

## WEEK 2: SHUNTING COMPETITIVE NETWORKS AND REPRESENTATION IN EARLY VISION

- 1) Brightness: Constancy and contrast
- 2) Shift property and Weber law
- 3) Featural noise suppression
- 4) Distance--dependent shunting networks
- 5) Another approach (Marr)
- 6) Recurrent competitive networks

## Preface

Recall that last week's lecture introduced several examples of visual phenomena (e.g. brightness constancy, brightness contrast, boundary completion, filling-in, . . . )

The goal of this week's class is to introduce a number of formalisms with properties needed to construct networks that can model those visual phenomena. Such properties include ratio processing, normalization, pattern choice (contrast enhancement) and pattern uniformization. This lecture will not in every case explicitly link each formal network property with associated visual phenomena; subsequent lectures will probe these connections, and related properties of biological networks, in detail.

This lecture also begins a comparison among "styles" of modeling, as exemplified by formalisms advocated by Grossberg and Marr.

## THE PLAN

Recall from Week 1 that the **simple shunting network equation**

$$\frac{dx_i}{dt} = -Ax_i + (B - x_i)I_i - x_i \sum_{k \neq i} I_k$$

was shown to factorize **pattern** and **energy** in its inputs

$$x_i = \theta_i \frac{BI}{A + I}$$

and to display *automatic gain control*, *ratio sensitivity*, and *normalization* properties.

## VISUAL ADAPTATION TO INPUT LEVEL

Automatic gain control can be said to be a type of *adaptation* to changes in visual input.

In "early" vision, two main types of adaptation occur:

- 1) **Bleaching**: Breaking down of pigment in photoreceptors due to exposure to light
- 2) **Background**: modulation of response to a small area when the input level in its surround is varied

Bleaching effects are **large** and important to our visual function, but they do not account for all perceptual effects.

We are here concerned with **background adaptation**.

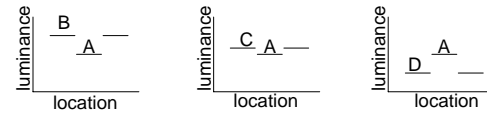
What in general is “adaptation”?

Is it “habituation”? Learning?  
Weight changes in a network (cf. “adaptive filter”)?

The human visual system’s operating range spans 15 orders of magnitude!

**BACKGROUND ADAPTATION AND SHIFT PROPERTY**

**Background adaptation** can be thought of as “appropriately” modulating the response to a given (absolute) input level depending on whether it’s background is high or low:



Intuitively, it seems that the luminance A should “count for more” when measured against background C, as opposed to background B, and even more when on background D.

The generalization of such behavior in a system, when plotted against continuous variations of background, results in “**the shift property.**”

**THE SHIFT PROPERTY, I**

The output of the simple shunting network:

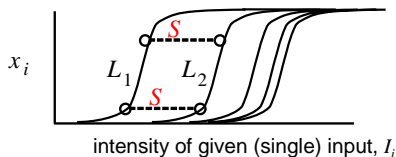
$$\frac{dx_i}{dt} = -Ax_i + (B - x_i)I_i - x_i \sum_{k \neq i} I_k$$

displays the shift property.

Define:

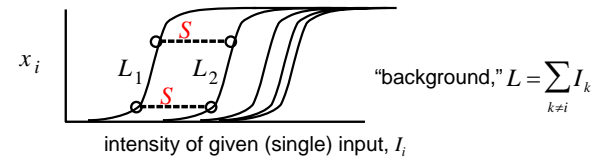
“background,”  $L = \sum_{k \neq i} I_k$

$L_2$  is a higher intensity background than  $L_1$



Response curves for all backgrounds are **congruent** -- just shifted copies of each other.

**THE SHIFT PROPERTY, II**

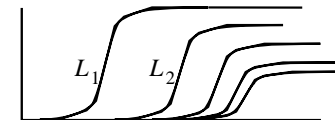


An increase of total background input energy to the network creates a strictly horizontal displacement of response curves.

**Automatic gain control:**

There is *no loss of sensitivity*, in the sense that nodes still employ their *entire output range*. (What about input range?)

As opposed to, for example:



Here “too much” inhibition from background reduces dynamic range of output.

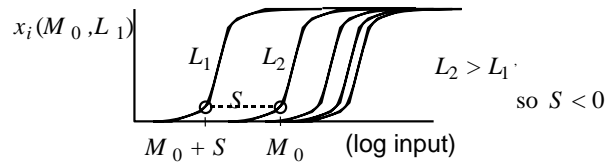
(This pattern is characteristic of “subtractive” inhibition).

**THE SHIFT PROPERTY, III**

More specifically, since  $x = \sum_k x_k = \frac{BI}{A+I} = \frac{BI}{A+I_i + \sum_{k \neq i} I_k}$

consider one input,  $M = \log_e(I_i)$ , and its background  $L = \sum_{k \neq i} I_k$

Then  $x_i(M, L) = \frac{Be^M}{A + e^M + L}$



**SHIFT PROPERTY:** There exists an  $S$  such that:

$$\forall M : x_i(M + S, L_1) = x_i(M, L_2) \quad S = \log_e \left( \frac{A + L_1}{A + L_2} \right)$$

**WEBER LAW**

*Definition:* The smallest increment,  $\Delta I$ , in some sensory stimulus (not necessarily visual) that can be reliably detected\* relative to some background level,  $I$ , is called the “just noticeable difference” (JND) for that background.

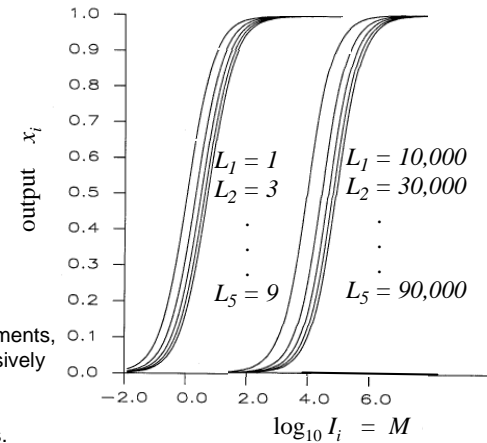
**Weber Law:** Over a wide range of some sensory dimension, the ratio of a JND for a background to that background is a constant; i.e.

$$\frac{\Delta I}{I} = C$$

*Note:* The *shift property* is a generalization of the *Weber law*

\* Important methodological issues concerning criteria for determining that some stimulus can be “reliably detected” will be discussed at a later time.

**EXAMPLES OF SHIFT PROPERTY**



Note that for two sequences of background increments, shifts get progressively smaller for linear increases in background levels.

**FECHNER'S LAW**

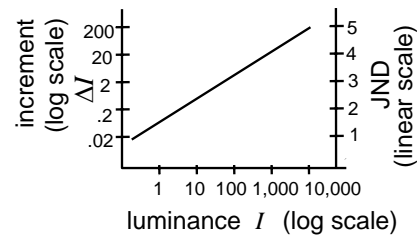
Quoting from a recent posting to CVNet:

<http://www.visionscience.com/mail/cvnet/2000/0314.html>

“As many of you are aware, October 22, 1850 is considered an important day in the history of psychology and sensory science. On that morning, Gustav Fechner formulated his famous quantitative relation between mental sensation and physical stimulus. **Fechner's law is  $S = K \log I$**  (the mental sensation varies as the logarithm of the material stimulus).”

## FECHNER'S LAW AND LOGARITHMIC PROCESSING

**Fechner** and **Weber** formulas are related, *if you assume that all JNDs are equal increments in sensation.*



## TRANSLATING GROSSBERG'S TO ENGLISH, PART 1 OF N

G '83 states that . . . while the shift property occurs in *logarithmic coordinates*, it is not due to *logarithmic processing*.

Translation:

There is no stage in network computations that performs an operation called: "Take the log of your input."

Remark:

This is not an empty distinction, as there are many computational models that *do* have such a stage.

## SUBTRACTIVE INHIBITION

Grossberg's "target" is (at times!) the *subtractive* (a.k.a. additive) modeling of "**lateral inhibition**" [Ratliff, 1965](#); [Cornsweet, 1970](#)

Put in a form closest to that of Grossberg's UMAP equation:

$$\frac{dx_i}{dt} = -Ax_i + BI_i - \sum_{k \neq i} I_k$$

At equilibrium:

$$x_i = \frac{BI_i - \sum_{k \neq i} I_k}{A}$$

Note: Unlike the case for the shunting network considered so far, inhibition **can** drive a node of *this* network **below zero**.

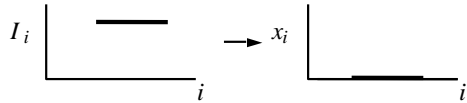
Having a node potential in a "computational" neural network go below zero is sometimes called "*hyperpolarization*," by analogy to a neuron *in vivo* being driven away from its resting level and further away from depolarization by some inhibitory input.

Note that in most "computational" or "artificial" neural modeling studies, a convention is employed whereby high cell activity is coded by positive numbers, and numbers below zero are assumed not to result in cell outputs -- notwithstanding the failure of direct analogy of such values to those of real neurons.

Note: Equations that are hybrids of those containing shunting and (input-dependent) subtractive inhibition are possible, as we will soon see.

**FEATURAL NOISE SUPPRESSION\***

Some networks can have parameters set so as to attenuate uniform regions of a pattern, which contain no information. ("Just a DC shift.")



Recall, for simple network:

$$x_i = \frac{(B+C)I}{A+I} \left( \theta_i - \frac{C}{B+C} \right)$$

Try:  $B = (n - 1)C$

Result: Uniform pattern is suppressed independent of intensity of  $I_i$ .

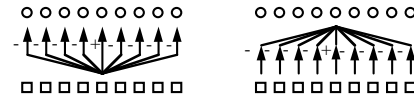
$$\frac{C}{B+C} = \frac{1}{n} = \theta_i$$

$$x_i = 0$$

**\*G jargon alert!** Featural noise suppression is the attenuation of signals in uniform regions of a pattern.

**NEED FOR DISTANCE-DEPENDENT INTERACTIONS**

To get the benefits of "featural noise suppression" for **parts** of a pattern, interactions among network nodes must **not** be uniform across the whole network, as for the equations considered so far, whose "network anatomies" are:



↪ dual representations ↪

Remember: All bottom nodes connect to all top ones.

NOTE: "**Center-surround**" is a misnomer for the anatomy of the previous panel, as there are **no neighborhood relations** or, equivalently for our purposes, **distance** measures -- as embodied in **connection weights** -- among nodes.

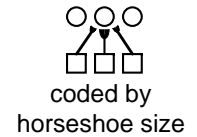
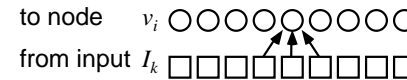
Thus, no node can be said to be in the "center" of any particular other nodes.

**DISTANCE-DEPENDENT KERNELS**

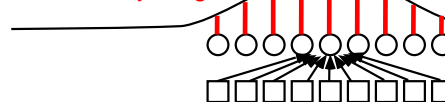
Consider 1-D cross-section for discrete kernels that code *unoriented* (radially symmetric) connections.



**Excitatory weights** (coded by height of bars)



**Inhibitory weights**



Note that for a **discrete** network we must sample (continuous) Gaussian weight functions.

**SHUNTING NETWORK\* WITH DISTANCE-DEPENDENT TERMS**

$$\frac{dx_i}{dt} = -Ax_i + (B - x_i) \sum_{k=1}^n I_k C_{ki} - (x_i + D_i) \sum_{k=1}^n I_k E_{ki}$$

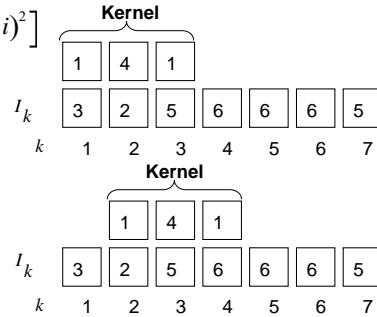
$$C_{ki} = C \exp[-\mu(k-i)^2]$$

$$E_{ki} = E \exp[-\nu(k-i)^2]$$

Example:

$$C_{ki} = \begin{cases} 4 & \text{if } k=i \\ 1 & \text{if } |k-i|=1 \\ 0 & \text{otherwise} \end{cases}$$

apart from "outer boundary effects"



**EQUILIBRIUM OF DISTANCE-DEPENDENT NETWORK**

Set  $\frac{dx_i}{dt} = 0$  and recall that  $I_k = I\theta_k$

$$x_i = \frac{I \sum_{k=1}^n \theta_k (BC_{ki} - DE_{ki})}{A + I \sum_{k=1}^n \theta_k (C_{ki} + E_{ki})}$$

**Hot tip:**

See Neumann, H. (1996), *Neural Networks*, 9(6), 921-936. (Suppl. reading, Week 3) for analysis of this equation.

Result:

Numerator:

**DOG**

(difference of Gaussians)

Denominator:

weighted **SOG**

(sum of Gaussians)

scaled against constant, **A**

**DISTANCE-DEPENDENT FAQS**

$$x_i = \frac{I \sum_{k=1}^n \theta_k (BC_{ki} - DE_{ki})}{A + I \sum_{k=1}^n \theta_k (C_{ki} + E_{ki})}$$

$$C_{ki} = C \exp[-\mu(k-i)^2]$$

$$E_{ki} = E \exp[-\nu(k-i)^2]$$

Should  $C = \frac{1}{\sigma\sqrt{2\pi}}$  ?

Should  $\mu = \frac{1}{2\sigma^2}$  ?

How do I choose the other constants?

How broad should "center" and "surround" be?

**FEATURAL NOISE SUPPRESSION**

Recall: "Featural noise suppression" is G's phrase for producing **zero output for homogeneous regions of input.**

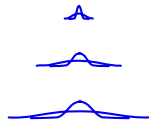
Theoretically, this occurs for distance-dependent equation *iff*:

$$B \sum_{k=1}^n C_{ki} \leq D \sum_{k=1}^n E_{ki}$$

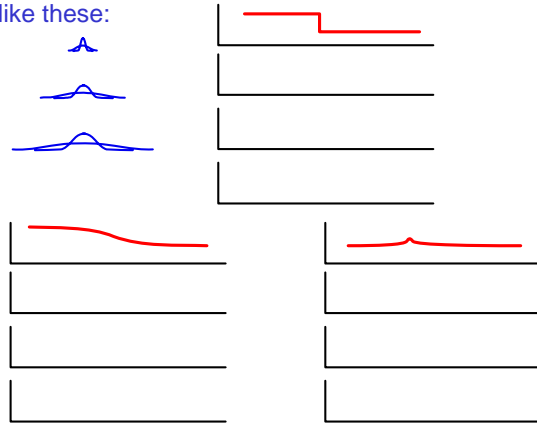
Provided that you rectify the output -- i.e. reset any negative value to zero.

### MENTAL CONVOLUTION

How would filters or detectors like these:



Respond to patterns like these?



### REFLECTANCES AND RATIOS

A distance-dependent network cannot possibly compute *Grossbergian reflectances*, which are defined over the set of all inputs.



"Reflectance" here depends on input strength there.

**feature or bug?**

Instead, compute ratios of *local* "reflectances", weighted by network parameters.

### Random Notes re: Distance-dependent shunting equation

In computational networks where a node's activity can assume negative values, *rectification* of output is often employed.

$x_i$  is sensitive to (Grossbergian) reflectances (ratios) in regions of rapid changes of input level\*.



"Gradual" changes -- **relative to**  $B, D, \mu, \nu$  -- are discounted.

Compare last two points to discussion of Retinex, brightness constancy, etc. in Lecture 1.

### BACKGROUND

Many **formalisms** have been applied to "explain" early vision. Many kinds of **psychophysical experiments** have been run. Much **physiological data** has been collected.

Yet . . . vision researchers do not have a consensus about how to characterize *early and middle\** vision.

\* "before\*\*" recognition of objects, but including *detection* of complex features (e.g. "T-junctions") **and** *perceptual organization*

### ALTERNATIVES TO NETWORK MODELS

*Near consensus:* Adaptation, contrast sensitivity, normalization, filtering, etc.

*Far from consensus:* Receptive field dynamics, parallel channels (including magno/parvo, ON/OFF, what/where, etc.), and more . . .

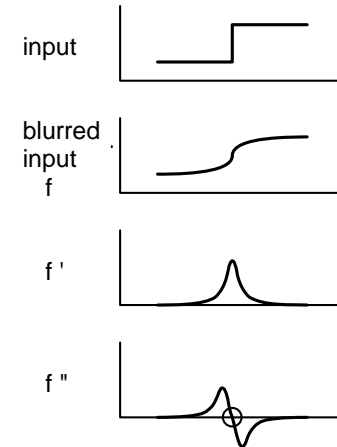
Where to start?

### MARR'S ZERO-CROSSINGS

Marr, 1980\*, recommends that image edges be detected by computing the zero-crossings of

$$\nabla^2 G * I$$

the *Laplacian of a Gaussian* filtering of the image data.



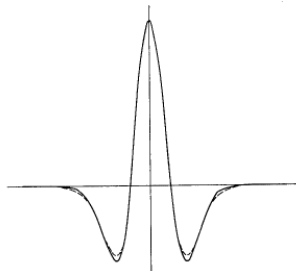
\*Note that this part of Marr's prescription is in Chapter 2 of his book, and is not excerpted in your reading from Yantis's book.

### DOGS AND LAPLACIANS

Marr also points out that

$$\nabla^2 G * I$$

can be well-approximated by a DOG, if the ratio of excitatory to inhibitory space constants is about 1:1.6.

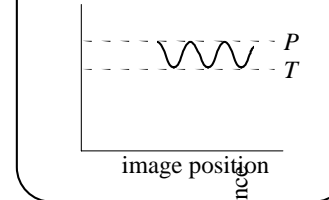


**Figure 2-16.** The best engineering approximation to  $\nabla^2 G$  (shown by the continuous line), obtained by using the difference of two Gaussians (DOG), occurs when the ratio of the inhibitory to excitatory space constraints is about 1:1.6. The DOG is shown here dotted. The two profiles are very similar. (Reprinted by permission from D. Marr and E. Hildreth, "Theory of edge detection," *Proc. R. Soc. Lond. B* 207, pp. 187-217.)

### DEFINITION OF CONTRAST

$$C = \frac{P - T}{P + T}$$

Michelson contrast



P: peak  
T: trough

The "**Michelson contrast**" formula is one of several equivalent forms of the same idea (Michelson, 1927).

E.g., one might instead choose an expression in which the mean of two luminances was in the denominator.

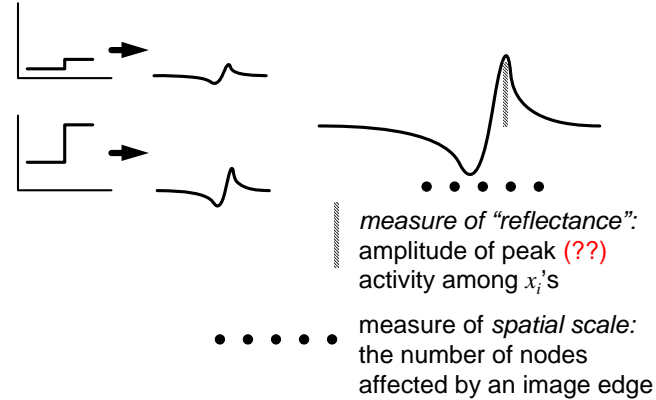
Do not confuse the word "contrast" in phrases like "Michelson contrast" or "stimulus contrast" or "image contrast" with its usage in phrases like "brightness contrast" or "color contrast."

The former usages refer to characteristics of a **visual stimulus** that can be *objectively measured* (e.g., with a photometer.)

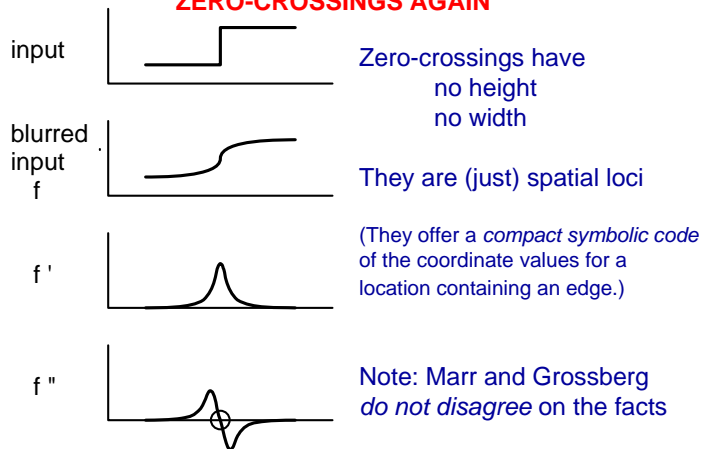
The latter usually refer to **perceptual phenomena**.

**GROSSBERG'S CRITIQUE OF ZERO-CROSSINGS**

Grossberg, 1983 prefers to retain the *entire output profile* of a shunting network with DOG kernels -- as opposed to the zero-crossings of DOG convolutions only:



**ZERO-CROSSINGS AGAIN**



**DAZZLING (FOR ITS TIME) DEMO**

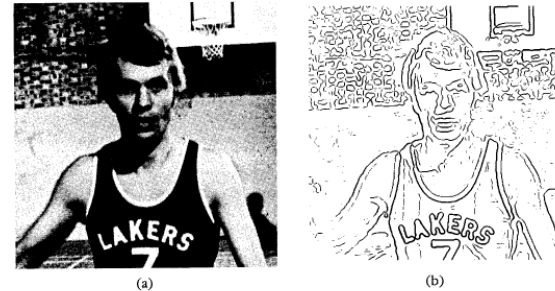
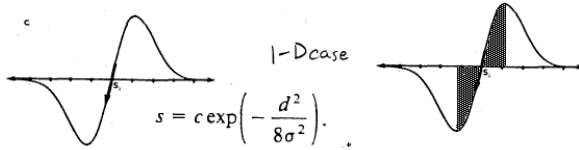


Figure 2-15. Another example of zero-crossings; here, the intensity of the lines has been made to vary with the *slope* of the zero-crossing, so that it is easier to see which lines correspond to the greater *contrast*. (Courtesy BBC Horizon.) [From Marr, 1982; emphasis added.]

**MODULARITY OF MARR'S APPROACH**



Hildreth, 1983  
CVGIP, 22, 1-27.

Huertas & Medioni, 1986  
IEEE PAMI, 8(5), 651.

Use *slope* at zero-crossings of various sized operators ( $\sigma$ ) to compute *contrast*.

Slope, or peak-to-peak amplitude, or area . . .

**Note:** While measures of contrast can be gotten, they are represented *separately* from the code for the zero-crossing itself.

**Disclaimers about the rest of the lecture**

Many students have pointed out that the material in the remainder of this lecture is redundant with the contents of (later weeks of) CN510.

Yes, I am aware of this redundancy.

- I feel it is appropriate to include this material here because:
- 1) the "gist" of the concepts are needed sooner in the semester in CN530 than they occur in CN510, and
  - 2) the material is of such fundamental importance that it would not hurt you to be exposed to it (at least) twice.

I am also aware that the arguments summarized in what follows are not easy to follow if you have not had coursework in dynamical systems. Still, ya gotta start somewhere . . .

**Recurrent Shunting Networks in Vision**

To join *grouping* with *coherent binding*, we need:

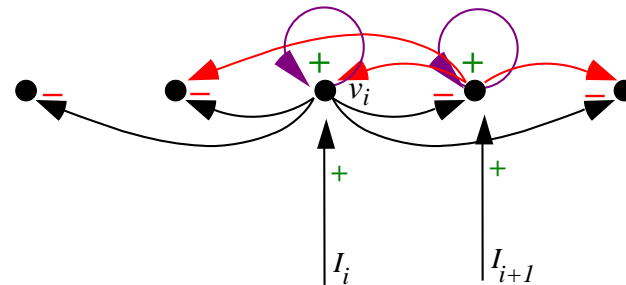
*spatial and orientational kernels*

multiple nested *layers with feedback loops*

Isn't earlier analysis of *feedforward shunting ON-center, OFF-surround network* enough?

**Noise-Saturation Dilemma -- Again!**

Need: *ON-center, OFF-surround with FEEDBACK*



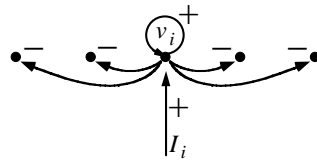
More complicated situation than feedforward case

Greater need for mathematical analysis to clarify . . .

See *Grossberg, 1973 -- supplementary reading, Week 2*

**ON-CENTER, OFF-SURROUND RECEPTIVE FIELDS**

Grossberg's 1973 Figure 5 looks like this



and is labeled "recurrent on-center, off-surround network."

The phrase "on-center, off-surround" has historically referred to the "receptive fields" of neurons, viewed functionally.

The relation of a neuron's **connectivity with other neurons** [ANATOMY] to that neuron's **receptive field** [PHYSIOLOGY] is tricky.

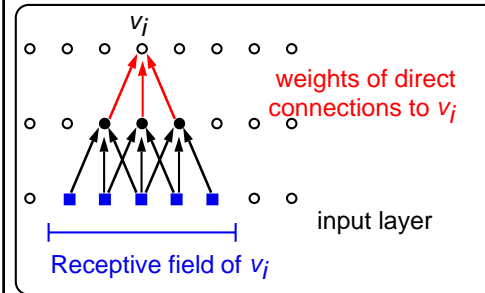
**MODELERS AND PHYSIOLOGISTS: "DIVIDED BY A COMMON LANGUAGE"**

**RECEPTIVE FIELD** -- functional

Where **on the retina** will stimulation yield a response at this (cortical) cell?

**KERNEL** -- structural

Which network cells send inputs directly to this cell?



Kernels are trivial for a modeler to specify, but are generally **not observable** for a physiologist!

**Feedback Shunting Networks**

Given a network's anatomy, its signal functions, parameter restrictions, and initial conditions, ask:

**STABILITY:** Is there **storage** of a pattern (**short-term memory**)?

**PATTERN TRANSFORMATION:**

What happens to initial activity pattern?

Is it preserved, destroyed, smoothed, contrast-enhanced, ...?

Plug: R. Abraham & C. Shaw  
*Dynamics: The geometry of behavior*  
 Aerial Press, Inc.  
 Box 1360, Santa Cruz, CA 95061  
 (408) 425-8629

**Properties of Recurrent Competitive Networks**

**Grossberg, 1973** (Week 3 of syllabus):

What happens to  $x$  (*total network activity*) as (time)  $t \rightarrow \infty$ ?

- Possibilities:*
- $x \rightarrow 0$  "collapse" of all activity
  - $x \rightarrow \infty$  network "blows up"
  - $x \rightarrow \text{constant}$  (stability) **storage!**
  - $x \rightarrow$  one of finitely many values
  - $x \rightarrow$  one of infinitely many (finite) values
  - $x$  oscillates
  - $x$  is chaotic (not in 1973!)

**Key result:**  
*Network anatomies (patterns of connections) and signal functions constrain outcomes.*

**Q:** Why this?! Why now?!  
**A:** To see what other kinds of representations (besides zero-crossings, etc.) are even possible!

### PATTERN VARIABLES

Definition of pattern variables (functions of time):  $X_i = \frac{x_i}{x}$   
 (compare with  $G$ 's definition of "reflectance")

What happens to *initial*  $X_i$ 's as  $t \rightarrow \infty$ ?

Possibilities:

Each  $X_i \rightarrow X_i$       *nothing* happens; i.e. "perfect storage"

Maximum  $X_i \rightarrow 1$       "**winner take all**" a.k.a. "**choice**"

All  $X_i$ 's  $\rightarrow 1/n$       "**uniformizing**" (cf. "filling-in"?)

where  $n$  = number of nodes in network

Some  $X_i$ 's  $\rightarrow 0$       "**quenching threshold**" yields **contrast enhancement** of activity of surviving nodes

### Factorize Pattern and Total Activity

Method of proof: Change variables to:

pattern:  $X_i = \frac{x_i}{x}$       total activity:  $x = \sum_{k=1}^n x_k$

feedback signal:  $f(w)$        $g(w) = \frac{f(w)}{w}$

Why  $g(w)$ ?

"How nonlinear IS it?"

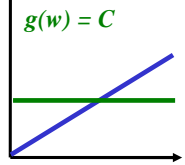
### Shape of Nonlinear Feedback

feedback signal:  $f(w)$

$$g(w) = \frac{f(w)}{w}$$

linear

e.g.,  $f(w) = Cw$

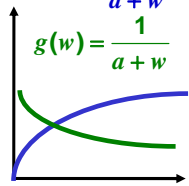


no advantage across size of  $x$

slower than linear

$$f(w) = \frac{w}{a+w}$$

$$g(w) = \frac{1}{a+w}$$

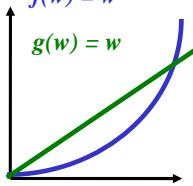


relatively stronger for small  $x$

faster than linear

$$f(w) = w^2$$

$$g(w) = w$$



relatively stronger for large  $x$

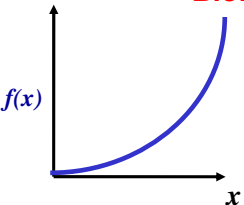
### A Series of Global Theorems

Grossberg, 1973, Studies in Applied Math

 Initial pattern $i$	$f$	$X_i(\infty) = \frac{x_i(\infty)}{\sum_j x_j(\infty)}$	
	 Linear	 Perfect storage of any pattern ( $\infty$ )	
	 Faster-than-linear	 Chooses max Winner-take-all	

CN530 F-2009 2- 49

### Biological Realism

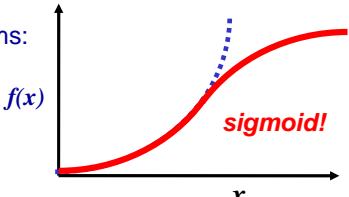


**Faster-than-linear** feedback signal function supports **noise suppression**

But, as  $x \rightarrow \infty$ ,  $f(x) \rightarrow \infty$   
**... not realistic**

Winner-take-all **noise suppression is too severe**  
 Network **only stores one feature**

One change solves both problems:



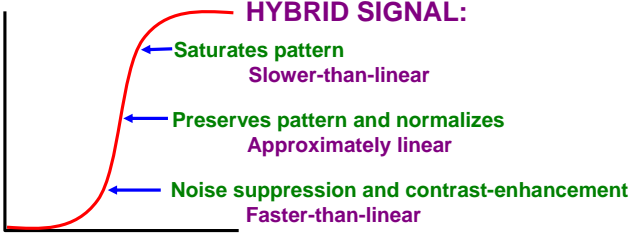
*sigmoid!*

CN530 F-2009 2- 50

### Sigmoid Signal Function

#### Distributed Processing and Noise Suppression

#### HYBRID SIGNAL:



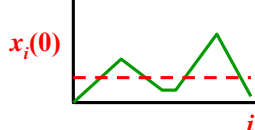
The **faster-than-linear** part **suppresses noise and starts to contrast-enhance the pattern**

As total activity **normalizes**, the **approximately linear** range is reached and **tends to store the partially contrast-enhanced pattern**

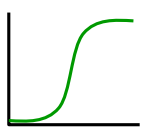
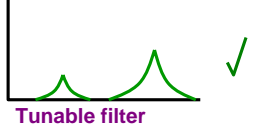
CN530 F-2009 2- 51

### Sigmoid Signal Function

#### Distributed processing and noise suppression



**Quenching Threshold (QT)**

$f$	$X_i(\infty)$	$x(\infty)$
 <b>Sigmoid</b>	 <b>Tunable filter</b> ✓	<b>Suppresses noise</b> ✓

**The QT can be dynamically tuned; e.g., pay attention better after unexpected event; choose max...**

CN530 F-2009 2- 52

Note that the quenching threshold for a network is a *dynamic quantity*. The **same** network with the **same** parameters will have a **different** quenching threshold for **different input patterns**.

An analogy to this is most readily seen in winner-take-all networks, where the second-largest node in one initial distribution (e.g. -- .1, .1, .1, .3, .4 -- might be driven to zero, whereas .3 might be the "winner that takes all" in another distribution, e.g., .2, .2, .3, .1, .2).

## **Onward!**

Next week, we will interrupt our modeling narrative for a “core dump” of some necessary physiological background.

Then we can return to modeling with an entirely more effective set of tools than we started with last week!

Result: It will take another few weeks for certain themes in the course to “gel,” but that’s to be expected, and previous generations of CN530-takers are living proof that survival is possible.