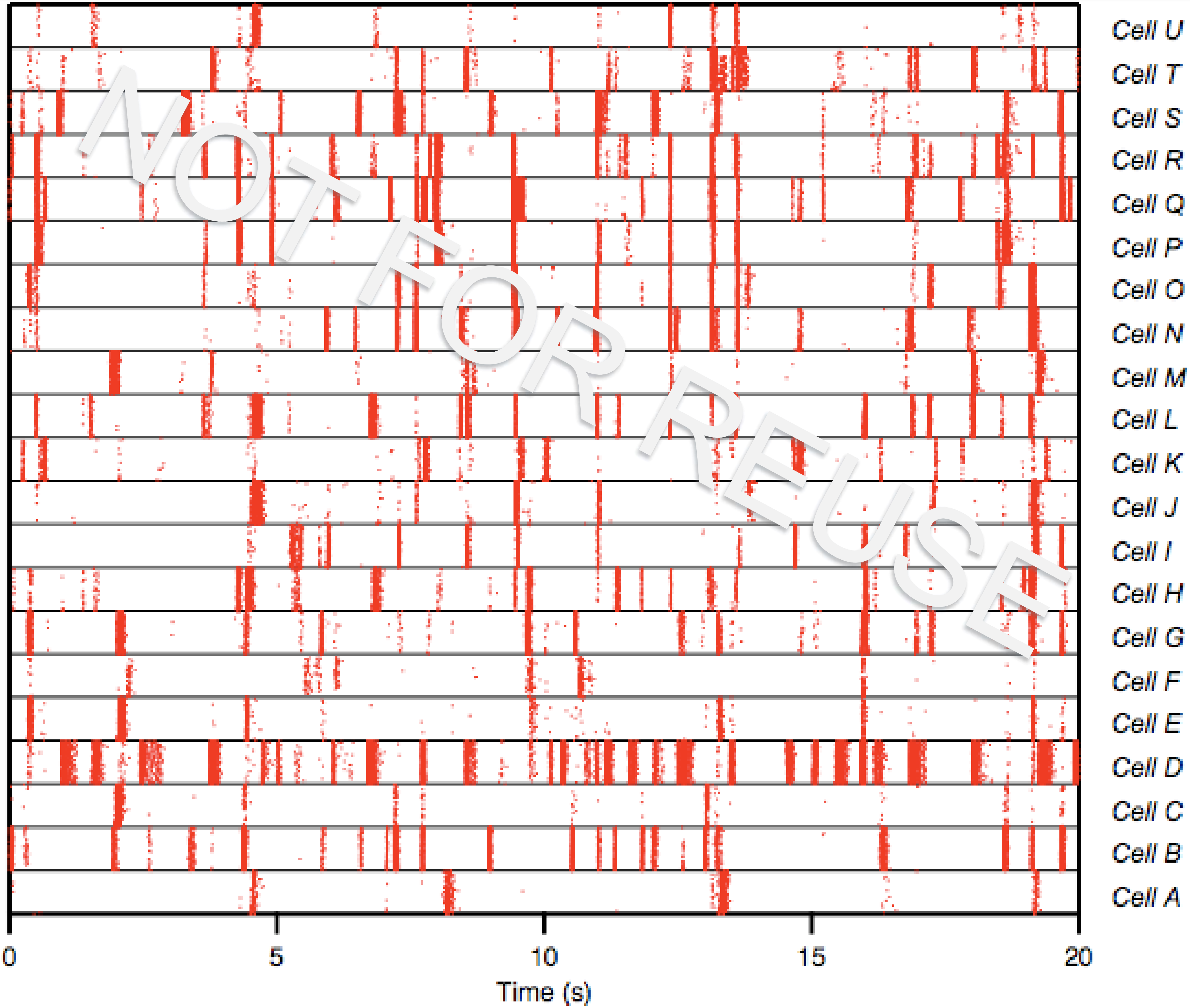
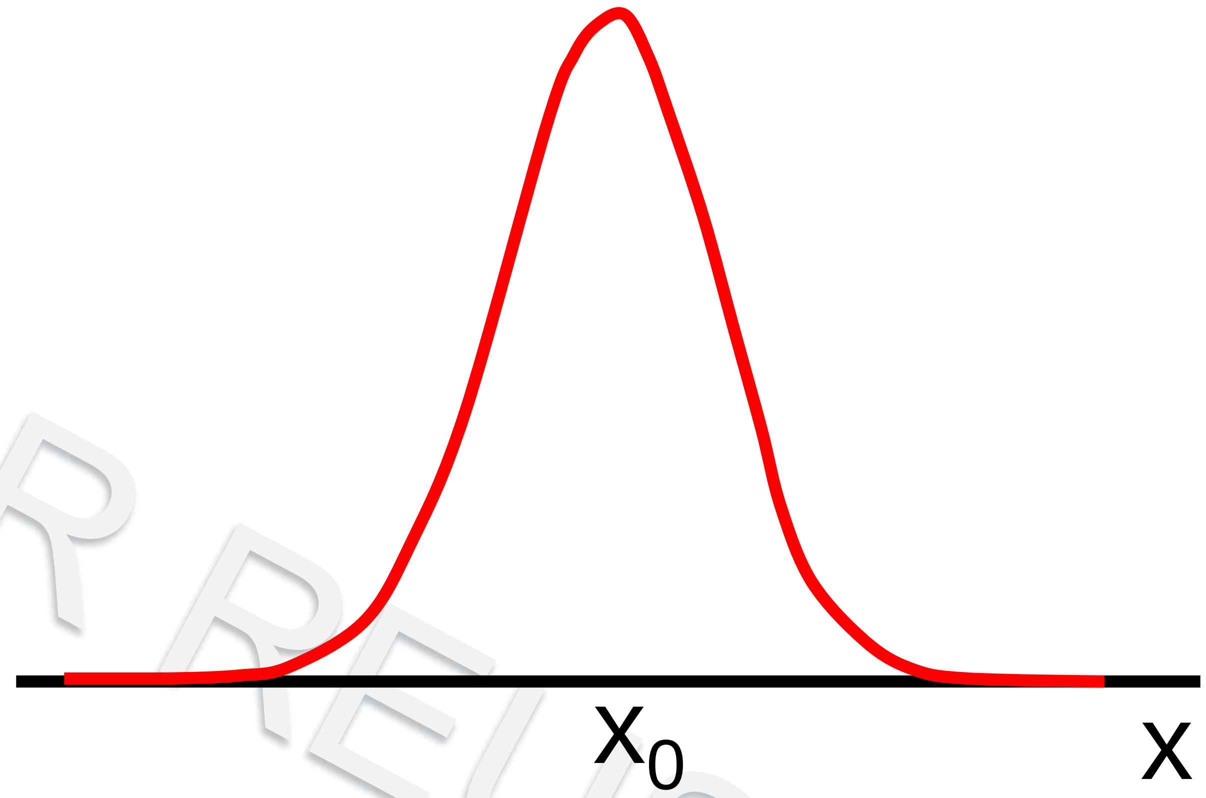
Better models



The magical Gaussian

$$P(x) = Ae^{-((x-x_0)^2/2\sigma^2)}$$



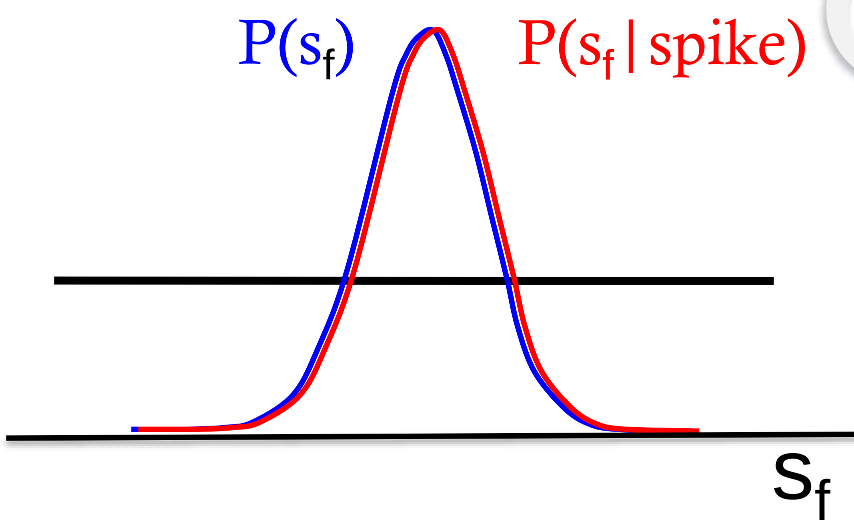
When have you found a good feature or features?

- When the input/output curve over your variable is *interesting*.
- How to quantify interesting?

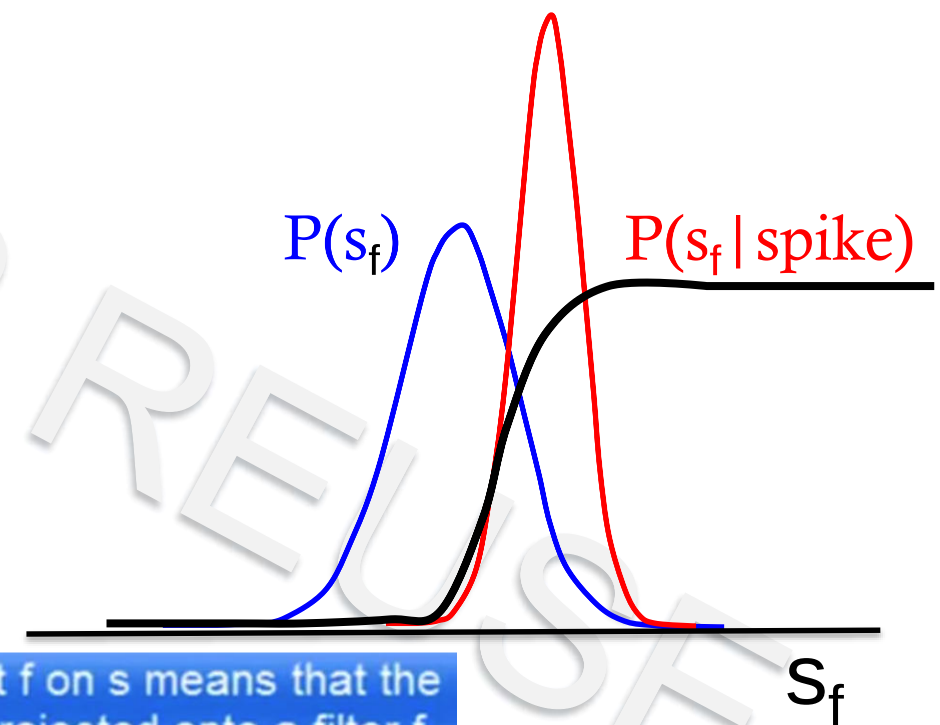
When have you done a good job?

Tuning curve: $P(spike|s_f) = P(s_f|spike)\, P(spike) \,/\, P(s_f)$

Boring: spikes unrelated to stimulus



Interesting: spikes are selective



The subscript f on s means that the stimulus is projected onto a filter f

So, instead of taking the average or doing PCA to find that filter, could we just go directly to these distributions, the prior and the conditional distribution, and ask, can I find an F? A choice of F, that when I project the stimulus onto it, that the conditional distribution and the prior are as different as possible.

Introducing the Kullback-Leibler divergence

$$D_{KL}(P(s),Q(s)) = \int ds\, P(s) \log_2 P(s)/Q(s)$$

也就是找到spike的时候最多的一个因素提供更多的信息

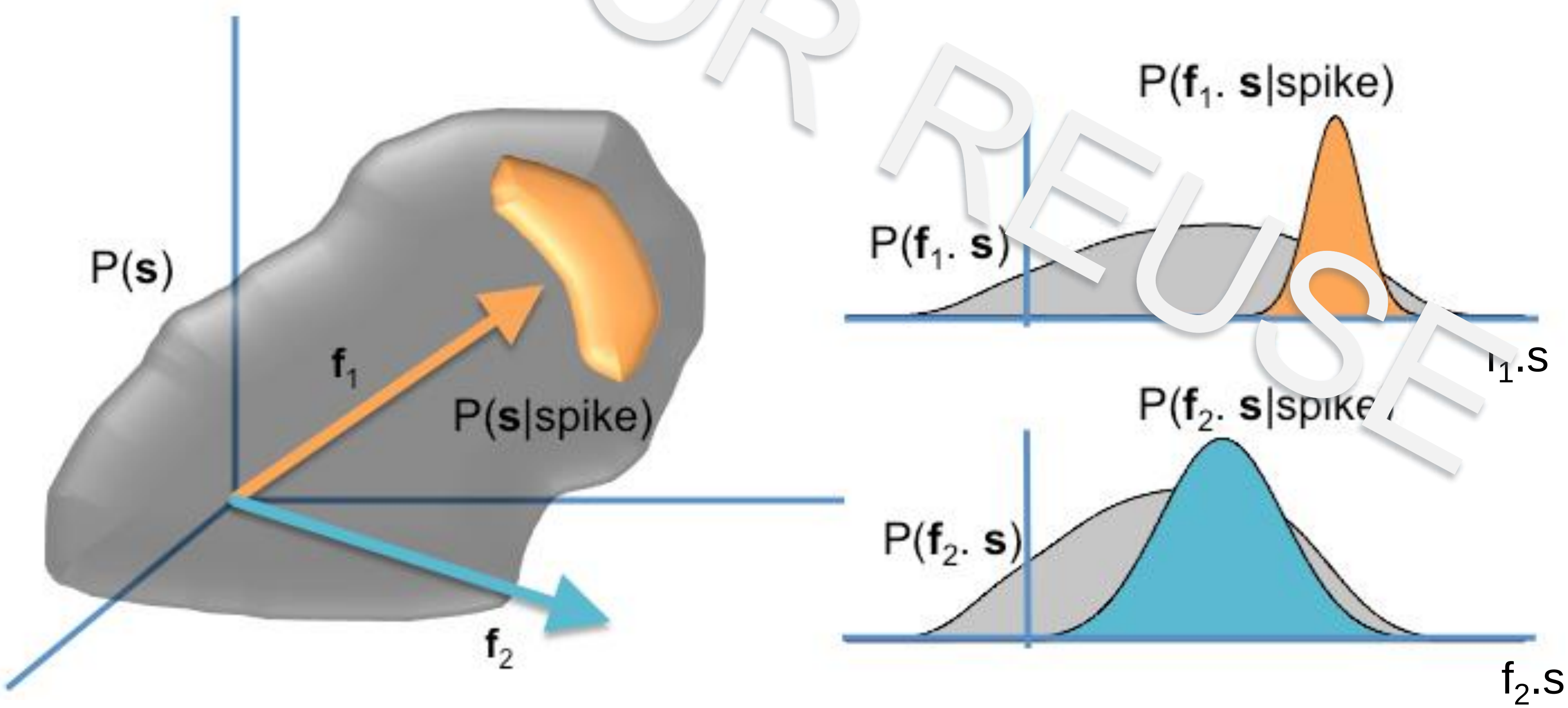
Goodness measure: $D_{KL}(P(s_f|spike), P(s_f))$

Maximally informative dimensions

Choose filter in order to maximize D_{KL} between spike-conditional and prior distributions

One can move around in this space and keep evaluating these two distributions, and search for an f that maximizes the difference between those two distributions.

This turns out to be equivalent to maximizing the mutual information between the spike and the stimulus. So we're trying to find a stimulus component that is as informative as possible.



Sharpee, Rust and Bialek, Neural Computation (2004)

Image from Fairhall, Barreiro, Shea-Brown (2012)

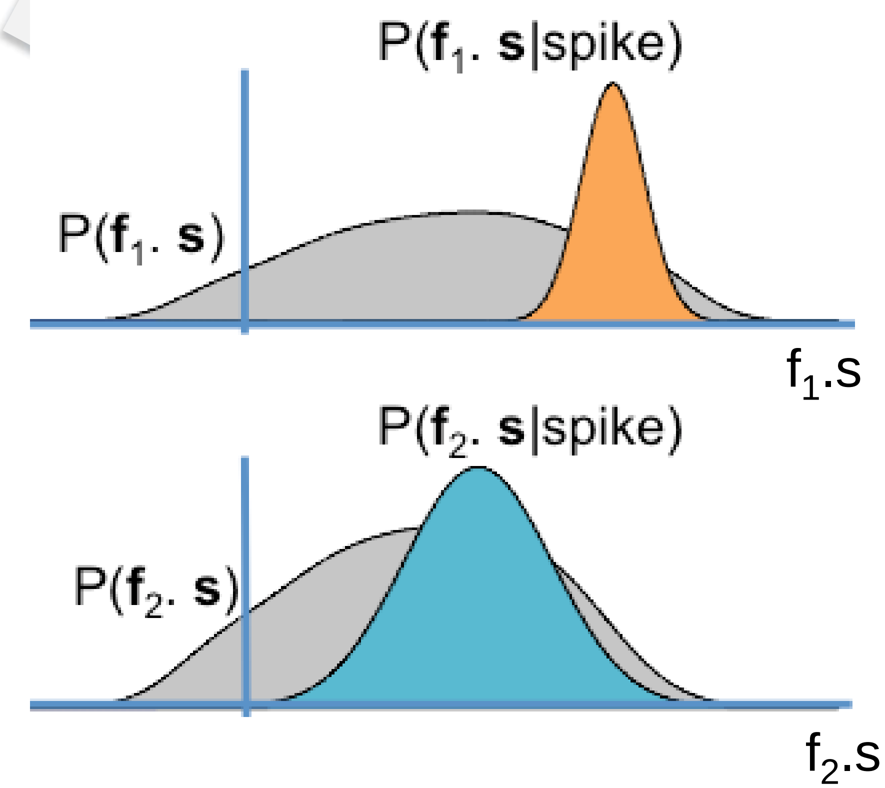
Maximally informative dimensions

Choose filter in order to maximize D_{KL} between spike-conditional and prior distributions

Equivalent to maximizing mutual information between stimulus and spike

Does not depend on white noise inputs

Can be used for deriving models from natural stimuli



Finding relevant features

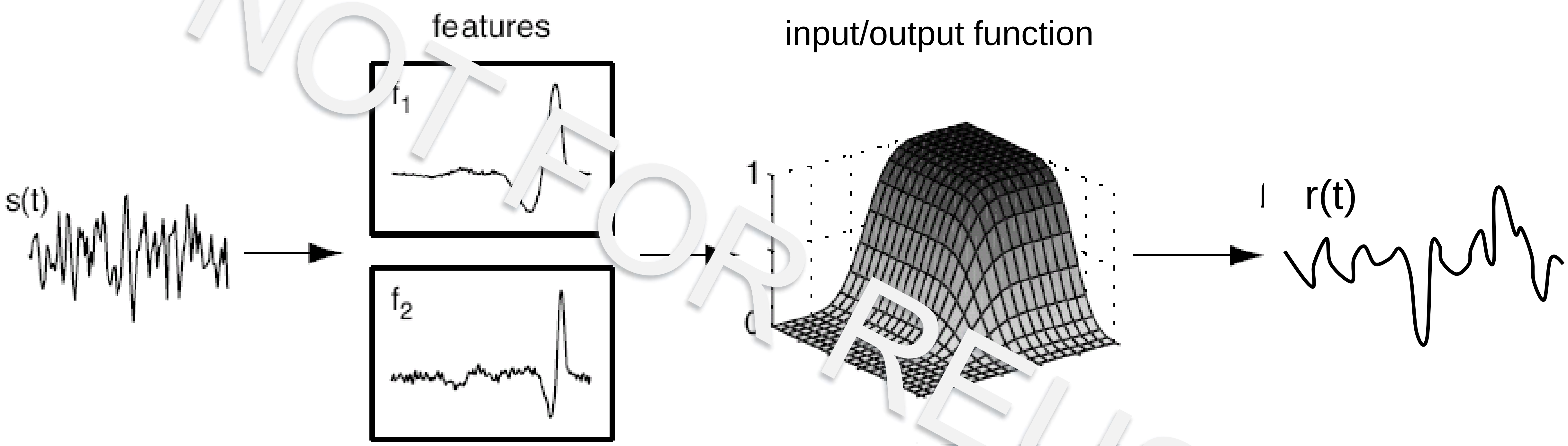
1. Single filter determined by the conditional average

2. A family of filters derived using PCA

3. Information theoretic methods use the whole distribution

   Removes requirement for Gaussian stimuli

Modeling the noise



Bernoulli trials

Binomial spiking

Distribution: $P_n[k] = ?$

Mean: <k> = ?

Variance: $\mathrm{Var}(k) = ?$

Poisson spiking

Distribution: $P_T[k] = (rT)^k \exp(-rT)/k!$

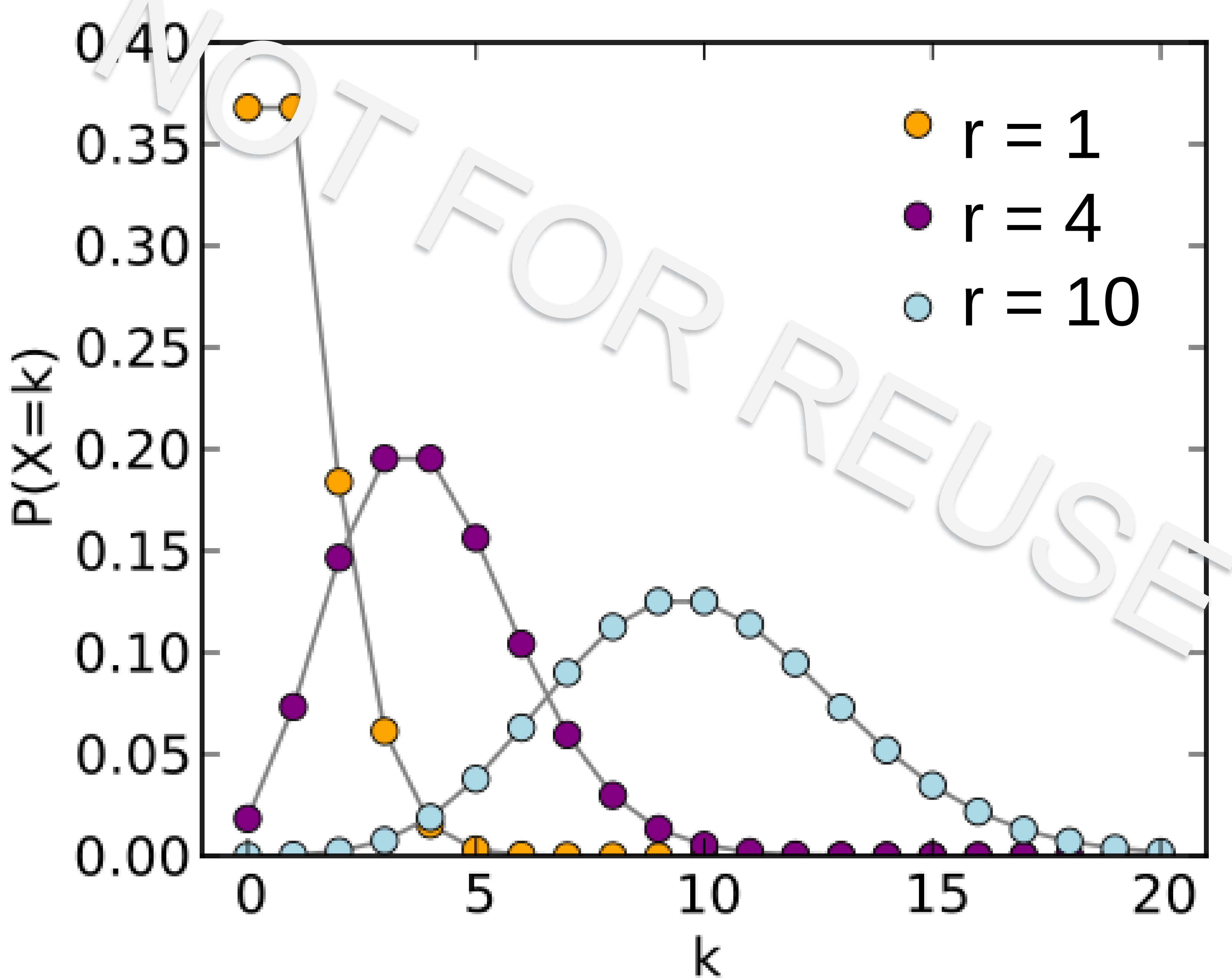
Mean: $\langle k \rangle = rT$

Variance: $\mathrm{Var}(k) = rT$

Fano factor: mean/var $F = 1$

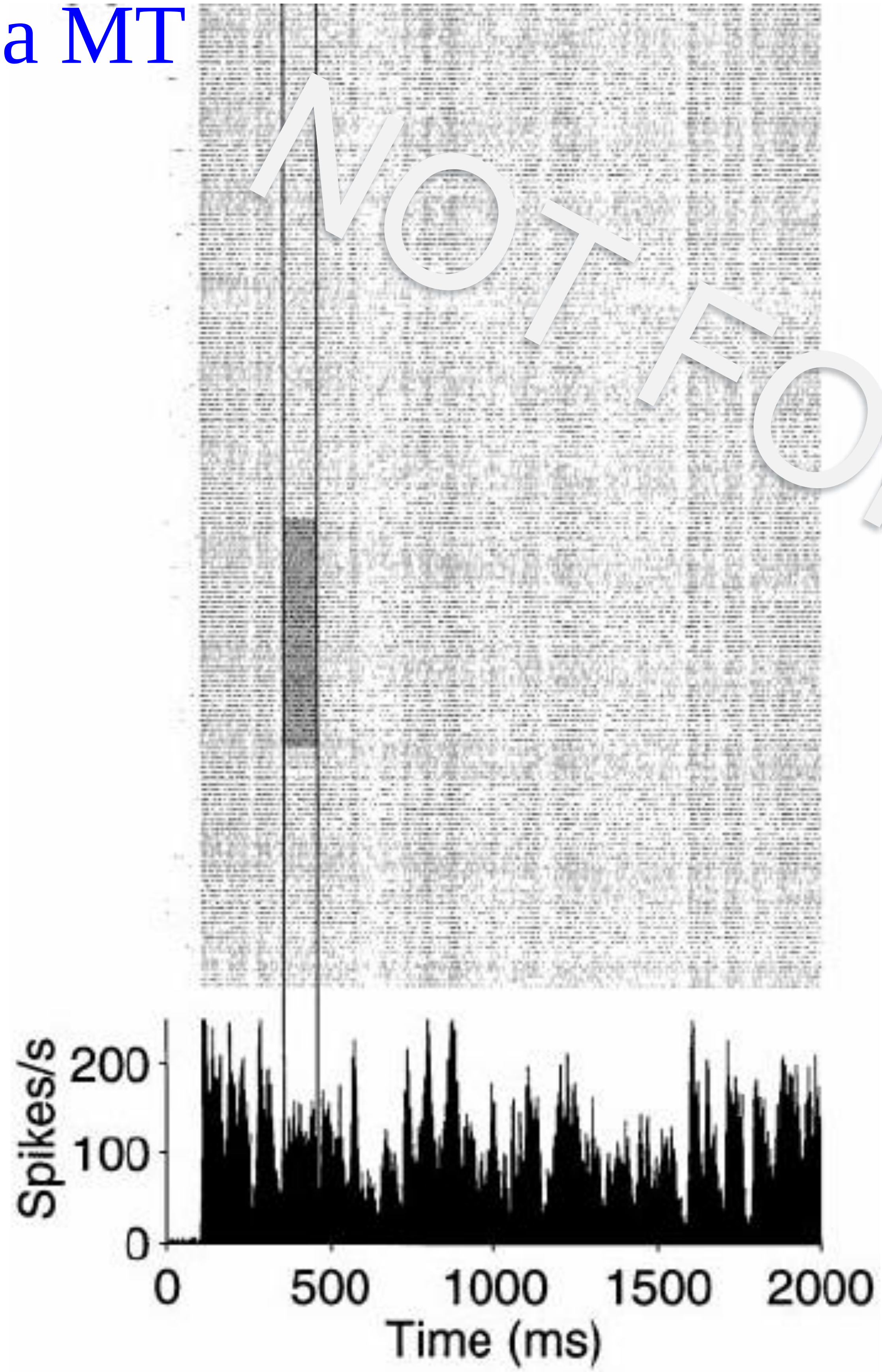
Interval distribution: $P(T) = r \exp(-rT)$

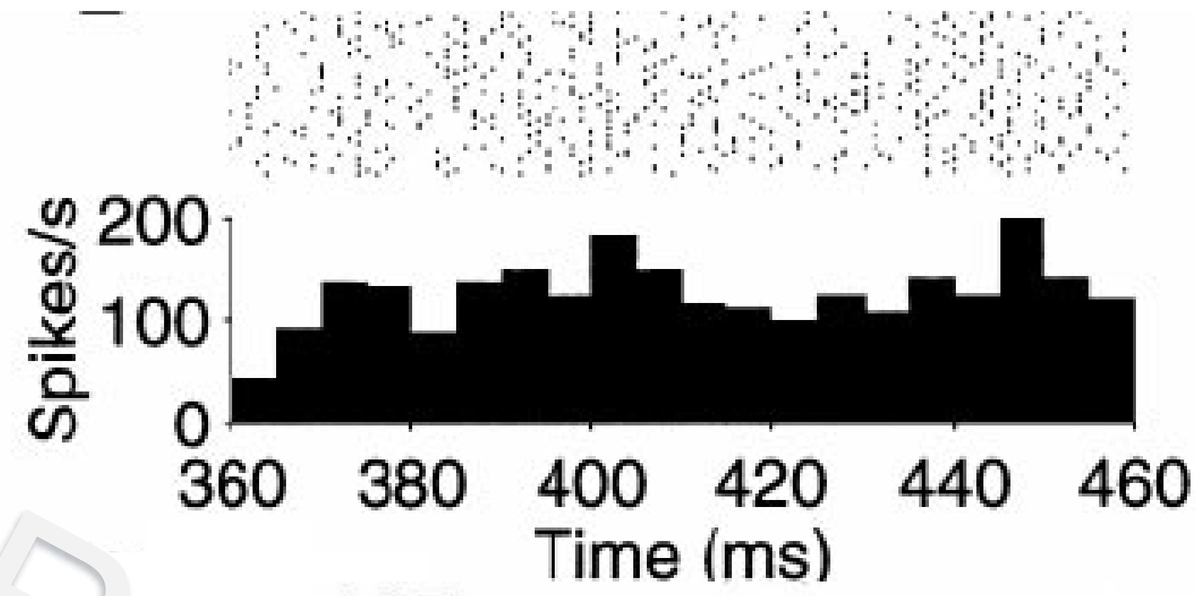
The Poisson distribution

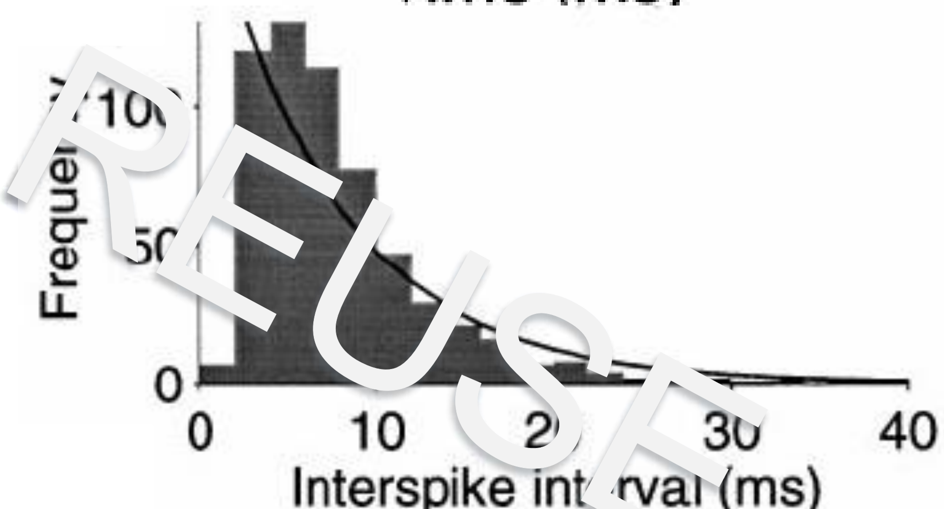


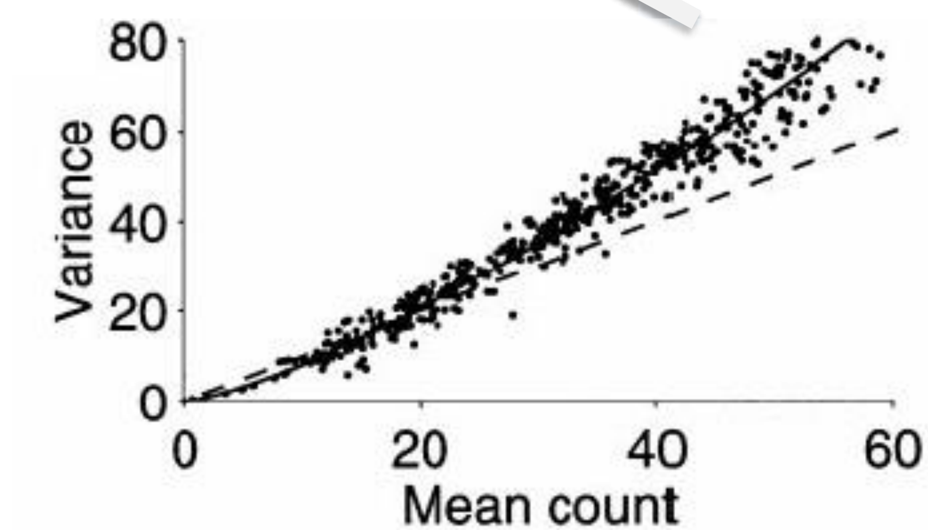
Poisson or not?

Area MT





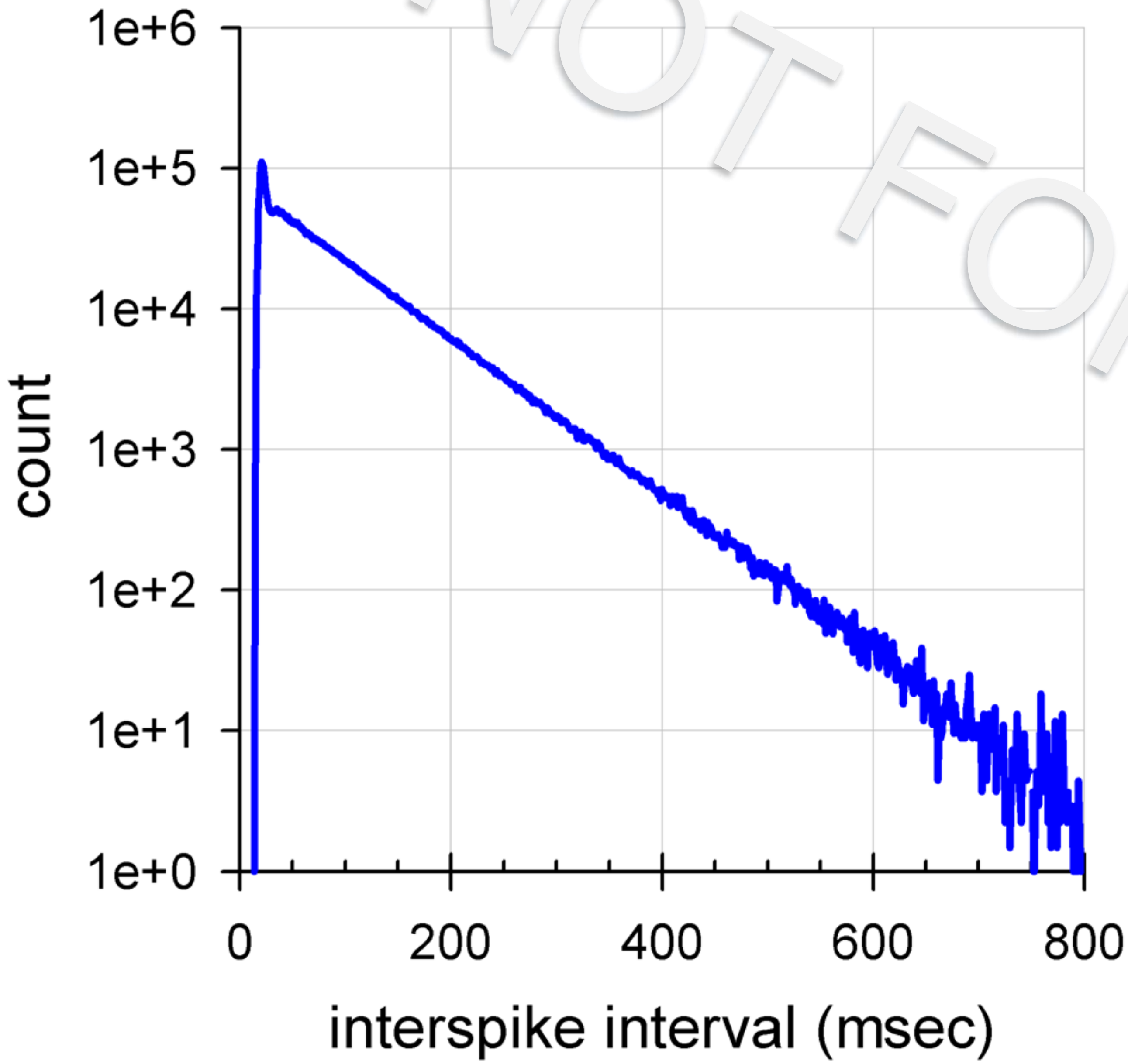


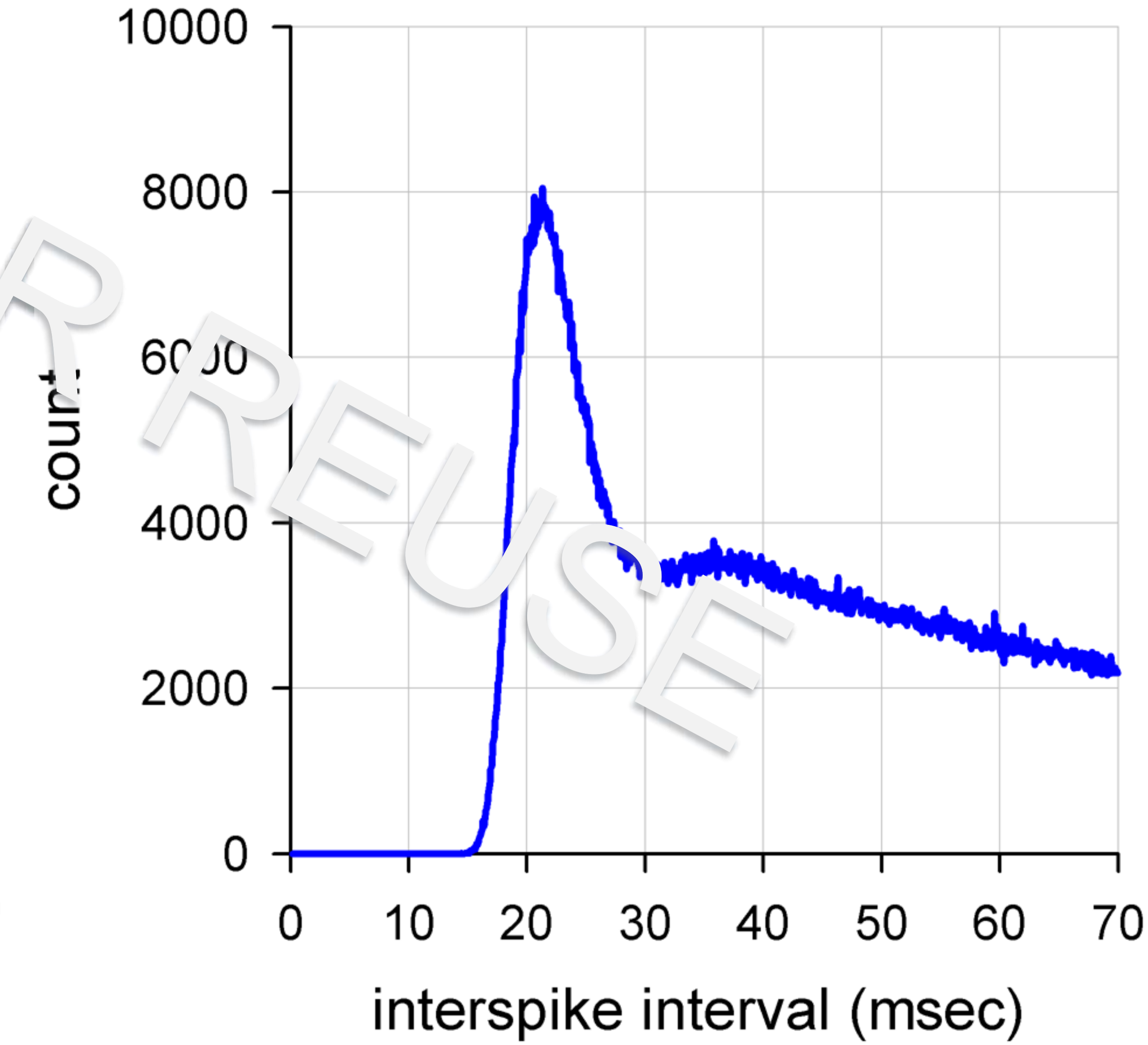


在很短的一个时间段内，fano factor很像1，似乎是poisson

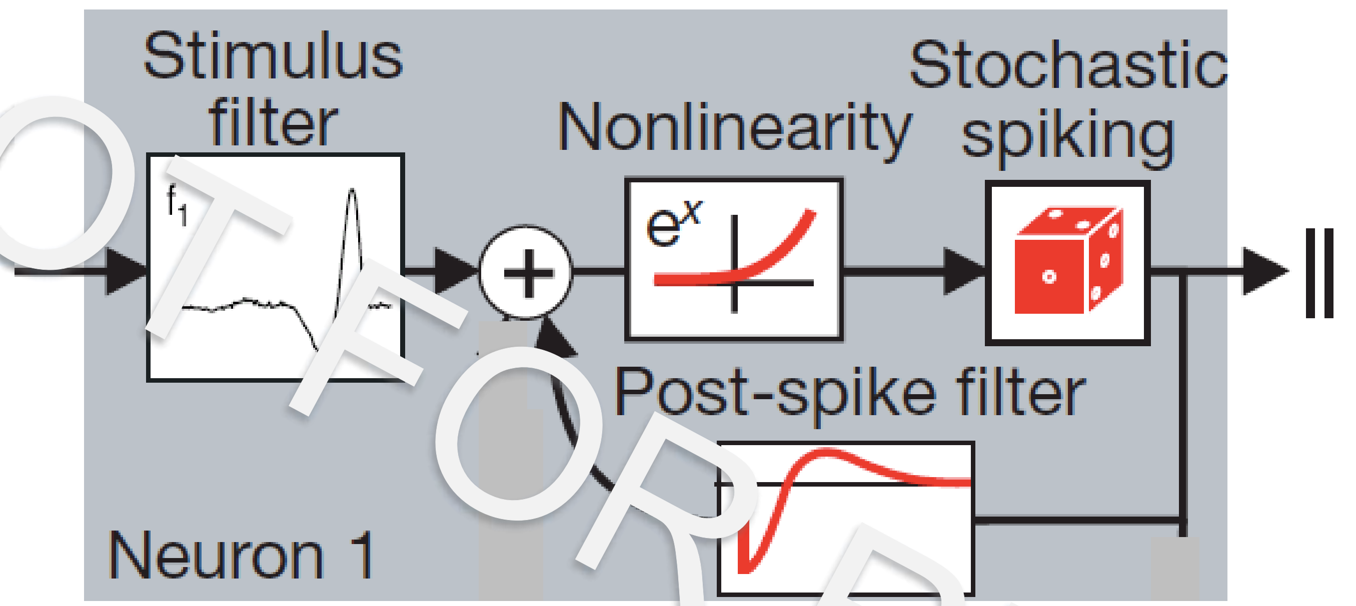
Shadlen and Newsome, 1998

Interspike interval distributions





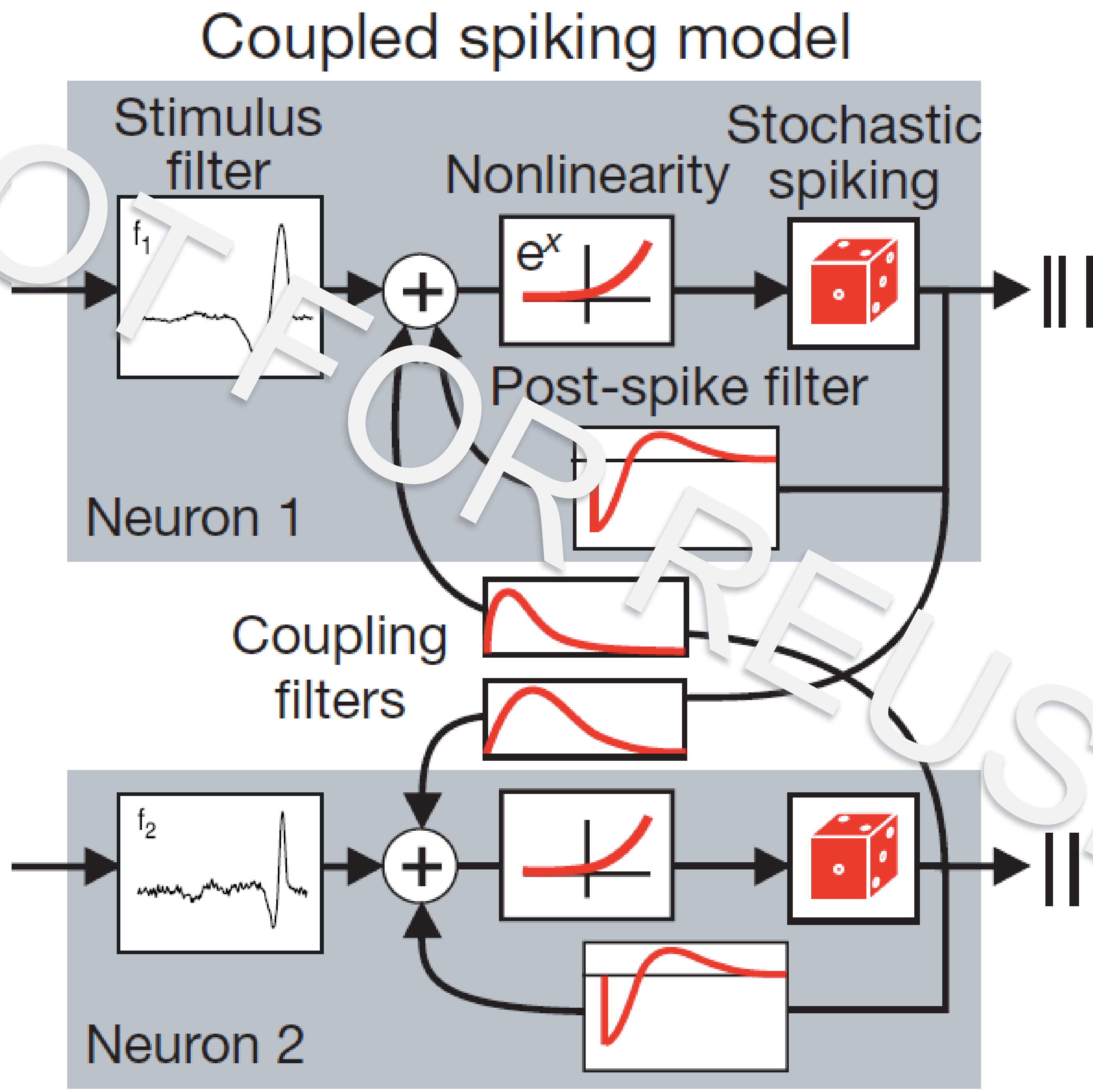
The generalized linear model



GLM: P(spike at t) ~ $\exp(f_1 * s + h_1 * r)$

Pillow et al., *Nature* 2008

But wait, there's more!



$$\text{GLM: } r(t) = g(f_1 * s + h_1 * r_1 + h_2 * r_2 + \ldots)$$

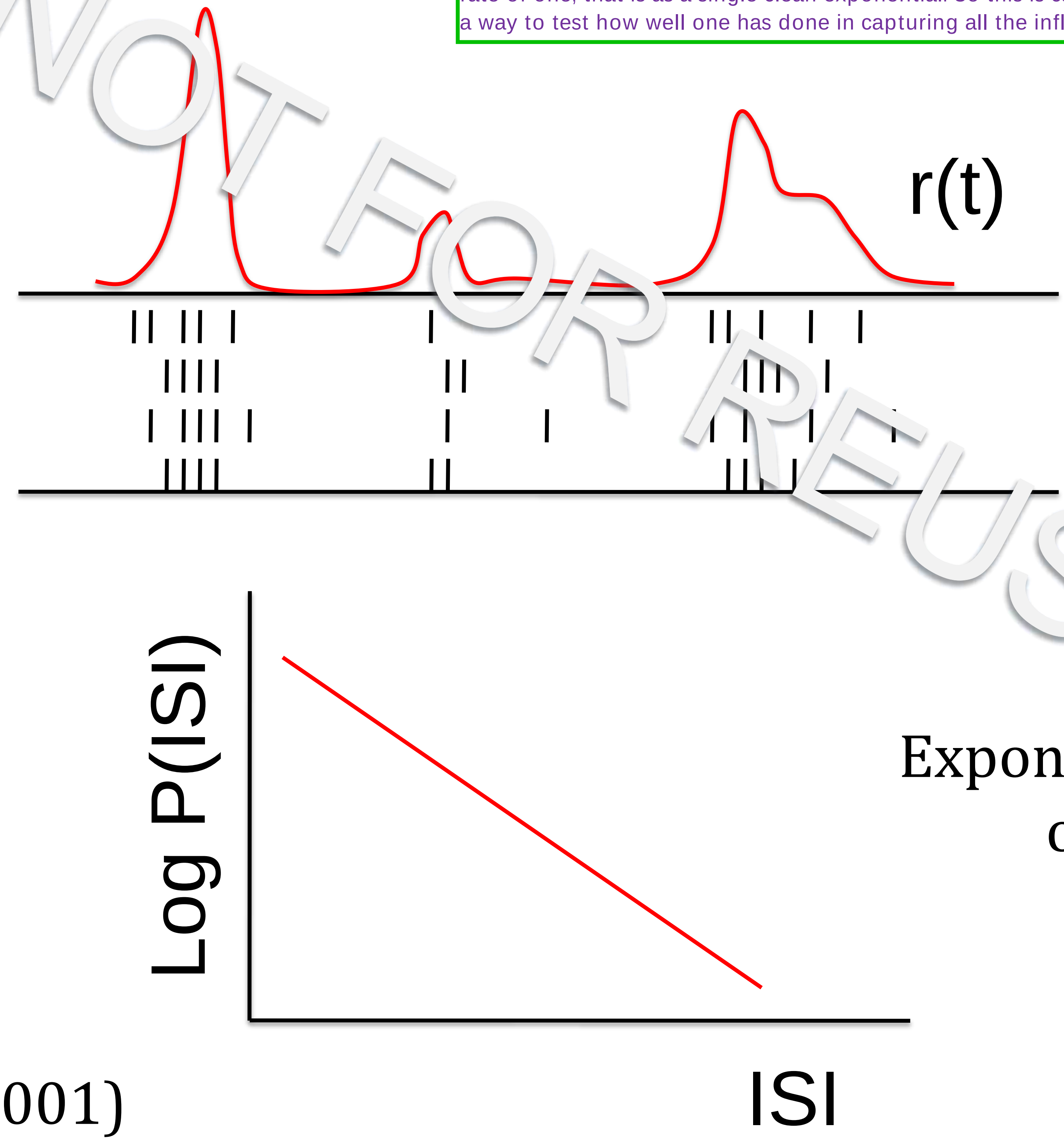
Pillow et al., *Nature* 2008

Time-rescaling theorem

We can use this Poisson nature of firing to test whether we have captured everything that we can about the inputs in our model.

Let's say we have a model like the GLM, where the output depends on many influences, on the stimulus, on the history of firing in the neuron that recoding from on the history of firing in, in other neurons as well. Then we can our output spike intervals and scale them by the firing rate that's predicted by the model. So we take these intervals times between successive spikes, we scale them by the firing rate that our model predicted given all the interactions that, that we've incorporated.

If this predicted rate does truly account for all the influences on the firing, even ones due to previous spiking, then these new scaled intervals should be distributed like a pure Poisson process, with an effective rate of one, that is as a single clean exponential. So this is called the Time-rescaling theorem and it's used as a way to test how well one has done in capturing all the influences on spiking with ones models.



Exponential distribution of *scaled* ISIs: $(t_{i-1-}\ t_i)\ r(t)$

Brown et al. (2001)

That's it for encoding!

That’s it for encoding!

Next week…

- Reading minds!
- Decoding methods