Synergy, redundancy, and independence in population codes, revisited.

or

Are correlations important?

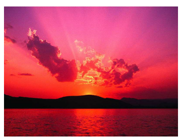
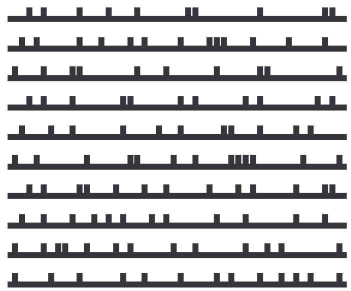
Peter Latham* and Sheila Nirenberg†

***Gatsby Computational Neuroscience Unit**
†University of California, Los Angeles

The neural coding problem

Estimate stimulus from responses:

s $\qquad\qquad$ $r_1, r_2, \ldots, r_n$

 →  → ?

Approach problem probabilistically:

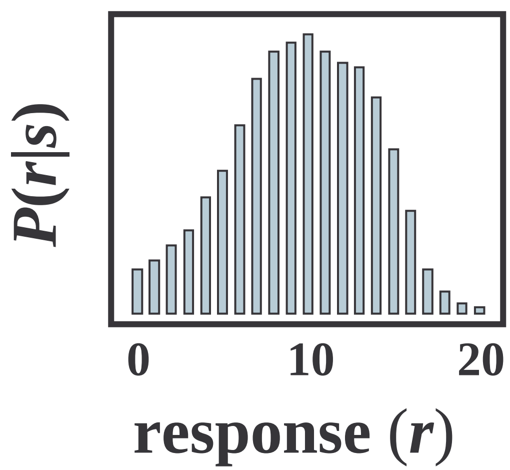
$$P(r_1, r_2, \ldots, r_n | s) \qquad\qquad P(s | r_1, r_2, \ldots, r_n)$$

Bayes

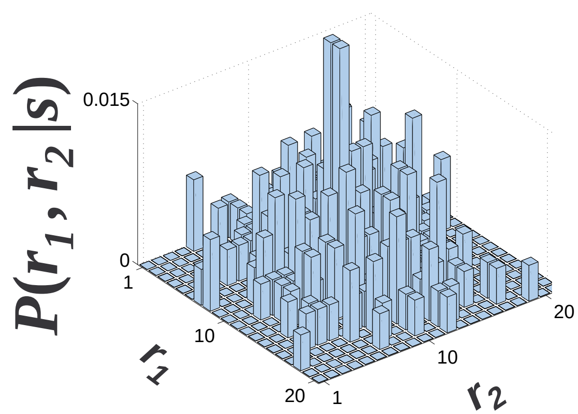
$$\frac{P(r_1, r_2, \ldots, r_n | s)\, P(s)}{P(r_1, r_2, \ldots, r_n)}$$

The problem is easy for one neuron (1-D) but harder for populations (≥ 2-D). Why? Because correlations force you to measure the probability in every bin.

easy in 1-D

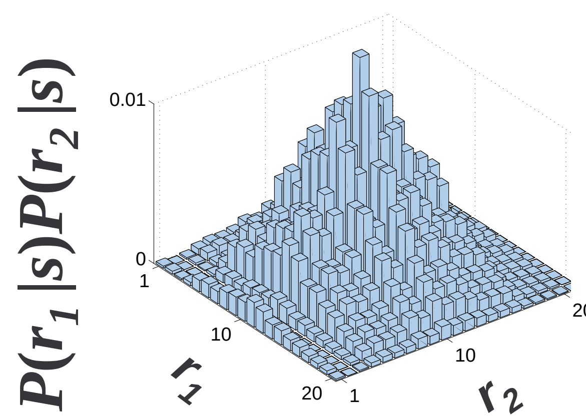


harder in 2-D



impossible in high-D.

"high" ~ 3.



Note: this problem disappears when the responses are uncorrelated.

- **If you want to understand how to decode spike trains, you have to figure out how to deal with correlations.**

- **The first step is to understand whether you need to deal with correlations.**

- **In other words, are correlations important?**

How to determine if correlations are important:

- **Get rid of them by treating the cells as though they were independent and then estimate the stimulus.**
- **If your estimate of the stimulus is different from the true estimate, then correlations are important. Otherwise they are not.**

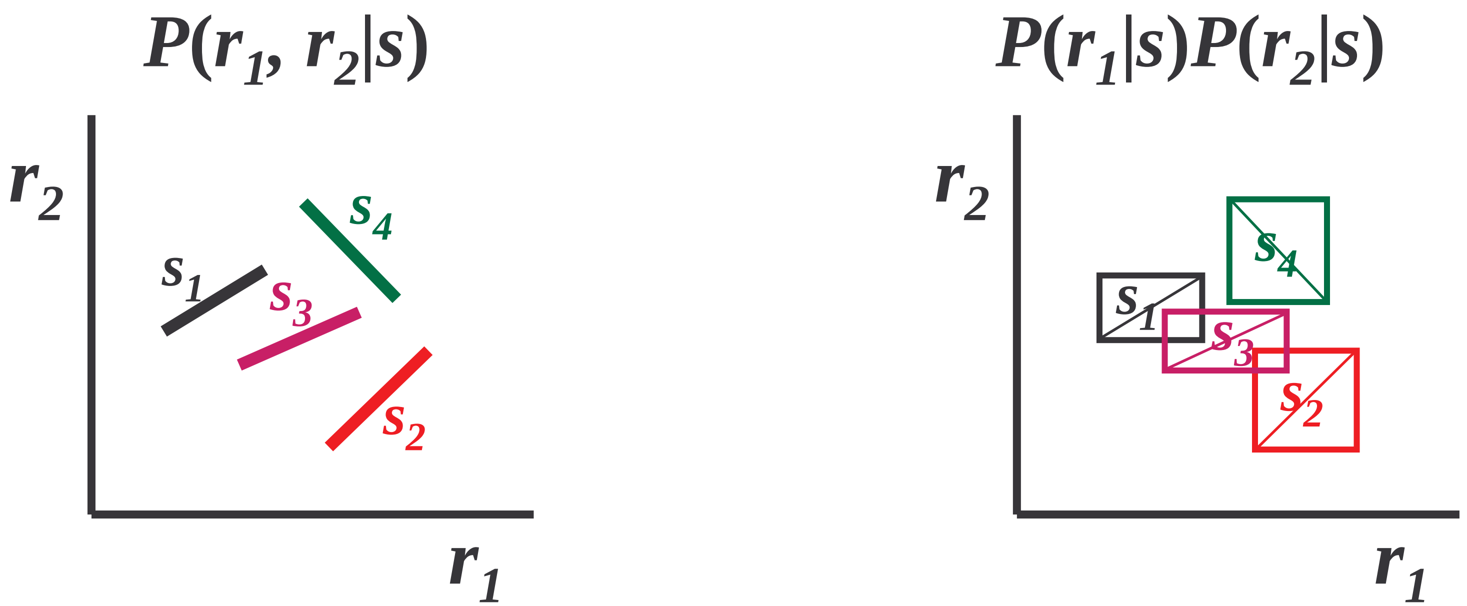
Formally, compare $P_{\text{ind}}(s|r_1, r_2)$ to $P(s|r_1, r_2)$, where

Independent response distribution

$$P_{\text{ind}}(s|r_1, r_2) = \frac{P(r_1|s)P(r_2|s)P(s)}{\sum_{s'} P(r_1|s')P(r_2|s')P(s')}$$

If $P_{\text{ind}}(s|r_1, r_2) \neq P(s|r_1, r_2)$, correlations are important for decoding.
If $P_{\text{ind}}(s|r_1, r_2) = P(s|r_1, r_2)$, correlations are not important.

One might wonder: how can $P_{ind}(s|r_1, r_2) = P(s|r_1, r_2)$ when neurons are correlated – i.e., when $P(r_1|s)P(r_2|s) \neq P(r_1, r_2|s)$?



Neurons are correlated, that is, $P(r_1|s)P(r_2|s) \neq P(r_1, r_2|s)$,

but correlations don't matter: $P_{ind}(s|r_1, r_2) = P(s|r_1, r_2)$.

Intuitively, the closer $P_{ind}(s|r_1, r_2)$ is to $P(s|r_1, r_2)$, the less important correlations are. We measure "close" using

$$\Delta I = \sum_{r_1,r_2,s} P(r_1, r_2)P(s|r_1, r_2) \log \frac{P(s|r_1, r_2)}{P_{ind}(s|r_1, r_2)}$$

= 0 if and only if $P_{ind}(s|r_1, r_2) = P(s|r_1, r_2)$

= penalty in yes/no questions for ignoring correlations

= upper bound on information loss.

If $\Delta I/I$ is small, then you don't lose much information if you treat the cells as independent.

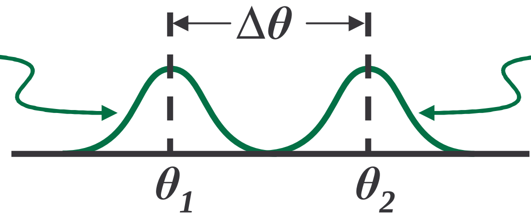
Quantifying information loss

Information is the log of the number of messages that can be transmitted over a noisy channel with vanishingly small probability of error.

An example: neurons coding for orientation.

You know: $P(\theta|r_1, r_2)$. You build: $\hat{\theta}(r_1, r_2)$ = optimal estimator.

distribution of $\hat{\theta}$ given θ_1 was presented.

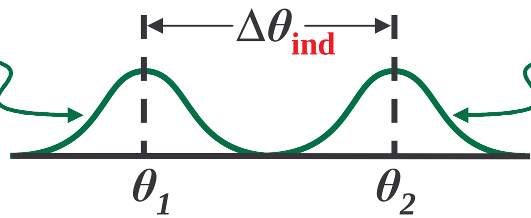


distribution of $\hat{\theta}$ given θ_2 was presented.

$$I \approx \log \frac{180}{\Delta\theta}$$

You know: $P_{ind}(\theta|r_1, r_2)$. You build: $\hat{\theta}_{ind}(r_1, r_2)$ = suboptimal estimator

distribution of $\hat{\theta}_{ind}$ given θ_1 was presented.



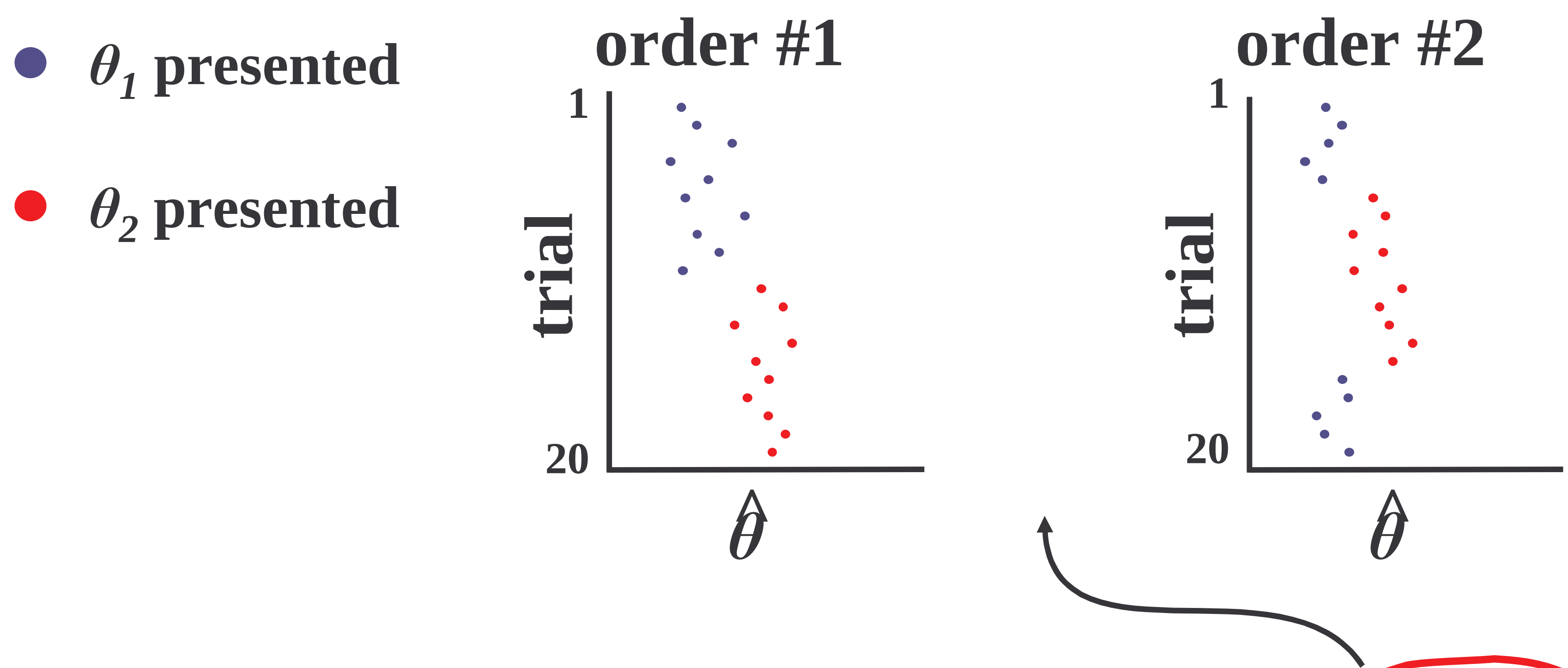
distribution of $\hat{\theta}_{ind}$ given θ_2 was presented.

$$I_{ind} \approx \log \frac{180}{\Delta\theta_{ind}}$$

Information loss: $I - I_{ind}$

if $\Delta\theta$ is large:

- Show multiple trials.
- Stimuli appear in only two possible orders.



$$I \approx \frac{\text{log of the number of distinct orders}}{n} = \frac{\log 2}{20}$$

rigorous as $n \to \infty$!!!!

Formal analysis: the general case

true distribution: $p(r|s)$

approximate distribution: $q(r|s)$

how many code words (a.k.a. orders) can you transmit using each?

$c(1) = s_1(1)\ s_2(1)\ s_3(1)\ \ldots\ s_n(1)$

$c(2) = s_1(2)\ s_2(2)\ s_3(2)\ \ldots\ s_n(2)$

$\vdots$

$c(w) = s_1(w)\ s_2(w)\ s_3(w)\ \ldots\ s_n(w)$

$\vdots$

code words (different ordering of stimuli)

trials

- **Observe $r_1\ r_2\ r_3\ \ldots\ r_n$; guess code word (guess w).**
- **More code words $\Rightarrow$ mistakes are more likely.**
- **You can transmit more code words without mistakes if use p to decode than if you use q.**
- **The difference tells you how much information you lose by using q rather than p.**

true probabilty: $p(w|r_1, r_2, r_3, \ldots, r_n) \sim \prod_i p(r_i|s_i(w))\, p(w)$

approx. probability: $q(w|r_1, r_2, r_3, \ldots, r_n) \sim \prod_i q(r_i|s_i(w))\, p(w)$

constant

decode: $\hat{w} = \arg\max_w \prod_i p(r_i|s_i(w))$ or $\prod_i q(r_i|s_i(w))$

want: $\prod_i p(r_i|s_i(w^*)) > \prod_i p(r_i|s_i(w))\ \forall w \neq w^*$

true code word

prob. error: $P_e[p,w] = \text{prob}\left\{\prod_i p(r_i|\, s_i(w)) > \prod_i p(r_i|s_i(w^*))\right\}$

$P_e[q,w] = \text{prob}\left\{\prod_i q(r_i|\, s_i(w)) > \prod_i q(r_i|s_i(w^*))\right\}$

number of code words that can be transmitted with vanishingly small probability of error ~ $1/P_e$.

$I[p] = \log(1/P_e[p,w])/n$

$I[q] = \log(1/P_e[q,w])/n$

definition

algebra algebra algebra

Information loss = $I[p] - I[q] \leq \Delta I$

Other measures

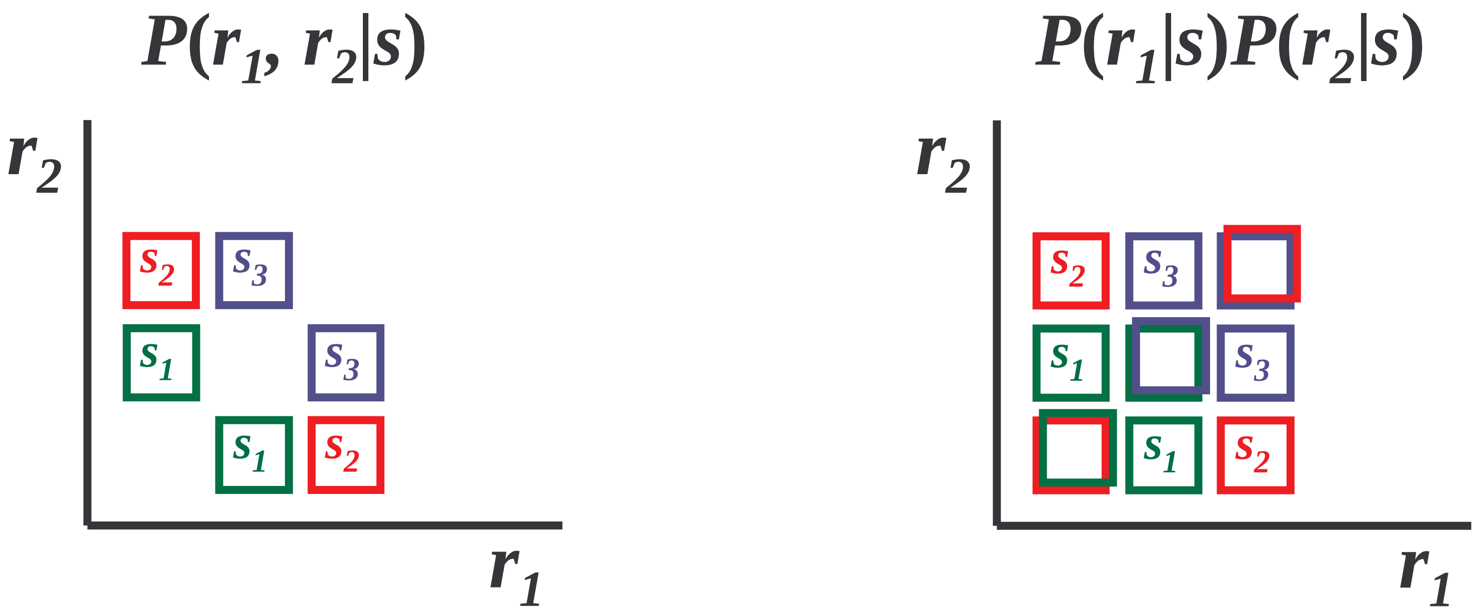
$$\Delta I_{synergy} = I(r_1, r_2;s) - I(r_1;s) - I(r_2;s)$$

Intuition: If responses taken together provide more information than the sum of the individual responses, which can only happen when correlations exist, then correlations "must" be important.

Good points:
- Cool(ish) name.
- Compelling intuition.

Bad points:
- Intuition is wrong: $\Delta I_{synergy}$ can't tell you whether correlations are important.

Example: A case where and you can decode **perfectly**, that is, $P_{ind}(s|r_1, r_2) = P(s|r_1, r_2)$ for all responses that occur, but, $\Delta I_{synergy} > 0$.



Schneidman, Bialek and Berry (2003) used this example to argue that $\Delta I_{synergy}$ is a good measure of whether or not correlations are important. We find this baffling.

$\Delta I_{synergy}$ can be: **zero, positive, negative** when $P_{ind}(s|r_1, r_2) = P(s|r_1, r_2)$ (Nirenberg and Latham, PNAS, 2003).

$\Delta I_{\text{shuffled}} = I_{\text{true}} - I_{\text{shuffled}}$

Information from neurons that saw the same stimulus but at different times (so that correlations are removed).

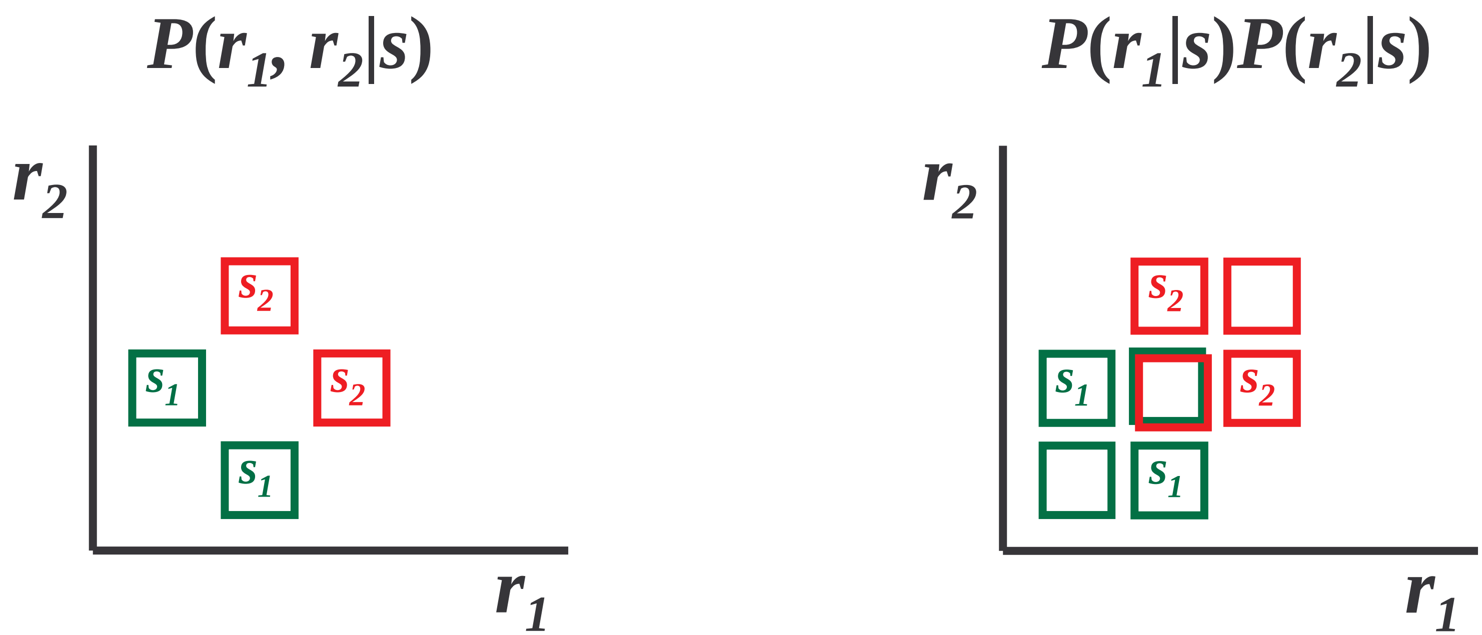
Intuition:

1. $I_{\text{shuffled}} > I_{\text{true}}$: Correlations hurt
2. $I_{\text{shuffled}} < I_{\text{true}}$: Correlations help
3. $I_{\text{shuffled}} = I_{\text{true}}$: Correlations don't matter

Good point:
- Can be used to answer high-level questions about neural code (what class of correlations increases information?).

Bad points:
- Intuition #3 is false; #1 and #2 are not so relevant, as they correspond to cases the brain doesn't see.

Example: A case where and you can decode perfectly, that is, $P_{\text{ind}}(s|r_1, r_2) = P(s|r_1, r_2)$ for all responses that occur, but, $\Delta I_{\text{shuffled}} > 0$.



$I = 1$ bit

$I_{\text{shuffled}} = 3/4$ bit

$\Delta Is_{\text{huffled}}$ can be: zero, positive, negative when $P_{\text{ind}}(s|r_1, r_2) = P(s|r_1, r_2)$ (Nirenberg and Latham, PNAS, 2003).

Summary #1

$\Delta I_{\text{shuffled}}$ and $\Delta I_{\text{synergy}}$ do not measure the importance of correlations for decoding – they are confounded.

ΔI does measure the importance of correlations for decoding:

- **$\Delta I = 0$ if and only if $P_{\text{ind}}(s|r_1, r_2) = P(s|r_1, r_2)$.**
- **ΔI is an upper bound on information loss.**

Summary #2

1. **Our goal was to answer the question:**

   Are correlations important for decoding?

2. **We developed a quantitative information-theoretic measure, ΔI, which is an upper bound on the information loss associated with ignoring correlations.**

3. **For pairs of neurons, $\Delta I/I$ is small, $< 12\%$, except in the LGN where it's 20-40%.**

4. **For larger populations, this is still an open question.**