Calculating information in spike trains

Two methods:

- Information in spike patterns
- Information in single spikes

Calculating mutual information

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

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

Grandma's famous mutual information recipe

Take one stimulus *s* and repeat many times to obtain $P(R|s)$.

Compute variability due to noise: *noise entropy* $H[R|s]$

Repeat for all s and average: $\Sigma_s P(s)\, H[R|s)]$.

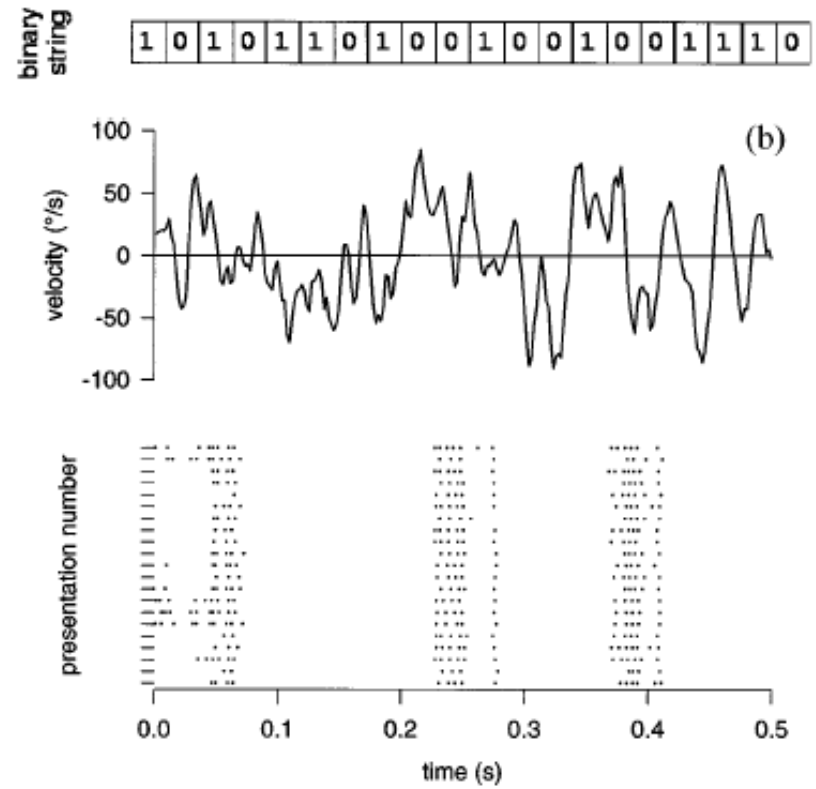
Compute $P(R) = \Sigma_s P(s)\, P(R|s)$ **and the total entropy** $H[R]$

Calculating information in spike patterns

So far only dealt with single spikes, or firing rates.

What information is carried by patterns of spikes?

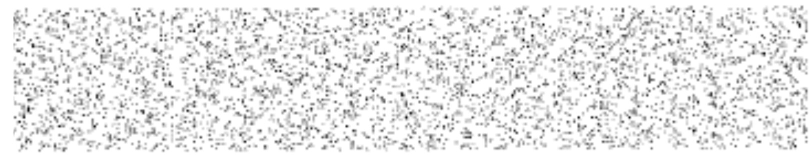
Analyze patterns of the code: how informative are they?



Strong et al., 1997

Calculating information in spike trains

Entropy:



- Binary words w with letter size Δt, length T.

- Compute $p(w_i)$

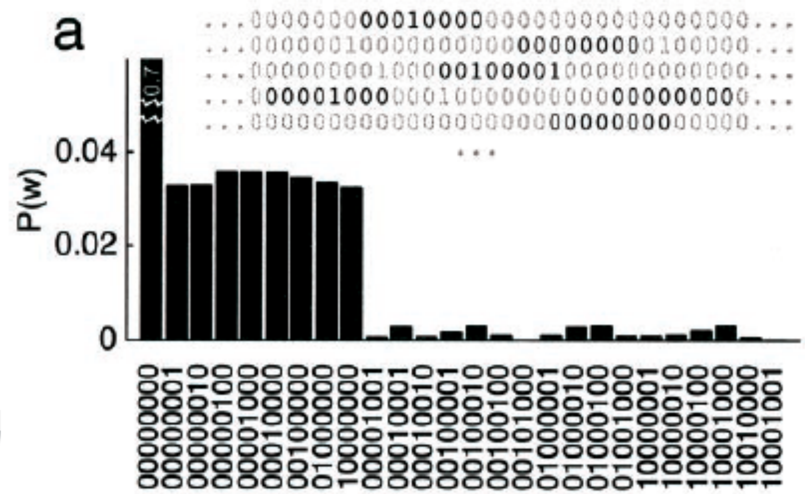
$$H[w] = -\sum p(w_i) \log_2 p(w_i)$$

以一个适当的window width(n) slide，count所有可能组合(2^n)

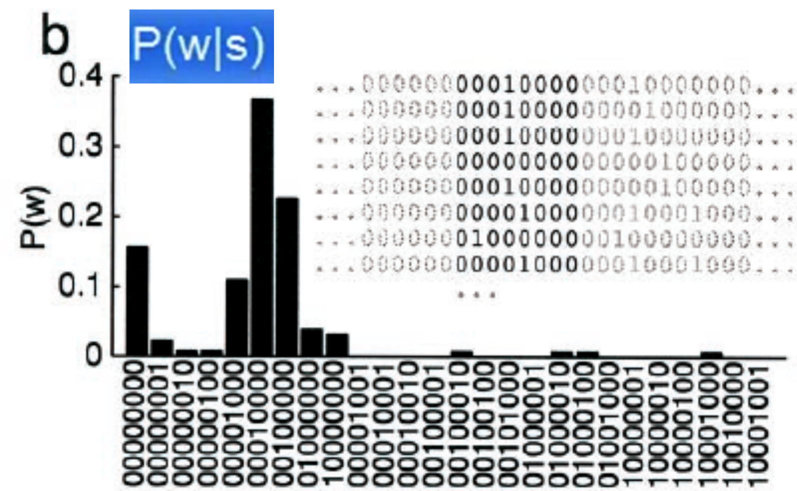
Strong et al., 1997; Reinagel and Reid, 2000

Calculating information in spike trains

Information :
difference between the total variability driven by stimuli and that due to noise, averaged over stimuli.



Here, the repeated stimulus is still random, but the same segment of random noise is repeated over and over.



Since we are talking about distributions of responses, we are really computing the reduction of uncertainty about responses due to knowing the stimulus;
we could have done this problem the other way around, by selecting each word, and looking for the distribution of stimuli that goes with that word.
Can you see practical reasons why this approach is simpler?

Reinagel and Reid, '00

Apply grandma's recipe

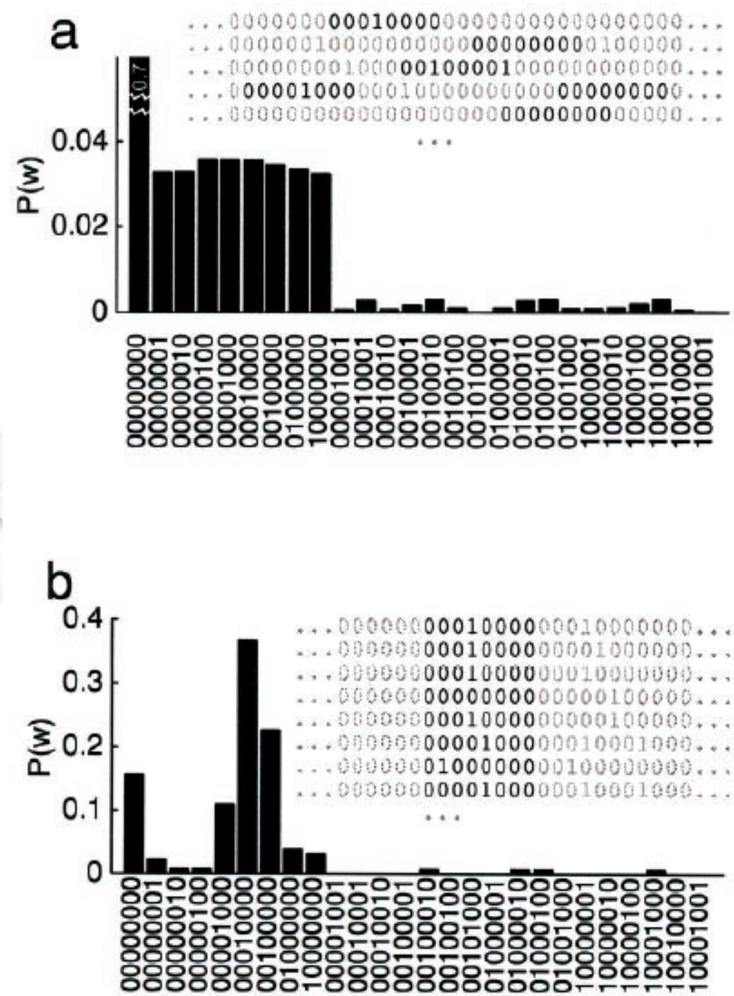
Take a stimulus sequence and repeat many times.

How to sample P(s)?
Average over s → average over time:
用随机的刺激，足够唱的时间，这样更好操作

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

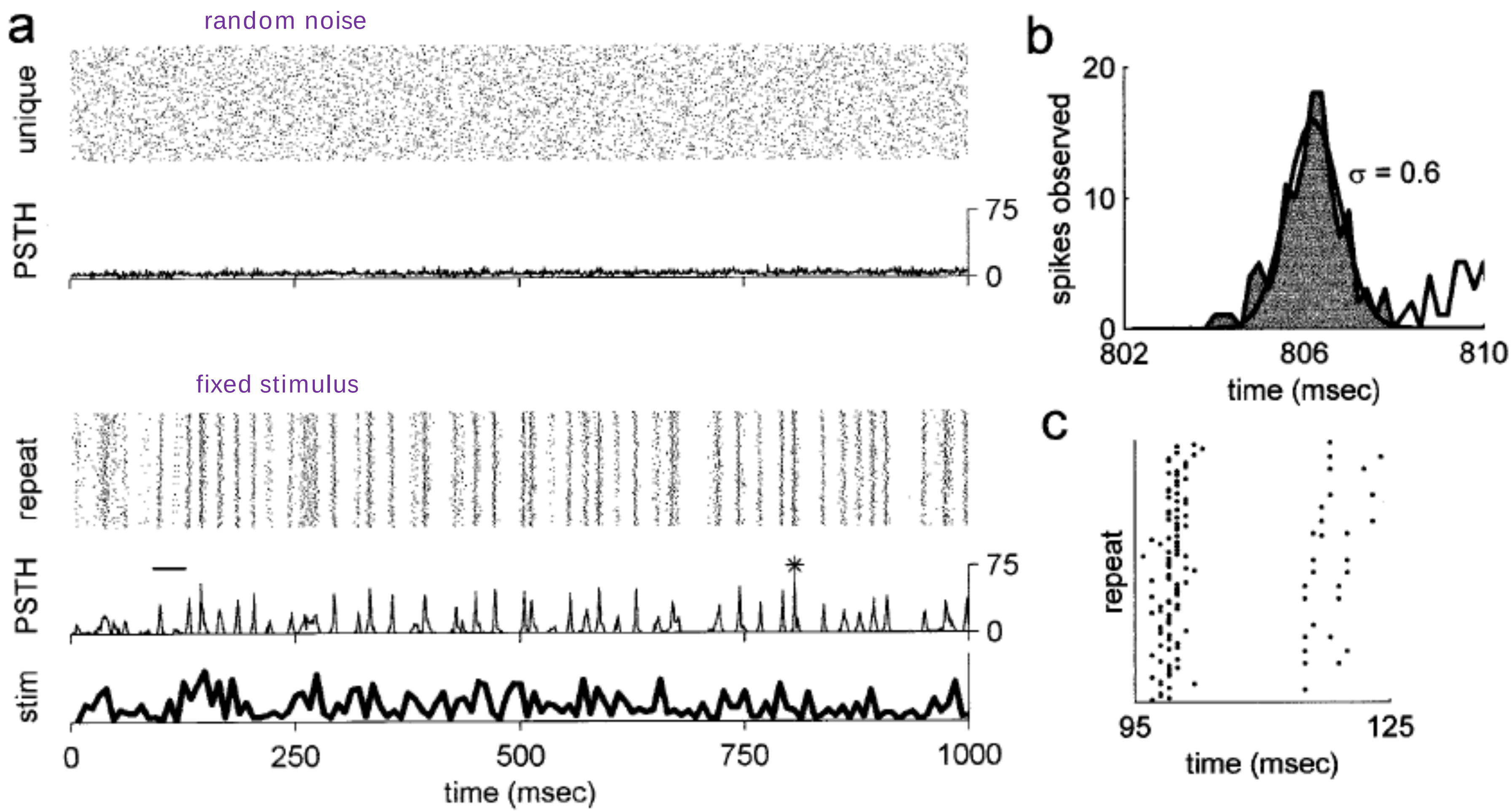
$$H_{noise} = \langle H[P(w|s_i)] \rangle_i.$$ time point i

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



Reinagel and Reid (2000)

Calculating information in the LGN



On what time scale, do these responses continue to convey information about the stimulus?

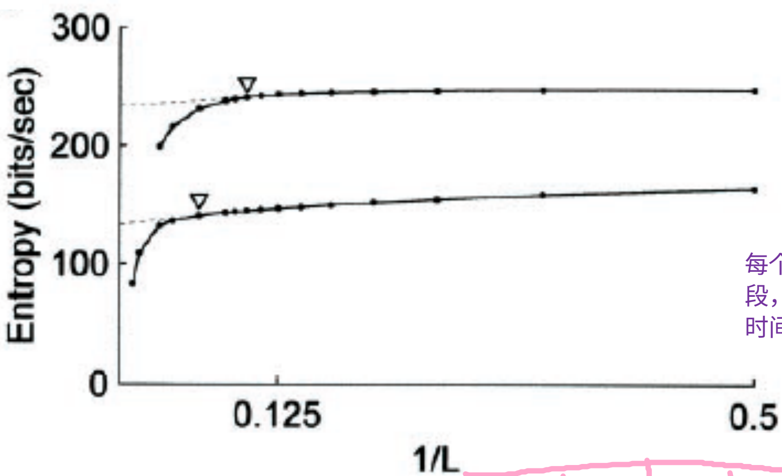
我们该以什么样的方式采样才能得到更多spike train的信息呢？
有两个变量可以操纵：
- pattern的长度
- 时间段

我们要在一个时间段，找pattern。然后得到count的分布

Reinagel and Reid (2000)

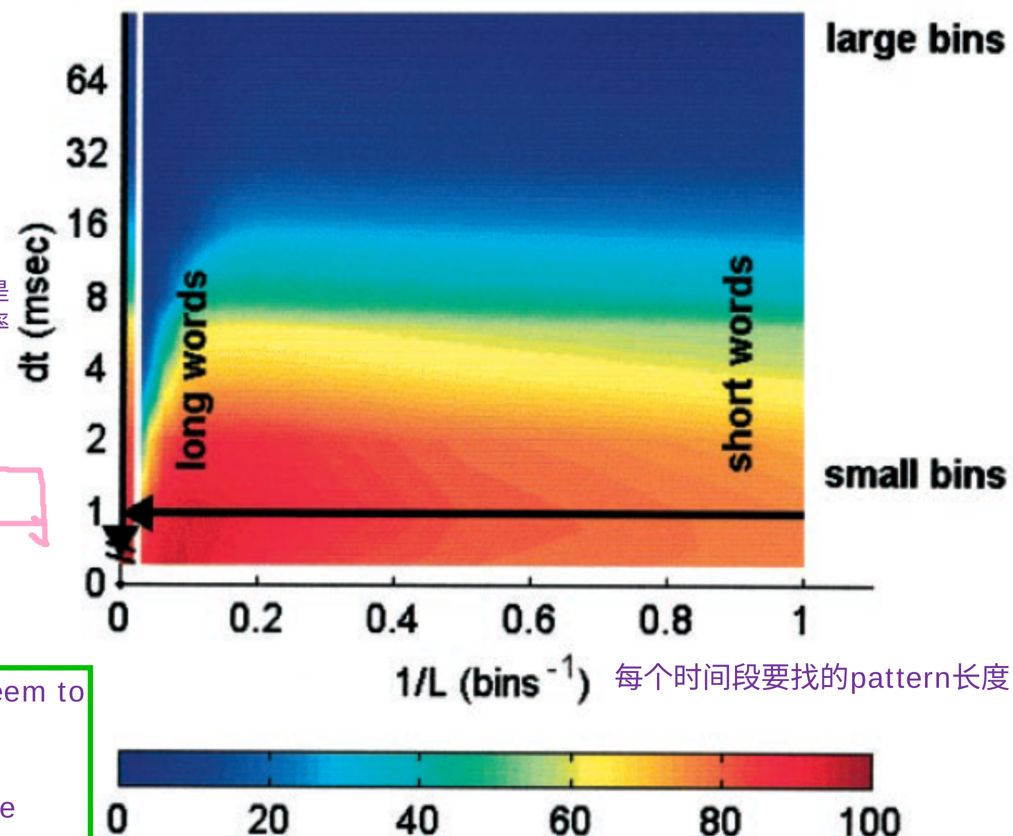
Learning about the LGN's code

如果采样太长，会存在问题。对于一个有限的spike train数据，你采样的长度越长。你有符合pattern的数据就越少。
So when one tries to estimate the entropy of the distribution of words of this length, it's very unlikely that you will have seen them all. And so not surprisingly, if you now look at the entropy, plotted as one over the word length. The entropy drops off at this limit indicating that the information is not completely sampled.



每个时间段，也就是时间分辨率

计算不同采样pattern长度的entropy



每个时间段要找的pattern长度

From the jitter in the spike train that we looked at, that they seem to be repeatable on a time scale of about 1 or 2 ms.

So that time scale dt corresponds to the time scale in which the jitter in the spike train. Still allows one to read that off as an encoding of the same stimulus.

It's going to quantify approximately what's the temporal with that one can discatize this spike train and still extract all the information about the stimulus that distinguishes it from other stimuli.

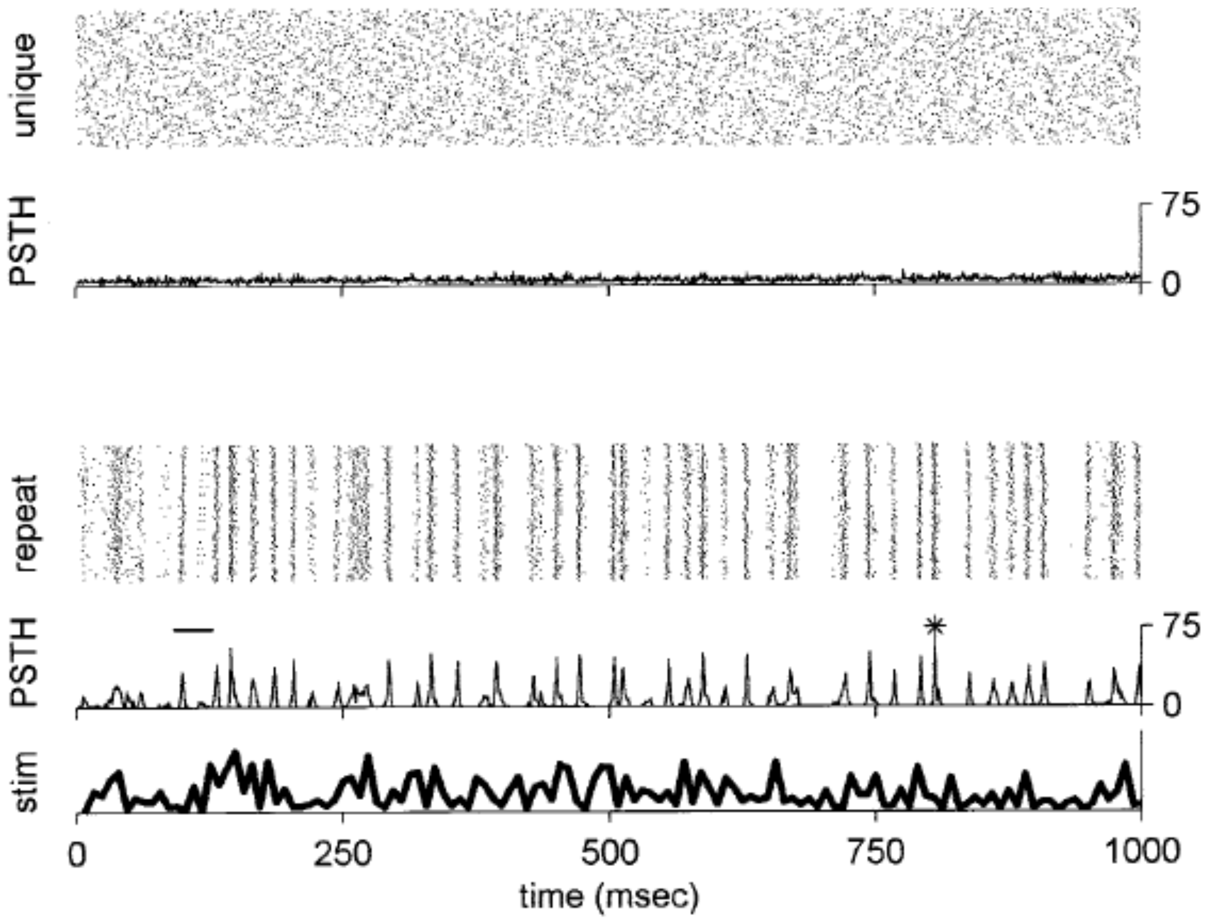
Reinagel and Reid (2000)

Sampling and bias

- Never enough data!
- Corrections for finite sample size
- Panzeri, Nemenman, ..

Information in single spikes

By how much does knowing that a particular stimulus occurred reduce the entropy of the response?

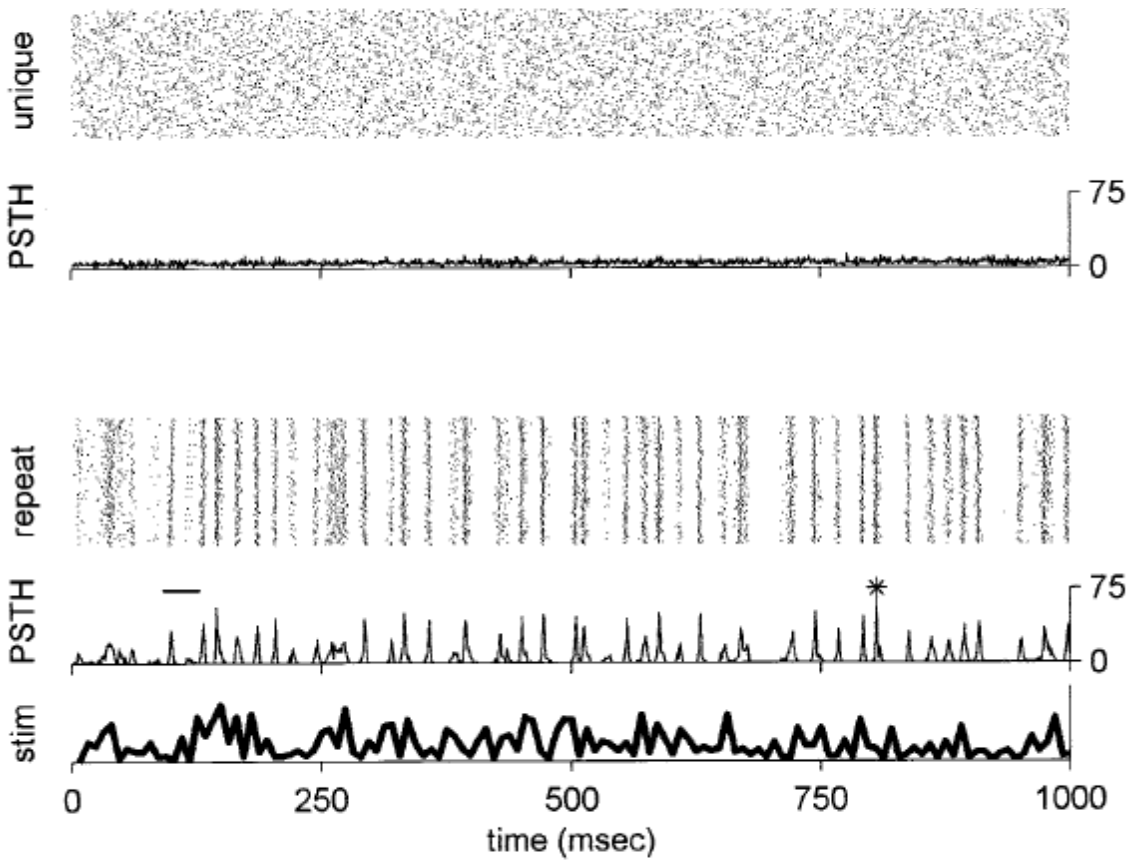


Brenner et al. (2000), data Reinagel and Reid (2000)

Information in single spikes

前一种是通过count pattern bin计算entropy
现在我们是通过firing rate计算entropy

我们不知道stimulus



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

Time varying firing rate caused by the changing stimulus

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

Information in single spikes

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

$$I(r,s) = -p\log p - (1-p)\log(1-p) +$$ ← Total

$$+\frac{1}{T}\int_0^T dt\,\left[p(t)\log p(t) + (1-p(t))\log(1-p(t))\right].$$ ← Noise

Every time *t* stands in for a sample of *s*

A time average is equivalent to averaging over the *s* ensemble.

Ergodicity

Information in single spikes

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

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

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

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

To get *information per spike*, divide by $\bar{r}\Delta t$:

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

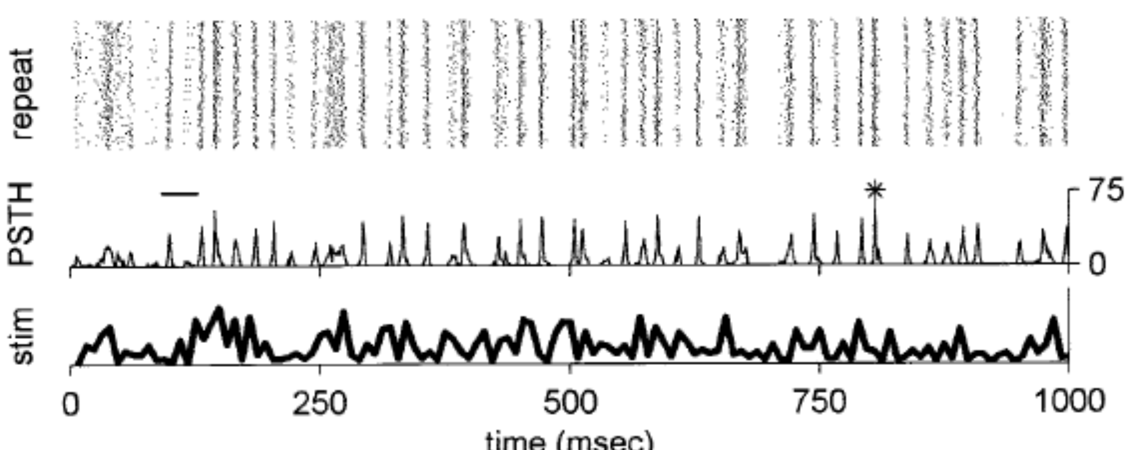
Information in single spikes

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

- No explicit stimulus dependence (no coding/decoding model)
- The rate *r* does not have to mean rate of spikes; rate of any event.

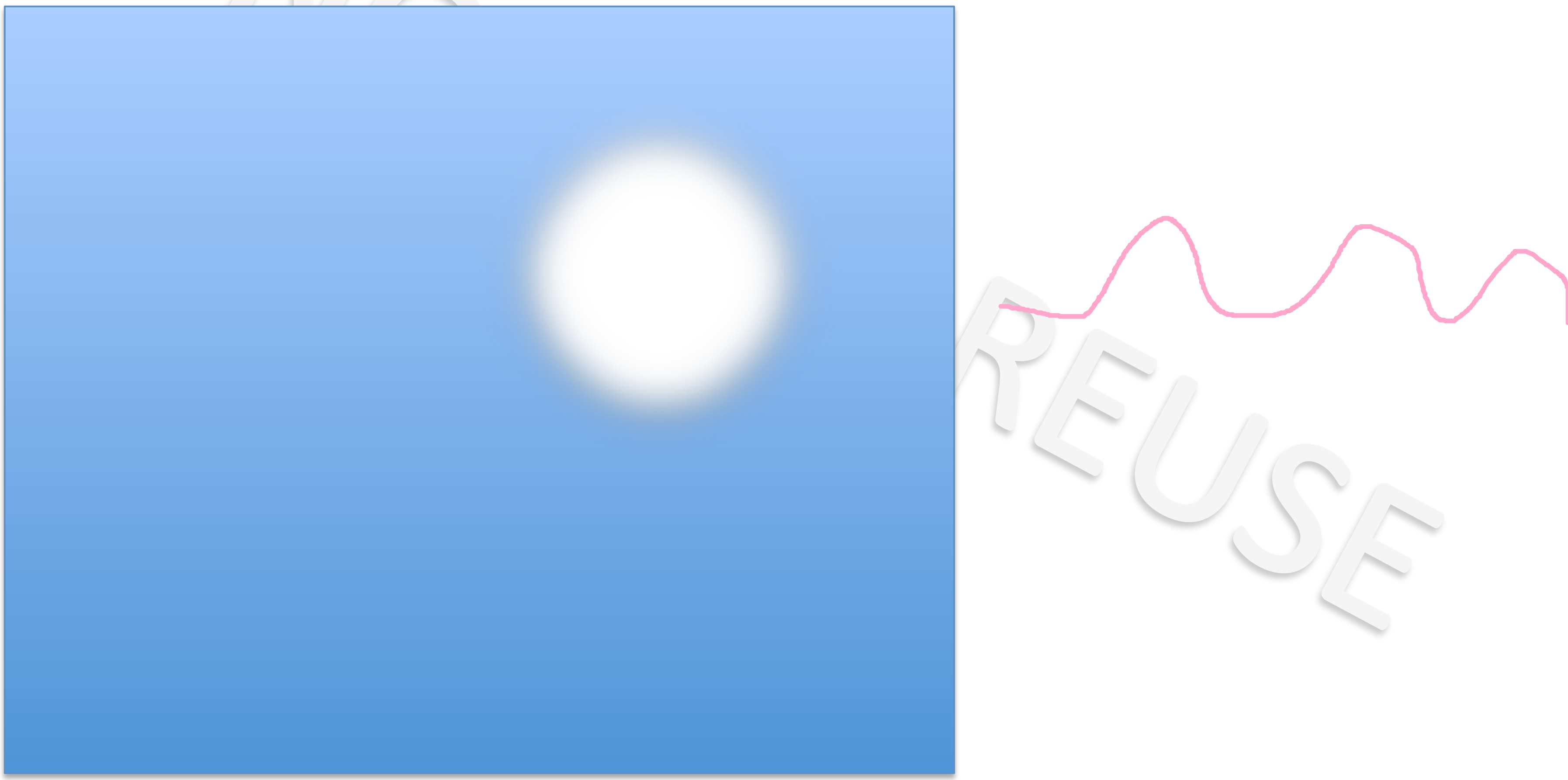
What limits information?

- ➢ spike precision, which blurs r(t) 如果我们的时间拉的太长，时间分辨率太低，就相当于一个总平均，什么信息都被平均掉了。
- ➢ the mean spike rate. 噪音大，这个也大。



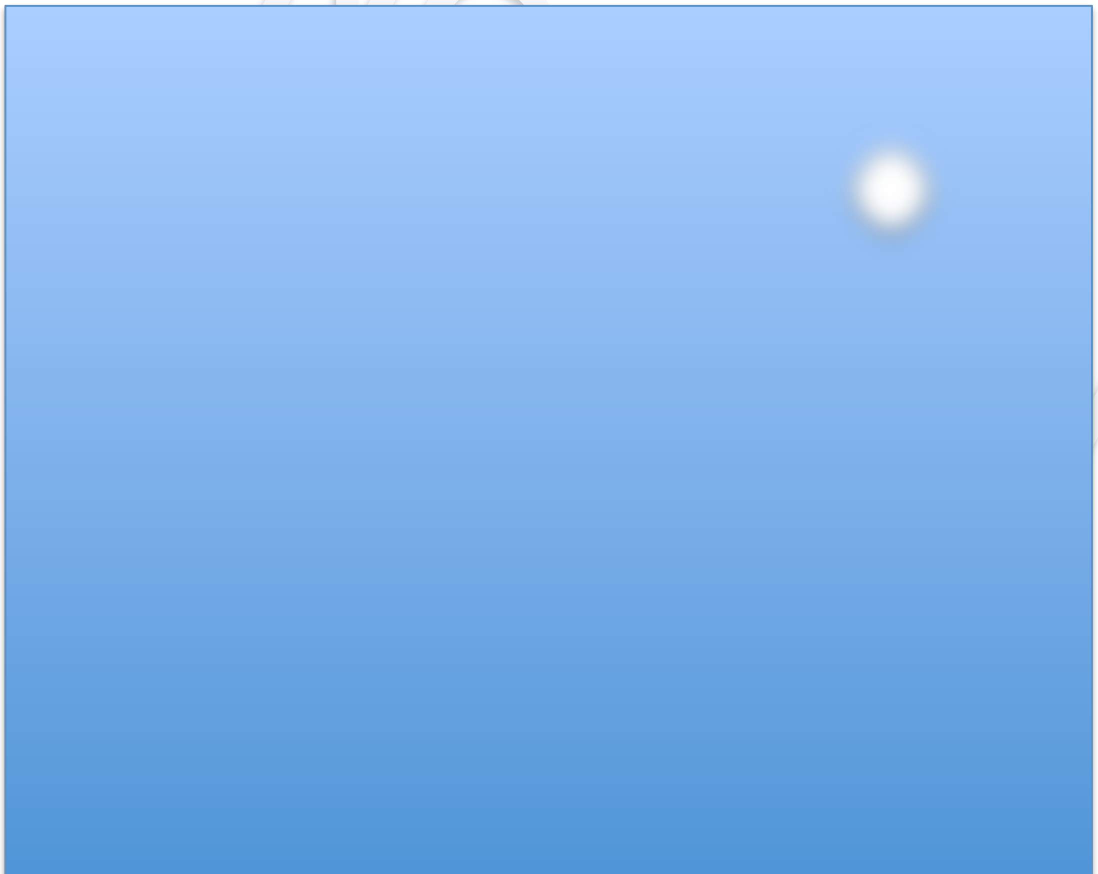
Information in single spikes

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



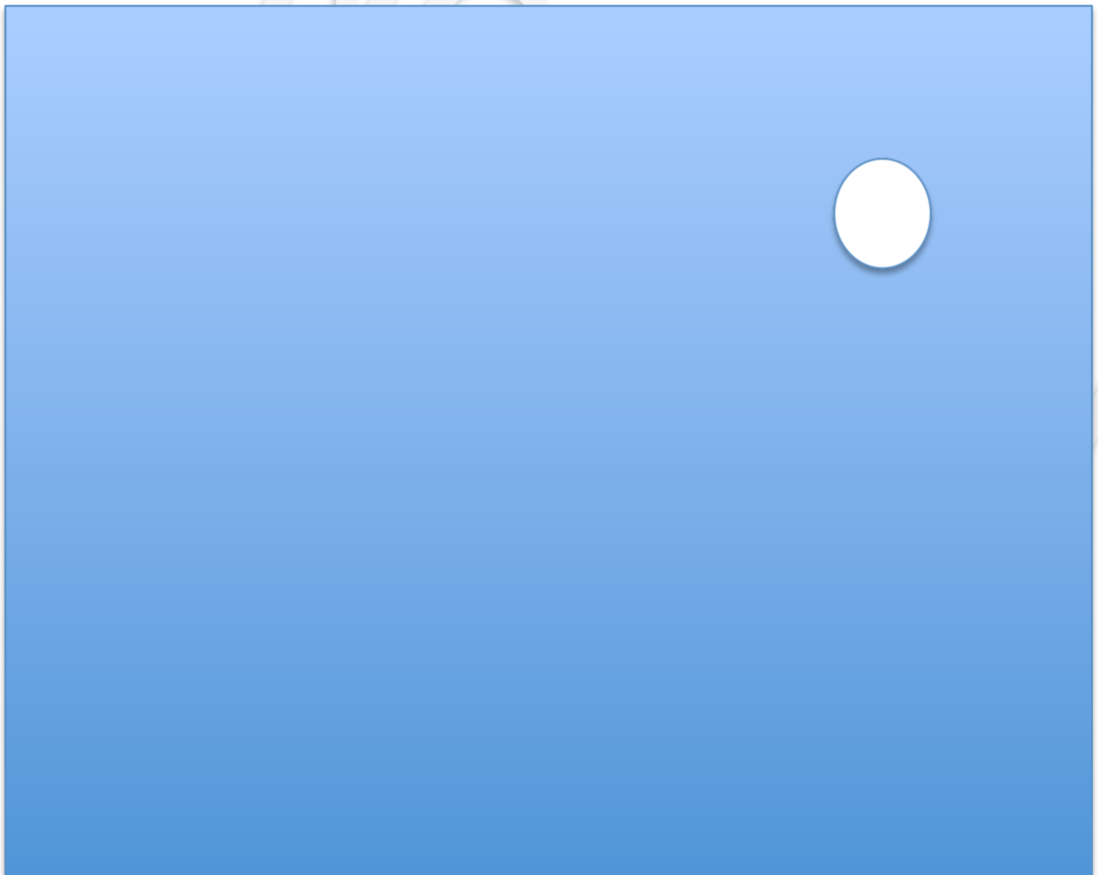
Information in single spikes

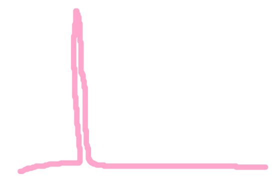
Information per spike: $$I(r,s) = \frac{1}{T}\int_0^T dt\,\frac{r(t)}{\bar{r}}\log\frac{r(t)}{\bar{r}}$$



Information in single spikes

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





Next up: information and coding efficiency

- What are the challenges posed by natural stimuli?
- What do information theoretic concepts suggest that neural systems should do?
- What principles seem to be at work in shaping the neural code?