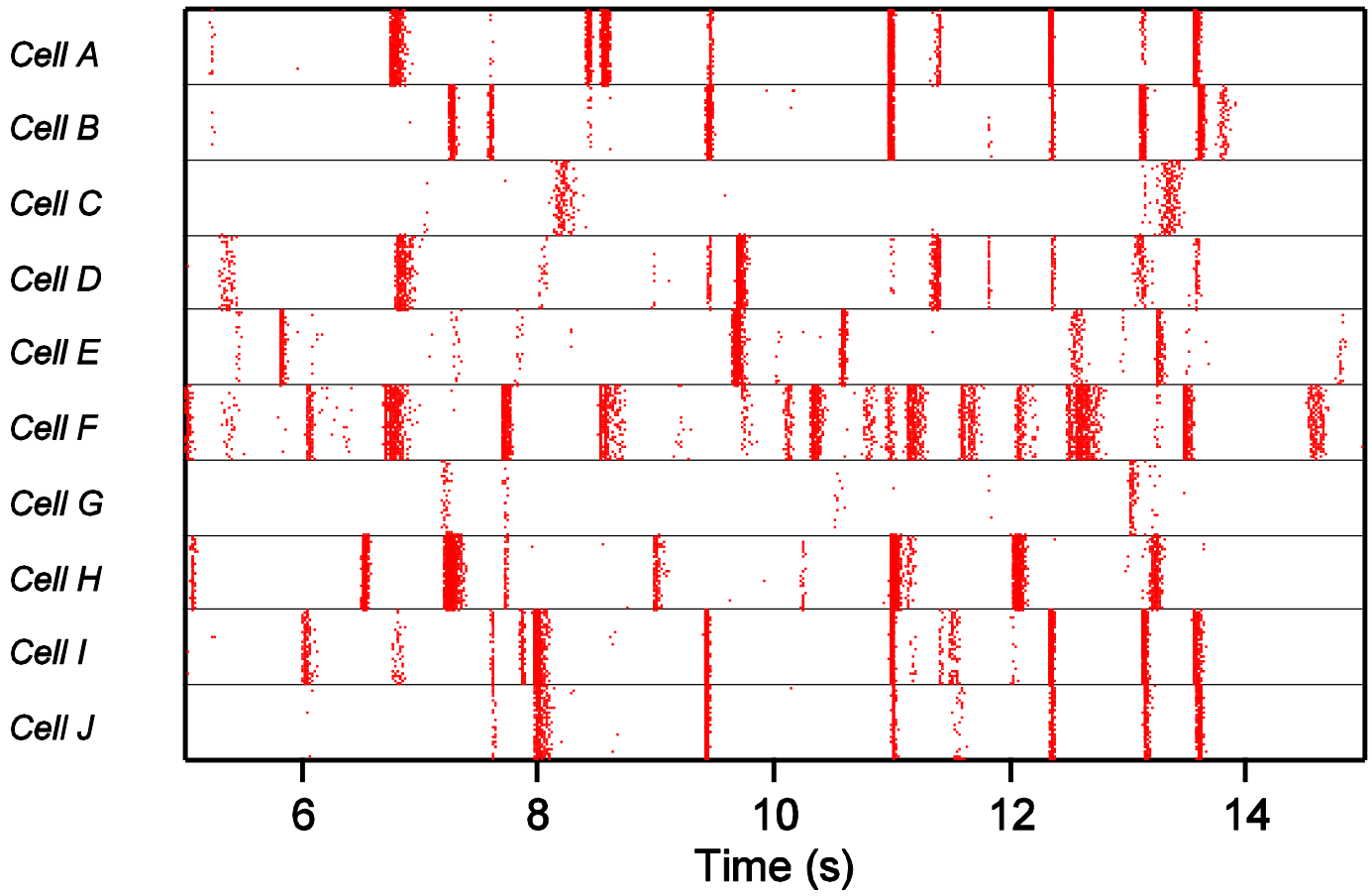
Evaluating neural codes



Entropy and information

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

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

Entropy and information

For a random variable X with distribution $p(x)$, entropy is given by

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

"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.

Entropy and information

For a random variable X with distribution $p(x)$, entropy is given by

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

"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.

Information quantifies how *independent* r and s are:

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

Entropy and information

→ 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*

Information is the difference between the total response entropy and the mean noise entropy:

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

Entropy and information

Information is symmetric in r and s

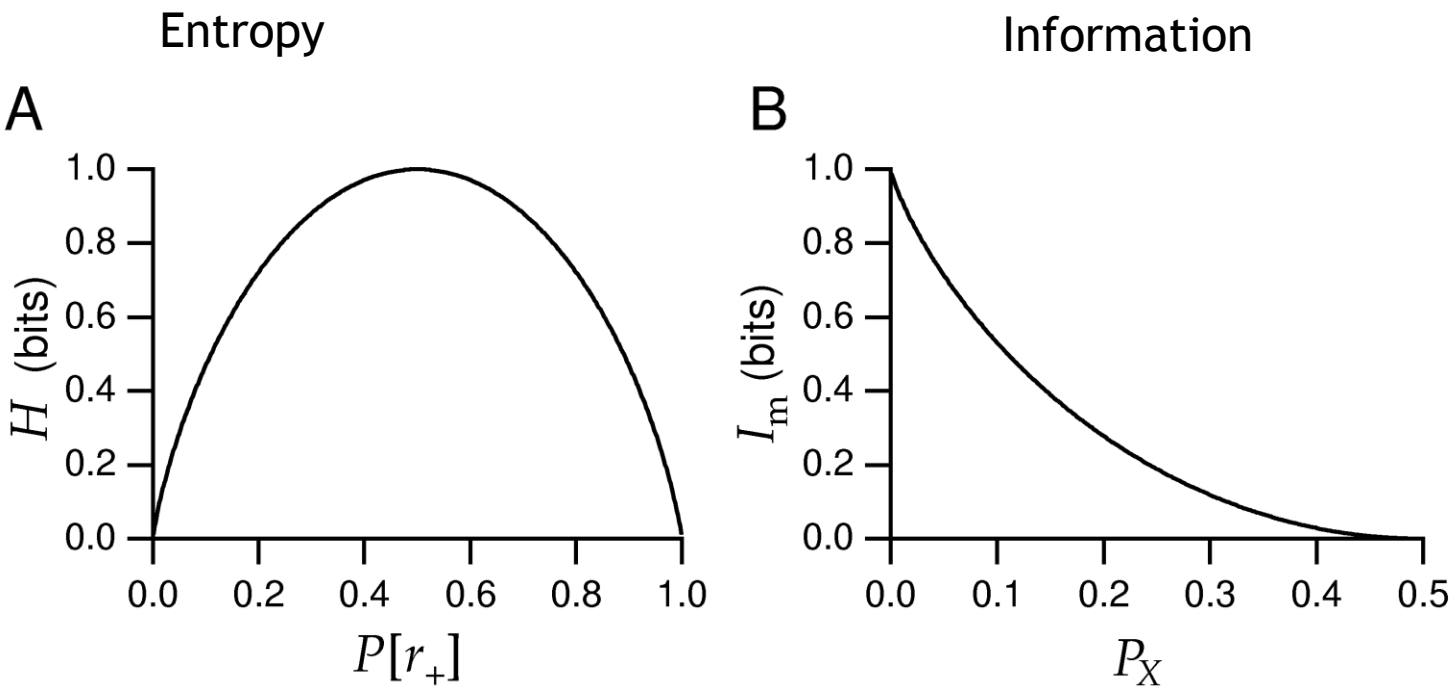
Examples:

response is unrelated to stimulus: $p[r|s] = p[r]$

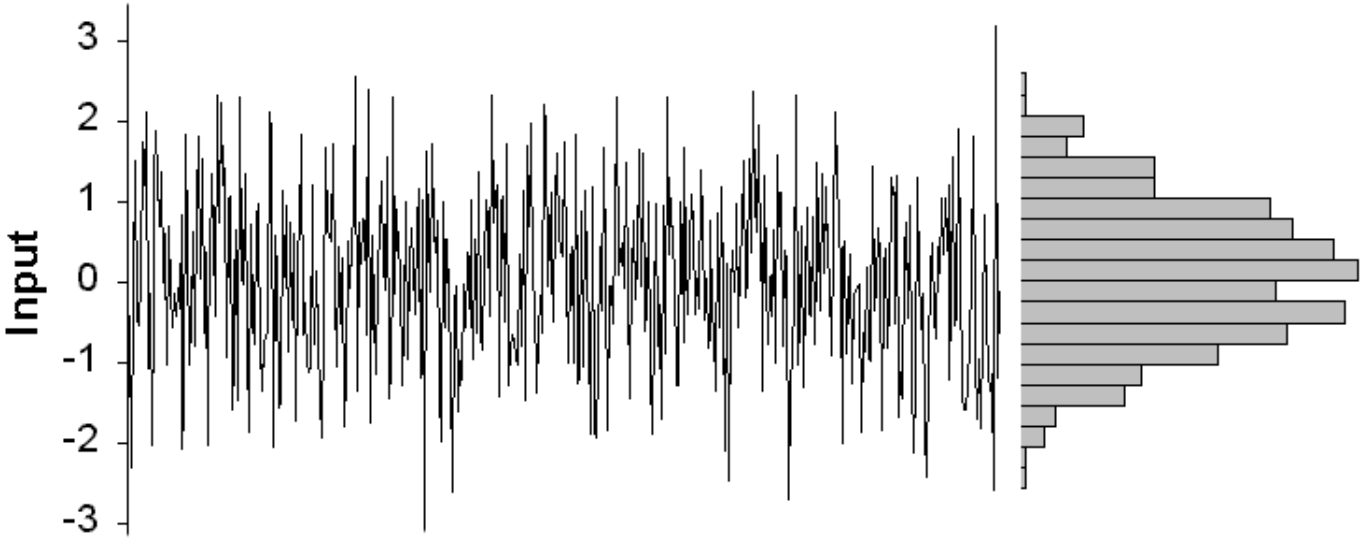
response is perfectly predicted by stimulus: $p[r|s] = \delta(r-r_s)$

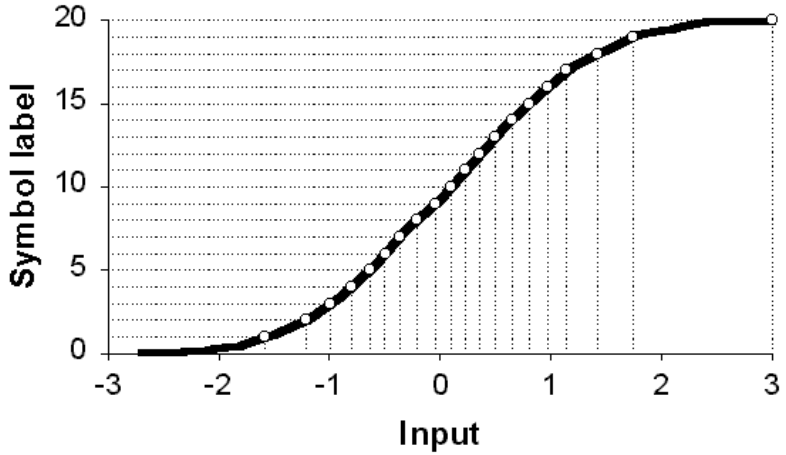
Maximising information

Binary distribution:



Efficient coding





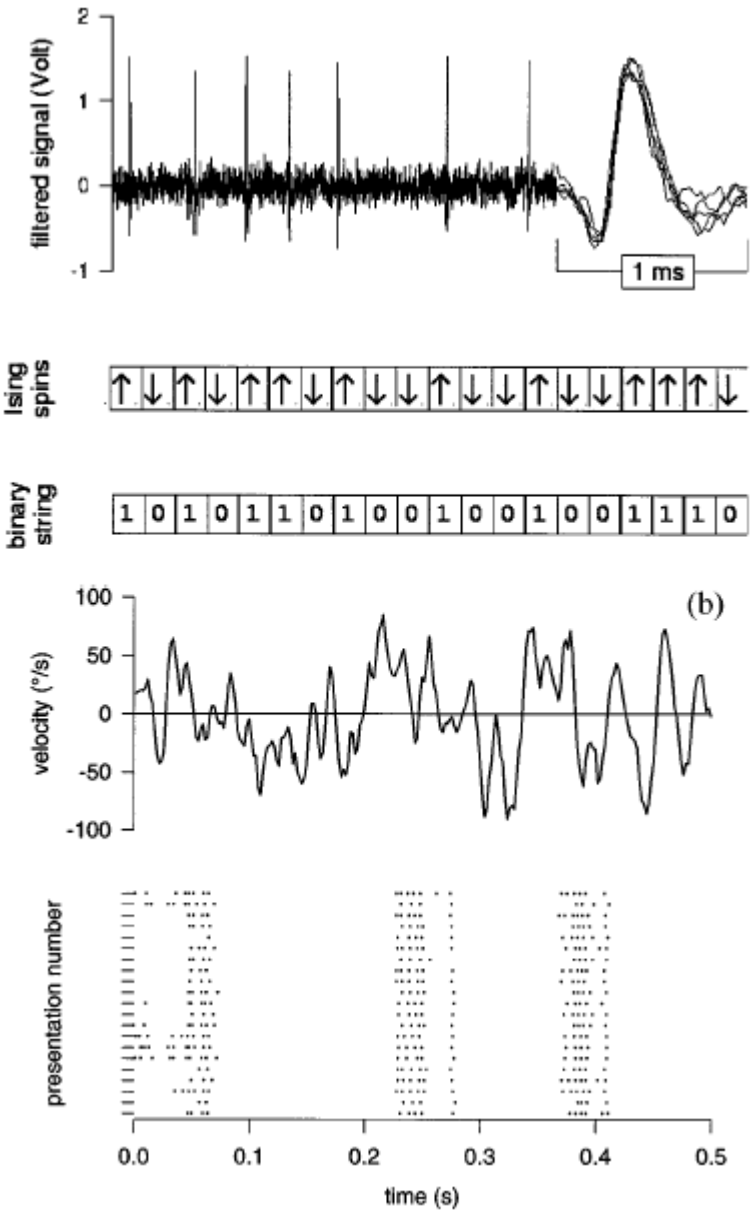
Information in single cells: the direct method

How can one compute the entropy and information of spike trains?

Discretize the spike train into binary words *w* with letter size $\Delta\tau$, length T. This takes into account correlations between spikes on timescales $T\Delta\tau$.

Compute $p_i = p(w_i)$, then the naïve entropy is

$$S_{\text{naive}}(T, \Delta\tau; \text{size}) = -\sum_i \tilde{p}_i \log_2 \tilde{p}_i \; ;$$



Strong et al., 1997; Panzeri et al.

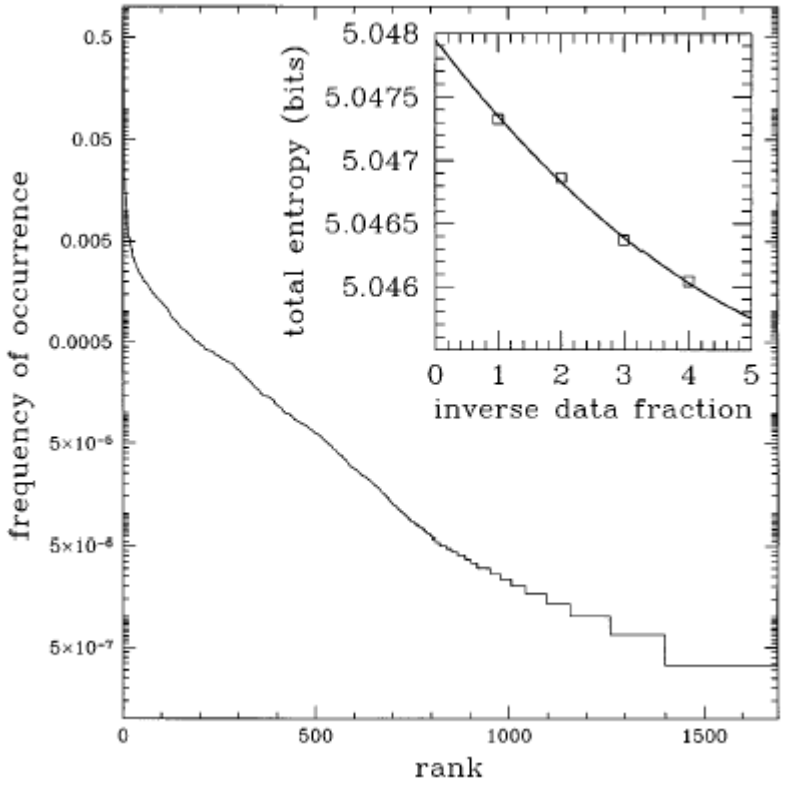
Information in single cells: the direct method

Many information calculations are limited by sampling: hard to determine P(w) and P(w|s)

Systematic bias from undersampling.

Correction for finite size effects:

$$S_{\mathrm{naive}}(T, \Delta\tau; \mathrm{size}) = S(T, \Delta\tau) + \frac{S_1(T, \Delta\tau)}{\mathrm{size}} + \frac{S_2(T, \Delta\tau)}{\mathrm{size}^2}.$$



Strong et al., 1997

Information in single cells: the direct method

Information is the difference between the variability driven by stimuli and that due to noise.

Take a stimulus sequence *s* and repeat many times.

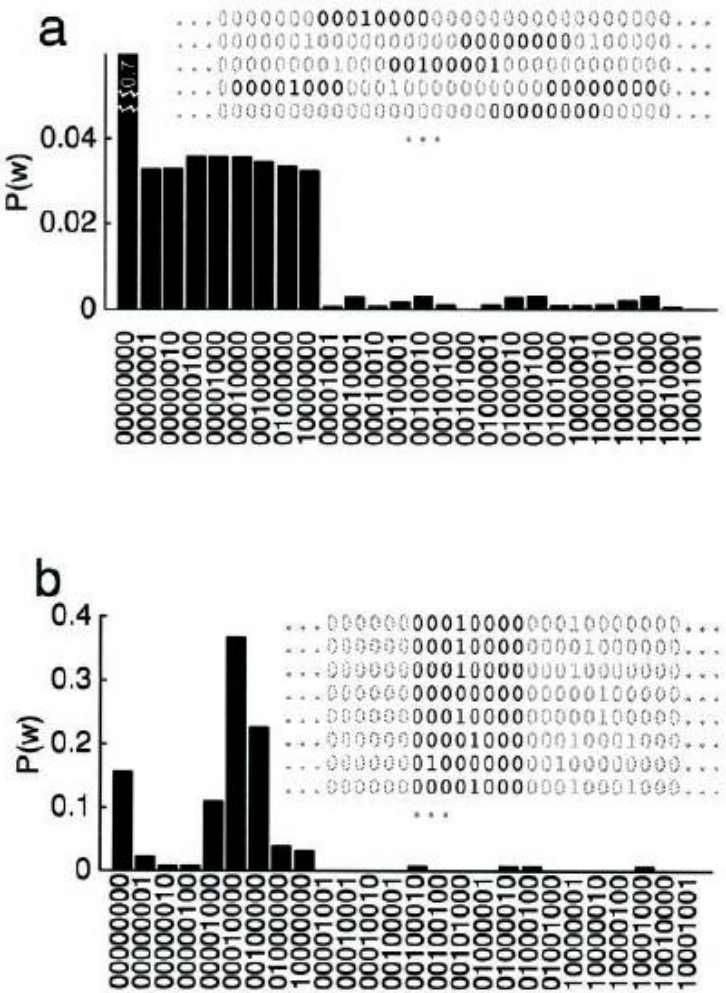
For each time in the repeated stimulus, get a set of words P(*w*|*s*(t)).

Should average over all *s* with weight P(*s*); instead, average over time:

$$H_{noise} = \langle H[P(w|s_i)] \rangle_i .$$

Choose length of repeated sequence long enough to sample the noise entropy adequately.

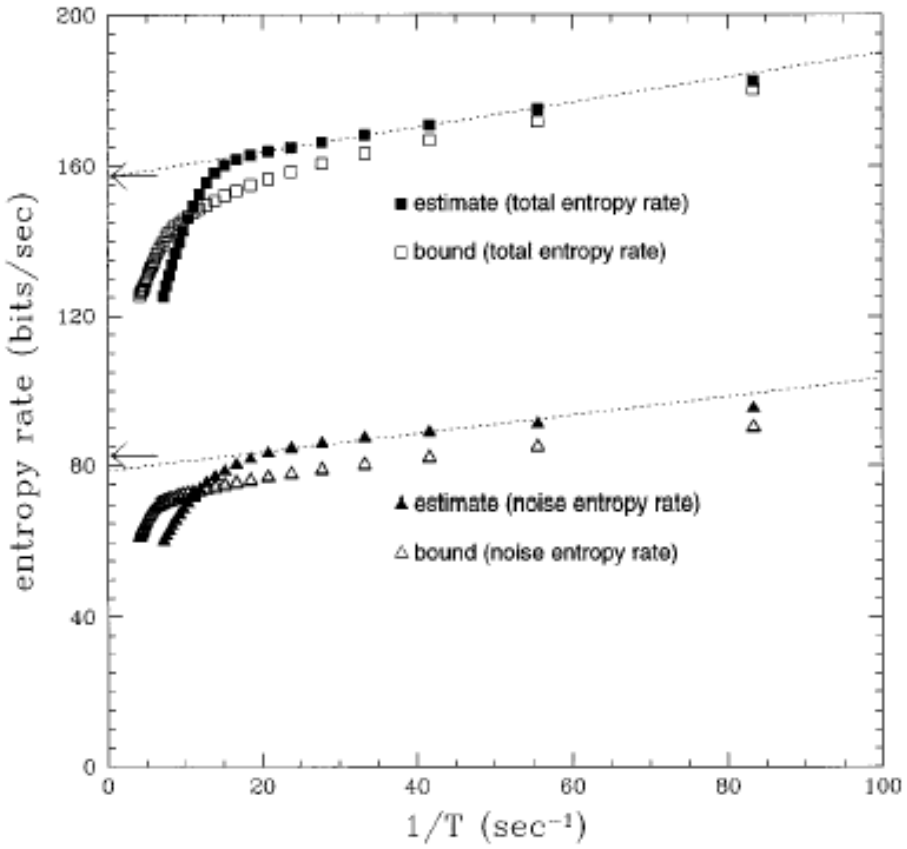
Finally, do as a function of word length T and extrapolate to infinite T.



Reinagel and Reid, '00

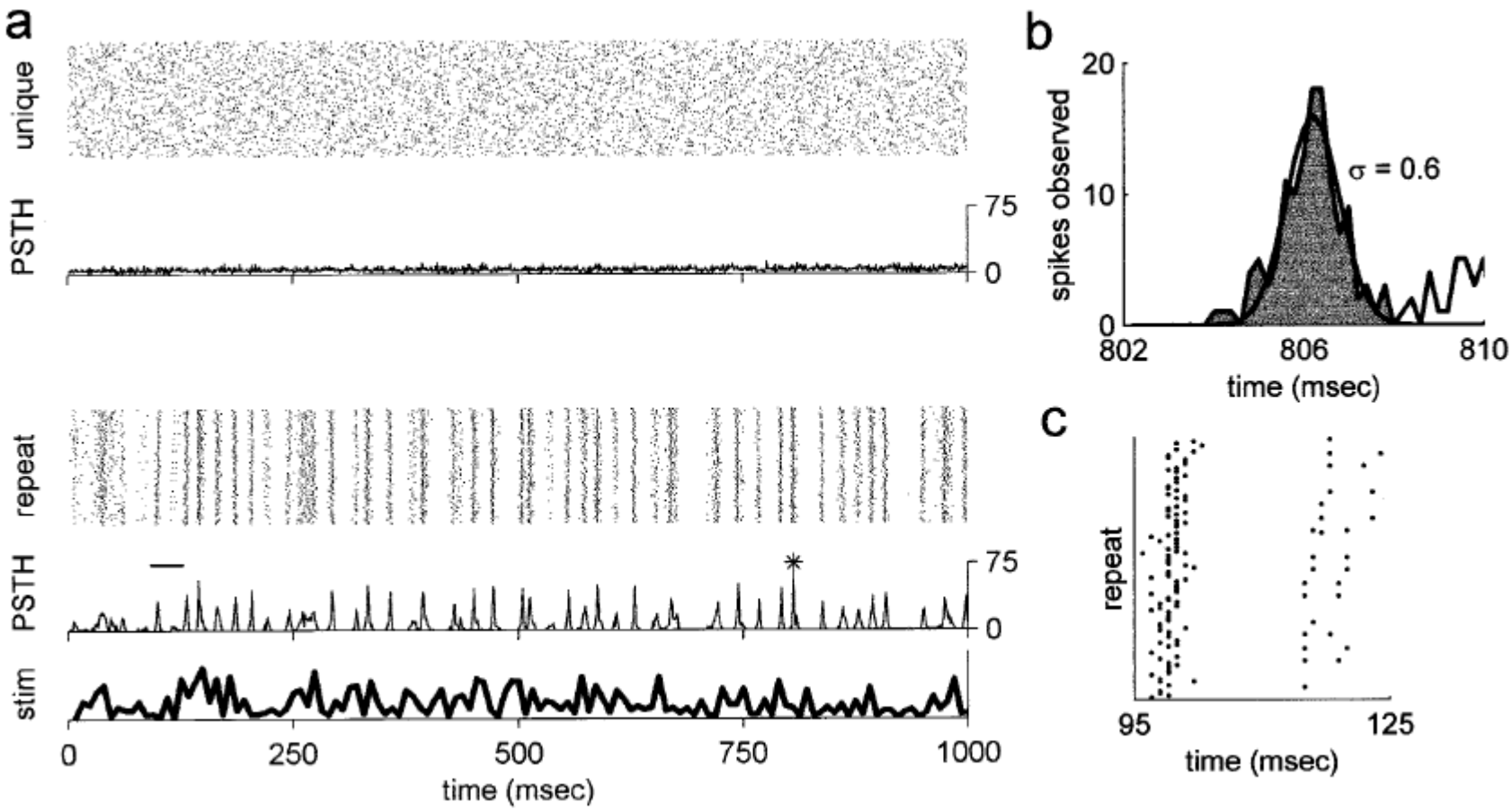
Information in single cells

Fly H1:
obtain information rate of
~80 bits/sec or 1-2 bits/spike.

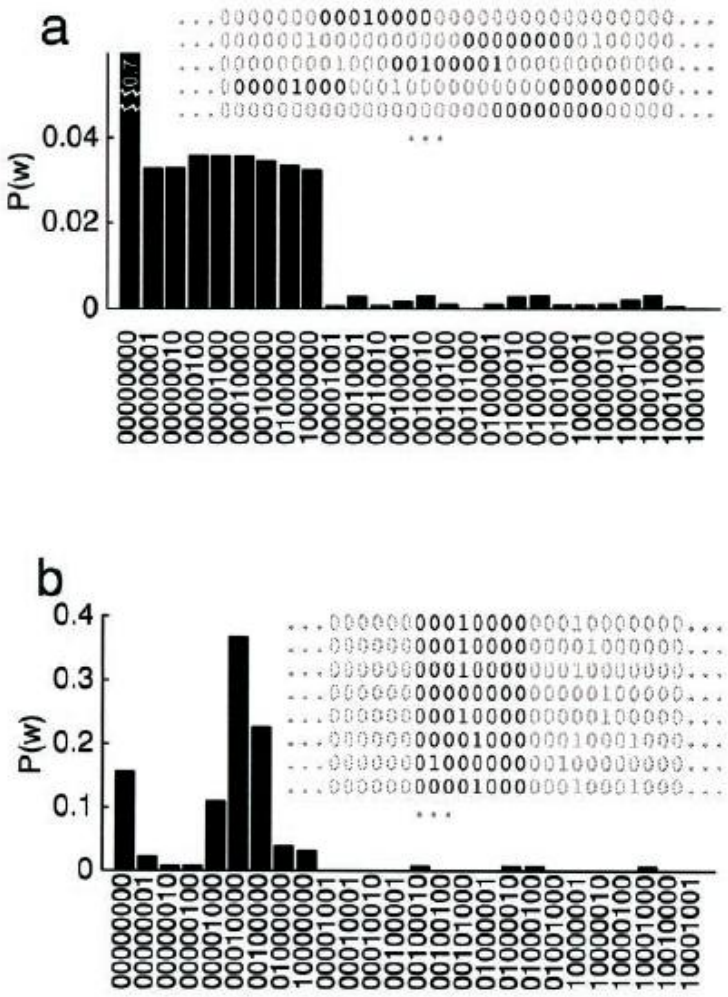


Information in single cells

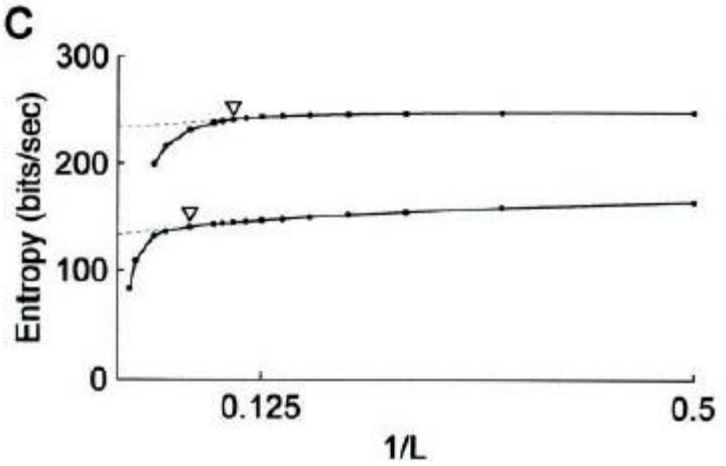
Another example: temporal coding in the LGN (Reinagel and Reid '00)



Information in single cells

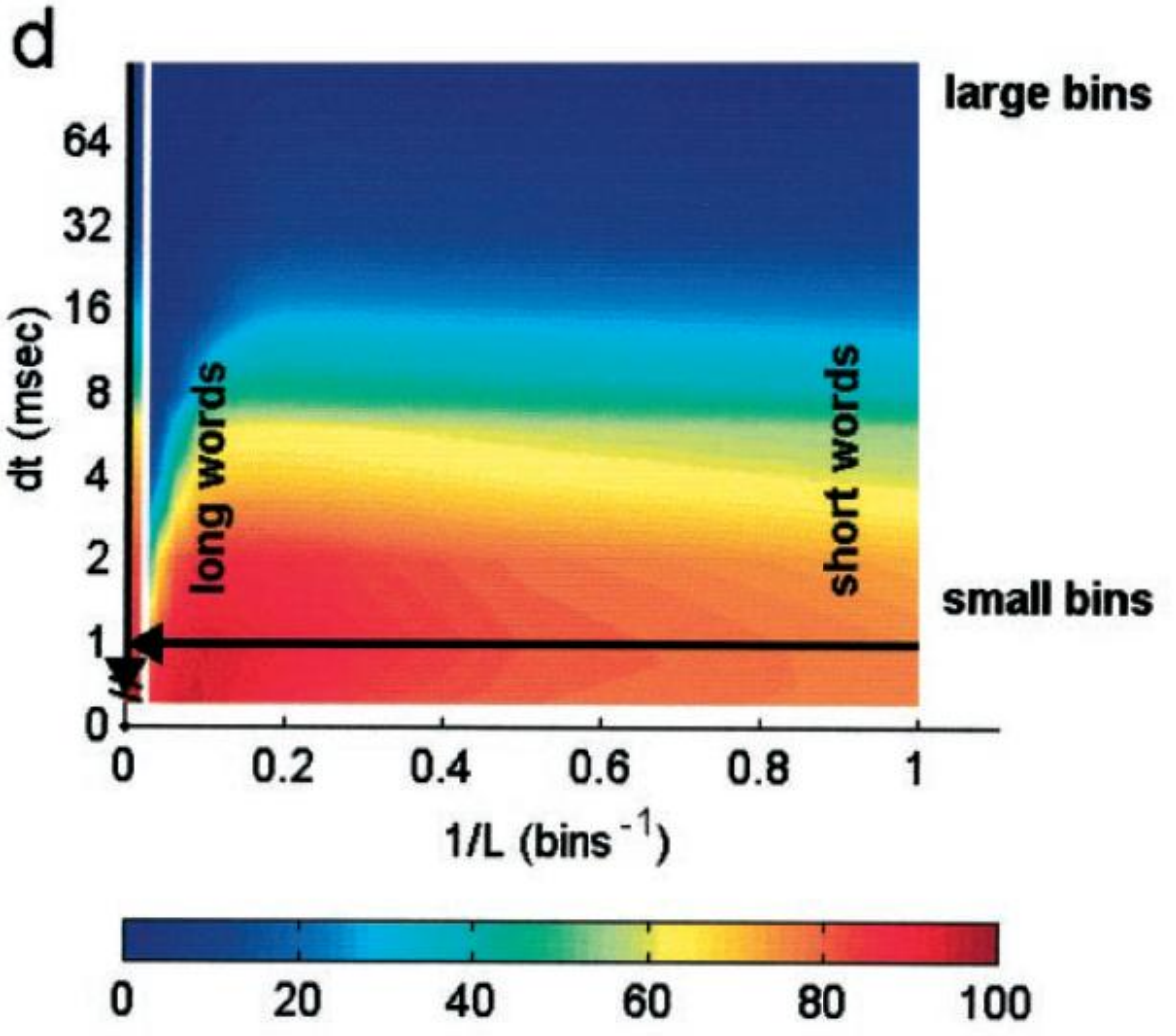


Apply the same procedure:
collect word distributions
for a random, then repeated stimulus.



Information in single cells

Use this to quantify how precise the code is,
and over what timescales correlations are important.



Information in single spikes

How much information does a single spike convey about the stimulus?

Key idea: the information that a spike gives about the stimulus is the reduction in entropy between the distribution of spike times not knowing the stimulus, and the distribution of times knowing the stimulus.

The response to an (arbitrary) stimulus sequence ***s*** is r(t).

Without knowing that the stimulus was ***s***, the probability of observing a spike in a given bin is proportional to $\bar{r}$, the mean rate, and the size of the bin.

Consider a bin Δt small enough that it can only contain a single spike. Then in the bin at time t,

$$\begin{aligned} P(r=1) &= \bar{r}\Delta t, \\ P(r=0) &= 1-\bar{r}\Delta t, \\ P(r=1|s) &= r(t)\Delta t, \\ P(r=0|s) &= 1-r(t)\Delta t. \end{aligned}$$

Information in single spikes

Now compute the entropy difference: $p = \bar{r}\Delta t$, $p(t) = r(t)\Delta t$

$$I(r,s) = -p\log p - (1-p)\log(1-p) + \quad \leftarrow \text{prior}$$

$$+\frac{1}{T}\int_0^T dt\ [p(t)\log p(t) + (1-p(t))\log(1-p(t))]\,. \quad \leftarrow \text{conditional}$$

Note substitution of a time average for an average over the *r* ensemble.

Assuming $p \ll 1, \log(1-p) \sim p$ and using $\frac{1}{T}\int_0^T dt\, p(t) \to p$

$$I(r,s) = \frac{1}{T}\int_0^T dt\, \Delta t\, r(t)\log\frac{r(t)}{\bar{r}} + Var(p(t))/2ln2 + O(p^3).$$

In terms of information per spike (divide by $\bar{r}\Delta t$):

$$I(r,s) = \frac{1}{T}\int_0^T dt\, \frac{r(t)}{\bar{r}}\log\frac{r(t)}{\bar{r}}$$

Information in single spikes

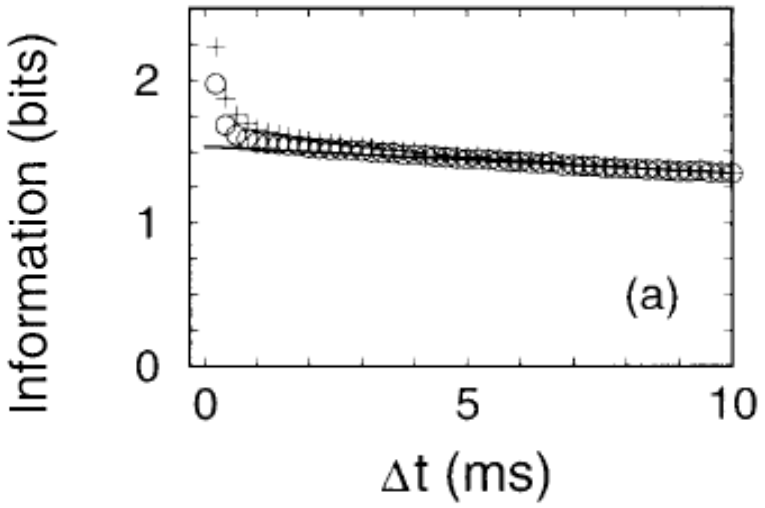
Given $$I(r,s) = \frac{1}{T}\int_0^T dt\, \frac{r(t)}{\bar{r}} \log \frac{r(t)}{\bar{r}}$$

note that:

- It doesn't depend explicitly on the stimulus
- The rate r does not have to mean rate of spikes; rate of any event.
- Information is limited by spike precision, which blurs r(t),
  and the mean spike rate.

Compute as a function of Δt:

Undersampled for small bins



Synergy and redundancy

How important is information in multispike patterns?

The information in any given event can be computed as:

$$I(E;s) = \left\langle \left(\frac{r_E(t)}{\bar{r}_E} \right) \log_2 \left(\frac{r_E(t)}{\bar{r}_E} \right) \right\rangle_s ,$$

Define the *synergy*, the information gained from the joint symbol:

$$\mathrm{Syn}[E_1, E_2; s] = I[E_1, E_2; s] - (I[E_1; s] + I[E_2; s]).$$

or equivalently,

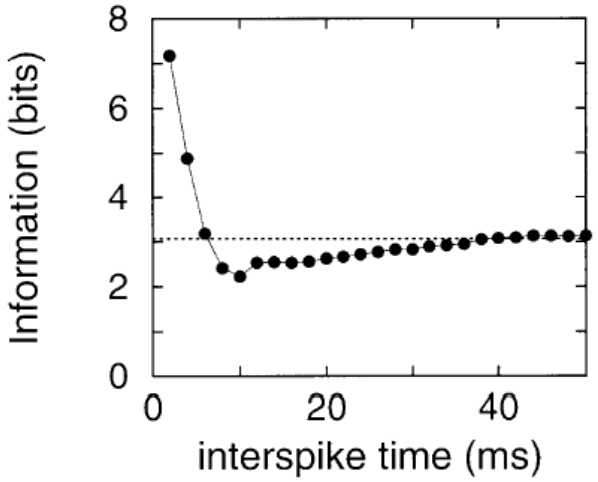
$$\mathrm{Syn}[E_1, E_2; s] = I[E_1; E_2|s] - I[E_1; E_2].$$

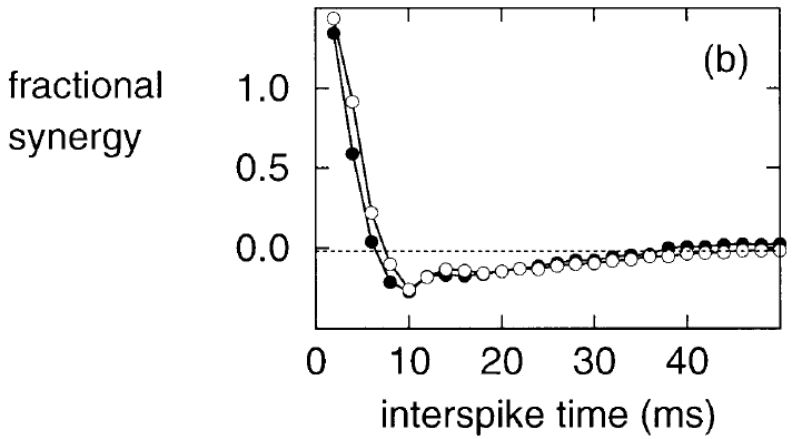
Negative synergy is called *redundancy*.

Brenner et al., '00.

Multispike patterns

In the identified neuron H1, compute information in a spike pair, separated by an interval dt:





$$\mathrm{Syn}[E_1, E_2; s] = I[E_1, E_2; s] - (I[E_1; s] + I[E_2; s]).$$

Brenner et al., ’00.

Using information to evaluate neural models

We can use the information about the stimulus to evaluate our reduced dimensionality models.

Using information to evaluate neural models

Information in timing of 1 spike:

$$I_{\text{one spike}} = \frac{1}{T}\int_0^T dt\, \frac{r(t)}{\bar{r}} \log_2 \left[\frac{r(t)}{\bar{r}}\right]$$

$$\frac{r(t)}{\bar{r}} = \frac{P(\text{spike at } t|\mathbf{s})}{P(\text{spike at } t)}$$

By definition

Using information to evaluate neural models

Given:

$$I_{\text{one spike}} = \frac{1}{T}\int_0^T dt\, \frac{r(t)}{\bar{r}} \log_2 \left[\frac{r(t)}{\bar{r}}\right]$$

$$\frac{r(t)}{\bar{r}} = \frac{P(\text{spike at } t|\mathbf{s})}{P(\text{spike at } t)} = \frac{P(\mathbf{s}|\text{spike at } t)}{P(\mathbf{s})}$$

By definition

Bayes' rule

Using information to evaluate neural models

Given: $$I_{\text{one spike}} = \frac{1}{T}\int_0^T dt\, \frac{r(t)}{\bar{r}} \log_2\left[\frac{r(t)}{\bar{r}}\right]$$

$$\frac{r(t)}{\bar{r}} = \frac{P(\text{spike at } t|\mathbf{s})}{P(\text{spike at } t)} = \frac{P(\mathbf{s}|\text{spike at } t)}{P(\mathbf{s})} \rightarrow \frac{P(s_1, s_2, s_3, ...|\text{spike at } t)}{P(s_1, s_2, s_3, ...)}$$

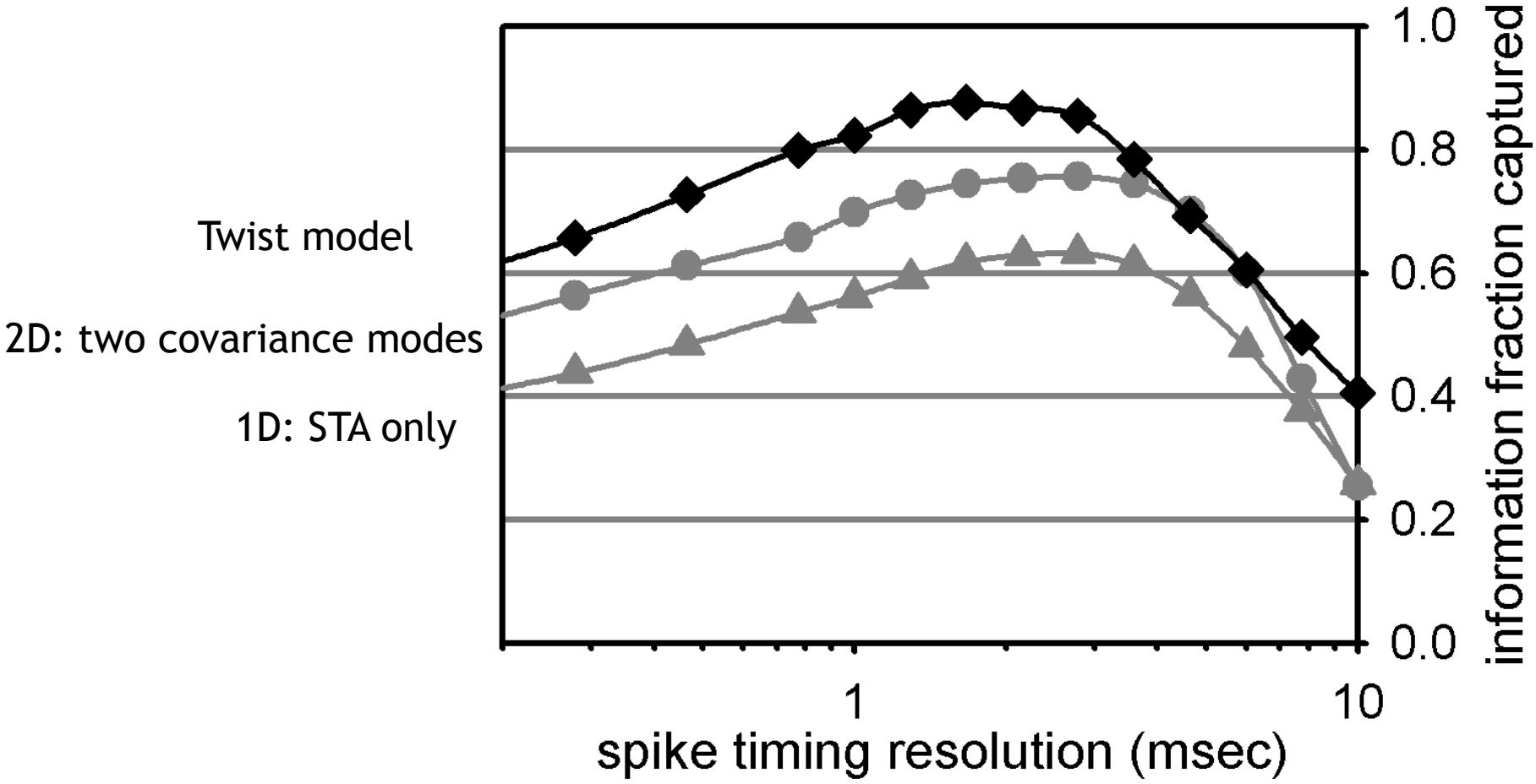
By definition Bayes' rule Dimensionality reduction

So the information in the K-dimensional model is evaluated using the distribution of projections:

$$I^K_{\text{one spike}} = \int d^K s\, P(s_1, ..., s_K|\text{spike at } t) \log_2\left[\frac{P(s_1, ..., s_K|\text{spike at } t)}{P(s_1, ..., s_K)}\right].$$

Using information to evaluate neural models

Here we used information to evaluate reduced models of the Hodgkin-Huxley neuron.



Adaptive coding

Just about every neuron adapts. Why?

- To stop the brain from tiring out
- To make better use of a limited dynamic range.
- To stop reporting already known facts

All reasonable ideas.

What does that mean for coding?
What part of the signal is the brain meant to read?

Adaptation can be mechanism for early sensory systems to make use of statistical Information about the environment.

How can the brain interpret an adaptive code?

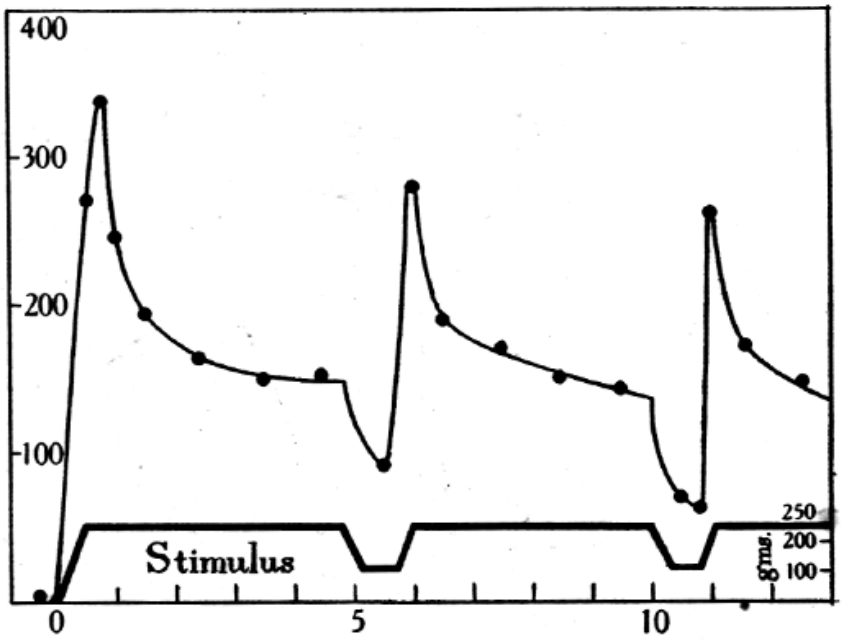


FIG. 29. IMPULSES PRODUCED BY PRESSURE ON CAT'S TOE-PAD. STIMULUS REMOVED PARTIALLY AND RENEWED.

From The Basis of Sensation, Adrian (1929)

Adaptation to stimulus statistics: efficient coding

Rate, or spike frequency adaptation is a classic form of adaptation.

Let's go back to the picture of neural computation we discussed before:
Can adapt both:

the system's *filters*
the *input/output relation* (threshold function)

Both are observed, and in both cases, the observed adaptations can be thought of as **increasing information transmission** through the system.

Information maximization as a principle of adaptive coding:

For optimum information transmission, coding strategy should adjust to the **statistics** of the inputs.

To compute the best strategy, have to impose constraints (Stemmler&Koch)

e.g. the variance of the output, or the maximum firing rate.

Adaptation of the input/output relation

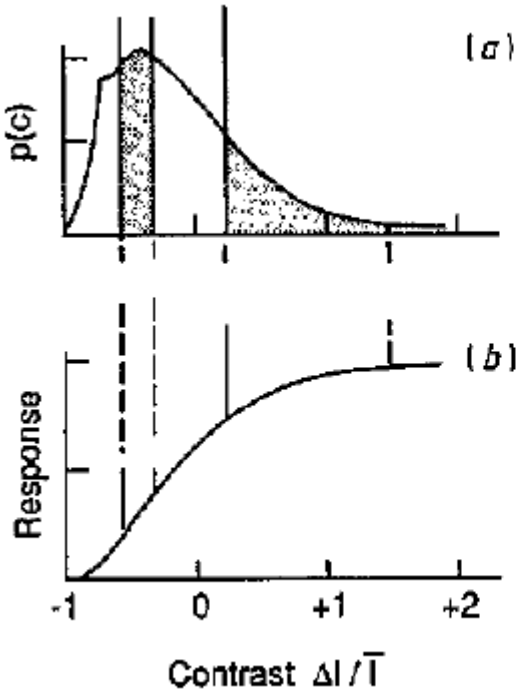
If we constrain the maximum, the solution for the distribution of output symbols is P(r) = constant = α.
Take the output to be a nonlinear transformation on the input: r = g(s).
From

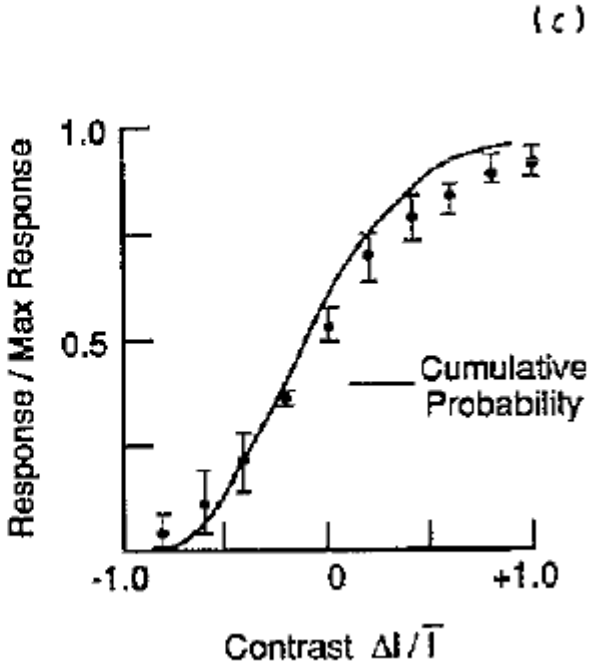
$$P(r)dr = P(s)ds$$

→ $$r = g(s) = \frac{1}{\alpha}\int_{-1}^{s} ds'\, P(s').$$

Fly LMC cells.

Measured contrast in natural scenes.





Laughlin ’81.

Dynamical adaptive coding

But is all adaptation to statistics on an evolutionary scale?

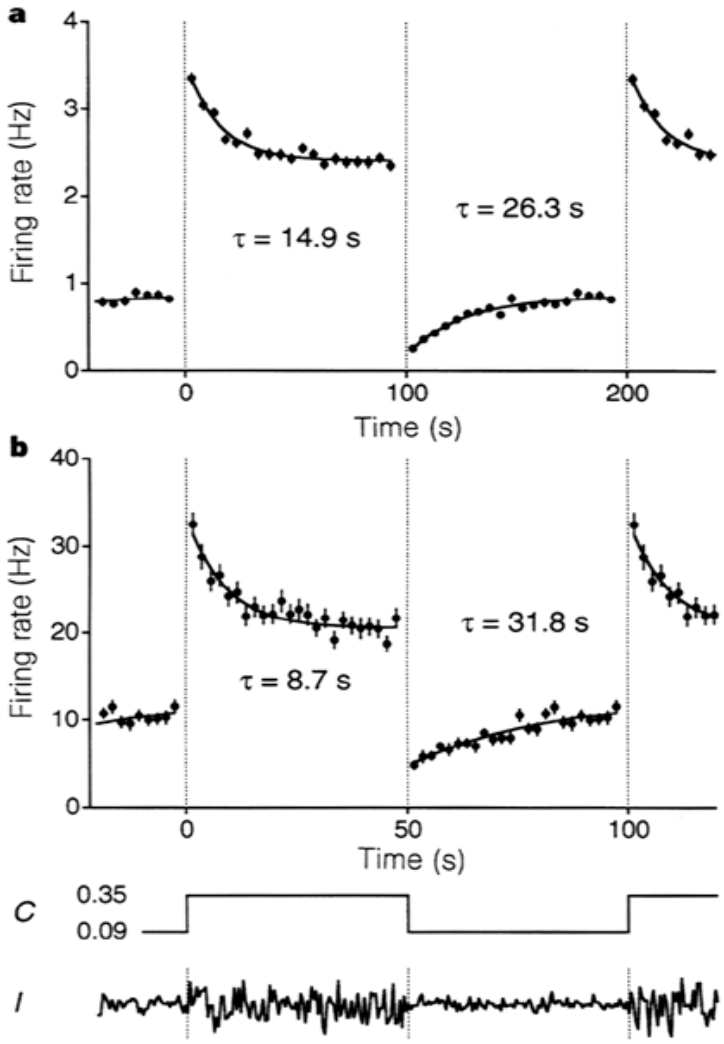
The world is highly fluctuating.
Light intensities vary by 10^{10} over a day.

Expect adaptation to statistics to happen dynamically, in real time.

Retina: observe adaptation to variance, or contrast, over 10s of seconds.

Surprisingly slow: contrast gain control effects after 100s of milliseconds.

Also observed adaptation to spatial scale on a similar timescale.

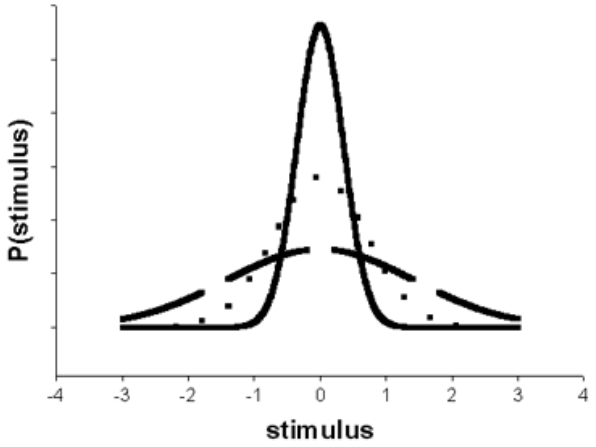


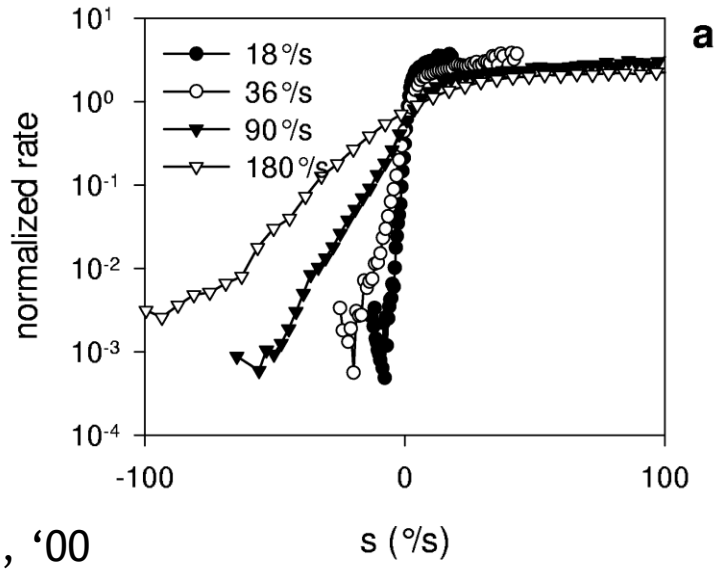
Smirnakis et al., ‘97

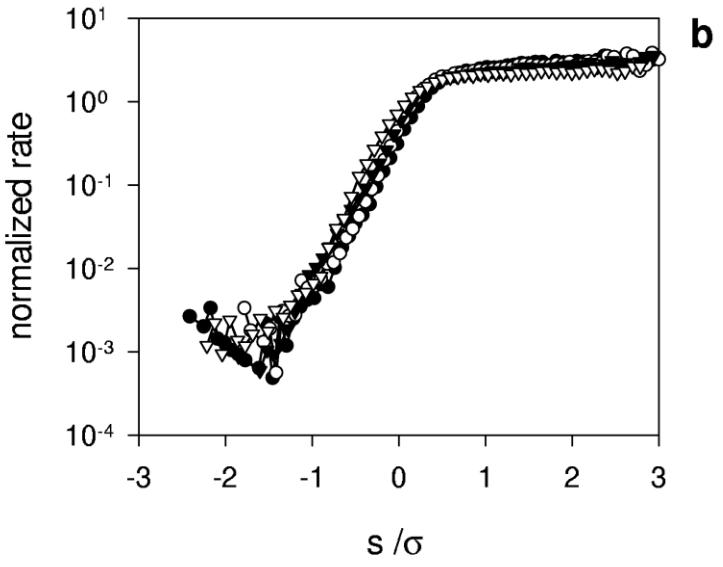
Dynamical adaptive coding

The H1 neuron of the fly visual system.

Rescales input/output relation with steady state stimulus statistics.



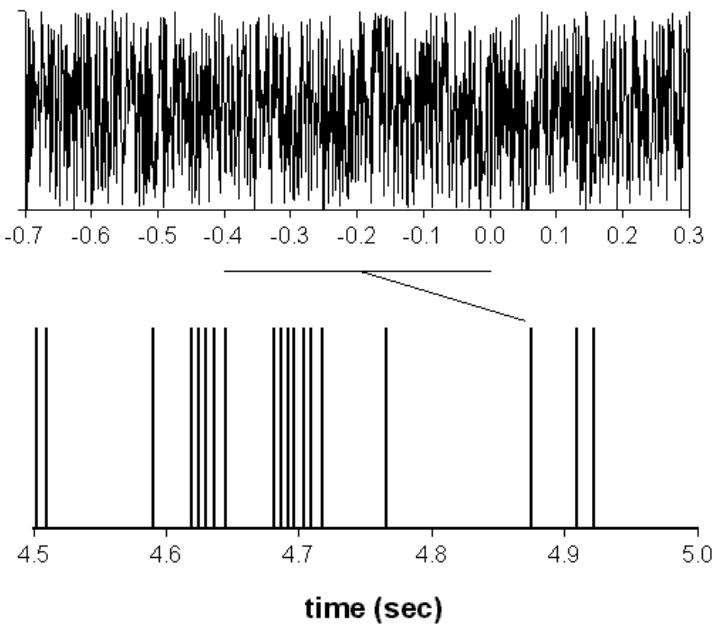


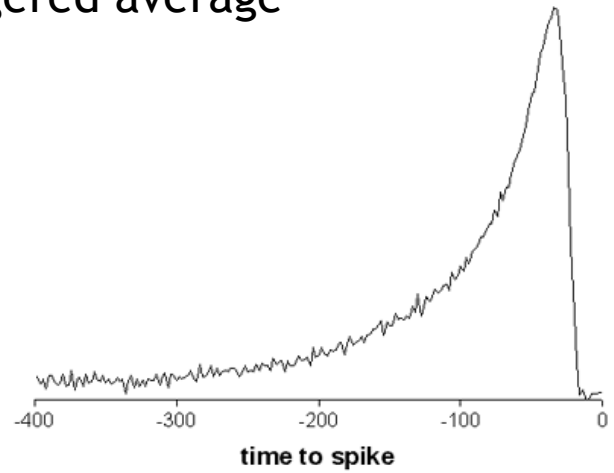


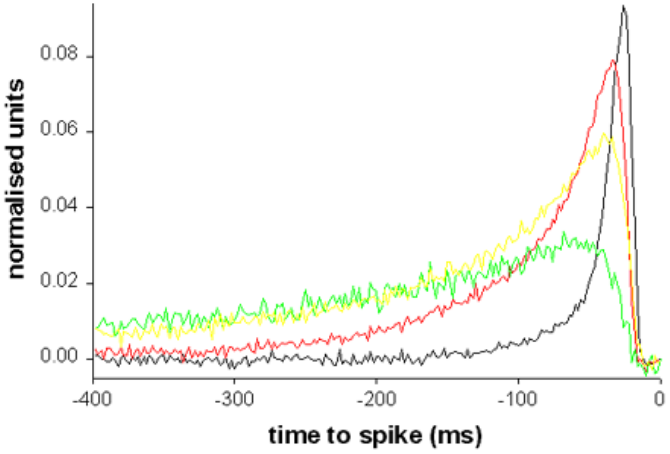
Brenner et al., ‘00

Dynamical adaptive coding

As we have seen already, extract the spike-triggered average







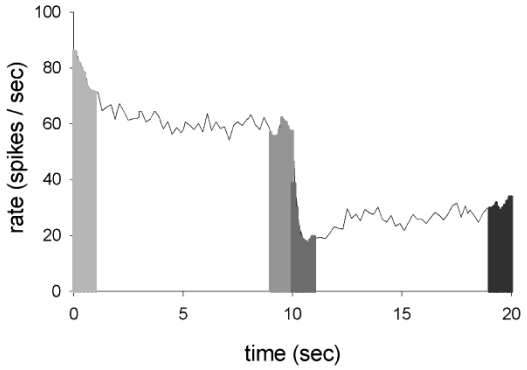
Dynamical adaptive coding

Compute the input/output relations, as we described before:

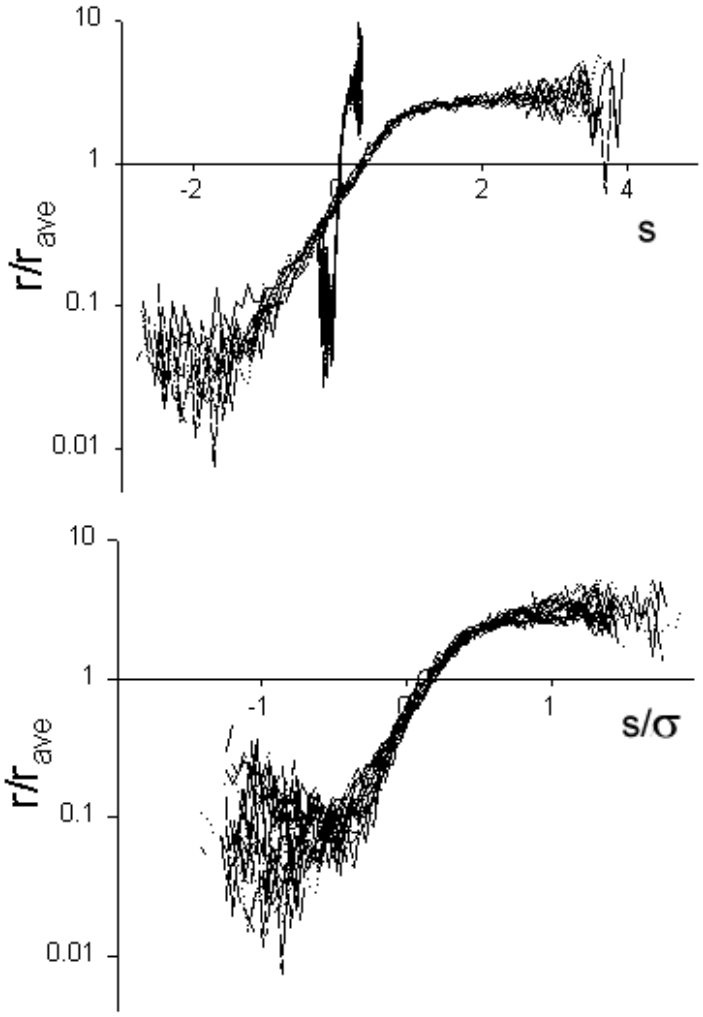
s = stim . STA;

$P(spike|s) = r_{ave}\, P(stim|s) / P(s)$

Do it at different times in variance modulation cycle.



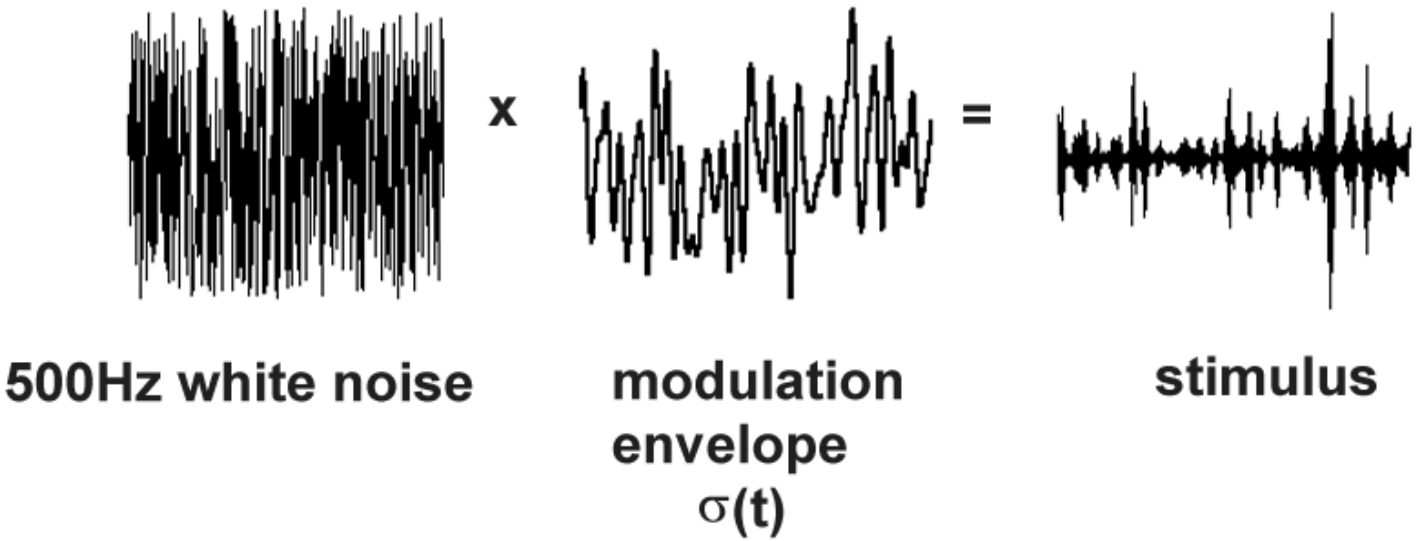
Find ongoing normalisation with respect to stimulus standard deviation



Dynamical adaptive coding: naturalistic stimuli

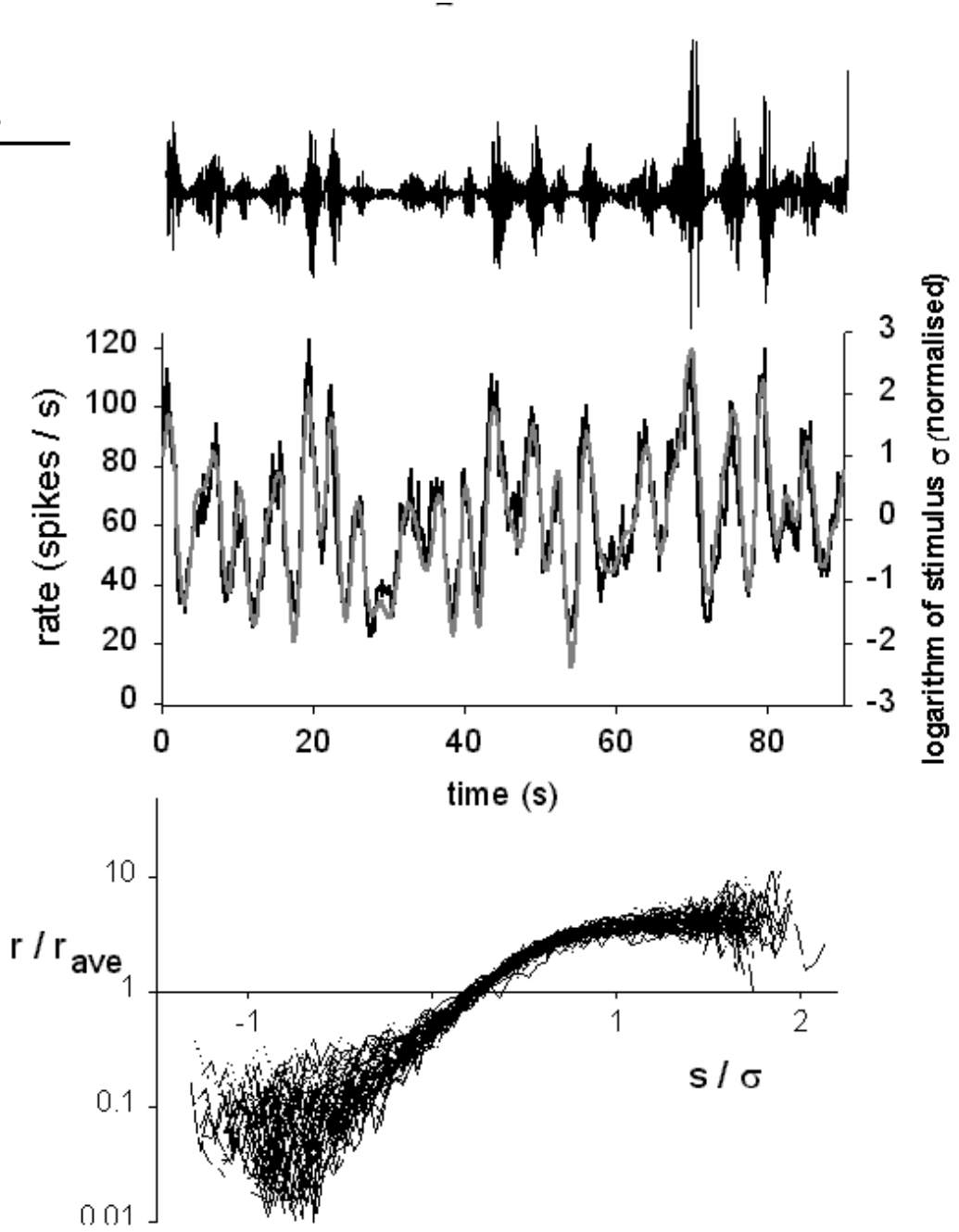
Take a more complex stimulus: randomly modulated white noise.

Not unlike natural stimuli (Ruderman and Bialek ’97)



Dynamical adaptive coding

Find continuous rescaling to variance envelope.



Dynamical information maximization

This should imply that information transmission is being maximized.
We can compute the information directly and observe the timescale.
How much information is available about the stimulus fluctuations?
Return to two-state switching experiment.

Method:

Present *n* different white noise sequences, randomly ordered, throughout the variance modulation.

Collect word responses indexed by time with respect to the cycle, P(w(t)).

Now divide according to probe identity, and compute

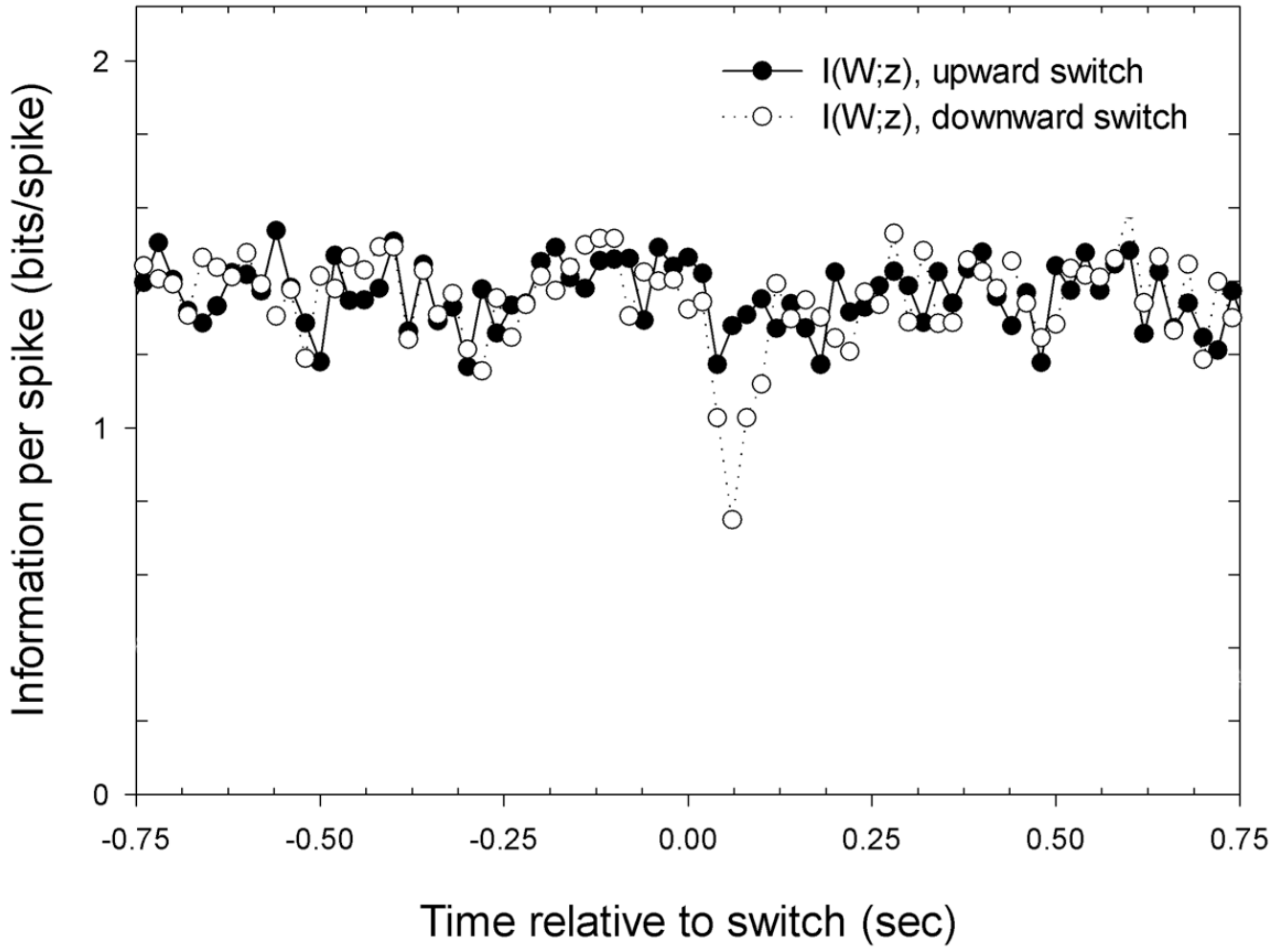
$$I_t(w;s) = H[P(w(t))] - \Sigma_i P(s_i)\, H[P(w(t)|s_i)] , \quad P(s_i) = 1/n;$$

Similarly, one can compute information about the variance:

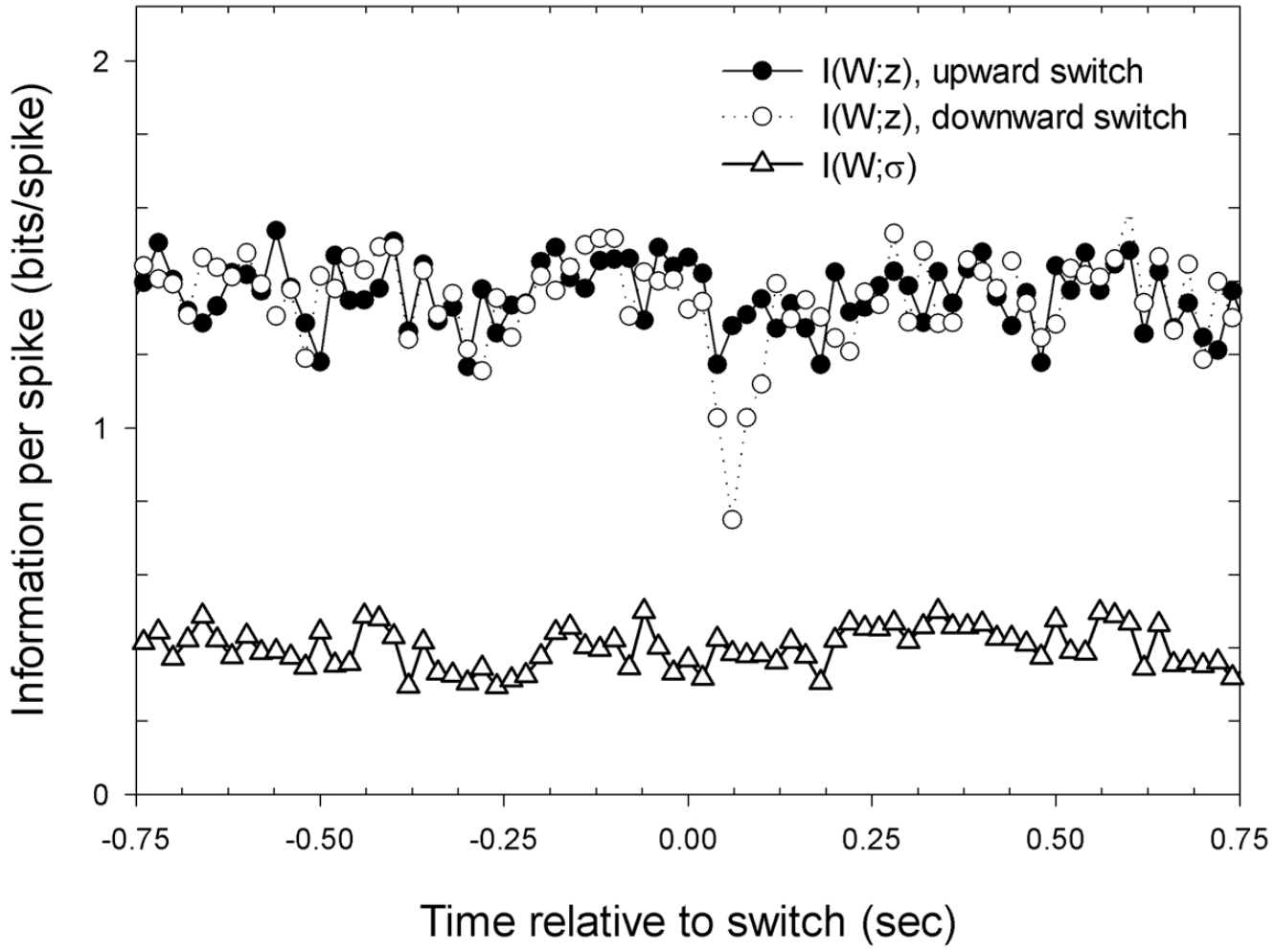
$$I_t(w;s) = H[P(w(t))] - \Sigma_i P(\sigma_i)\, H[P(w(t)|\sigma_i)] , \quad P(\sigma_i) = ½;$$

Convert to information/spike by dividing at each time by mean # of spikes.

Tracking information in time



Tracking information in time: the variance



Other fascinating subjects

Maximising information transmission through optimal filtering
→ predicting receptive fields

Populations: synergy and redundancy in codes involving many neurons
→ multi-information

Sparseness: the new efficient coding