WEEK 2: SHUNTING COMPETITIVE NETWORKS AND REPRESENTATION IN EARLY VISION

1) The noise--saturation dilemma
2) Reflectances and ratios; shunting and mass action
3) Brightness: Constancy and contrast
4) Shift property and Weber law
5) Retinal physiology
6) Hyperpolarization and featural noise suppression
7) Distance--dependent shunting networks

THE PLAN

Feedforward shunting competitive networks
Distance-dependent interactions

Functional view of network properties: motivated by physiology described abstractly.

Try to understand certain molar perceptual (or “computational”) properties with reference to simplified micro-interactions.

COMPETITION

• a maximum and a minimum number of sites (processes) can turn on or off;
• each site's process is essentially “all or none.”

Key
- off (unexcited site)
- on (excited site)

Competition is ubiquitous in the biological world.

Why competition on the cellular level?

What is a cell, computationally speaking?

Infinity does not exist in biology; so our neural models should not assume that it does -- not even implicitly.
**NOISE-SATURATION DILEMMA**

Pattern registration by cellular systems

Given the constraints of a finite number of sites with all-or-none processes, how does nature prevent the following?

Let: \( I_n = x^n \)

- \( I \) stands for intensity, or energy of input.
- \( x \) stands for the amplitude of the input signal.

<table>
<thead>
<tr>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>X4</th>
<th>X5</th>
<th>X6</th>
<th>X7</th>
</tr>
</thead>
<tbody>
<tr>
<td>I1</td>
<td>I2</td>
<td>I3</td>
<td>I4</td>
<td>I5</td>
<td>I6</td>
<td>I7</td>
</tr>
</tbody>
</table>

**Effective operating range**

Range in which input signals can not be distinguished from endogenous noise.

**Saturation range**


**NOISE-SATURATION DILEMMA, AGAIN**

![Graphs showing energy levels and saturation]

**NEED:** automatic gain control (retuning of sensitivity)

**HARDWARE AND WETWARE**

The conventional engineering solution for an automatic gain control . . .

In a neural network:

1) Let each input source fluctuate in amplitude;
2) Let the total number of convergent sources fluctuate during pattern processing.

But: Each node has a fixed operating range.

(Output stays within upper and lower bounds.)

Grossberg, 1973:

If \( x_i \)'s are sensitive to small inputs, why don't they saturate in response to large inputs?

If \( x_i \)'s are sensitive to large inputs, why don't small inputs get lost in endogenous noise?
The previous panel’s point can be put like this:

Normalization is easy; knowing when and where to normalize and when and where to not normalize is hard.

A current leading theory of lightness perception (Gilchrist et al.) is based on interactions among “frameworks” -- i.e. regions of the scene within and between which ratios of luminance are to be compared.

Good news: The formula’s for computing the ratios and related quantities are pretty well worked out.

Bad news: There is nothing like an effective procedure for determining what the frameworks are for even modestly complex “scenes.”

### PHYSICAL REFLECTANCE

Classical definition:

the ratio of reflected \( R \) to incident \( I \) light for a Lambertian (i.e., ideal, perfectly diffusing) surface.

\[ 0 < R/I < 1 \] for each surface

a.k.a: albedo

COMPARE:

Grossberg's definition of reflectance of network inputs:

Input to a node: \( I_i \) or \( I_i(t) \) for \( i = 1, \ldots n \)

Total input: \( I = \sum_j I_j \)

“Reflectance” of an input: \( \theta_i = \frac{I_i}{I} \)

### RATIO SENSITIVITY IN SHUNTING NETWORKS

PROBLEM: How to remain sensitive to input ratios (e.g. \( I_j/I_k \)) as the total input, \( I \), gets larger and larger?

Notation:

\( v_i \) name of the \( i \)th node

\( x_i \) potential or activity level of the \( i \)th node

\( f(x_i) \) output signal function of the \( i \)th node

(frequently involving, e.g., a rectification, thresholding, or sigmoidal transform)

What about Grossberg’s phrase “cell or population”?
CONSTRUCTION OF SIMPLE SHUNTING NETWORK, PART 2

Assume: (1) Inputs perturb cell sites, and (2) activity at each site is “all or none.”

Graphical convention: Each little circle stands for a site.

Activity

\[ x_1(t) \quad x_2(t) \quad x_3(t) \quad \ldots \quad x_n(t) \]

Inhibitory inputs affect only these: \( x_i(t) \) excited sites at time \( t \)

Excitatory inputs affect only these: \( B - x_i(t) \) unexcited sites at time \( t \)

Why (Go Forth and) Multiply?

Assume:

Activity of excited sites spontaneously decays to equilibrium (first term.)

Activity of unexcited sites is increased by inputs, in proportion to size of input and number of unexcited sites (mass action*, connoted by multiplication in second term).

\[
\frac{dx_i}{dt} = -Ax_i + (B - x_i)I_i
\]

* implies statistical independence, therefore multiplication:

Expected value: \( E(X \text{ and } Y) = E(X)E(Y) \)

Mass Action

Reaction rate for particles of two types in a medium?

Analogy: Consider a chemical reaction occurring among particles of two types that are sparsely distributed in a fluid container (e.g. yeast and sugar, for fermentation).

Reaction occurs only upon collision of particles of different types.

Random process: Rate of reaction is proportional to product of the concentrations of the two types of particles.

In network equation, multiply input size by measure of available (inactive) sites, . . .

because there are many sites, and a single input is thought of as arriving at particular sites randomly along similarly numerous, “all or none” channels.
NEED FOR INHIBITION

Consider: \[ \frac{dx_i}{dt} = -Ax_i + (B - x_i)I_i \]

To compute equilibrium response, set: \[ \frac{dx_i}{dt} = 0 \]

Solution: \[ x_i = \frac{B\theta_i I_i}{A + \theta_i I} \]

Remember: \[ \theta_i = \frac{I_i}{\sum_k I_k} \]

If \( I \) is small relative to \( A \), inputs are lost in noise,* because \( A \) of denominator dominates.

If \( I \) is large relative to \( A \), all nodes saturate at the value \( B \).

* CHEATING ALERT: There's no “noise” in the equation.

SHUNTING INHIBITION

Next, suppose that excited sites are inhibited by mass action:

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

Here, each node receives “equal weight” inhibitory input from all input sources except the one with the same index value.

NOTE: The use of indices does not here imply a functional spatial ordering (“neighbors”) among nodes.

COOPERATIVE AND COMPETITIVE NETWORKS

Network constructed so far “didn’t have a chance.”

To be sensitive to Grossbergian reflectances, each node’s output must be influenced by all of the inputs, by definition:

\[ \theta_i = \frac{I_i}{\sum_k I_k} = \frac{I_i}{I_i + \sum_{k \neq i} I_k} \]

by definition desired effect on \( x_i \)

Increasing \( I_i \) tends to increase \( \theta_i \) COOPERATIVE

Increasing \( I_j (j \neq i) \) tends to decrease \( \theta_i \) COMPETITIVE

G: “on-center, off-surround” network anatomy (Kuffler, 1953)

CAREFUL: There’s no “distance” between nodes in this network!

SHUNTING INHIBITION

Now, at equilibrium: \[ x_i = \theta_i \frac{Bl}{A + I} \]

i.e., \( x_i \rightarrow \theta_i \) as \( I \rightarrow \infty \)

PATTERN* ENERGY

“factorization”

*The word “pattern” here means a normalized vector quantity; elsewhere, the word may be used to mean any vector quantity.

QUIZ: Sensitivity of node potential to (Grossbergian) reflectances is said to “model” brightness constancy. Why?
NORMALIZATION: CONSERVATION OF TOTAL ACTIVITY

As total input increases, what happens to total activity in the network?

\[ x = \sum_k x_k = \frac{B I}{A + I}, \text{ since } \sum_k \theta_k = 1 \]

Result, total network activity asymptotically approaches a constant value, \( B \).

This fact is said to “model” (or “explain”) brightness contrast. Why?

VISUAL ADAPTATION TO INPUT LEVEL

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

Although bleaching effects are large, they do not account for all perceptual effects.

We are concerned with **background adaptation**.

Why “contrast”?

If the network is presented with a “gray” -- or, more properly, medium luminance -- target on a “black” (low luminance) background, the target has a larger Grossbergian reflectance than that same target on a “white” (high luminance) background.

The normalization property prevents all node activities from rising proportionately to create absolutely veridical (linear) responses to inputs regardless of background level.

Aside: When a variable that generally employs an index (to indicate which node in a network is being referred to) appears without any index, **SUMMATION** over that index (or a typo) has occurred.

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!
**THE SHIFT PROPERTY, I**

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

**EXAMPLES OF SHIFT PROPERTY**

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

**THE SHIFT PROPERTY, II**

More specifically, since \( x = \sum_k x_k = \frac{BI}{A+I} = \frac{BI}{A+I + \sum I_k} \)

consider one input, \( M = \log_e(I_i) \), and its background \( L = \sum I_k \)

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

\( x_i(M_0, L_1) \)

\( L_2 > L_1 \)

so \( S < 0 \)

**SHIFT PROPERTY:** There exists an \( S \) such that:

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

**SUPPLEMENTARY MATERIAL Follows in Four Panels**

Note: The next four panels contain material from ancient manuscripts recently discovered in the ruins of what is believed to have been a monastery of a zealous neural modeling sect on the site of what is presently the BU Math Department.

They contain some algebraic derivations. While I will not “lecture” on these panels, and while you will not be expected to produce these derivations on the midterm exam, “what could it hurt, to study them”? 
DERIVATION OF SHIFT PROPERTY

Find S such that
\[ x_i (M + S, L_1) \equiv x_i (M, L_2) \]

\[ \frac{B e^{M+S}}{A + e^{M+S} + L_1} \equiv \frac{B e^M}{A + e^M + L_2} \]

\[ \frac{e^S (A + e^M + L_2)}{A + e^M + L_1} = \frac{e^S (A + L_2)}{A + L_1} \]

\[ e^S = \frac{A + L_1}{A + L_2} \]

\[ S = \log_e \left( \frac{A + L_1}{A + L_2} \right) \]

Note that value of S is independent of \( M = \log_e I \).

DERIVATION OF SHIFT SIZES

Shifts get smaller as background intensity increases (for linear increases in \( L \)).

Let \( L_n = nL \)

\[ L = \sum_{k=1}^{n+1} I_k \]

That is, let background intensity increase.

\[ S = \log_e \left( \frac{A + (n+1)L}{A + nL} \right) \]

\[ = \log_e \frac{A (n+1) + L}{A n + L} \]

As \( n \to \infty \), \( S \to \log_e \left( \frac{L}{L} \right) = 0 \).

DERIVATION OF WEBER LAW PROPERTY, I

Derivation of "Weber law\(^{20}\) (Gershberg, 1982)

Section 2.3

\[ \frac{d}{dt} x_i = -Ax_i + (B - x_i)I_i - x_i \sum_{j \neq i} I_j \]

Let \( I_i = K + \Delta I \)

\[ I_j = K, \text{ for } j \neq i \]

before increment

Then:

\[ x_i = B \left( I_i + \Delta I \right) \]

\[ \frac{A + \Delta I}{A + I + \Delta I} \approx \frac{K + \Delta I}{K + I} \]

If \( I >> \Delta I \) and \( n >> 1 \)

\[ \frac{K + \Delta I}{nK + \Delta I} = \frac{I + \Delta I}{I + \Delta I} = 1 \]

If \( x_i \) is detectable when it exceeds some threshold \( \Gamma \),

\[ \sum -D \approx \frac{A I \Delta I}{I} \]

a constant

DERIVATION OF WEBER LAW PROPERTY, II

\[ x_i = B \left( I_i + \Delta I \right) \left( \frac{K + \Delta I}{K + I} \right) \]

If \( I >> A \)

\[ \frac{B (I_i + \Delta I)}{A + I + \Delta I} \approx B \]

(from last page)

So:

\[ x_i \approx B \left( \frac{\Delta I}{I} + D \right) \]

If \( x_i \) is detectable when it exceeds some threshold \( \Gamma \),

\[ \sum \approx B \left( \frac{\Delta I}{I} + D \right) \]

\[ \frac{\sum - D}{B} \approx \frac{\Delta I}{I} \]
**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.

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

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

---

**FECHNER’S LAW**

Quoting from a recent posting to CVNet:


"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 GROSSBERGSESE 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 (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.

---

**BIOLOGICAL NETWORKS: RETINAL PROCESSING**

Werblin’s (1970) study of mudpuppy retina:

- Receptors
- Horizontal cells
- Bipolar cells

*intracellular* electrode (records “graded” potentials)

**Q:** Do bipolars spike?

Findings:

1. Center-surround processing
2. Weber law in intermediate input range
3. No hyperpolarization from surround*
4. Shift property

* Suggests shunting inhibition. Why?

---

**WERBLIN (1970) DATA**

Above: Responses to light increments by a single ganglion cell as background is raised by two log units.

Shift is “imperfect,” with some loss of dynamic range.
ASIDE: HUMAN RETINAL PATHWAYS (1 OF 3)

Reading physio figures:

How is time indicated?

How is cell potential (or firing rate) indicated?

Read these diagrams with “computational” questions in mind. E.g., why have separate ON and OFF pathways?

Note: this edition makes no mention of possible “cross-talk” between ON and OFF channels in the retina.

ASIDE: ON/OFF CROSSTALK? (2 OF 3)

In 1991 CNS student (and, subsequently, CNS Professor) Paolo Gaudiano predicted on theoretical grounds that there should be convergent input from ON and OFF bipolar cells onto each of ON and OFF ganglion cells.

This was a genuine “prediction;” i.e., the kind that points forward in time.

ASIDE: NOW YOU SEE IT; NOW YOU DON’T (3 OF 3)

Compare the figure shown two panels ago with the corresponding one from the 3rd edition of K, S, & J:

What gives?

a) Crosstalk does not exist after all?

b) K, S, & J has been “dumbed down”?

c) Neither of the above?

ELABORATIONS OF SHUNTING NETWORK: HYPERPOLARIZATION

Hyperpolarization and, more generally, the effects of subtractive inhibition, can occur in a so-called “shunting” network, provided that the passive (“resting”) potential, $V^P$, is different from $V^-$, the potential that is farthest from generating a node’s output,

$$\frac{dx_i}{dt} = -A x_i + (B - x_i) I_i - (x_i + C) \sum_{k \neq i} I_k$$
**HODGKIN/HUXLEY EQUATIONS AND SHUNTING NETWORKS**

Membrane equations of physiology (Hodgkin and Huxley, 1952; Carpenter, 1981)
of the following (simplified) form

\[
C \frac{\partial V}{\partial t} = (V^+ - V)g^+ + (V^- - V)g^- + (V^p - V)g^p
\]

\[
\begin{align*}
\text{Na}^+ \text{ channel} & \quad V^+ \rightarrow B \\
\text{K}^+ \text{ channel} & \quad C \rightarrow 1 \\
\text{Cl}^- \text{ channel} & \quad V^- \rightarrow D \\
\text{passive} & \quad V^p \rightarrow 0 \\
& \quad g^+ \rightarrow A \\
& \quad g^- \rightarrow I_i + \sum_{k \neq i} I_k \\
& \quad g^p \rightarrow \sum_{k \neq i} I_k
\end{align*}
\]

inspired the simple shunting network equation:

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

HYPERPOLARIZATION AND “ADAPTATION LEVEL”

For hyperpolarizing network equation

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

with \( C > 0 \),

at equilibrium:

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

\( G \) refers to \( \frac{C}{B+C} \) as the “adaptation level” that must be exceeded by \( \theta_i \) in order to trigger a positive response for node \( x_i \).

Note: In vivo, generally \( B \gg C \), so \( \frac{C}{B+C} \ll 1 \)

That is, the range between saturation level and resting level (“zero”) is generally greater than the range between resting level and maximum hyperpolarization level.

**G: “SYMMETRY-BREAKING MECHANISM”**

**DYNAMICS (typical):**

The positive (output producing) range is extended (amplified).

The negative (hyperpolarized) range is compressed.

For non-distance dependent network with hyperpolarization:

\[
\begin{align*}
\text{extended} & \quad B/C = (n-1)/1 \\
\text{compressed} & \quad \text{i.e. an input has to be greater than } 1/n\text{ in order for output to be positive.}
\end{align*}
\]

**GEOMETRY (typical):**

The spatial range of sensitivity to excitatory signals is narrower.

The spatial range of sensitivity to inhibitory signals is broader.
FEATURAL NOISE SUPPRESSION*

A network with hyperpolarization can have parameters set so as to attenuate uniform regions of a pattern, which contain no information. (“Just a DC shift.”)

\[ I_i \xrightarrow{\text{hyper}} x_i \]

Recall, for simple network:

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

Result: Uniform pattern is suppressed independent of intensity of \( I_i \).

\[ C \left( \frac{1}{n} = \theta_i \right) \]

\[ x_i = 0 \]

*G jargon alert!

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](image)

Remember:
All bottom nodes connect to all top ones.

MULTI-CHANNEL NETWORKS*

Consider a network anatomy such as:

![Multi-channel network](image)

No more fan-out than to single nearest neighbor is implied here.

Here one can meaningfully speak of “center-surround” processing, and perform featural noise suppression for uniform regions of a globally nonuniform pattern (a.k.a. “edge enhancement.”)

* Jargon alert: See next panel.

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 in the “center” of particular others.
G’s use of the phrase “multi-channel” for distance-dependent (localized) spatial interactions in a network is not the same as what most of the vision world calls “multiple channels” of spatial frequency tuning. In place of the latter concept, Grossberg typically refers to “multiple scales” -- although multiple scales are not viewed as being sensitive to spatial frequency as such, but to a complex of attributes, including spatial frequency.

**DISTANCE-DEPENDENT KERNELS**

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

**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:

\[
\begin{array}{cccccccc}
1 & 4 & 1 \\
3 & 2 & 5 & 6 & 6 & 6 & 5 & 7
\end{array}
\]

\[
\begin{array}{cccccccc}
1 & 4 & 1 \\
3 & 2 & 5 & 6 & 6 & 6 & 5 & 7
\end{array}
\]

apart from “outer boundary effects”

* Note: This network has both subtractive and shunting inhibitory terms.

**EQUILIBRIUM OF DISTANCE-DEPENDENT NETWORK**

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

\[
I \sum_{k=1}^{n} \theta_k (BC_{ki} - DE_{ki})
\]

\[
= \frac{x_i}{A + I \sum_{k=1}^{n} \theta_k (C_{ki} + E_{ki})}
\]

**Result:**

\[\text{DOG} \quad \text{(difference of Gaussians)}\]

\[\text{SOG} \quad \text{(sum of Gaussians)}\]

scaled against constant, \(A\)

DISTANCE-DEPENDENT FAQs

\[
I \sum_{k=1}^{n} \theta_k (BC_{ki} - DE_{ki}) \quad C_{ki} = C \exp\left[-\mu(k-i)^2\right] \\
A + I \sum_{k=1}^{n} \theta_k (C_{ki} + E_{ki}) \quad E_{ki} = E \exp\left[-\nu(k-i)^2\right]
\]

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?

Notes: Rectification of output is generally assumed.

\( x_i \) 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.

*Yes, but, WHAT about the pattern of \( x_i \)’s demonstrates that sensitivity?

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:

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

In practice, you must pay attention to truncation effects when coding discrete approximations of Gaussians.

To get a numerical zero output (after rectification) at a given node, it may be necessary for the input pattern to be homogeneous over as many network locations as the size of the discrete inhibitory kernel.

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.

SHUNTING FOR FUN AND PROFIT

Consider this figure from Waxman et al., (1997), *Neural Networks*, 10(1), 1-6:

Quoting from caption of Figure 1:
“(a) High-end of 12-bit dynamic range; (b) low-end of 12-bit dynamic range; (c) entire 8-bit dynamic range after center-surround shunting neural processing of original 12-bit imagery. All objects are clearly visible without saturation at either end of the compressed dynamic range.”

“BIOLOGICAL PLAUSIBILITY” OF SHUNTING IN CORTEX*


...looked for conductance changes as evidence for shunting, (with respect to directional selectivity)

*Conclusion:* Shunting inhibition is too small to measure.


...looked for linearity of synaptic summation

*Conclusion:* Synaptic summation is linear-- ergo no shunting.

Subsequent papers rule out shunting on dendritic spines and other loopholes.

Note that “shunting networks” are only incompatible with this data in an extreme parameter range -- i.e. when $D=0$, the shunting network is shunting (only); nonzero $D$ values admit subtractive (nonshunting) inhibition.

* Thanks to David Somers for info in this and next three panels.
POINTS TO PONDER, CIRCA 1998

1. Until 1998, it appeared that shunting inhibition was not performed by cortical neurons. This is, perhaps, still believed by many experimentalists and biases them against “shunting” models.

2. To a physiologist, “shunting inhibition” is purely silent, meaning that if you stimulate an inhibitory fiber alone, no hyperpolarization of post-synaptic cells will occur. Biophysically this means that the reversal potential of inhibition is at the resting potential (or \( D = 0 \)) and, more importantly, that large conductance changes occur. The large conductance changes have a divisive or normalizing effect on excitatory current. Typical, real conductance changes are thought to be on the order of 20-30%. “Heeger’s Normalization Model (i.e., a shunting network) requires 400% conductance changes to achieve normalization and hence did not appear to be “plausible.”

3. Conductance change is equivalent to the change in the size of the denominator of the equilibrium solution of the shunting equation. So, the data appear to rule out significant inhibitory normalization via the denominator (even if \( D \) is not exactly zero).

4. Something like normalization clearly occurs in cortex, but at a slower time scale than spiking (seconds, rather than milliseconds).

CORTICAL SHUNTING REDUX!

But wait! Just when cortical shunting appeared at death’s door, Judith Hirsch (Rockefeller) and Lyle Borg-Graham (France) have independently reported measuring conductance changes on the order of 200 -- 300% at recent conferences!

See:


For further info and updates on cortical shunting, contact David Somers (CNS alumnus, B.U. Psychology professor, smart guy, and all-around swell person) at: somers@bu.edu.

LEADING CANDIDATE MECHANISMS FOR CORTICAL NORMALIZATION (STILL CIRCA 1998)

1) synaptic depression at active synapses
2) metabotropic (non-ionotropic) glutamate receptors
3) intrinsic modulation (normalization) of synaptic currents.

It is also possible that the network is very clever and somehow balances hyperpolarizing (non-shunting) inhibition with excitation.

Bottom lines:

1) Normalization is important, and does occur in cortex.
2) “Shunting network model” has simple and intuitive way of achieving normalization.
3) Apparently, something other than shunting as such may be performing normalization in cortex (on a relatively slow time scale.)

Connections between Psychophysics, Electrophysiology, and Modeling

1. Can a psychophysics experiment suggest an electrophysiology test?
2. Can psychophysics suggest a modeling idea?
3. When are the electrophysiology results incorporated into the modeling ideas?
4. How can modeling ideas give a better view of electrophysiological results?
Connections between Psychophysics, Electrophysiology, and Modeling

Psychophysics Experiments ↔ Electrophysiology Experiments

Modeling Idea (a) ↔ Modeling Idea (b)

Example 1:
A triangle occluding three circles

Example 2:
A grouping across scenic positions not receiving contrastive input

Psychophysics Experiments ↔ Electrophysiology Experiments?

Oriented receptive fields

Oriented receptive fields have a limited spatial range

[Diagrams showing various arrangements and responses]

Psychophysics Experiments ↔ Electrophysiology Experiments?

Illusory contours and Electrophysiology
von der Heydt (1984)

There are cells in V2 that respond to illusory contours when their receptive fields are just on the illusory part

When inducers are offset the response disappears
Corresponding Modeling Idea
Cohen and Grossberg (1984)
Grossberg and Mingolla (1985)

The old model of bipole has two lobes.
Both lobes have to have excitatory inputs for bipole activity.
Result: Interpolation, Yes
Extrapolation, No

Modeling Idea

Psychophysics Experiments
Electrophysiology Experiments

Laminar Organization of Cerebral Cortex

Cerebral cortex

1. Molecular layer
2. External granular layer
3. External pyramidal layer
4. Internal granular layer
5. Internal pyramidal layer
6. Multiform layer

In vitro preparation of layer 2/3 of Ferret visual cortex.
Caveat: Layer 2/3 is in isolation, getting sound inference needs more care.

Initial positive area of activity is followed by the expanding ring of suppression.

More Electrophysiology Data

The panels are shown in sequence to give you a better impression of excitation and inhibition dynamics.
Another activity spread after stimulation

Boundary representation that emerges from long range excitation and short range inhibition

Contribution to Modeling Idea

Electrophysiology Experiments

Modeling Idea

Clusters of long range positive connections accompanied by short range inhibitory interneurons

Does it remind you of an older modeling idea?
Emergent Older Modeling Idea  
Grossberg and Raizada, 2000

The following model of bipole grouping cells emerged from the psychophysical property (interpolation, yes, extrapolation, no) + Following electrophysiological data

2/3 pyramidal cells to 2/3 pyramidal cells  
Bosking et al. (1997, shrew), Schmidt et al. (1997, cat)

2/3 pyramidal cells to 2/3 inhibitory interneurons  

2/3 inhibitory interneurons to 2/3 inhibitory interneurons  
Tamas et al. (1998, cat)

Compare this circuit with the previous panel circuit. What are differences? At which level this difference may generate discrepancy?

Aperture Problem

What is the difference between the direction of movement inside the aperture and outside it? This may cause trouble!

Aperture Problem: Any local motion detector can measure only the component of motion perpendicular to a contour that extends beyond its field of view.

It is especially true for small V1 receptive fields: 0.2 – 0.5 degree of visual angel. MT receptive field size are around 3-5 times larger than V1 cells at the same eccentricity.

The length of each bar was always 3 degree, significantly longer than corresponding receptive fields in V1, but smaller than the excitatory receptive parafoveal and peripheral MT.

This is such to encourage the even excitation by endpoints.
Direction Tuning Polar Curve

Average activity of the cell at each direction is plotted in a polar coordinate and connecting them by straight lines.

Results 1
Resolving the aperture problem takes time:

Earlier
Later

MT cell has a ~150 ms time delay to resolve aperture problem, but it can solve it finally.

Results 2

Lorenceau J., et al., 1993

Seeing MT Cells!
Lorenceau J., et al., 1993
From Electrophysiology to Psychophysics

We can see our MT cells: Again psychophysics!

Lorenceau J., et al., 1993

The tilted bar moves horizontally back and forth, but we illusively observe a temporary upward/downward movement of the middle point that turns to real direction.

Psychophysics Problem
Electrophysiology Data

Close the Loop by Modeling

Remember the inward illusory contour mechanism:

One idea could be that the detection delay of the real direction is due to the same inward mechanism from line ends to the middle (?)

Grossberg and Mingolla (1993)

Psychophysics Experiments
Modeling Idea
Electrophysiology Experiments

da Vinci and Classic Stereopsis

In (a,1), occlusion configuration blocks the view of the object, therefore stereogram (a,2) can reproduce the viewing condition.

In (b,1), both eyes see the farther object with their line of sight having different distance from the nearer object edge.

(a)
(1)
(2)
L R
L R

(b)
(1)
(2)
L R
L R

Anatomy: Regions with Different Eye of Origin

Many primary visual cortex cells still input from one eye
da Vinci stereopsis, relies on the eye of origin, therefore, originates in V1
Shimojo and Nakayama (1990)

Classic stereopsis, relies on the eye of origin too, therefore originates in V1
Poggio (1991)

Open Question for Modelers
Grossberg and Howe (2003)