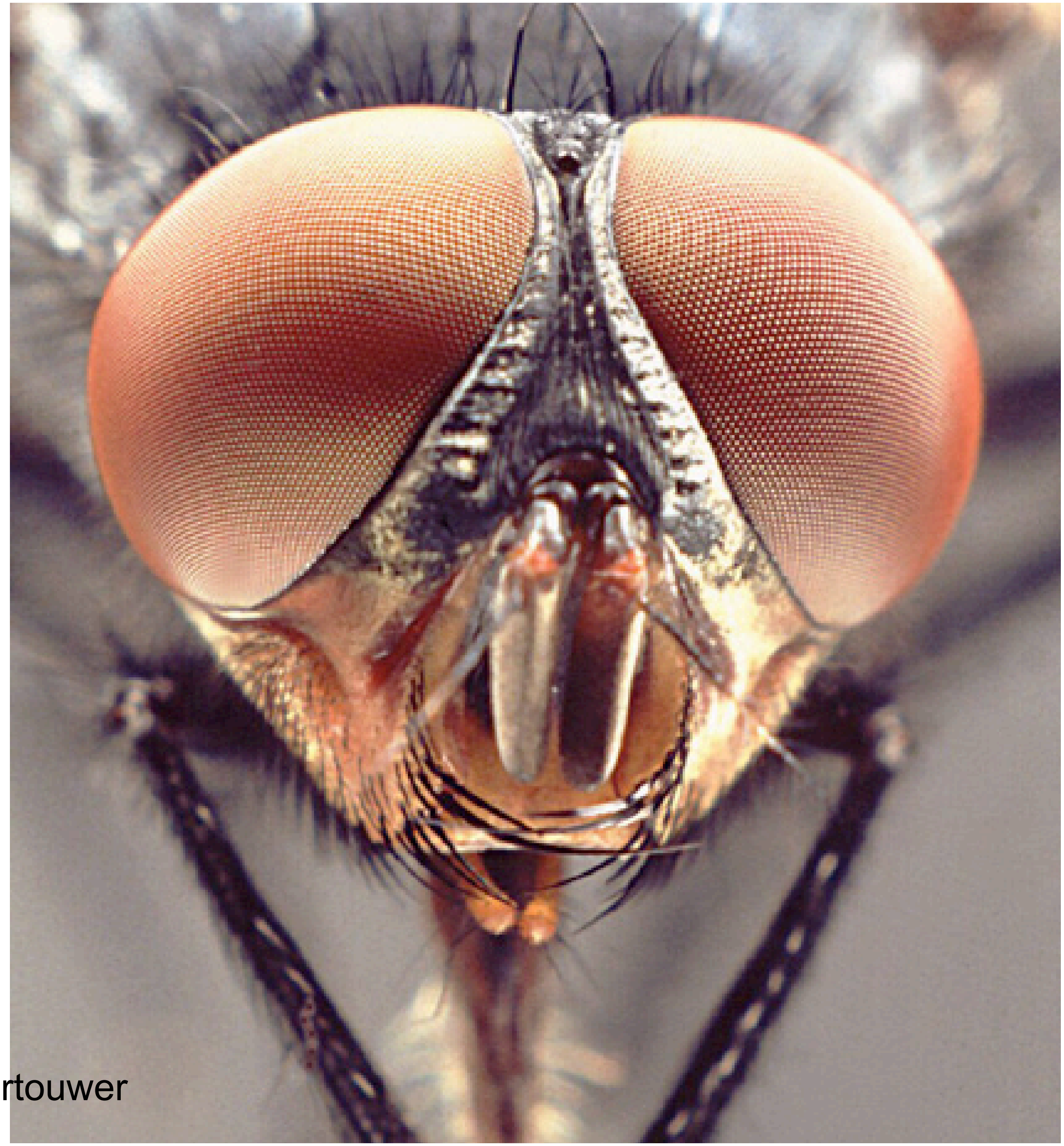
CSHL, 06/10/05

How much does a fly know about its world?

Ilya Nemenman
(JCSB, Columbia)
and
William Bialek (Princeton)
Rob de Ruyter van Steveninck (Indiana)

http://sourceforge.net/projects/nsb-entropy



H. L. Leertouwer

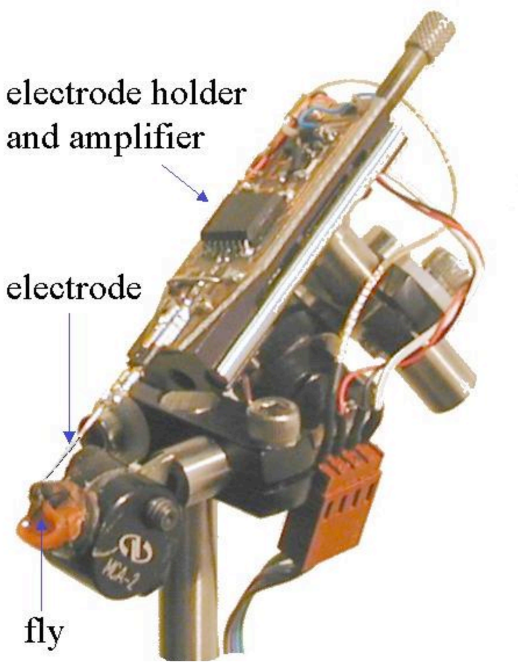
Why fly as a neurocomputing model system?

- 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

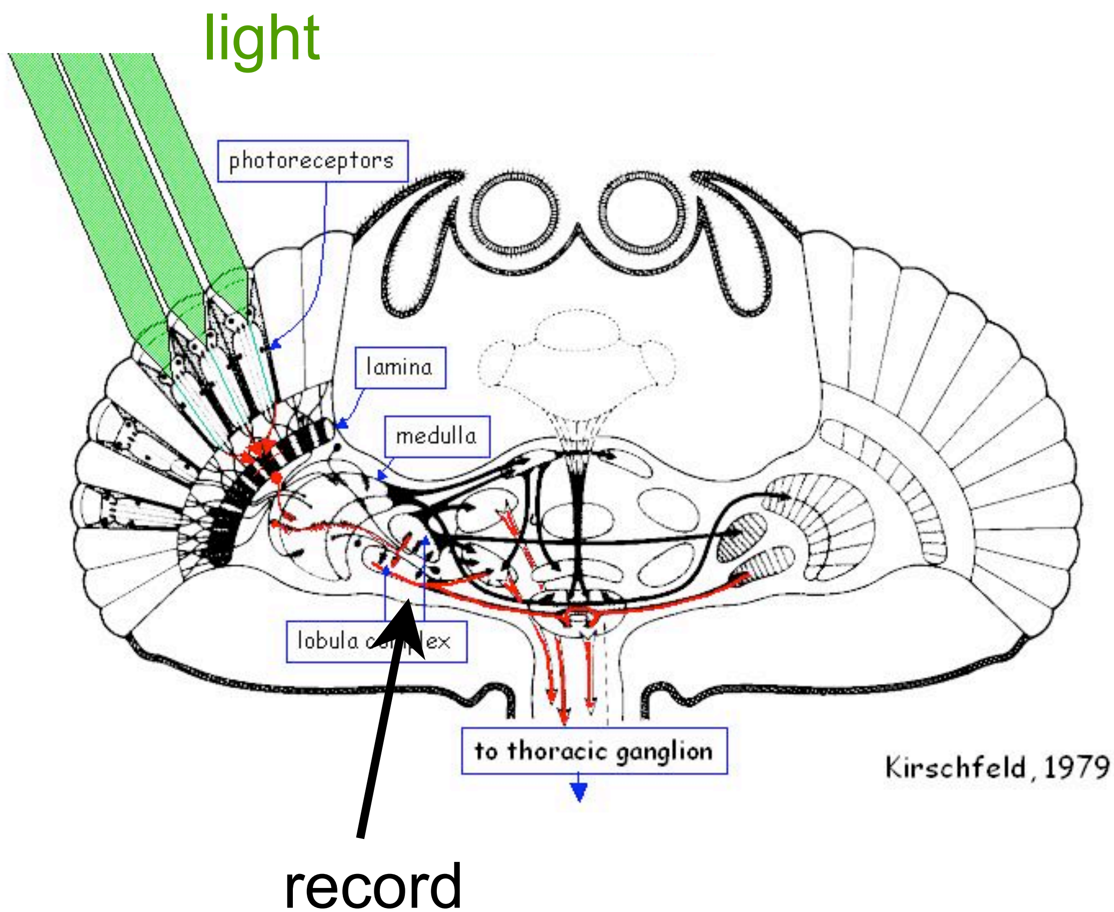
Questions

- Can we understand the code?
- Which features of it are important?
  - Rate of precise timing (how precise)?
  - Synergy between spikes?
- What/how much does the fly know?
- Is there an evidence for optimality?

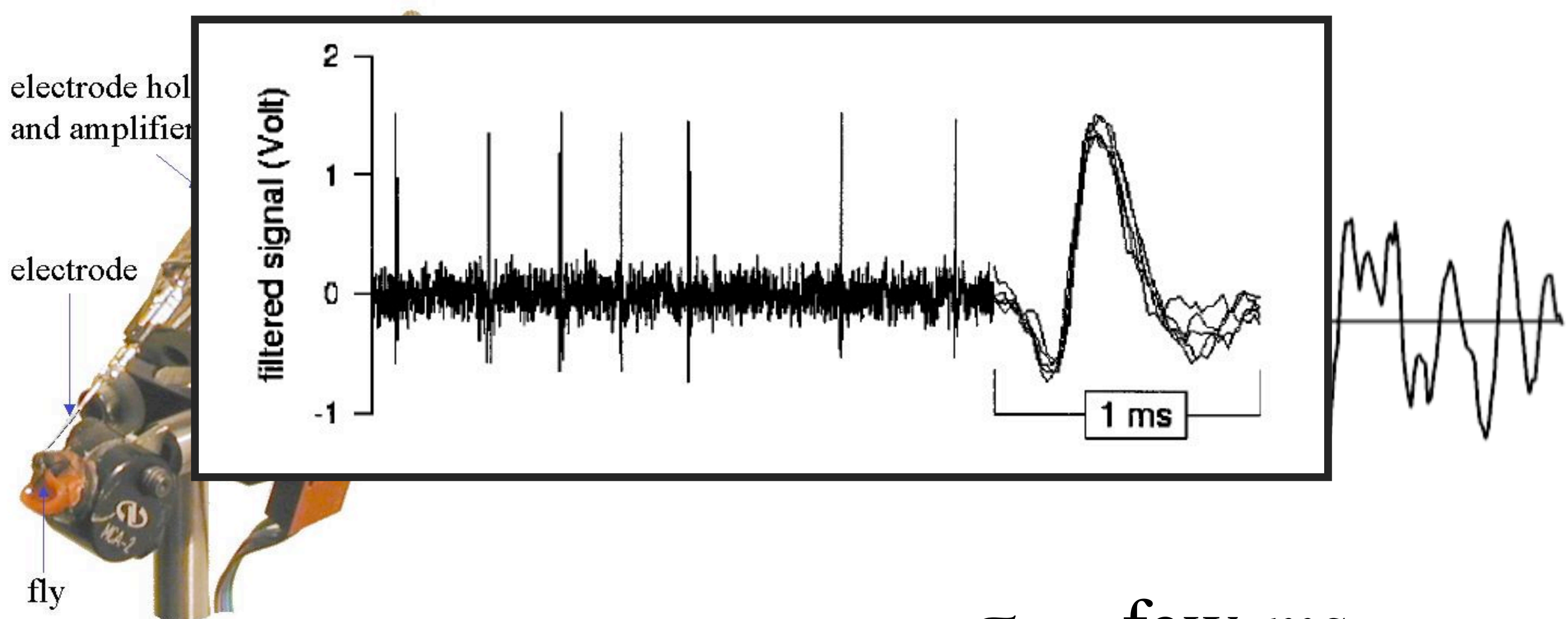
Recording from fly’s H1



(Lewen et al, 2001)



Motion estimation in fly H1



$\tau =$ few ms

(Strong et al., 1998)

Decoding a simple spike train

$$P(t_i \mid s(t)) \sim Poisson[r(s(t_i)]$$

nonlinear

$$P[\{t_i\} \mid s(t)] = \frac{1}{N!} \exp\left[-\int r(s(t))\, dt\right] \prod_{i=1}^{N} r(s(t_i))$$

$$P[s(t)] \propto \exp\left[-\frac{1}{4\tau_c} \int dt \left(\tau_c^2 \dot{s}^2 + s^2\right)\right]$$

$$s_{est}(t_0) = \int [ds]\ P[s(t) \mid \{t_i\}]\ s(t_0) = \int [ds]\ \frac{\boldsymbol{P[\{t_i\} \mid s]P[s]}}{\boldsymbol{Z}}\ s(t_0)$$

(Bialek, Zee, 1990)

Linear decoding for sparse spikes (cluster expansion)

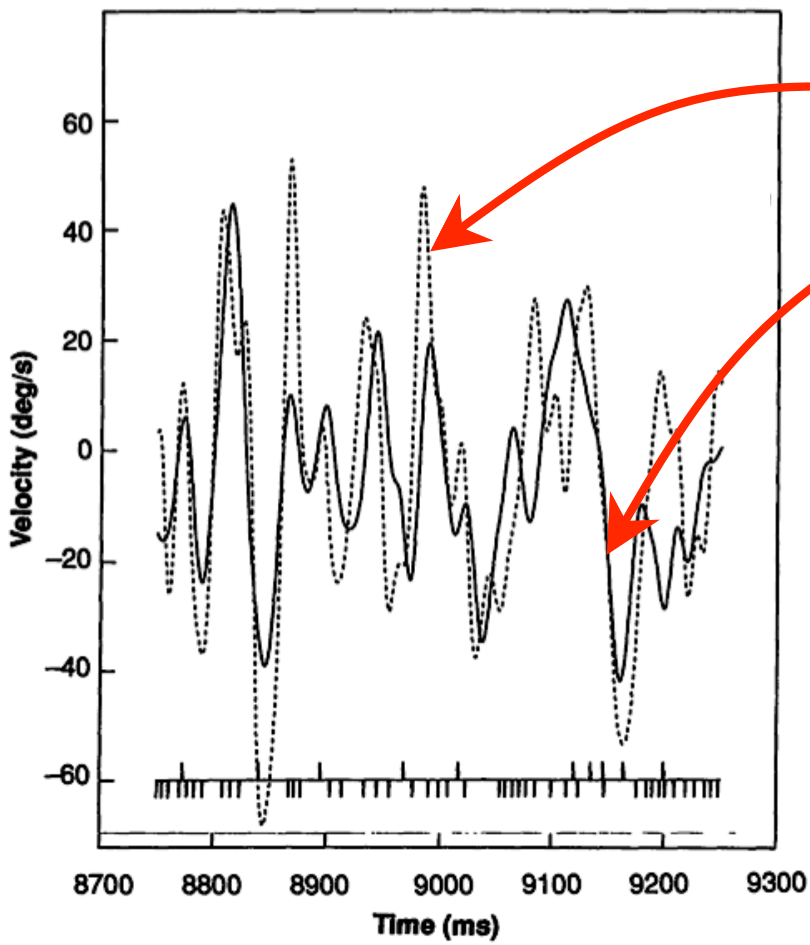
$$s_{est}(t_0) = \frac{\left\langle s(t_0)\prod_{i=1}^{N} r(s(t_i)) \right\rangle_{prior}}{\left\langle \prod_{i=1}^{N} r(s(t_i)) \right\rangle_{prior}}$$

Stimulus couples spikes; but the strength of the coupling drops with $\sim (t_i - t_{i+1})/\tau$ (very fast varying mean field)

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

(Bialek, Zee, 1990)

Linear decoding



stimulus

reconstruction

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

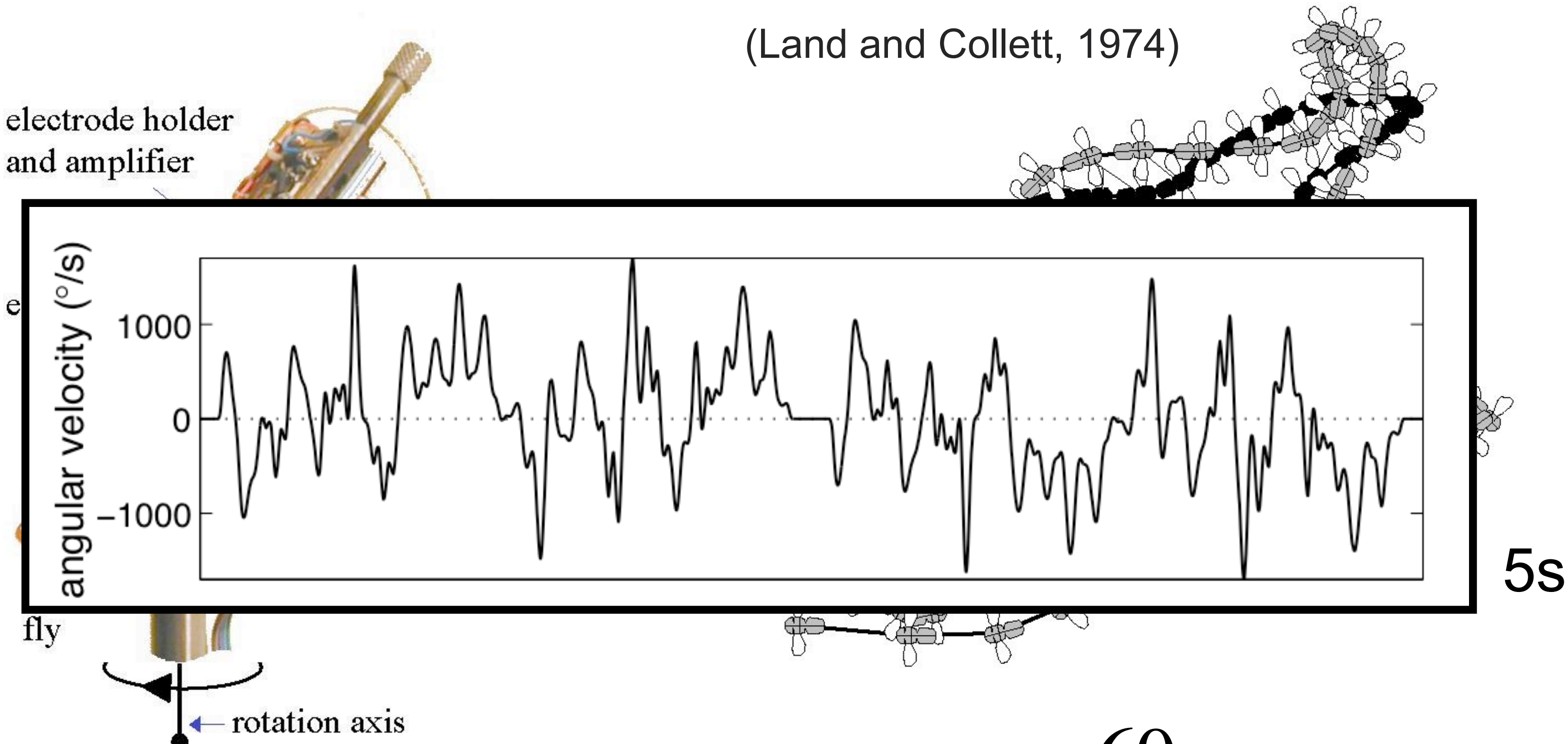
Position of each spike within ~2ms matters!

But what if …

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

Natural stimuli

(Land and Collett, 1974)



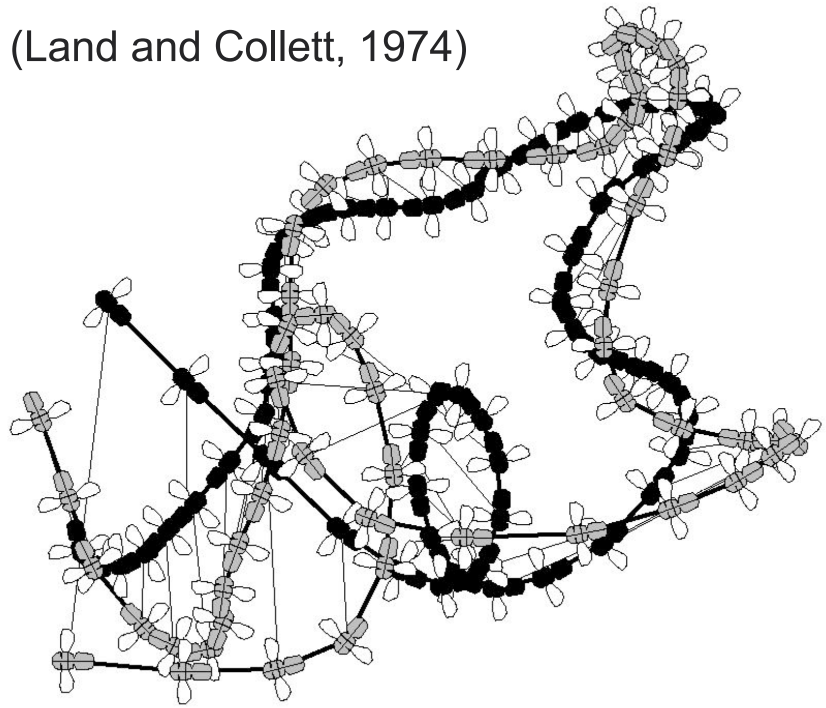
(Lewen et al, 2001)

$$\tau = 60ms$$
$$\text{response} = 30ms$$

Natural stimuli

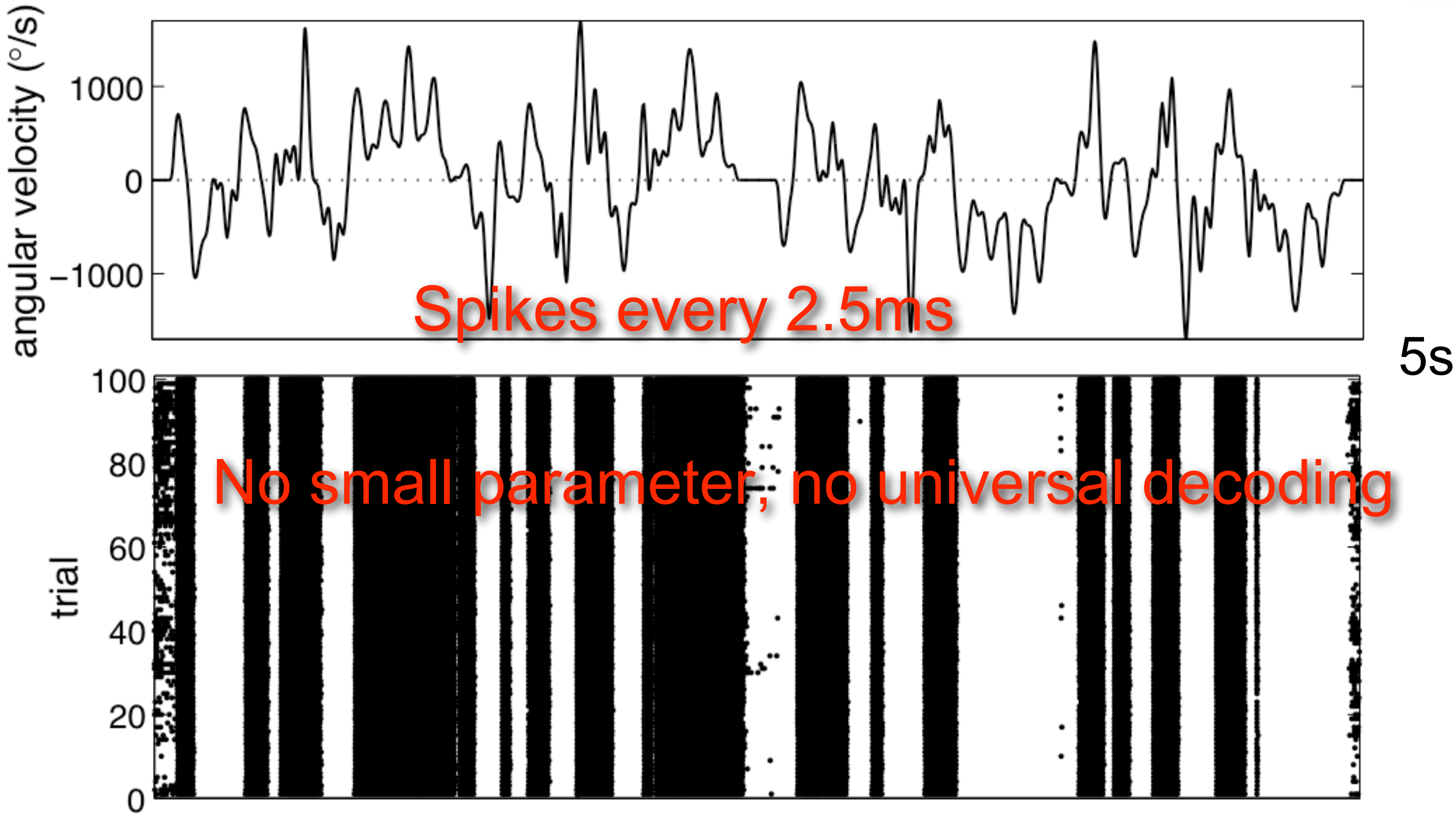
- ~2 ms resolution known to be important for white noise stimuli
- Could such “brisk” spikes be due to ~1 ms correlations in stimulus?
- What if stimulus has natural correlations?

(Land and Collett, 1974)

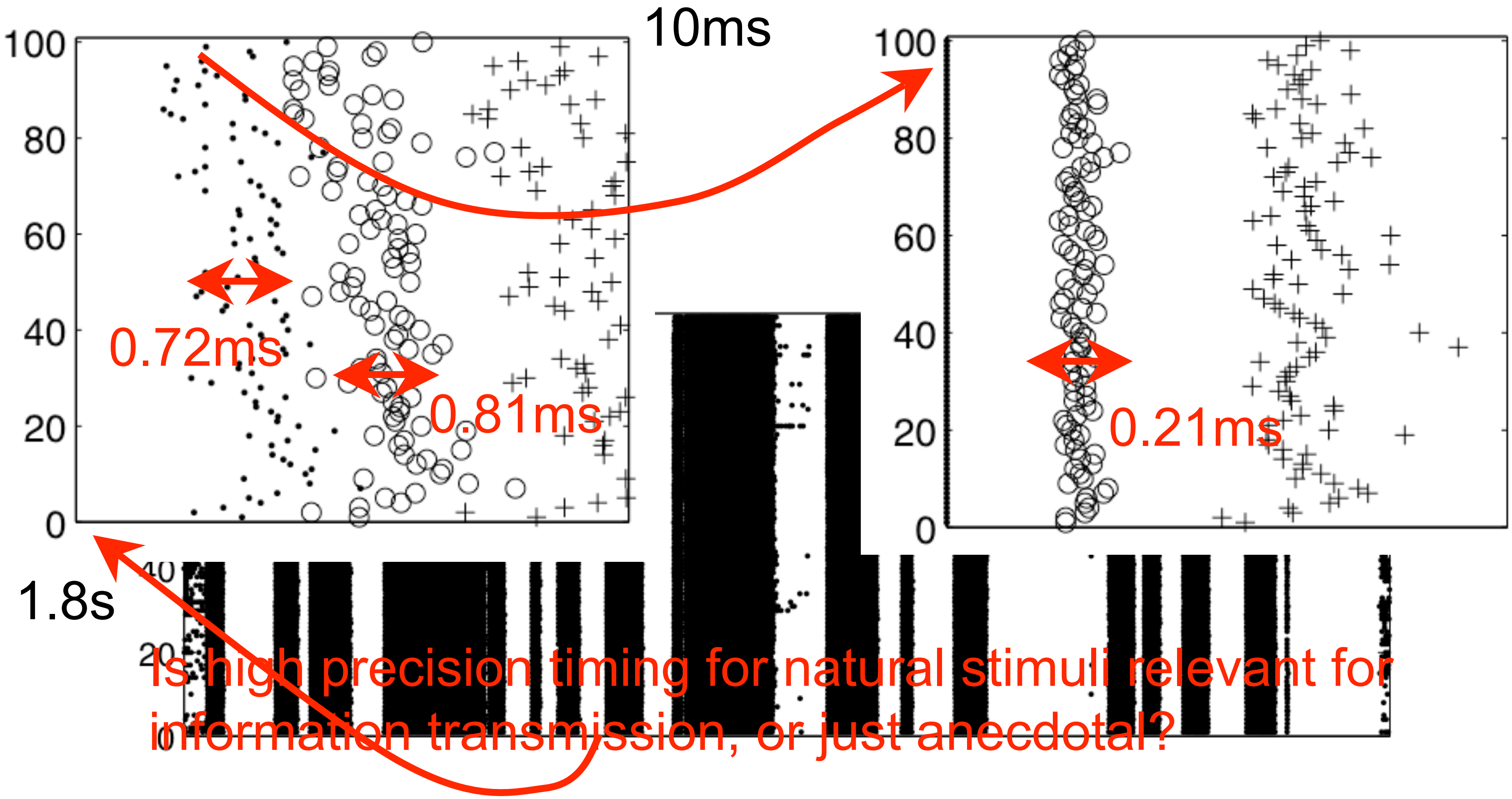


$$\tau = 60ms$$
$$\text{response} = 30ms$$

Natural stimulus and response



Highly repeatable spikes (not rate coding)



Is high precision timing for natural stimuli relevant for information transmission, or just anecdotal?

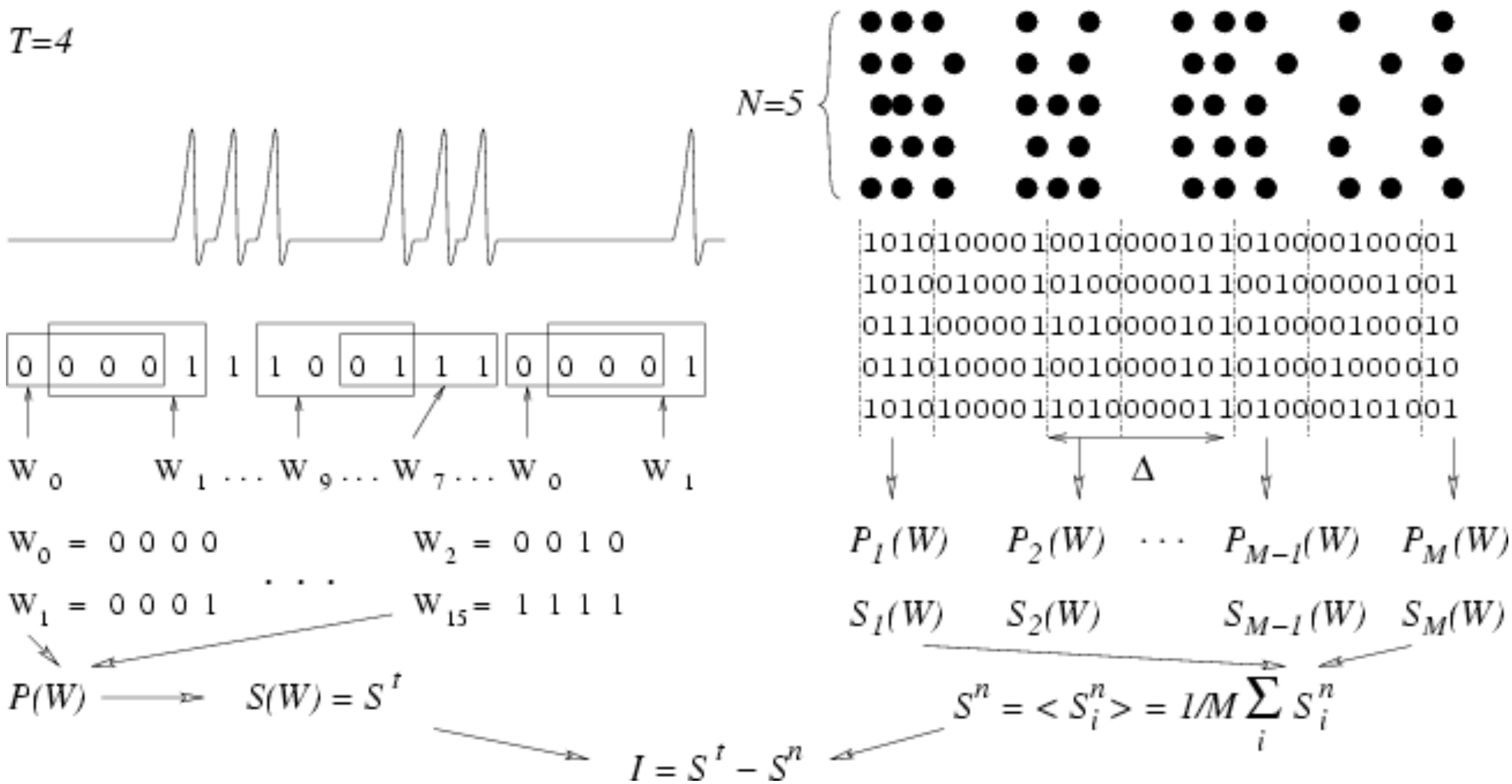
How to characterize coding without an explicit decoding ?

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

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

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

Experiment design



(Strong et al., 1998)

Problems

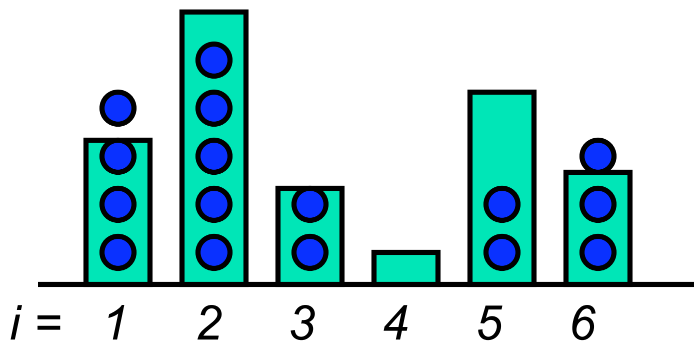
- Total of about 10-15 min of recordings (limited by stationarity of the outside world)
- At most 200 repetitions
- Stimulus correlation of 60ms: only 10000 independent samples (repeated or nonrepeated)
- Need to sample words of length 30 ms (behavioral) to 60 ms (stimulus) at resolution down to 0.2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

Undersampling 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)

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

!

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

Undersampling and entropy/MI estimation

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

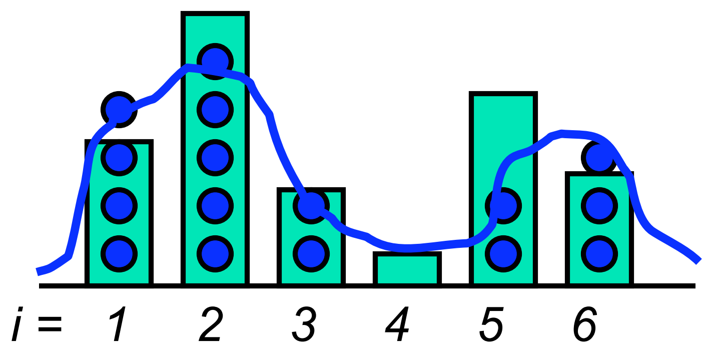
log *K*

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

Fluctuations underestimate entropies and overestimate mutual informations.

(Need smoothing.)

Correct smoothing possible



$$S \le \log N$$

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 $K/N \le 1$ 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 for

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

$$S \sim 2 \log N_c$$

Time of first coincidence

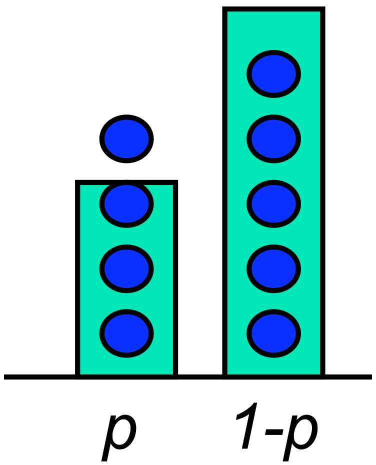
Can make estimates for square-root-fewer samples!
Can this be extended to nonuniform cases?

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

What is unknown?

Binomial distribution:

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

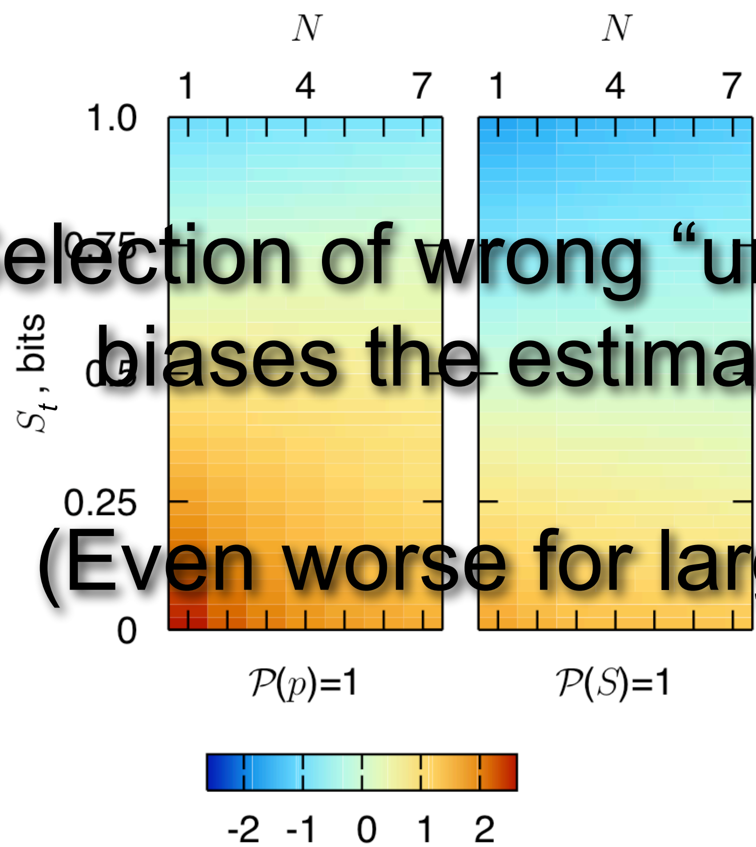


Assume (Bayes)

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*.)

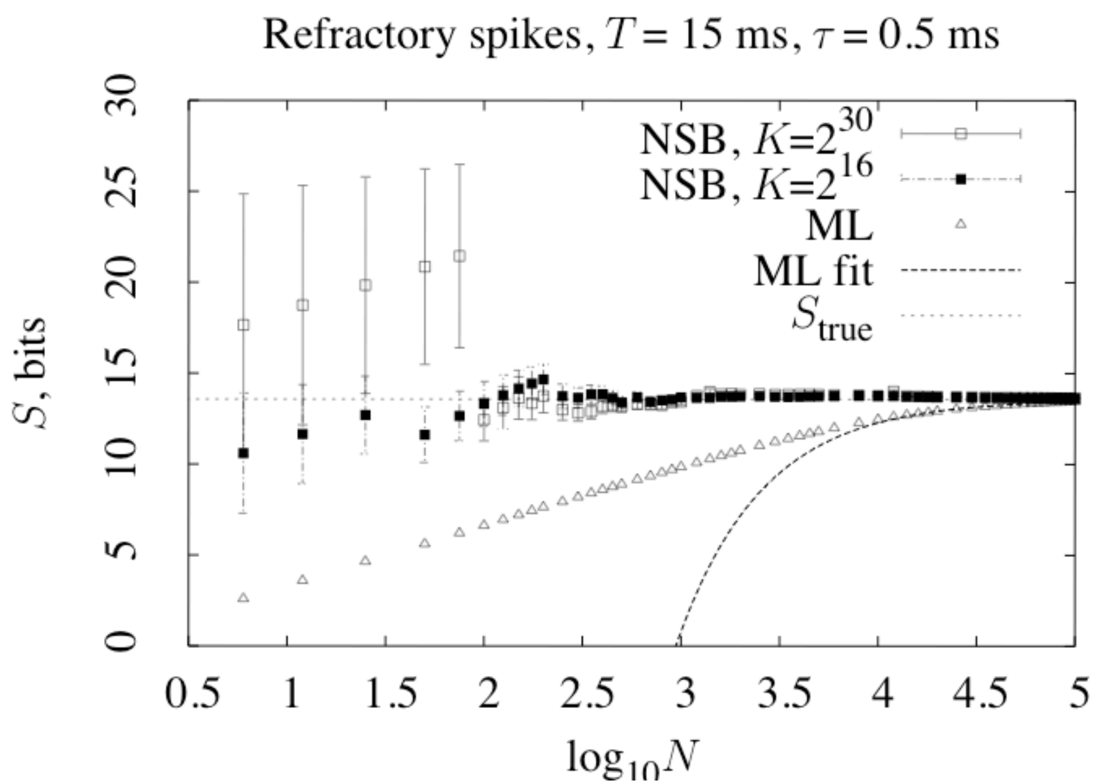
One possible uniformization strategy for *S* (NSB)

- 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, Nemenman 2003)

Synthetic test

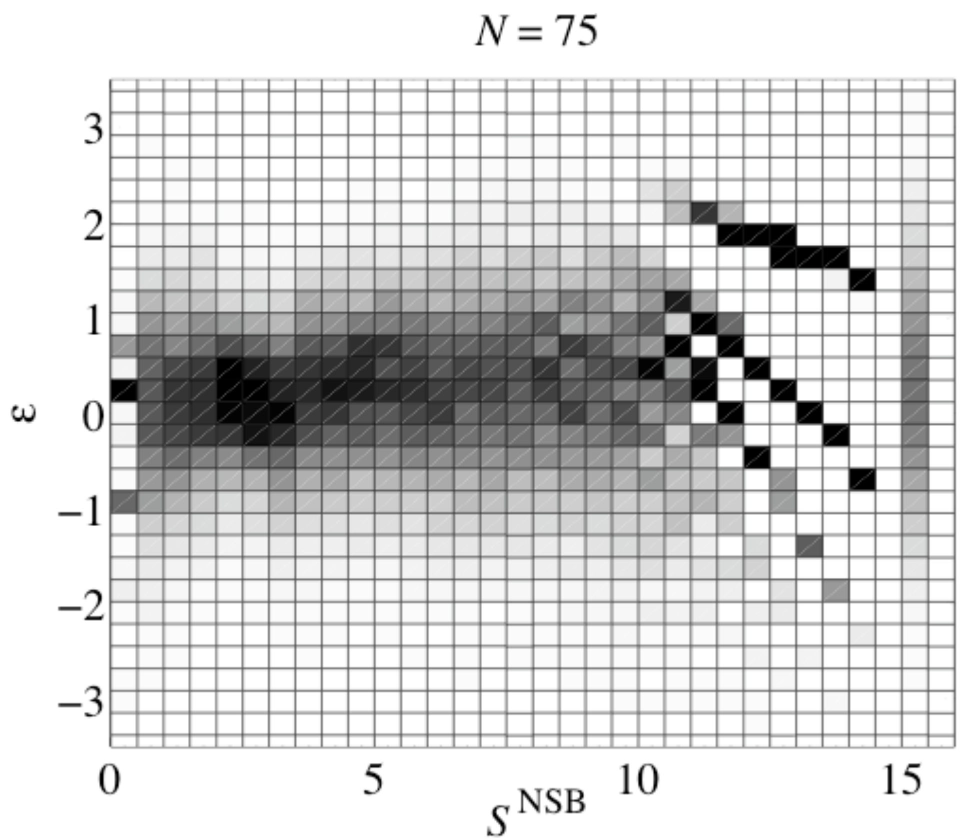
Refractory Poisson, rate 0.26 spikes/ms, refractory period 1.8 ms, T=15ms, discretization 0.5ms, true entropy 13.57 bits.



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

(Nemenman et al. 2004)

Natural data (all *S*)



$$\varepsilon = \frac{S^{NSB}(N) - S}{\delta S^{NSB}(N)}$$

$$\approx \frac{S^{NSB}(N) - S(N = \max)}{\delta S^{NSB}(N)}$$

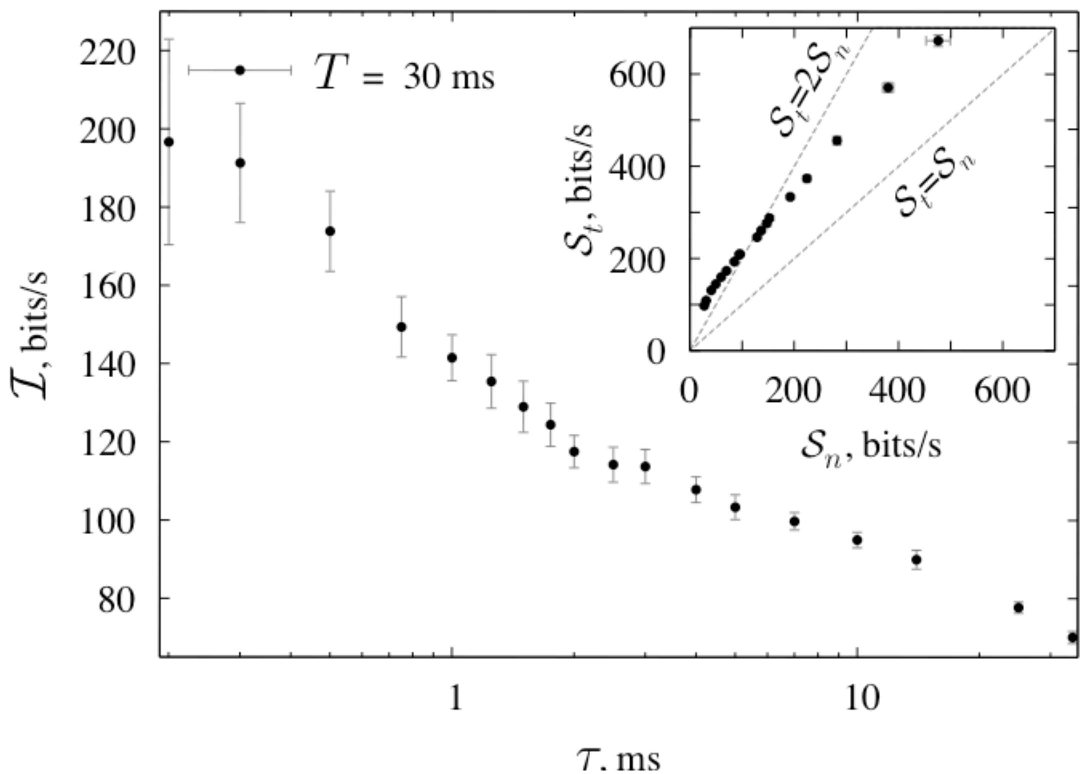
Max=196 repeats

(Nemenman et al. 2004)

Neural code: What remains hidden?

- Given entropy of slices, find the mean noise entropy with error bars (slice entropies are correlated and bimodal).
- Samples for total entropy are also correlated and have long tailed Zipf plots.
- For very fine discretizations and T~30ms need extrapolation.

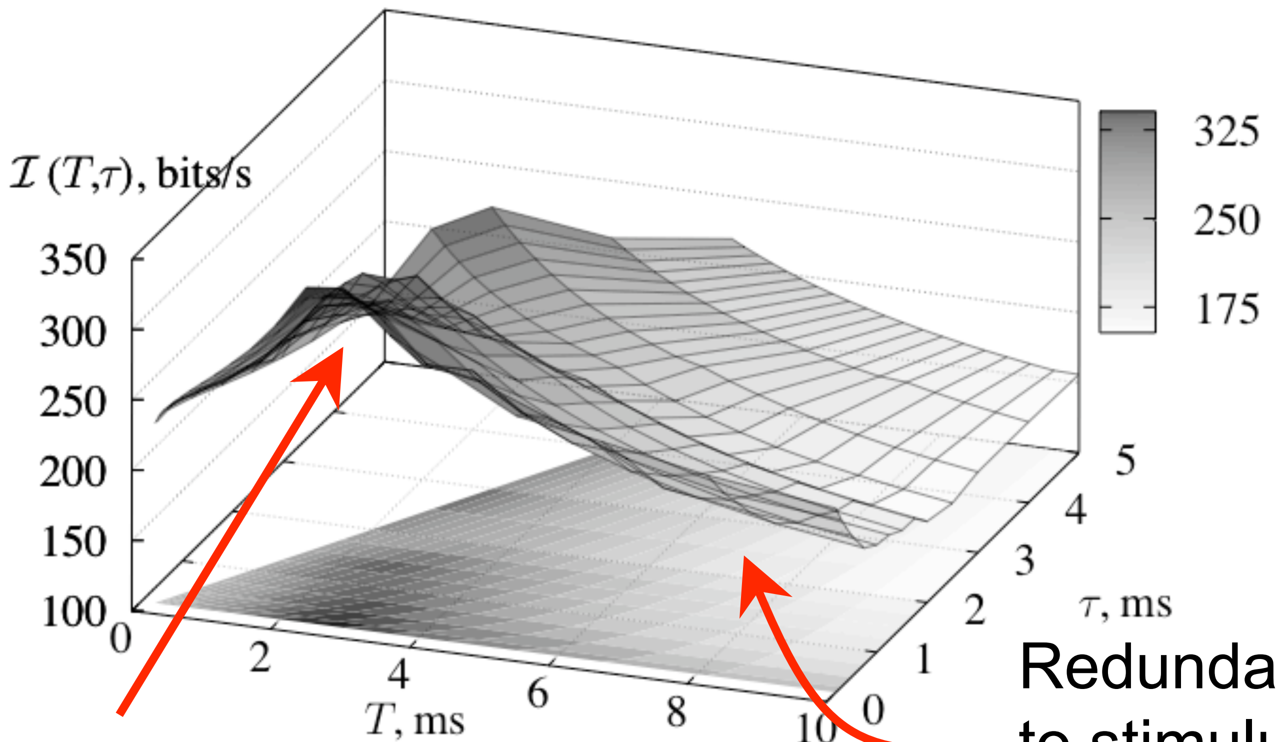
Information rate at T=30ms



0.2 ms -- comparable to channel opening/ closing noise and experimental noise.

- Information present up to τ =0.2-0.3 ms
- 30% more information at τ<1ms. Encoding by refractoriness?
- ~1 bit/spike at 170 spikes/s and low-entropy correlated stimulus. Design principle?
- Efficiency >50% for τ >1ms, and ~75% at 30ms. Optimized for natural statistics?

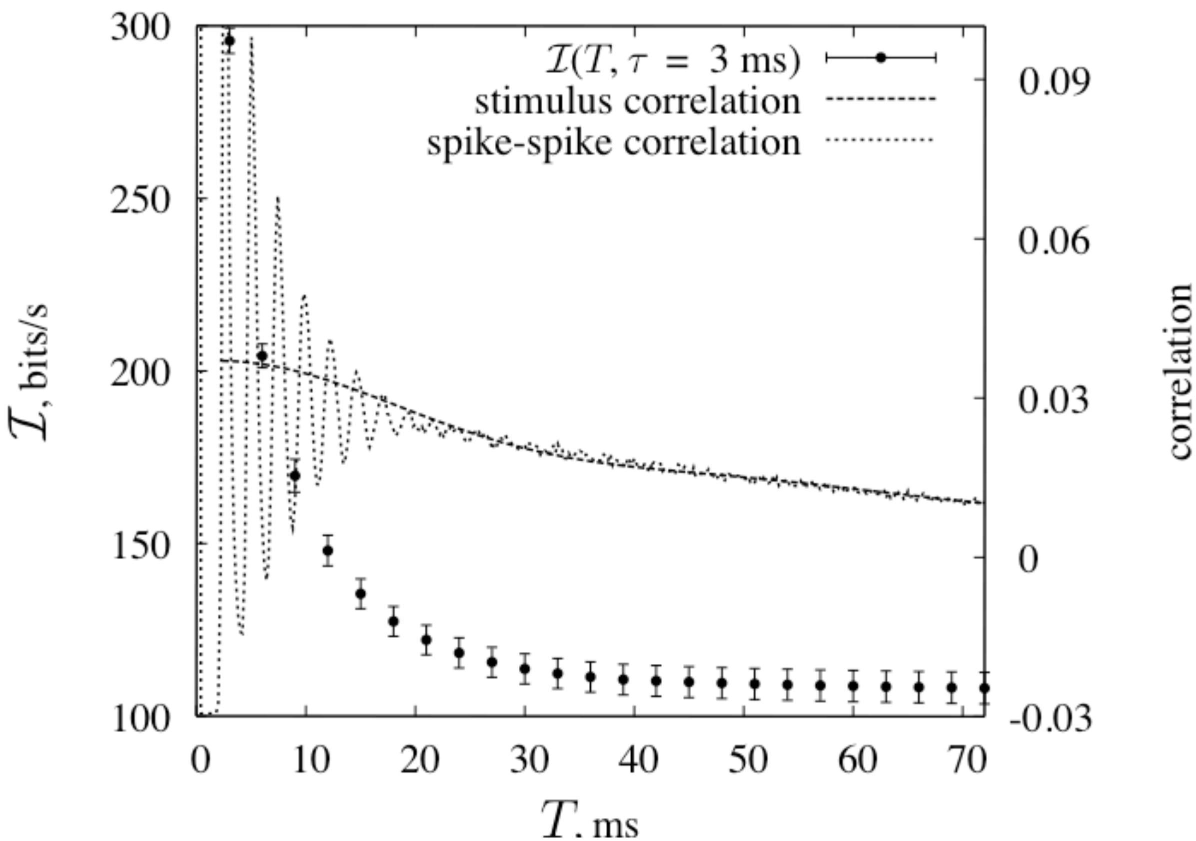
Synergy from spike combinations



Spike pairs

Redundancy due to stimulus

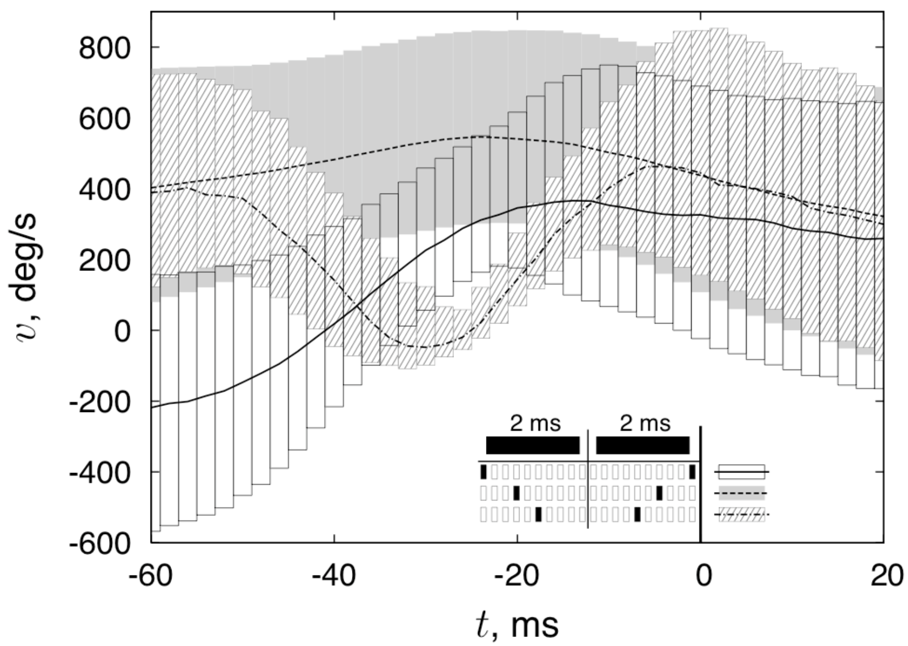
New bits (optimized code)



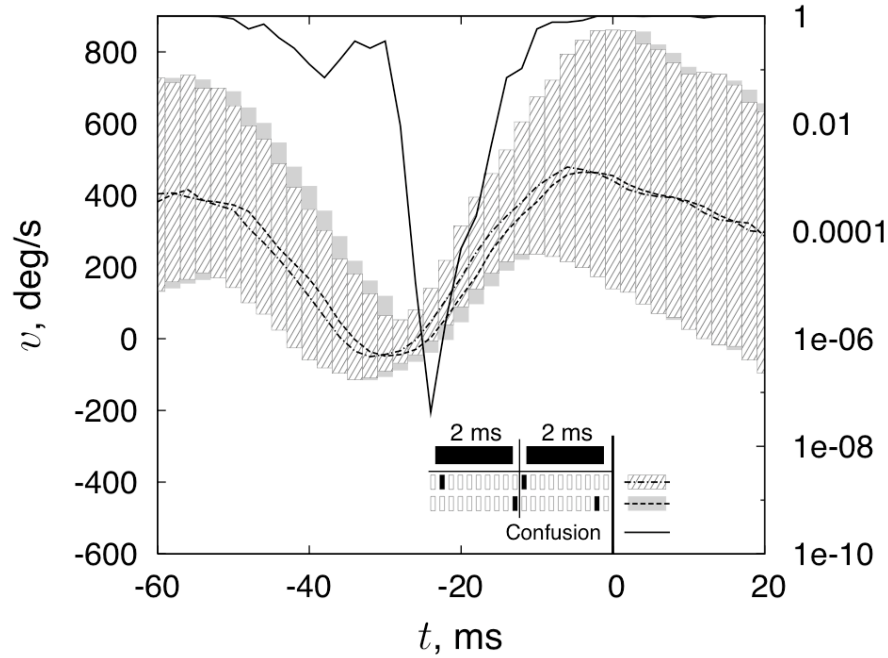
- Spikes are very regular (15 beats) WKB decoder? Interspike potential?
- CF at half its value, but fly gets new bits every 30 ms
- Independent info (even though entropies are T dependent).

Behaviorally optimized code!

Information about…



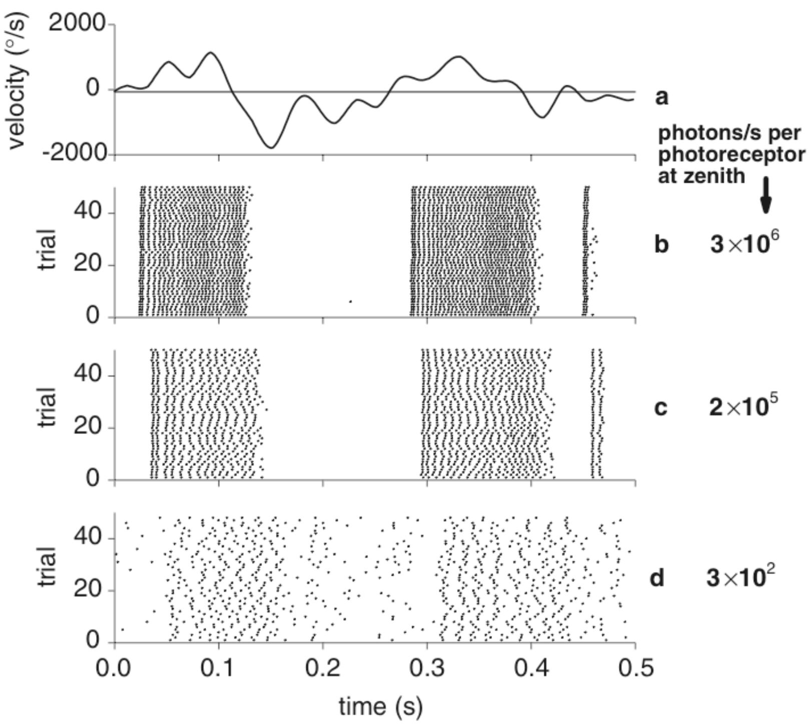
Signal shape



Zero-crossings time

Best estimation at 25ms delay. Little time for reaction.

Precision is limited by physical noise sources

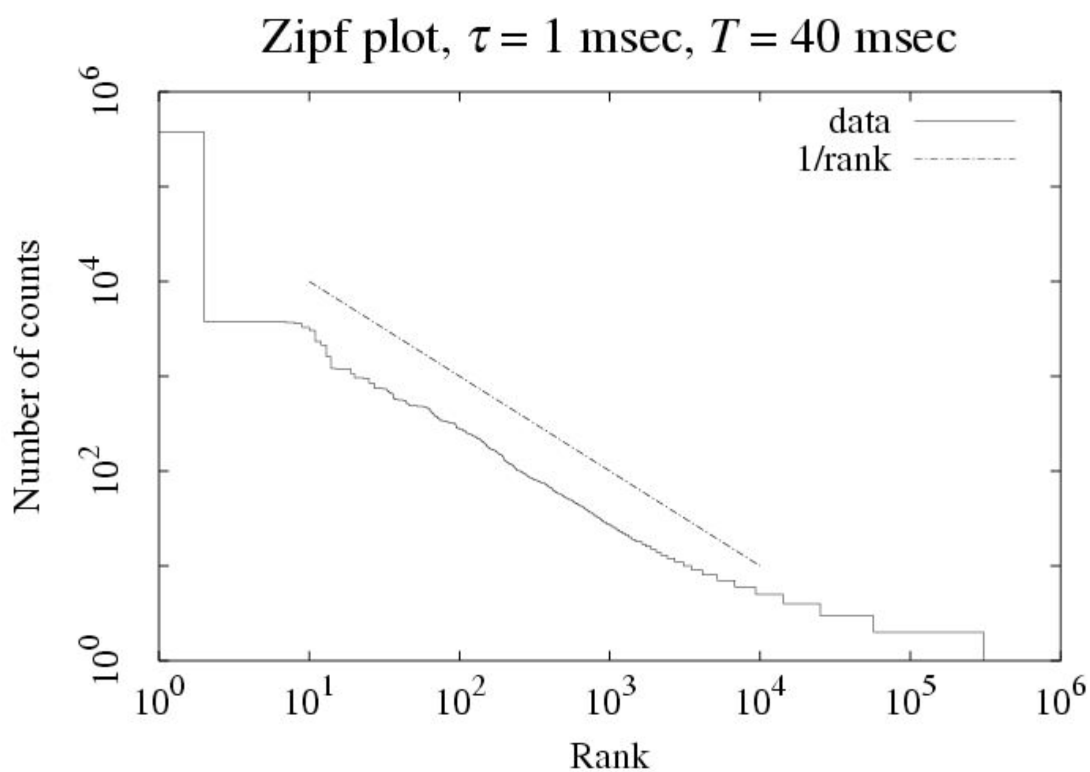


(Lewen, et al 2001)

We see evidence for lowering of the information rate with the light intensity dropping 0.3 log unit from its midday value.

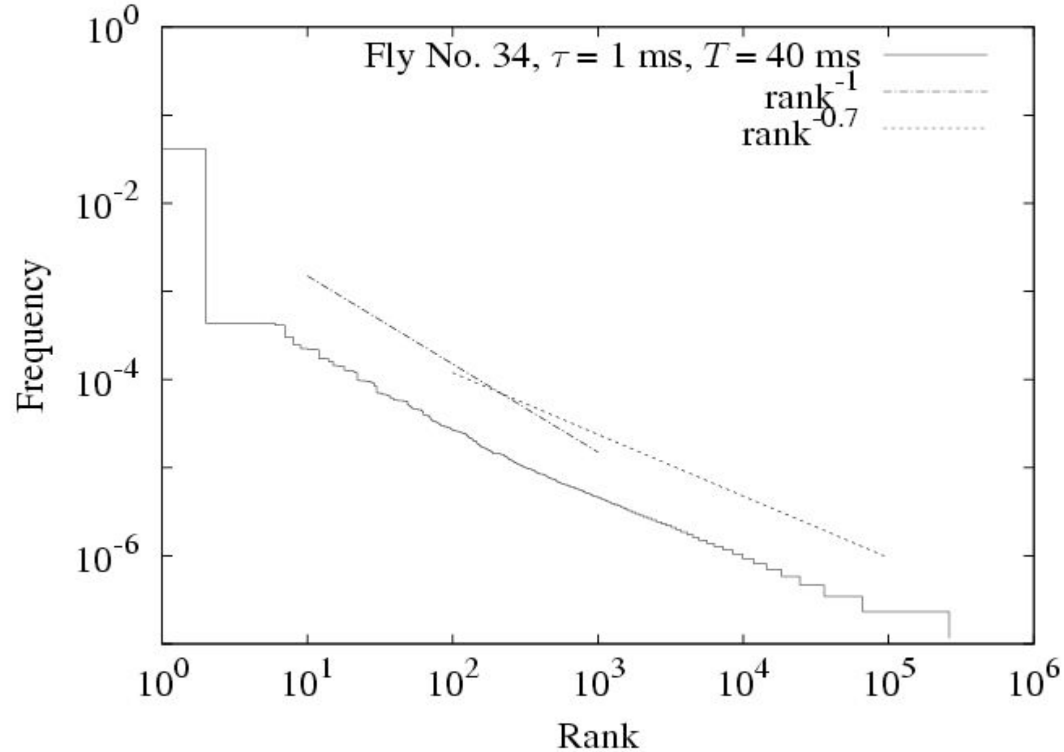
A very intelligent fly

- One often considers a 1/f rank-order plot as a sign of intelligence.
- But…



A very intelligent fly

- One often considers a 1/f rank-order plot as a sign of intelligence.
- But…



Zipf law may be a result of complexity of the world, not the language.