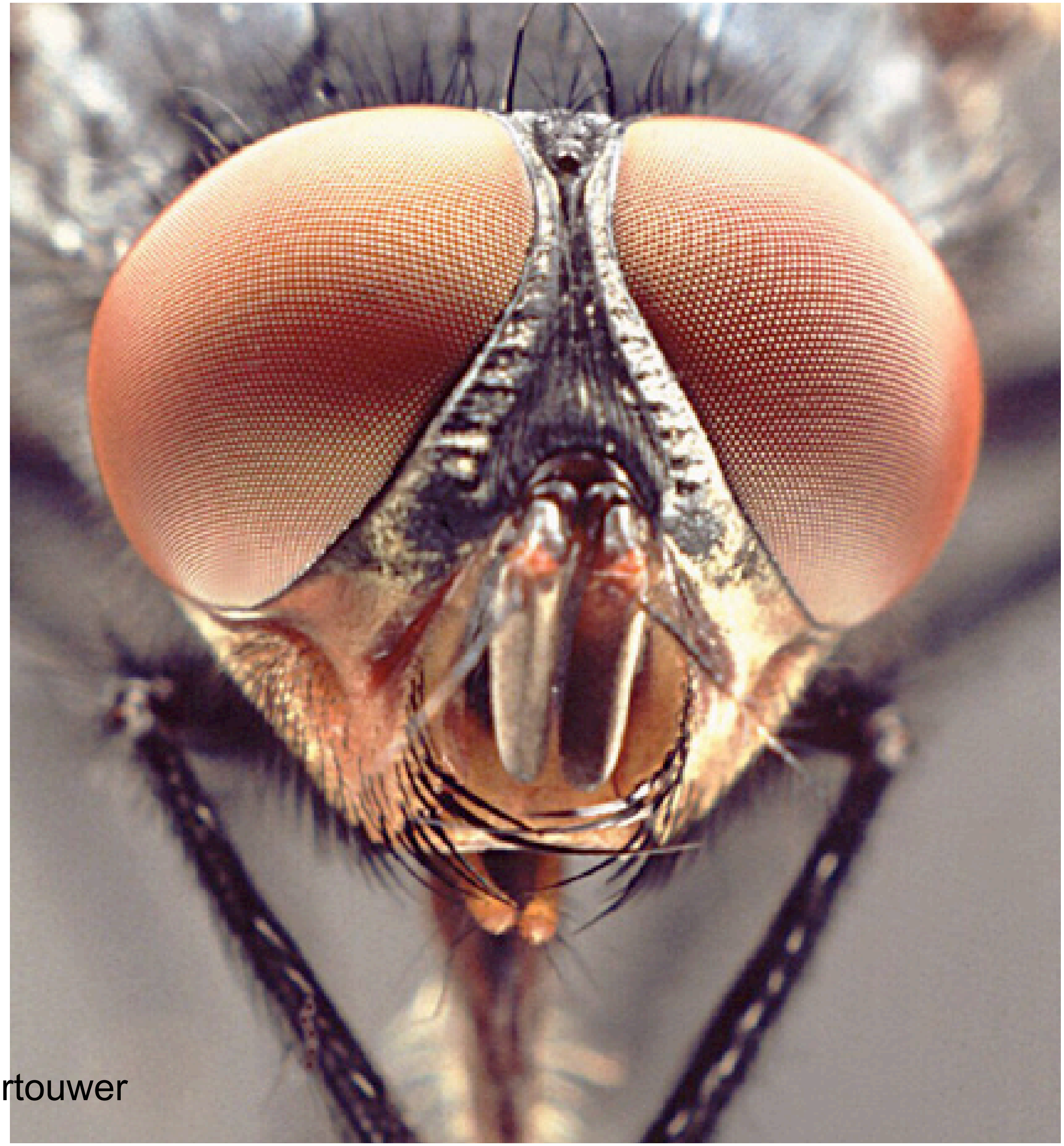
Neural Coding of Natural Stimuli: Information at Sub-Millisecond Resolution

Ilya Nemenman

CCS-3/CNLS, LANL

Geoffrey Lewen (Hun School of Princeton)
William Bialek (Princeton)
Rob de Ruyter van Steveninck (Indiana)



H. L. Leertouwer

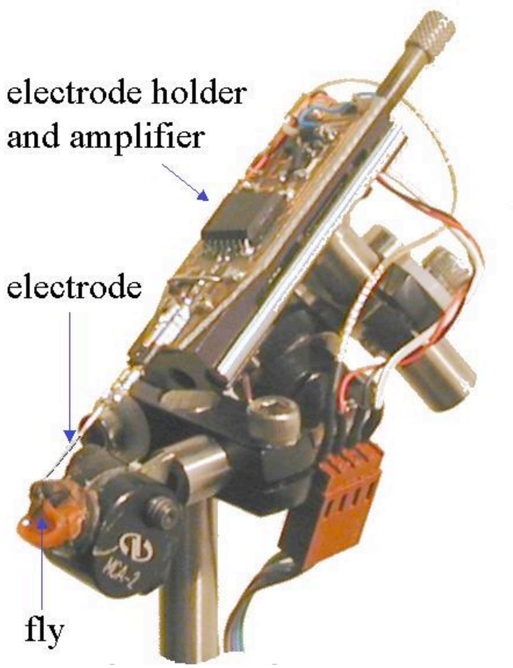
Why fly as a neurocomputing model system?

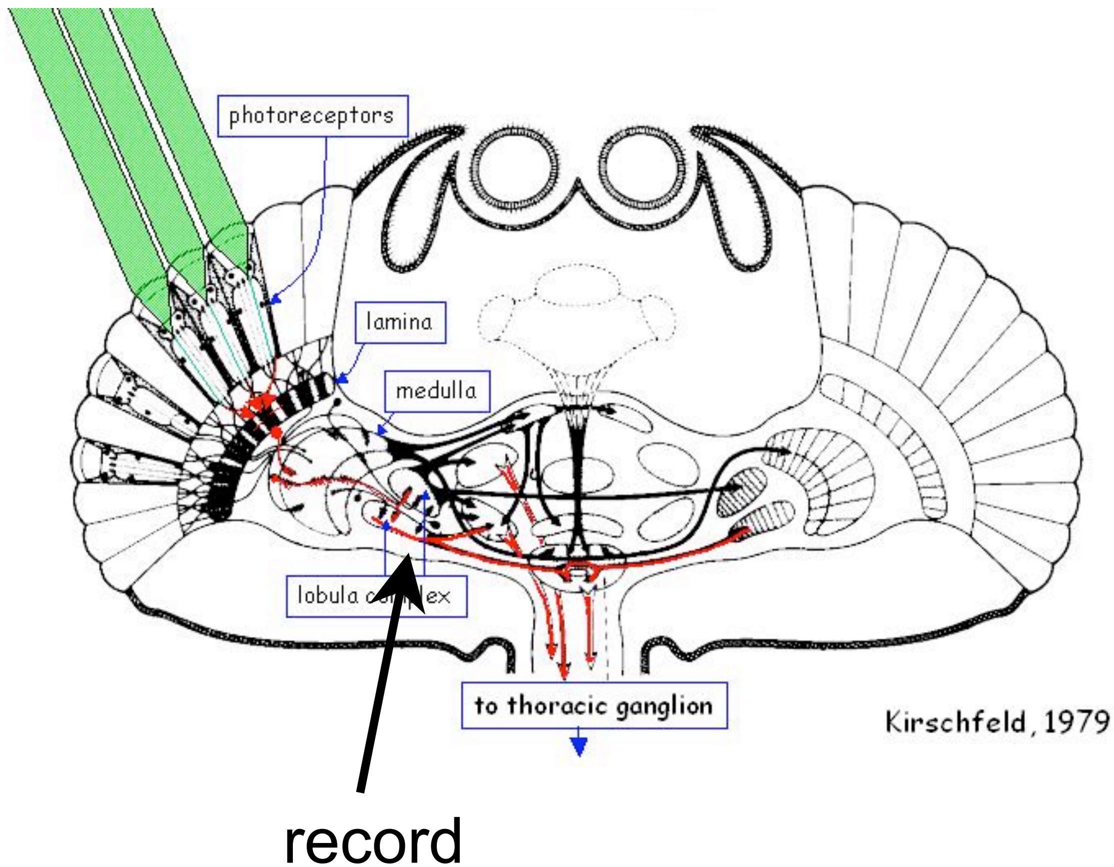
- Can record for long times
- Named neurons with known functions
- Nontrivial computation (motion estimation)
- Vision (specifically, motion estimation) is behaviorally important
- Possible to generate natural stimuli

Questions

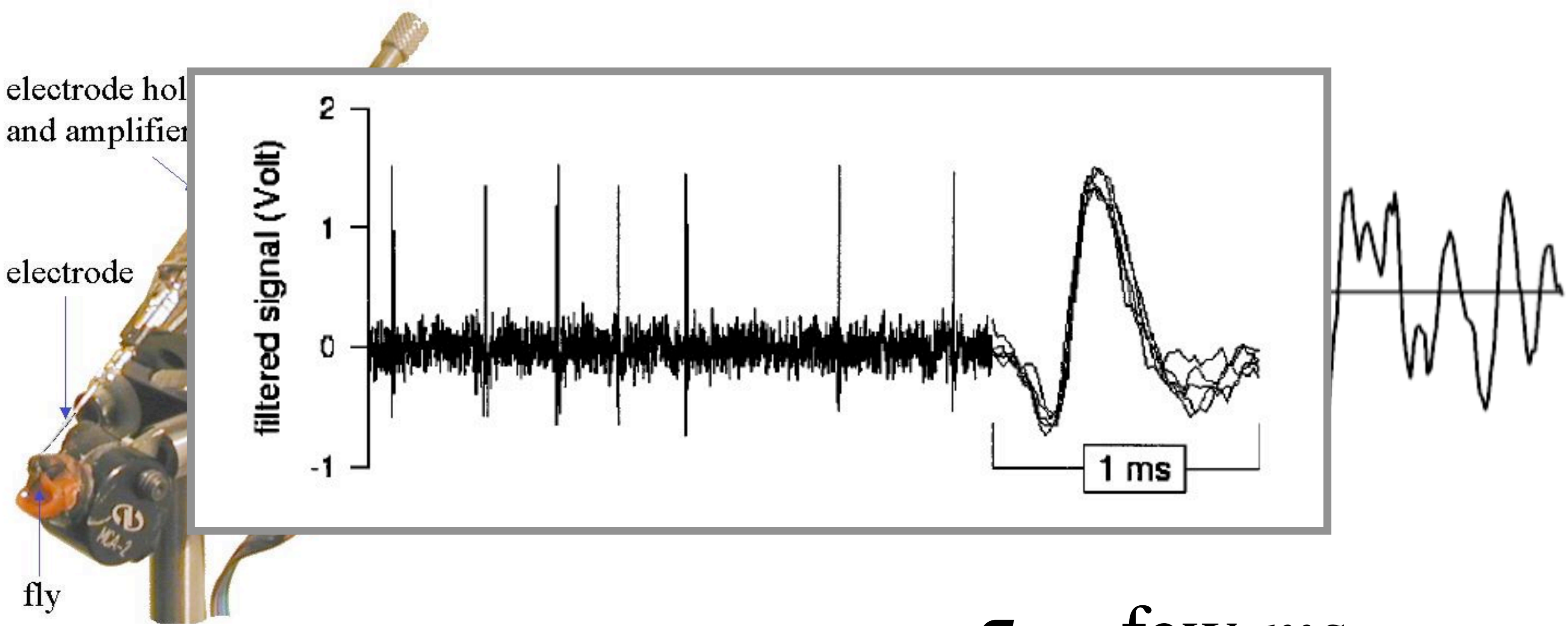
- Can we understand the code?
  - Which features of it are important?
  - Rate or precise timing (how precise)?
  - Temporal decorrelation?
  - Coding table?
  - How much does the fly know?
- Is there an evidence for optimality?

Recording from fly's H1





Motion estimation in fly H1



$\tau =$ few ms

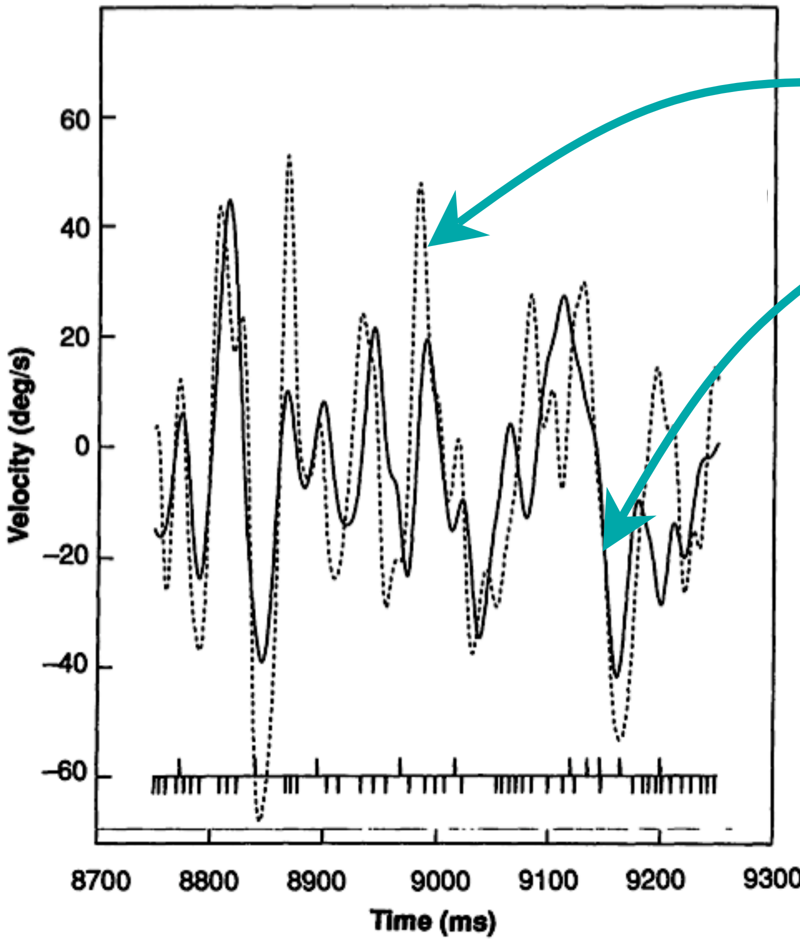
(Strong et al., 1998)

Linear decoding for sparse spikes

Small parameter $\sim \tau / (t_i - t_{i+1})$ allows to build linear decoding schemes even for nonlinearly encoded stimuli.

$$s_{est}(t) = \sum_i f_1(t - t_i) + \sum_{ij} f_2(t - t_i, t - t_j) + \ldots$$

Linear decoding



stimulus

reconstruction

$$\langle t_{i+1} - t_i \rangle = 30ms$$

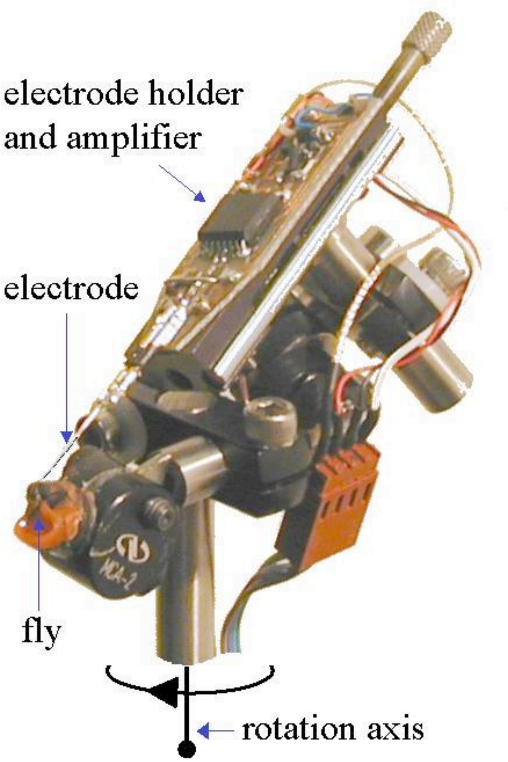
Position of each spike within ~2ms matters!

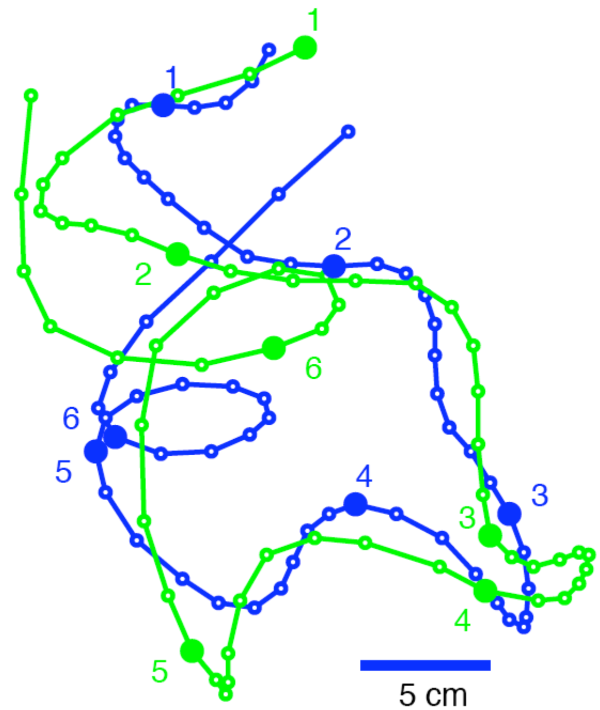
(Bialek et al. 1991, Strong, et al, 1998)

But…

- Slow signals: rate code
- Fast (white) signals: 2 ms resolution important
- Could such ~1 ms precise spikes be due to ~1 ms correlations in stimulus?
- What if stimulus has natural correlations?

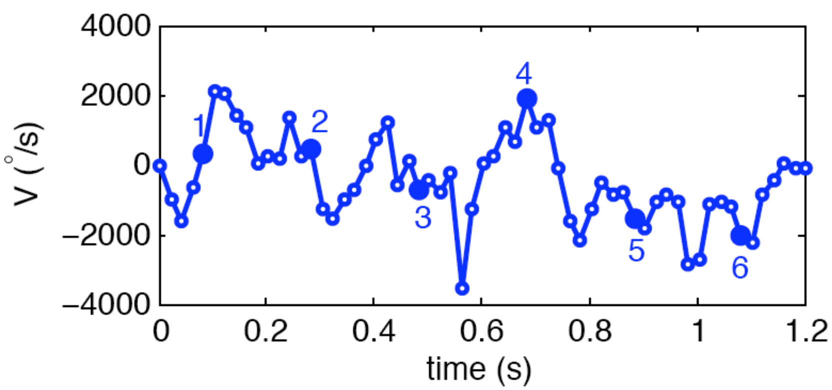
Natural stimuli

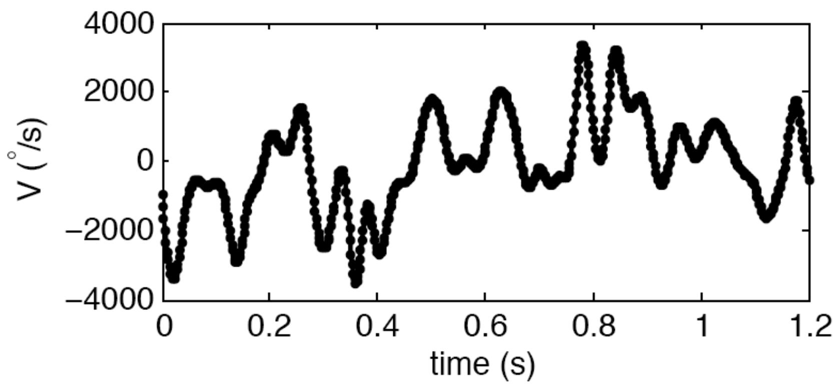


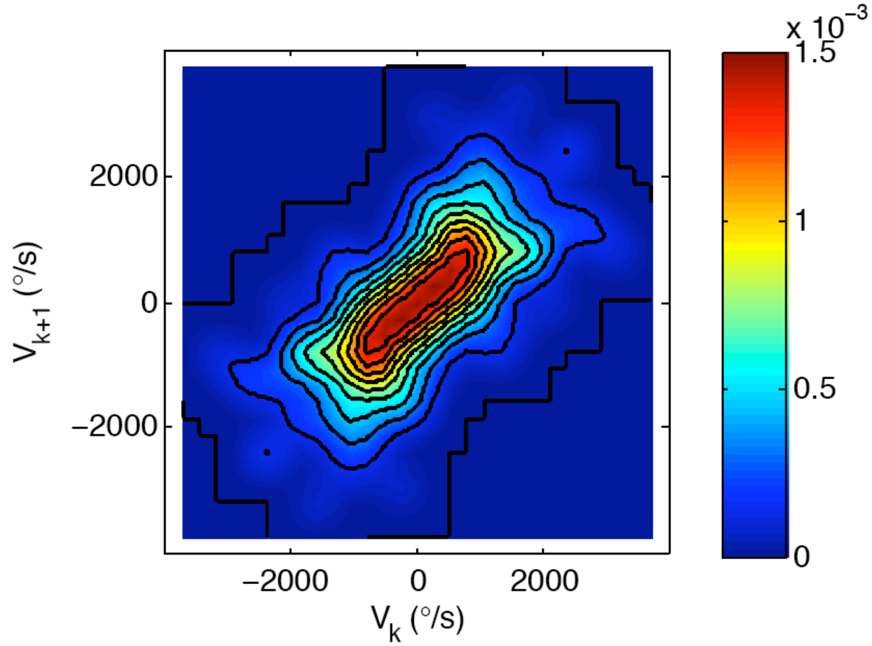


Land and Collett, 1974

Natural stimuli

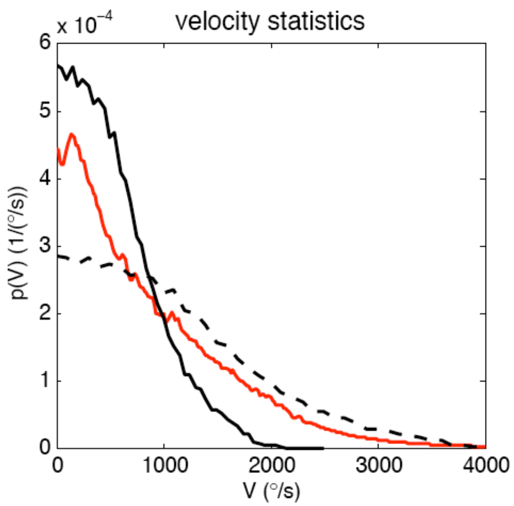


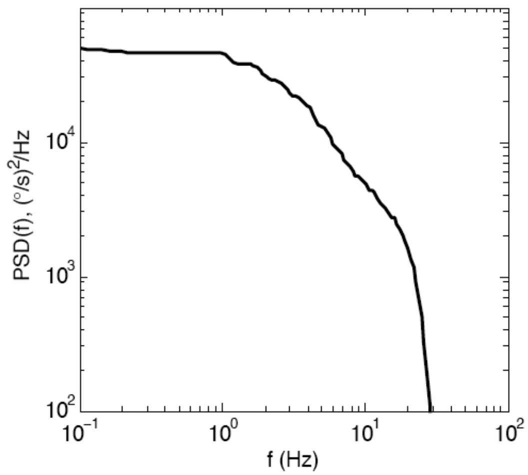


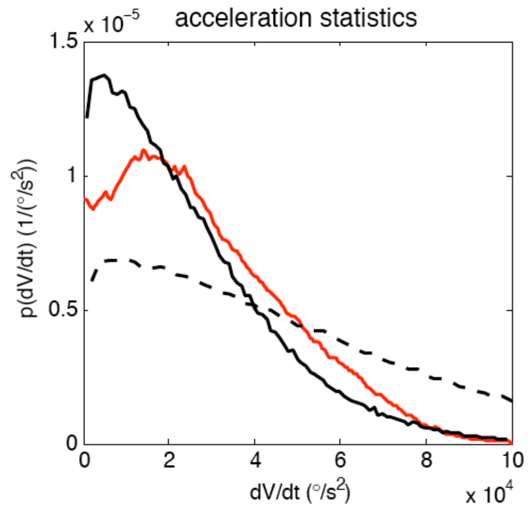


$$\tau = 60\text{ms}$$

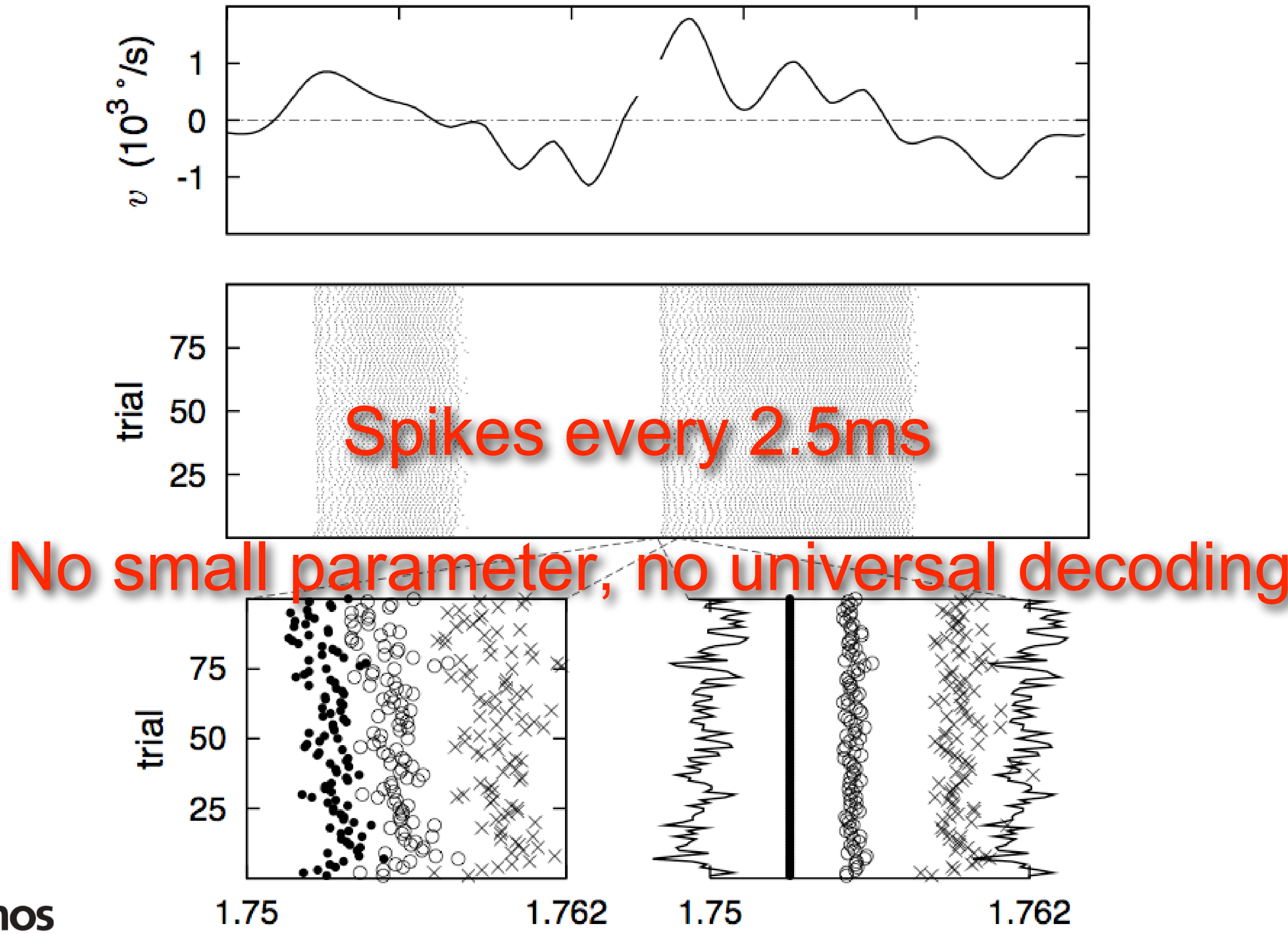
Natural stimuli: tests







Natural stimulus and response



Not rate coding?

Is high timing precision (0.2 ms for first spike, and 0.1 ms for intervals) for natural stimuli relevant for information transmission, or just anecdotal?

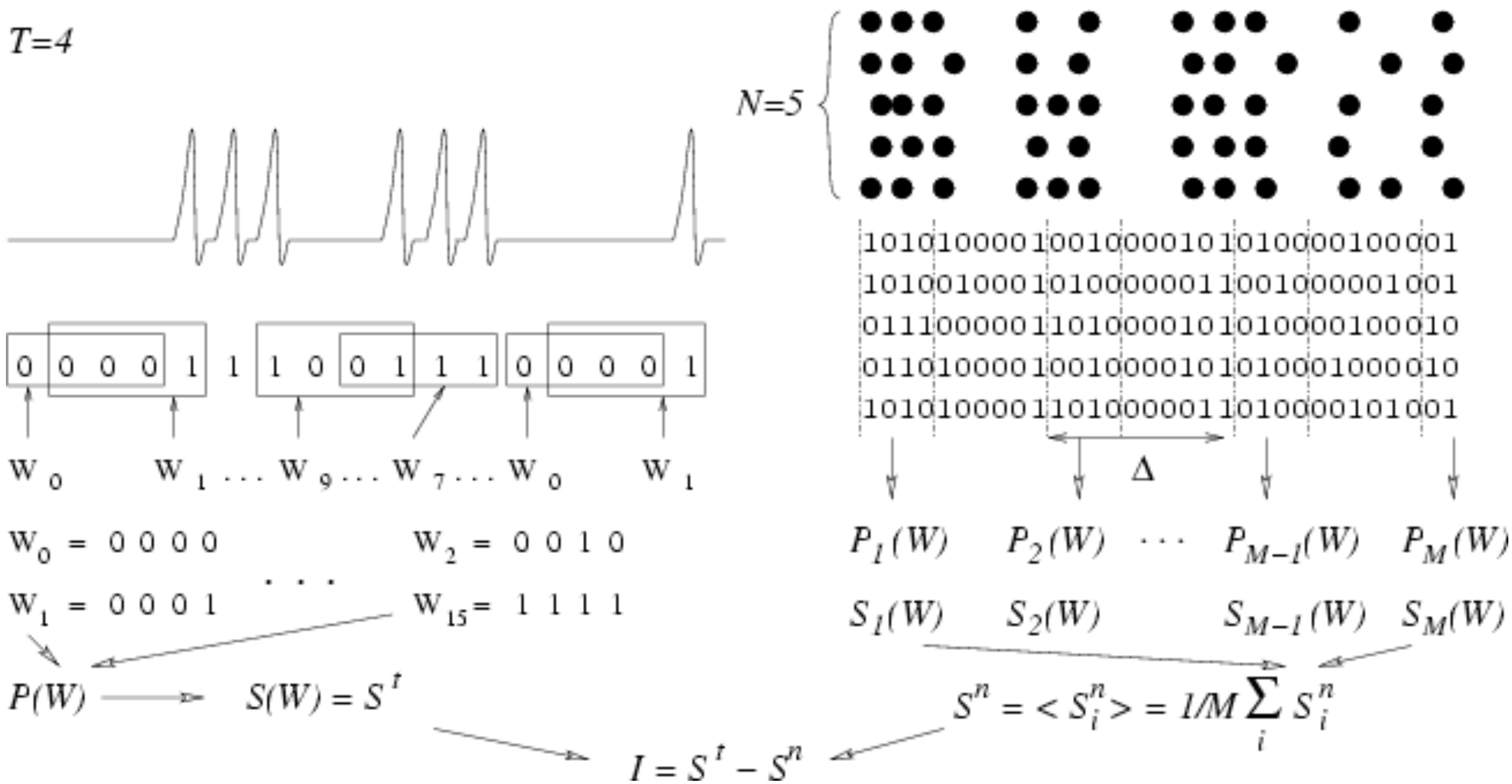
Characterize coding without explicit decoding

$$S[x] = -\sum_{x} p(x) \log p(x), \qquad x = s, \{t_i\}$$

$$I[s,\{t_i\}] = \sum_{s\{t_i\}} p(s,\{t_i\}) \log \frac{p(s,\{t_i\})}{p(s)\,p(\{t_i\})}$$

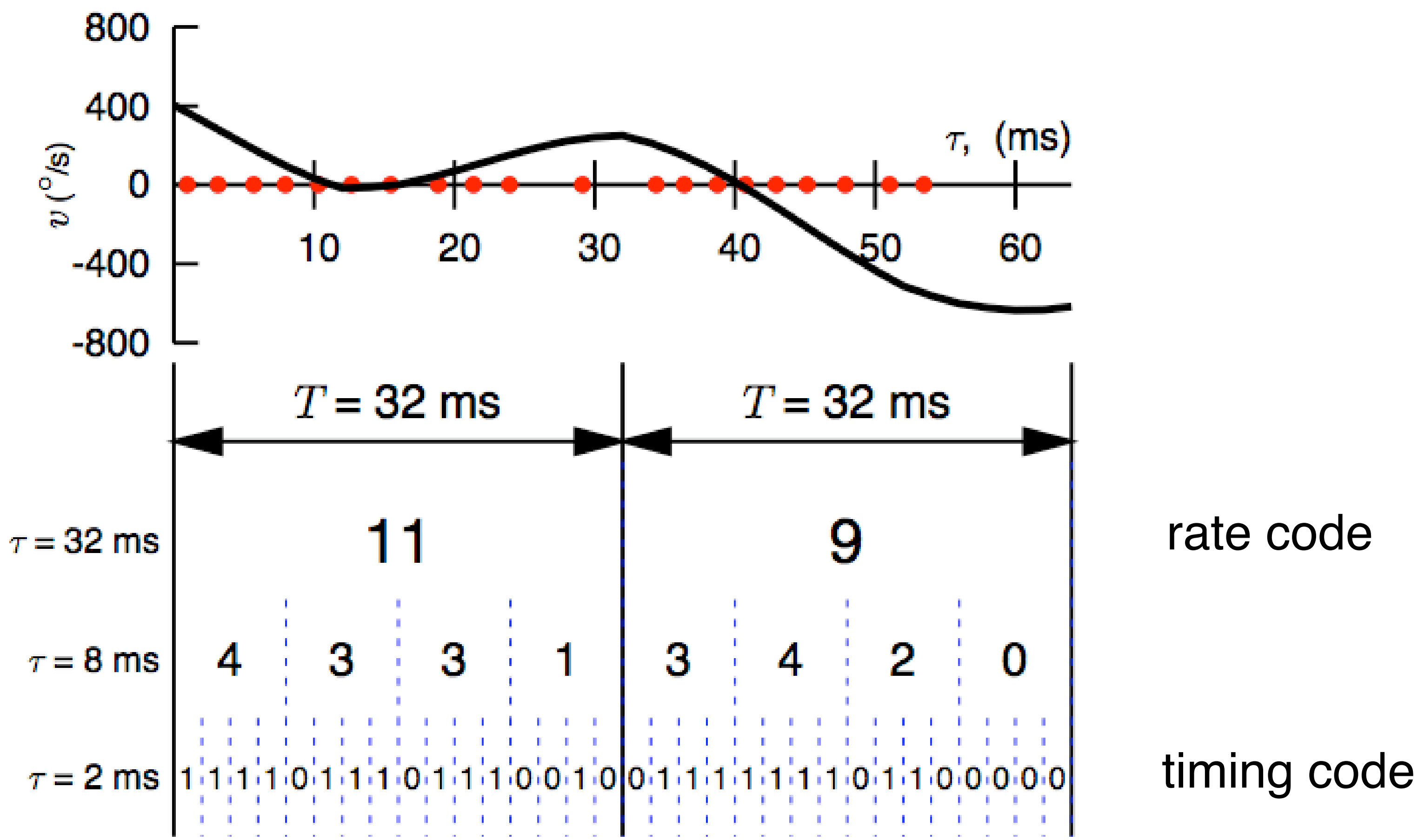
- Captures all dependencies (zero *iff* joint probabilities factorize)
- Reparameterization invariant
- Unique metric-independent measure of "how related"

Experiment design



(Strong et al., 1998)

Experiment design:
probing precise spike timing



(Strong et al., 1998)

Problems

- Total of about 10-15 min of recordings (limited by stationarity of the outside world)
- At most 200 repetitions
- Stimulus correlation of 60ms: only 10000 independent samples (repeated or nonrepeated)
- Need to sample words of length 30 ms (behavioral) to 60 ms (stimulus) at resolution down to 0.2 ms (binary words of length up to ~100).

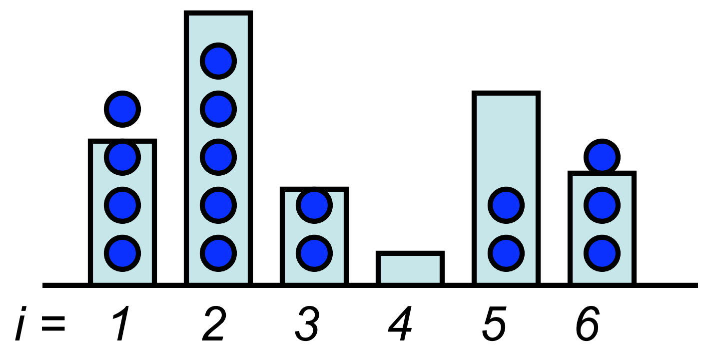
Undersampling!

Undersampling and entropy/MI estimation

Maximum likelihood estimation:

$$p_i, \quad i = 1 \ldots K$$

(K - # of bins)



$$p_i^{ML} = \frac{n_i}{N}$$

(N - sample size)

$$S_{ML} = -\sum_i \frac{n_i}{N} \log \frac{n_i}{N}$$

!

$$\langle S_{ML} \rangle \le -\sum_i \frac{\langle n_i \rangle}{N} \log \frac{\langle n_i \rangle}{N} = S$$

Undersampling and entropy/MI estimation

$$\langle S_{ML} \rangle \leq -\sum_i \frac{\langle n_i \rangle}{N} \log \frac{\langle n_i \rangle}{N} = S$$

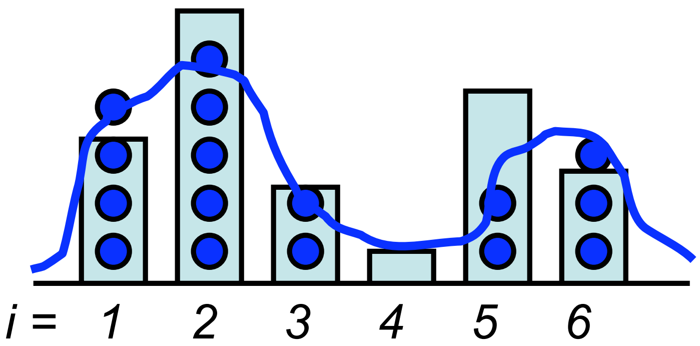
log K

$$\text{bias} \propto -\frac{2^S}{N} \gg (\text{variance})^{1/2} \propto \frac{1}{\sqrt{N}}$$

Fluctuations underestimate entropies and overestimate mutual informations.

(Need “smoothing”)

Correct smoothing possible



$$S \leq \log N$$

Incorrect smoothing -- over- or underestimation.

13 bits for NR, 6-7 bits for R

Even refractory Poisson process at this T, τ has over 15-20 bits of entropy!

For estimation of entropy at $K / N \leq 1$ see:
Grassberger 1989, 2003, Antos and Kontoyiannins 2002, Wyner and Foster 2003, Batu et al. 2002, Paninski 2003, Panzeri and Treves 1996, Strong et al. 1998

What if *S*>log*N* ?

But there is hope (Ma, 1981):

For uniform *K*-bin distribution the first coincidence occurs at

$$N_c \sim \sqrt{K} = \sqrt{2^S}$$

$$S \sim 2 \log N_c$$

Time of first coincidence

Can make estimates for square-root-fewer samples!

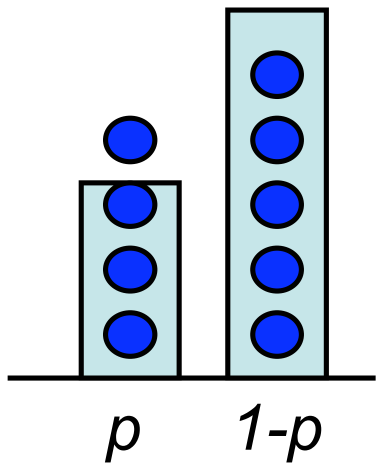
Can this be extended to nonuniform cases?

- Assumptions needed (won't work always)
- Estimate entropies without estimating distributions.

Generalizing Ma: What is unknown?

Binomial distribution:

$$S = -p \log p - (1-p)\log(1-p)$$

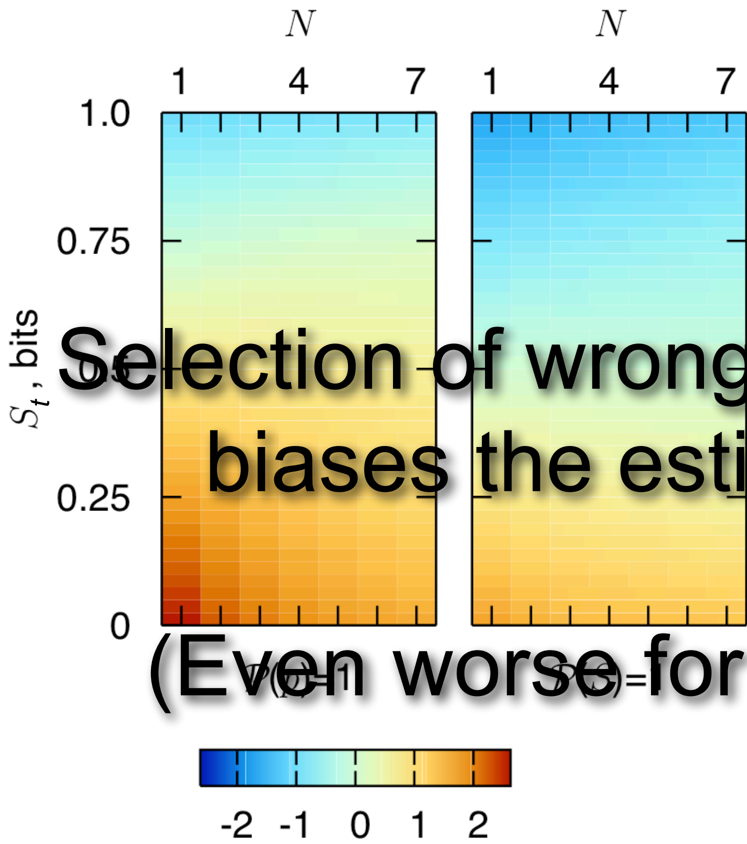


Assume (Bayes)

uniform (no assumptions)

p S

What is unknown?



$$\varepsilon = \left\langle \frac{S_{est} - S_{true}}{\delta S_{est}} \right\rangle$$

Selection of wrong "unknown" biases the estimation.

(Even worse for large *K*.)

One possible uniformization strategy for *S* (NSB)

- Posterior variance scales as $1/\sqrt{N}$
- Little bias, except in some known cases.
- Counts coincidences and works in Ma regime (if works).
- Is guaranteed correct for large *N*.
- Allows infinite # of bins.

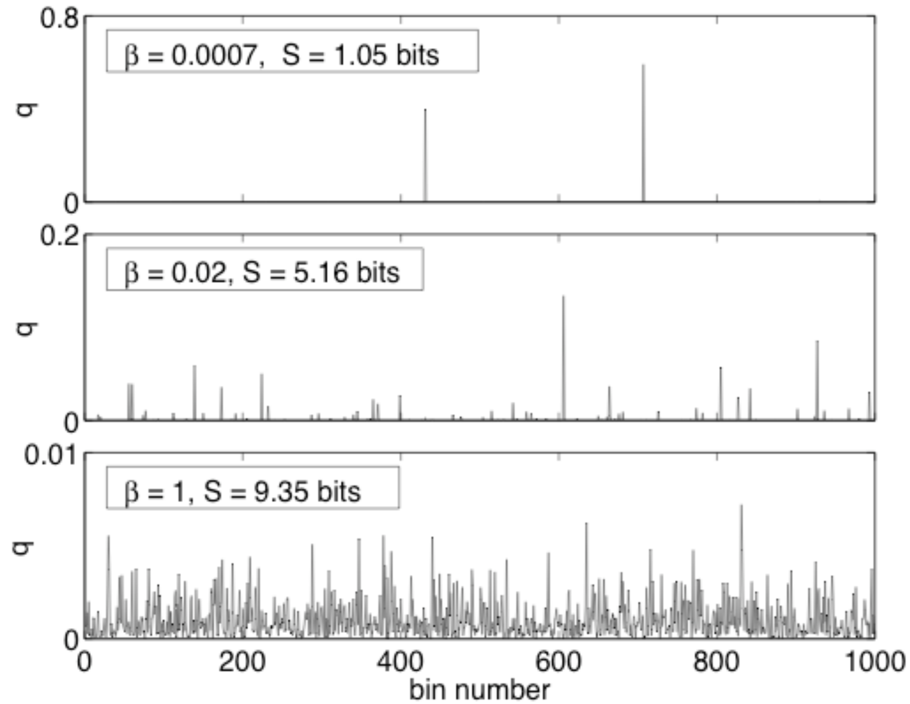
(Nemenman et al. 2002-2007)

For large *K* the problem is extreme (*S* known a priori)

$$P_\beta(\{q_i\}) = \frac{1}{Z(\beta)}\,\delta\!\left(1-\sum_{i=1}^{K} q_i\right)\prod_{i=1}^{K} q_i^{\beta-1}$$

Dirichlet priors, a.k.a., adding pseudocounts (include the uniform prior, the ML prior, and others).

Inference is analytic

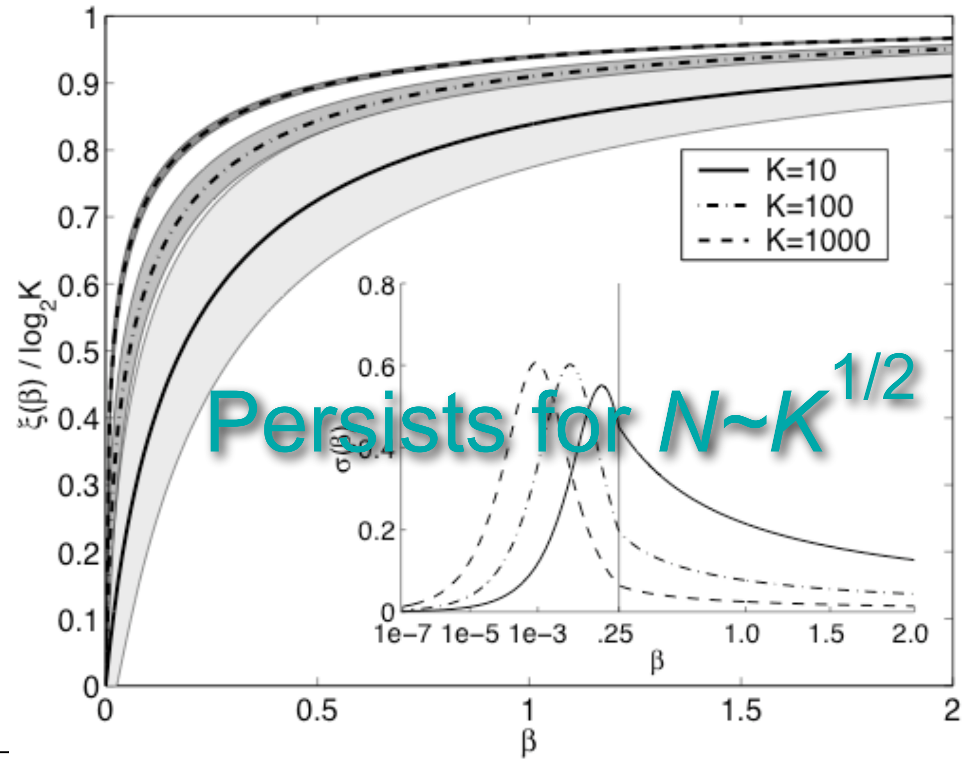


For large *K* the problem is extreme (*S* known a priori)

$$P_\beta(\{q_i\}) = \frac{1}{Z(\beta)}\, \delta\left(1 - \sum_{i=1}^{K} q_i\right) \prod_{i=1}^{K} q_i^{\beta-1}$$

Dirichlet priors, a.k.a., adding pseudocounts (include the uniform prior, the ML prior, and others).

Inference is analytic.



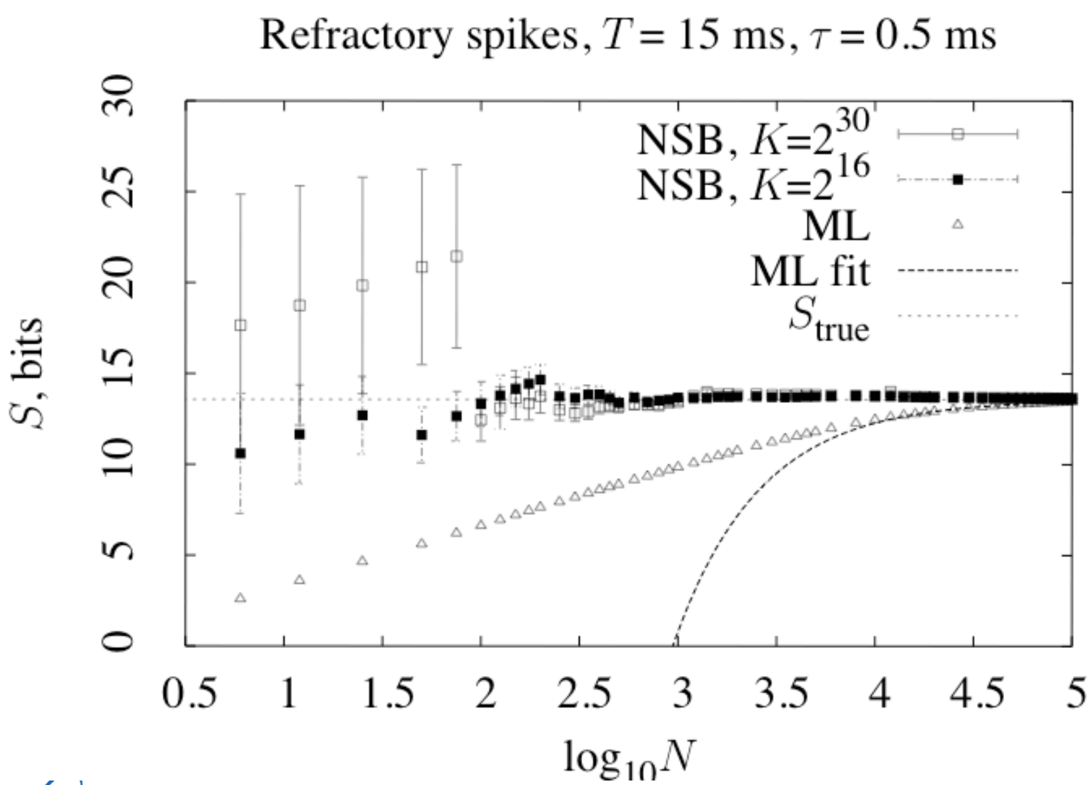
Uniformize on *S*

$$P_\beta(\{q_i\},\beta) = \frac{1}{Z}\,\delta\left(1-\sum_{i=1}^{K} q_i\right)\prod_{i=1}^{K} q_i^{\beta}\,\left.\frac{dS}{d\beta}\right|_{N=0} P(S|_{N=0})$$

- A delta-function sliding along the a priori entropy expectation.
- This is also Bayesian model selection (small β large phase space)
- Have error bars (dominated by posterior variance in β, not at fixed β).

Synthetic test (same for natural data)

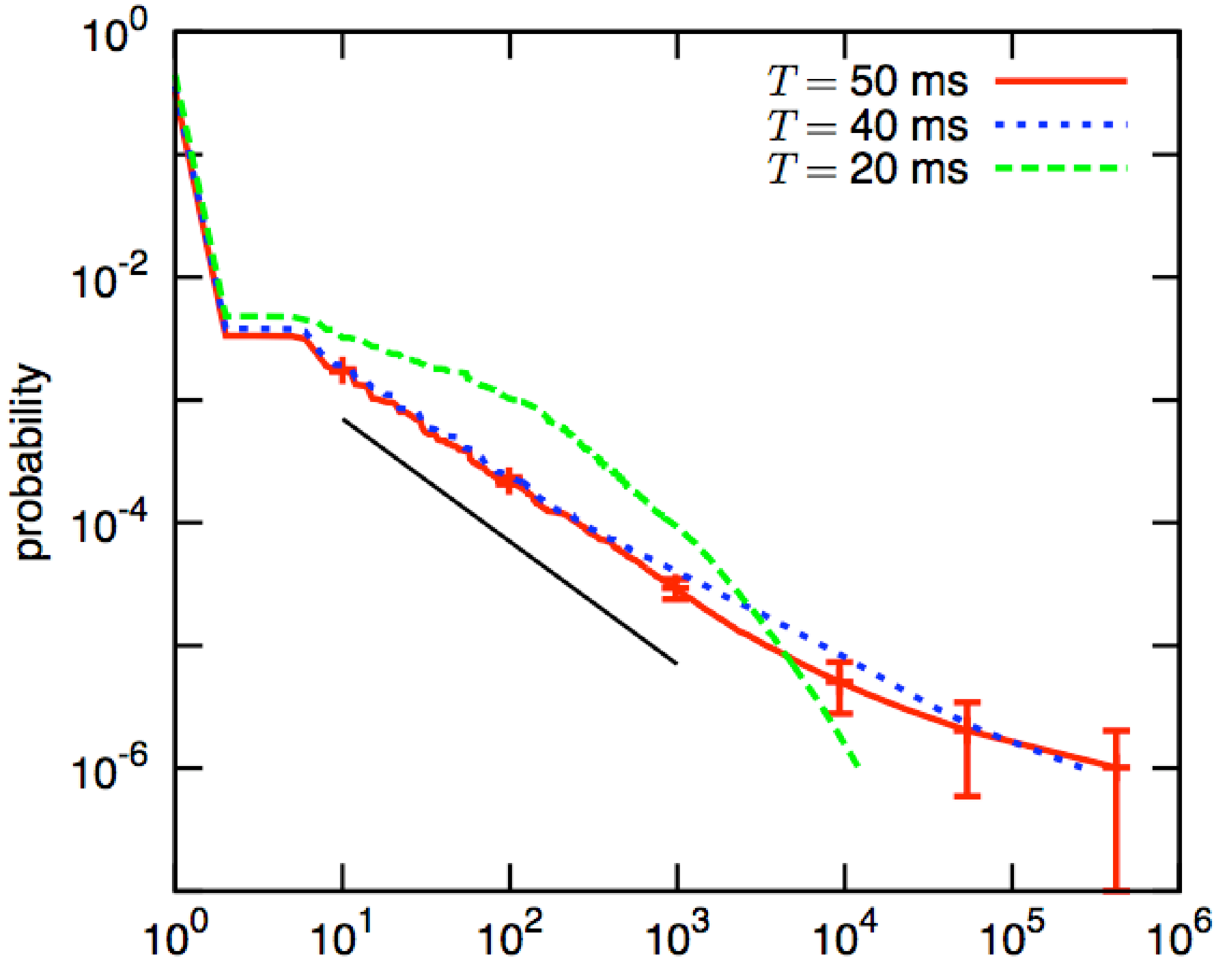
Refractory Poisson, rate 0.26 spikes/ms, refractory period 1.8 ms, T=15ms, discretization 0.5ms, true entropy 13.57 bits.



- Estimator is unbiased if consistent and self-consistent.
- Always do this check.

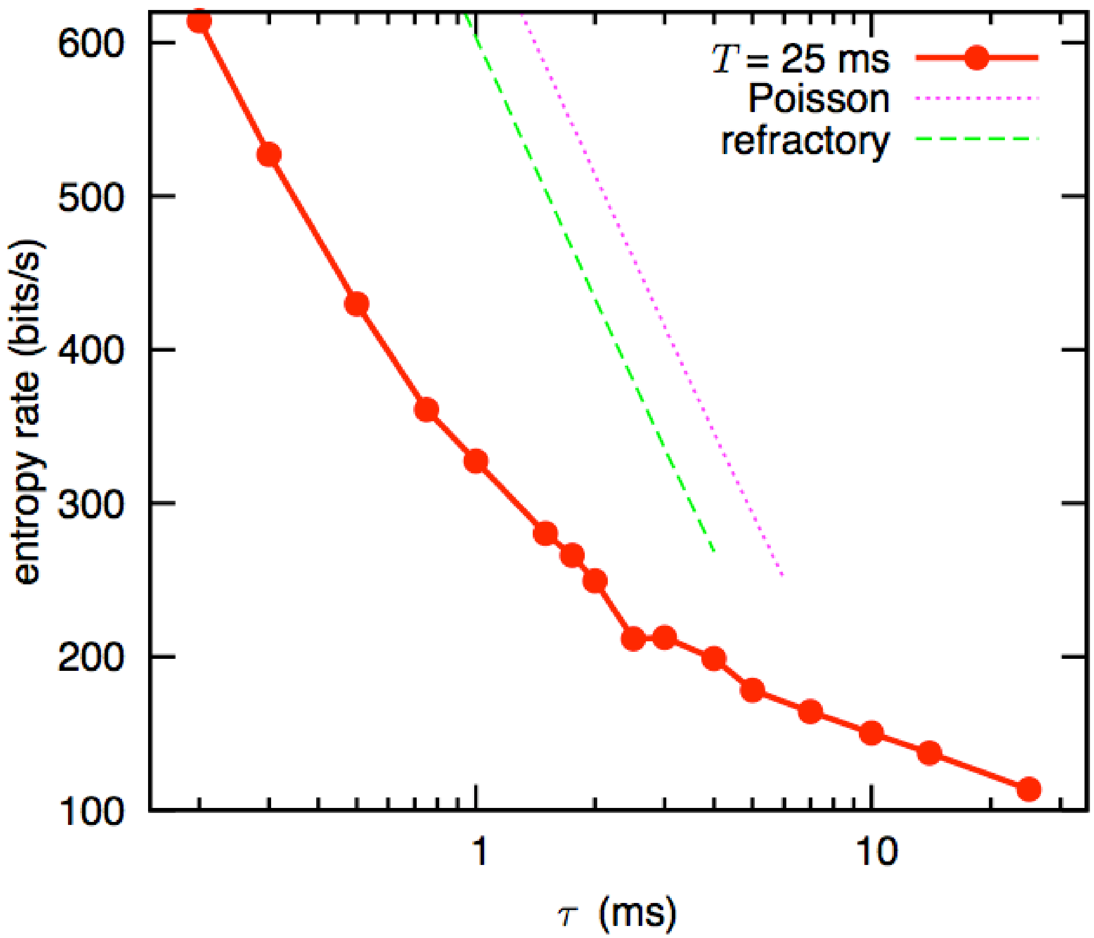
(Nemenman et al. 2004)

Exploring the code: long tails



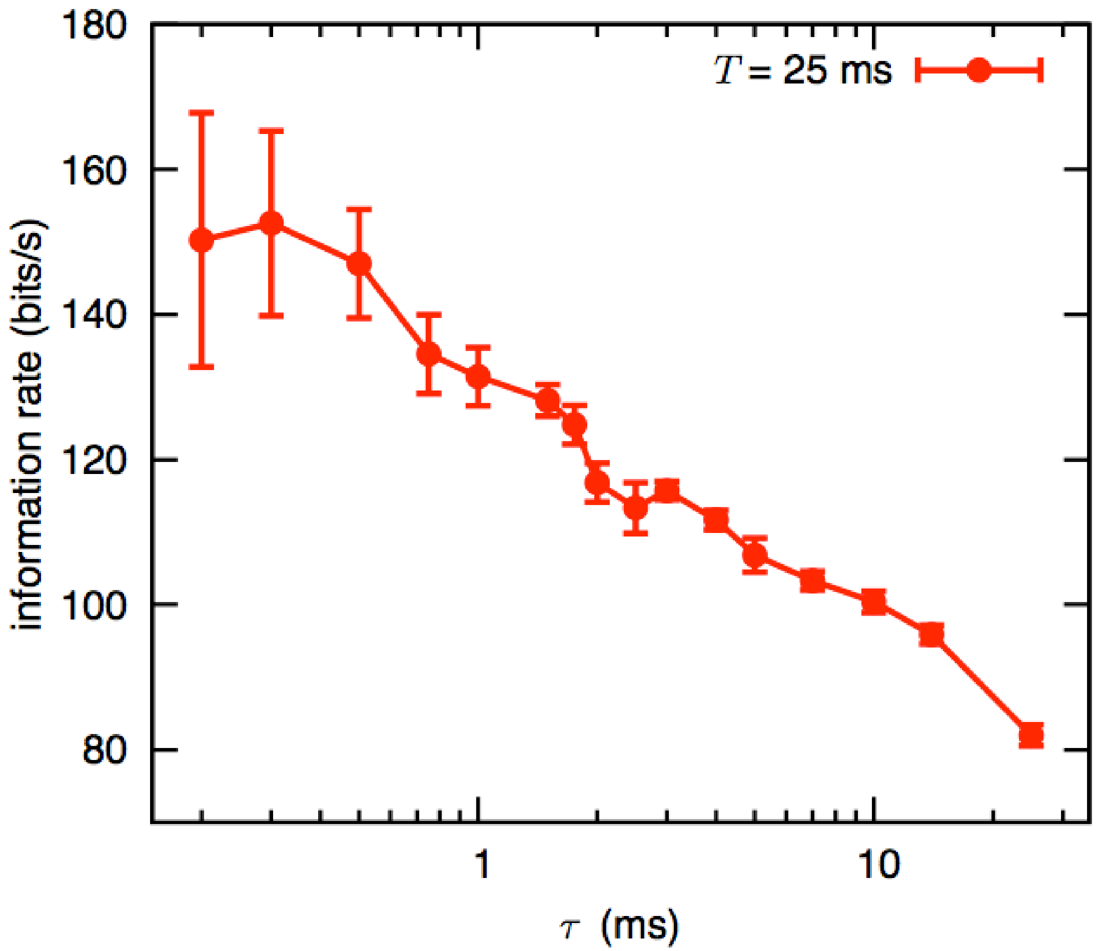
This causes complications

Exploring the code: high entropy



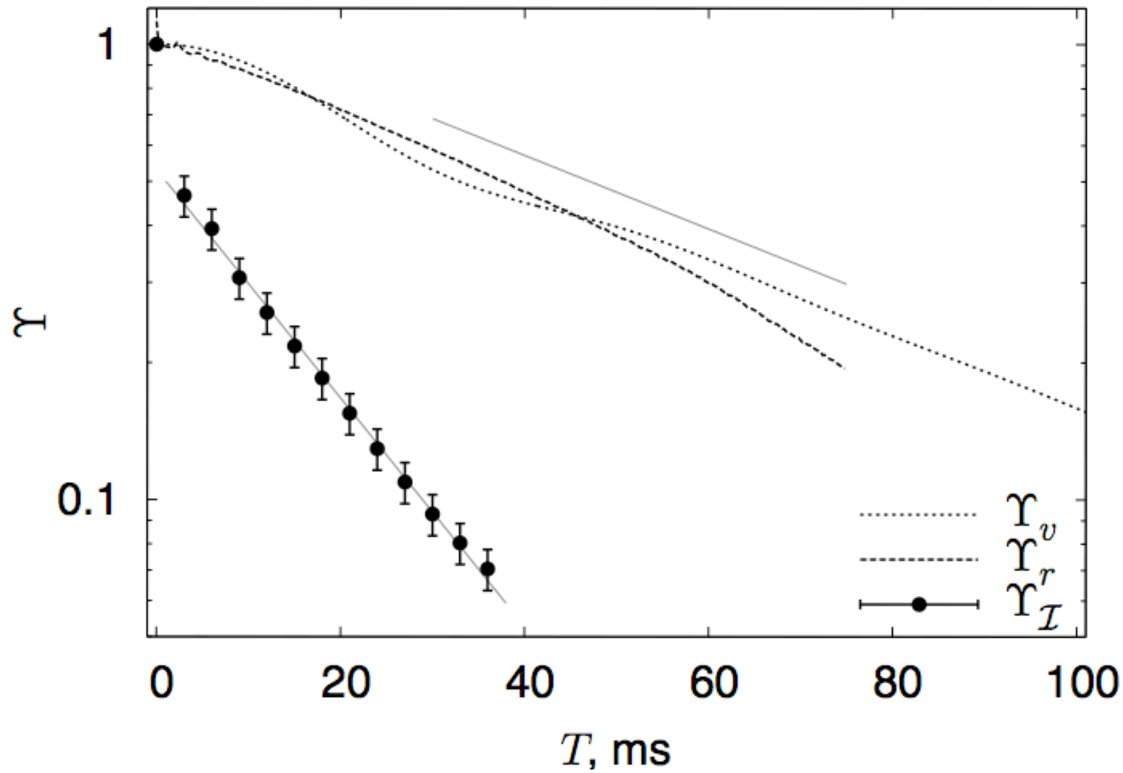
But is all this entropy used in coding?

Information rate at *T*=25ms



- Rate grows up to τ =0.2-0.3 ms
- 30% more information at τ<1ms.
- ~1 bit/spike at 150 spikes/s and low-entropy correlated stimulus. Design principle?
- 0.2 ms - comparable to channel opening/ closing noise and experimental noise.

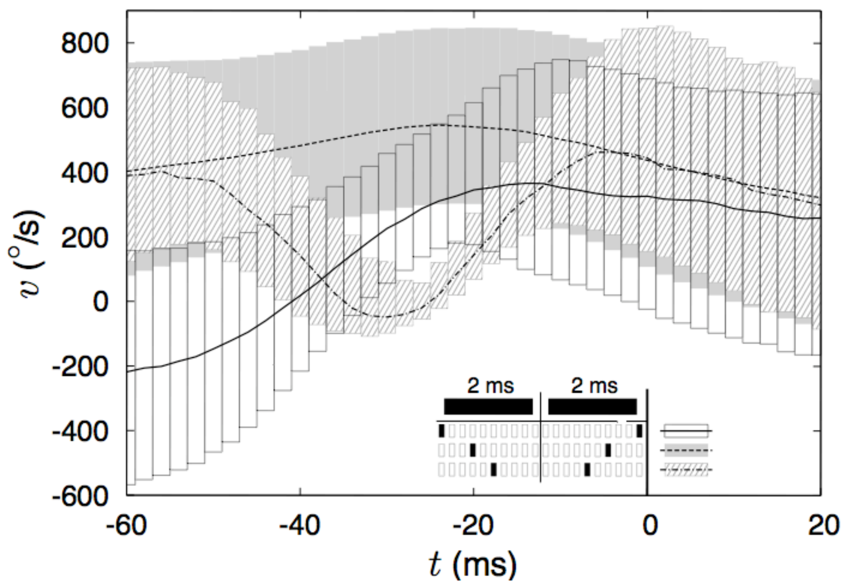
New bits:
Decorrelation in the time domain



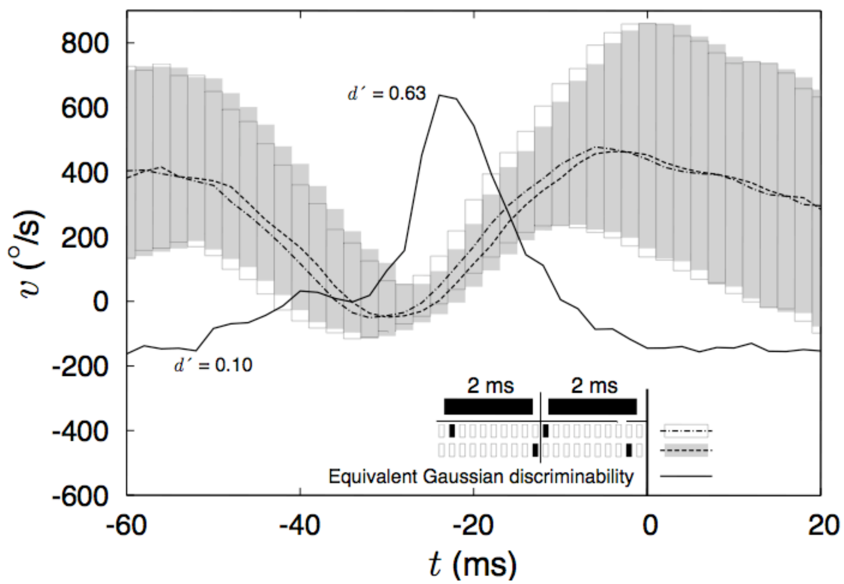
- Corr. func. at half its maximum value (for stimulus and rate), but fly gets new bits every 25 ms
- Not a simple delta-code
- Behaviorally optimized code
- Pretty amazing!

$$\Upsilon = \frac{2I(T) - I(2T)}{I(2T)}$$

Information about…



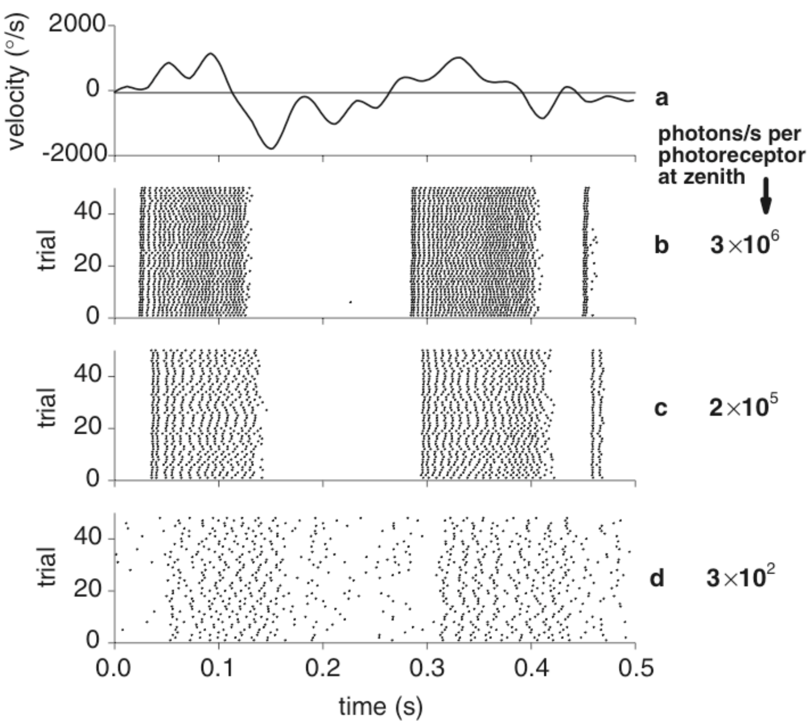
Signal shape



Zero-crossings time

Best estimation at 24 ms delay. Little time for reaction.

Precision is limited by physical noise sources



(Lewen, et al 2001)

$$T = 6 \text{ ms}$$

$$\tau = 0.2 \text{ ms}$$

$$1.1 \cdot 10^6 \text{ ph/(s} \cdot \text{rec)} \pm 3\%$$

$$I^+ - I^- = 0.0204 \pm 0.0108 \text{ bits}$$

$$p = 6\% \text{ (and much smaller)}$$