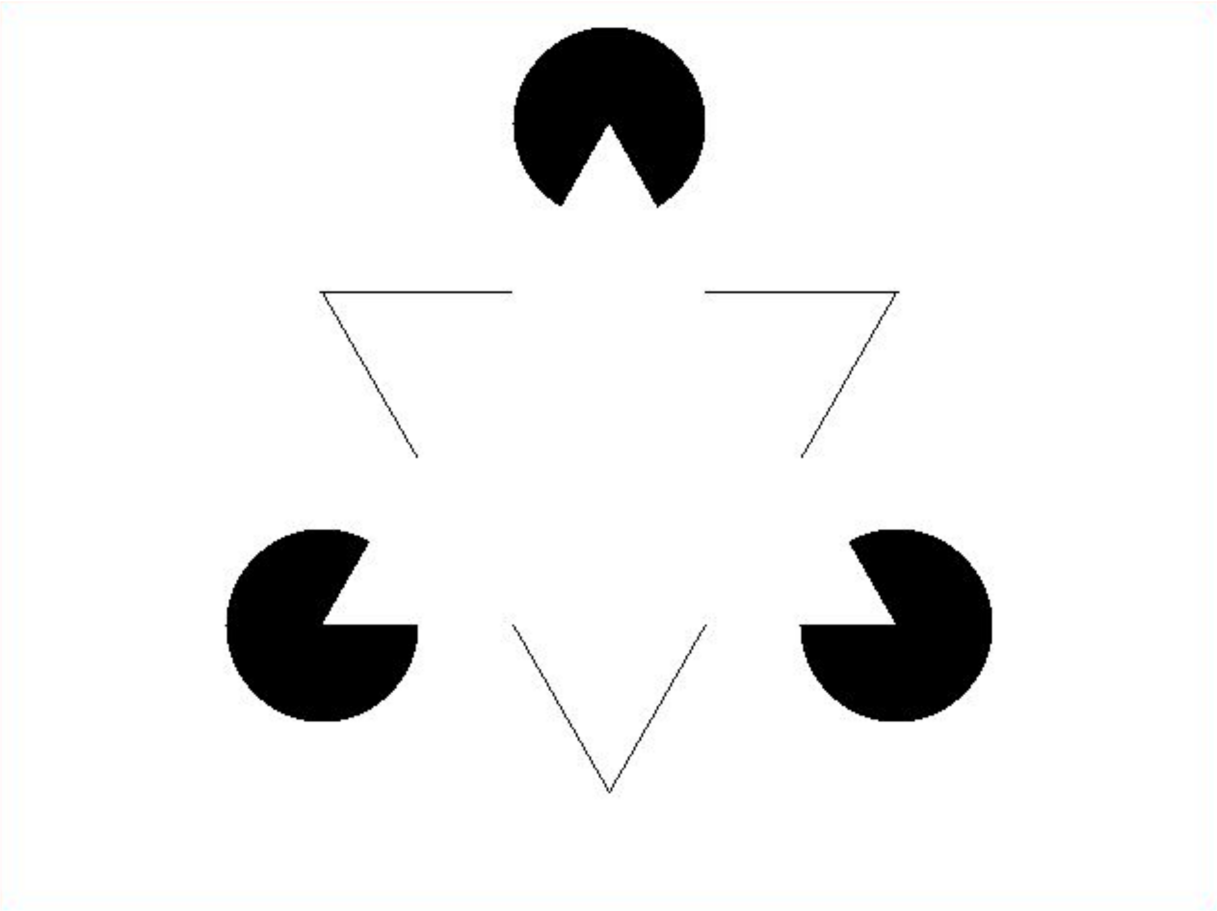
And also "Science et Méthode" (1908).

20. Completion and Gestalt theory

Phenomena studied by Gestalt theory are very striking.

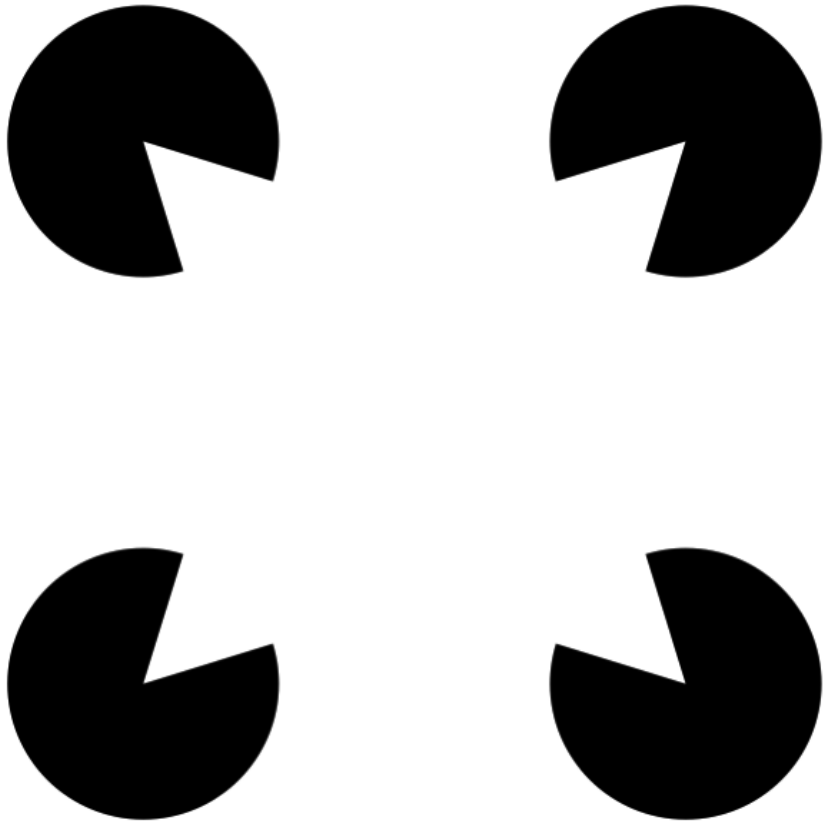
Consider for example this well-known Kanizsa triangle. *Local cues* as pacmen and end-points induce *very long-range global illusory contours* (what is called "modal completion").

21. Kanizsa triangle



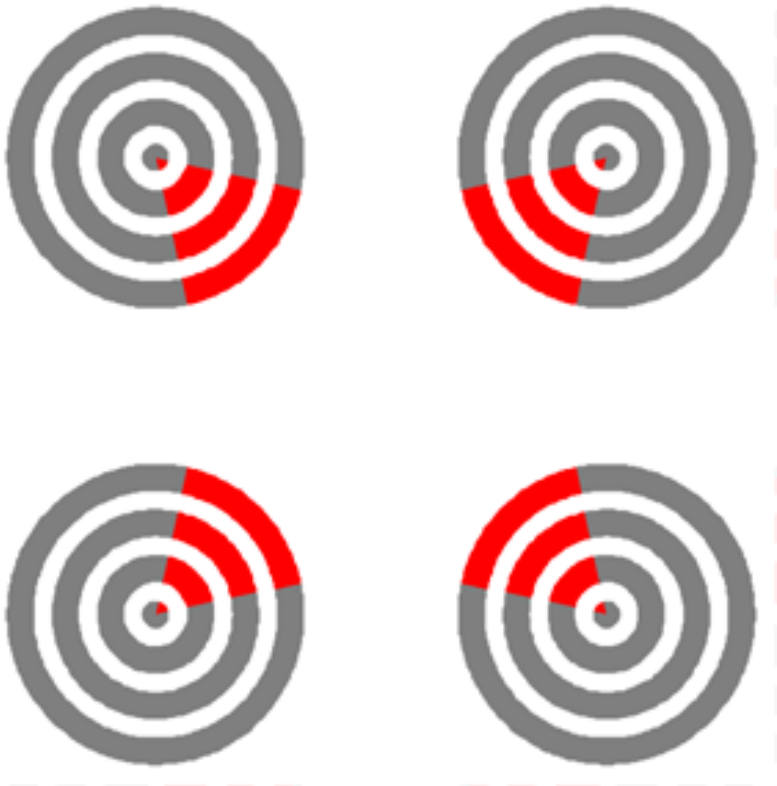
22. Curved Kanizsa square

Illusory contours are particularly interesting when they are *curved*.



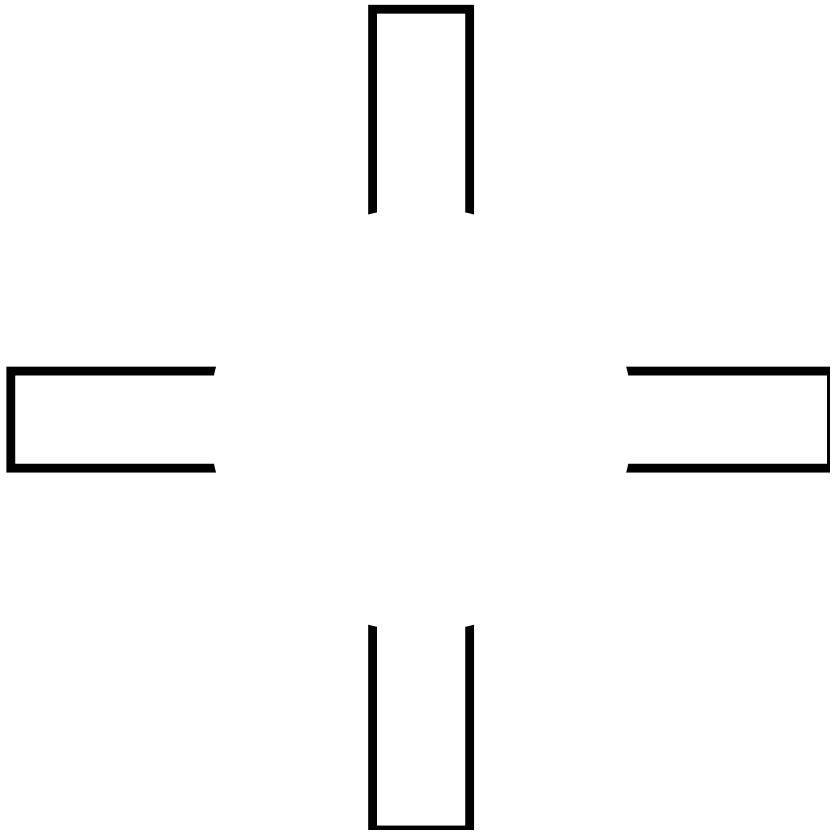
23. Watercolor (neon) effect

Furthermore, these contours act as *boundaries for a diffusion of color* inside the square (what is called the "neon" or "watercolor effect"). It is not easily seen on a screen but can be measured with adequate psychophysical methods.



24. Koffka cross

Consider also the Koffka cross:



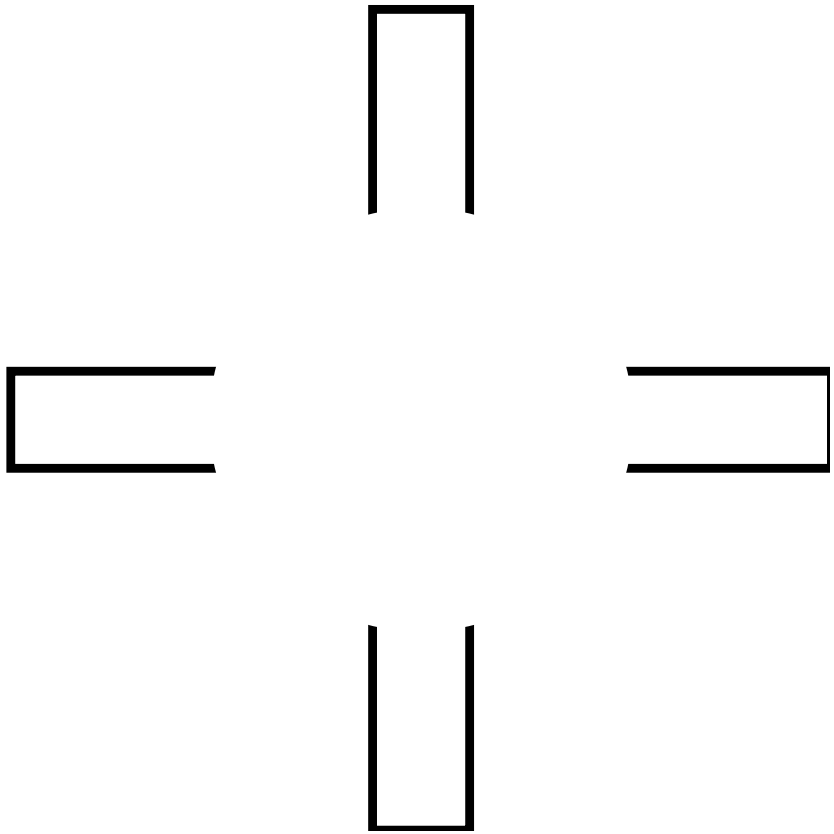
25. Bistability

The end points activate locally (area $V2$ is necessary) the orthogonal orientations.

Moreover, subjects perceive alternately circles and squares, which means that there exists a *competition* between two completion strategies:

- *circle*: illusory contours with a maximal diffusion of curvature,
- *square*: piecewise linear illusory contours (curvature $= 0$) with corners (singularities of curvature $= \infty$).

26. Koffka cross, *bis repetita*

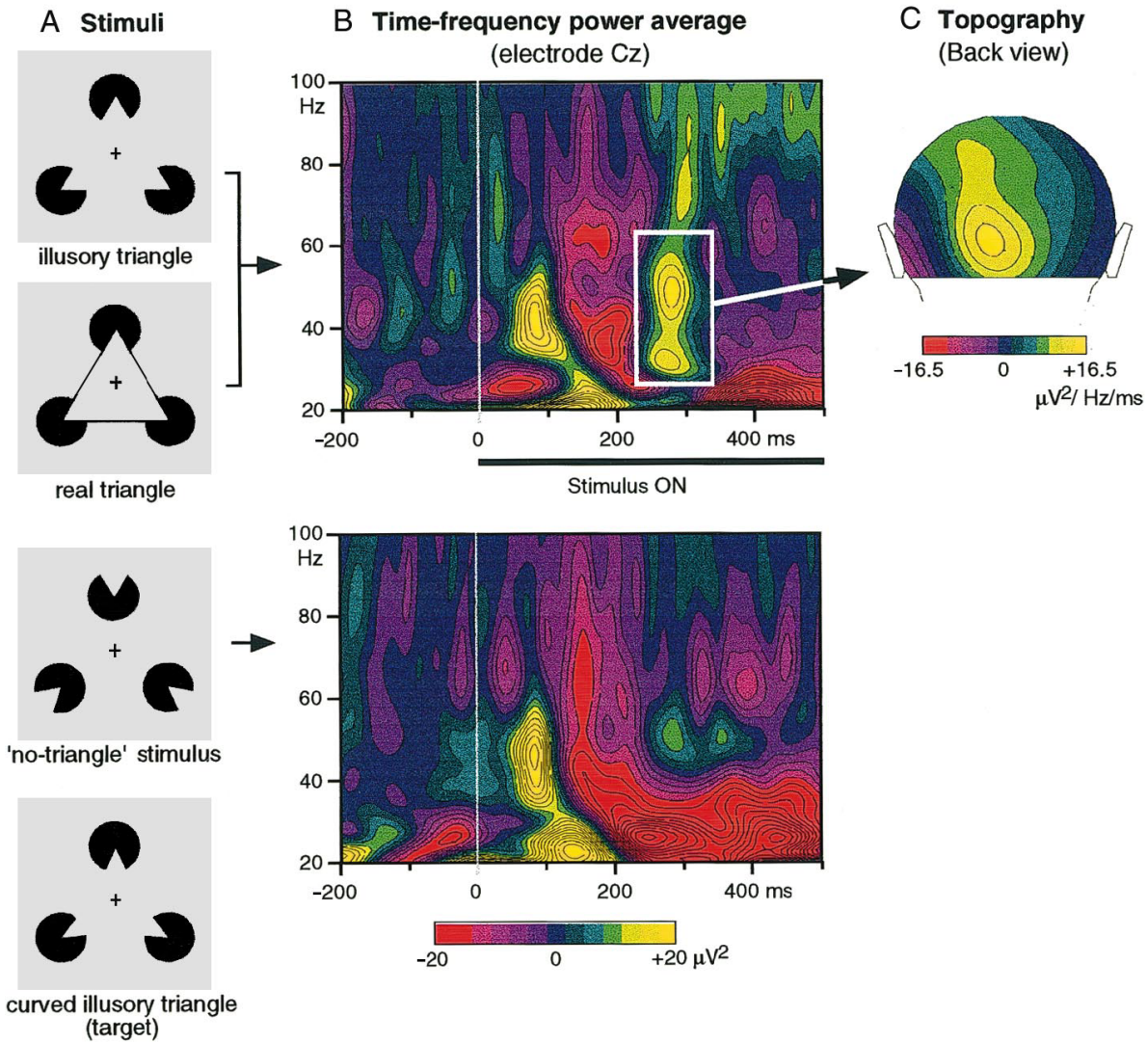


27. Illusory contours are neurally real. I

Illusory contours are *neurally real* (Catherine Tallon-Baudry, 1999).

We consider the EEG of an illusory triangle and of a real triangle (three pacmen with global coherence, top of the following figure) and a no-triangle stimulus (three pacmen with no global coherence, bottom of the figure).

28. EEG image


A Stimuli
illusory triangle
real triangle
'no-triangle' stimulus
curved illusory triangle
(target)
B Time-frequency power average
(electrode Cz)
100
Hz
80
60
40
20
-200
0
200
400 ms
Stimulus ON
-20
0
+20 μV2
C Topography
(Back view)
-16.5
0
+16.5
μV2/ Hz/ms

29. Illusory contours are neurally real. II

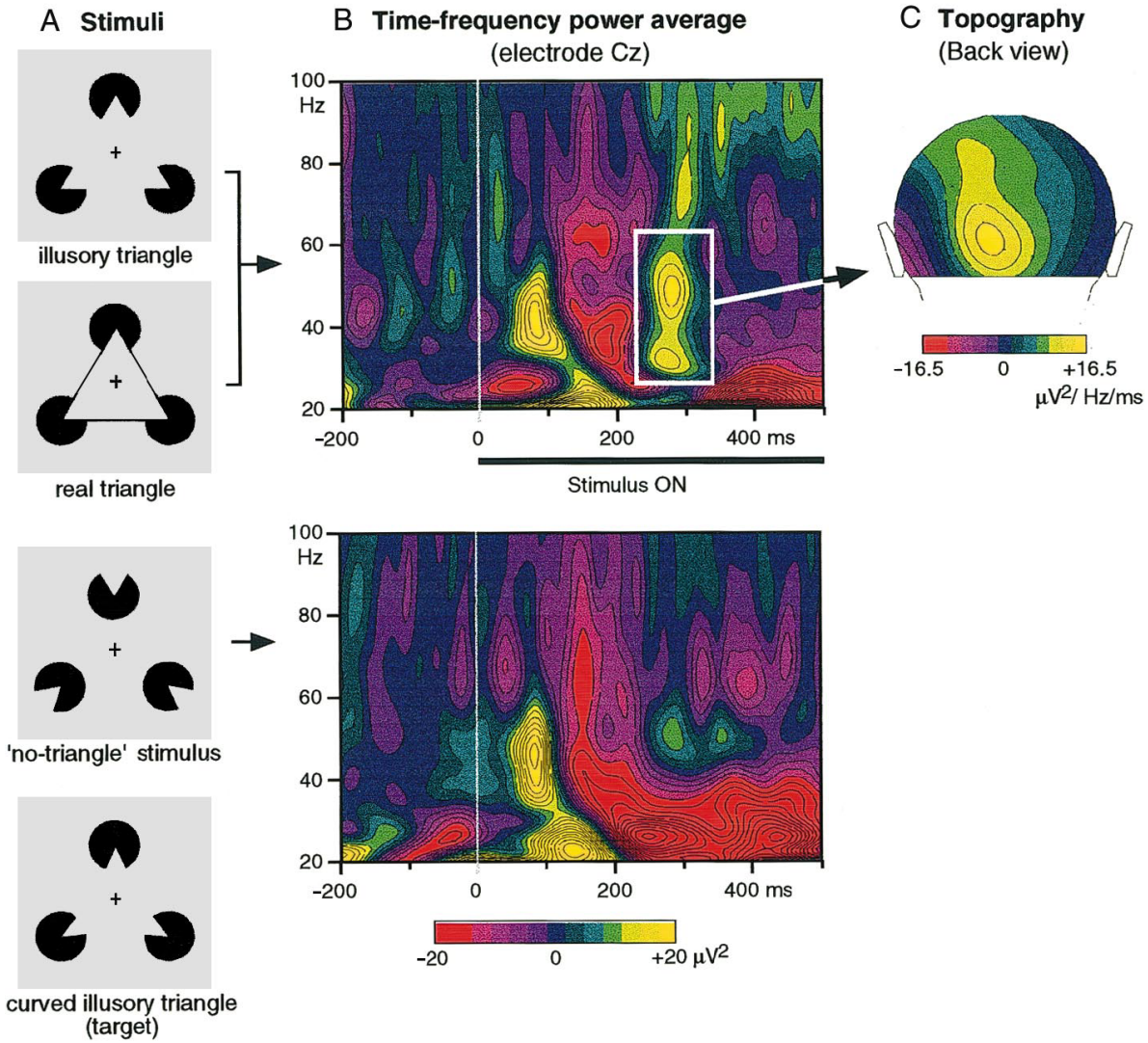
We observe

> *"two successive bursts of oscillatory activities [...]. A first burst at about* $100ms$ *and* $40Hz$*. It showed no difference between stimulus types. A second burst around* $280ms$ *and* 30-$60Hz$*. It is most prominent in response to coherent stimuli."*

Yet, the second activation burst, which corresponds to the integration of local sensory data into a global percept, is *the same* for both types of triangles, whether real or illusory,

30. EEG image (*bis repetita*)

The second burst is framed.



31. The antinomy of perceptual geometry. I

QUESTION:
How the visual brain can be a neural geometric engine?
How differential and integration routines can be neurally implemented?

Visual cortical neurons are very local detectors and even "point-like processors".

They can only code a single numerical value by means of their "firing rate".

They emit action potentials (spikes) defining their "rate coding".

32. The antinomy of perceptual geometry. II

Such point-like processors cannot implement *directly* classical differential and integration routines (the classical versions of "differentiation" and "integration" do not work).

They cannot "compute" anything.

They can only be more or less active and propagate their activity along more or less inhibitory or excitatory connections.

But *maybe* they can implement a *geometrization* of differential calculus in complex networks of point-like processors *if the connectivity of the network is so specific that, for it, to be activated turns out to be equivalent to computing differential and integration routines*.

33. Binding hypothesis

We must emphasize here the neurophysiological principle called "binding hypothesis".

The global coherent features of percepts emerge from pixelized stimuli when the involved neurons *synchronize* their activity through their connections: the local data detected by the neurons become "bound" together .

"Binding hypothesis" :
connections $\longrightarrow$ synchronization $\longrightarrow$ binding $\longrightarrow$ propagation of coherent activity $\longrightarrow$ saliency $\longrightarrow$ pop out of global features.
No connections, no synchronization $\longleftarrow$ no binding $\longleftarrow$ no propagation of coherent activity $\longleftarrow$ no saliency $\longleftarrow$ no pop-out of global features.

34. The antinomy of perceptual geometry. III

My hypothesis is that we can use to solve the antinomy the long and deep tradition of *geometrizing* the differential calculus: Pfaff, Jacobi, Clebsch, Grassmann, Frobenius, Lie, Engel, Darboux, Cartan, Weyl, Goursat.

In neurogeometry, this type of geometry (jet-spaces, differential forms, contact structures, etc.) becomes "internal", "immanent", neurally implemented, and explains the "external", "transcendent", ideally formalized, perceptual geometry.

ANSWER to the QUESTION:
Main hypothesis: the visual brain can be a neurally implemented Lie-Cartan geometric engine!

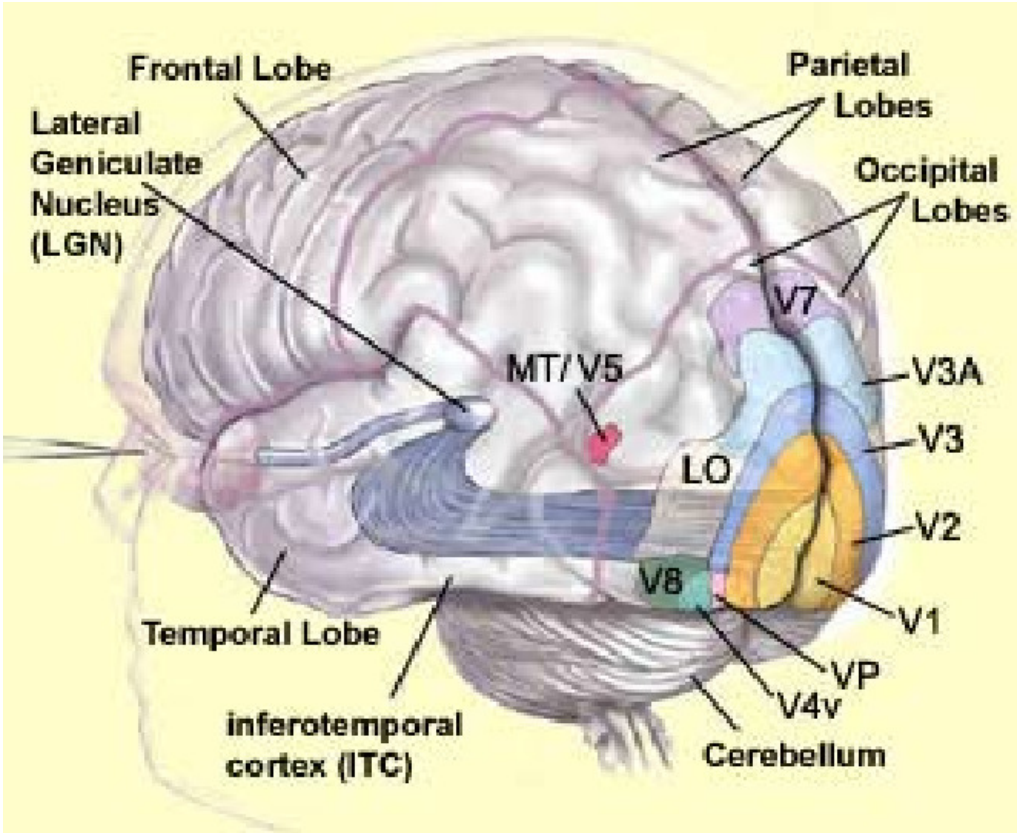
35. Jan Koenderink

Jan Koenderink strongly emphasized this crucial point:

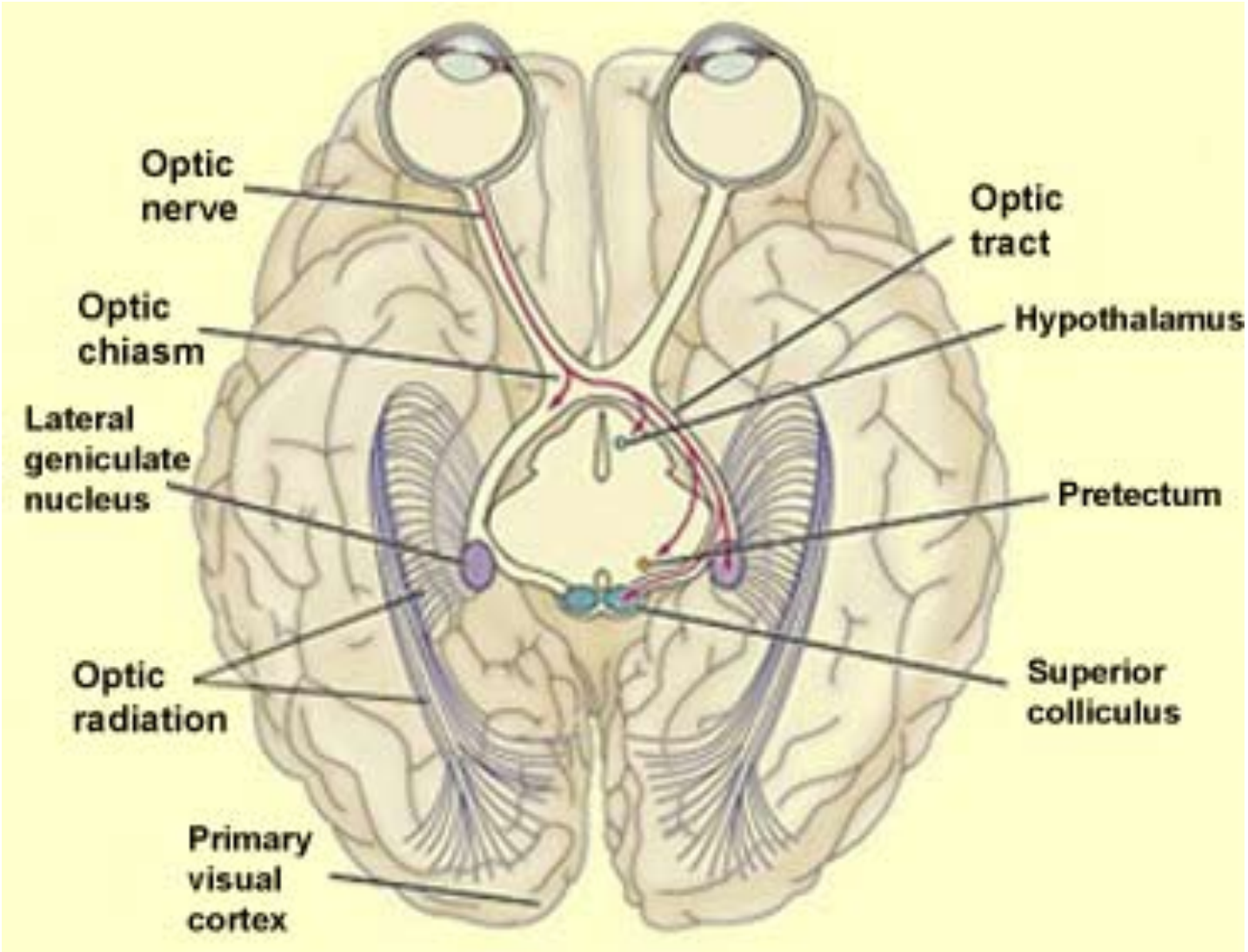
> *"Geometrical features [are] multilocal objects, i.e., in order to compute [boundary or curvature] the processor would have to look at different positions simultaneously, whereas in the case of jets it could establish a format that provides the information by addressing a single location. Routines accessing a single location may aptly be called point processors, those accessing multiple locations array processors. The difference is crucial in the sense that point processors need no geometrical expertise at all, whereas array processors do (e.g., they have to know the environment or neighbours of a given location)."*

36. The visual brain. I

Here are two images of the human brain. They show the neural pathways from the retina to the lateral geniculate nucleus (LGN thalamic relay) and then to the occipital primary visual cortex (area $V1$).



37. The visual brain. II



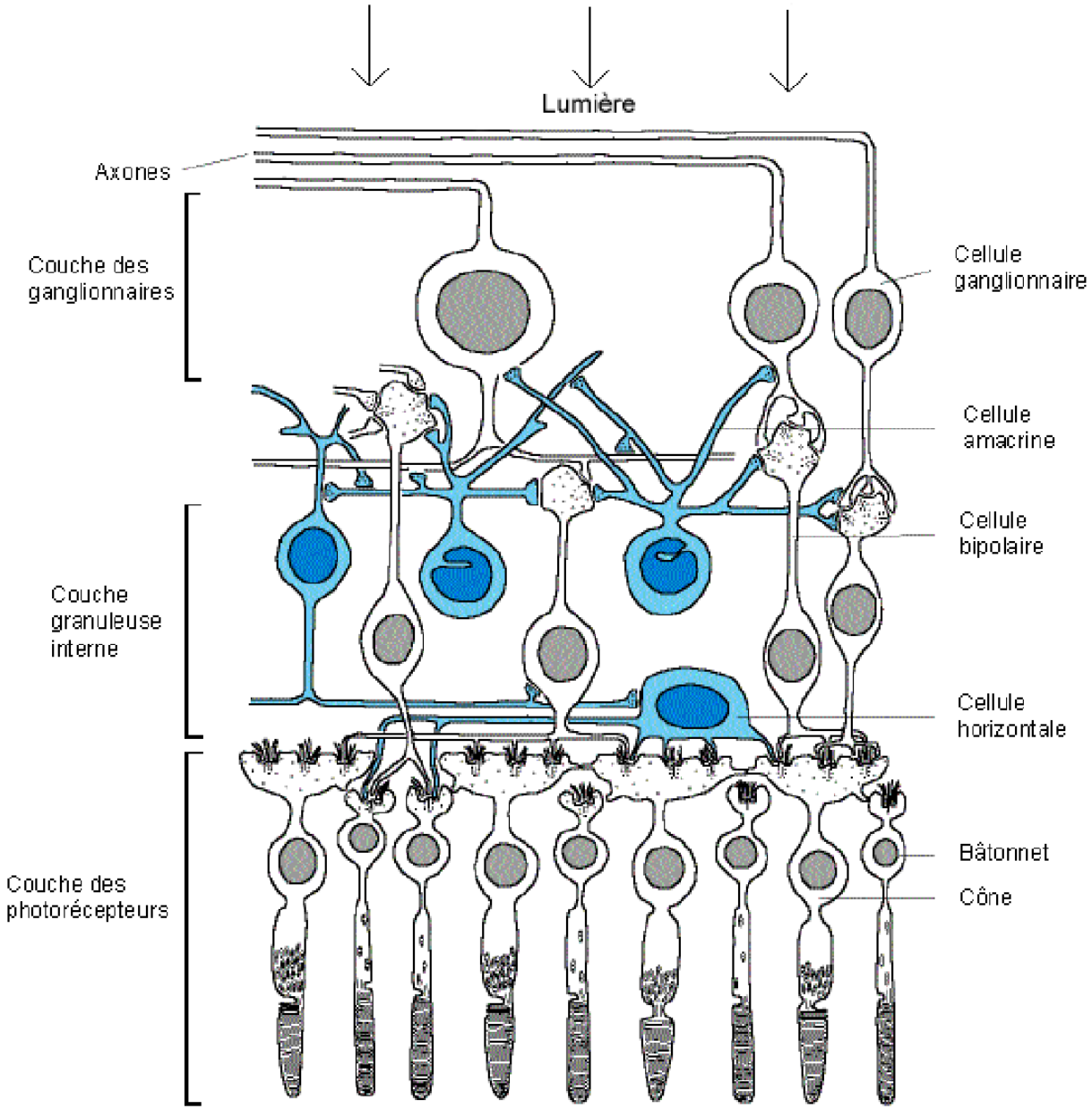
38. Retina layers

For the retina, there are already complicated connections between 5 layers.

1. Photoreceptors, few μm and around 6×10^6 cones and 120×10^6 rods in humans.
2. Horizontal cells,
3. Bipolar cells,
4. Amacrine cells,
5. Ganglion cells (around) 1.5×10^6. Axons $\Longrightarrow$ optical nerve.

The compression ratio ("degree of convergence") of the photoreceptors onto the ganglion cells is of the order of 100.

39. Retina image



40. Receptive fields and receptive profiles

In a very rough linear approximation, neurons act as filters on the optical signal transduced by the photoreceptors.

They are connected to a small domain D of the photoreceptor layer, called their *receptive field.*

Their *receptive profile* is their transfer function as filter. It is a function $\phi(x, y)$ on D whose value is > 0 or < 0 according to the fact that the impulse response of the neuron to a Dirac light impulse at (x, y) is excitatory (ON) or inhibitory (OFF).

The receptive profiles are well modeled by derivatives of Gaussian G or by Gabor patches (*i.e.* sinusoidal functions modulated by Gaussians G).

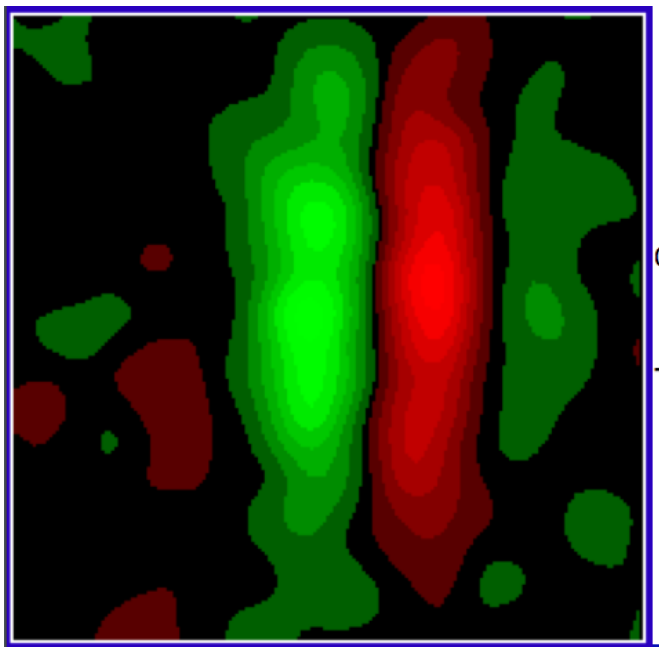
The width σ of G defines a *scale* and the neuron is a *point-like processor* at this scale.

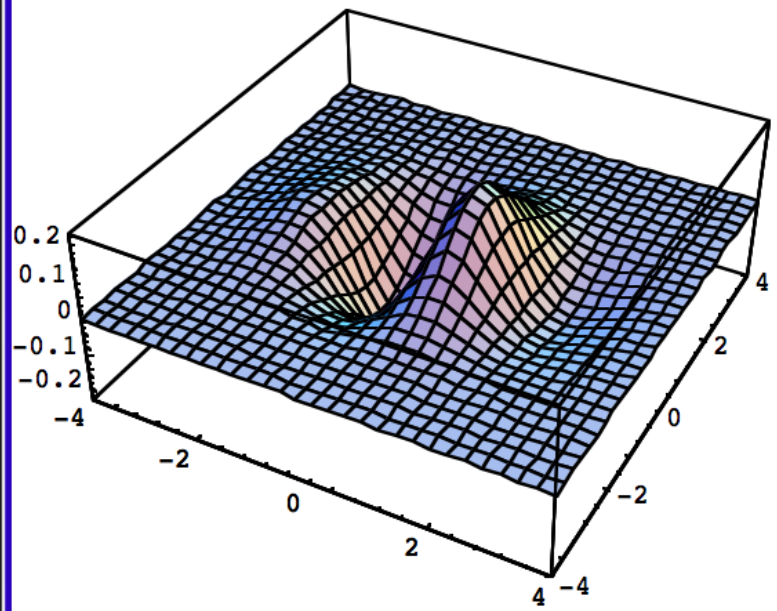
41. Example of receptive field

Level sets of receptive profiles can be recorded. It is an experimental "tour de force".

Here is an example from a cortical neuron in $V1$ (Gregory DeAngelis).

Left: Level sets ON (excitatory, red)/OFF (inhibitory, green) zones. Right: model using a third derivative $\varphi(x,y) = \frac{\partial^3 G}{\partial x^3}$).





42. Wavelet analysis

The filtering of the signal is like a *wavelet analysis* (a spatially localized and multiscale Fourier analysis).

It is a compromise between a spatial representation and a frequency representation. We discussed this point very early with Stéphane Mallat.

The receptive fields of the ganglion cells are like *Gaussian Laplacians*.

David Marr explained in *Vision* (1982) that their function was to detect local discontinuities.

Since Marr, vision has been considered one of the earliest empirical sources of wavelet analysis.

43. Coherent states and harmonic analysis. I

Wavelets are related to the notion of a *coherent state* (well known in quantum mechanics).

1. We want to analyse signals considered as vectors of a Hilbert space $\mathcal{H}$ (here $L^2\left(\mathbb{R}^2\right)$).

2. A locally compact group G is available (here the Euclidean group $SE(2)$ of direct isometries of the plane) which acts on $\mathcal{H}$ through a unitary irreducible square-integrable representation.

3. A basic receptive profile $\varphi_0 \in \mathcal{H}$ is also available, which is well localised both in the position space and in the Fourier space.

44. Coherent states and harmonic analysis. II

4. We suppose that φ_0 satisfies a condition of "admissibility" and we take the orbit $\{\varphi_g\}_{g\in G}$ of φ_0 under the action of G. We get a coherent state.

5. We perform the harmonic analysis of the signals $f \in \mathcal{H}$ using these coherent states. This allows to represent signals as superpositions of elementary functions and this way, in our neurogeometrical context, to *measure* them neurophysiologically.

Giovanna Citti and Alessandro Sarti worked a lot on this $SE(2)$ harmonic analysis.

45. Efficient coding

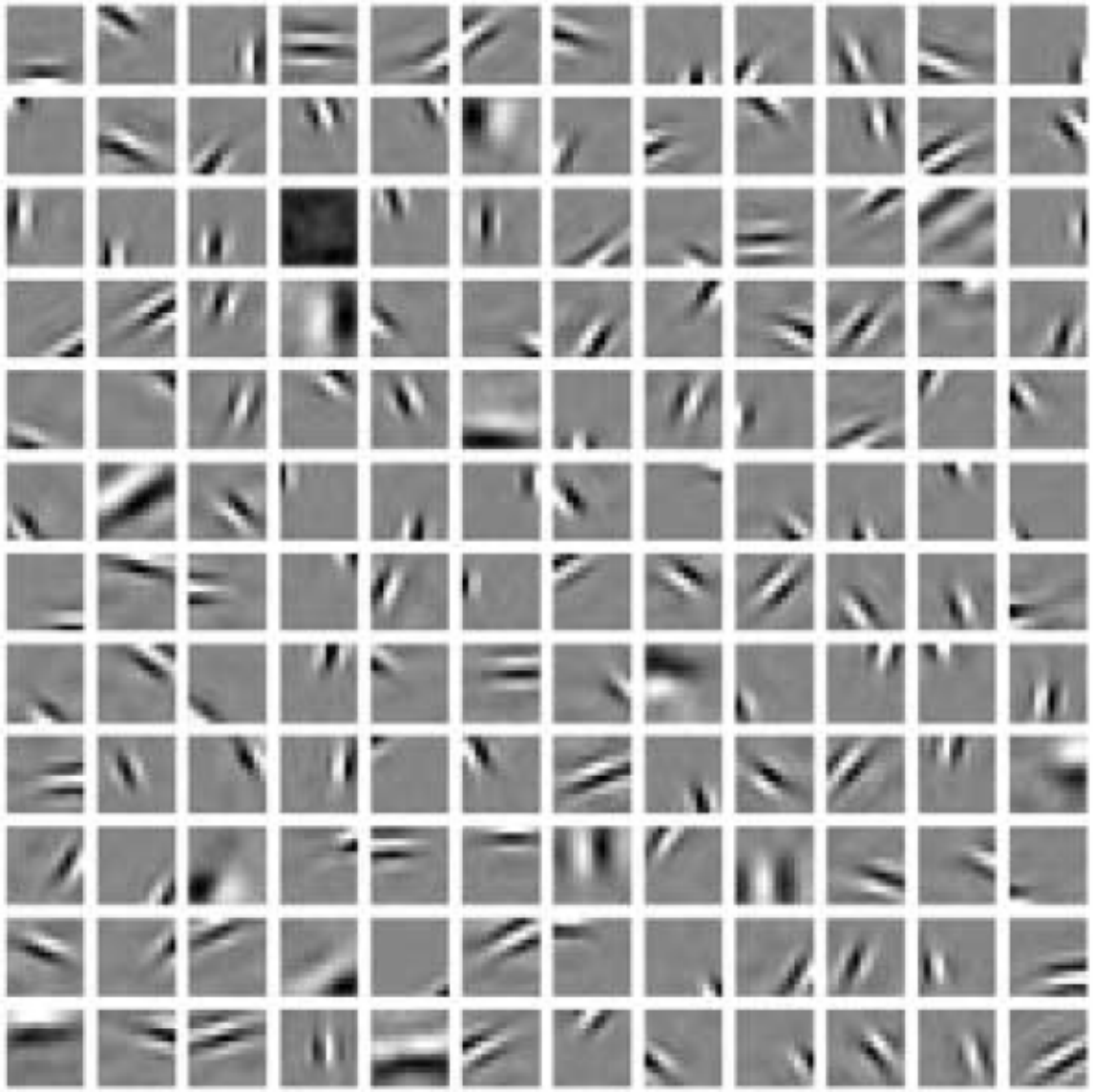
These receptive profiles can be retrieved from general hypotheses of information theory. They optimize the efficiency of the neural representation of information.

The "pixelized" transduction by photoreceptors is extremely redundant and fundamentally inefficient.

The system has to *compress* it using a "sparse" coding decorrelating the signal and reflecting the statistics of *natural* images, which is very particular.

Optimal sparse representations for the class of natural images lead to basic functions which are very similar to the receptive profiles of visual neurons (Bruno Olshausen, David Field, Karol Gregor, Yann Le Cun, Eeno Simoncelli).

46. Efficient coding (image)



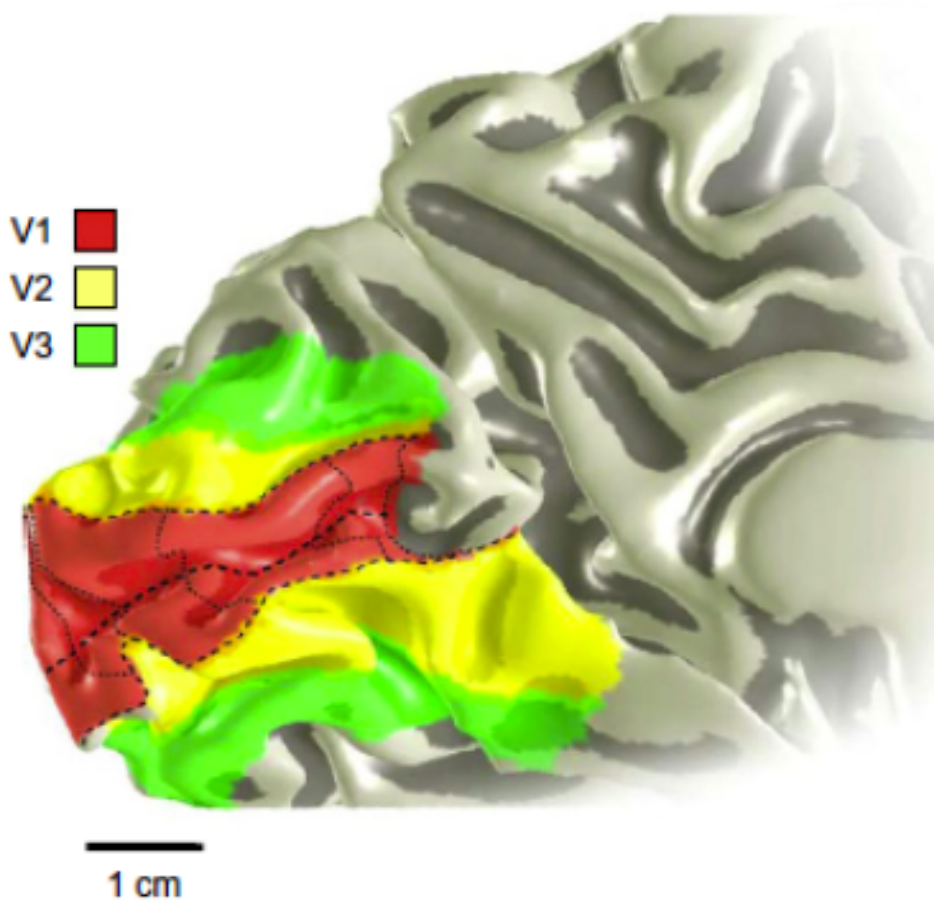
47. Area V1

From now on I will focus on the first primary area V1 (cat's area 17).
I won't say anything about

1. the feedback of $V1$ onto the LGN;
2. the post-$V1$ processing of visual images by other cortical areas.

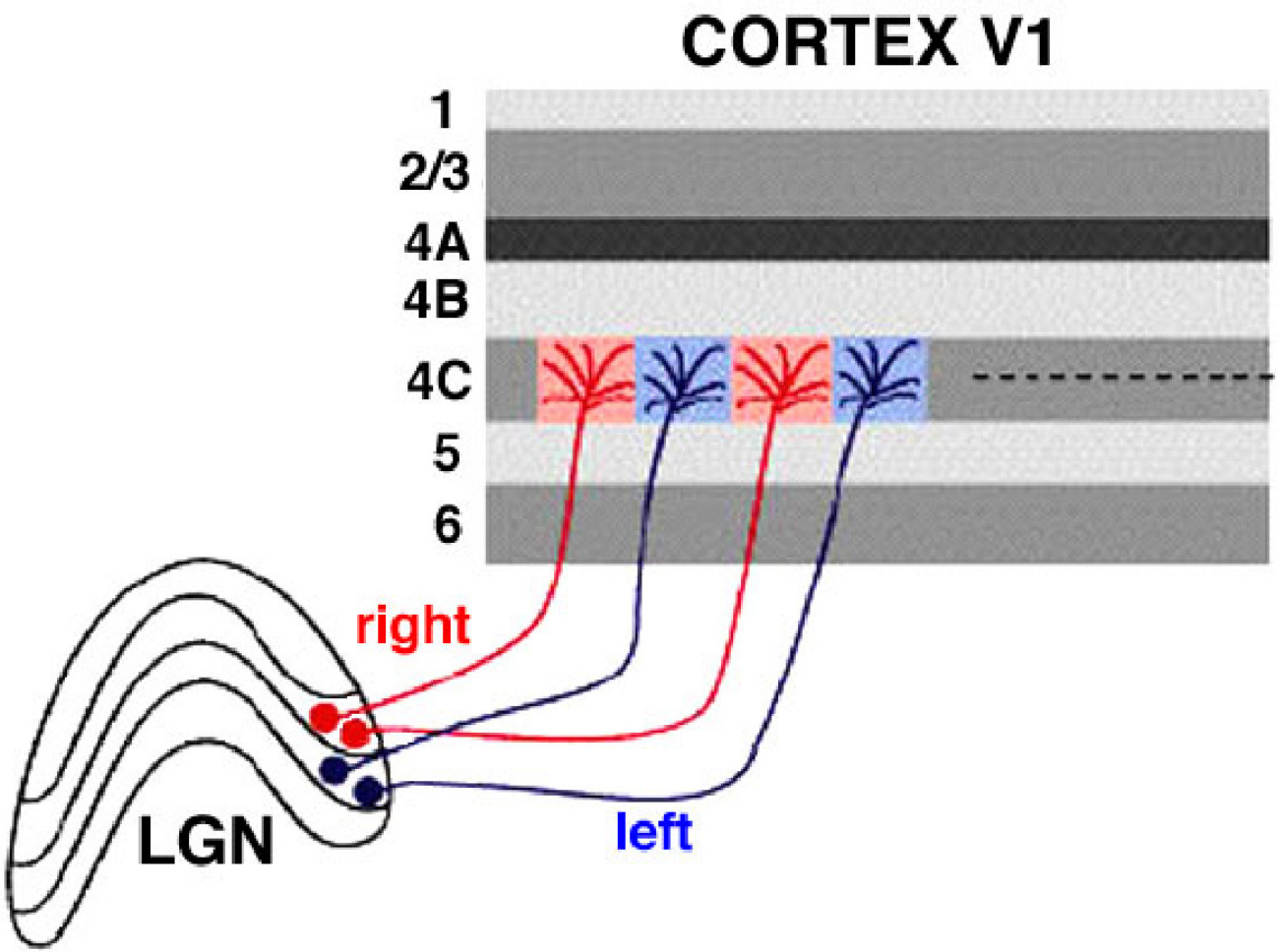
This restriction is of course a drastic limitation: reality is much more complex!

This image shows how $V1$ is surrounded by $V2$ and $V3$.



49. $LGN \longrightarrow V1$ (layer 4)

The retino-geniculo-cortical pathway projects the retina onto the layer 4 of $V1$:



50. Hubel & Wiesel crucial experiments

The crucial discovery of a functional architecture of $V1$ was made in 1959 by David Hubel and Torsten Wiesel (1981 Nobel Prize).

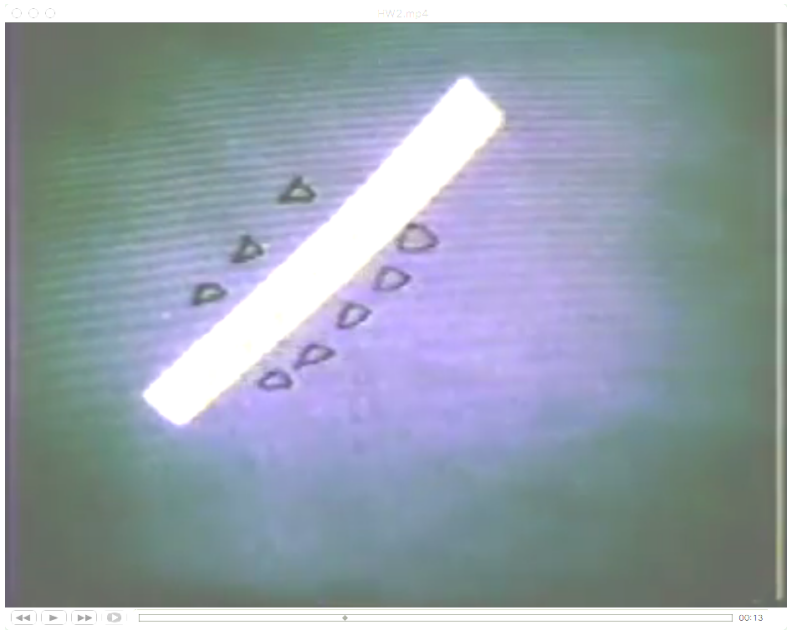
They implanted electrodes into cortical neurons and recorded the sound transduction of their activity.

By scanning the visual field with a small bar, they were able to locate very precisely the receptive fields.

They then discovered, almost by accident (it is a typical case of serendipity), that some of these neurons (they called "simple") were activated only when the bar had a specific well-defined orientation (modulo π).

51. Images of H&W

Here are two images of a 40s breakthrough recording.
Left: a bar aligned along the prefered orientation (noisy firing).
Right: a bar orthogonal to the prefered orientation (quiet, no firing)



52. Detecting contact elements

We can say that, at its proper scale (the size of its receptive fild), a simple neuron is a point-like processor detecting a pair (a, p) of a position a (its receptive field) and an orientation p at a.

In other words, it detects (at its proper scale) a *contact element*.

53. Improving H&W's experiments

Hubel and Wiesel analyzed the dependence of the orientation tuning curve w.r.t. the position of the bar, its orientation, its thickness, its length, its contrast, etc.

After them, other types of neurons have been studied, "complex" neurons with no privileged orientation, etc. And also detectors of other local geometric cues than mere orientations.

Today, many experiments explore the response to more complex stimuli, in particular pieces of natural images.

54. Conformal retinotopy

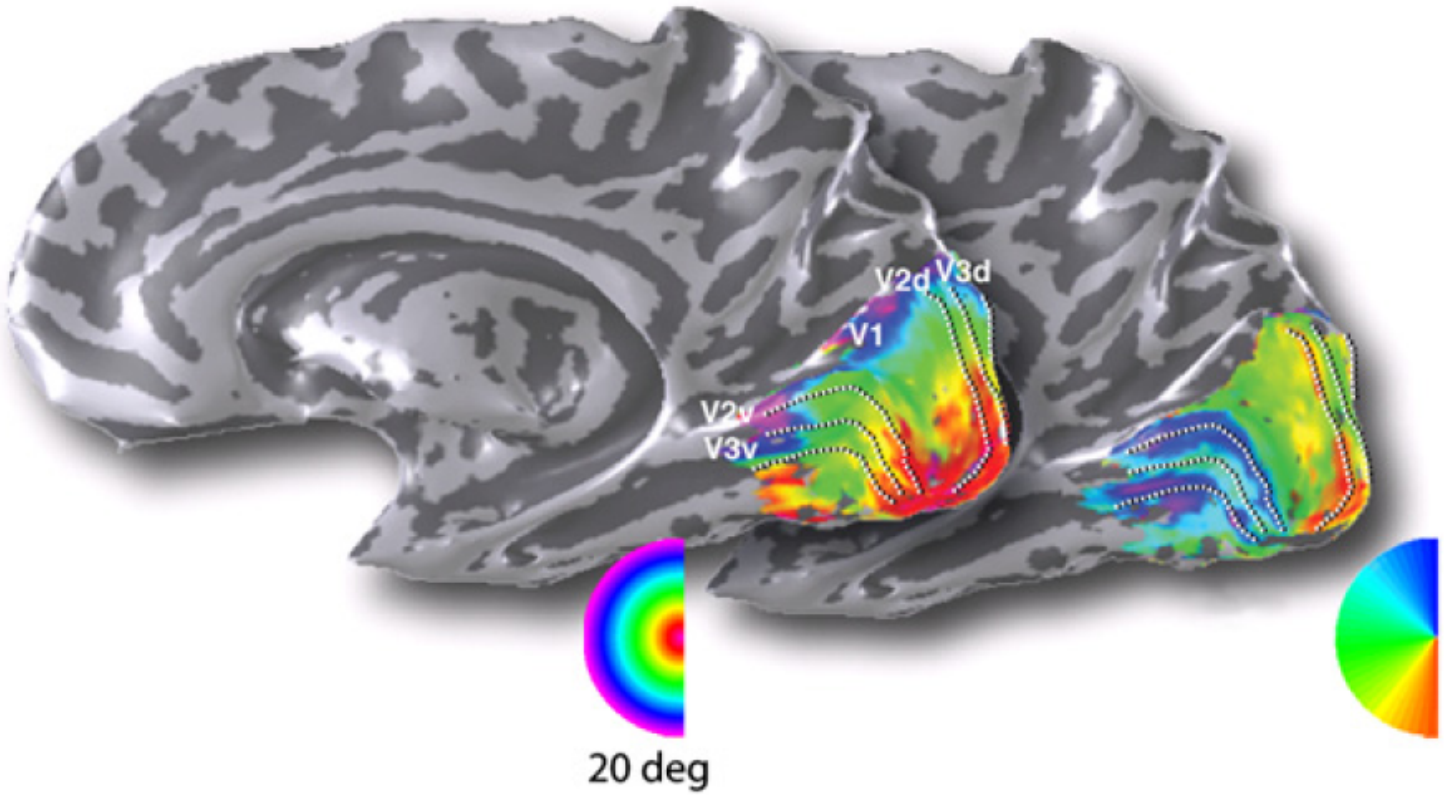
Let us now come to the key discovery of a neuronal implementation of *geometrical structures* structuring the set of contact elements. It is carried out by connections between neurons.

The first fundamental fact is that the retino-geniculo-cortical pathway projects the retina onto $V1$ in a well-behaved way called "retinotopy".

It can be shown experimentally that the retinotopic map is a *conformal* isomorphism.

The following figure shows an fMRI of the retinotopic projection of a (human) visual hemifield onto the corresponding $V1$ hemisphere. Concentric circles and rays are coded by colors.

55. fMRI image



56. Axon guidance and chemotaxis

The genetic control of retinotopy is a fascinating morphogenetic process.

It requires extremely accurate targeting and axon guidance.

A key mechanism is *axon chemotaxis*, which occurs through gradients of chemo-attractive and chemo-repulsive molecules like the "ephrins" with their Eph receptors, the "semaphorins" for the axons of the corpus callosum, "Netrin 1' for the thalamocortical pathways, "slits" for the optic chiasm and the corpus callosum, etc.

The idea of chemoaffinity introduced by Roger Sperry in the 1950s asserts that there is an address system, i.e., molecular tags, distributed in complementary gradients on the axons and their targets, which determine the specificity of the axon connections in topographical maps.

57. Eph / ephrin gradients

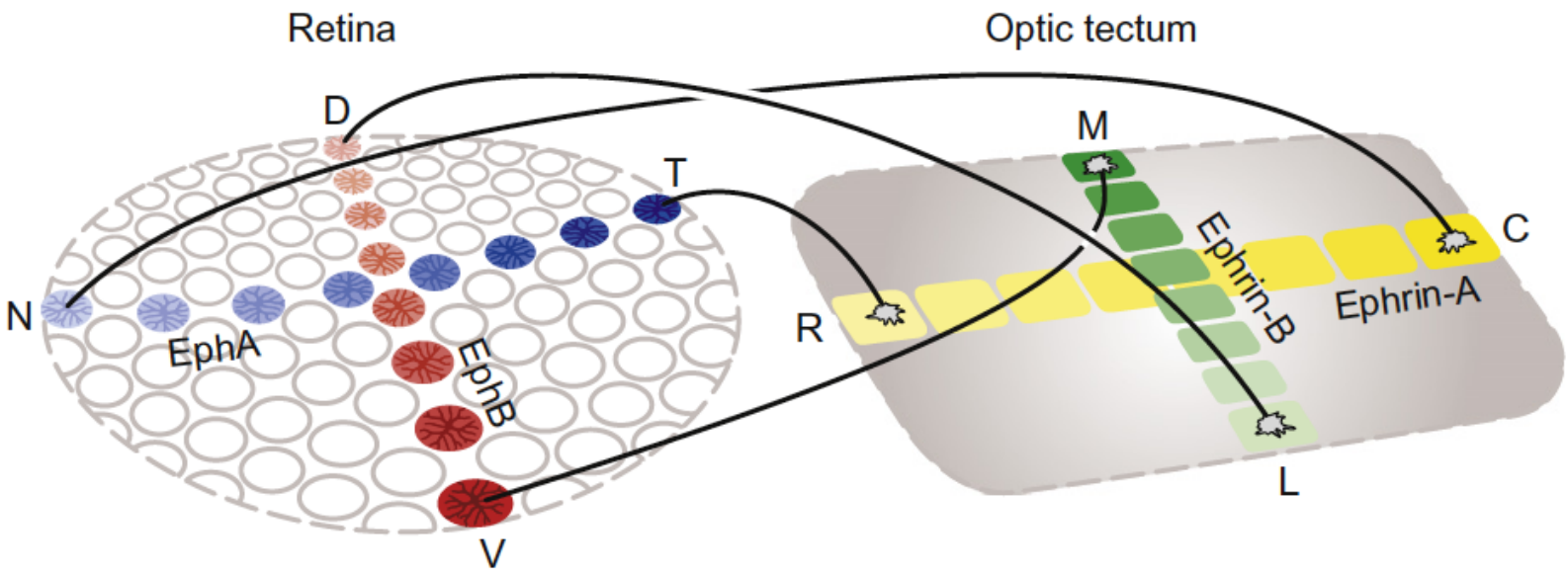
The following figure due to Benjamin Reese schematizes the gradient-controlled dynamics for the retina $\longrightarrow$ tectum projection in the non mammalian brain of the chicken (the optic tectum is the analogous of the superior colliculus in mammals).

It is much simpler than the projection $LGN \longrightarrow V1$ but is already a fine example.

It shows how the increasing naso$\nearrow$temporal ($N \nearrow T$) *EphA* and dorso$\nearrow$ventral ($D \nearrow V$) *EphB* gradients of the ganglion cell layer allows the axon growth cone of these cells to reach a precise position in the tectum defined by the increasing complementary gradients of rostro-caudal ($R \nearrow C$) *ephrin-A* and lateromedial ($L \nearrow M$) *ephrin-B* with opposite *EphA* and *EphB* gradients.

58. Reese's image

The retinotopic map $N \to C$, $T \to R$, $D \to L$, $V \to M$.



59. The gap reality $\longrightarrow$ model

This shows the immense gap that separates genetically controlled neurophysiological reality from mathematical modeling.

Modeling layer 4 of $V1$ as a plane endowed with a differentiable structure and, what's more, a conformal or metric structure, is immensely non-trivial.

60. The functional architecture

We will simplify the situation by assuming that the retinotopic conformal map is simply the identity.

Simple neurons of $V1$ are parametrized by triples (a, p) where $a = (x, y)$ is a position on the retina (identified with $\mathbb{R}^2$) and p is an orientation (modulo π) at a.

So, simple cells of $V1$ constitute a *field of orientations*.

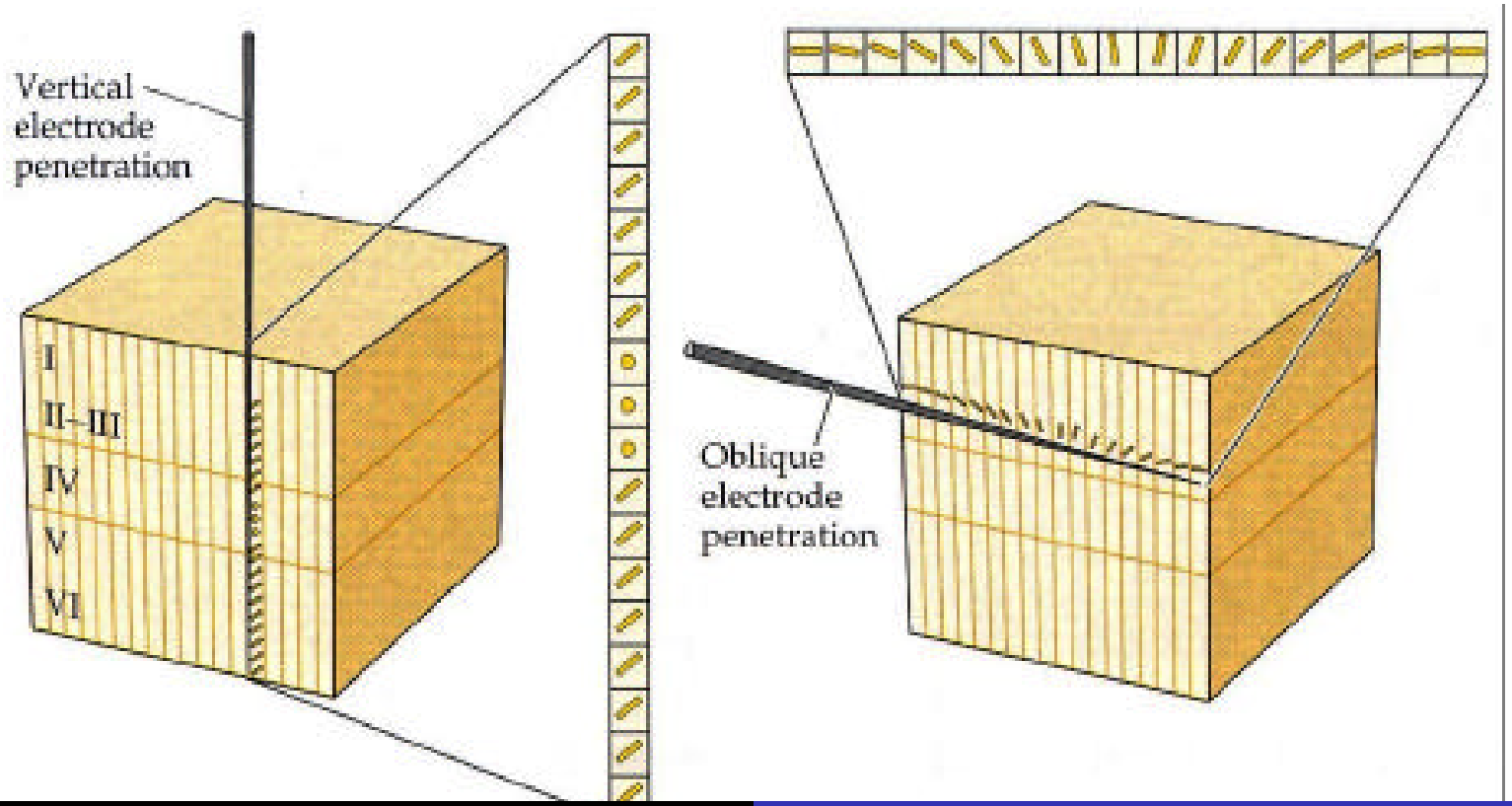
This field is the basis of the "functional architecture" of $V1$.

What could be its structure?

61. H&W discovery

H&W discovery of this structure is schematized in the following figure. It was a great discovery.

Neurons detecting all the orientations p at the same position $a \in \mathbb{R}^2$ constitute an anatomically well delimited small neural module called an "orientation hypercolumn".



62. Orientation hypercolumns

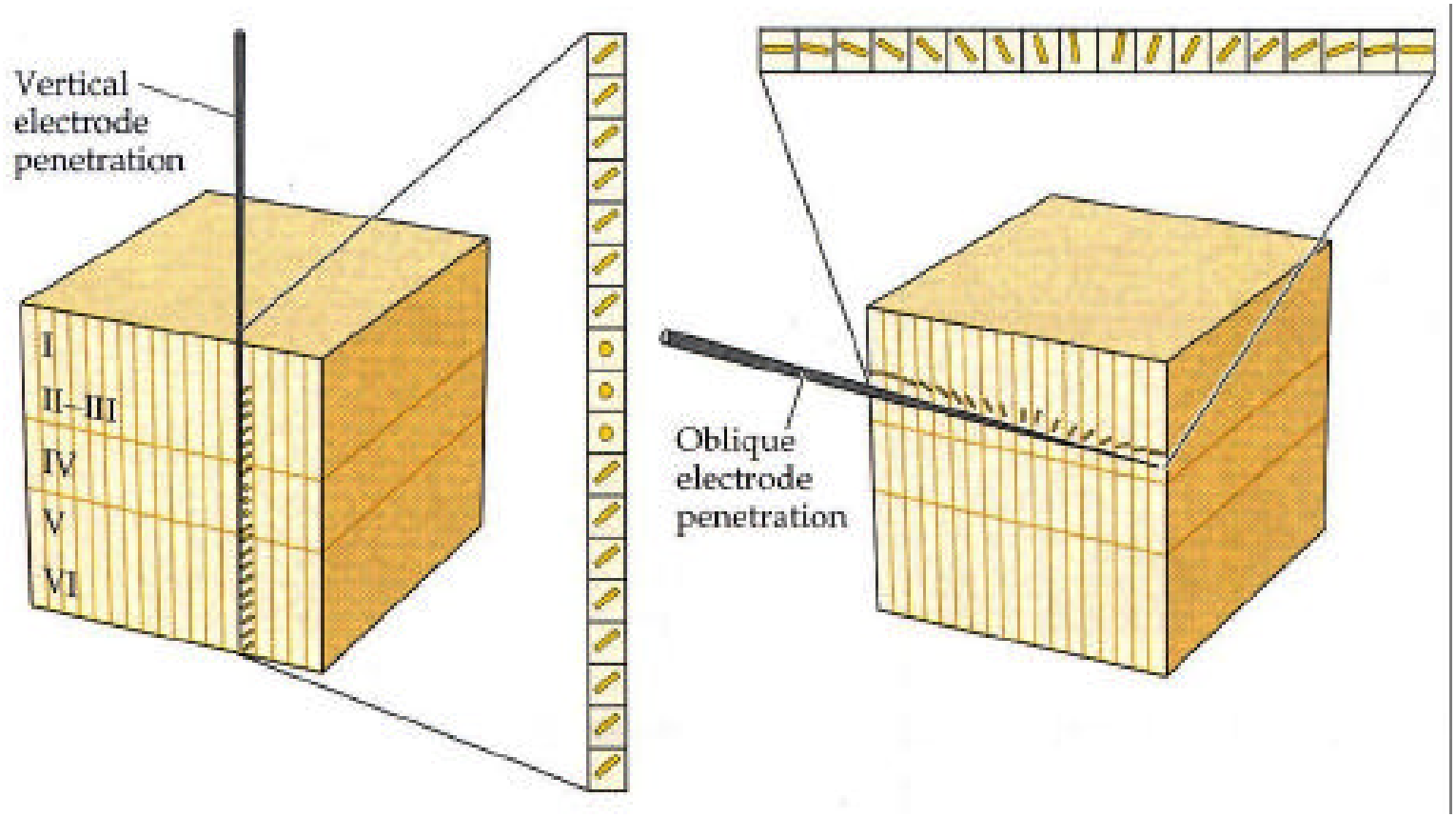
1. Along a "vertical" electrode penetration, *orthogonal* to the cortical surface, the detected contact element (a, p) remains essentially *constant*.

This "vertical" redundancy defines *orientation columns*, which are micromodules of about 20μm.

2. But, along an "oblique" electrode penetration, almost *parallel* to the cortical surface, the prefered orientation p varies from 0° to 180° in steps of about 10°.

This "horizontal" grouping of columns defines *orientation hypercolumns*, which are neuronal micromodules from 200μm to 1mm wide.

63. H&W discovery (*bis repetita*)



64. Braitenberg abduction

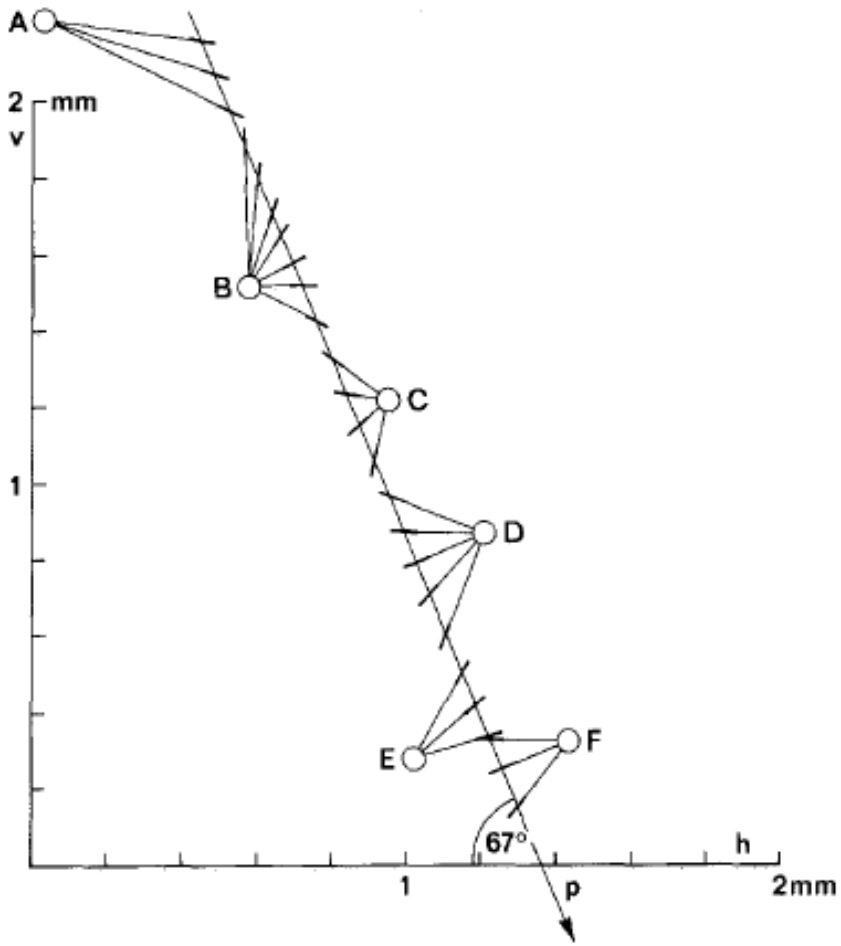
The first global reconstruction of an orientation field from the sparse local data provided by electrodes was *infered* abductively in 1979 by Valentino and Carla Braitenberg.

This was long before the introduction of modern *in vivo* optical imaging techniques. They claimed:

> *"We believe that the most natural explanation of the facts observed would be in terms of orientations arranged with circular symmetry around centers, either radially or along concentric circles."*

65. Orientation centers

The introduction of centers of orientation explained the observed *inversion of chirality* of orientations along a linear penetration.



66. Swindale's abduction

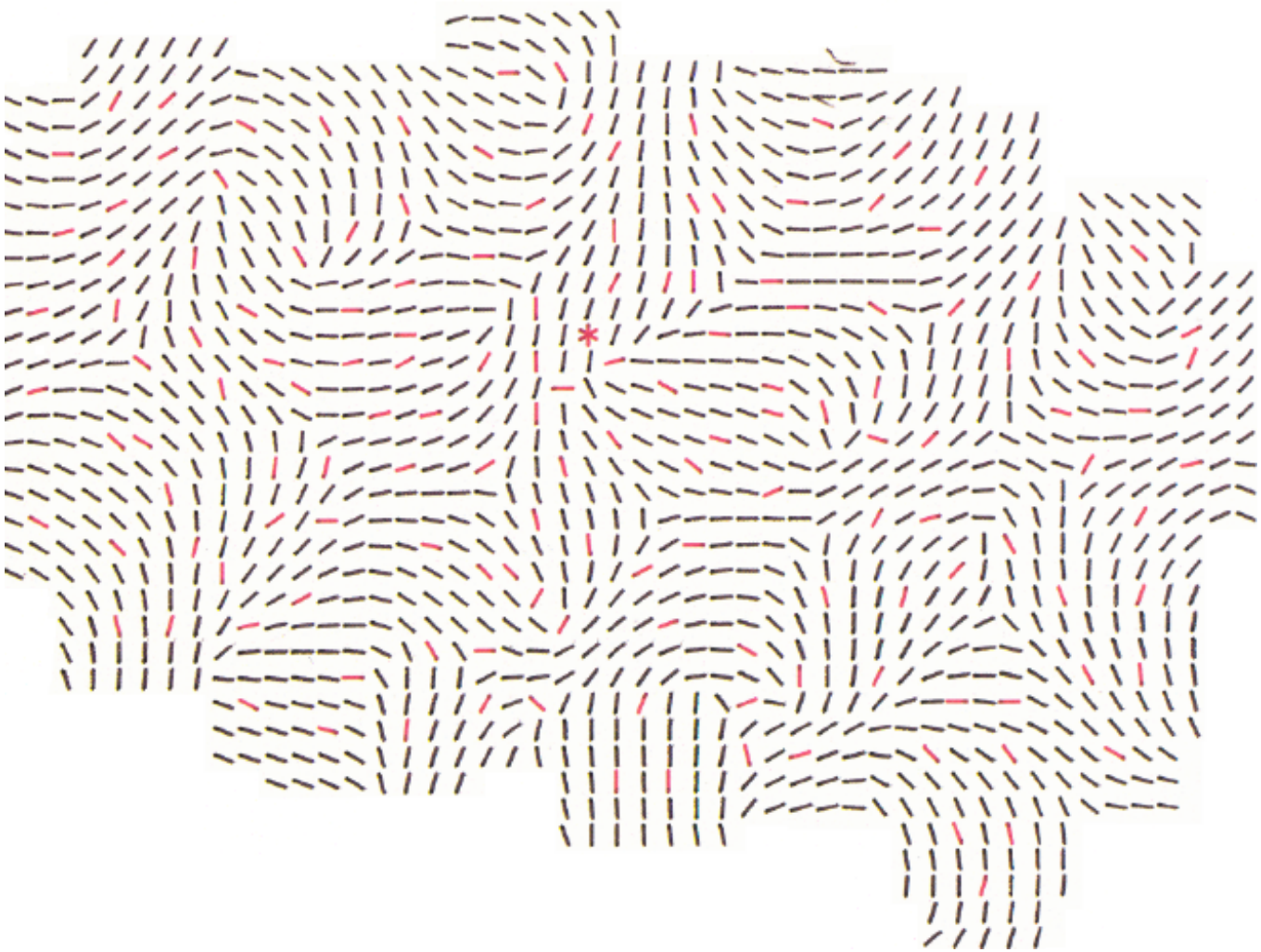
After Braitenberg, in an astonishing 1987 paper (still before the advent of optical imaging techniques), Nicholas Swindale reconstructed the "spatial layout" of the orientation map.

What are now called *pinwheels*.

He thus confirmed Braitenberg's abduction.

Using Fourier transform techniques, he succeeded in *interpolating* between the prefered orientations measured at the different sites and reconstructed the "fine grained" map shown in the following figure.

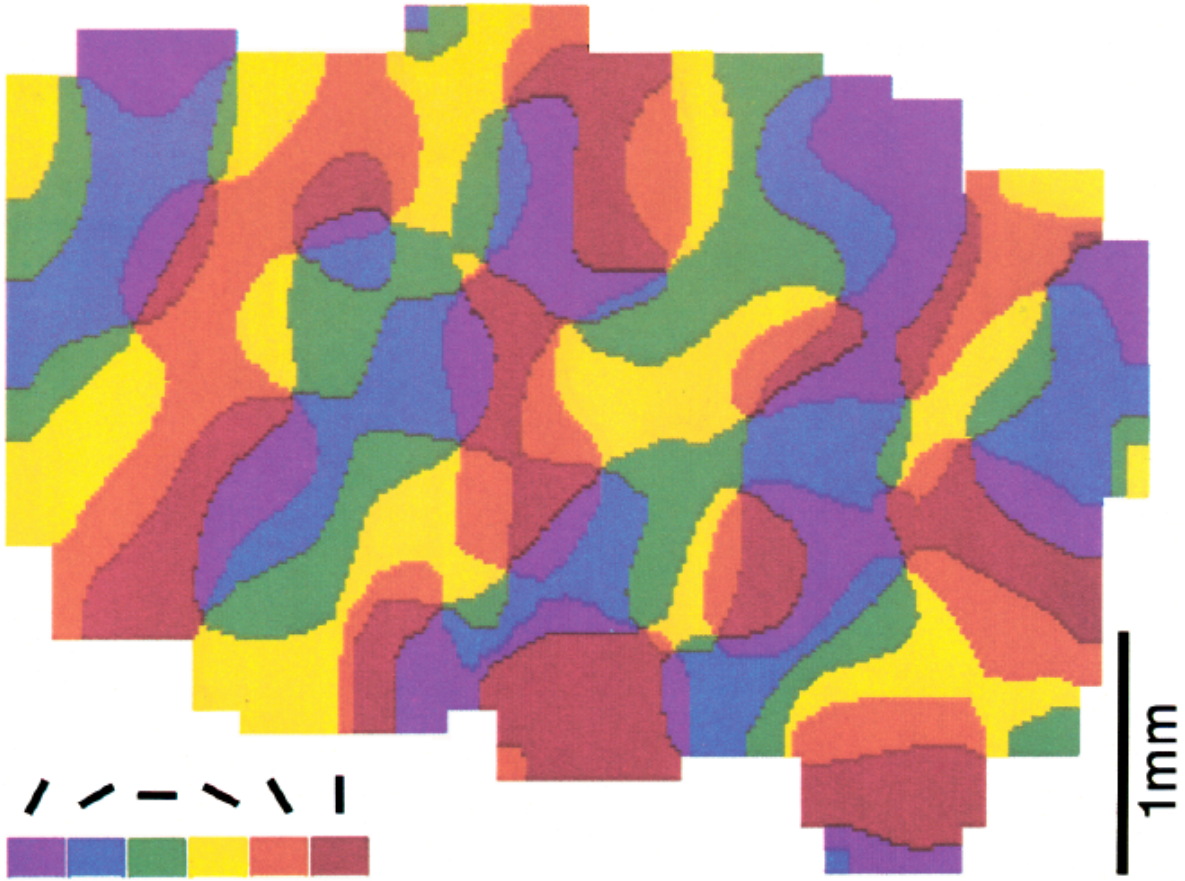
67. Swindale's image 1



68. Swindale's image 2

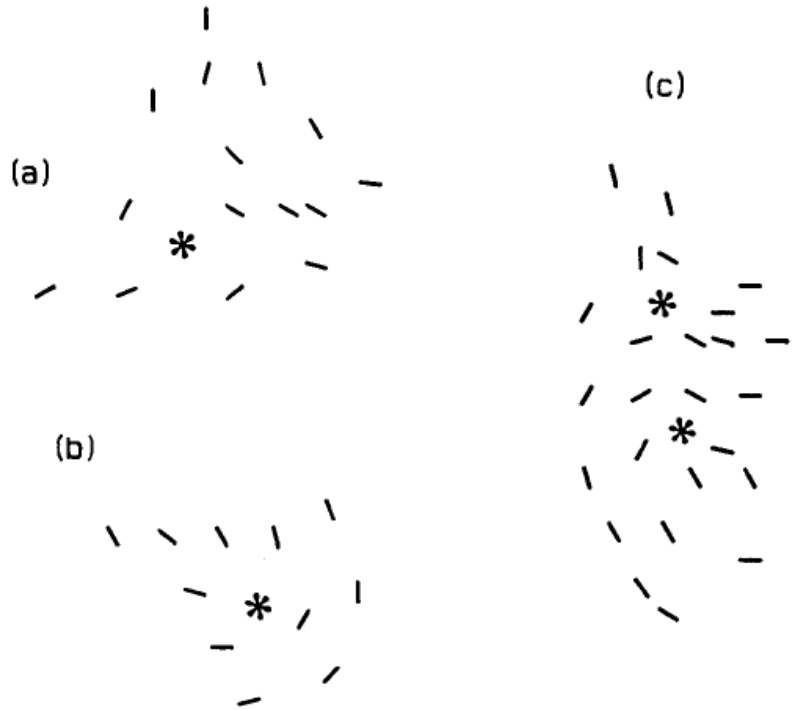
Using a color code for directions, he got an orientation map.

This is a theoretical reconstruction and not an empirical observation.



69. Swindale's image 3

He even reconstructed the possible singularities of the orientation field: they can be end points or triple points.



70. In vivo optical imaging. I

Braitenberg's and Swindale's abductions have been strikingly confirmed in the 1990s by brain imagery.

In the early 90s, Amiram Grinvald and Tobias Bonhöffer introduced new revolutionary techniques of "in vivo optical imaging based on activity-dependent intrinsic signals".

They used the fact that the metabolic activity of cortical layers change their optical properties (differential absorption of oxyhemoglobin or deoxyhemoglobin whose fluorescence is an index of the local depolarisation of neurons).

This enables to acquire in vivo images of the activity of the superficial cortical layers.

71. In vivo optical imaging, II

As Kenichi Ohki and Clay Reid have pointed out,

> *"optical imaging revolutionized the study of functional architecture by showing the overall geometry of functional maps."*

The scale of observation is a "meso"-scale. For a true "micro"-scale observation at the level of single neurons, you need more recent techniques such as "two-photon confocal microscopy" (Ohki 2006).

72. Orientation maps

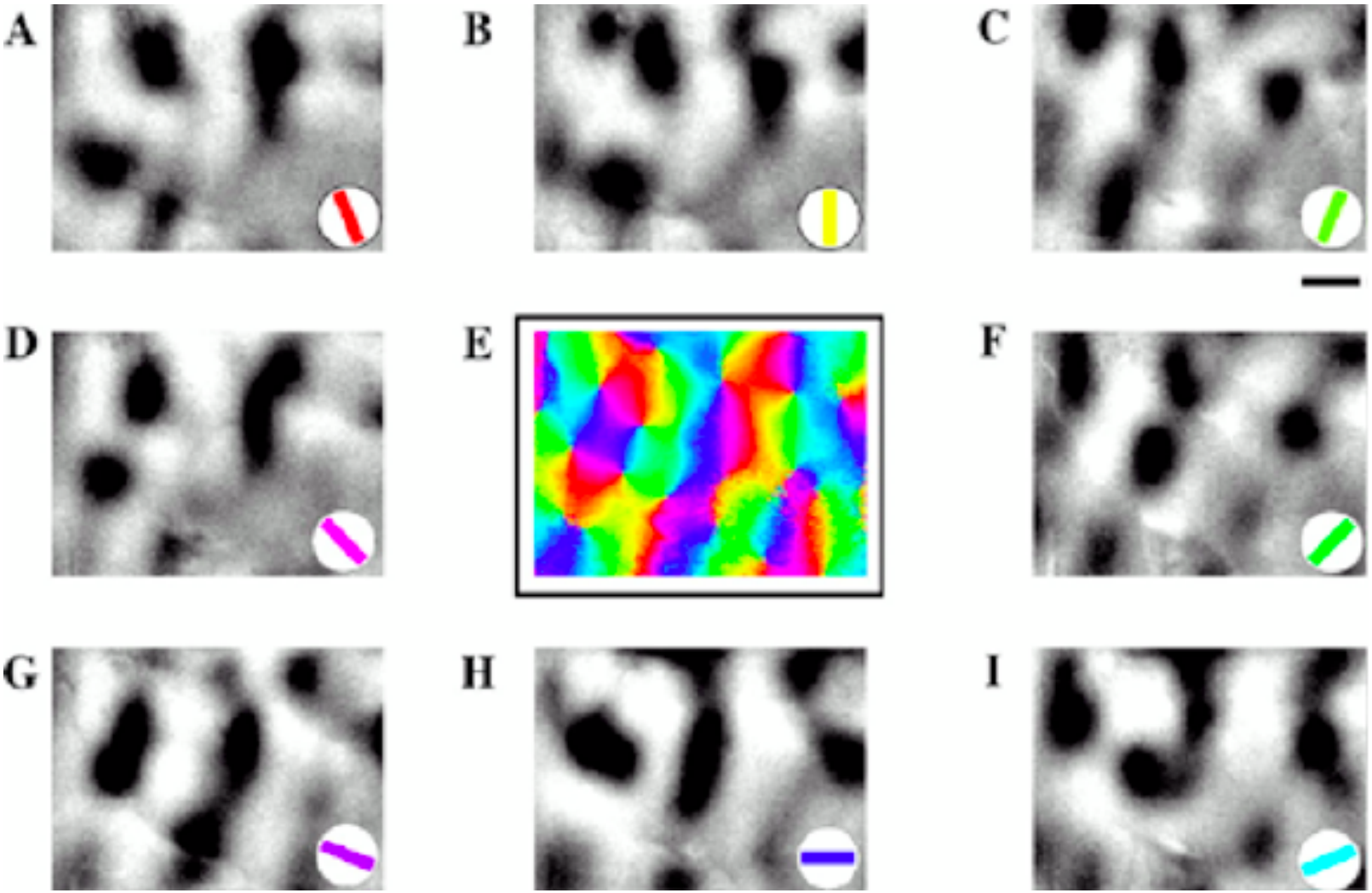
Here is the functional architecture of the area $V1$ of a tree-shrew (tupaya) obtained by "in vivo optical imaging" (William Bosking with David Fitzpatrick's team at Duke University).

They used "grattings", that is large grids of parallel dark stripes translated in the visual field.

For every orientation (coded by the bottom-right color) they got a global map of activity (dark = active).

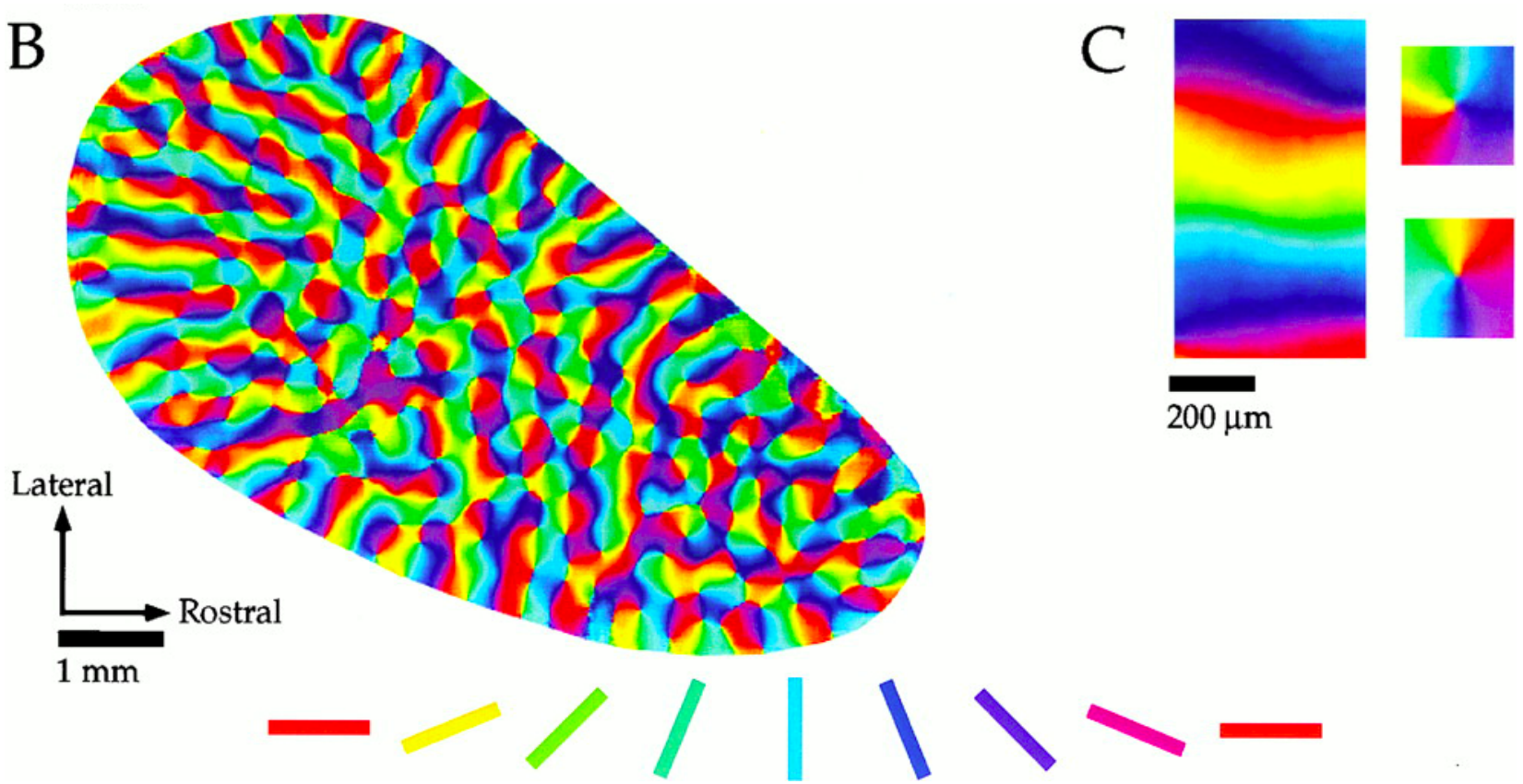
This is an empirical observation and not a theoretical reconstruction.

73. Orientation maps. Image



74. Pinwheels

When you code the orientations by colors and you superpose the maps, you get Braitenberg-Swindale *pinwheels* !

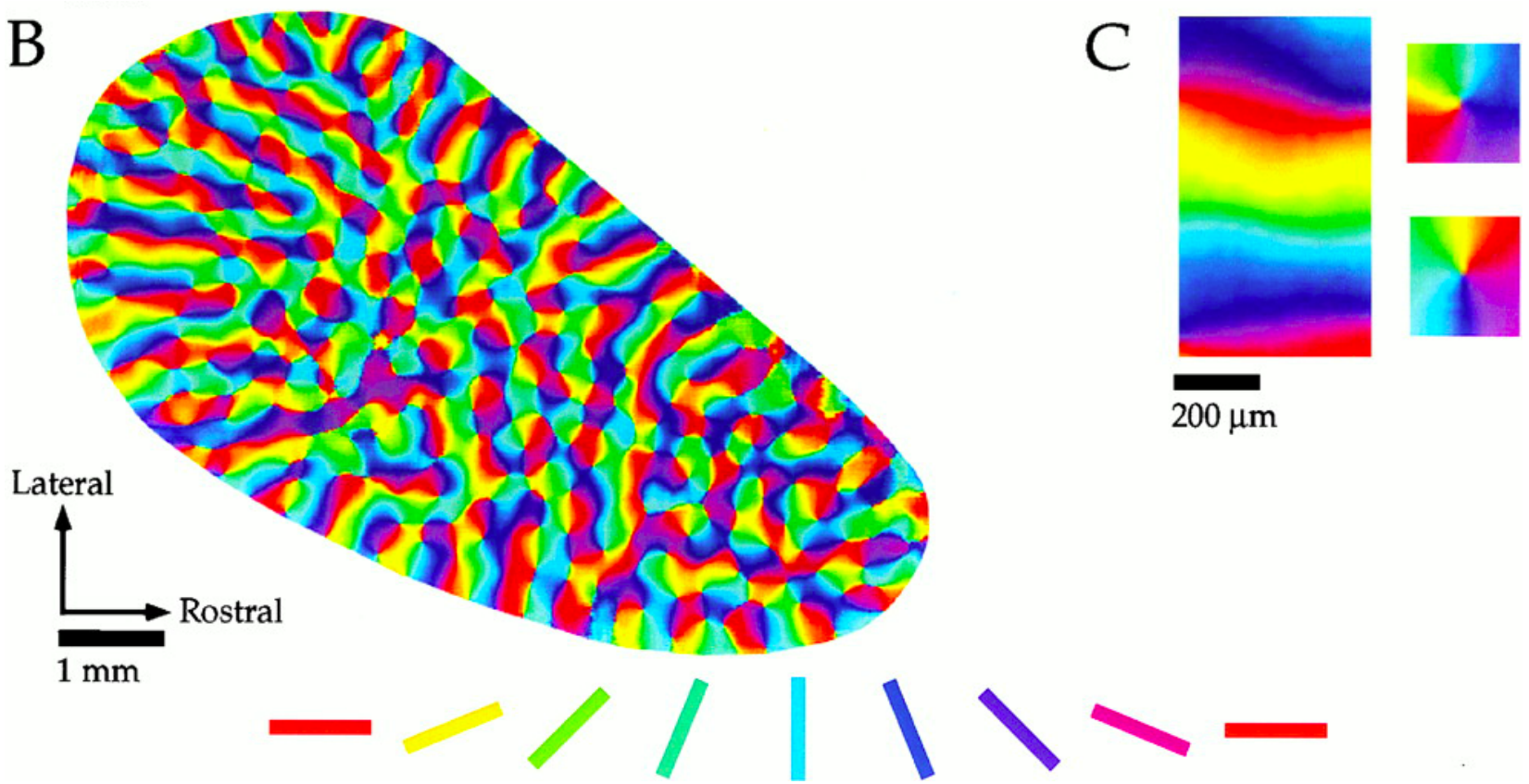


75. Pinwheels' structure

- The plane is $V1$,
- A colored point represents the mean of a small group of real neurons (meso-scale).
- Colors code for the preferred orientation at each point.
- The field of isochromatic lines (i.e. iso-orientation lines) is organized by a lattice of *singular points* (pinwheels) where all orientations meet (distant about 1200μm in cats and about 600μm in primates).
- There exist a "mesh" of the lattice of pinwheels (a sort of characteristic length).
- Pinwheels have a chirality.
- Adjacent pinwheels have opposed chirality.

76. Pinwheels (*bis repetita*)

Right: two pinwheels of opposed chiralities.

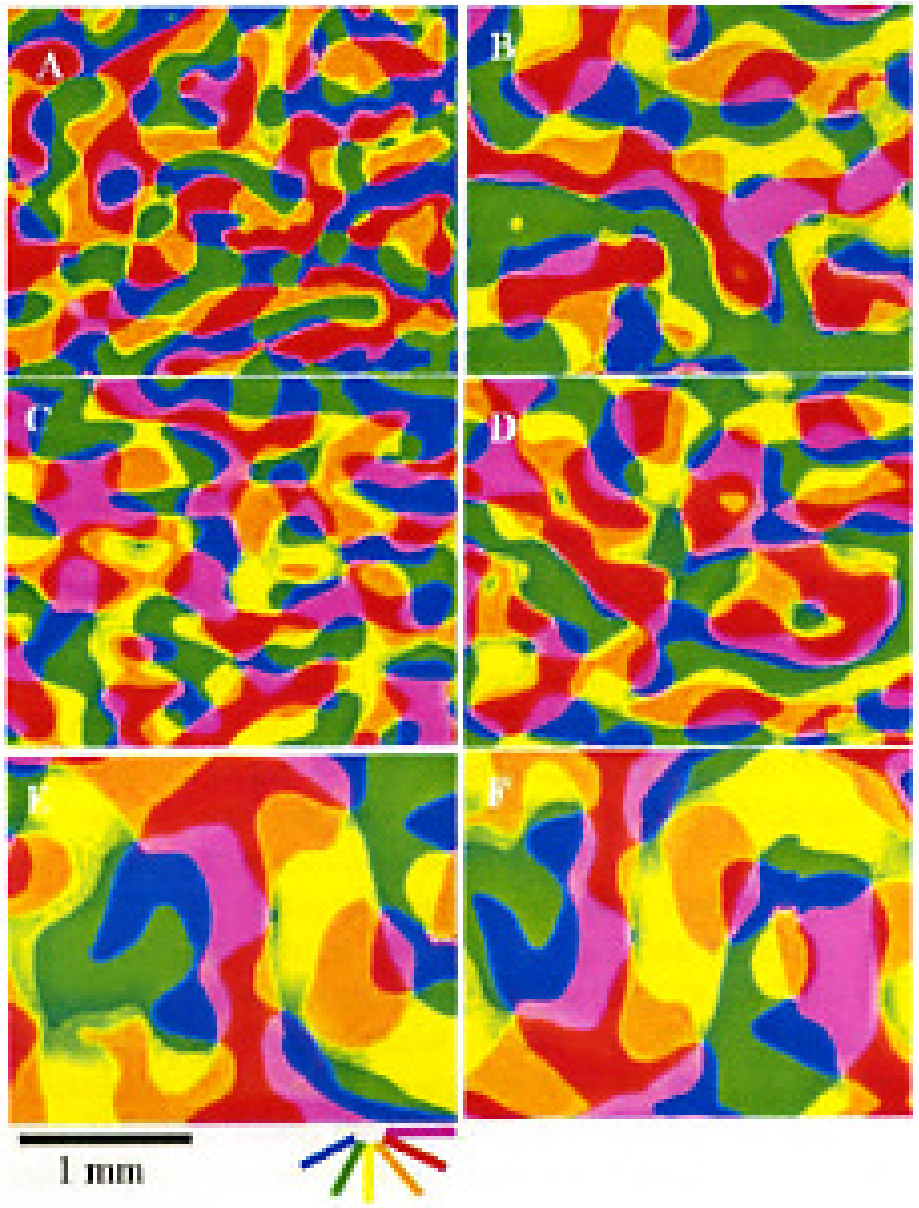


77. Interspecific functional architecture

A pinwheel organisation can be found in many species: cat, primate (marmoset), tupaya (tree shrew), prosimian Bush Baby, tawny owl, etc.

The following figure shows pinwheels in the $V1$ and $V2$ areas of the cat, (A) and (B), the marmoset (C) and (D), and the tawny owl (E) and (F).

78. Pinwheels in different species



79. End points and triple points

If θ is the angle of a pinwheel ray, the associated orientation varies as $\alpha \pm \theta/2$ according to the chirality.

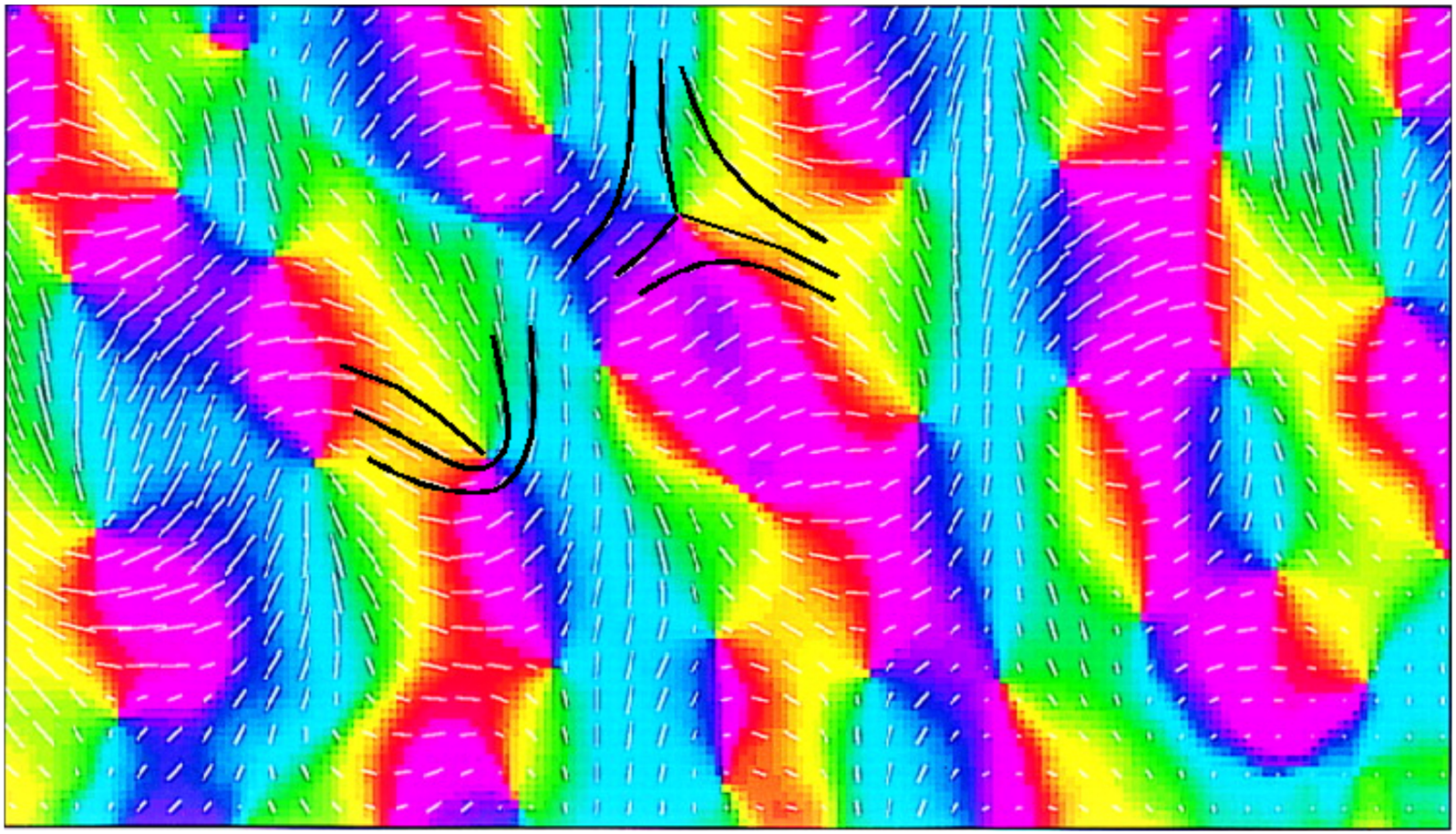
In the following picture due to Shmuel (cat's area 17), orientations are coded by colors but are also represented by small white segments.

We observe very well the two types of generic singularities of 1D foliations in the plane:

- end points ($\alpha + \theta/2 = \theta$);
- triple points ($\alpha - \theta/2 = \theta$).

They correspond to the two possible chiralities of pinwheels.

80. Shmuel's orientation map



81. Wolf-Geisel models

There exist beautiful models of pinwheels. They are analogous to *dislocations of phase fields* in optics (see Michael Berry and Mark Dennis' works).

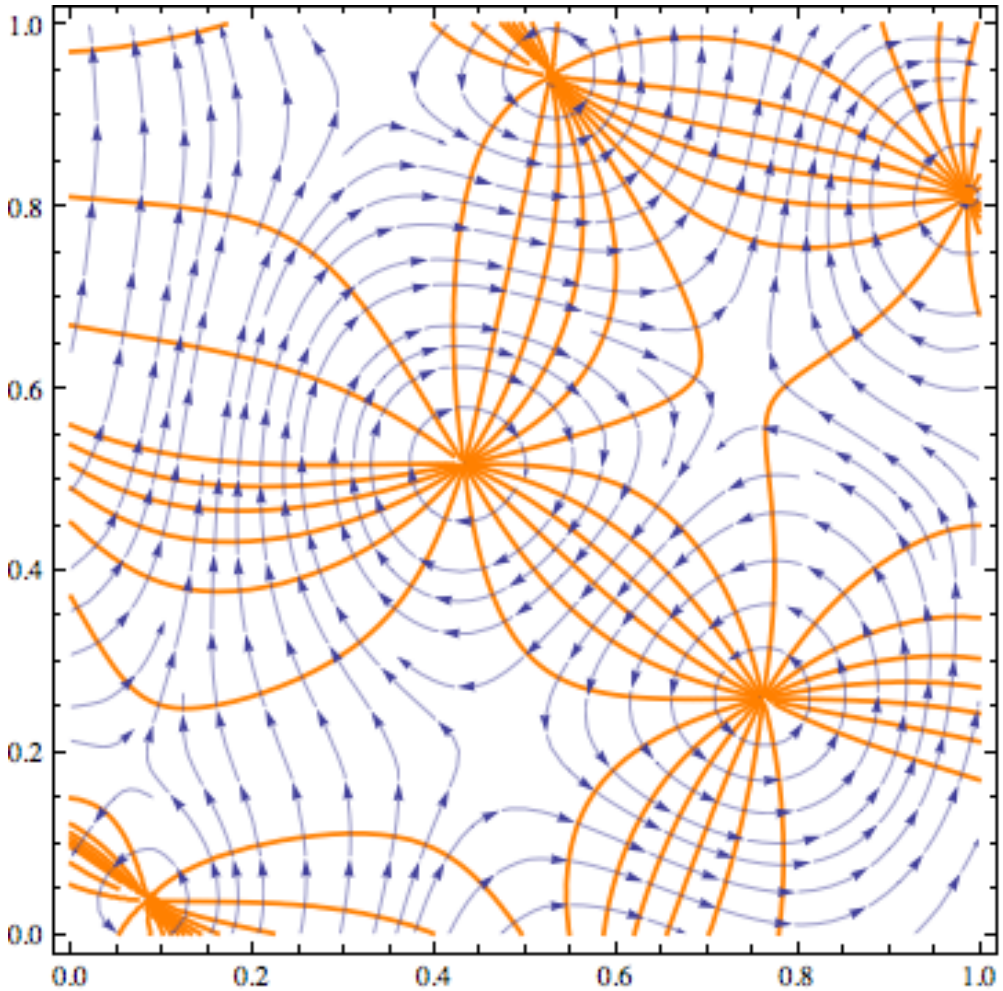
Fred Wolf and Theo Geisel proposed such models using a *complex valued* field (of the real variables $a = (x, y)$ but not necessarily of the complex variable $a + iy$):

$$Z : \mathbb{R}^2 \to \mathbb{C},\ a = \rho e^{i\theta} \mapsto r(a)\, e^{i\varphi(a)} = X(a) + iY(a)$$

where $a = (\rho, \theta)$ are polar coordinates, where the spatial phase $\varphi(a)$ codes the orientation ($\varphi(a)$ varies as $\pm\theta/2$ near singular points) and the modulus $r(a)$ codes the orientation selectivity.

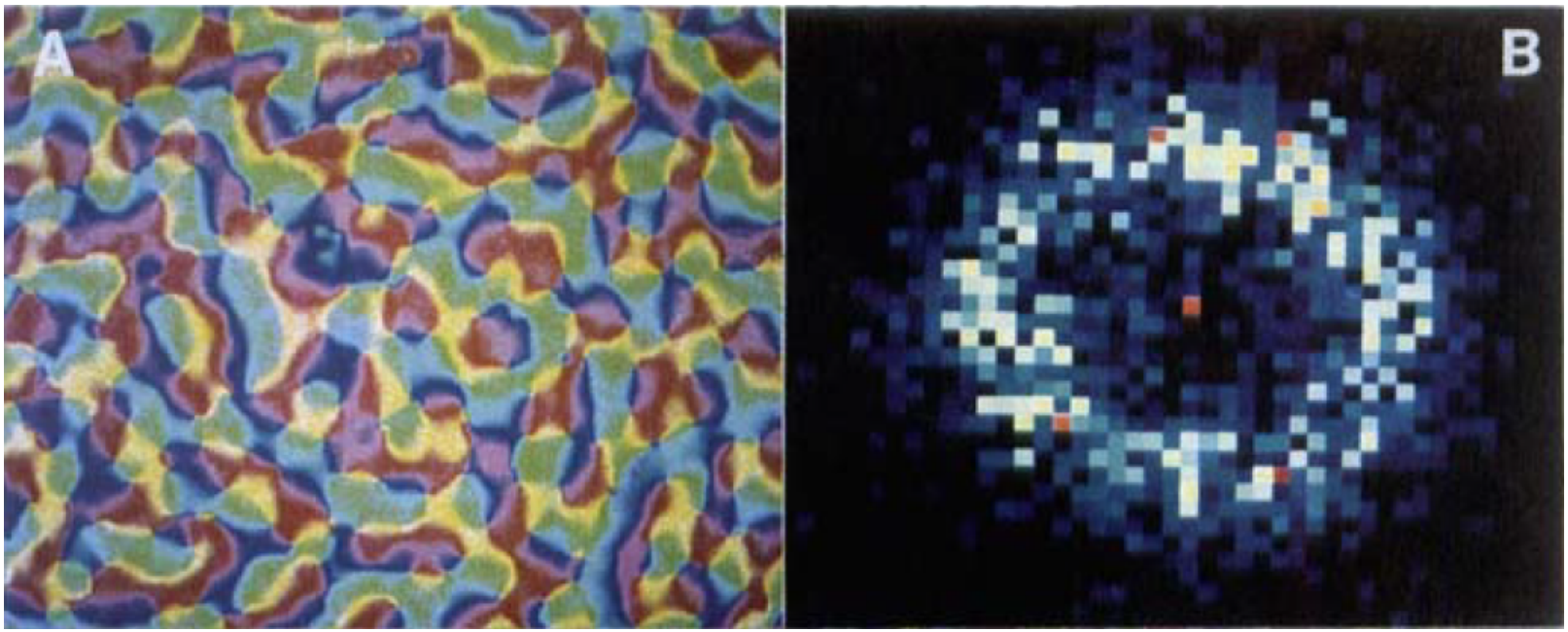
82. Phase fields

Here is an example of a phase field (using values of one of Berry's examples).



83. Periodicity of pinwheels

Orientation maps (image A: left) manifest a periodicity. The *power spectrum* (the Fourier transform of the autocorrelation function of the map) is concentrated on a *ring* of average radius $k_0 = \frac{2\pi}{\Lambda_0}$ (image B: right)



This means that, ideally, such phase fields with a characteristic length are superpositions of plane waves with random phases but sharing *the same* wave number k (k is the modulus of the wave vector κ).

They are solutions of the Helmholtz equation:

$$\Delta Z + k^2 Z = 0 \ .$$

85. Statistics of pinwheels and Gaussian fields

The statistics of singularities of Helmholtz phase fields is a beautiful subject. Under some simplifying hypotheses, they can be modeled by *Gaussian fields*.

In that case, you can prove that the *density of singularities* is $d = \frac{k^2}{4\pi} = \frac{\pi}{\Lambda^2}$ with $\Lambda = \frac{2\pi}{k}$.

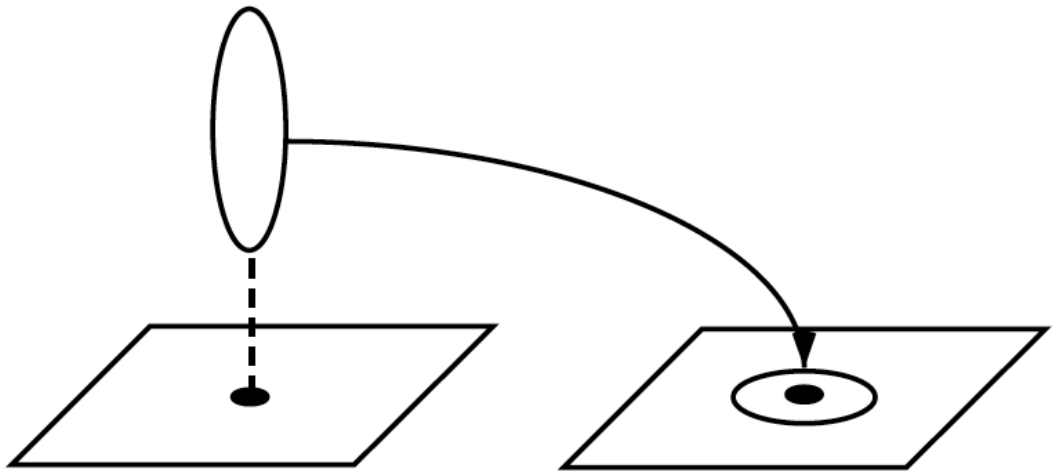
It is remarkable that this density $\frac{\pi}{\Lambda^2}$ was also found *empirically* in pinwheels map.

86. Pinwheels as blowing-up

In one way, pinwheels are *blowing-up* of points a_i and the orientation field is the closure of a section σ of the fiber bundle $\pi : \mathbb{V} = \mathbb{R}^2 \times \mathbb{P}^1 \to \mathbb{R}^2$ defined over the open subset $\mathbb{R}^2 - \{a_i\}$.

Over the a_i the closure of σ is the "exceptionnal" fiber $\mathbb{P}^1_{a_i}$.

These exceptional fibers $\mathbb{P}^1_{a_i}$ are "contracted" and "folded" onto small neighborhoods of the base points a_i.



87. V1 as fiber bundle

There is therefore a $3D \longrightarrow 2D$ dimensional collapse : an orientation map is, in a way, a geometric object of "intermediate" dimension between 2 and 3.

At the limit, when all the points of the base plane $\mathbb{R}^2$ are blown-up in parallel, we get the fiber bundle $\pi : \mathbb{V} = \mathbb{R}^2 \times \mathbb{P}^1 \longrightarrow \mathbb{R}^2$.

So $\mathbb{V}$ can be considered as an *idealized continuous approximation* of the concrete $V1$.

88. Geometrical interpretation

The fiber bundle $\pi : \mathbb{V} = \mathbb{R}^2 \times \mathbb{P}^1 \to \mathbb{R}^2$, which is an abstract *ideal* 3D geometric structure, is *materially* implemented in a 2D neural hardware.

This is mathematically trivial but not at all neurophysiologically trivial as a result of biological evolution.

89. Fiber bundles and "engrafted" variables

By the way, the *intuition* (not the mathematical concept) of a fiber bundle was explicit in Hubel with his key concept of "engrafted" variables:

> "What the cortex does is map not just two but many variables on its two-dimensional surface. It does so by selecting as the basic parameters the two variables that specify the visual field coordinates (...), and on this map it engrafts other variables, such as orientation..."

90. Contact elements and 1-jets

Now, the fiber bundle $\mathbb{V} = \mathbb{R}^2 \times \mathbb{P}^1$ is $\mathbb{R}^2 \times \mathbb{S}^1$ modulo π and $\mathbb{R}^2 \times \mathbb{S}^1$ has an affine open subset which is the space $\mathbb{V}_J$ of *1-jets of smooth plane curves*.

As 1-jets can be processed by neural "point-like processors", we get a first answer to our initial question: "How the visual brain can be a neural geometric engine?"

Low dimensional *jet spaces* are neurally implemented and jet spaces are "prolongations" (in Cartan's sense) dedicated to integrability.

91. The contact structure of 1-jets

Let us recall the contact structure of the space of 1-jets.

A skew curve

$$\Gamma = v(s) = (a(s), p(s)) = (x(s), y(s), p(s))$$

in $\mathbb{V}_J$ is the *Legendrian lift* of its projection $\gamma = a(s)$ onto the base plane $\mathbb{R}^2$

- iff $p(s)$ is the tangent $p = dy/dx$ to the curve γ at the point $a(s)$,
- iff it is an *integral curve* of the *contact structure* $\mathcal{C} = ker(\omega)$ of $\mathbb{V}_J$, where ω is the 1-form

$$\omega = dy - pdx$$

92. The distribution of contact planes

The distribution $\mathcal{C}$ of contact tangent planes is maximally non integrable since the 3-form

$$\omega \wedge d\omega = (-pdx + dy) \wedge dx \wedge dp = -dx \wedge dy \wedge dp \ .$$

is a volume form, which is the opposite of the Frobenius integrability condition $\omega \wedge d\omega = 0$.

So, even if there exists a lot of integral *curves* of $\mathcal{C}$ (Legendrian lifts), there exists no integral *surface*.

93. The polarized Heisenberg group

Let's remember that the contact structure $\mathcal{C}$ is left-invariant for a group law making $\mathbb{V}_J$ isomorphic to the *polarized Heisenberg group* $\mathbb{H}_{pol}$.

$$(x, y, p).(x', y', p') = (x + x', y + y' + px', p + p') \ .$$

Its Lie algebra is generated by the basis of left-invariant fields $X_1 = \frac{\partial}{\partial x} + p\frac{\partial}{\partial y} = (1, p, 0)$ and $X_2 = \frac{\partial}{\partial p} = (0, 0, 1)$ with $[X_1, X_2] = (0, -1, 0) = -\frac{\partial}{\partial y} = -X_3$ (the other brackets $= 0$).

The basis $\{X_1, X_2\}$ of the distribution $\mathcal{C}$ is *bracket generating* (*i.e.* Lie-generates the whole tangent bundle $T\mathbb{V}_J$) (Hörmander condition).

$\mathbb{V}_J = \mathbb{H}_{pol}$ is a *nilpotent* group of step 2 (a Carnot group).

94. The contact structure of $SE(2)$

This can be generalized to the Euclidean group $SE(2)$.

The contact form of $SE(2)$ is

$$\omega_S = \cos(\theta)\, dy - \sin(\theta)\, dx$$

The contact planes are spanned by the tangent vectors $X_1 = \cos(\theta)\frac{\partial}{\partial x} + \sin(\theta)\frac{\partial}{\partial y}$ and $X_2 = \frac{\partial}{\partial \theta}$ with Lie bracket $[X_1, X_2] = \sin(\theta)\frac{\partial}{\partial x} - \cos(\theta)\frac{\partial}{\partial y} = -X_3$.

95. The two models

Contrary to the polarised Heisenberg case, the X_j constitute an Euclidean orthonormal basis and are therefore more natural.

The distribution $\mathcal{C}$ of contact planes is still bracket generating (Hörmander condition). But $SE(2)$ is no longer nilpotent. The Carnot group $\mathbb{V}_J = \mathbb{H}_{pol}$ is its "*tangent cone*", its "nilpotentisation".

96. Neural contact structure

The very key point, which is another striking *experimental* discovery, is that the contact structure of $\mathbb{V}$ is *implemented* in a *specific class* of neural connections.

Orientation hypercolumns correspond to the "vertical" retino-geniculo-cortical connectivity.

But cortical neurons of $V1$ are also connected by "horizontal" cortico-cortical connections *inside the cortical layer itself*.

They are long-ranged (up to 6-8 mm), excitatory, slow (about $0.2\,\mathrm{m}\,/\,\mathrm{s}$) and distributed in a very anisotropic and "patchy" way.

97. Necessity of a parallel transport

Such a second system of connections is necessary to explain perceptual geometry.

The "vertical" retinotopic structure is not sufficient.

To implement a *global* coherence, the visual system must be able to *compare* two retinotopically neighboring orientation hypercolumns P_a and P_b over two different base points a and b.

It is a problem of *parallel transport* that requires long-range "horizontal" cortico-cortical connections.

98. Horizontal cortico-cortical connections

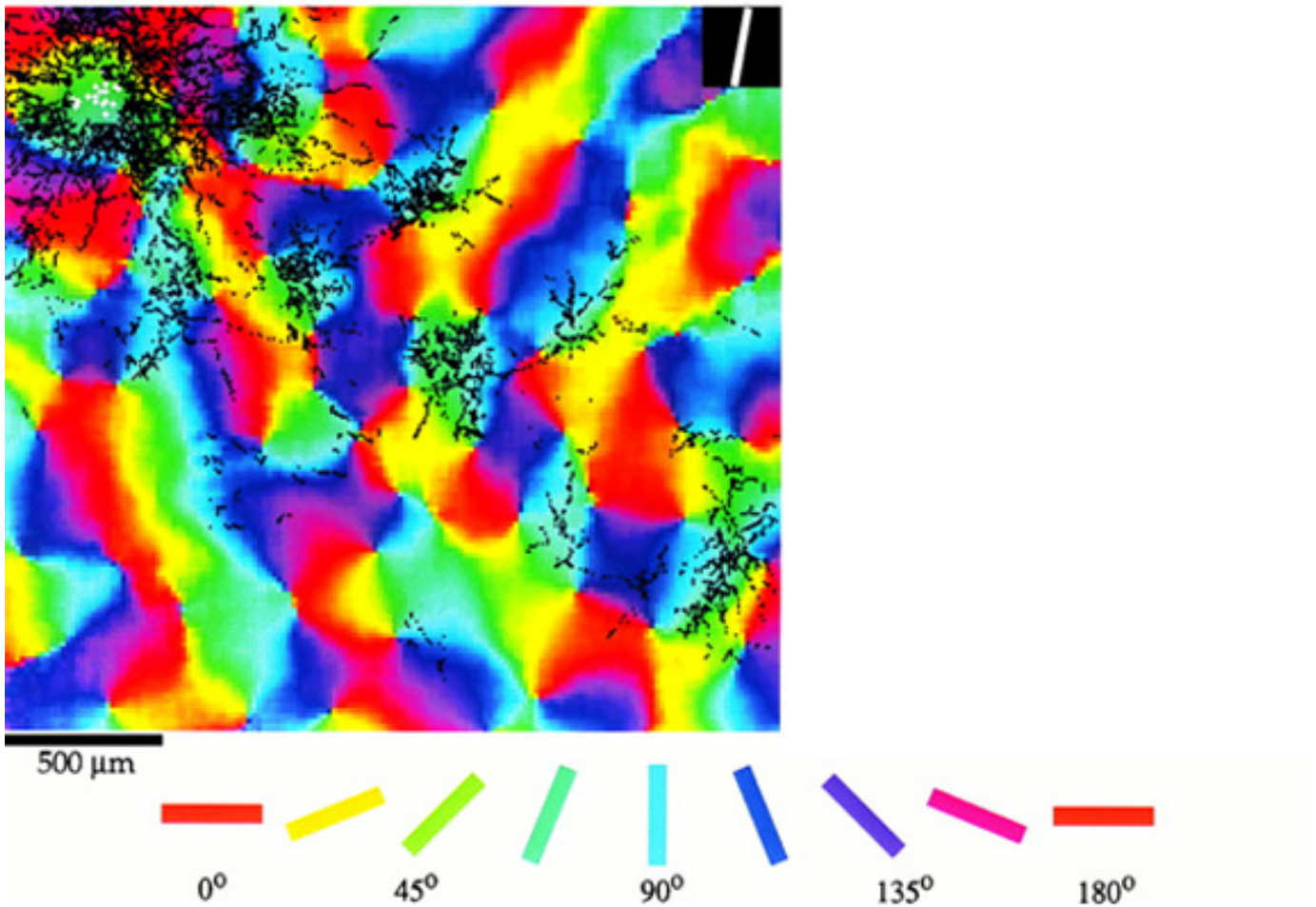
The following image (also due to Bosking *et al.*) shows how a marker (biocytin) locally injected in a zone of about $100\mu m$ of the layer $2/3$ of $V1$ of a tupaya (tree shrew) diffuses along horizontal connections (black marks) in a selective, "patchy", anisotropic way.

Short-ranged diffusion is isotropic and corresponds to *intra*-hypercolumnar inhibitory connections.

On the contrary, long-ranged diffusion is highly anisotropic, and corresponds to excitatory *inter*-hypercolumnar connections

99. Biocytin diffusion

The injection site is upper-left in a blue-green domain.

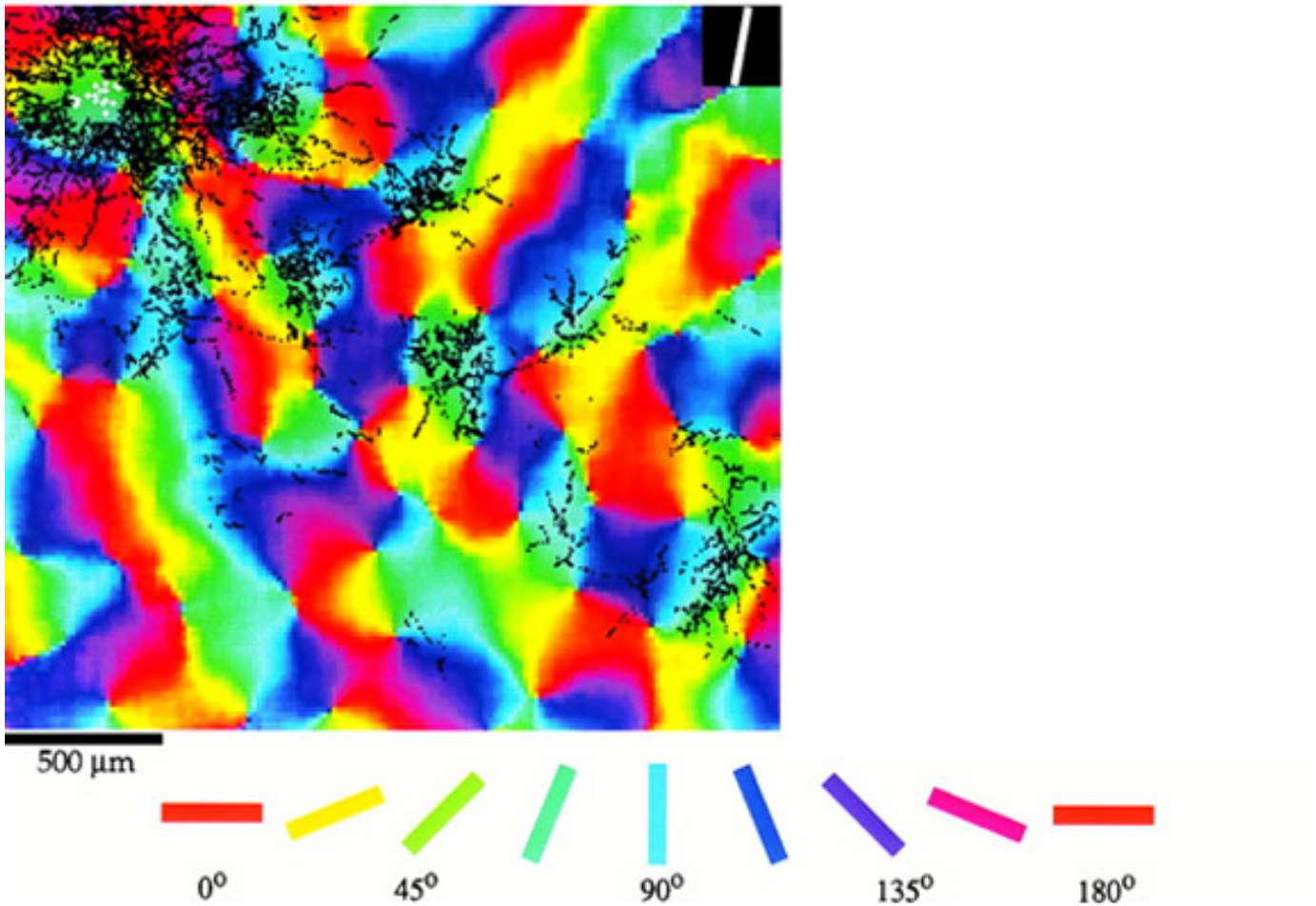


100. Parallelism and co-axiality

There are two main results:

1. The marked axons and synaptic buttons cluster in domains of the *same* blue-green color (same orientation) as the injection site, which means that horizontal connections connect neurons detecting approximately parallel orientation and therefore implement neurally a *parallel transport*.
2. Furthermore, the striking global clustering along the top-left $\longrightarrow$ bottom-right diagonal shows that *almost all* blue-green cells are *without* any marked connections! This crucial empirical fact means that horizontal connections connect neurons detecting not only almost *parallel* but also almost *aligned* "co-axial" orientations.

101. Biocytin diffusion (*bis repetita*)



102. The key result

"The system of long-range horizontal connections can be summarized as preferentially linking neurons with co-oriented, co-axially aligned receptive fields." (W. Bosking)

This means that a chain of simple neurons (a_i, p_i) is a chain of *"horizontally" connected* simple neurons iff it is a discretization of the Legendrian lift of a not too curved base curve interpolating between the (a_i).

So, this means that,

up to some bound on curvature, the contact structure $\mathcal{C}$ is neurally implemented in $V1$.

103. William Hoffman

Let me underline that the hypothesis that the notions of *jet* and *contact structure* must be involved in a natural way in visual perception was already explicitly formulated in 1989 by William Hoffman in his pioneering paper "The visual cortex is a contact bundle".

It was before "in vivo" optical imaging.

104. Association field

These *neurophysiological* results are corroborated by *psychophysical* experiments on curve integration.

In a breakthrough paper of 1993, David Field, Anthony Hayes and Robert Hess considered approximately aligned segments $v_i = (a_i, p_i)$ embedded in a background of randomly distributed segments lacking any global structure

However, subjects perceive very well a global alignment.

This striking phenomenon of *pop-out* (perceptual *saliency*) is due to a *low-level integration* processing: there exist local neurophysiological *binding* rules which let a global perceptual organization emerge.

We recognize here the "binding hypothesis" explaining saliency by the propagation of coherent synchronized activity.

105. Joint position–orientation constraints

After a lot of experimental measures, Field, Hayes and Hess concluded that the pop-out comes from a *specific connectivity*, which they called *association field*.

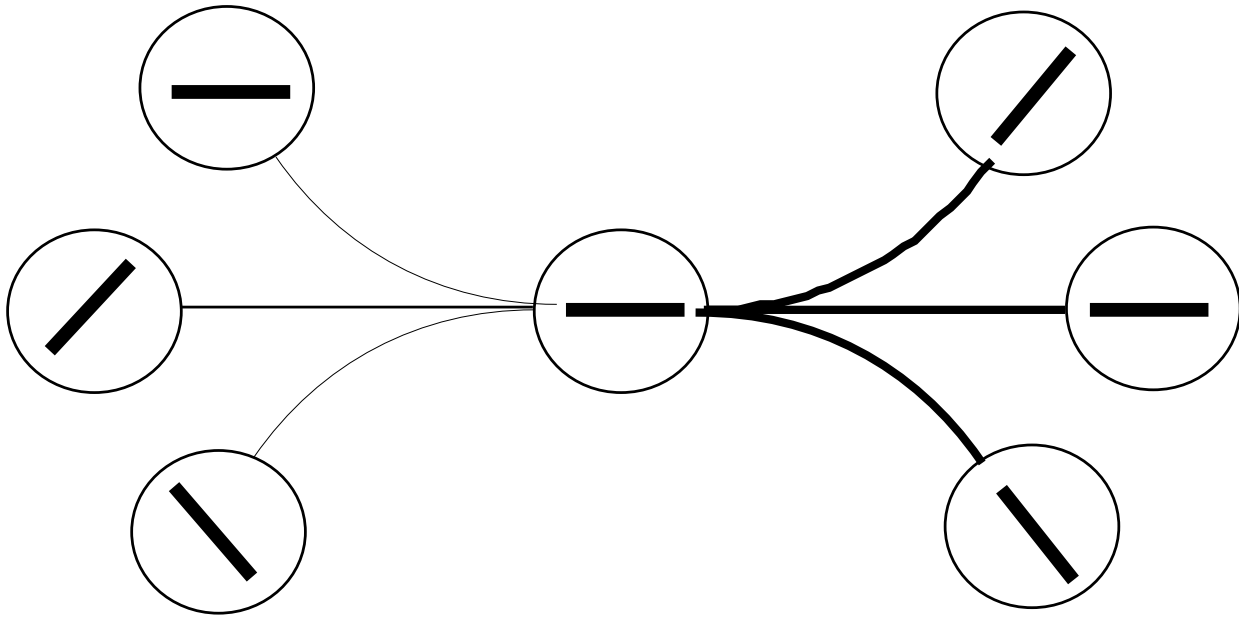
This connectivity is defined by what they called "joint conditions on positions and orientations":

1. two elements (a_1, p_1) and (a_2, p_2) are connected if one can *interpolate* between them a curve γ (which is tangent to p_1 and p_2 at a_1 and a_2, respectively) that is not too curved;
2. otherwise the two elements are not connected.

It has been shown later (in particular by Jean Lorenceau) that the "joint constraints on positions and orientations" correspond to neural "horizontal" cortico-cortical connections.

106. Schema of the association field

Thick lines: connected elements, thin lines: non connected elements.



This specifies one of the fundamental laws of Gestalt theory called "good continuation" ("Gesetz der guten Fortsetzung" or "Gesetz der durchgehenden Linie").

107. Two vocabularies

So, we get a correspondance between two vocabularies, a neurophysiological one and a mathematical one.

simple neurons	(scaled) contact elements (a, p)
retin-geni-corti. retinotopy	base space $\mathbb{R}^2$ of positions a
basic / engrafted variables	fiber bundle $\mathbb{R}^2 \times P \to \mathbb{R}^2$
orientation hypercolumns and pinwheels	1-jet space $J^1 \subsetneq \mathbb{R}^2 \times P \to \mathbb{R}^2$
– long-range horizontal connections, – "co-oriented, co-axially aligned RFs", – "joint constraints on positions and orientations" – "good continuation"	Contact structure

108. Variational models for illusory contours

Let us come back now to *curved* illusory contours. *Variational models* have been introduced since the late 70s.

1. Shimon Ullman (1976) explained that

   "A network with the local property of trying to keep the contours 'as straight as possible' can produce curves possessing the global property of minimizing total curvature."

2. Berthold Horn introduced in 1983 "the curves of least energy".

These models minimize an energy along curves *in the base plane*.

109. Mumford's elastica model. I

The best known is the *elastica* model proposed in 1992 by David Mumford.

The energy to minimize is:

$$E = \int_\gamma (\alpha\kappa^2 + \beta) ds$$

where γ is a smooth curve in $\mathbb{R}^2$.

110. Mumford's elastica model. II

Mumford argued that an illusory contour is a chain of pairs (a_i, p_i) along which the loss of activity is as weak as possible.

But leaks can have a double origin:

1. leaks proportional to the number of elements of the chain lead to the term $\int_\gamma \beta ds$ with a constant factor β;
2. leaks due to curvature and equal to the sum of the deflections of orientation between consecutive elements lead to the term $\int_\gamma \alpha \kappa^2 ds$

111. A stochastic approach

Mumford developed also a *stochastic* explanation of his model.

His hypothesis is that the curvature $\kappa(s)$ of the curve γ is a *white noise*.

As $\kappa(s) = \dot{\theta}(s)$, this means that $\theta(s)$ is a *Brownian motion*.

If we further suppose that the length of γ is a random variable with an exponential law, then elastica are the most probable curves.

112. Geodesic models

But for *neural* models (and not only 2D image processing) it is natural to work in $V1$, that is with the *contact structure* and the *Legendrian lifts*.

It is why I proposed in the 90s variational models in the fiber bundles $\mathbb{V}$ ($\mathbb{V}_J = \mathbb{H}_{pol}$, $\mathbb{R}^2 \times \mathbb{P}^1$, $\mathbb{R}^2 \times \mathbb{S}^1$, or $SE(2)$) modeling $V1$.

It is here that sub-Riemannian geometry, finally, fully comes on stage.

The natural idea is to introduce sub-Riemannian metrics on $\mathbb{V}$ and look at geodesic models for curve completion and illusory modal contours.

113. Thom's morphodynamical turn in the 70s-80s

Some historical landmarks.
My contribution to this story has only been

1. to recognize the contact structure of 1-jet space in pinwheels maps and horizontal connectivity;
2. to renew the variational hypothesis in this new context;
3. to articulate the two hitherto completely heterogeneous universes of differential geometry and neurophysiology of perception.

In the late 60s and in the 70s, I worked a lot on singularity theory and differential geometry with René Thom in a circle where I met Bernard Teissier, Alain Chenciner, Jean-Pierre Bourguignon, and later Daniel Bennequin and Marc Chaperon.

114. Cognitive neurosciences in the 80-90s

I was introduced to neuroscience through the extraordinary interdisciplinary network created in 1990 by the eminent specialist in vision Michel Imbert.

It favored collaborations with specialists of neurosciences, physiology and psychophysics like Michel Imbert himself, Yves Frégnac, Alain Berthoz, or Jean Lorenceau.

In the 90s, a working group organized many conferences at the "Fondation des Treilles" and also at the Institut Henri Poincaré and Oberwolfach, with Bernard Teissier, Jean-Michel Morel, David Mumford, Gérard Toulouse, Stéphane Mallat, Yves Frégnac, Jean Lorenceau, Olivier Faugeras, Giuseppe Longo.

115. Sub-Riemannian geometry in the 2000

In 2000, Alessandro Sarti, who worked since 1997 at Berkeley with James Sethian on Kanizsa illusory contours joined me and we started a cooperation involving also his colleague from Bologna Giovanna Citti.

They have worked in particular on coherent states and the harmonic analysis of SE (2).

At the same time, my colleague Helena Frankowska put me in touch with the SISSA, and I met Andrei Agrachev and his group of control theory and sub-Riemannian geometry (Jean-Paul Gauthier, Ugo Boscain, Yuri Sachkov).

I remember a discussion with Andrei at the IHP where he explained to us the pendulum equation underlying the geodesics of $SE(2)$!

116. The sub-Riemannian land

And so, after wavelets and coherent states, after jet spaces and contact structures, after variational models using the classical method of Lagrange multipliers and Robert Bryant and Phillip Griffiths' analysis of integrals $\int(\kappa^2/2)ds$, I set out for the sub-Riemannian enchanted land ...

Some other of my guides have been Misha Gromov, Richard Montgomery, André Bellaïche, Jean-Pierre Pansu, John Mitchell, Robert Strichartz, Anatoly Vershik.

117. The sub-Riemannian $\mathbb{H}_{pol}$

The sub-Riemannian geometry of the non polarized Heisenberg group $\mathbb{H}$ has been explained in the 1980s by Richard Beals, Bernard Gaveau and Peter Greiner.

It can easily be adapted to the polarized $\mathbb{H}_{pol}$.

As you know, geodesics are the projections on $\mathbb{H}_{pol}$ of the Hamiltonian on the cotangent space

$$H(x, y, p, \xi^*, \eta^*, \pi^*) = \frac{1}{2}\left[(\xi^* + p\eta^*)^2 + \pi^{*2}\right] .$$

where $(x, y, p) = q$ are coordinates in $\mathbb{H}_{pol}$ and (ξ^*, η^*, π^*) coordinates in the cotangent space $T_q^*\mathbb{H}_{pol}$.

118. The sub-Riemannian wavefront

The sphere S and the wave front W (with radius $\sqrt{2}$) are given by the equations (where (x, p) are expressed in polar coordinates with module $\frac{|\sin(\varphi)|}{\varphi}$, $\varphi > 0$)

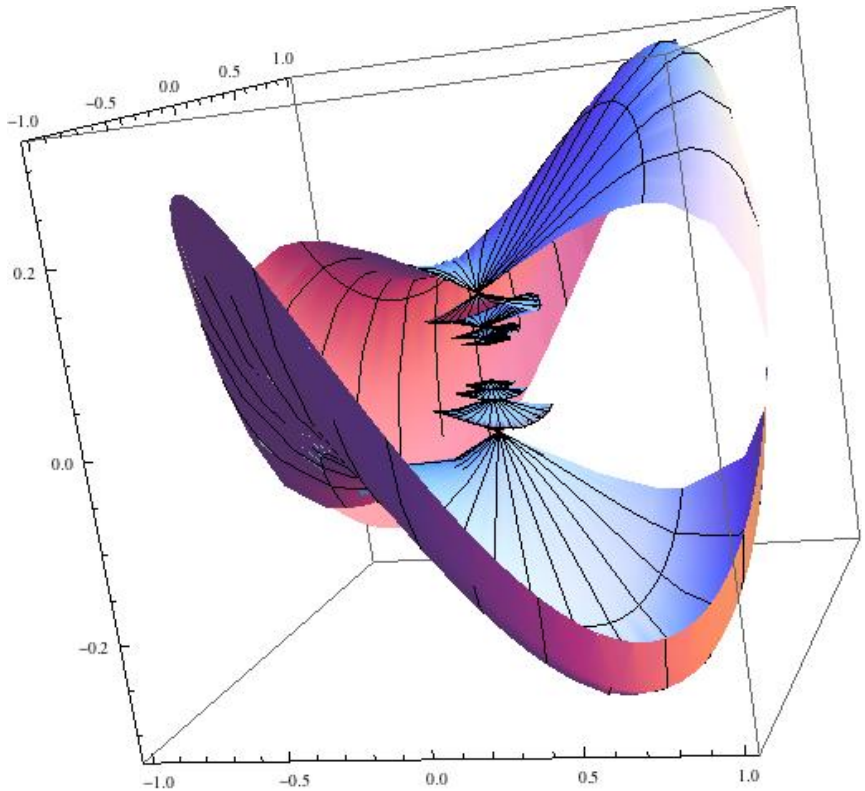
$$\begin{cases} x_1 = \frac{|\sin(\varphi)|}{\varphi} \cos(\theta) \\ p_1 = \frac{|\sin(\varphi)|}{\varphi} \sin(\theta) \\ y_1 = \frac{1}{2} x_1 p_1 + \frac{\varphi - \sin(\varphi)\cos(\varphi)}{4\varphi^2} \\ = \frac{1}{2} \frac{\sin^2(\varphi)}{\varphi^2} \cos(\theta)\sin(\theta) + \frac{\varphi - \cos(\varphi)\sin(\varphi)}{4\varphi^2} \\ = \frac{\varphi + 2\sin^2(\varphi)\cos(\theta)\sin(\theta) - \cos(\varphi)\sin(\varphi)}{4\varphi^2} \end{cases}$$

They are displayed in the following figure.

Such a complex behavior is impossible in Riemannian geometry.

119. Image of the SR sphere and wave-front

The external surface is the sub-Riemannian sphere S. It has a saddle form with singularities at the intersections with the y-axis. The internal part is $W - S$. It presents smaller and smaller circles of cusp singularities which converge to 0.



120. Inpainting and SR diffusion

I will conclude this presentation with the problem of *inpainting*, that is the completion of corrupted images.

It is natural to use *diffusion* along the horizontal connections, that is the sub-Riemannian hypoelliptic Laplacian $\Delta = X_1^2 + X_2^2$, and the sub-Riemannian heat kernel.

121. SR expertise

I learned a lot on these problems

- with Giovanna and Luca Capogna,
- with Andrei, Jean-Paul, Ugo, and their "intrinsic" formulation of the sub-Riemannian Laplacian and their general theorem for unimodular Lie groups of dimension 3,
- Yuri for the elastica and the $SE(2)$ conjugate points and cut locus,
- with the SRGI seminar: let me thanks warmly all of you and in particular Emmanuel Trélat and Davide Barilari.

122. Lifting level sets and filling-in the gaps

Given an image of intensity function $I(x, y)$, we can consider the Legendrian lifts of its level curves in $\mathbb{V}$.

We get a surface Σ in $\mathbb{V}$.

Let's suppose that the image is corrupted and contains a gap Λ.

To restore the image and to fill-in Λ, the idea is to use the *highly anisotropic* sub-Riemannian diffusion on $\mathbb{V}$.

The idea is conceptually simple but computationnally difficult.

123. Gauthier-Prandi inpainting

In their *SpringerBrief* on the sub-Riemannian neurogeometrical model (2018), Jean-Paul Gauthier and Dario Prandi deeply improved the computationnal efficiency of the model.

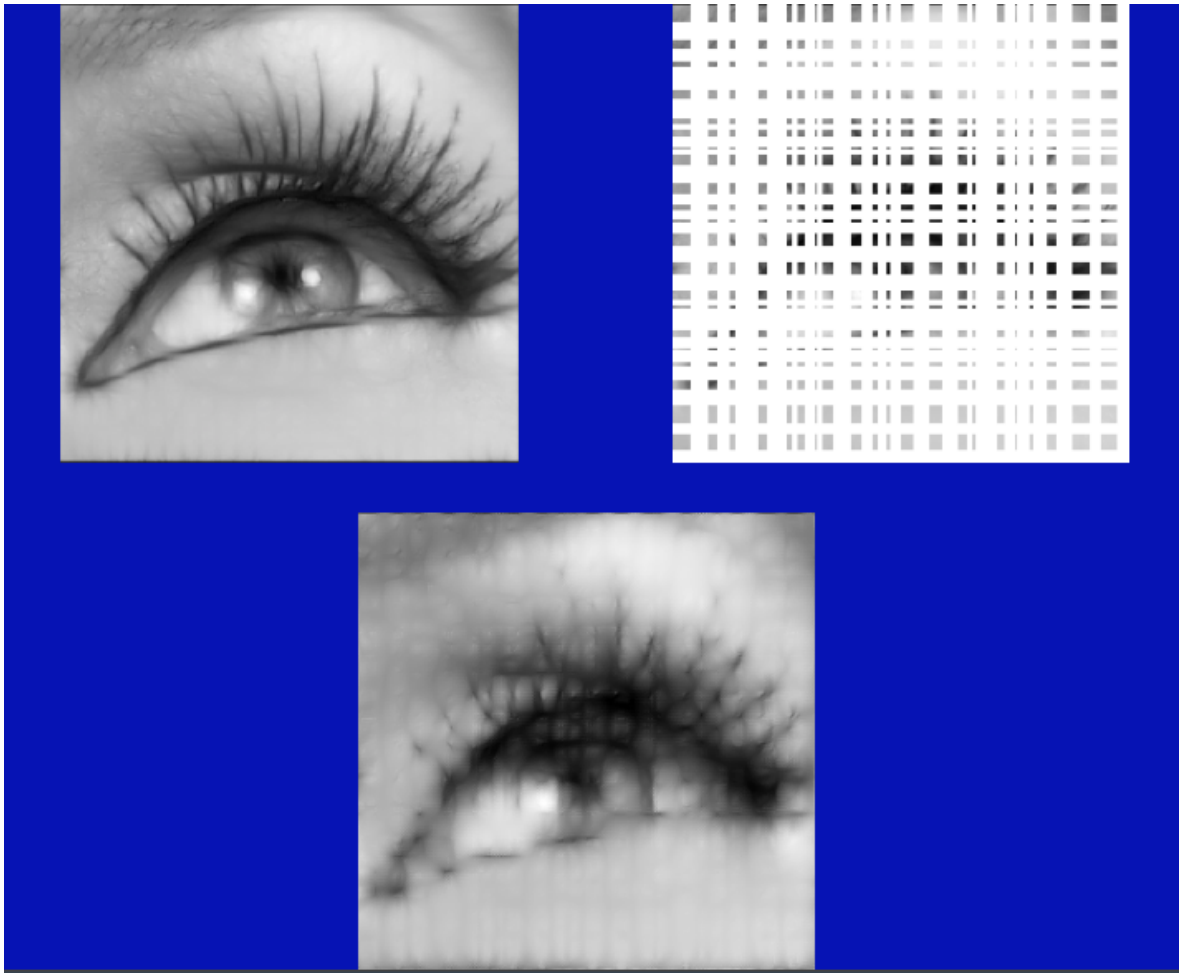
They used a *semi-discrete* version which is a left-invariant sub-Riemannian structure no longer over $SE(2)$ but over $SE(2, N)$ ($SE(2)$ restricted to a *finite* number of rotations).

In the following display, they start with an image almost completely concealed by a grid. The residual information is very sparse and scattered.

They apply sub-Riemannian diffusion until the grid has vanished.

Despite its dramatic corruption, the geometry of the initial image can be restored very correctly.

Top-left: initial image, top-right: highly corrupted image, bottom: restored image.



125. Further readings

Recently (Winter 2017) Matilde Marcolli (who worked with Misha Gromov and Alain Connes) devoted with her colleague Doris Tsao a very nice course on many of these topics (dynamical models of the neuron, receptive fields and Gabor patches, conformal geometry, contact geometry, segmentation and tracking, bistable images, neural codes, neural networks, deep learning).

In March 2020, they co-organized a "Focus Program" at the Fields Institute (Toronto).

Here is the presentation of their course. Look at Gromov's quotation.

126. Geometry of Neuroscience at Caltech

"It is unimaginable that a viable representation of the brain/mind interface is possible without being incorporated into a broader mathematical framework."
-Mikhail Gromov, "Structures, Learning, and Ergosystems"

127. “die der Geometrie zum Grunde liegen”

A “broader mathematical framework” could be sub-Riemannian geometry.

Using the nice expression “from within” that Gromov used for sub-Riemannian geometry, I could say that neurogeometry seeks to model “from within” the *neuronal hardwares* of the *brain* functional architectures implementing the *mind* ideal softwares of visual perception.

It adds new foundations for what Riemann and Helmholtz called the Hypotheses and Facts underlying Geometry:

die der Geometrie zum Grunde liegen.