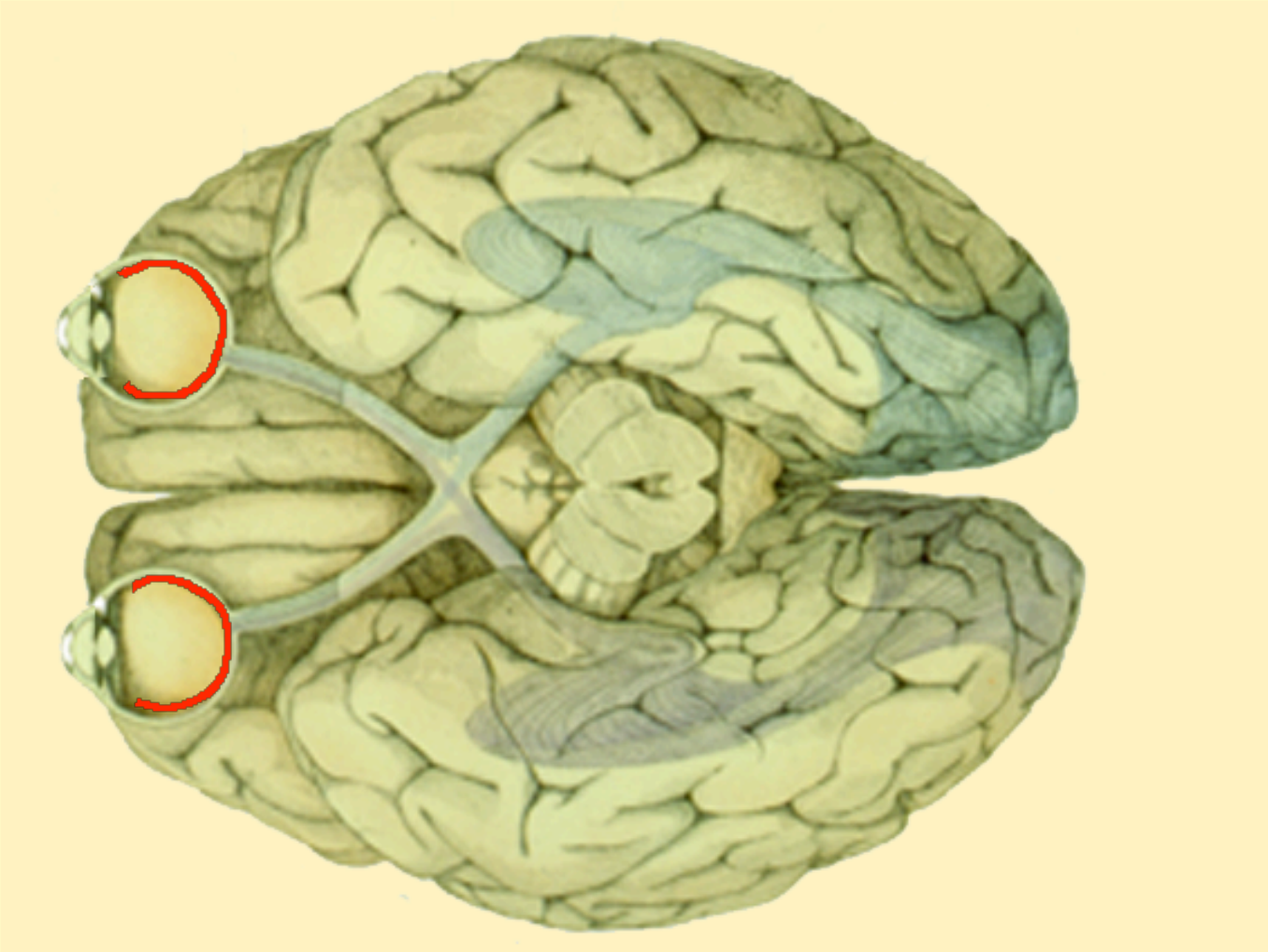
Bright vs. dark in natural scenes:
Implications for retinal structure

- Vijay Balasubramanian (Penn)
- David Brainard (Penn)
- Harry Kao (Penn)
- Charles "Dutch" Ratliff (Penn)
- Peter Sterling (Penn)

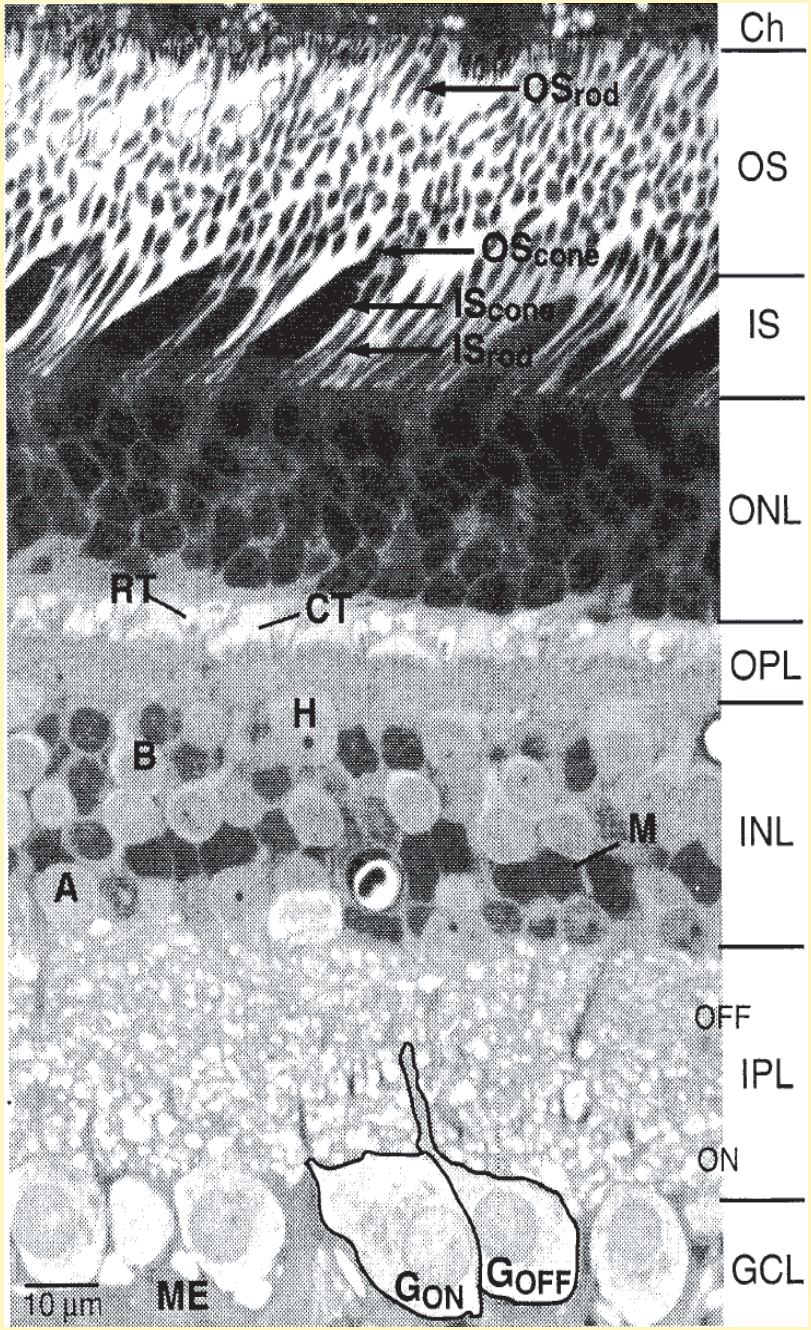
Also:
- Michael Berry (Princeton)
- Michael Freed (Penn)
- Don Kimber (Xerox)
- Kristin Koch (Penn)
- Judith McLean (Penn)

The Eye and The Brain



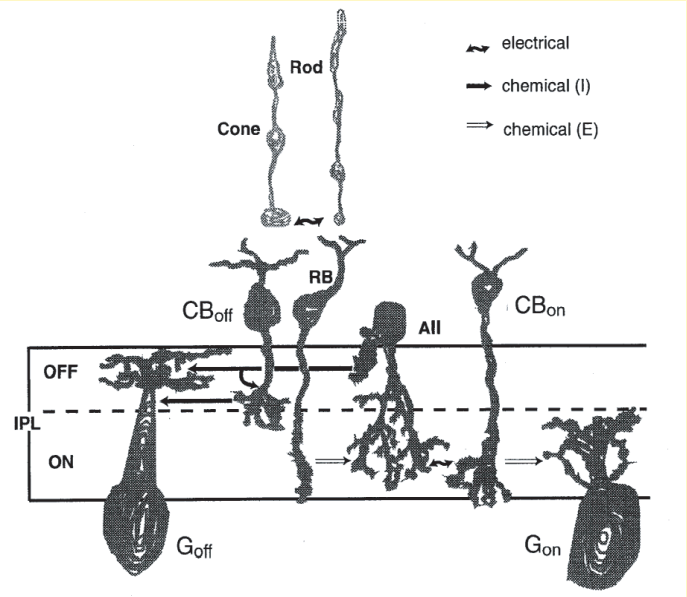
D. Hubel

THE RETINA



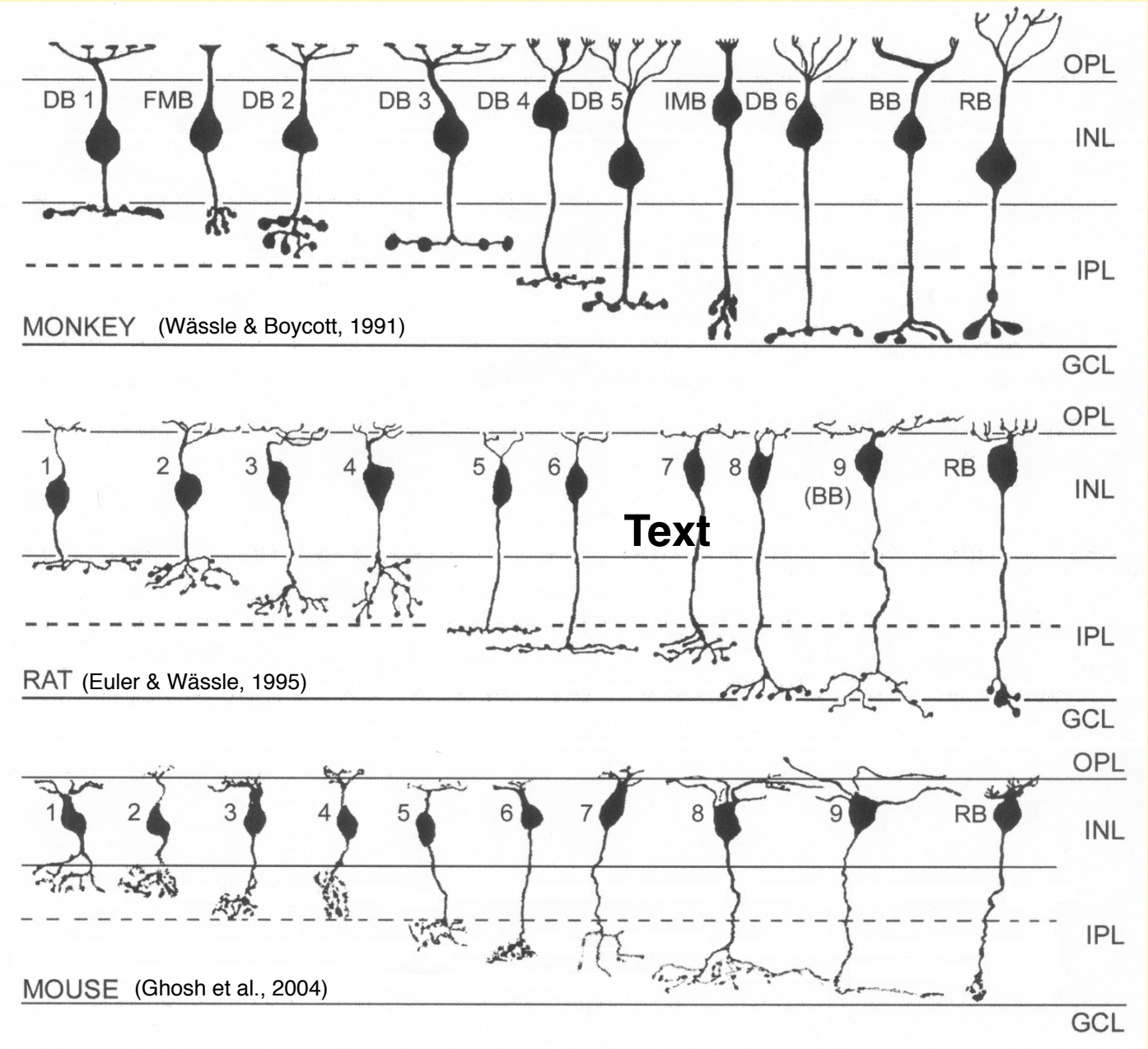
Section through monkey retina ~5mm from the fovea (Sterling 1999).

OS Outer segments
IS Inner segments
ONL Outer nuclear layer
INL Inner nuclear layer
CT Cone terminal
RT Rod terminal
OPL Outer plexiform layer
IPL Inner plexiform layer
GCL Ganglion cell layer
B Bipolar cell
H Horizontal cells
A Amacrine cells



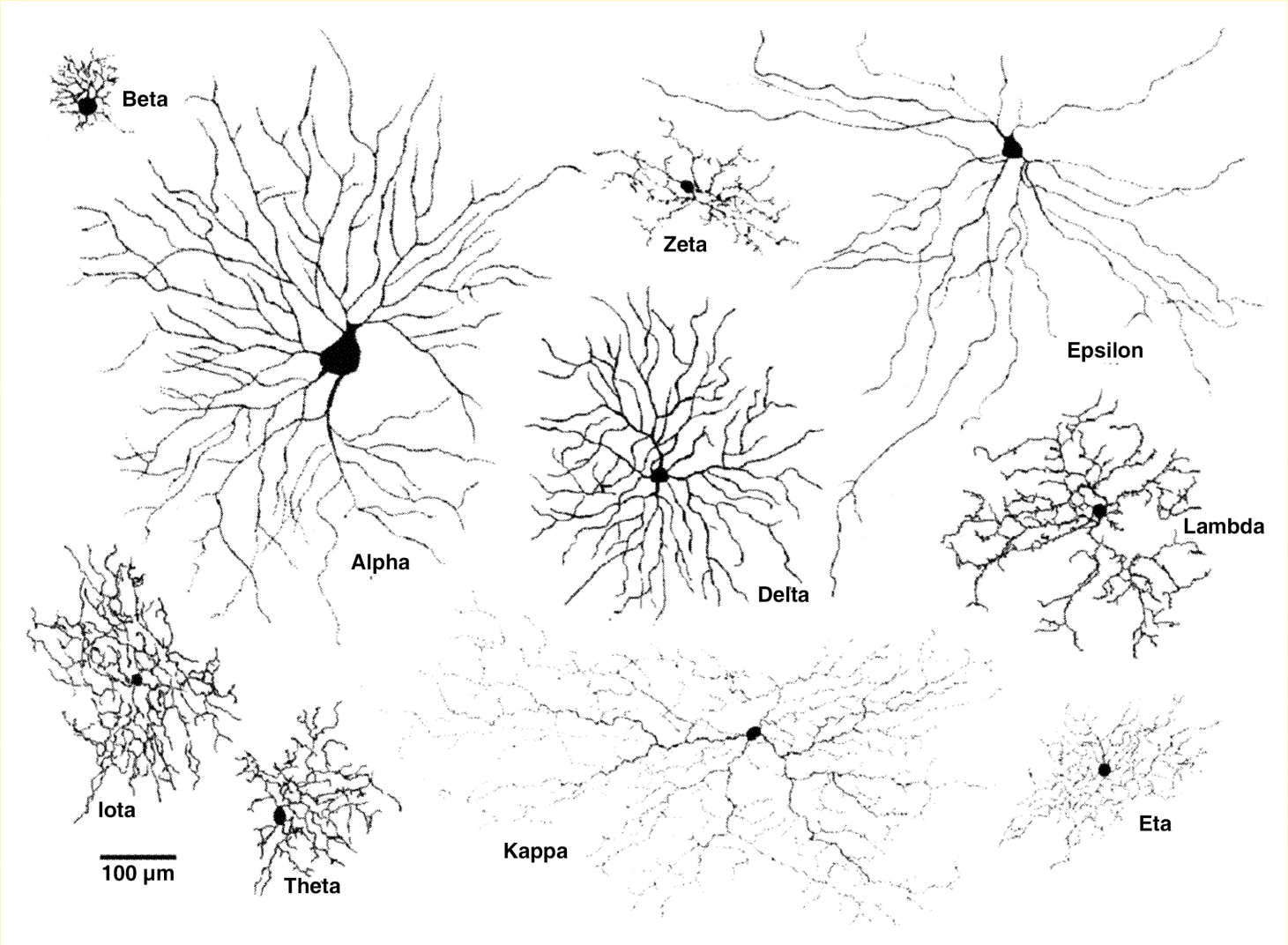
Schematic of retinal connectivity. ON and OFF cells arborize into separate sublayers of IPL. (Sterling 2004)

bipolar neurons: ten types



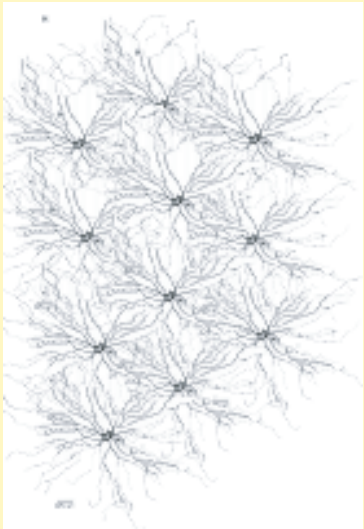
Ganglion Cells

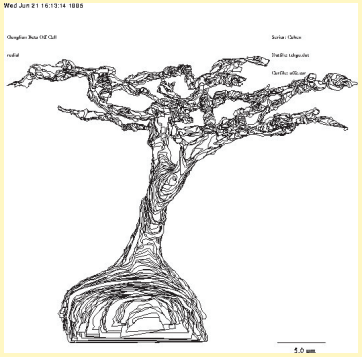
alpha: '*brisk-transient*'
beta: '*brisk-sustained*'
rest: '*sluggish*'



Masland, 2001
(Berson)

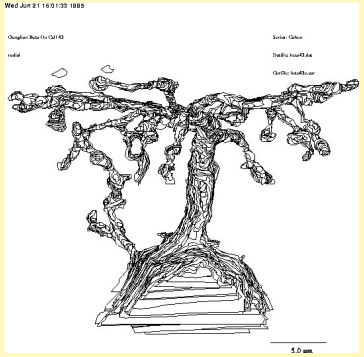
ON and OFF ganglion cells have evolved asymmetrically



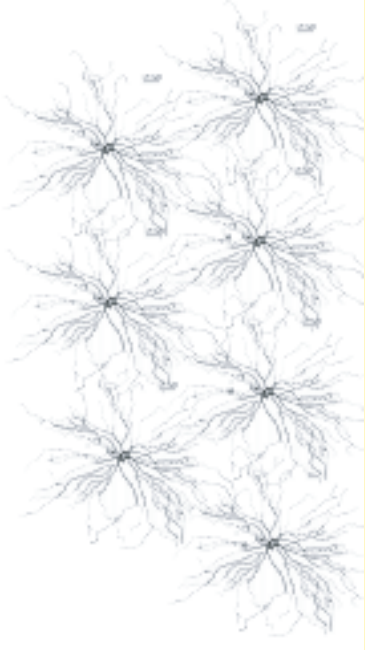


OFF-Beta cell

OFF (ON) cells respond to dark (bright) patches



ON-Beta cell



Across cell types and species OFF cells compared to ON cells:

- are 20-50% smaller
- are more numerous (e.g. 2.75:1 in guinea pig)
- have greater receptive field overlap (20% OFF vs 3% ON)
- sample at higher resolution with denser dendritic arbor
- respond to contrast with great rectification

Chichlinsky & Kalmar, 2002; Zhagloul, Boahen & Demb, 2003; Dacey & Peterson 1992; Kao & Sterling, unpublished

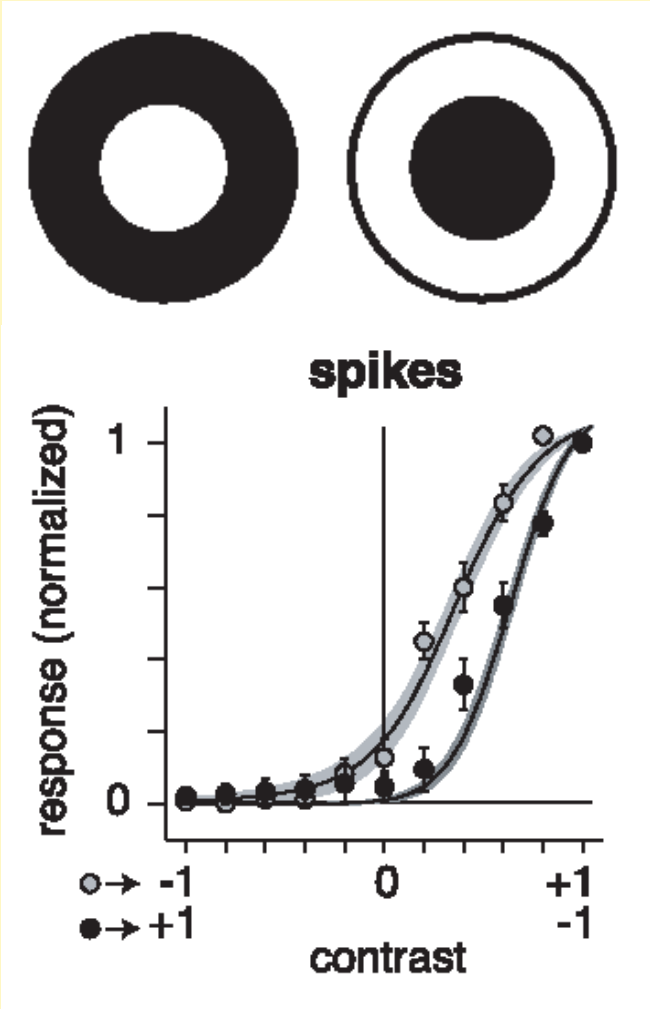
HYPOTHESIS

Structural and functional asymmetries
of OFF vs. ON channels
arise from asymmetries between
DARK vs. BRIGHT regions of natural stimuli

ON and OFF cells respond to local contrast

Preferred Stimulus

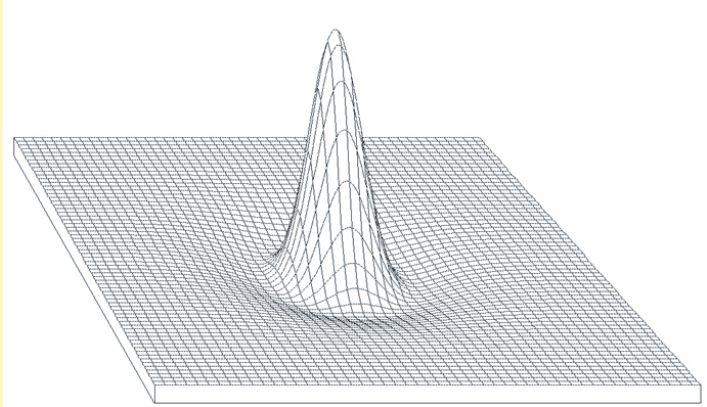
ON OFF



Zaghloul, Boahen & Demb (2003). ON vs. OFF spiking response to a spot centered over the dendritic arbor under photopic conditions. Response to flashes confirms that OFF cells are more rectified

Simple Contrast Model

Difference of Gaussians



$$\text{Contrast} = \frac{\int G_c(\vec{x}-\vec{x}^0)\, I(\vec{x}) - \int G_s(\vec{x}-\vec{x}^0)\, I(\vec{x})}{\int G_s(\vec{x}-\vec{x}^0)\, I(\vec{x})}$$

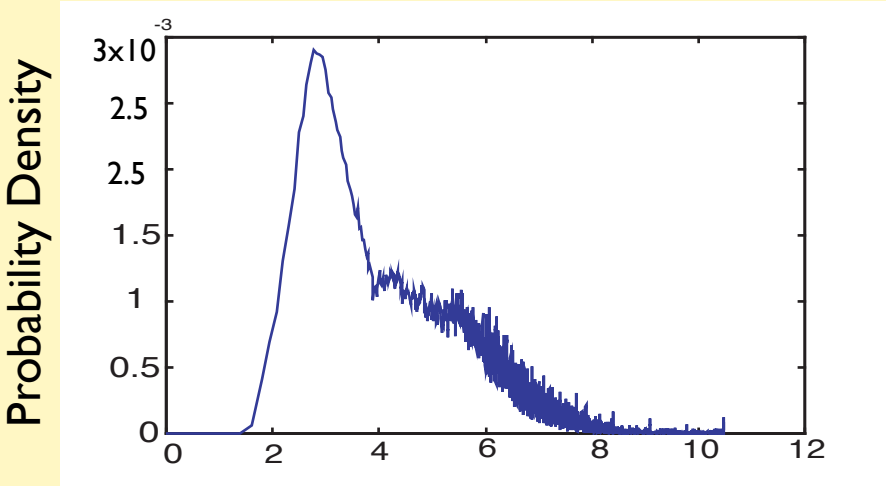
- Divisive normalization of intensity. (Tadmore and Tolhurst, 2002)
- Equal weight center and surround -- constant intensity gives zero contrast.
- Center radius = distance from center where weighting becomes negative
- Surround radius fixed at 5 times center radius

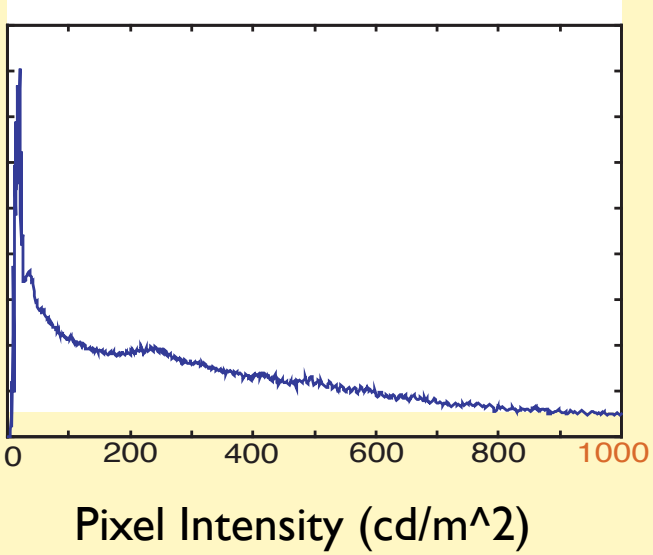
Images from van Hateren's natural stimulus collection

J.H. van Hateren & A. van der Schaaf, Proc. Roy. Soc. Lond. B (1998).
Photos taken near Groningen, Netherlands

- monochromatic
- 12 bit, 1536x12034
- Linear in intensity

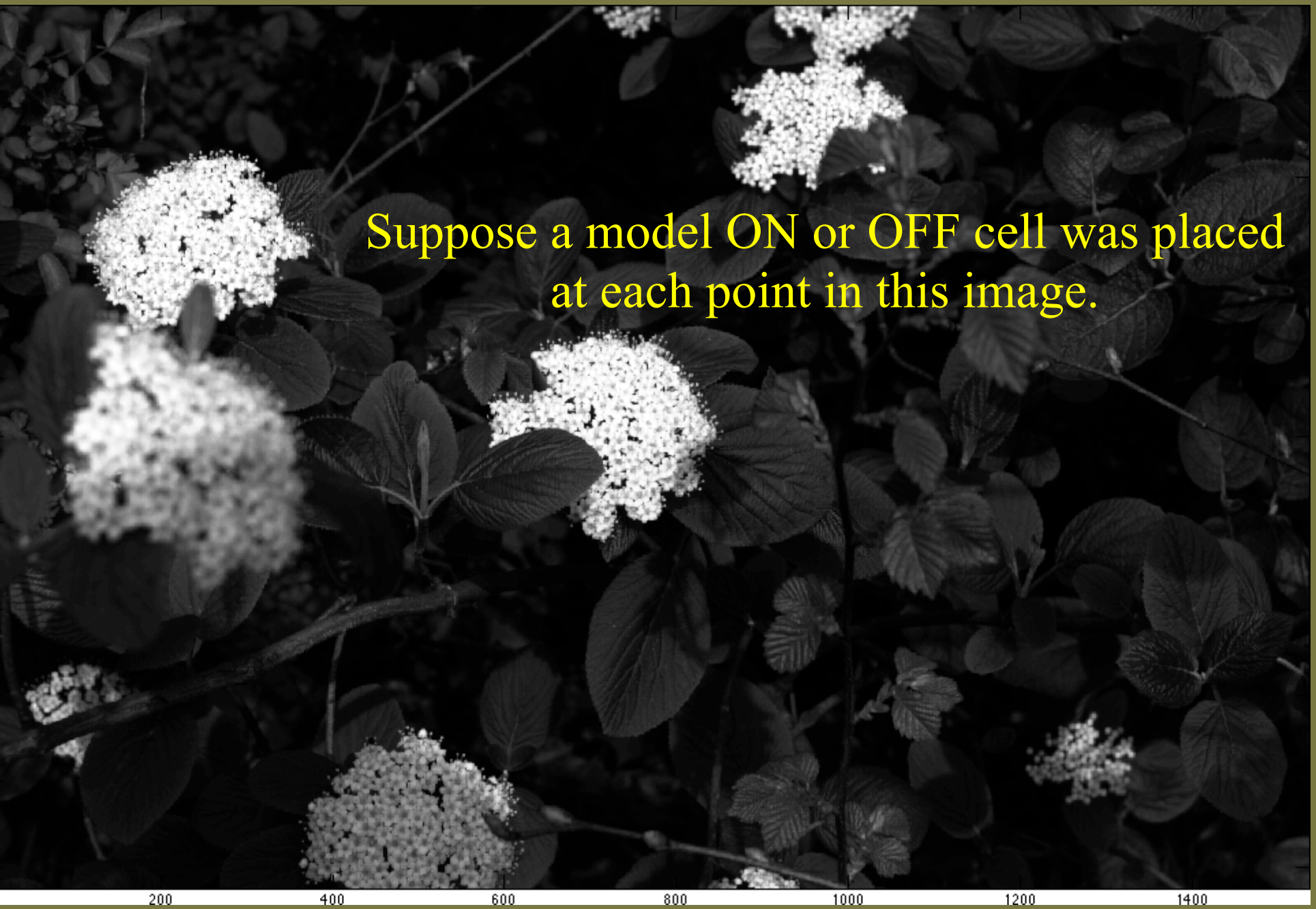
Intensity distributions from 100 natural scenes

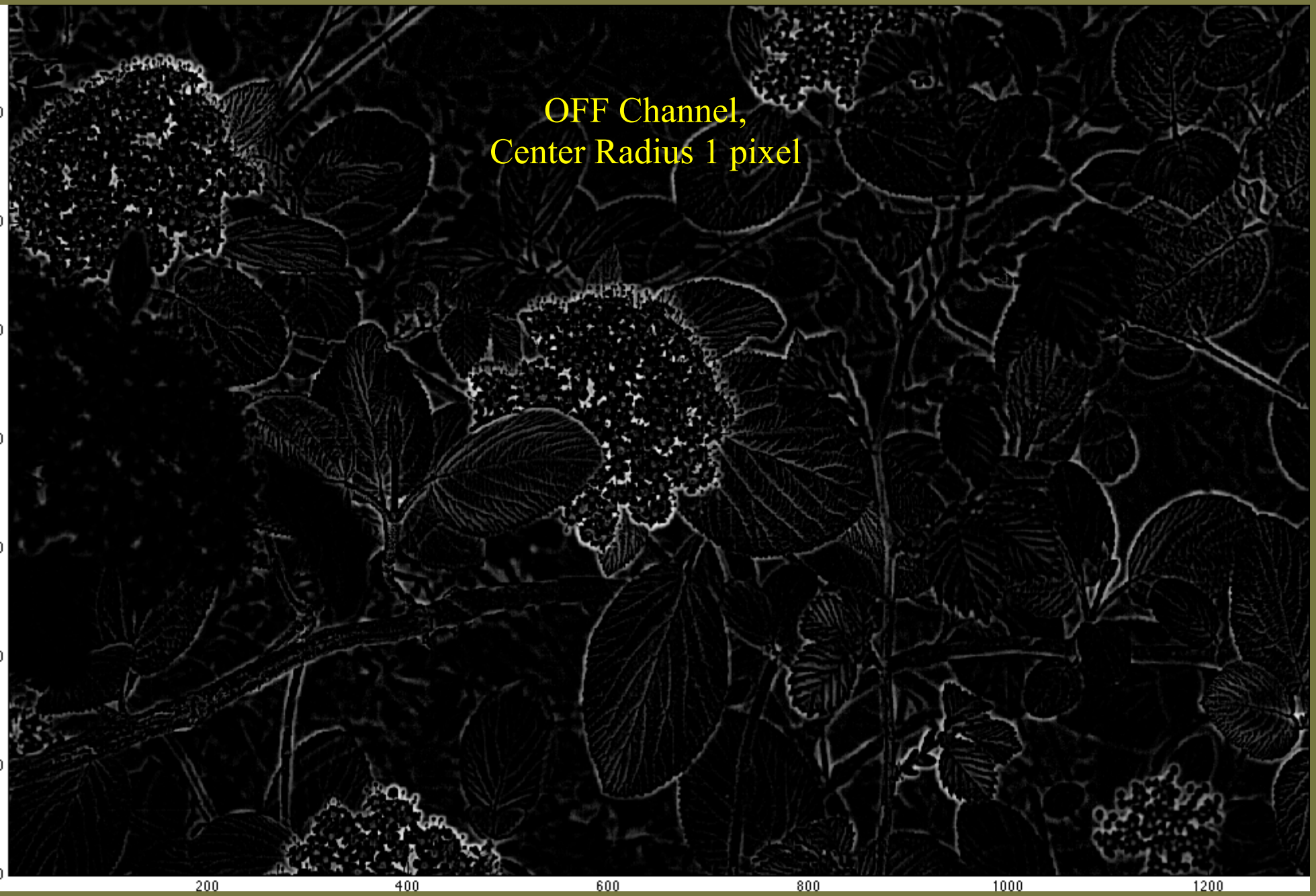


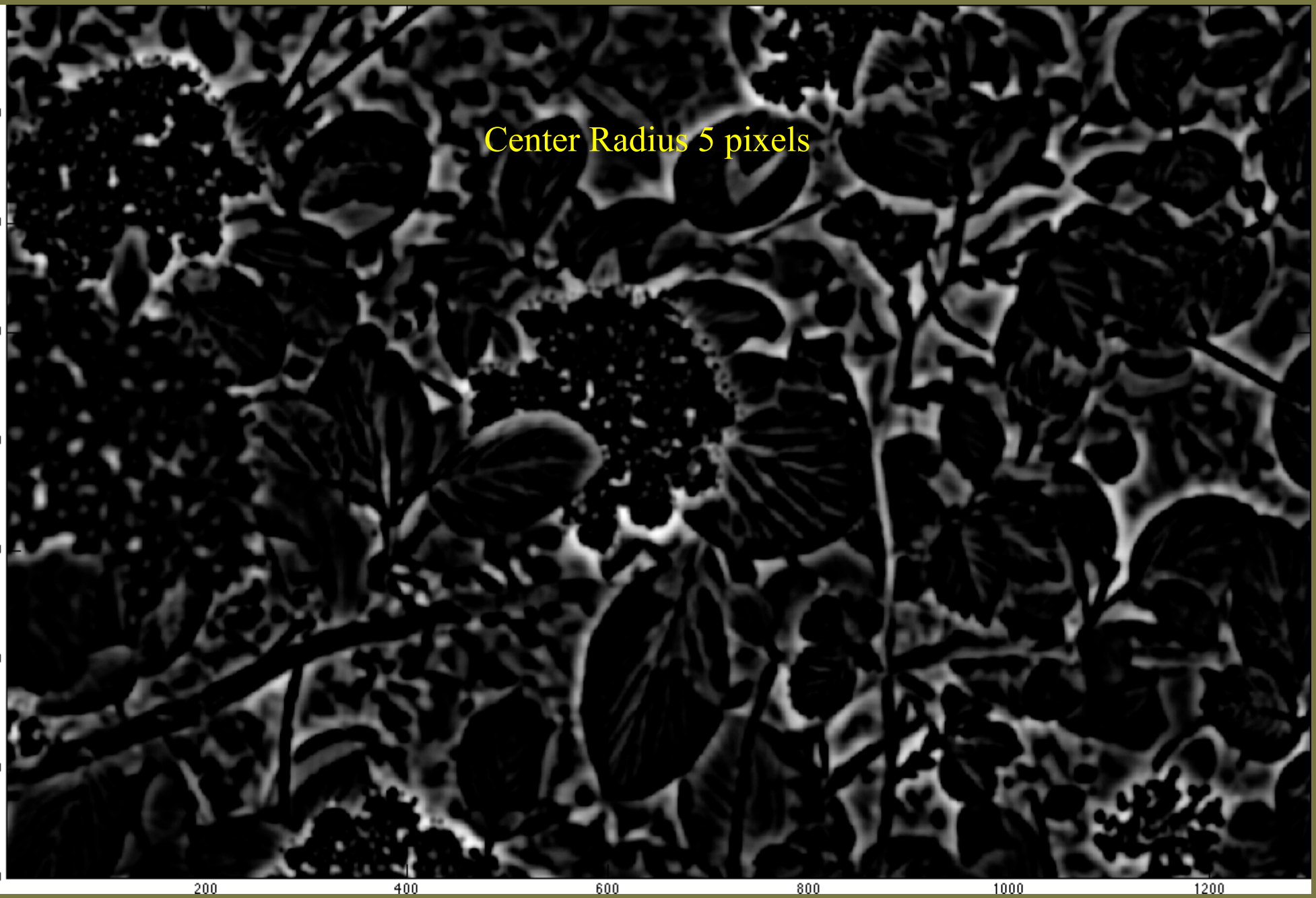


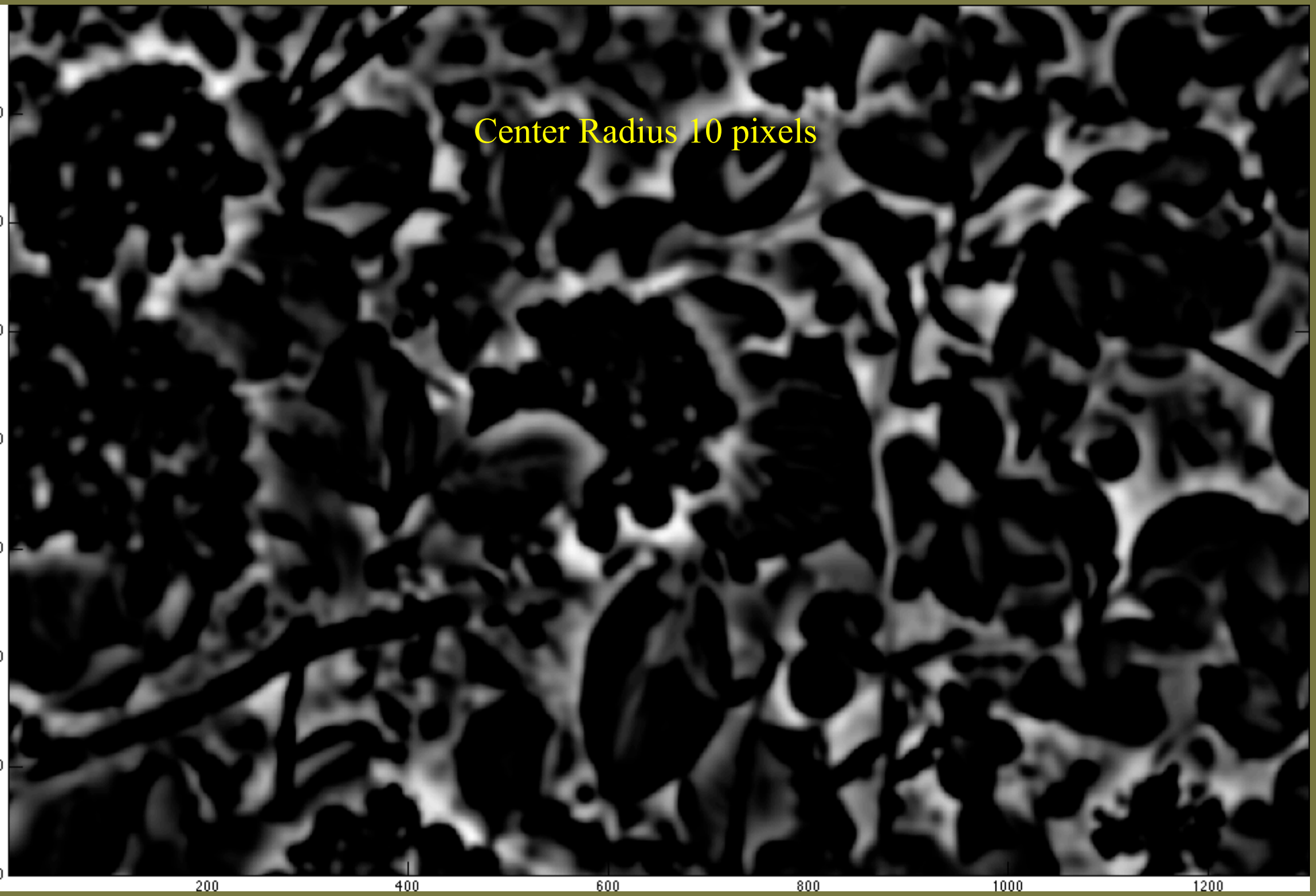
The intensity distribution is skewed around the median and has a very long tail.

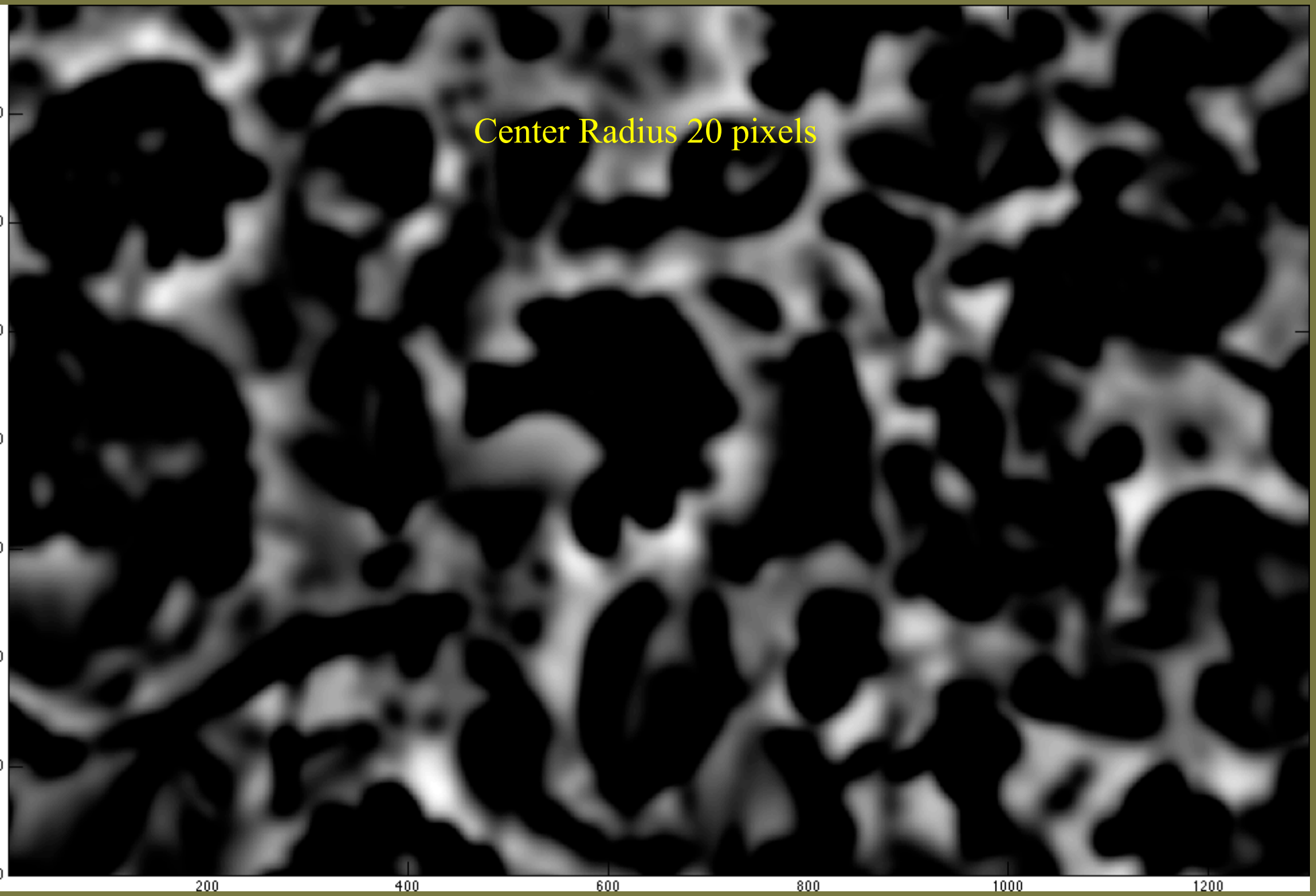
Individual scenes have distributions that can be quite different from the ensemble average.

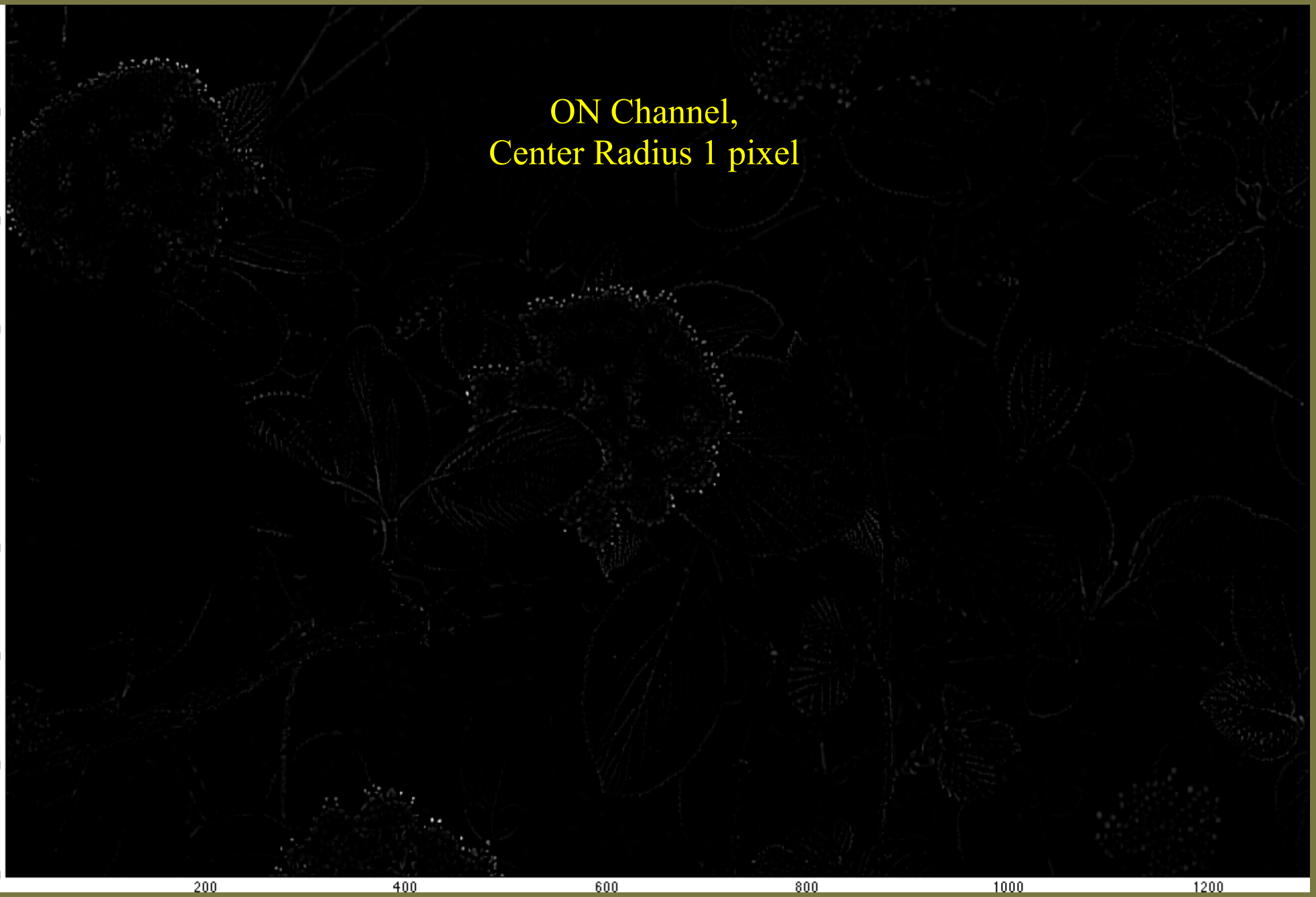

Suppose a model ON or OFF cell was placed
at each point in this image.
200
400
600
800
1000
1200
1400

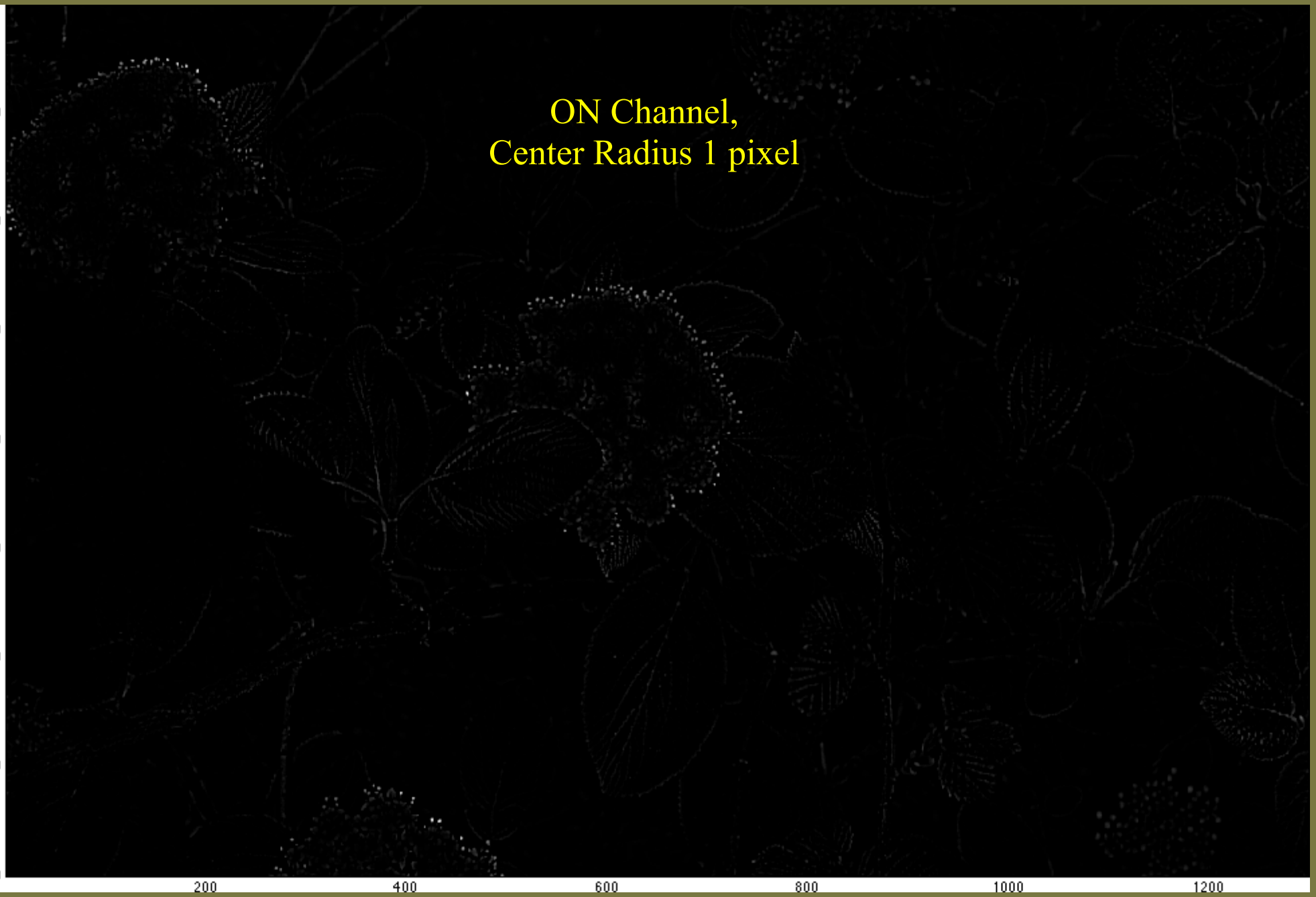

OFF Channel,
Center Radius 1 pixel
200
400
600
800
1000
1200

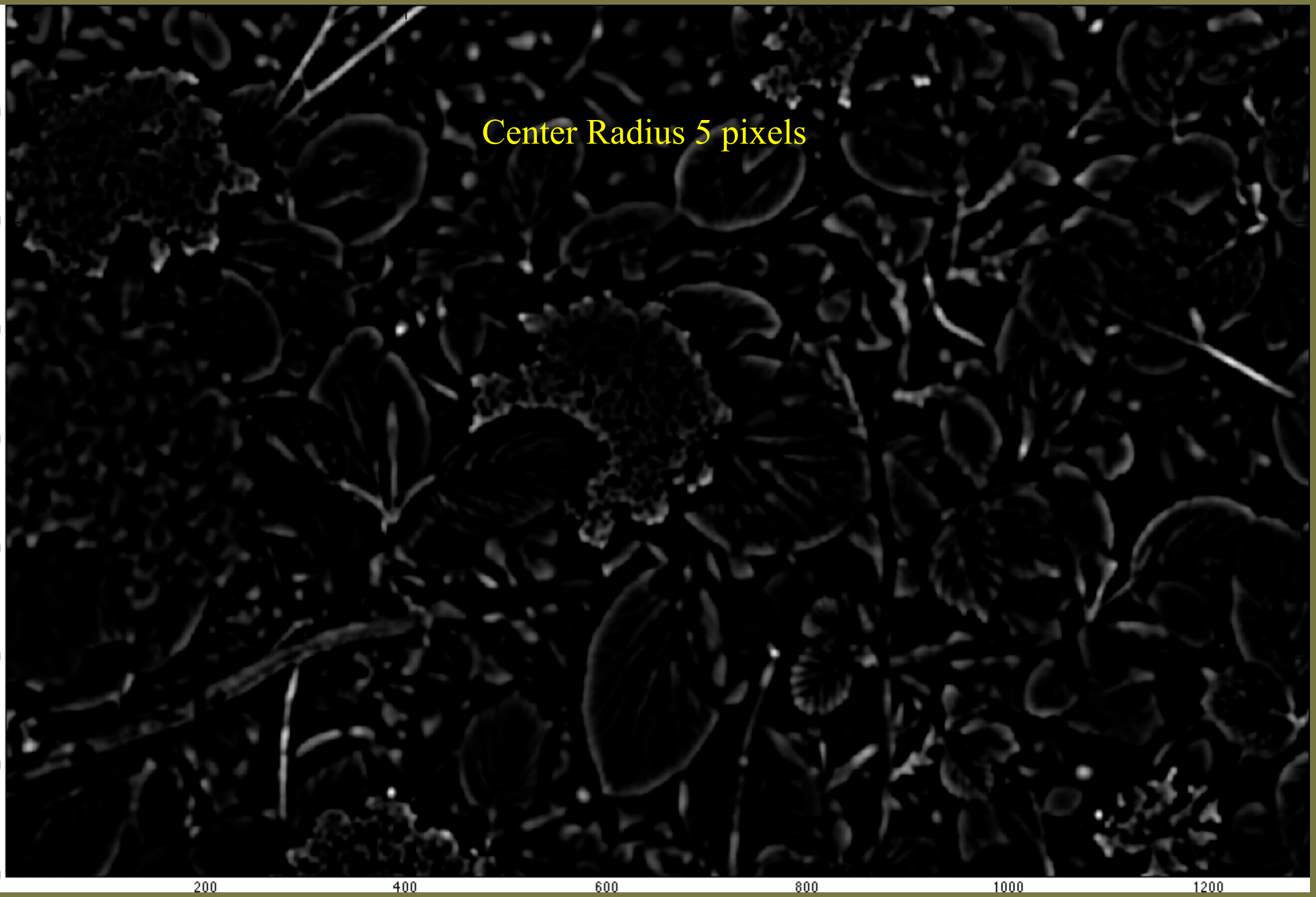

Center Radius 5 pixels
200
400
600
800
1000
1200

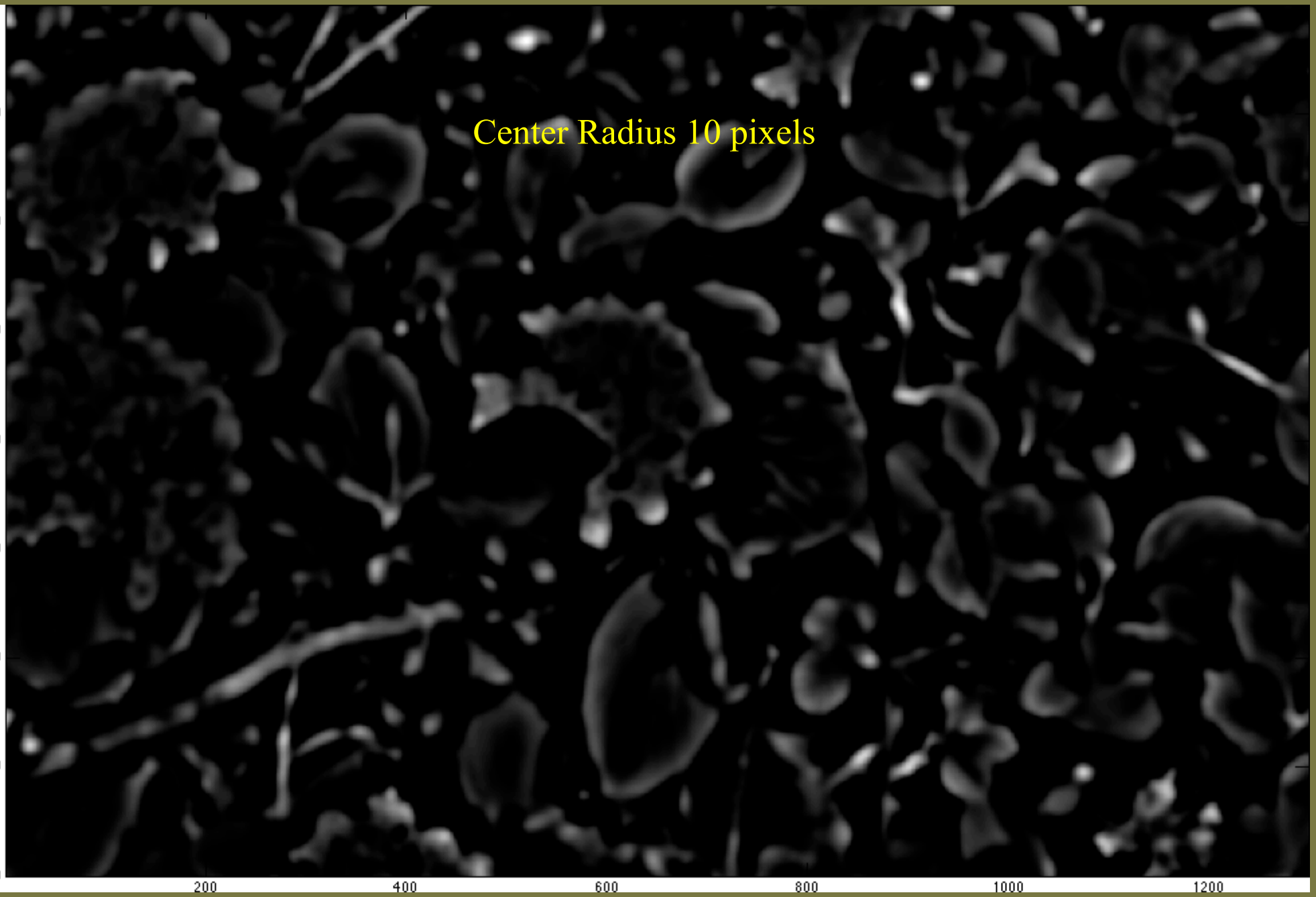

Center Radius 10 pixels
200
400
600
800
1000
1200

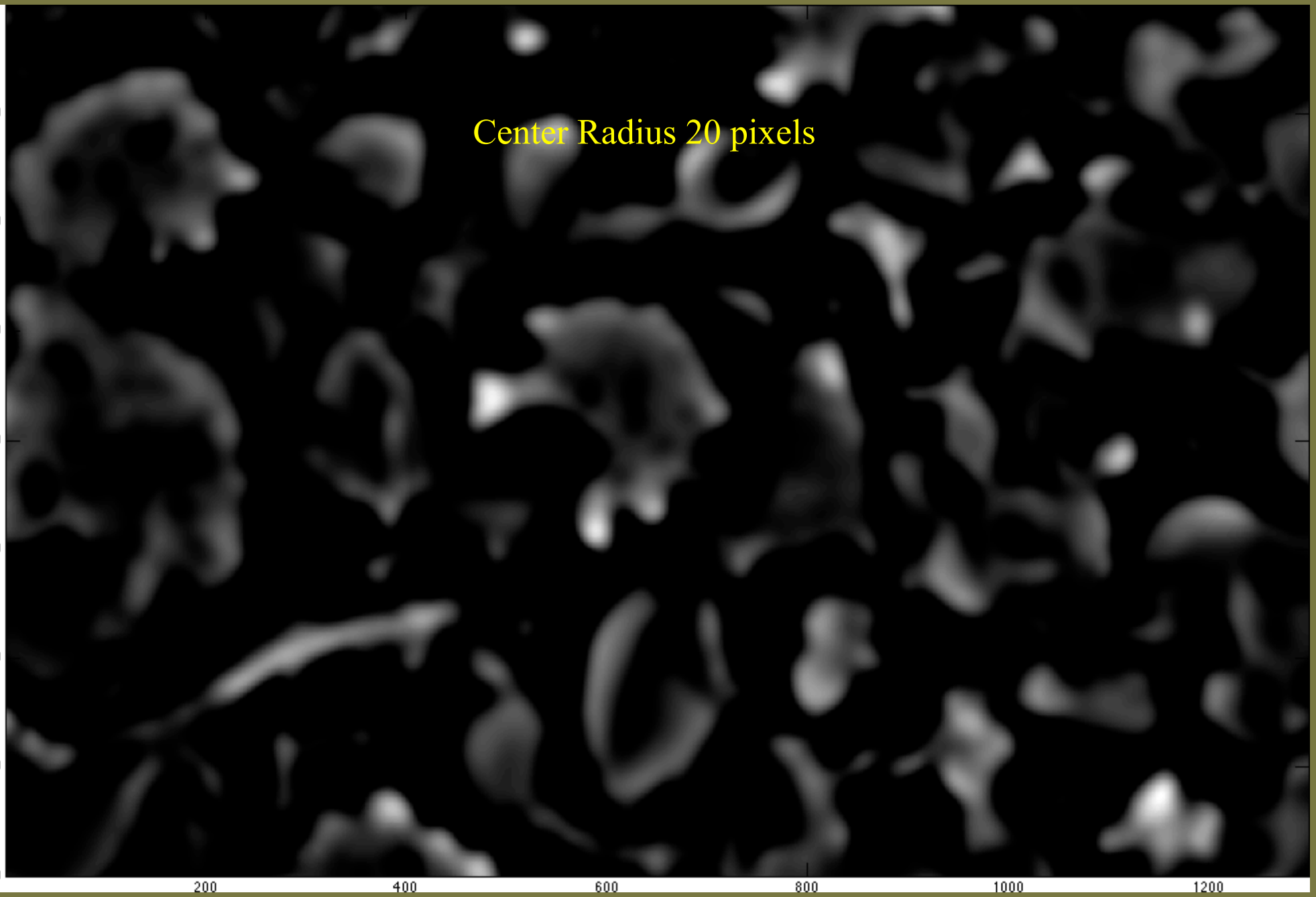

Center Radius 20 pixels
200
400
600
800
1000
1200


ON Channel,
Center Radius 1 pixel
200
400
600
800
1000
1200


ON Channel,
Center Radius 1 pixel
200
400
600
800
1000
1200

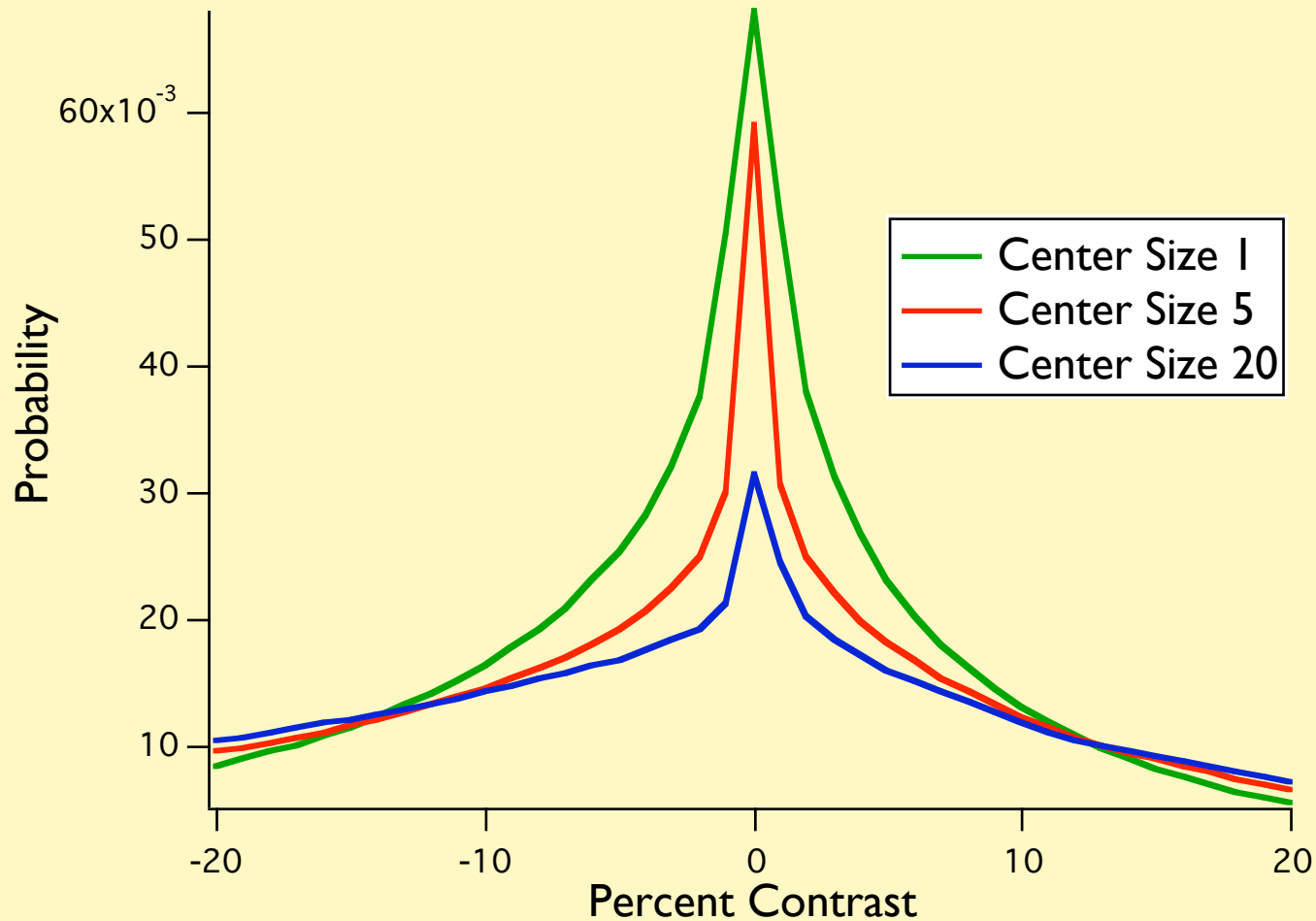

Center Radius 5 pixels
200
400
600
800
1000
1200


Center Radius 10 pixels
200
400
600
800
1000
1200


Center Radius 20 pixels
200
400
600
800
1000
1200

Distributions of local contrast in 100 natural images for DOG filters of different sizes

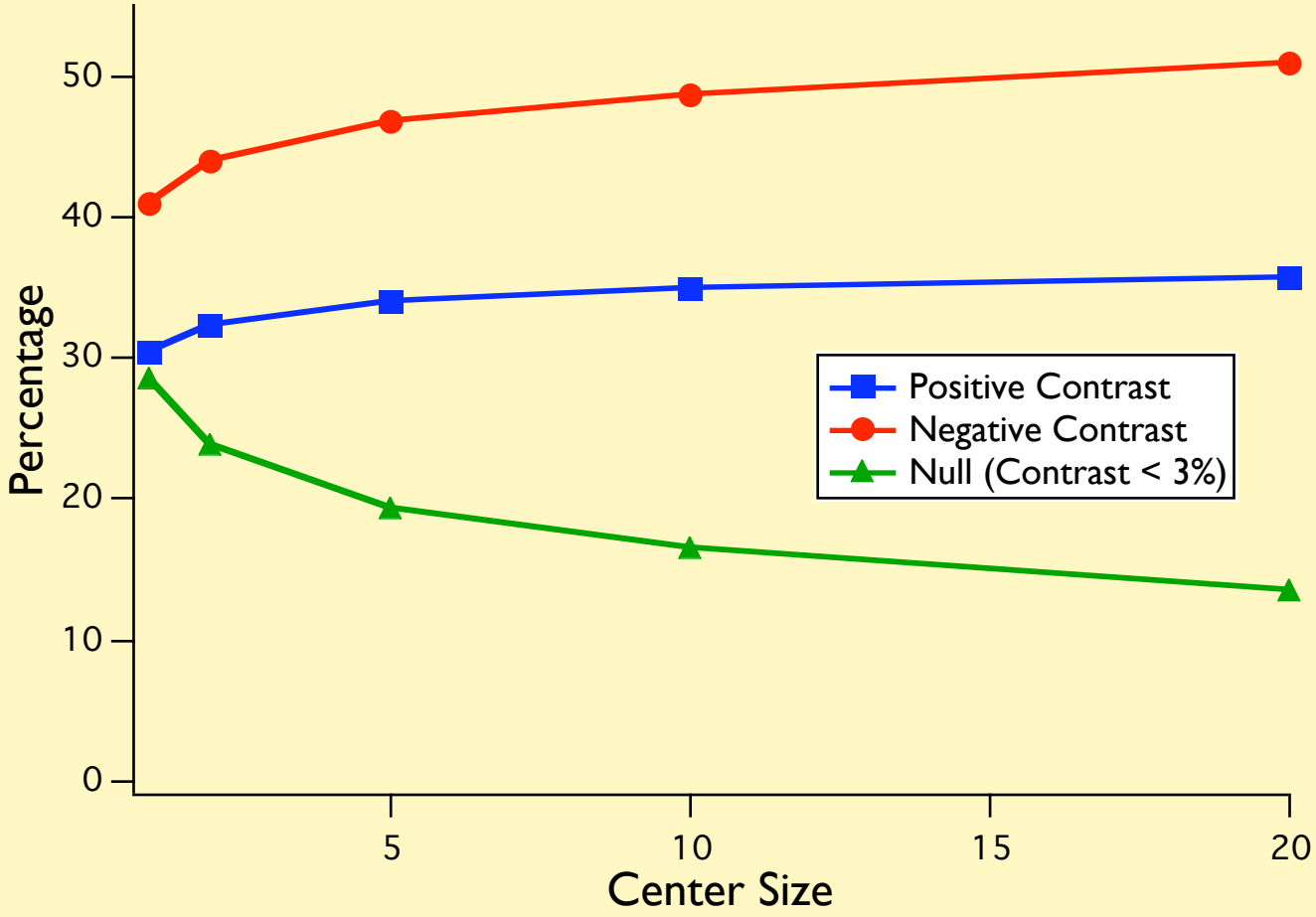
Skew towards negative contrasts



Smaller filters observe more small contrasts

- A sparser, more energy efficient code would result from splitting into separate bright and dark channels. (But this requires more cells to cover the scenes and thus more packing space. A tradeoff?)
- The skew suggests that more resources should be devoted to processing dark patches.

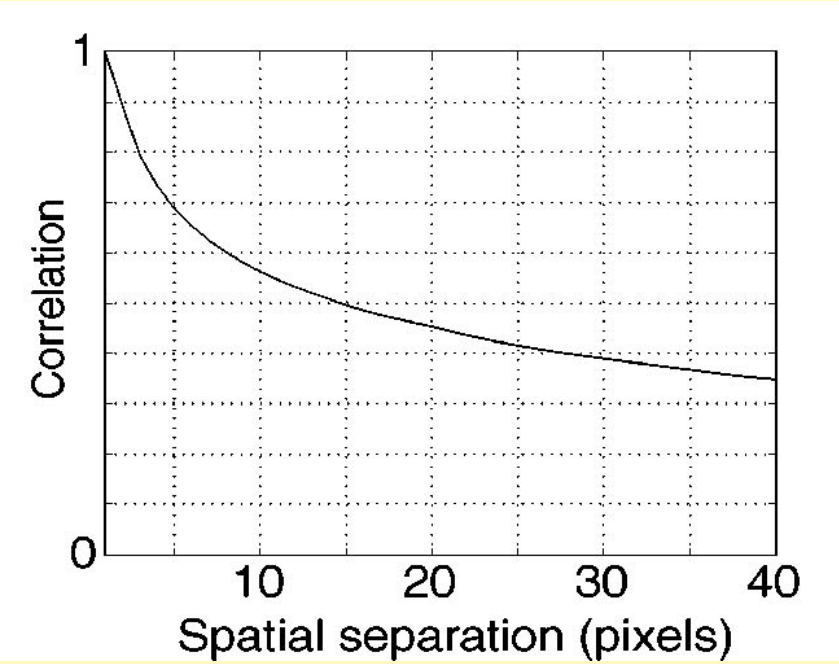
More Dark Contrast In Natural Images



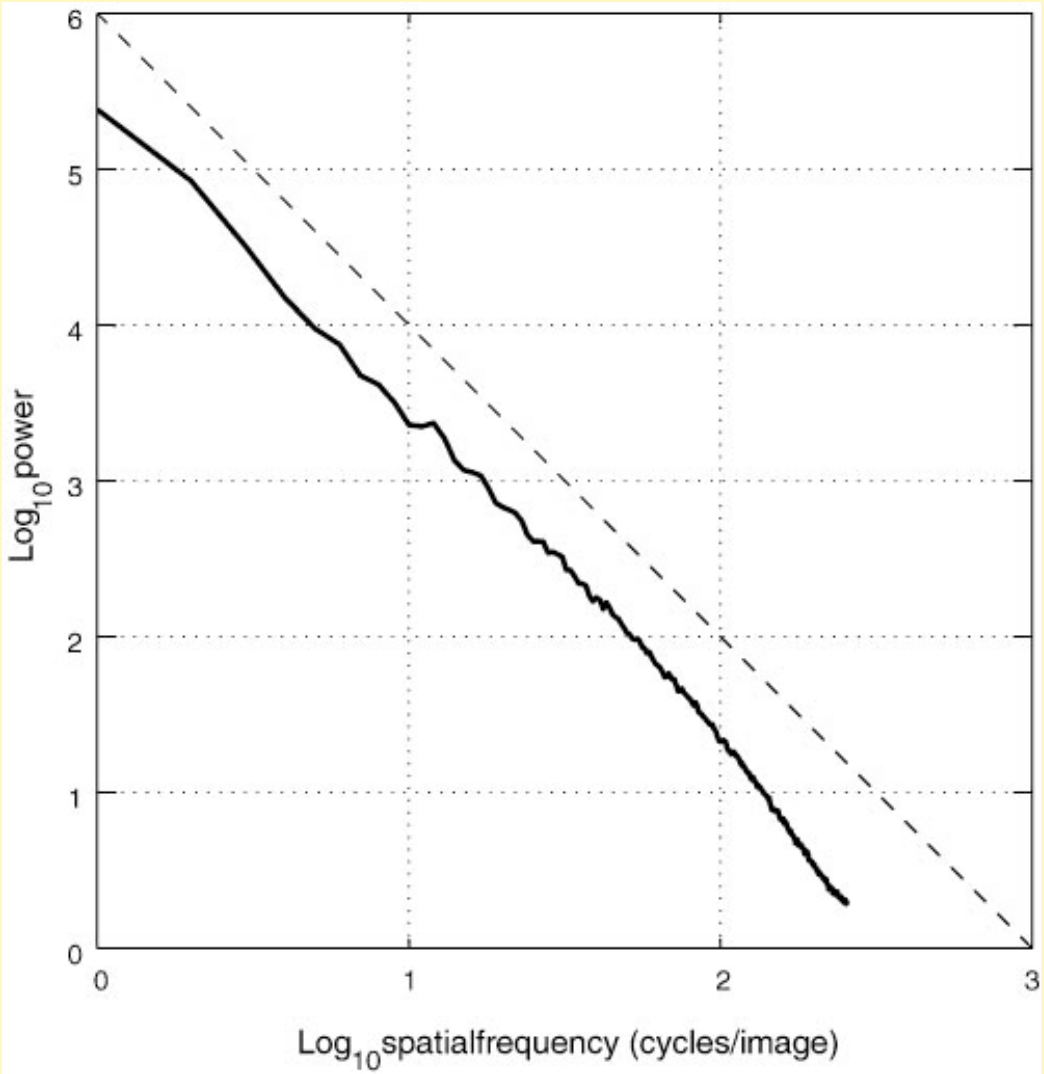
At all measured scales there are more negative contrasts than positive, suggesting that there should be more OFF cells than ON cells.

The Statistical Structure of Natural Scenes

(see e.g., Kersten (1987), Field (1987), Daugman (1989), Ruderman and Bialek (1994))



Autocorrelation function of intensities as a function of spatial separation. (Figure from Simoncelli & Olshausen, 2001)

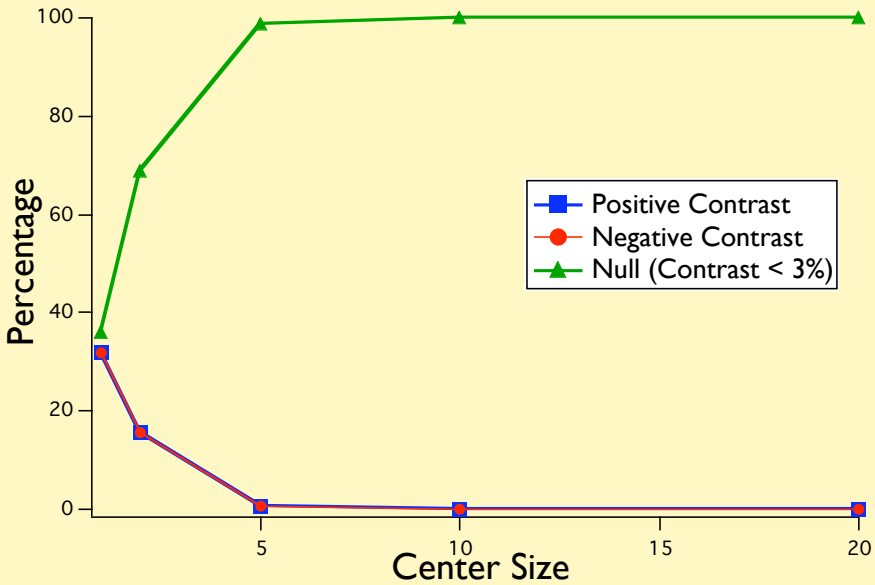


The Fourier transform of the autocorrelation function, averaged over orientations scales as 1/f^2.
(Figure from Simoncelli & Olshausen, 2001)

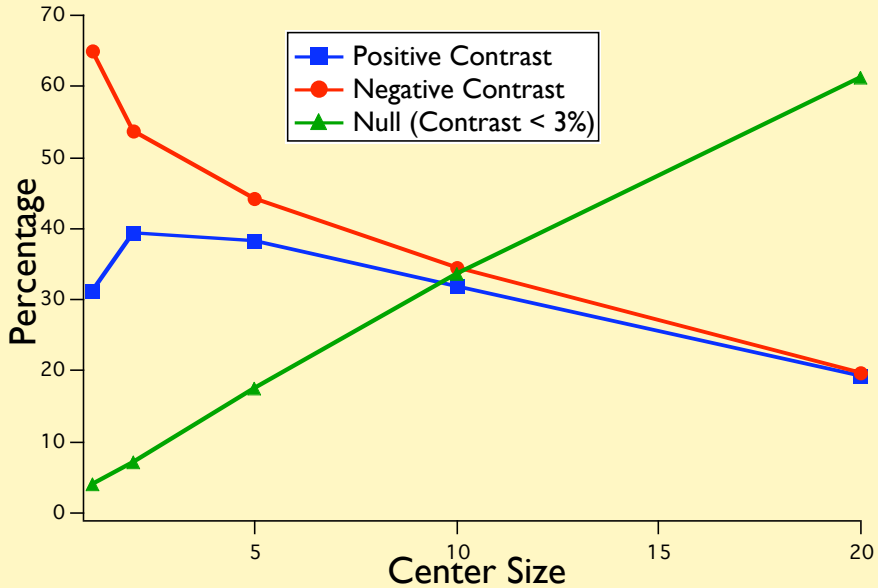
Can the distribution of natural intensities, along with power law correlations explain the distribution of positive/negative contrasts in natural scenes?

Contrast Proportions in Artificial Images

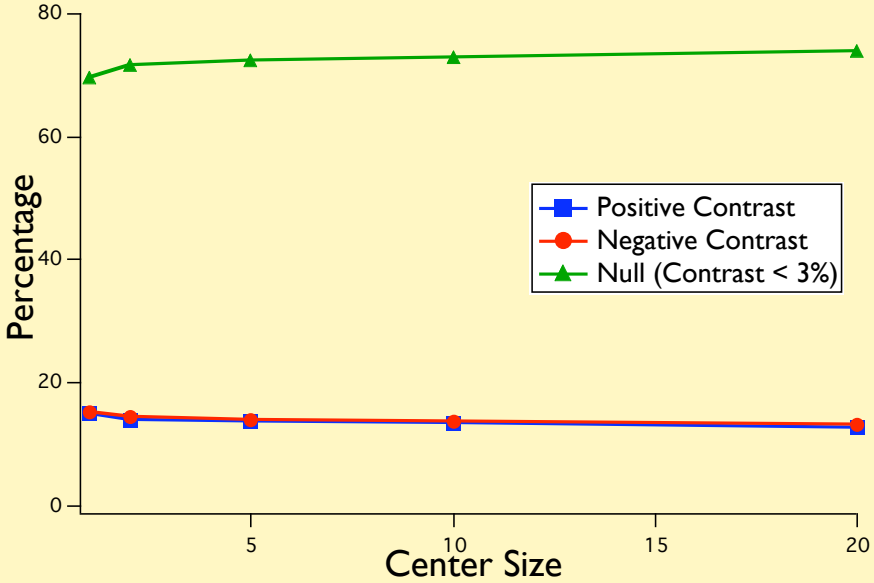
Gaussian Noise



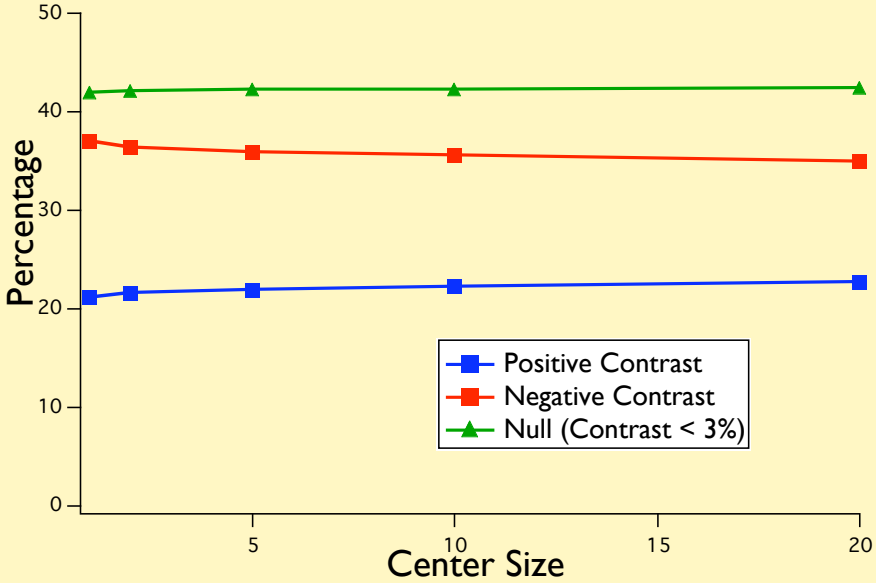
Natural Noise



"Pink" Gaussian Noise

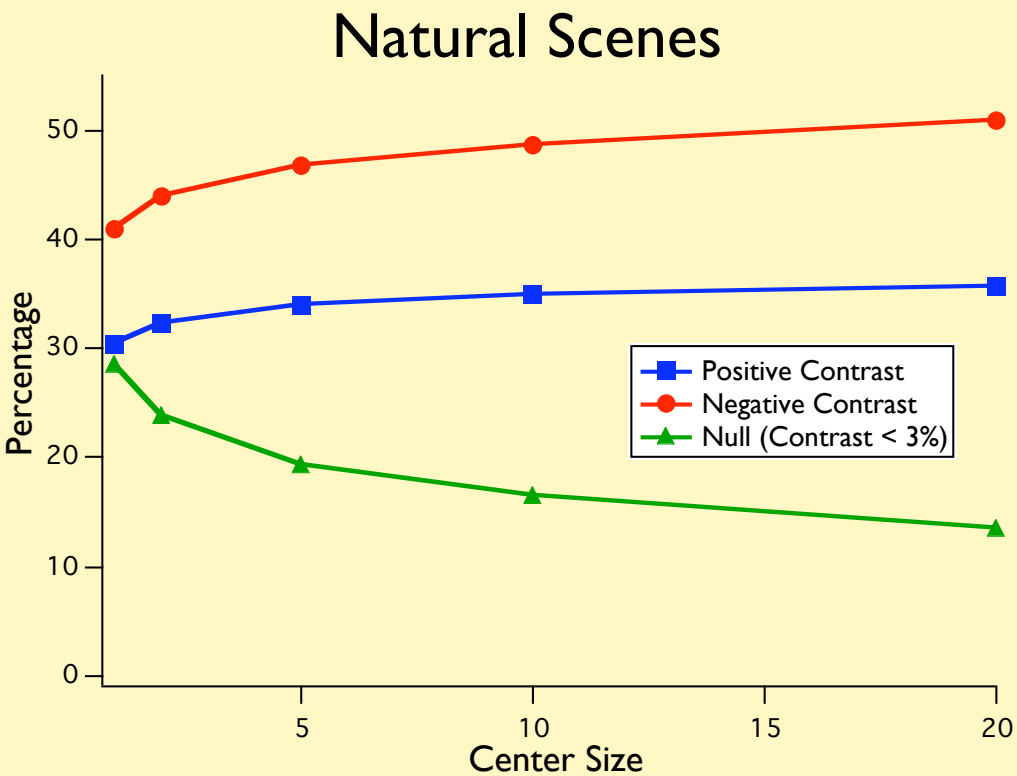


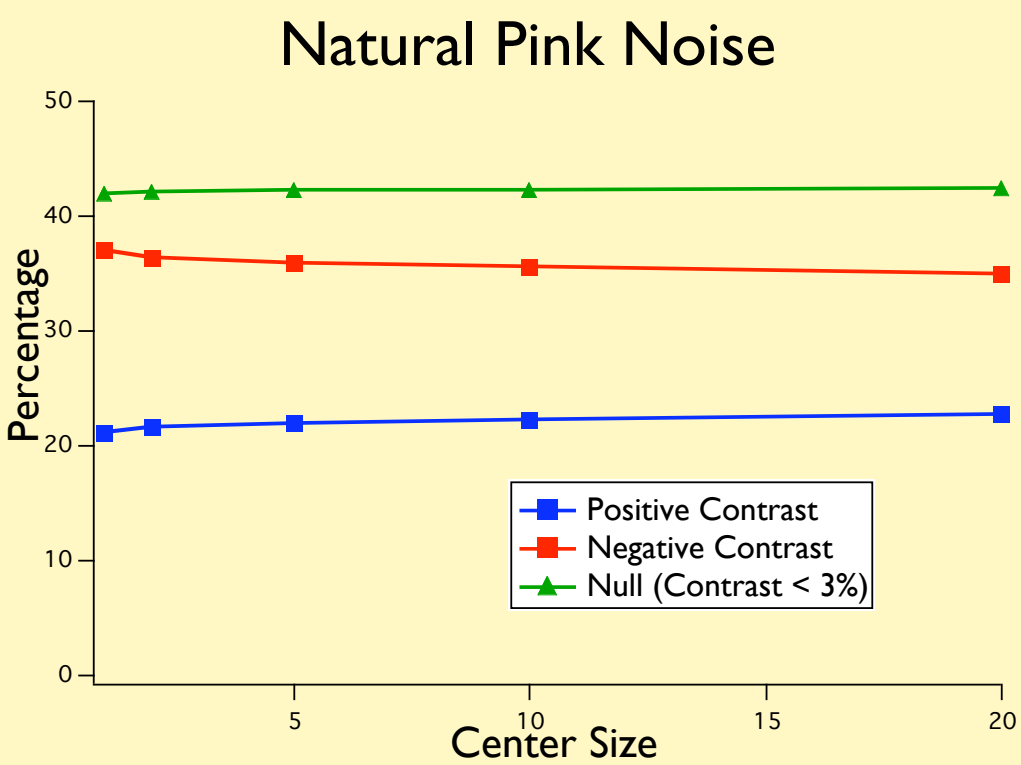
Natural "Pink" Noise



Thanks to Ben Backus for the pink noise images.

Natural Scenes vs. Natural Pink Noise





The natural intensity distribution and scale invariance do not suffice to fully explain the natural scene contrast measurements.

Some thoughts

- Phase information in the auto-correlation function is important

- Higher order statistics are important

- Do texture synthesis models lead to images with the bright/dark structure of natural scenes? (e.g, de Bonet and Viola (1997), Portilla and Simoncelli (1998).)

- The contrast measurements could be performed on individual scenes which have been remapped to preserve the individual intensity distribution and power law correlations and only then averaged over the ensemble. (Variations from the mean intensity distribution are important.)

Back to the retina...

- To determine how much resource should be committed to bright vs. dark processing we really want to know how *useful* the information in bright and dark patches is to behaviour.

- Unfortunately such utility functions are difficult to determine.

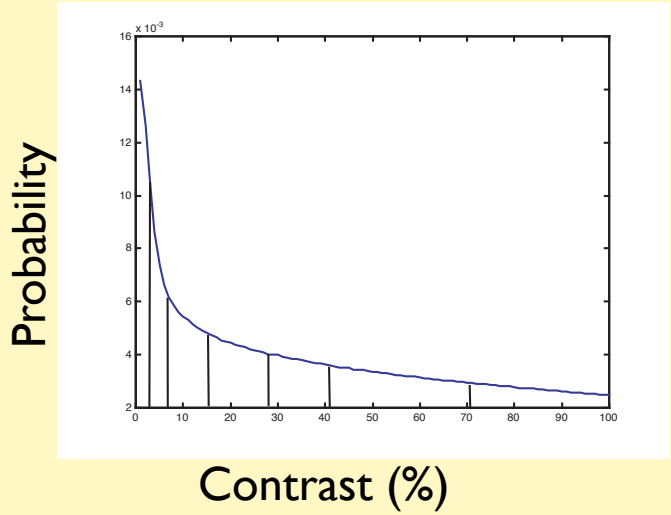
- However, we should expect that utility increases monotonically with the *amount* of information (number of bits).

Thus we explore how much information there is bright vs. dark components of natural scenes.

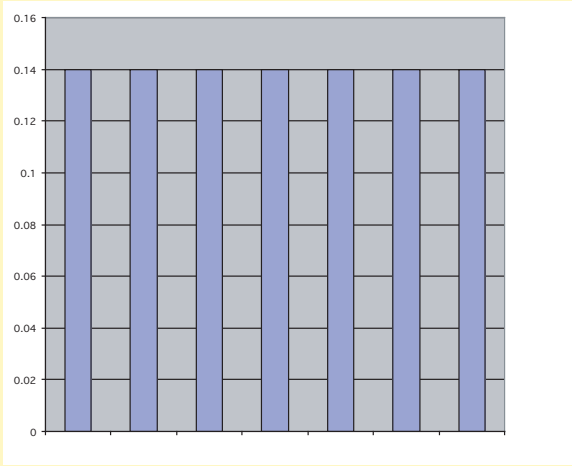
Mapping contrasts into discrete levels

Contrast is a continuous variable, but real cells have finitely many discriminable contrast levels. Maximum information is obtained by "histogram equalization". Laughlin (1976) showed that fly LMC performs such a procedure.

Positive contrast distribution

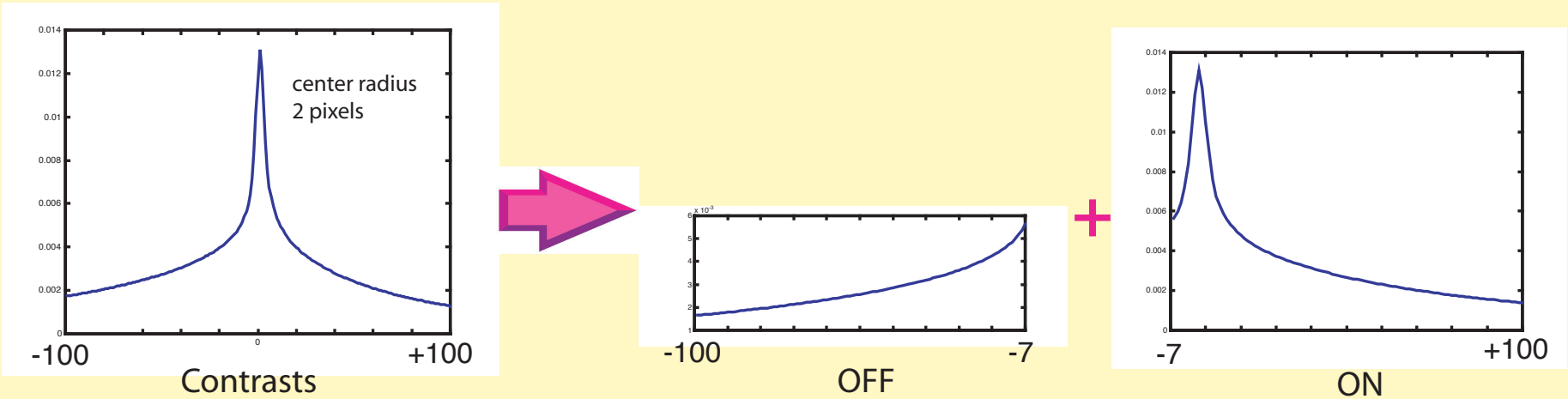


Equally probably contrast levels



- Measured in this way, entropy in the joint distribution of contrast levels of multiple filters depends on the number of discriminable levels.
- Thus we will be interested in trends as a function of the geometry of the contrast filter mosaic, not the absolute information measures.
- We could use "binless" estimates of entropy. (Bialek, Nemenman, Tishby; Victor)

Relevance to asymmetric ON/OFF rectification?



- A system transmitting the natural image contrast distribution using histogram equalization (or some efficient coding variant) must devote more contrast levels to encode negative contrasts.

- Thus, if an equal number of contrast levels is efficiently encoded by both ON and OFF channels, OFF channels should be more rectified, as they are.

- Are spatial contrasts efficiently encoded in retina? How many levels can ON and OFF cells distinguish? Effects of contrast adaptation?

Information transmitted by contrast filter arrays

Because of statistical dependencies in natural scenes, we must examine mutual information between arrays of nearby filters and the scene. Assuming independence will overestimate information.

The entropy of the joint distribution of a collection of contrasts $p(C_1, \cdots C_N)$ is

$$H(\{C\}) = -\sum_{\{C\}} p(C_1, \cdots C_N) \log_2 p(C_1, \cdots C_N)$$

The mutual information carried by $\{C_1, \cdots C_N\}$ about the the pixel intensities in the scene $\{i_1, \cdots i_k\}$ is

$$I(\{i\};\{C\}) = H(\{C\}) - H(\{C\}|\{i\}) = H(\{C\}) - H_{\text{noise}}$$

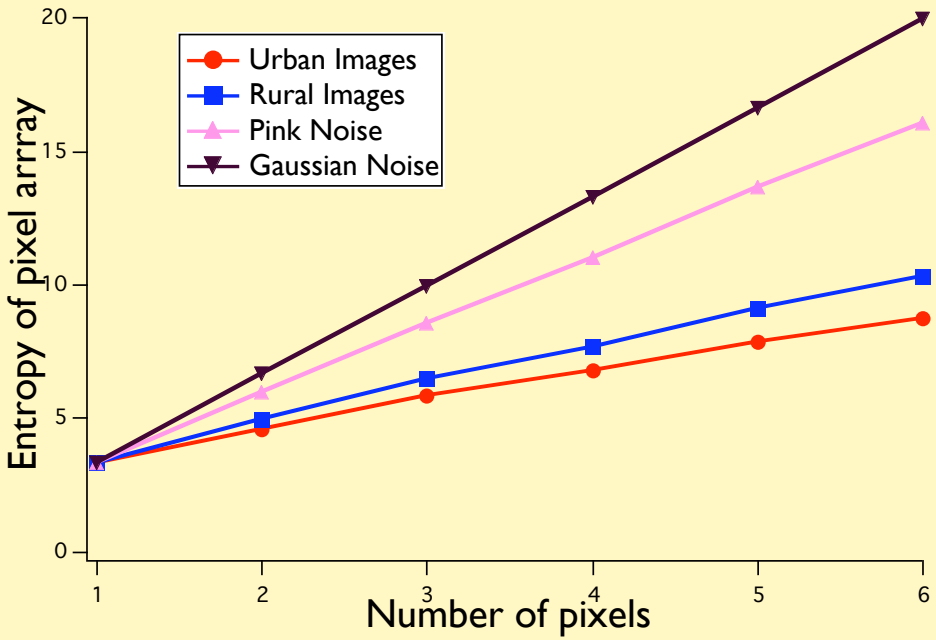
There is no noise in our setting (so far). Thus,

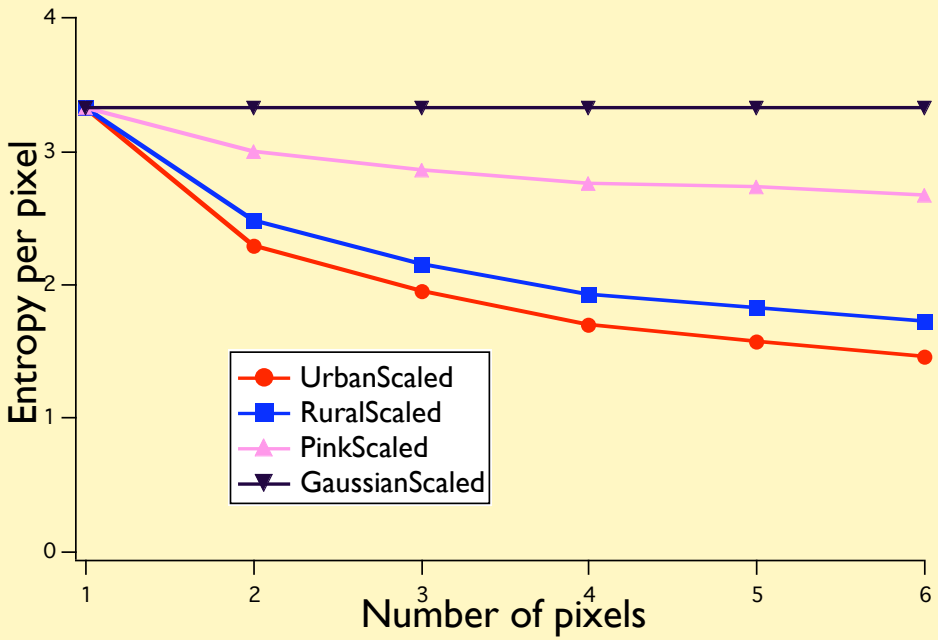
$$I(\{i\};\{C\}) = H(\{C\})$$

❂ Serious data limitations due to combinatorial explosion of the number of possible responses of N filters. (For 10 contrast levels this is 10^N.)

❂ Could use more efficient techniques to estimate mutual information or at least set a lower bound (e.g. talks by Tishby and Bialek in this workshop).

Information in intensity arrays



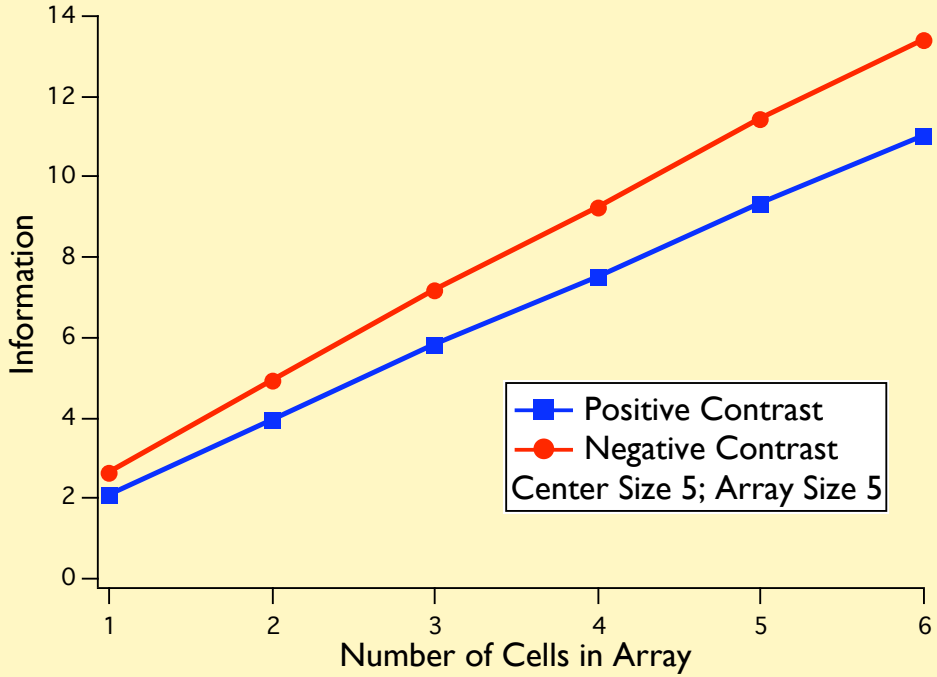


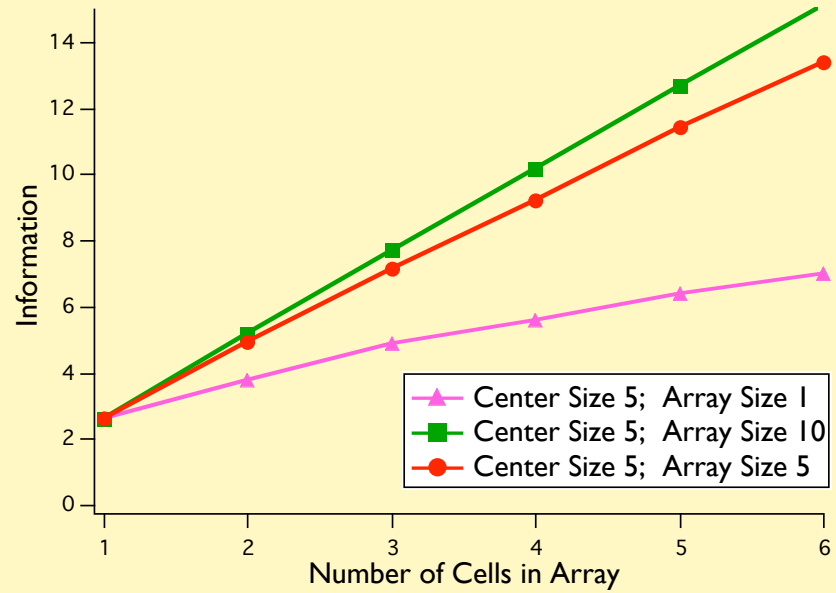
- Used 10 intensity levels and histogram equalization. Gaussian and natural noise will give the same results.

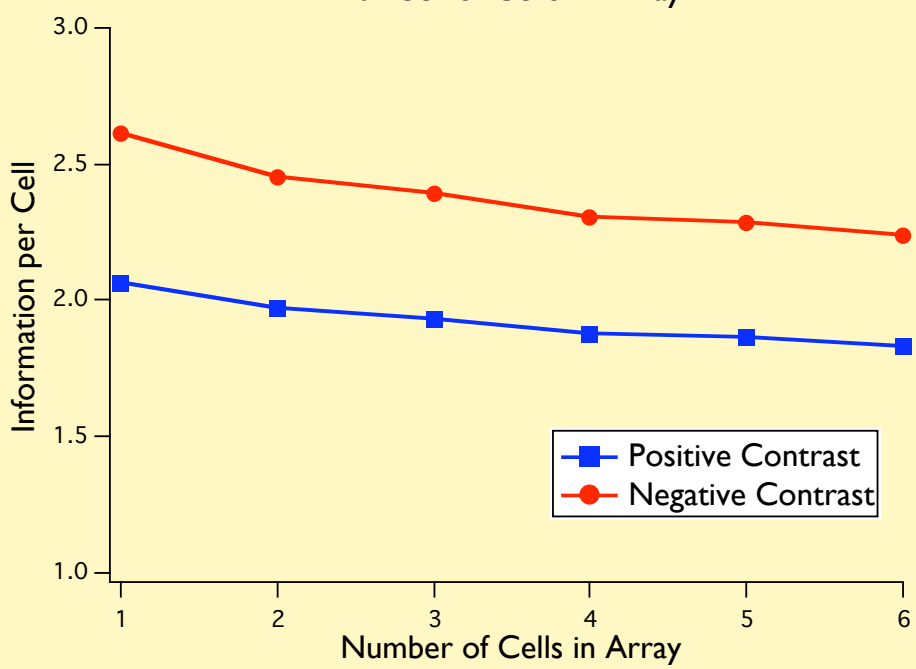
- Natural scenes carry much less information than pink noise. Hence additional correlations beyond the power law are important.

- Cities are less informative than villages. (Who knew?) Are there important environmental differences in information content?

Information in contrast arrays



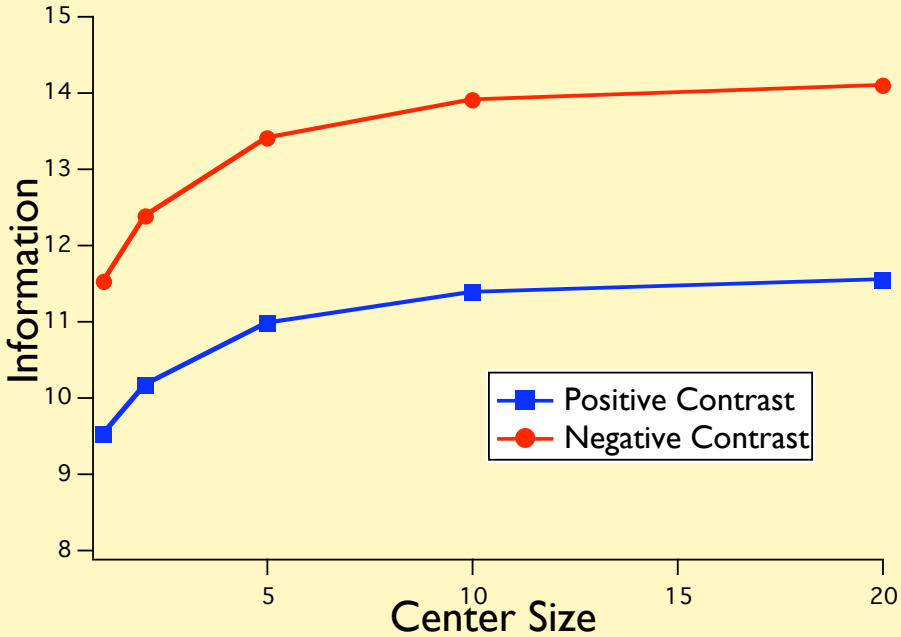




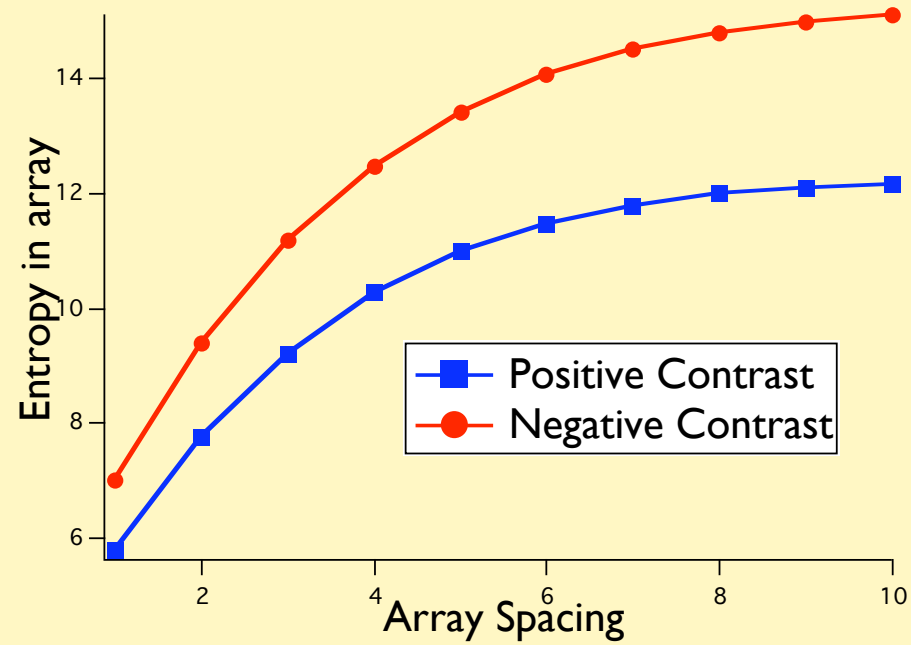
- Redundancy increases with overlap and with number of cells.

- Arrays of 6 cells capture most of the redundancy. Contrast filters have (substantially) decorrelated the intensity stimulus.

More Dark Information in Natural Scenes



Information from 6 filter arrays with center size = array size



Information from 6 filter arrays with center size = 5 pixels

- Contrast discretized to 10 levels. Increased array spacing decreases redundancy.

- At all scale sizes and array spacings there is more information in the negative contrast components of a natural scene.
- Since retinal ON channels are larger and have less receptive field overlap than OFF channels, their information transmission is more equalized than with symmetric channels.

Is there a principle of "information equalization" here?

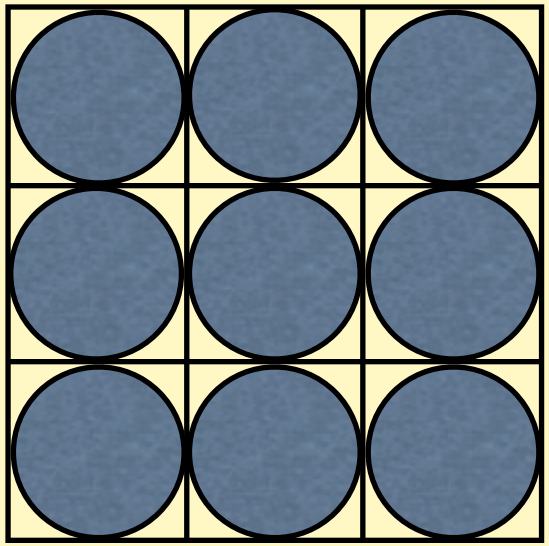
Anatomical evidence?: Although OFF centers are smaller, the dendritic arbors of OFF cells are denser and have roughly the same wire length as ON cells. They also make roughly the same number of synapses with bipolar cells (Kao and Sterling, unpublished).

Is there a tradeoff between between (asymmetric) information transmission in bright and dark channels and resource consumption that leads to information equalization, or some other design principle that can predict the detailed structure and function (size, overlap, rectification etc.) of ON vs. OFF channels?

A first stab at an optimization principle

Suppose resource constraints limit us to have N cells in total. Equivalently assume that ON and OFF cells have an equal cost to construct and maintain and that there is a constraint on the maximum cost.

Let N_{ON} be the number of ON cells ($N_{OFF} = (N - N_{ON})$) and assume that both ON and OFF cell mosaics tile the retina with Center Size = Array Spacing. (Overlapping surrounds.) How should N_{ON} be chosen?



Let i_{ON} be the information per cell in the ON array as a function of the number of ON cells (N_{ON}) under these tiling conditions. (Similarly for the OFF array.)

The total information transmitted is:

$$I = N_{ON}\, i_{ON} + (N - N_{ON})\, i_{OFF}$$

Maximize Information Transfer

Maximize $I = N_{ON}\, i_{ON} + (N - N_{ON})\, i_{OFF}$ as a function of N_{ON}

$$\frac{\partial I}{\partial N_{\mathrm{ON}}} = 0 \qquad \Longrightarrow \qquad (i_{\mathrm{ON}} - i_{\mathrm{OFF}}) = N_{\mathrm{OFF}} \frac{\partial i_{\mathrm{OFF}}}{\partial N_{\mathrm{OFF}}} - N_{\mathrm{ON}} \frac{\partial i_{\mathrm{ON}}}{\partial N_{\mathrm{ON}}}$$

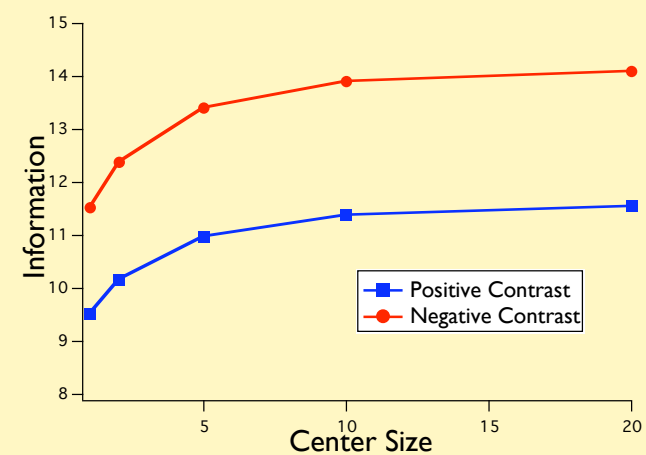
- When the number of cells is very large, as it is in retina, information per cell will be nearly independent of the the number of cells since increasing the number of cells by 1 will result in a small change in array geometry.

- Thus the right hand side of the optimization condition nearly vanishes, giving $i_{ON} - i_{OFF} = 0$

- Optimization equalized the information per cell in the ON and OFF arrays

- This implies under these tiling conditions, and with the assumption of equal cost, there should be fewer ON cells and they will be larger

Information in 6 filter center size = array spacing arrays



Plan

(1) Identify which properties of natural scenes fully account for the bright/dark asymmetries. E.g., consider texture models.

(2) Environmental dependence of our results?

(3) Other contrast models.

(4) Tradeoffs between space usage, energy usage and information transmission in ON/OFF channels?

(5) Effects of noise. Information content of small center filters will be greatly reduced. Essential to compare with data -- e.g, energy efficient codes.

(6) How do the statistics of eye movements, e.g, saccades, affect the design of ON/OFF channels? Contrast adaptation?

(7) Retinal ganglion cell types have individual functions -- e.g., local edge detection, motion etc. How does that affect the ON/OFF dichotomy.

Broader Picture

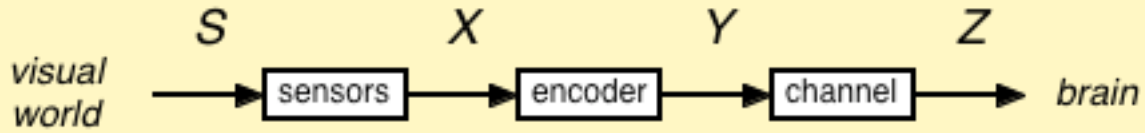
- The design of the brain is immensely heterogeneous and diverse. For example, the retina expresses a very large number of types of cell wired in (more or less) stereotyped ways.

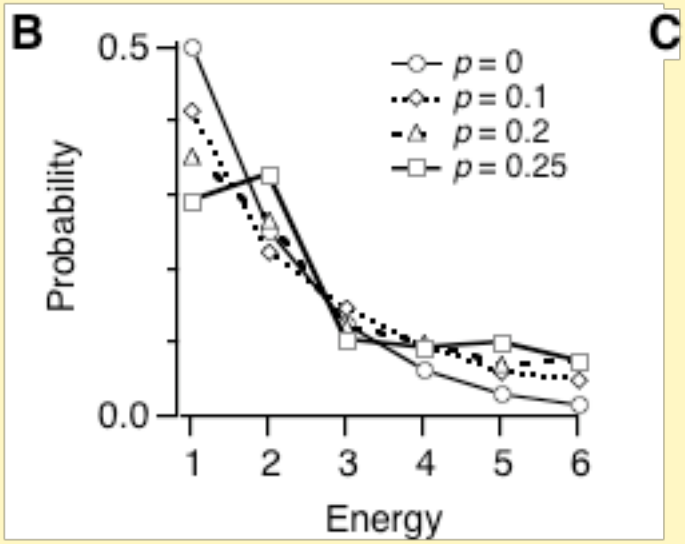
- Why this kind of design? Why so may cell types? An arbitrary computer can be built from NAND gates. Why are the wiring diagrams the way they are?

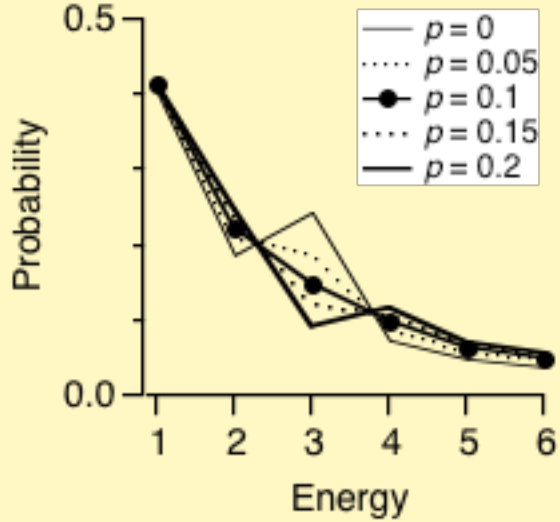
- We are going after the hypothesis that the heterogeneous design exists to efficiently extract and process useful information from natural stimuli, in the presence of spatial, energetic, temporal and evolutionary constraints.

- Conversely, this hypothesis can function as a predictive design principle.

- Our work is intended as a foothold on this cliff face, studying how the asymmetric segregation of bright/dark processing can be understood in these terms.







$$C(E) = \max_{q(y);\bar{E}\leq E}\left[-\sum_j q_j \ln q_j + \sum_{jk} q_j Q_{k|j} \log P_{j|k}\right]$$

$$Q = \begin{pmatrix} 1-2p & 2p & 0 & 0 & 0 & 0 \\ p & 1-2p & p & 0 & 0 & 0 \\ 0 & p & 1-2p & p & 0 & 0 \\ 0 & 0 & p & 1-2p & p & 0 \\ 0 & 0 & 0 & p & 1-2p & p \\ 0 & 0 & 0 & 0 & 2p & 1-2p \end{pmatrix}.$$

The effect of noise on energy efficient codes.

The distribution of symbols in the energy efficient code is strongly affected by the noise in the system.

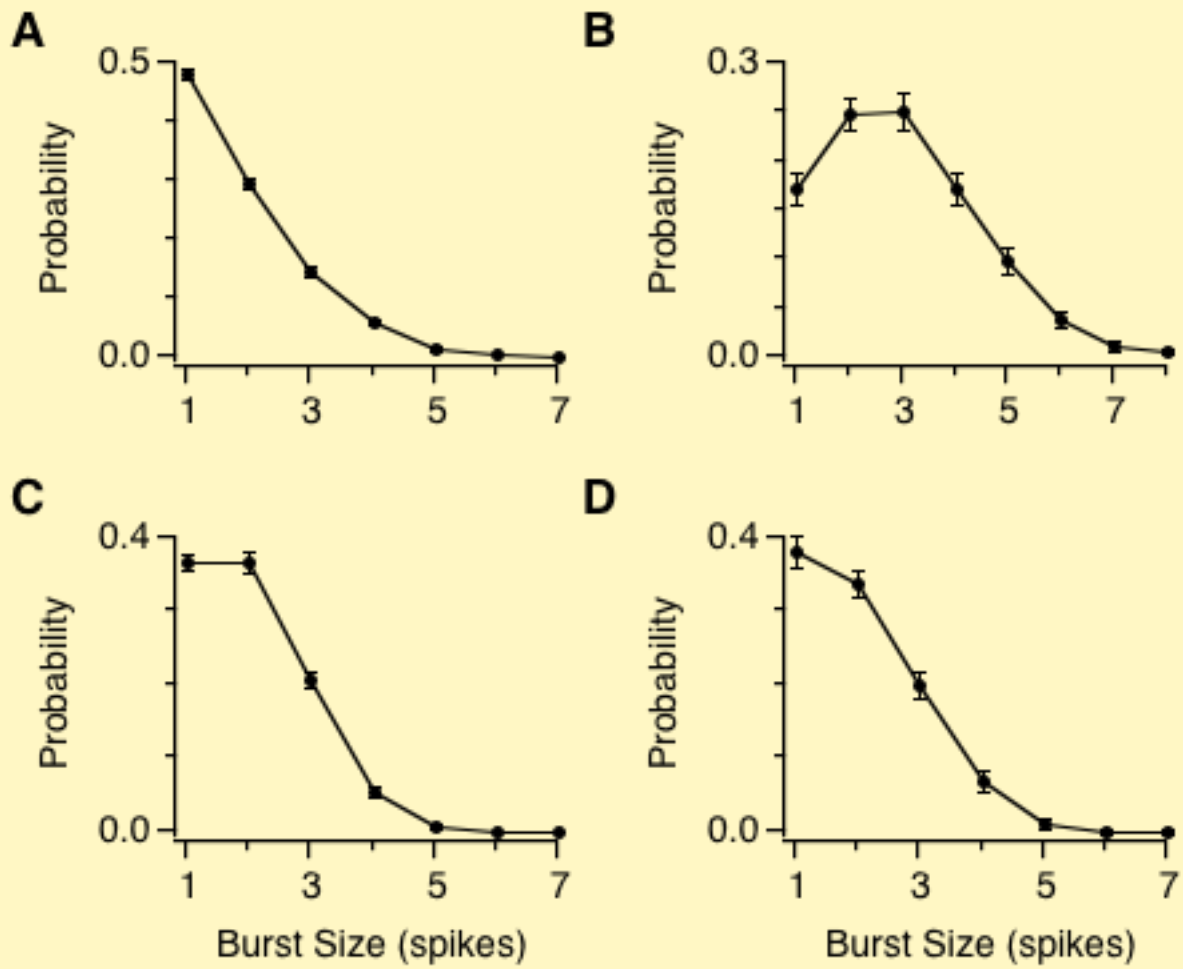
Thus noise must be carefully measured and accounted for in biological applications

Balasubramanian, Kimber & Berry, Neural Computation 13, 2001.

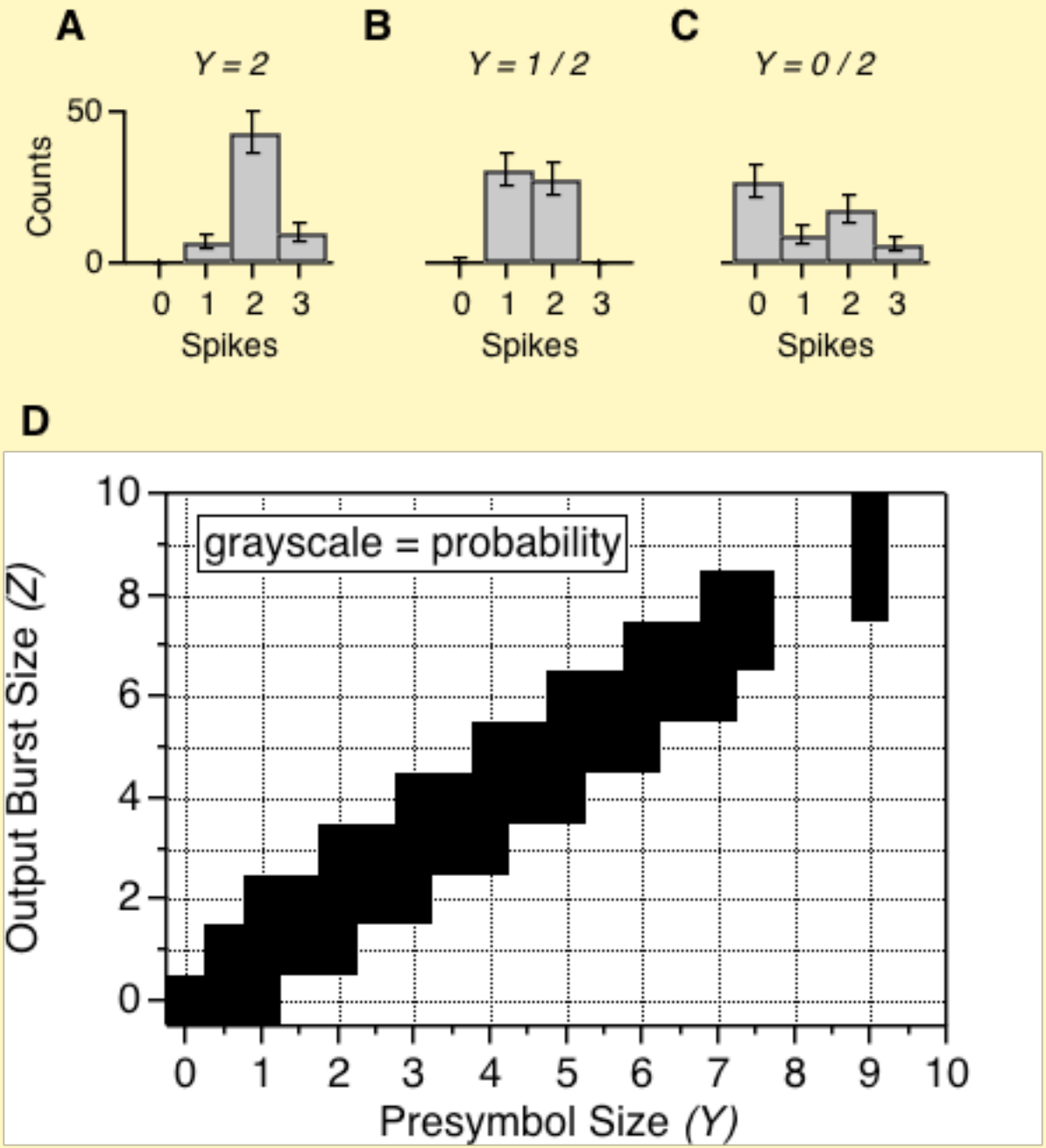
Balasubramanian & Berry, Network, 13, 2002.

de Polavieja, J. Theor. Biol 13, 2002.

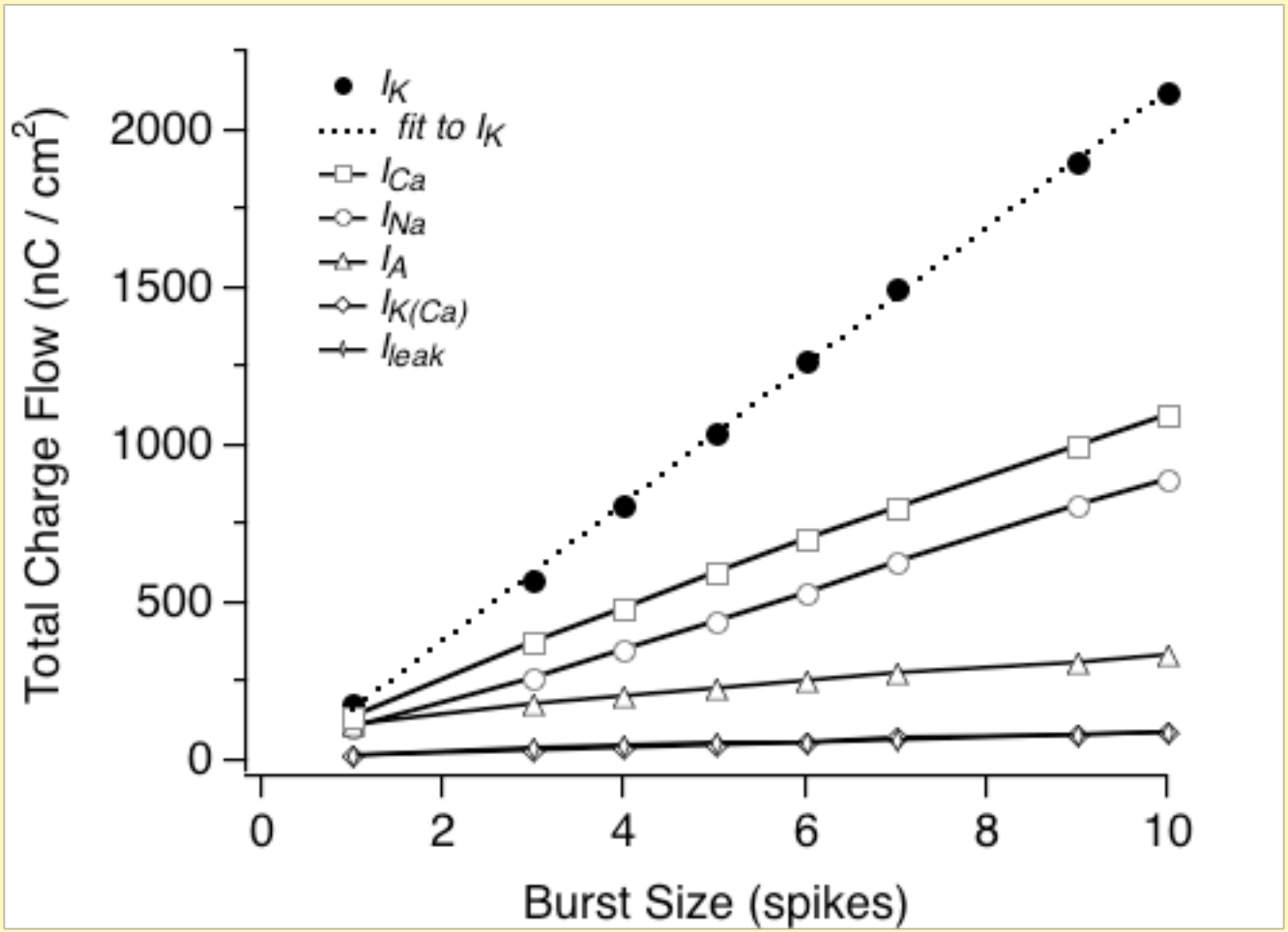
Burst Size Distributions in Salamander Retina



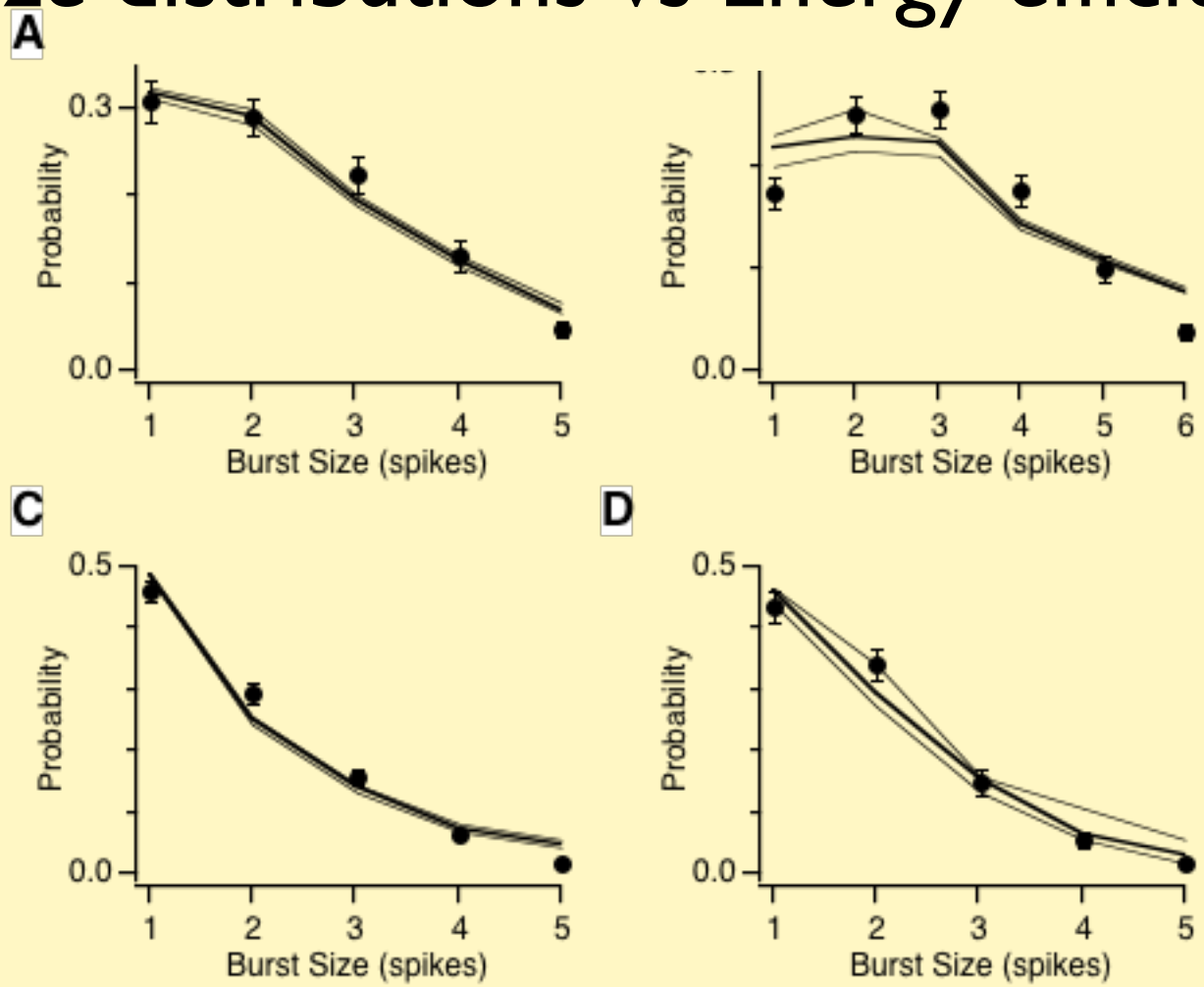
Empirically measured noise in burst sizes



The cost of burst is linear in the number of spikes



Burst size distributions vs Energy efficient codes



Excellent fits for all cells. Qualitative structure at large burst size is due to suppression of energetically expensive bursts. Qualitative structure at small burst size is dues to suppression of noisy bursts in an efficient code. (Also see related work by de Polavieja.)