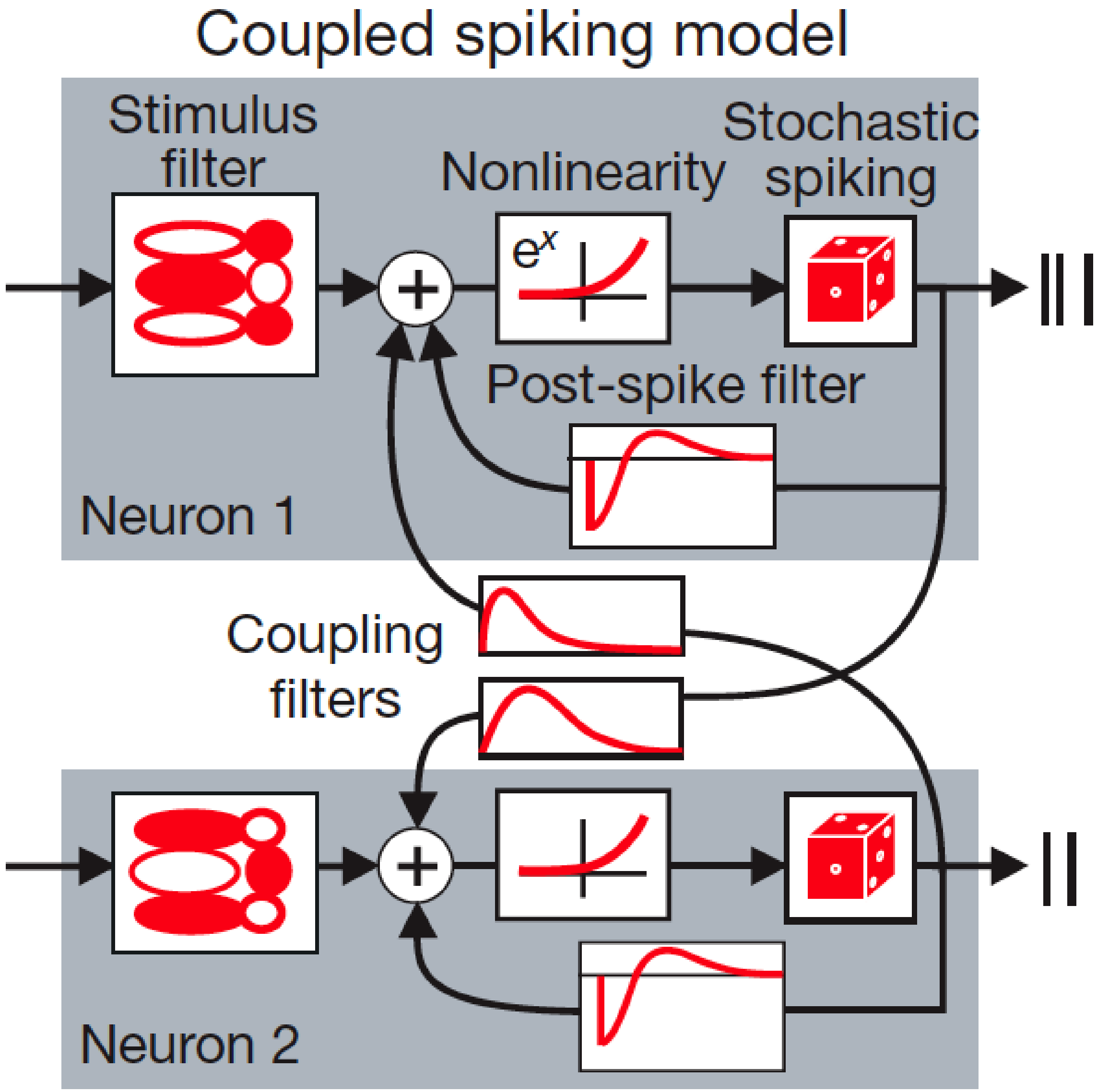
Encoding or decoding



Decoding

How well can we learn what the stimulus is by looking at the neural responses?

We will discuss two approaches:

- devise and evaluate explicit algorithms for extracting a stimulus estimate

- directly quantify the relationship between stimulus and response using information theory

The optimal linear estimator

Let's start with a rate response, r(t) and a stimulus, s(t).

The optimal linear estimator is closest to satisfying

$$r_{\text{est}}(t) = \bar{r} + \int d\tau\, s(t-\tau)K(\tau)$$

Want to solve for K. Multiply by s(t-**τ**') and integrate over t:

$$\int dt\, s(t-\tau')r(t) = \int dt \int d\tau s(t-\tau')s(t-\tau)K(\tau)$$

The optimal linear estimator

$$\int dt\, s(t-\tau')r(t) = \int dt \int d\tau s(t-\tau')s(t-\tau)K(\tau)$$

→ produced terms which are simply correlation functions:

$$C_{rs}(-\tau') = \int d\tau\, C_{ss}(\tau'-\tau)K(\tau)$$

Given a convolution, Fourier transform:

$$\int d\tau'\, e^{i\omega\tau'} C_{rs}(-\tau') = \int d\tau' e^{i\omega\tau'} \int d\tau\, C_{ss}(\tau'-\tau)K(\tau)$$

Now we have a straightforward algebraic equation for K(ω):

$$\tilde{C}_{rs}(-\omega) = \tilde{C}_{ss}(\omega)K(\omega)$$

Solving for K(t),

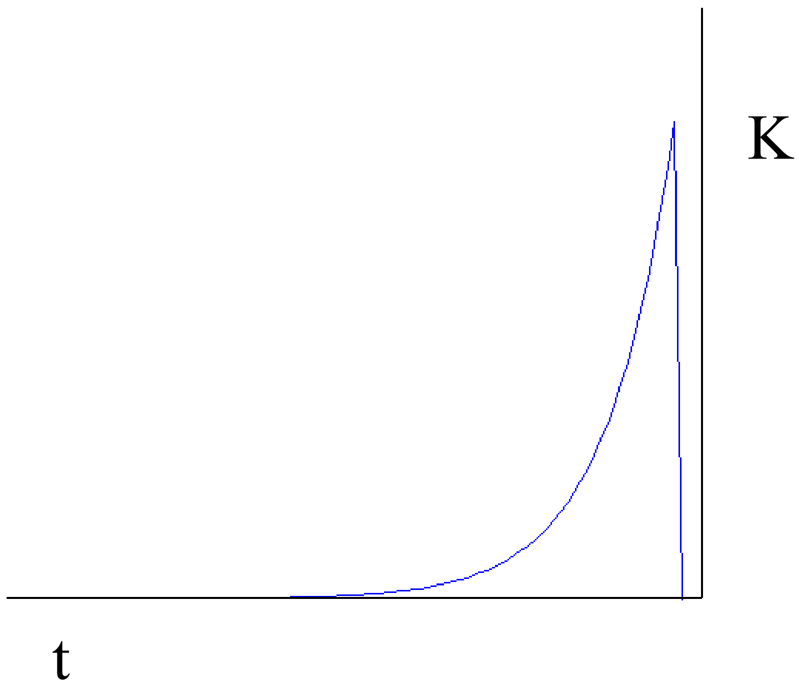
$$K(t) = \frac{1}{2\pi}\int d\omega\, e^{-i\omega t}\frac{\tilde{C}_{rs}(-\omega)}{\tilde{C}_{ss}(\omega)}$$

The optimal linear estimator

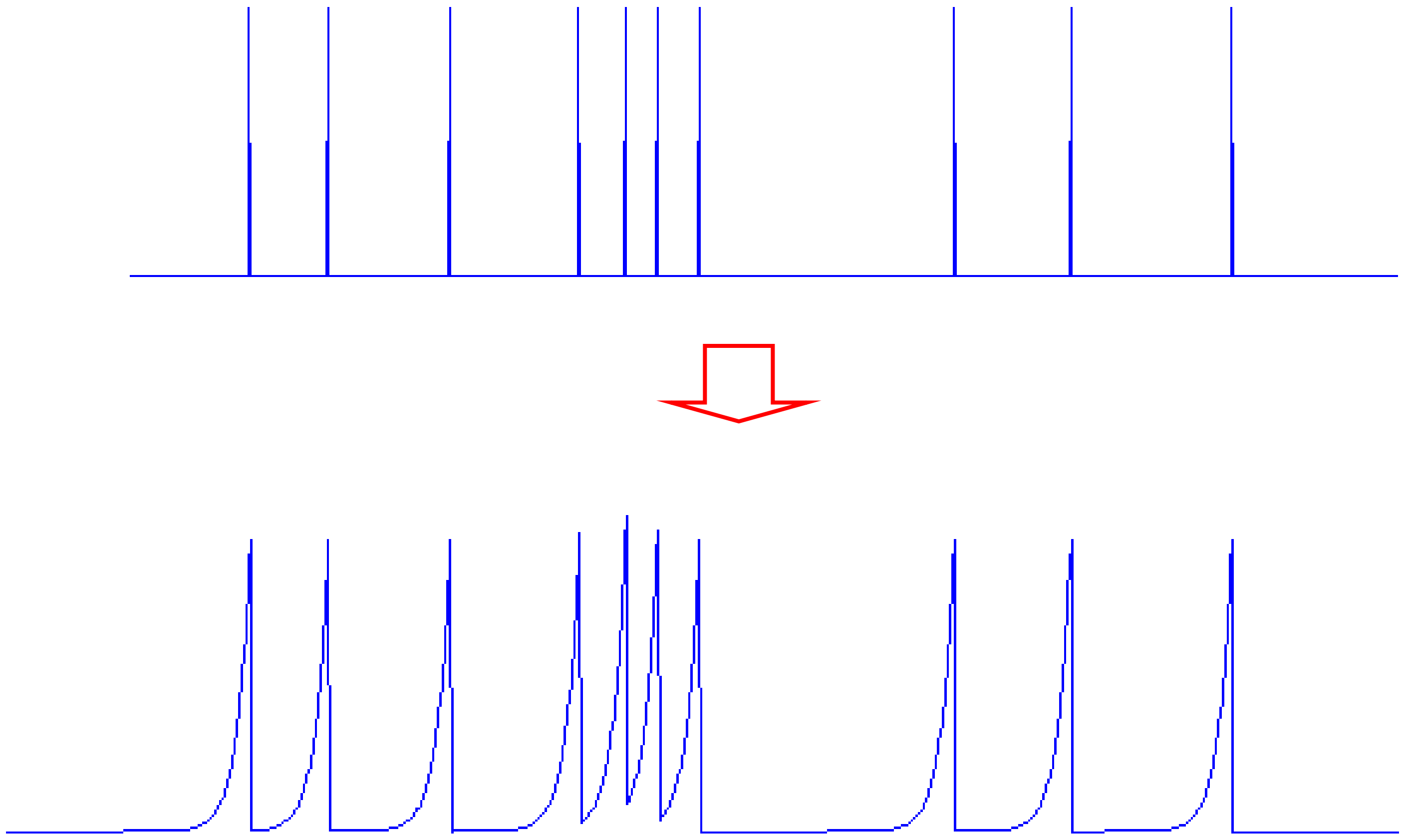
$$K(t) = \frac{1}{2\pi}\int d\omega\, e^{-i\omega t}\frac{\tilde{C}_{rs}(-\omega)}{\tilde{C}_{ss}(\omega)}$$

For white noise, the correlation function $C_{ss}(\tau) = \sigma^2\,\delta(\tau)$,
So $K(\tau)$ is simply $C_{rs}(\tau)$.

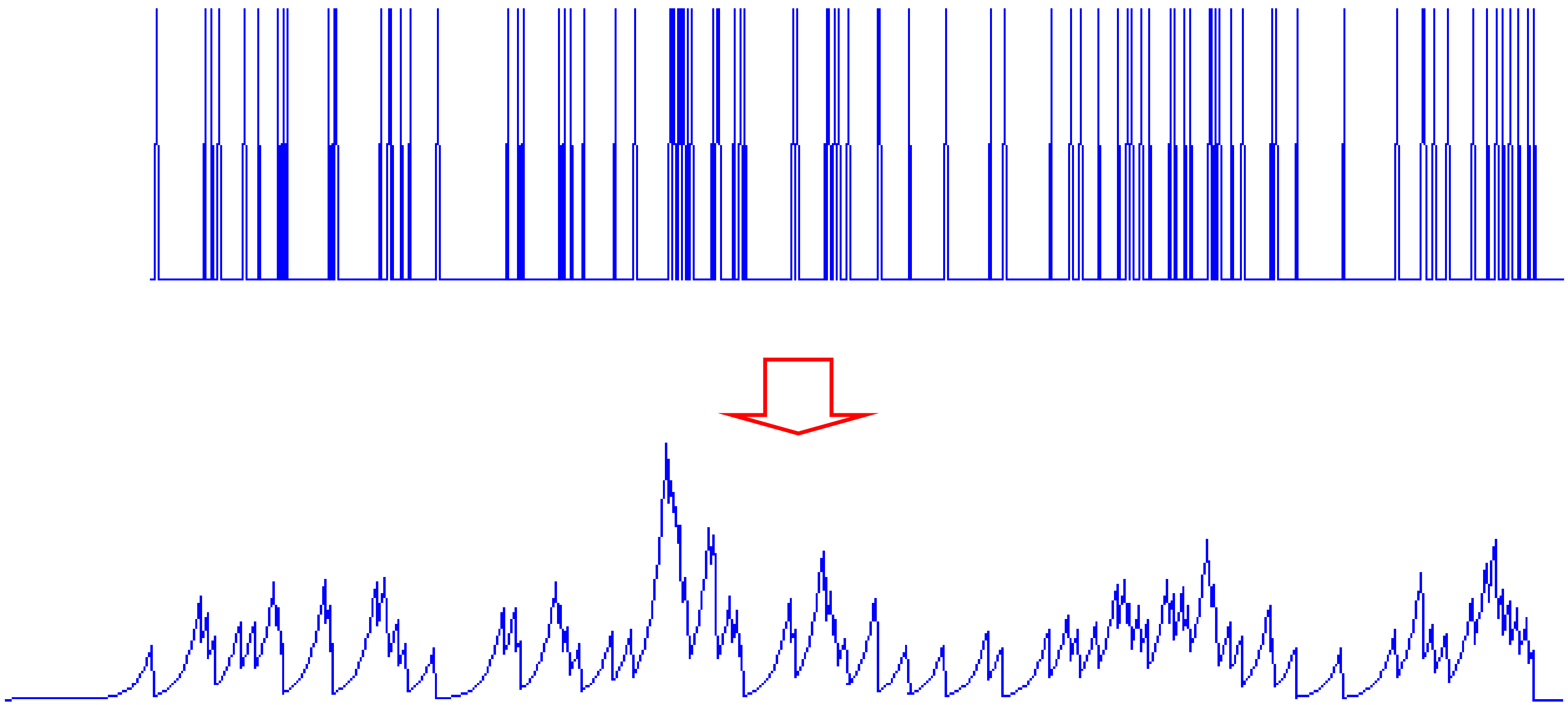
Stimulus reconstruction



Stimulus reconstruction



Stimulus reconstruction

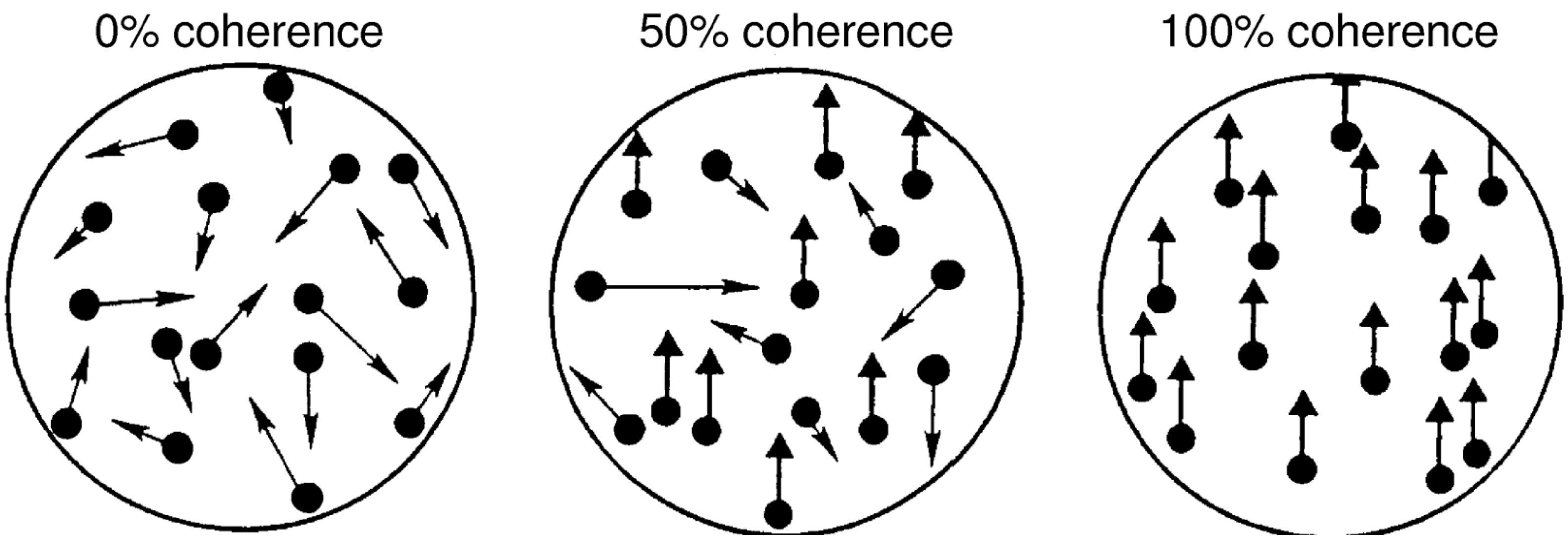


Reading minds: the LGN

Yang Dan, UC Berkeley

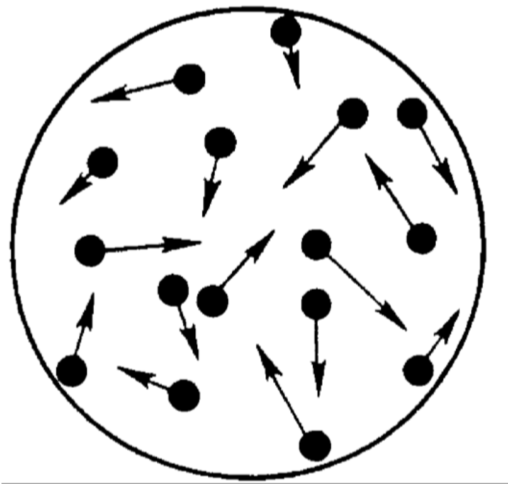
Other decoding approaches

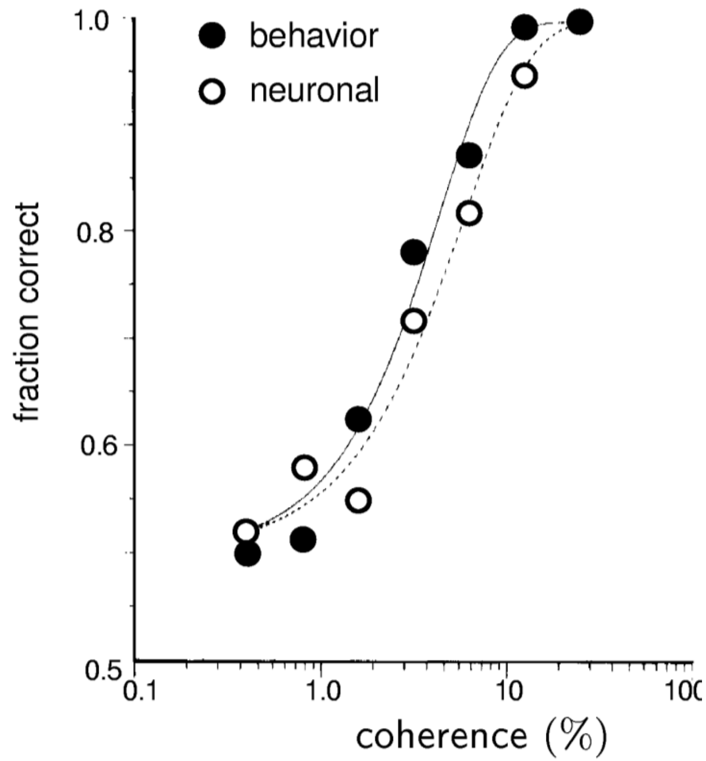
Binary choice tasks



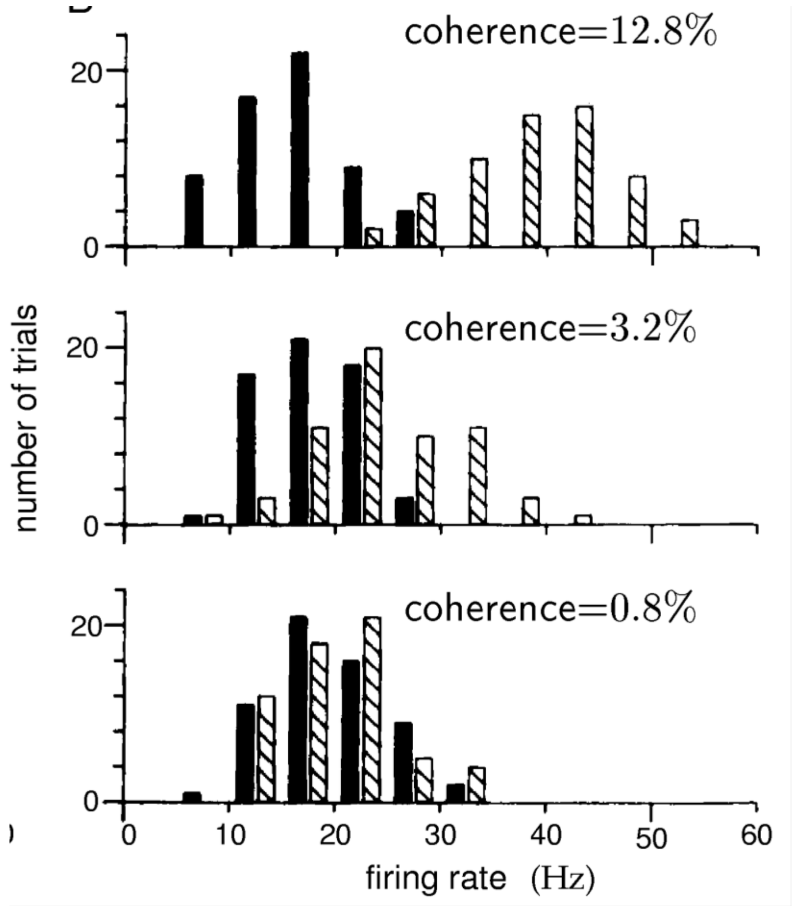
Britten et al. ‘92: measured both behavior + neural responses

Behavioral performance



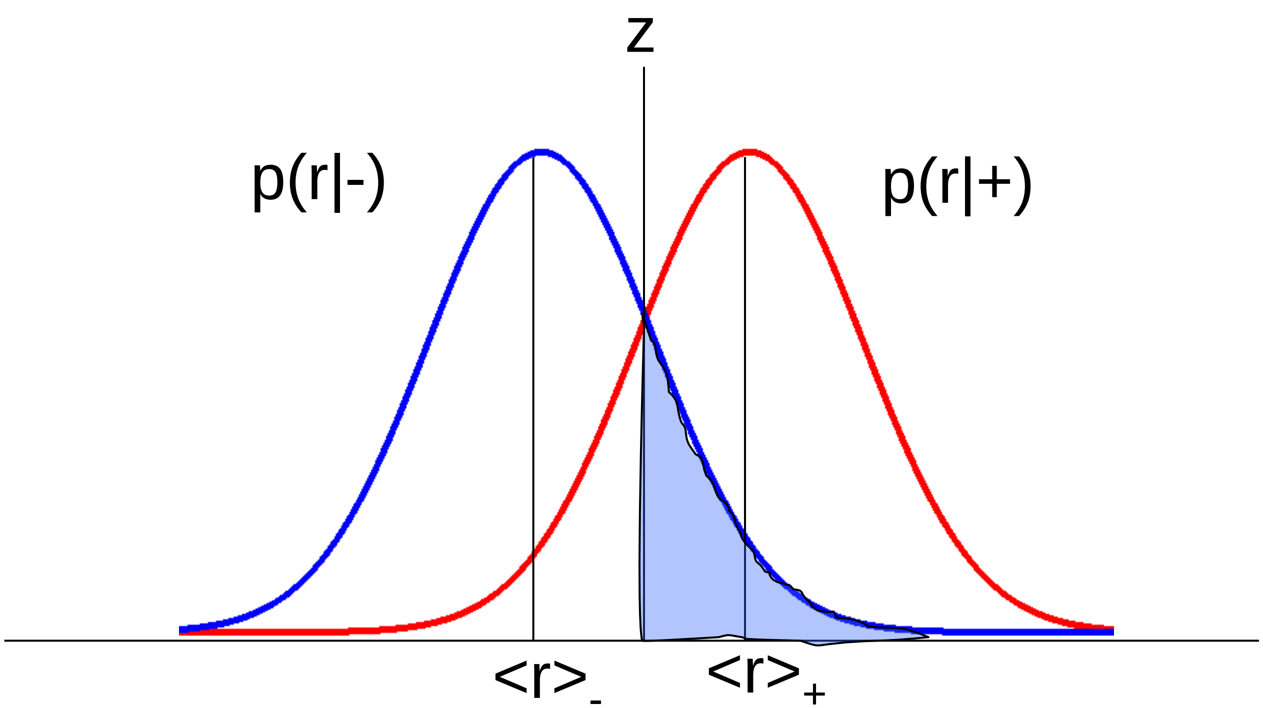


Predictable from neural activity?



Discriminability: $d' = (\langle r \rangle_+ - \langle r \rangle_-) / \sigma_r$

Signal detection theory



Decoding corresponds to comparing test, r, to threshold, z.

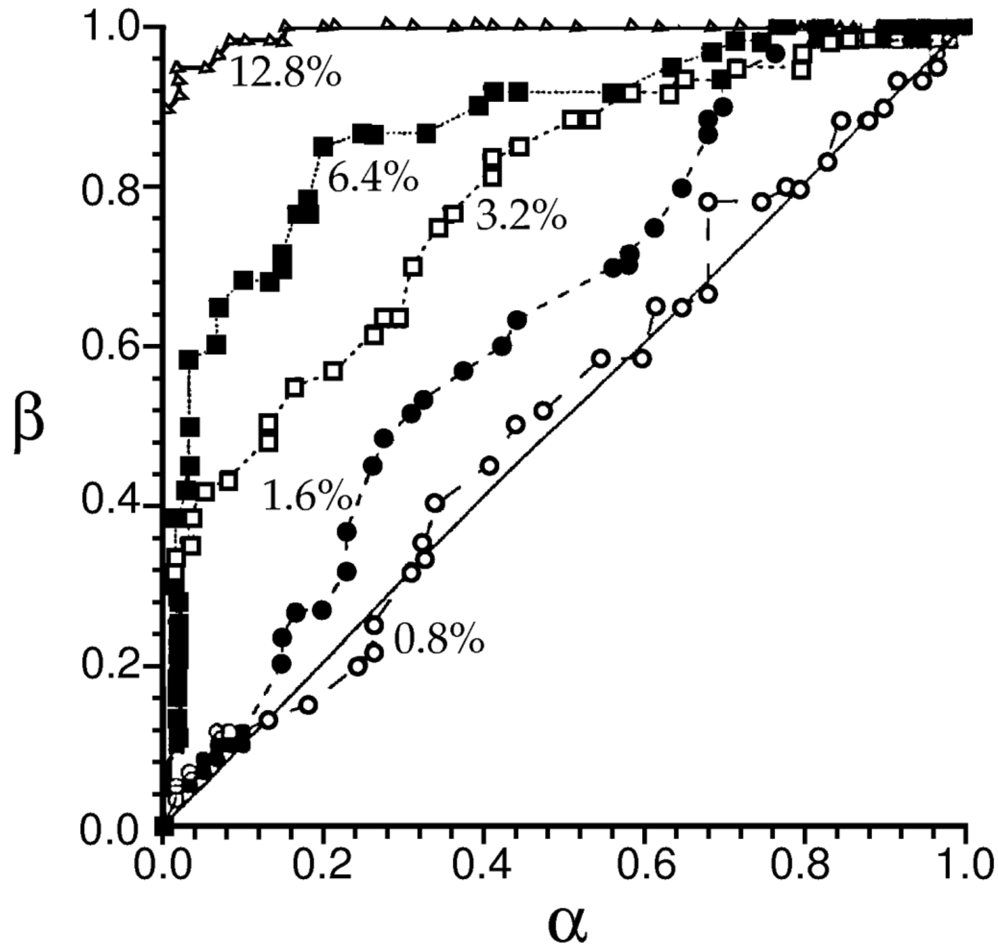
$\alpha(z) = P[\, r \geq z|-]$ false alarm rate, "size"

$\beta(z) = P[\, r \geq z|+]$ hit rate, "power"

Find z by maximizing $P[\text{correct}] = p[+]\,\beta(z) + p[-](1 - \alpha(z))$

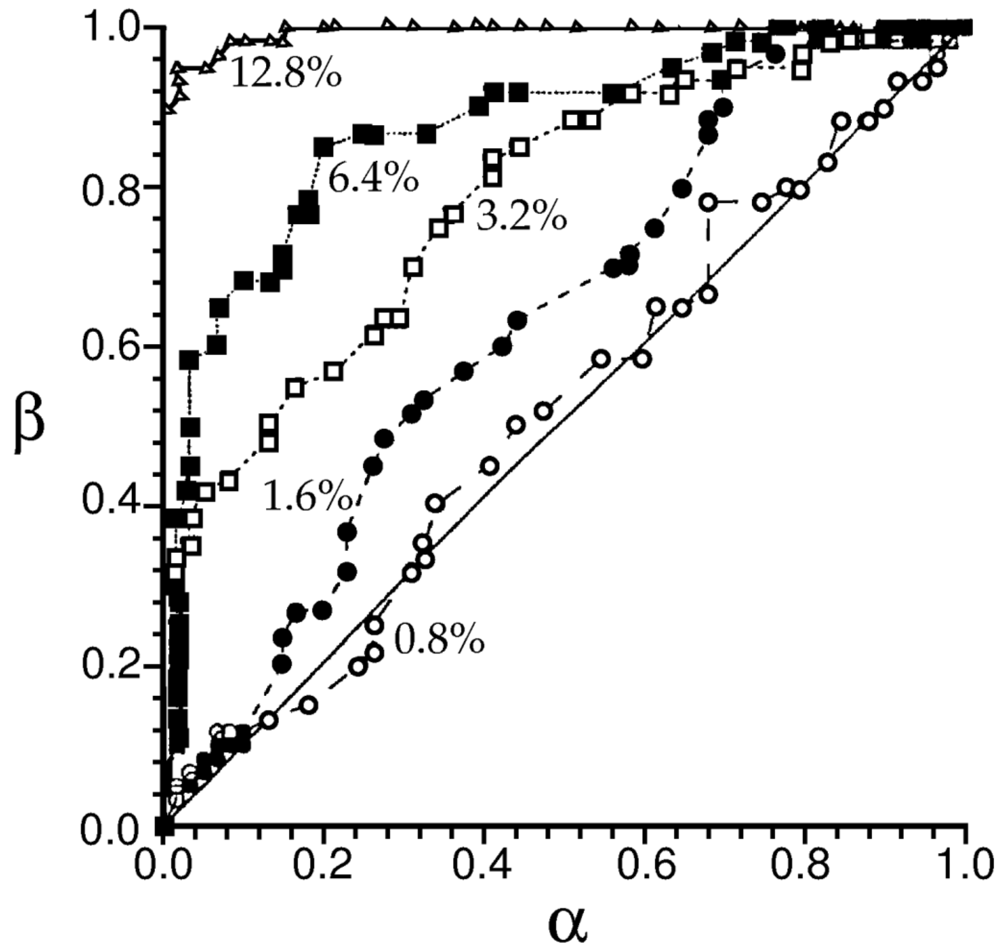
ROC curves

summarize performance of test for different thresholds z



Want $\beta \rightarrow 1$, $\alpha \rightarrow 0$.

ROC: two alternative forced choice



Threshold *z* is the result from the first presentation
The area under the ROC curve corresponds to P[correct]

Is there a better test to use than *r*?

The optimal test function is the *likelihood ratio*,

$$l(r) = p[r|+] \;/\; p[r|-].$$

(Neyman-Pearson lemma)

Recall $\alpha(z) = P[\, r \geq z|-]$ false alarm rate, "size"

$\beta(z) = P[\, r \geq z|+]$ hit rate, "power"

Then

$$l(z) = (d\beta/dz) \;/\; (d\alpha/dz) \;=\; d\beta/d\alpha$$

i.e. slope of ROC curve

The logistic function

If p[r|+] and p[r|-] are both Gaussian, one can show that

$$P[\text{correct}] = \tfrac{1}{2}\,\text{erfc}(-d'/2).$$

To interpret results as two-alternative forced choice, need simultaneous responses from
“+ neuron” and from “– neuron”.

Simulate “- neuron” responses from same neuron in response to – stimulus.

Ideal observer: performs as area under ROC curve.

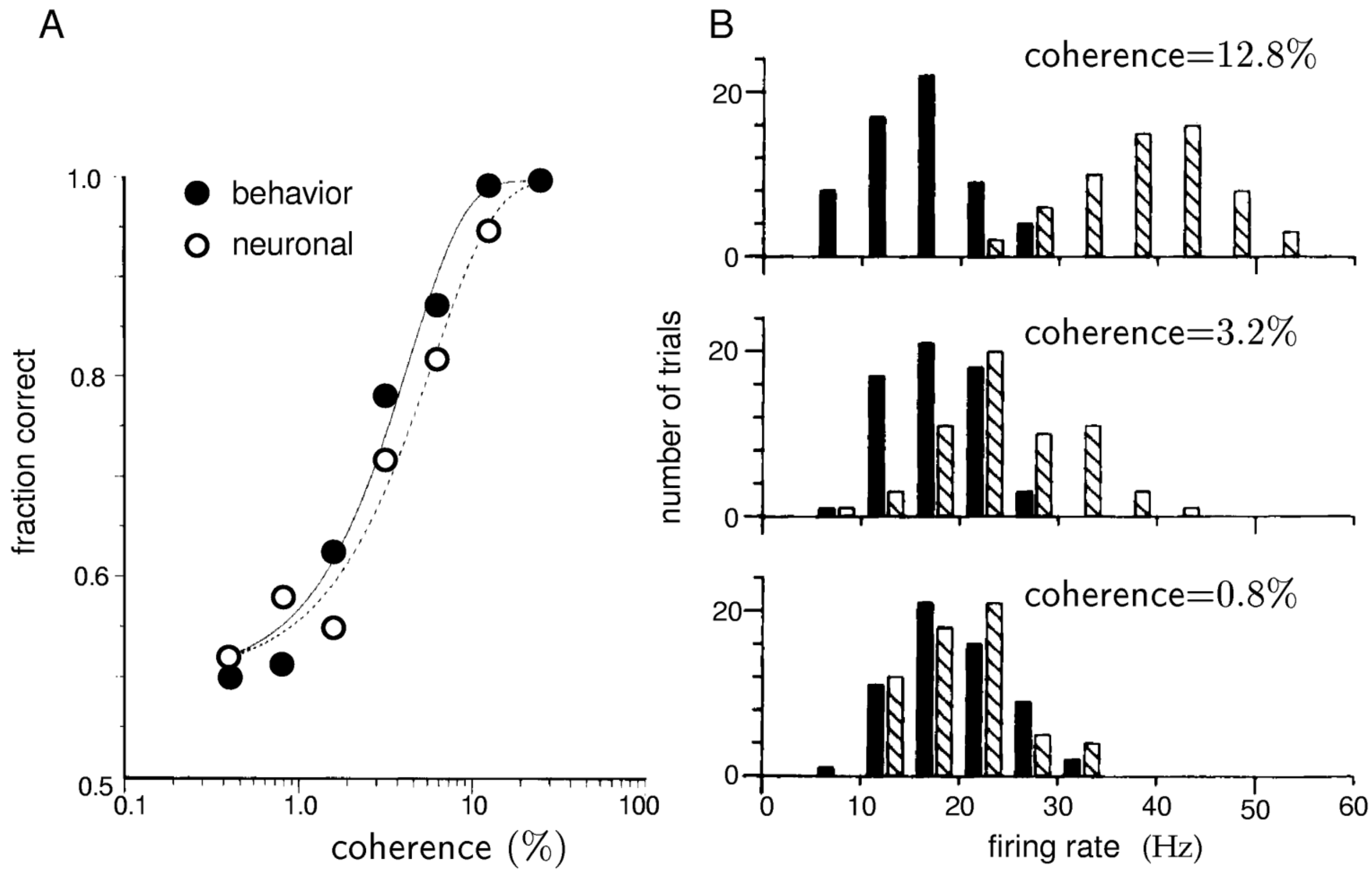
More d’

Again, if p[r|-] and p[r|+] are Gaussian,
and p[+] and p[-] are equal,

$$P[+|r] = 1/ [1 + \exp(-d' (r - \langle r \rangle)/ s)].$$

→ d’ is the slope of the sigmoidal fitted to P[+|r]

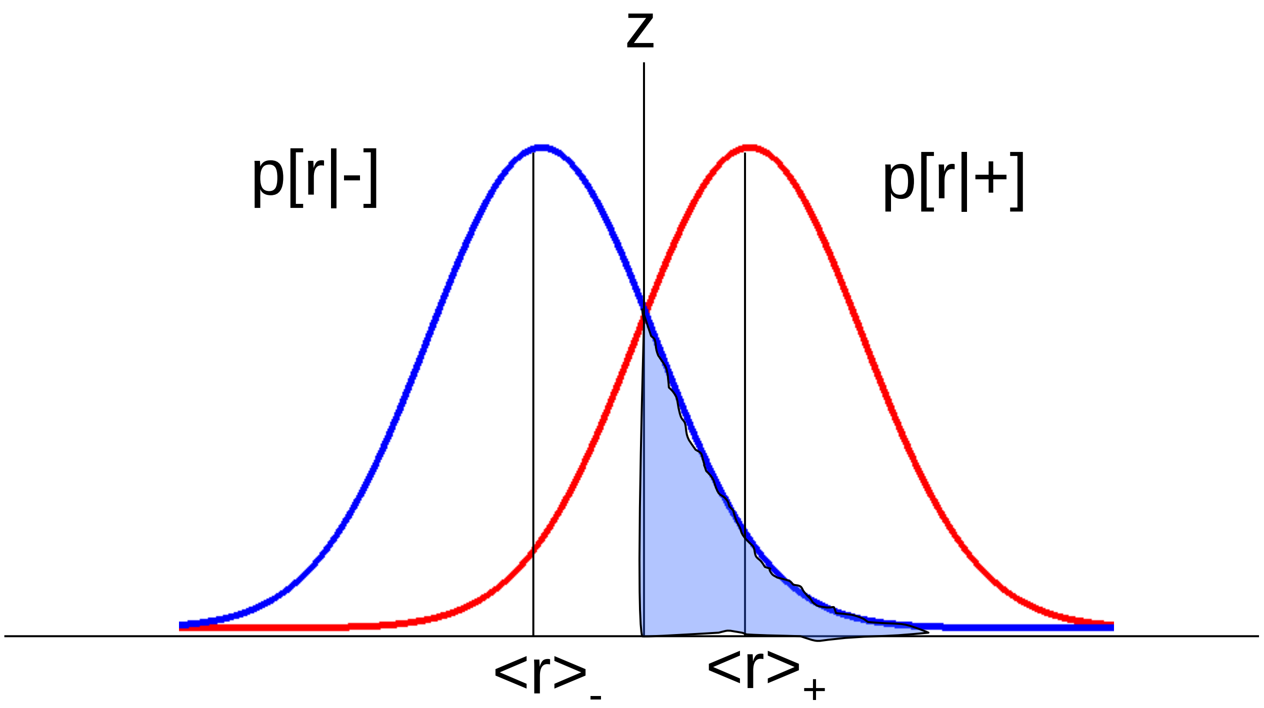
Neurons vs organisms



Close correspondence between neural and behaviour..

Why so many neurons? Correlations limit performance.

Priors

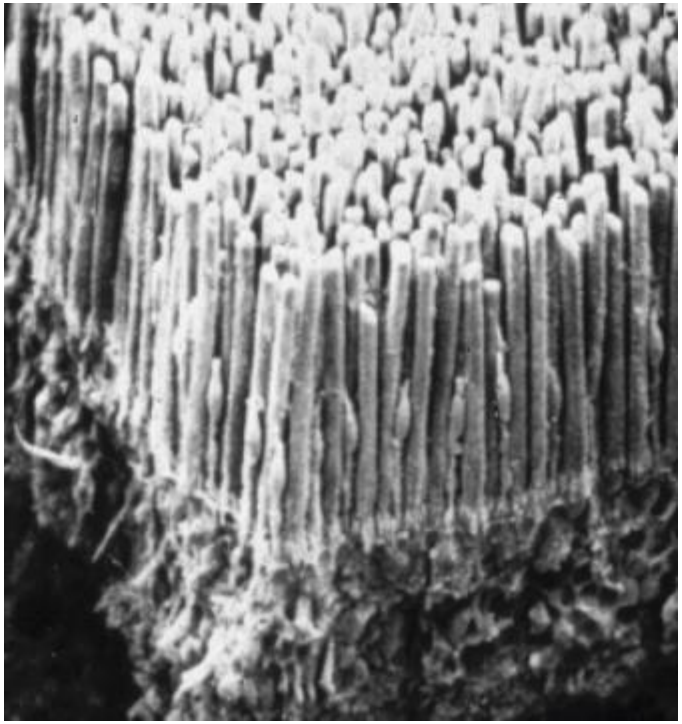


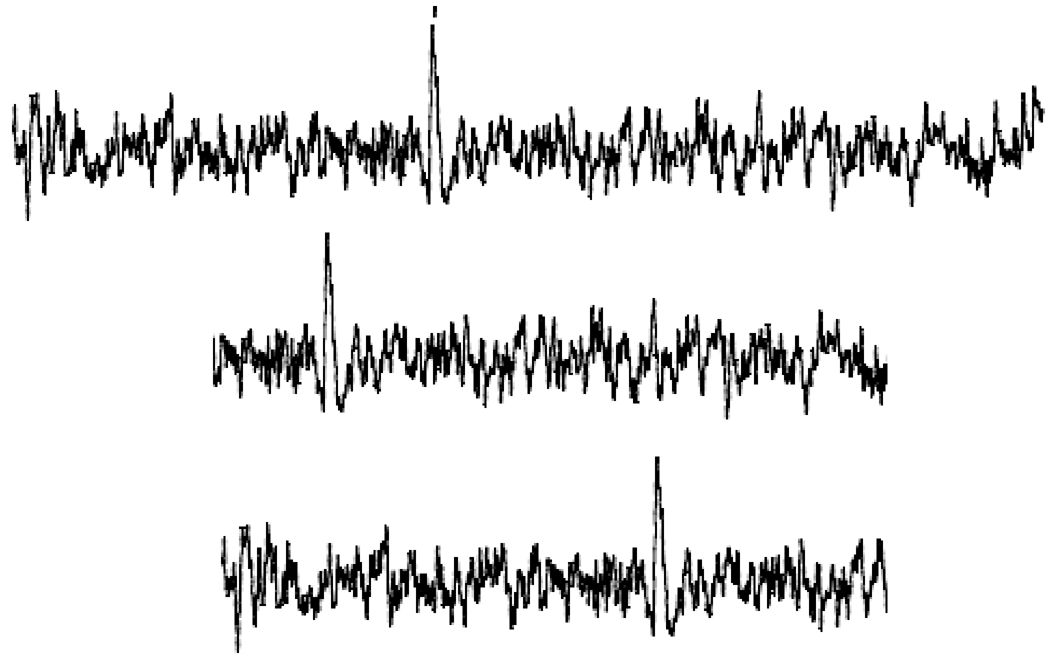
Role of *priors*:

Find z by maximizing P[correct] = p[+] $\beta(z)$ + p[-]$(1 - \alpha(z))$

The wind or a tiger?

Classification of noisy data: single photon responses

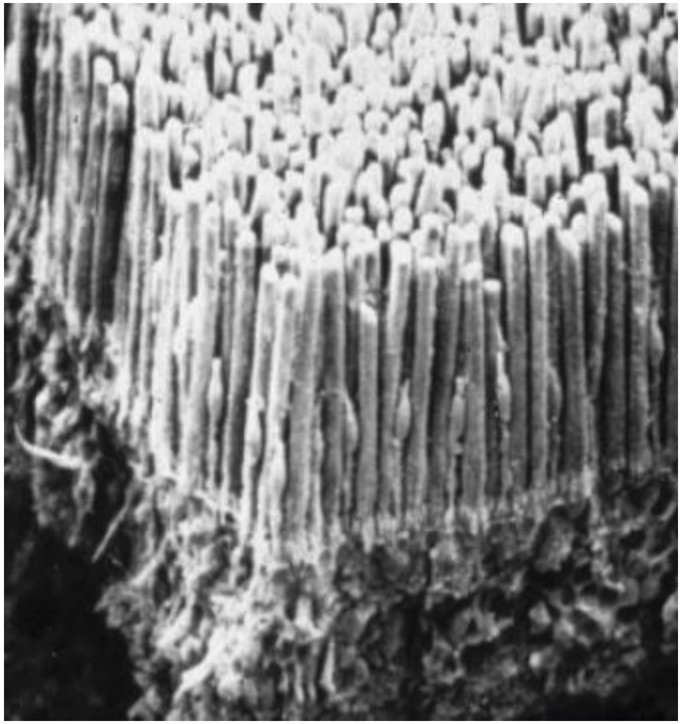


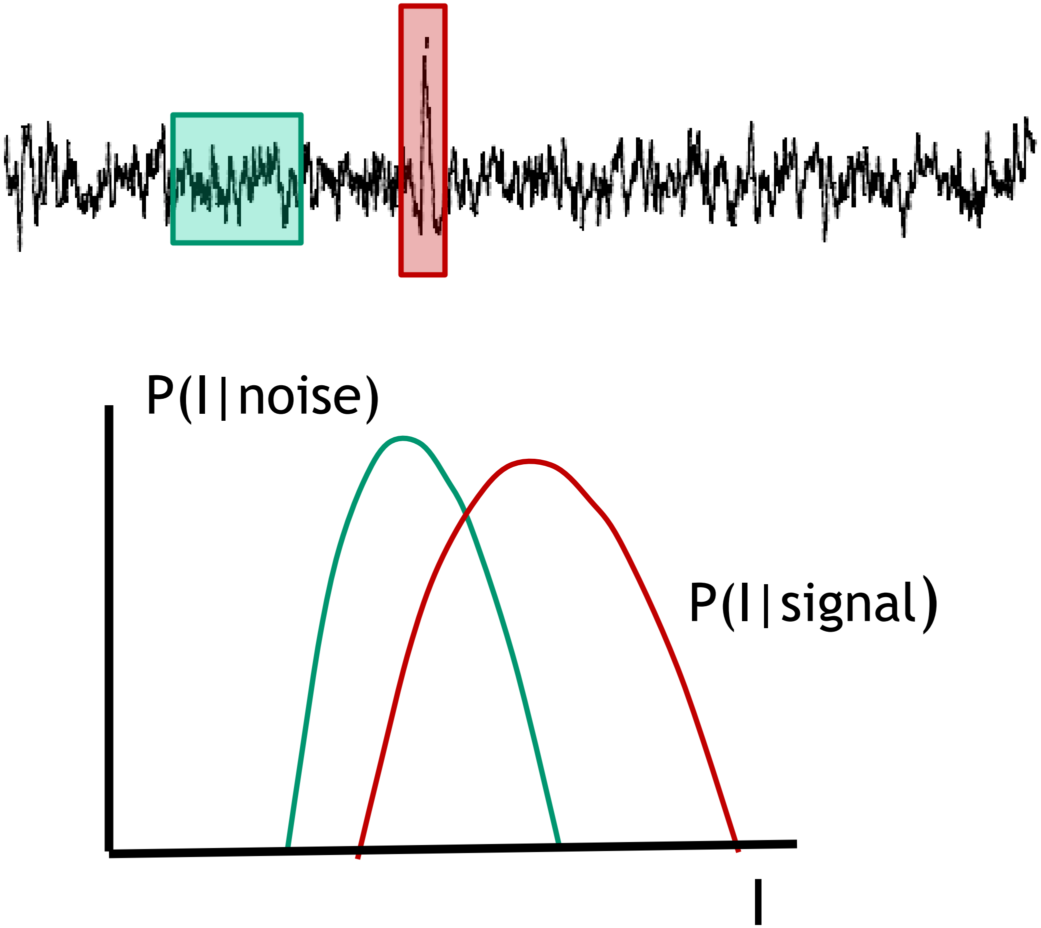


Rieke

Nonlinear separation of signal and noise

Classification of noisy data: single photon responses

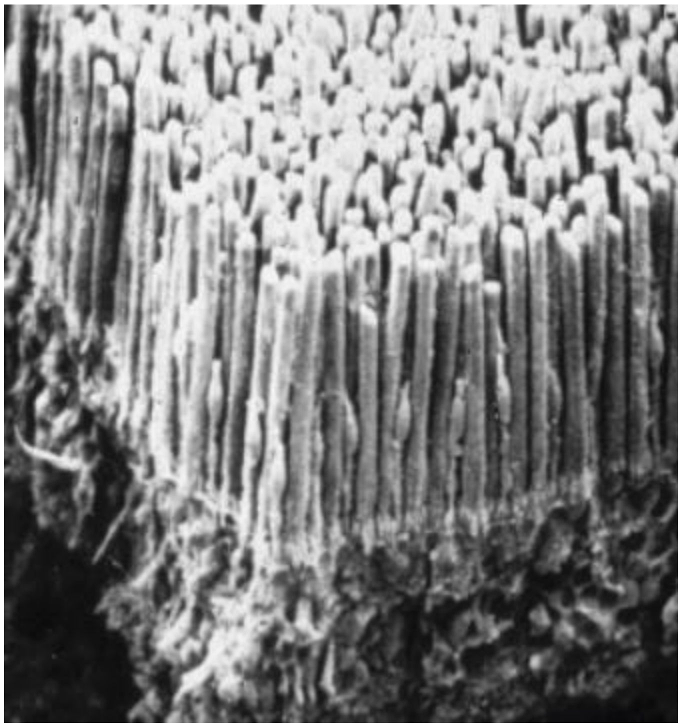


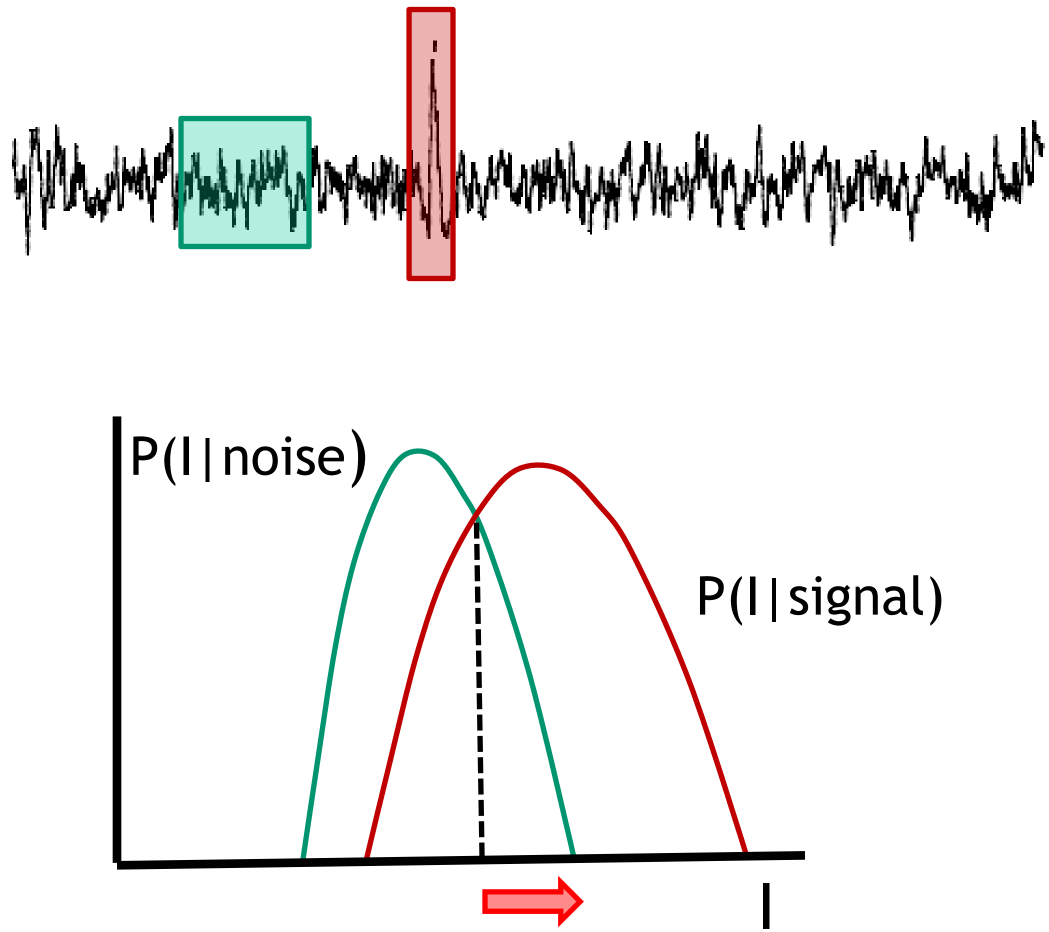


Rieke

Nonlinear separation of signal and noise

Classification of noisy data: single photon responses

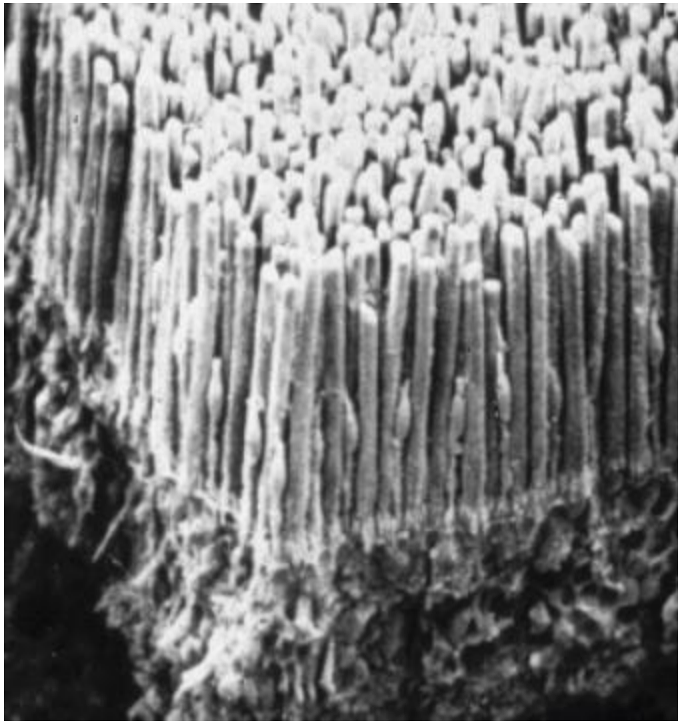


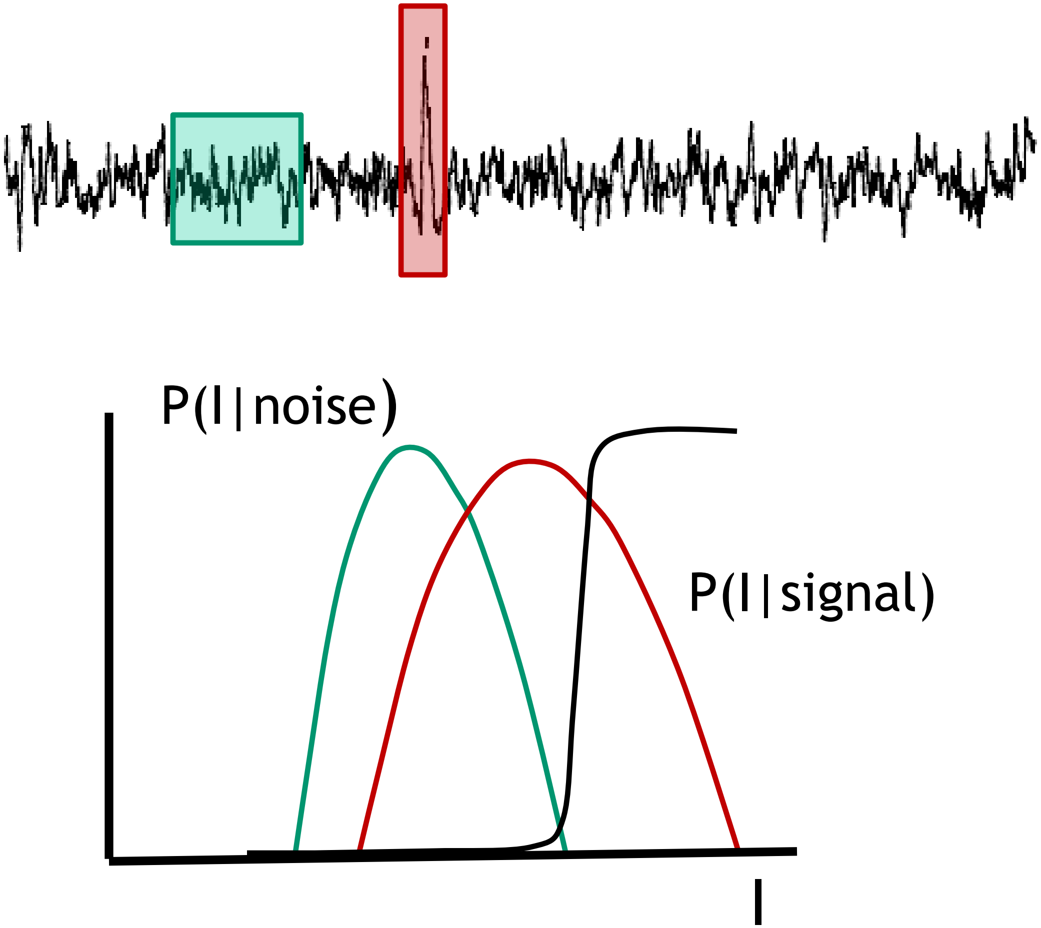


Rieke

Nonlinear separation of signal and noise

Classification of noisy data: single photon responses

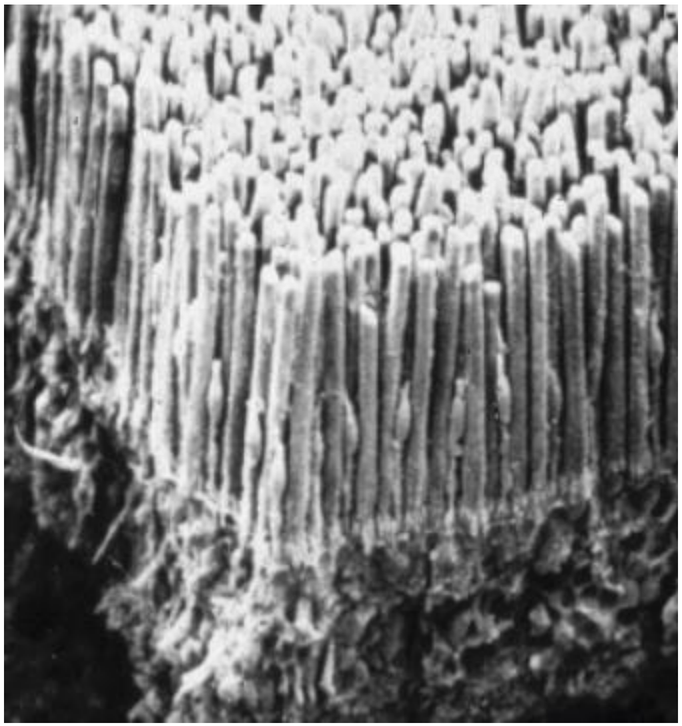


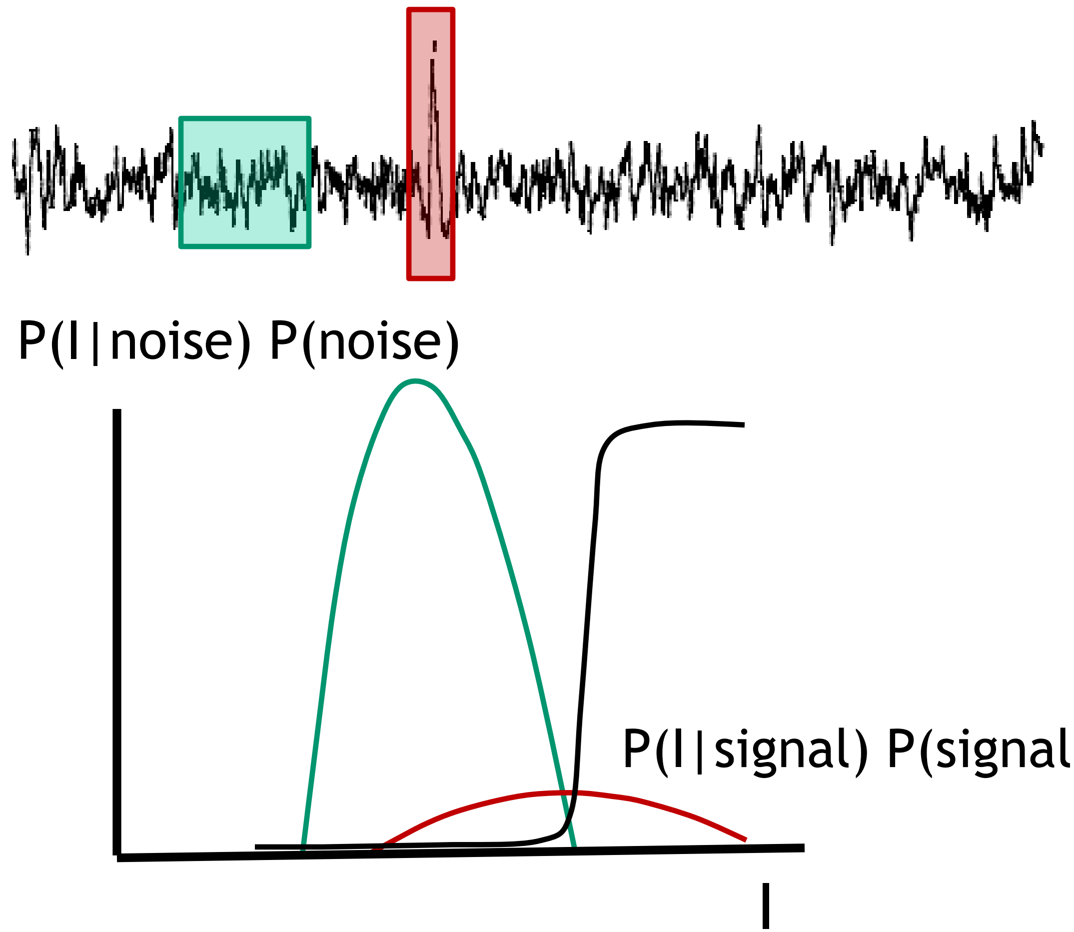


Rieke

Nonlinear separation of signal and noise

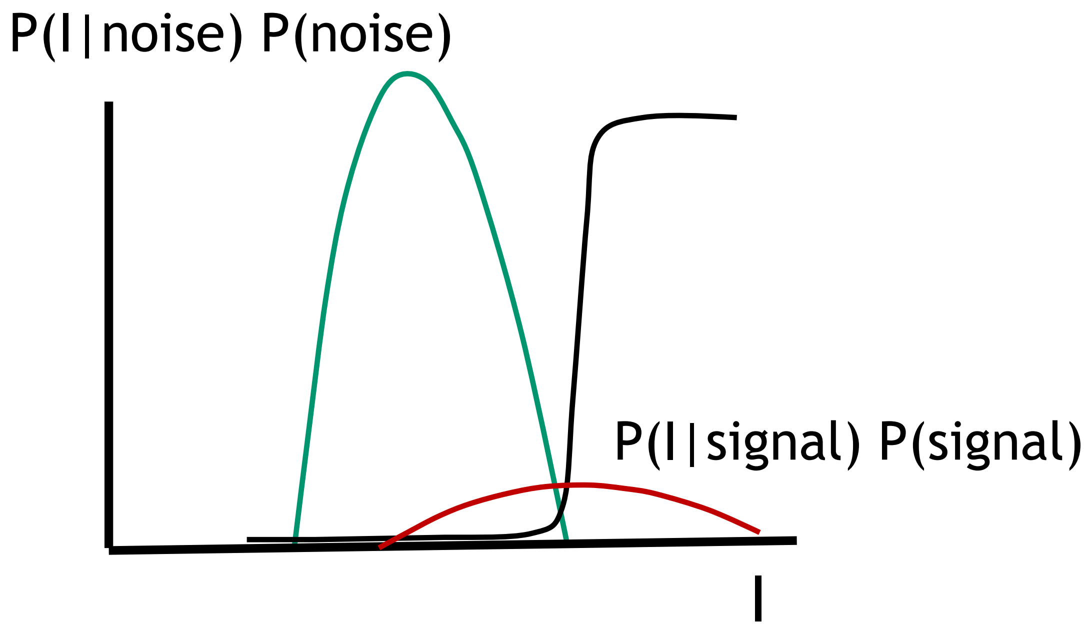
Classification of noisy data: single photon responses

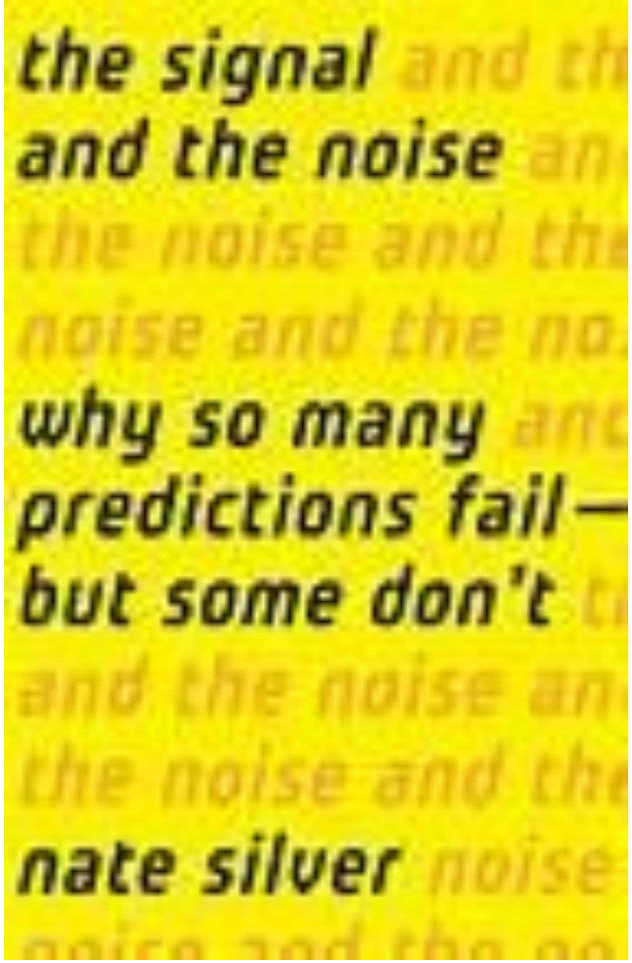




Rieke

How about costs?



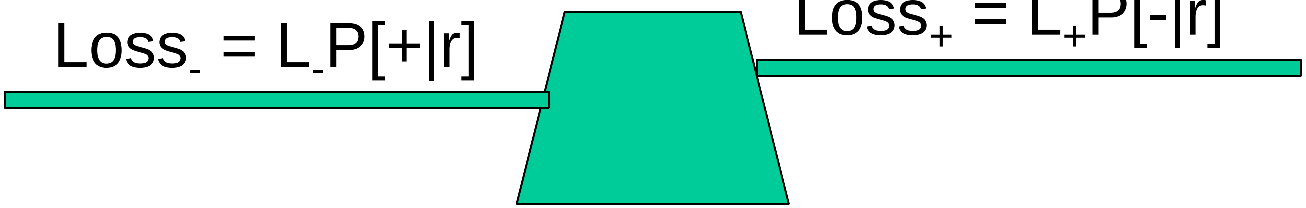


Building in cost

Penalty for incorrect answer: L_+, L_-

For an observation r, what is the expected **loss**?

$Loss_- = L_- P[+|r]$ $\qquad$ $Loss_+ = L_+ P[-|r]$



Cut your losses: answer + when $Loss_+ < Loss_-$

i.e. $\qquad L_+ P[-|r] < L_- P[+|r]$.

Using Bayes', $\qquad P[+|r] = p[r|+]P[+]/p(r)$;

$P[-|r] = p[r|-]P[-]/p(r)$;

$\rightarrow$ $l(r) = p[r|+]/p[r|-] > L_+ P[-] / L_- P[+]$.

Relationship of likelihood to tuning curves

For small stimulus differences s and s + δs

$$\frac{p[r|s+\delta s]}{p[r|s]} \sim \frac{p[r|s] + \delta s \partial p[r|s]/\partial s}{p[r|s]}$$

$$= 1 + \delta s \frac{\partial \ln p[r|s]}{\partial s}.$$

→ like comparing $Z(r) = \frac{\partial \ln p[r|s]}{\partial s}$

to threshold $(z-1)/\delta s$

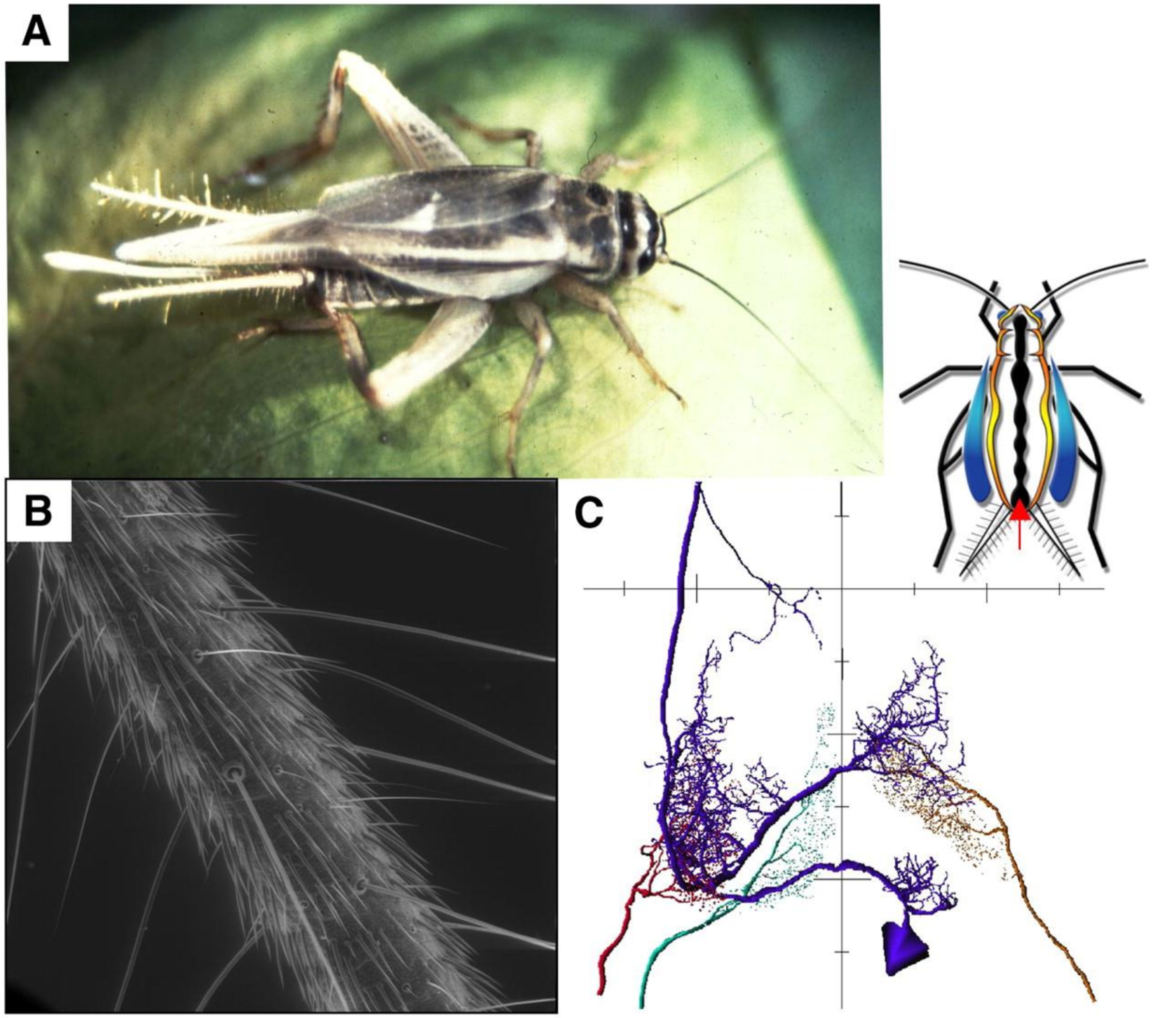
Decoding from many neurons: population codes

- Population code formulation

- Methods for decoding:
  - → population vector
  - → Bayesian inference
  - → maximum likelihood
  - → maximum a posteriori

- Fisher information

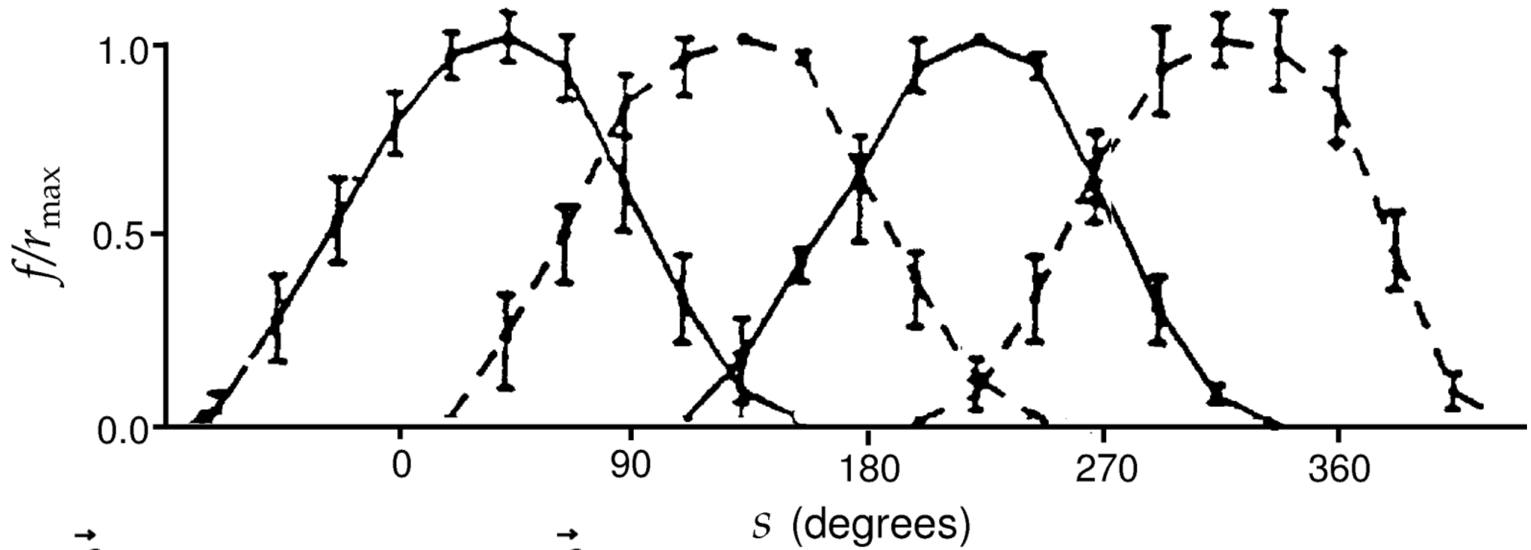
Cricket cercal cells

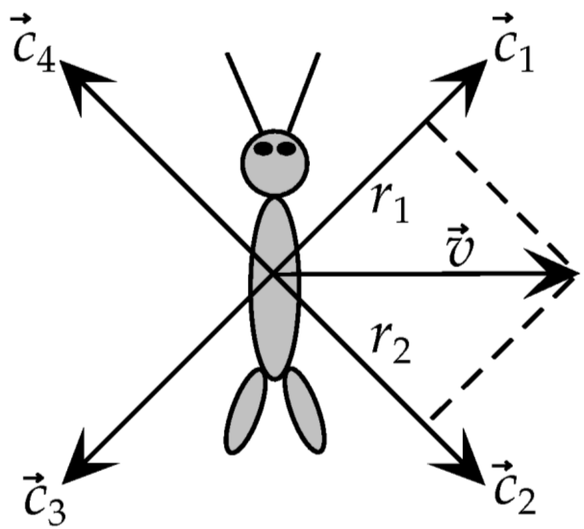


Jacobs G A et al. J Exp Biol 2008;211:1819-1828

The Journal of Experimental Biology

Cricket cercal cells



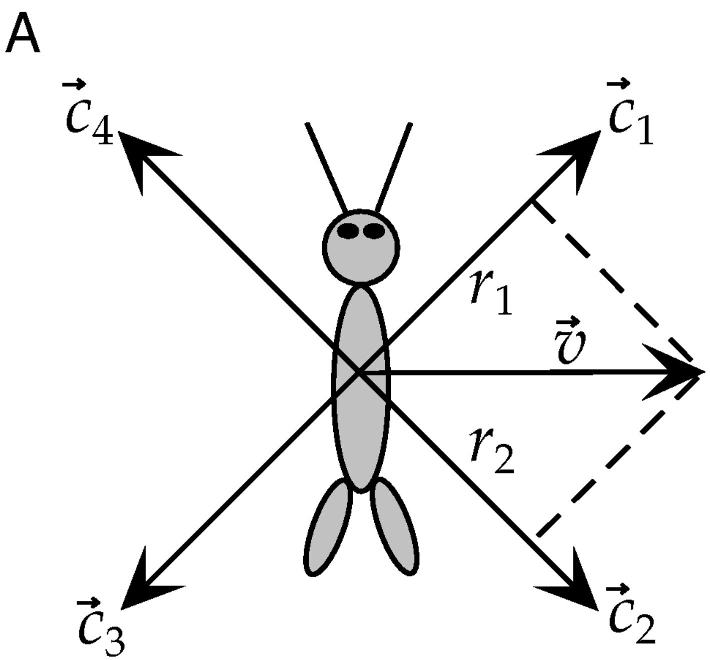


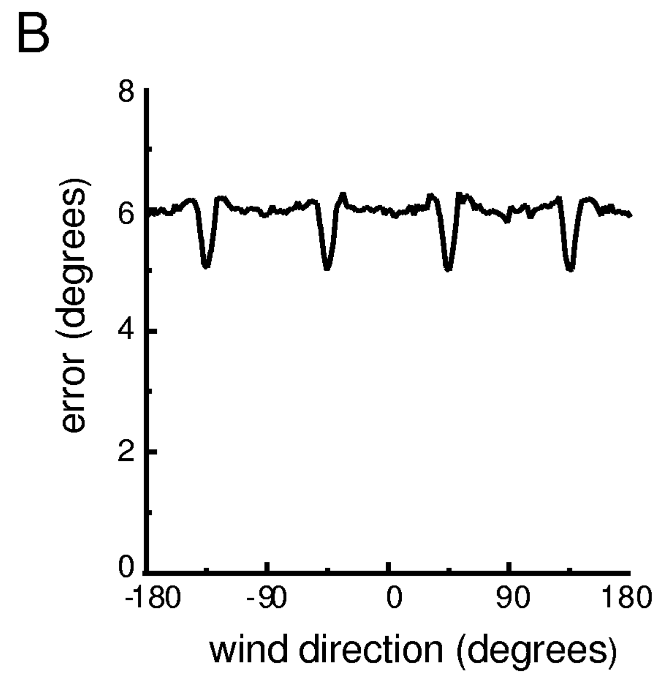
$$\left(\frac{f(s)}{r_{\max}}\right)_a = [\cos(s - s_a)]_+$$

$$\left(\frac{f(s)}{r_{\max}}\right)_a = [\vec{v} \cdot \vec{c}_a]_+$$

Population vector

$$\vec{v}_{\text{pop}} = \sum_{a=1}^{4} \left(\frac{r}{r_{\max}} \right)_a \vec{c}_a$$

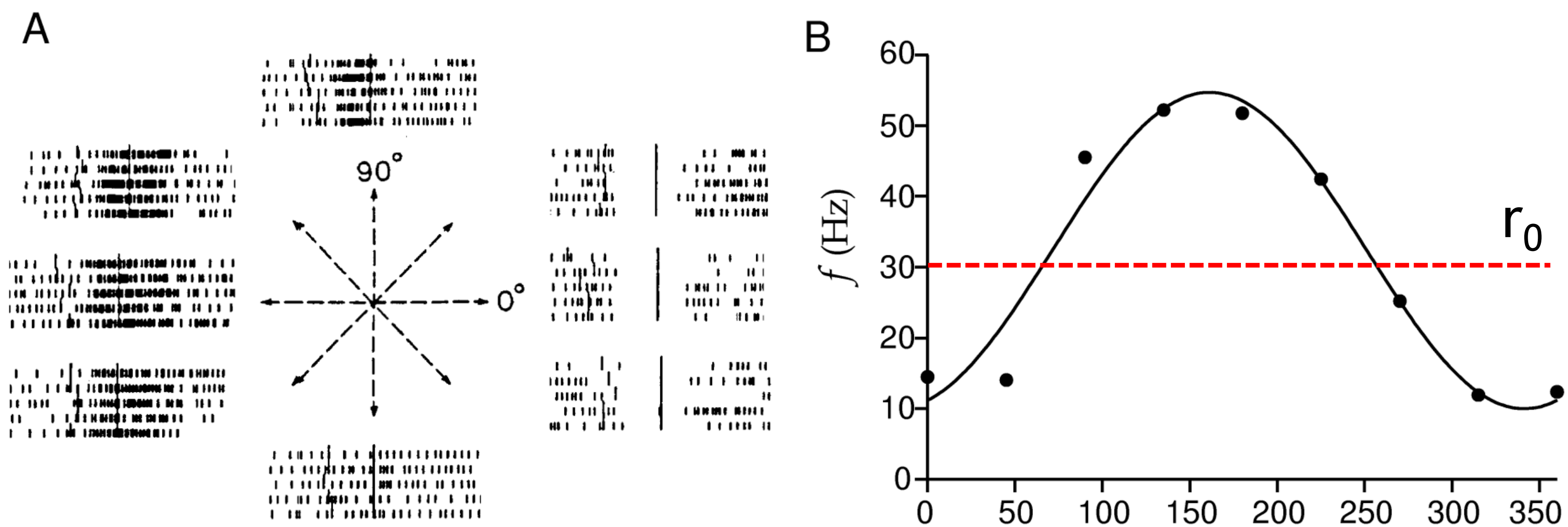




RMS error in estimate

Theunissen & Miller, 1991

Population coding in M1



Hand reaching direction

Cosine tuning curve of a motor cortical neuron

Population coding in M1

Cosine tuning:

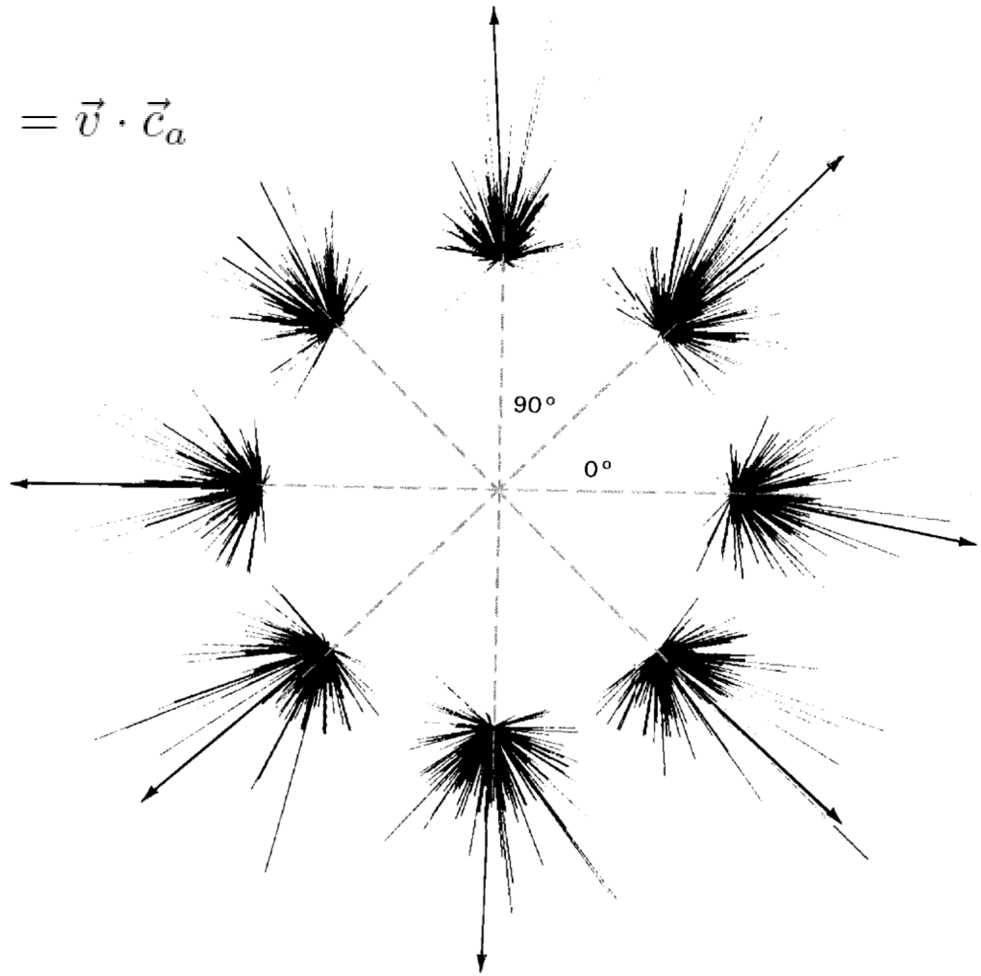
$$\left(\frac{\langle r\rangle - r_0}{r_{\max}}\right)_a = \left(\frac{f(s) - r_0}{r_{\max}}\right)_a = \vec{v}\cdot\vec{c}_a$$

Pop. vector:

$$\vec{v}_{\text{pop}} = \sum_{a=1}^{N}\left(\frac{r - r_0}{r_{\max}}\right)\vec{c}_a$$

For sufficiently large N,

$$\langle\vec{v}_{\text{pop}}\rangle = \sum_{a=1}^{N}(\vec{v}\cdot\vec{c}_a)\vec{c}_a$$



is parallel to the direction of arm movement

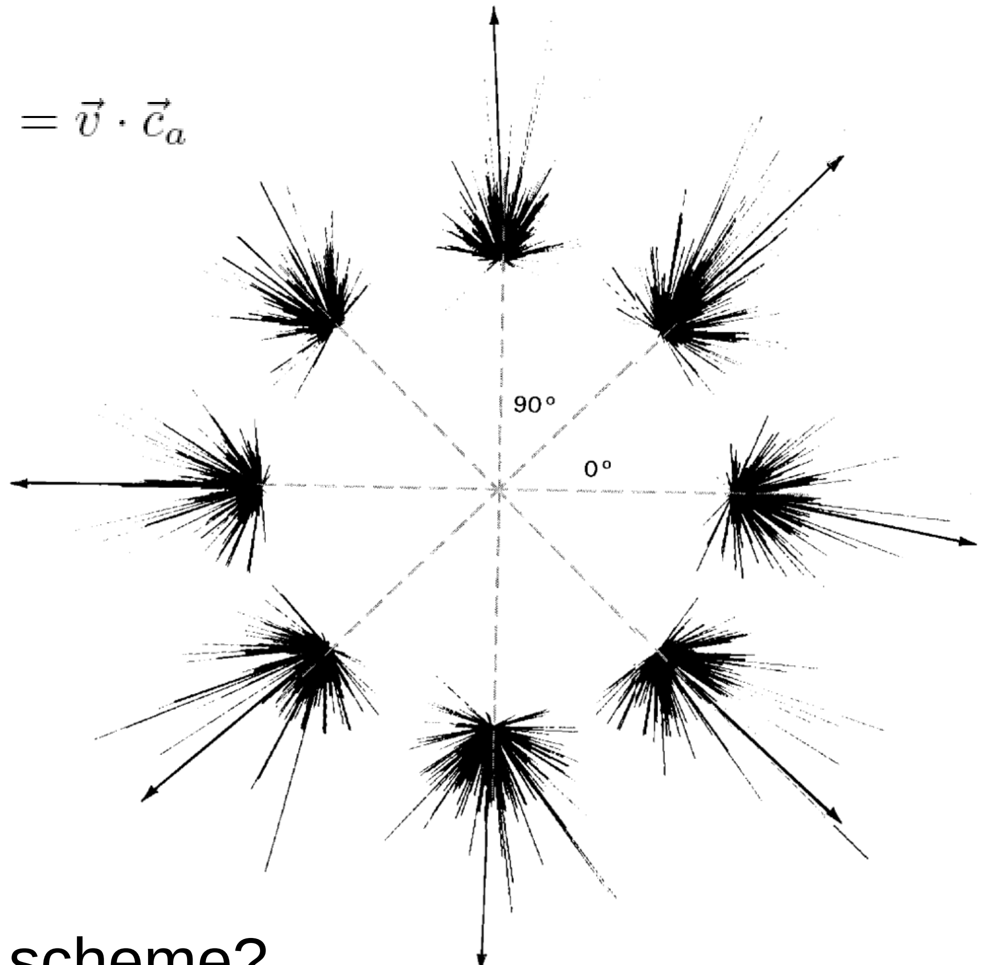
Population coding in M1

Cosine tuning:

$$\left(\frac{\langle r\rangle - r_0}{r_{\max}}\right)_a = \left(\frac{f(s) - r_0}{r_{\max}}\right)_a = \vec{v}\cdot\vec{c}_a$$

Pop. vector:

$$\vec{v}_{\text{pop}} = \sum_{a=1}^{N}\left(\frac{r - r_0}{r_{\max}}\right)\vec{c}_a$$



Difficulties with this coding scheme?

Is this the best one can do?

The population vector is neither general nor optimal.

"Optimal":

make use of all information in the stimulus/response distributions

Bayesian inference

Bayes' law:

likelihood function

conditional distribution

$$p[s|\mathbf{r}] = \frac{p[\mathbf{r}|s]\,p[s]}{p[\mathbf{r}]}$$

prior distribution

a posteriori distribution

marginal distribution

Bayesian estimation

Want an estimator s_{Bayes}

Introduce a cost function, $L(s,s_{Bayes})$; minimize mean cost.

$$\int ds\, L(s, s_{\text{bayes}})p[s|\mathbf{r}]$$

For least squares cost, $L(s,s_{Bayes}) = (s - s_{Bayes})^2$;
solution is the conditional mean.

$$s_{\text{bayes}} = \int ds\, p[s|\mathbf{r}]s$$

Bayesian inference

By Bayes' law,

likelihood function

$$p[s|\mathbf{r}] = \frac{p[\mathbf{r}|s]p[s]}{p[\mathbf{r}]}$$

a posteriori distribution

Maximum likelihood

Find maximum of P[r|s] over s

More generally, probability of the data given the "model"

"Model" = stimulus

assume parametric form for tuning curve

Bayesian inference

By Bayes' law,

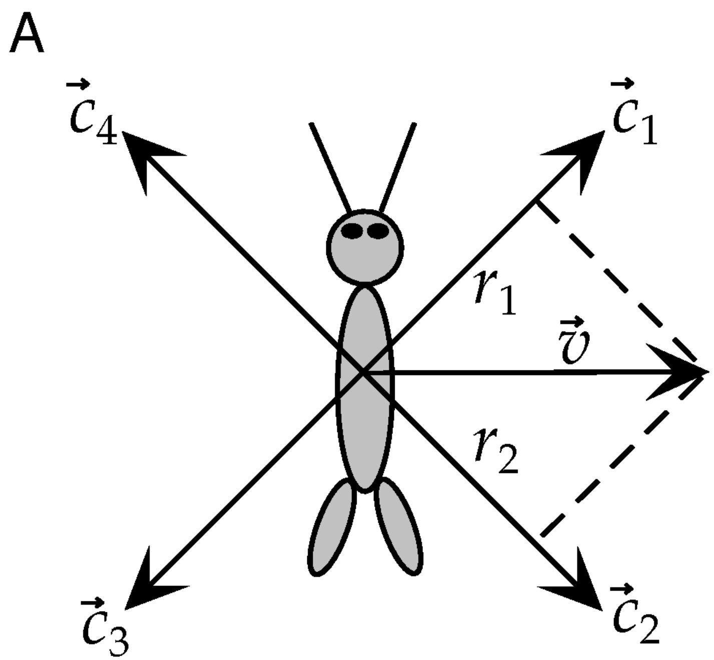
likelihood function

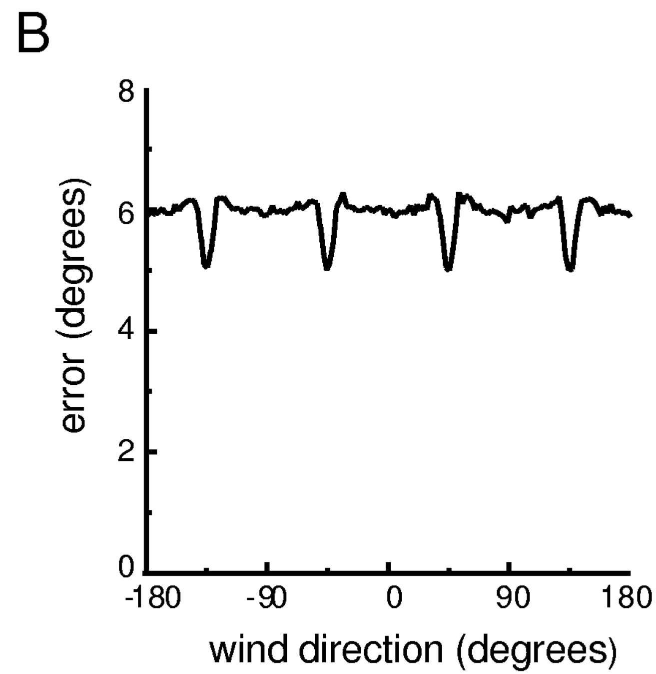
$$p[s|\mathbf{r}] = \frac{p[\mathbf{r}|s]\,p[s]}{p[\mathbf{r}]}$$

a posteriori distribution

Population vector

$$\vec{v}_{\text{pop}} = \sum_{a=1}^{4} \left(\frac{r}{r_{\text{max}}} \right)_a \vec{c}_a$$





RMS error in estimate

Theunissen & Miller, 1991

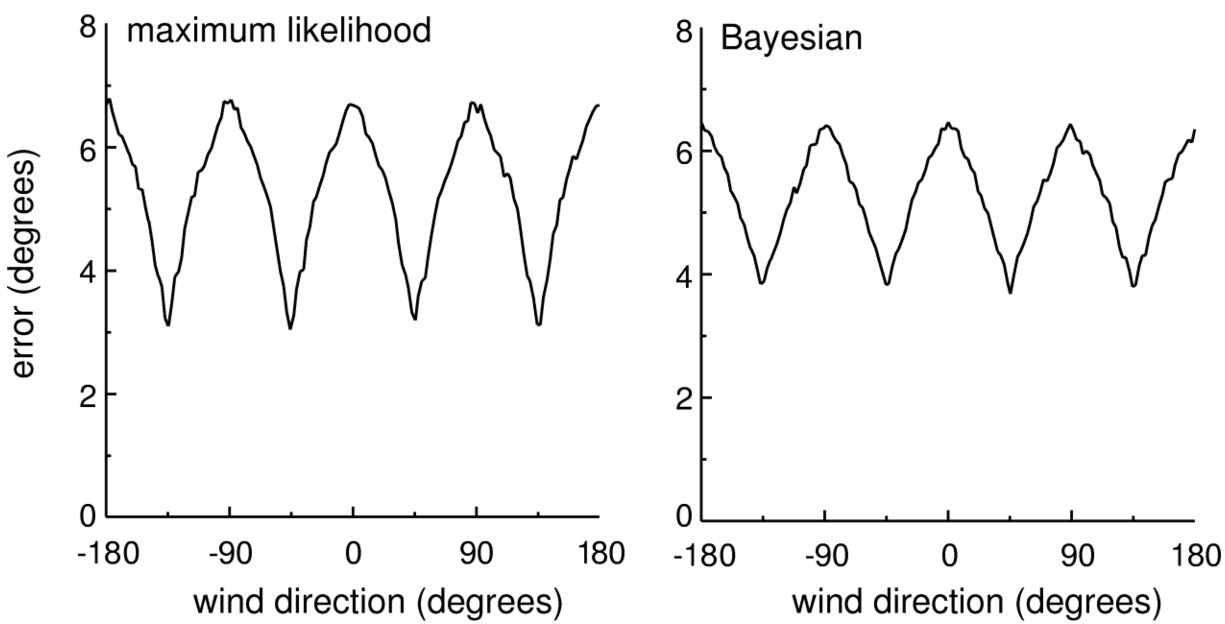
MAP and ML

ML: s* which maximizes p[r|s]

MAP: s* which maximizes p[s|r]

Difference is the role of the prior: differ by factor p[s]/p[r]

For cercal data:



Decoding an arbitrary continuous stimulus

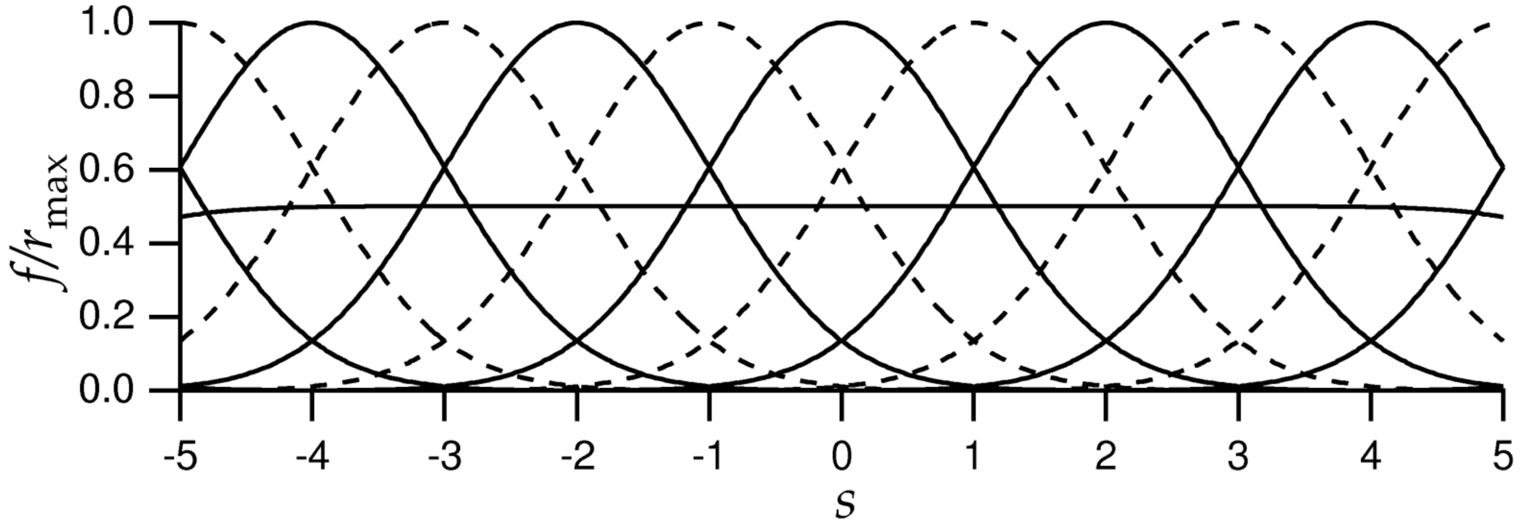
Work through a specific example

- assume independence
- assume Poisson firing

Noise model: Poisson distribution

$$P_T[k] = (\lambda T)^k \exp(-\lambda T)/k!$$

Decoding an arbitrary continuous stimulus



E.g. Gaussian tuning curves

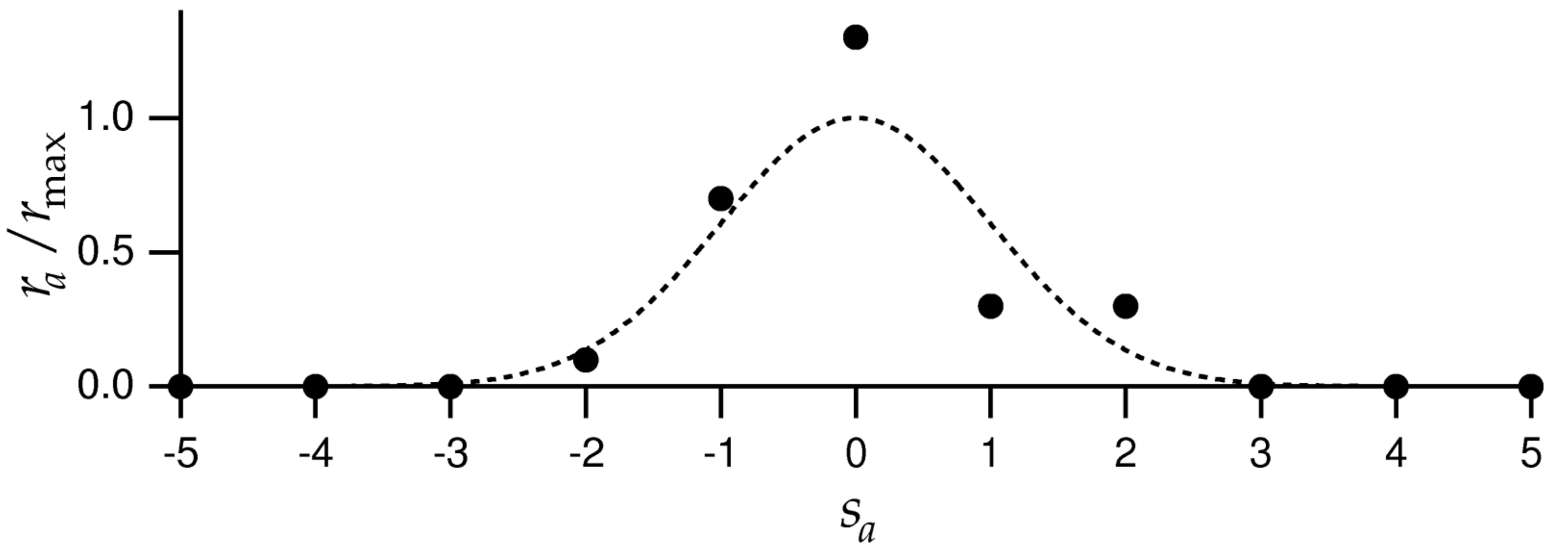
$$f_a(s) = r_{\max} \exp\left(-\frac{1}{2} \left[\frac{(s - s_a)}{\sigma_a} \right]^2 \right)$$

$$\sum_{a=1}^{N} f_a(s) \;\; \text{const.}$$

Need to know full P[**r**|s]

Assume Poisson:
$$P[r_a|s] = \frac{(f_a(s)T)^{r_a T}}{(r_a T)!}\exp(-f_a(s)T)$$

Assume independent:
$$P[\mathbf{r}|s] = \prod_{a=1}^{N} \frac{(f_a(s)T)^{r_a T}}{(r_a T)!}\exp(-f_a(s)T)$$



Population response of 11 cells with Gaussian tuning curves

ML

Apply ML: maximize ln P[**r**|s] with respect to s

$$\ln P[\mathbf{r}|s] = T \sum_{a=1}^{N} r_a \ln(f_a(s)) + \ldots$$

Set derivative to zero, use sum = constant

$$\sum_{a=1}^{N} r_a \frac{f'(s^*)}{f(s^*)} = 0$$

From Gaussianity of tuning curves,

$$s^* = \frac{\sum r_a s_a/\sigma_a^2}{\sum r_a/\sigma_a^2}$$

If all σ same

$$s^* = \frac{\sum r_a s_a}{\sum r_a}$$

MAP

Apply MAP: maximise ln p[s|**r**] with respect to s

$$\ln p[s|\mathbf{r}] = \ln P[\mathbf{r}|s] + \ln p[s] - \ln P[\mathbf{r}]$$

$$\ln p[s|\mathbf{r}] = T \sum_{a=1}^{N} r_a \ln(f_a(s)) + \ln p[s] + \ldots$$

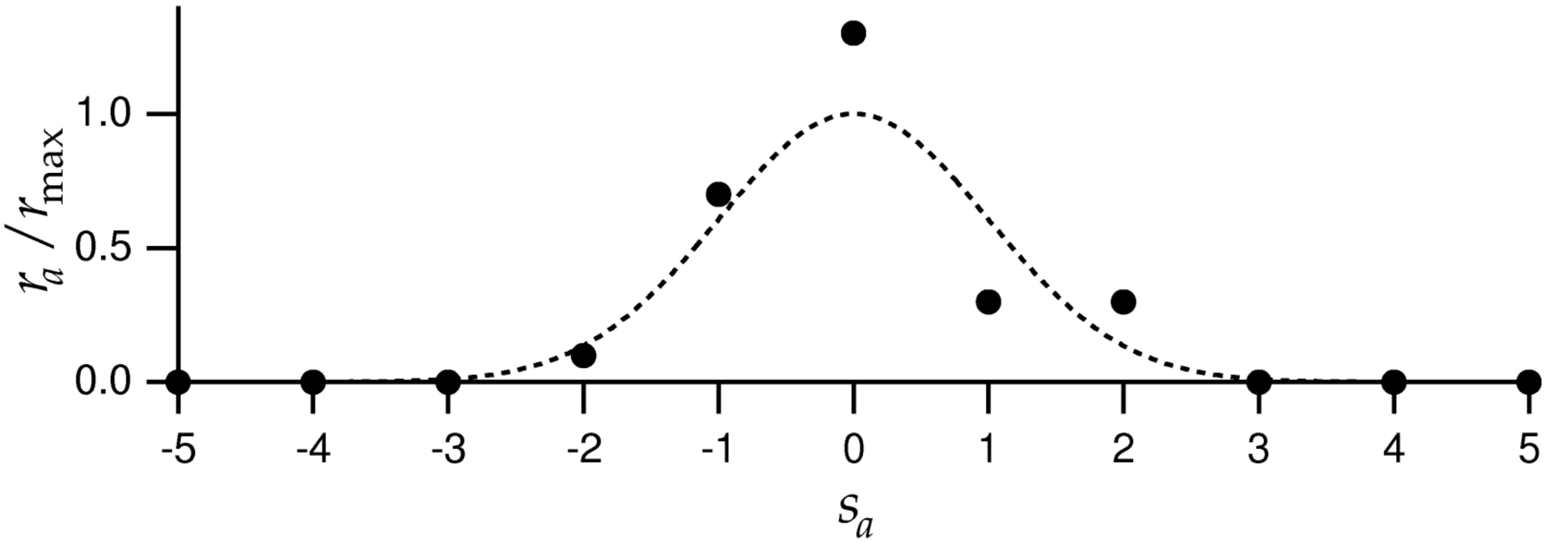
Set derivative to zero, use sum = constant

$$\sum_{a=1}^{N} r_a \frac{f'(s^*)}{f(s^*)} + \frac{p'[s]}{p[s]} = 0$$

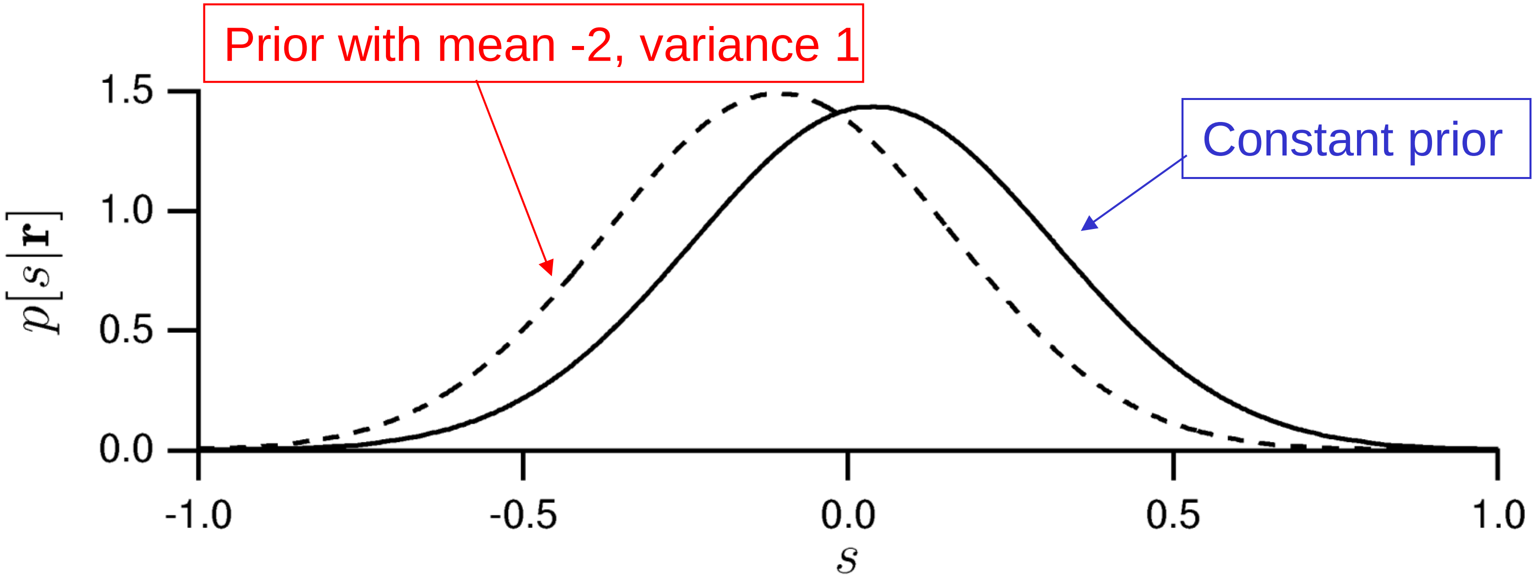
From Gaussianity of tuning curves,

$$s^* = \frac{T \sum r_a s_a / \sigma_a^2 + s_{\text{prior}} / \sigma_{\text{prior}}^2}{T \sum r_a / \sigma_a^2 + 1/\sigma_{\text{prior}}^2}$$

Given this data:



MAP:



How good is our estimate?

For stimulus s, have estimated s_{est}

Bias: $$b_{\text{est}}(s) = \langle s_{\text{est}} - s \rangle$$

Variance: $$\sigma^2_{\text{est}}(s) = \langle (s_{\text{est}} - \langle s_{\text{est}} \rangle)^2 \rangle$$

Mean square error:

$$\langle (s_{\text{est}} - s)^2 \rangle = \langle (s_{\text{est}} - \langle s_{\text{est}} \rangle + b_{\text{est}}(s))^2 \rangle = \sigma^2_{\text{est}}(s) + b^2_{\text{est}}(s).$$

Cramer-Rao bound: $$\sigma^2_{\text{est}} \geq \frac{(1 + b'_{\text{est}})^2}{I_{\text{F}}(s)}$$

Fisher information

(ML is unbiased: b = b' = 0)

Fisher information

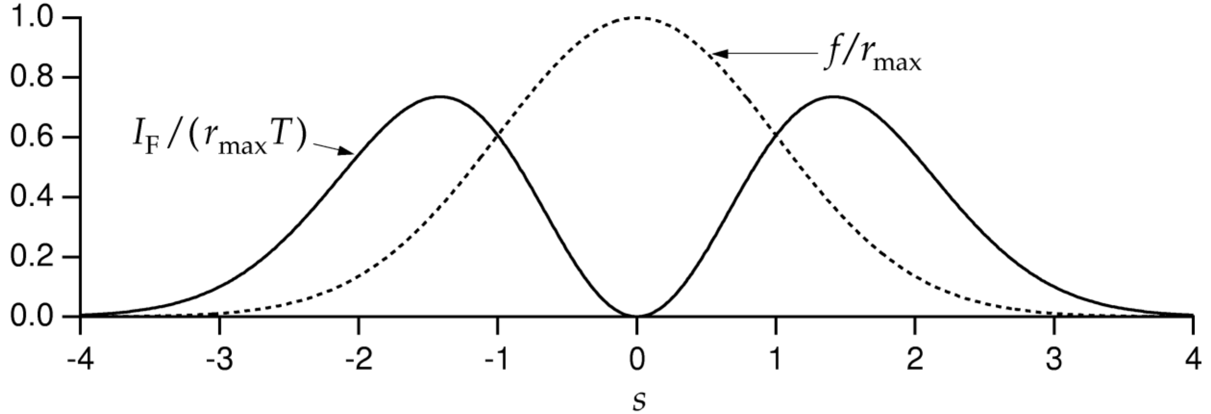
$$I_{\mathrm{F}}(s) = \left\langle -\frac{\partial^2 \ln p[\mathbf{r}|s]}{\partial^2 s} \right\rangle = \int d\mathbf{r}\, p[\mathbf{r}|s] \left(-\frac{\partial^2 \ln p[\mathbf{r}|s]}{\partial s^2} \right)$$

Alternatively:

$$I_{\mathrm{F}}(s) = \left\langle \left(\frac{\partial \ln p[\mathbf{r}|s]}{\partial s} \right)^2 \right\rangle = \int d\mathbf{r}\, p[\mathbf{r}|s] \left(\frac{\partial \ln p[\mathbf{r}|s]}{\partial s} \right)^2$$

Quantifies local stimulus discriminability

Fisher information for Gaussian tuning curves



For the Gaussian tuning curves w/Poisson statistics:

$$I_F(s) = \left\langle \left(\frac{d^2 \ln P[\mathbf{r}|s]}{ds^2} \right) \right\rangle = T \sum_{a=1}^{N} \langle r_a \rangle \left(\left(\frac{f'_a(s)}{f_a(s)} \right)^2 - \frac{f''_a(s)}{f_a(s)} \right)$$

Are narrow or broad tuning curves better?

$$I_{\mathrm{F}} = T \sum_{a=1}^{N} \frac{r_{\max}(s - s_a)^2}{\sigma_r^4} \exp\left(-\frac{1}{2} \left(\frac{s - s_a}{\sigma_r} \right)^2 \right)$$

Approximate: $I_{\mathrm{F}} \sim \frac{\sqrt{2\pi}\rho_s \sigma_r r_{\max} T}{\sigma_r^2}.$

Thus, $I_{\mathrm{F}} \sim 1/\sigma_r$ → Narrow tuning curves are better

But not in higher dimensions!

$$I_{\mathrm{F}} \sim (2\pi)^{D/2} D \rho_s \sigma_r^{D-2} r_{\max} T$$

..what happens in 2D?

Fisher information and discrimination

Recall d' = mean difference/standard deviation

Can also decode and discriminate using decoded values.

Trying to discriminate s and s+Δs:

Difference in ML estimate is Δs (unbiased)
variance in estimate is $1/I_F(s)$.

→

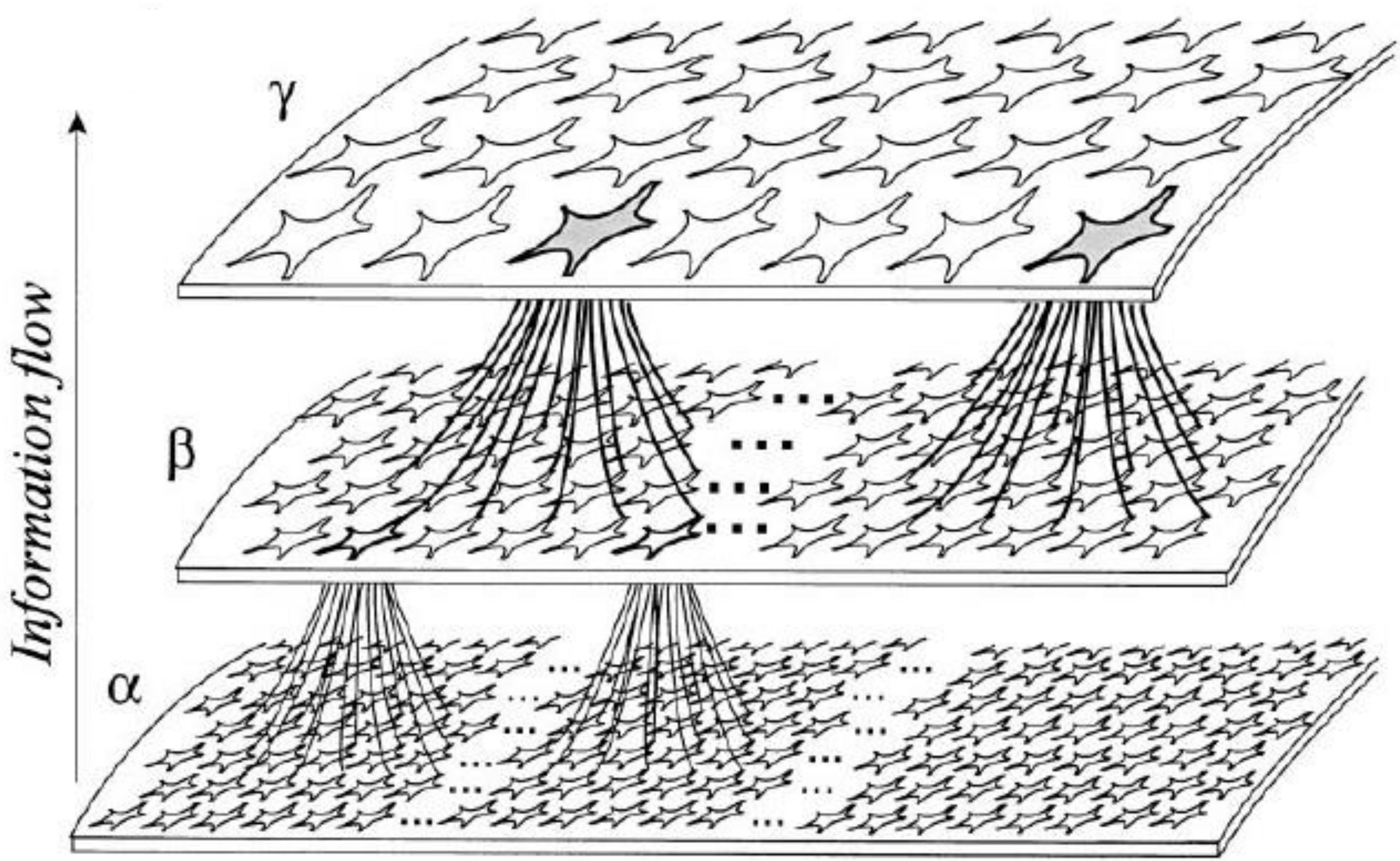
$$d' = \Delta s \sqrt{I_{\mathrm{F}}(s)}$$

Limitations of these approaches

- Tuning curve/mean firing rate

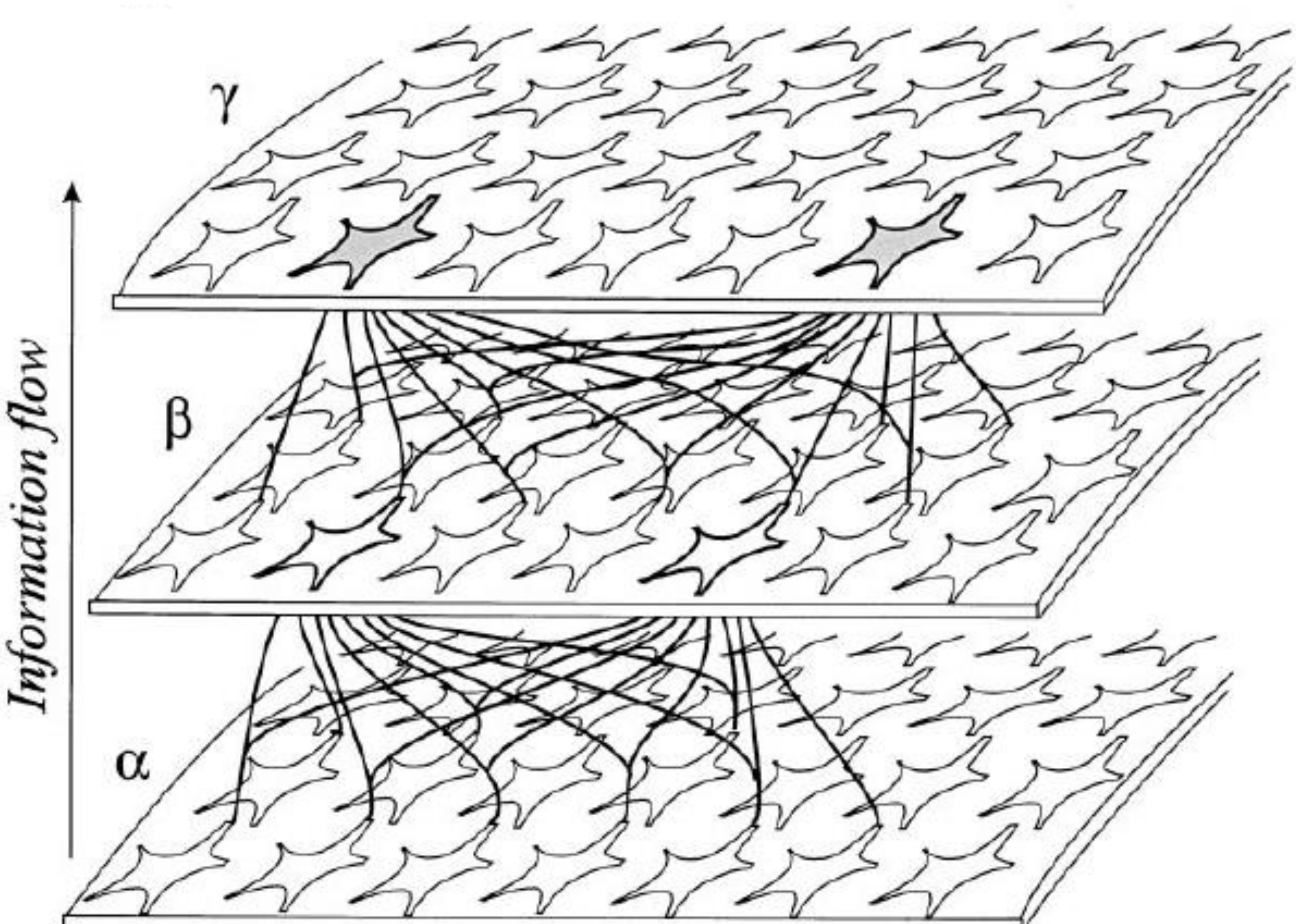
- Correlations in the population

The importance of correlation

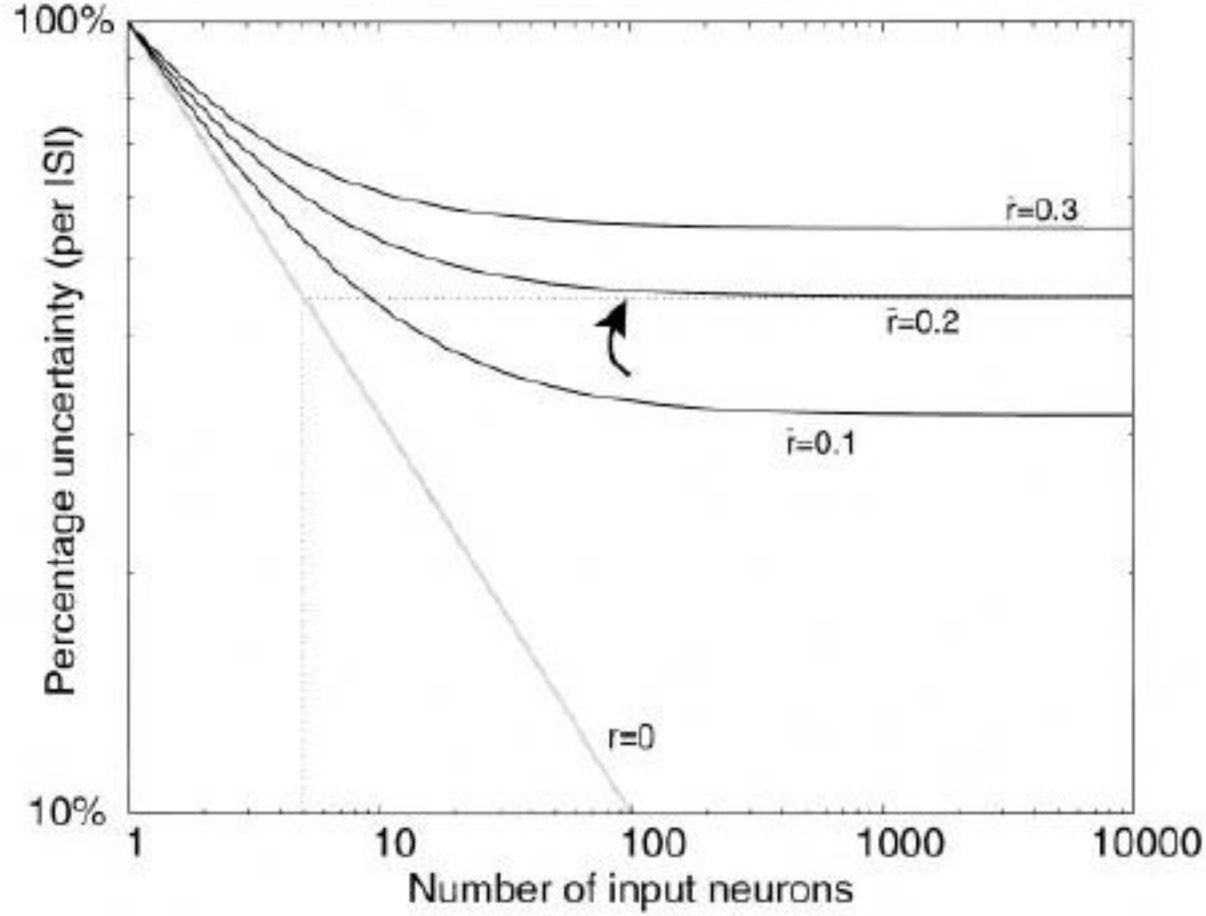


Shadlen and Newsome, '98

The importance of correlation



The importance of correlation



Entropy and Shannon information

Model-based vs model free

Entropy and Shannon information

For a random variable *X* with distribution *p(x),* the **entropy** is

$$H[X] = - \Sigma_x \, p(x) \log_2 p(x)$$

Information is defined as

$$I[X] = - \log_2 p(x)$$

Mutual information

Typically, "information" = *mutual information*:

how much knowing the value of one random variable *r* (the response) reduces uncertainty about another random variable *s* (the stimulus).

Variability in response is due both to different **stimuli** and to **noise**.
How much response variability is "useful", i.e. can represent different messages, depends on the noise. Noise can be specific to a given stimulus.

Mutual information

Information quantifies how *independent* r and s are:

$$I(s;r) = D_{KL}\,[P(r,s), P(r)P(s)]$$

Alternatively:

$$I(s;r) = H[P(r)] - \Sigma_s\, P(s)\, H[P(r|s)]\ .$$

Mutual information

Mutual information is the difference between
the total response entropy and the mean noise entropy:

$$I(s;r) = H[P(r)] - \Sigma_s P(s)\, H[P(r|s)]\ .$$

→ Need to know the conditional distribution $P(s|r)$ or $P(r|s)$.

Take a particular stimulus $s=s_0$ and repeat many times to obtain $P(r|s_0)$.
Compute variability due to noise: *noise entropy*

Mutual information

Information is symmetric in r and s

Examples:

response is unrelated to stimulus: p[r|s] = ?, MI = ?

response is perfectly predicted by stimulus: p[r|s] = ?

Simple example

r_+ encodes stimulus +, r_- encodes stimulus -

but with a probability of error:

$P(r_+|+) = 1- p$

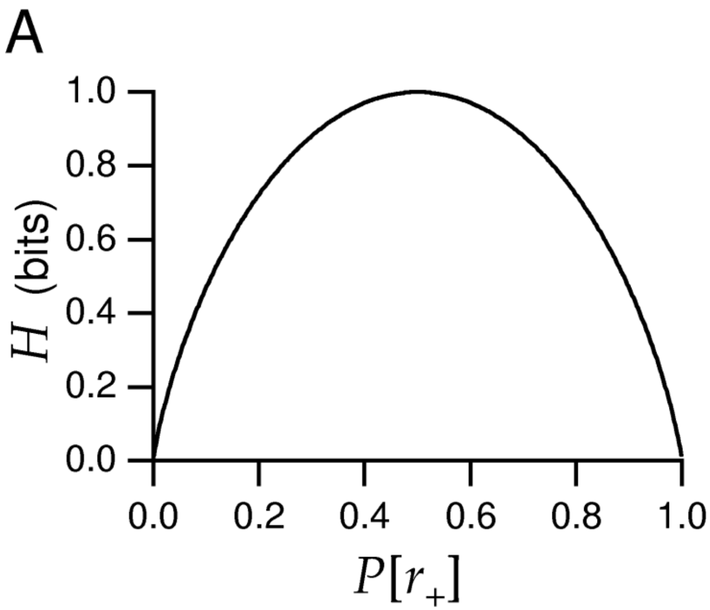
$P(r_-|-) = 1- p$

What is the response entropy H[p]?

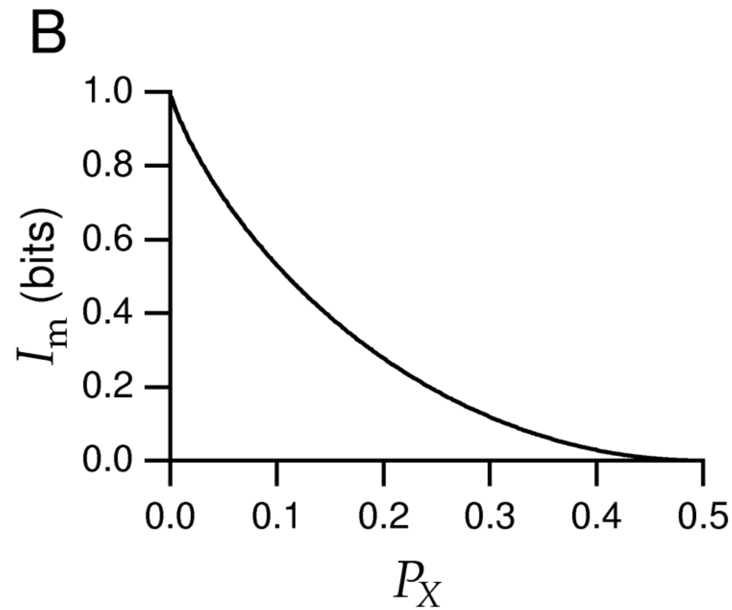
What is the noise entropy?

Entropy and Shannon information

Entropy



Information



$H[p] = -p_+ \log p_+ - (1-p_+)\log(1-p_+)$

When $p_+ = ½$,

$H[P(r|s)] = -p \log p - (1-p)\log(1-p)$