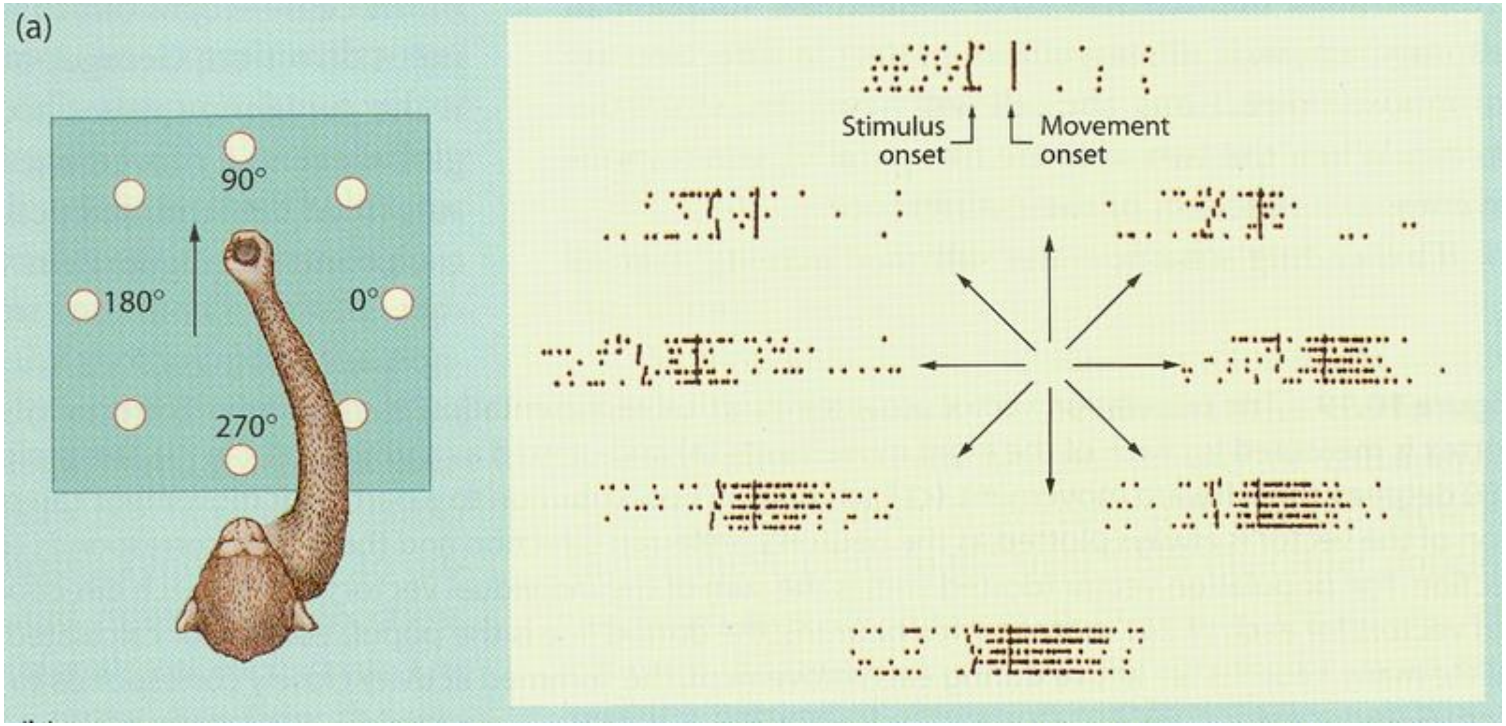
Neural Coding: Population Code

吴思
心理与认知科学学院
IDG/McGovern 脑科学研究所
北大-清华联合生命科学中心
北京大学

Neural Responses for Moving Direction

- In the experiment, the monkey was guided to move the lever in the center of apparatus to one of eight peripheral locations.

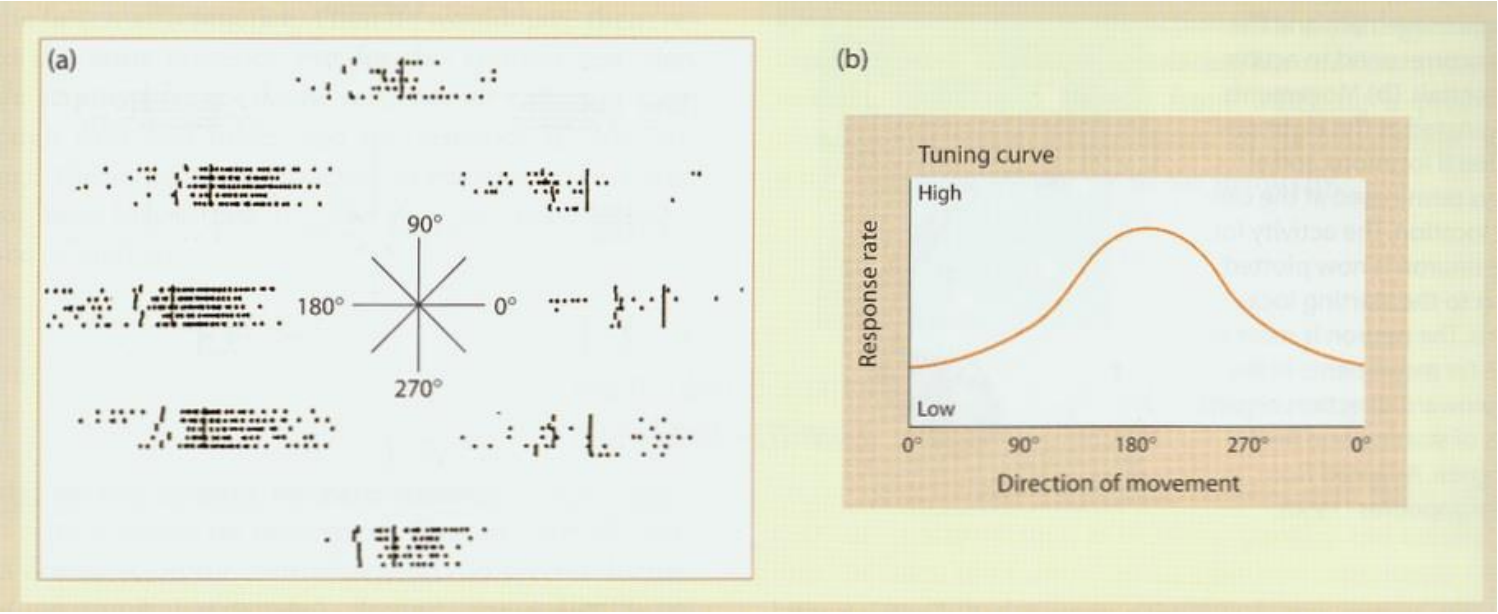
- Neural activities in the motor area were recorded.



Georgepolous et al, Science 1987

Preferred stimulus and tuning function

- Preferred stimulus: A neuron is associated with a stimulus value to which it has the maximum response
- Tuning function: the functional map between neural activity (measured by mean firing rate) and stimulus values
- Population representation: a group of neurons' receptive fields cover the whole stimulus value space
- Noise always there



Population Vector

Georgopoulos's **Population Vector**

$$\vec{v} = \sum_i \frac{r_i}{Z} \vec{c}_i$$

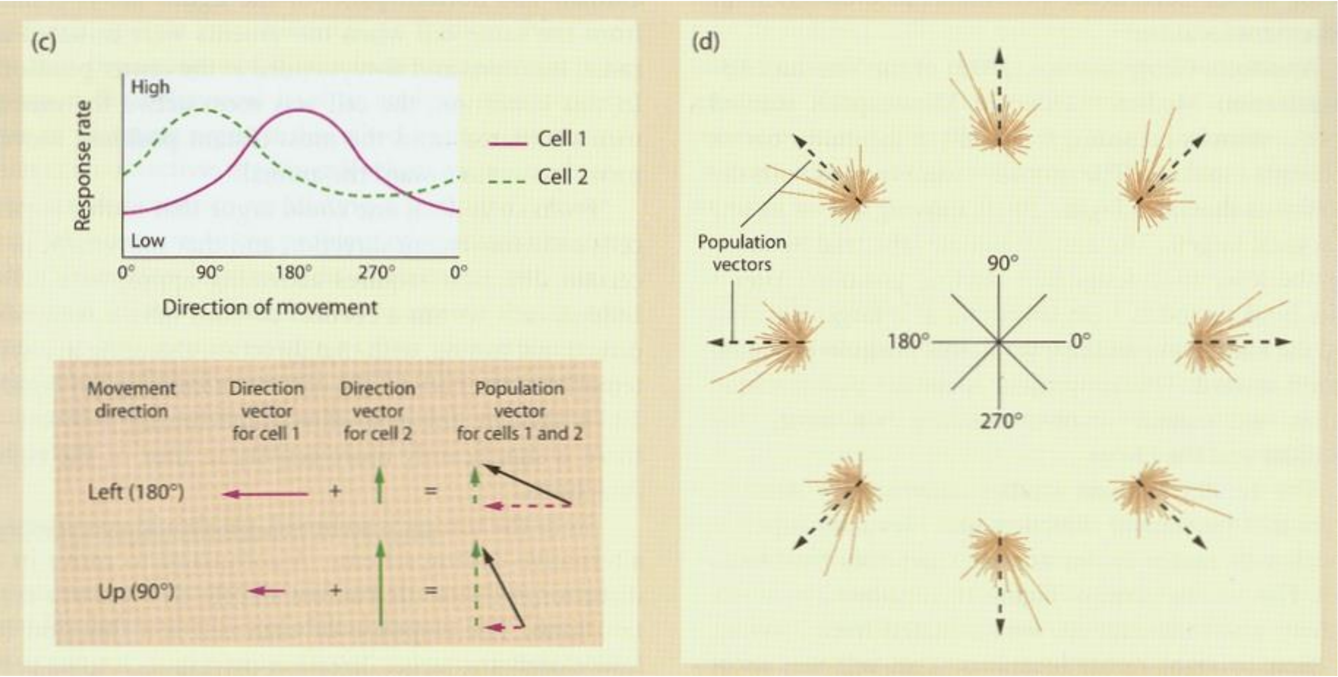
$\vec{v}$: the decoded moving direction

$\vec{c}_i$: the preferred direction of the ith neuron

$r_i \vec{c}_i$: the contribution from the ith neuron

$Z = \sum_i r_i$ the normalization factor

An illustration of Population Vector



About Population Code

- Robust to damage in individual neurons
- Robust to noise in the activity of neuron
- Hold for head direction, orientation, spatial location, …
- A general framework of neural information processing?

Mathematical Formulation (1)

- The smooth bell-shape tuning curve is often modeled by the Gaussian (or cosin) function:

$$f_i(x) \sim \exp[-(x-c_i)^2/2a^2]$$

$f_i(x)$: mean firing rate of the ith neuron

x: stimulus

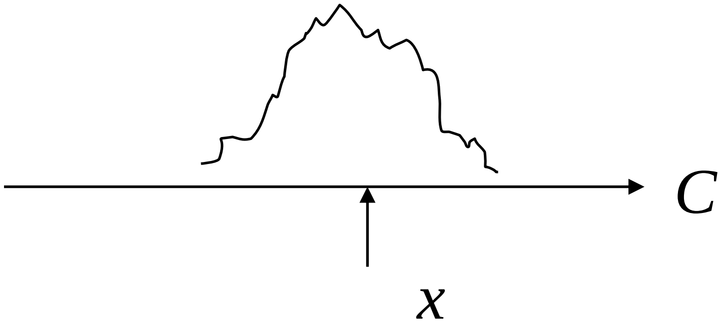
c_i: neuronal preferred stimulus

a: tuning width

- For neural activity in a single trial

$r_i = f_i(x) + \varepsilon_i$

ε_i: a random number of zero mean



Note: both tuning function and population activity are of the bell-shape

Mathematical Formulation (2)

- The encoding phase

$$x \rightarrow \mathbf{r} = \{r_i\}, \quad \text{for } i = 1, \dots, N, \qquad N \text{ is the number of neurons}$$
$$r_i = f_i(x) + \varepsilon_i$$

The encoding process is specified by the conditional probability $p(\mathbf{r}|x)$, i.e., the probability that the neural activities $\mathbf{r}$ is generated given the stimulus x is presented.

- The decoding phase

$$\mathbf{r} \xrightarrow{\text{infer}} x$$

Population vector is one of many inference strategies.

Some Theoretical Issues

- How much information about the stimulus is encoded in a population code?

- What is the most efficient decoding strategy?

- What is the biologically plausible decoding strategy?

- What is the effect of neural correlation on population coding?

Fisher Information & Cramer-Rao Bound

$$F = -\int d\mathbf{r}\, p(\mathbf{r}|x)\nabla\nabla \ln p\,(\mathbf{r}|x) = \int d\mathbf{r}\, p(\mathbf{r}|x)[\nabla \ln p\,(\mathbf{r}|x)]^2$$

$$\nabla p(\mathbf{r}|x) = \frac{dp(\mathbf{r}|x)}{dx}$$

$$< (\hat{x} - x)^2 > \geq \frac{1}{F}$$

F: the Fisher information, the amount of stimulus information contained in the population code.

The Cramer-Rao bound: the error of any unbiased decoder, measured by variance, is larger than the inverse of the Fisher information.

Asymptotical Efficiency of MLP

Maximum Likelihood Inference: $\hat{x} = Max_x \ln p\,(\mathbf{r}|x)$

Suppose x^* is the true stimulus,

$\hat{x}$ is the estimation based on the observation $\mathbf{r}$, i.e., $\nabla \ln p\,(\mathbf{r}|\hat{x}) = 0$

First−order Tayler expansion at x^* gives rise to

$\nabla \ln p\,(\mathbf{r}|x^*) + \nabla\nabla \ln p\,(\mathbf{r}|x^*)(\hat{x} - x^*) \approx 0$

$$(\hat{x} - x^*) \approx \frac{\nabla \ln p\,(\mathbf{r}|x^*)}{-\nabla\nabla \ln p\,(\mathbf{r}|x^*)} = \frac{R_1}{R_2}$$

Consider i.i.d of data, in the large N limit, $\mathrm{R}_2 \approx F$

$$\langle R_1 \rangle = 0,\ \langle R_1^2 \rangle = \int d\mathbf{r} p(\mathbf{r}|x^*)[\nabla \ln p\,(\mathbf{r}|x^*)]^2 = F$$

Thus

$$\langle (\hat{x} - x^*) \rangle \approx 0; \quad \langle (\hat{x} - x^*)^2 \rangle \approx \frac{1}{F}$$

Ex.: Calculating the Fisher Information

- Neuronal responses satisfy Poisson statistics,

$$p(n_i|x) = \frac{(f_i t)^{n_i}}{n_i!} e^{-f_i t}$$
$$f_i(x) = Ae^{-(c_i-x)^2/2a^2}$$

- Neurons fire independently,

$$p(\mathbf{n}|x) = \prod_i p(n_i|x)$$

- The Fisher information

$$F = \sqrt{2\pi}At\rho/a$$

Population Decoding Strategies

- Center of Mass (equivalent to population vector)

$$\hat{x} = \frac{\sum_i r_i c_i}{\sum_i r_i}$$

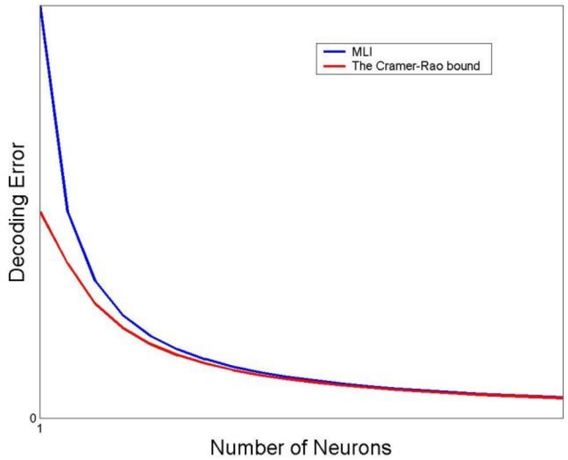
Accuracy may be low

Maximum Likelihood Inference

$$\hat{x} = \max_z \log p\,(r|z)$$

$p(r|z)$ is the conditional probability of observing r given the stimulus value z

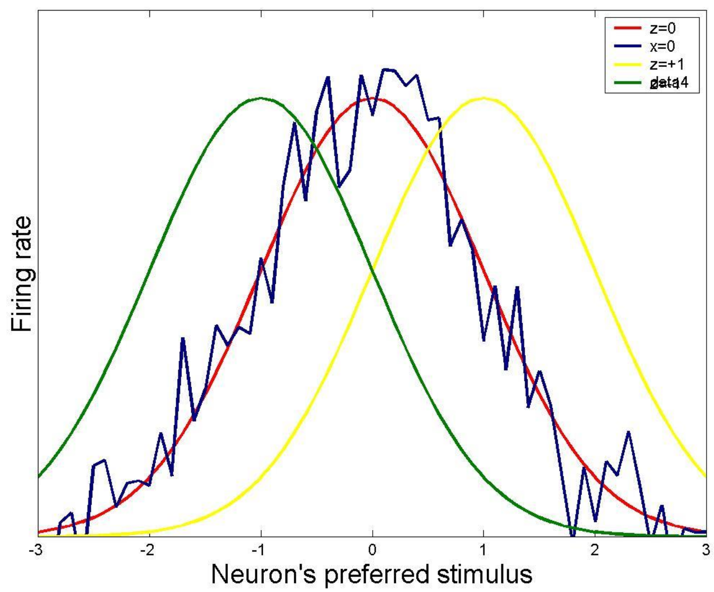
MLI is **asymptotically efficient** in many cases. This means that when the number of neurons is sufficiently large, its decoding error reaches the lower bound, i.e., the inverse of the Fisher information.



Computational cost may be high

Template Matching

$$\hat{x} = \max_z \sum_i r_i f_i(z)$$



- The noisy bump is the population activity when the stimulus $x=0$.
- Among three positions, the red one ($z \approx 0$) has the maximum overlap with the observed data.

Performance of template-matching

- MLI for independent Gaussian noise

For independent Gaussian noise,

$$p(\mathbf{r}|z) = \prod_i p(r_i|z) \propto \prod_i \exp[-(r_i - f_i(z))^2/2\sigma^2]$$

Thus

$$\hat{x} = \max_z \log p\,(\mathbf{r}|z)$$

$$= \max_z \sum_i -[r_i - f_i(z)]^2$$

$$= \max_z \sum_i r_i f_i(z) \longrightarrow \text{Template-matching}$$

To get the last equality, we have used the condition

$\sum_i f_i(z) \approx$ constant, which is true when the number of neuron is large

Unfaithful Decoding

- Maximum likelihood based on an unfaithful model:

$$\hat{x} = \max_x \log q\,(\mathbf{r}|x)$$

$q(\mathbf{r}|x) \neq p(\mathbf{r}|x)$

$p(\mathbf{r}|x)$: **the real encoding process**

$q(\mathbf{r}|x)$: **the presumed decoding model**

- Motivations:
  - The real encoding process is unknown to the neural estimator;
  - A simplified decoding model may greatly decease computation cost, and meanwhile achieve a reasonable decoding accuracy.

Wu et al., Nips, 1999; Neural Computation, 2001

Accuracy of unfaithful decoding

Suppose x^* is the true stimulus,

$\hat{x}$ is the estimation based on the observation $\mathbf{r}$, i.e., $\nabla \ln q\,(\mathbf{r}|\hat{x}) = 0$

First−order Tayler expansion at x^* gives rise to

$$\nabla \ln q\,(\mathbf{r}|x^*) + \nabla\nabla \ln q\,(\mathbf{r}|x^*)(\hat{x} - x^*) \approx 0$$

$$(\hat{x} - x^*) \approx \frac{\nabla \ln q\,(\mathbf{r}|x^*)}{-\nabla\nabla \ln q\,(\mathbf{r}|x^*)} = \frac{R_1}{R_2}$$

Consider i.i.d of data, in the large N limit, $\mathrm{R}_2 \approx -\int d\mathbf{r}p(\mathbf{r}|x^*)\nabla\nabla \ln q\,(\mathbf{r}|x^*)$

$$\langle R_1 \rangle = 0,\ \langle R_1^2 \rangle = \int d\mathbf{r}p(\mathbf{r}|x^*)[\nabla \ln q\,(\mathbf{r}|x^*)]^2$$

Thus

$$\left\langle (\hat{x} - x^*)^2 \right\rangle \approx \frac{\int d\mathbf{r}p(\mathbf{r}|x^*)[\nabla \ln q\,(\mathbf{r}|x^*)]^2}{-\int d\mathbf{r}p(\mathbf{r}|x^*)\nabla\nabla \ln q\,(\mathbf{r}|x^*)}$$

An Example of Unfaithful Decoding

- Assume neuronal responses satisfy independent Poisson distribution,

$$q(\mathbf{n}|x) = \prod_i q(n_i|x)$$
$$= \prod_i \frac{(f_i t)^{n_i}}{n_i!} e^{-f_i t}$$
$$f_i(x) = A e^{-(c_i - x)^2/2a^2}$$

- The maximum likelihood based on the unfaithful model

$$\hat{z} = \max_x \log q\,(\mathbf{n}|x)$$
$$= \frac{n_i c_i}{n_i}$$

→ Center of Mass

An example of Unfaithful Decoding

- **The real encoding process**

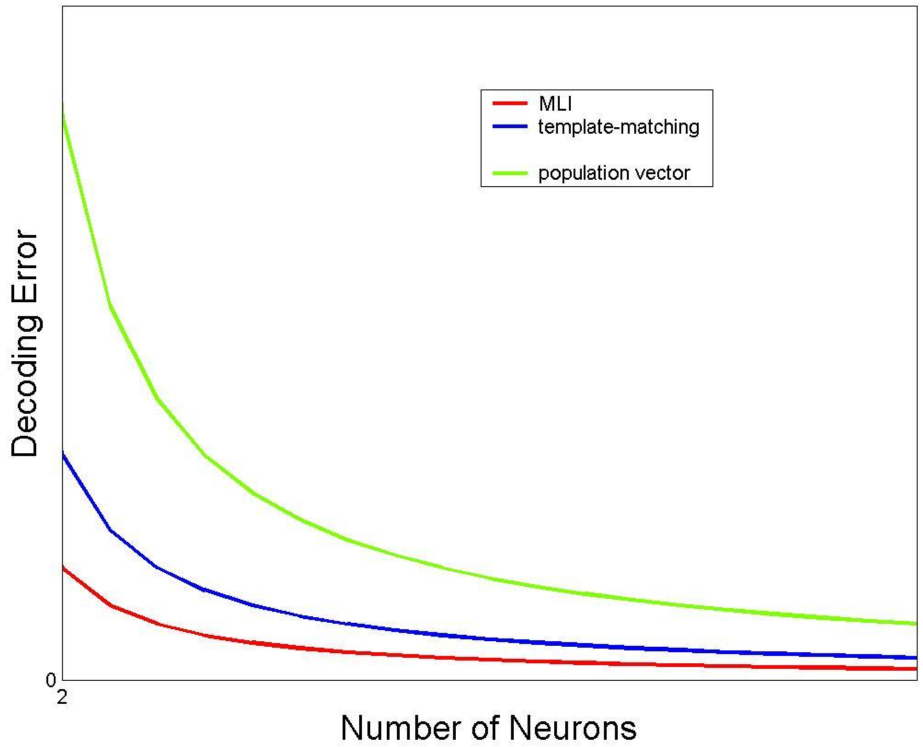
$$p(\{r_i\}|x) \propto \exp[-\sum_{ij}(r_i - f_i)A_{ij}^{-1}(r_j - f_j)/2\sigma^2]$$

- **An unfaithful model—neglecting the correlation**

$$q(\{r_i\}|x) \propto \exp[-\sum_i(r_i - f_i)^2/2\sigma^2]$$

$$\begin{aligned}\hat{x} &= \mathrm{Max}_x \log q\,(\{r_i\}|x)\\ &= \mathrm{Max} \textstyle\sum_i r_i f_i \Rightarrow \text{templat}-\text{matching}\end{aligned}$$

Typical performances of three methods



In term of decoding error
Center of Mass > Template-matching > MLI

Bayesian Inference

- Bayes' theorem

$$P(x|\mathbf{r}) = P(\mathbf{r}|x)P(x)/P(\mathbf{r})$$

- Maximum a Posterior (MAP)

$$\hat{x} = \mathbf{argmax}_x[\ln P\,(\mathbf{r}|x) + \ln P\,(x)]$$

Why Bayesian Decoding

- **Two simple scenarios for Bayesian decoding:**

1. The same or similar stimulus has been presented through multiple steps;
2. During a single presentation, neural signal has been sampled multiple times.

In both cases, the brain gets a rough estimation of stimulus in each step, which could be naturally used as prior knowledge for consecutive decoding.

Sequential Bayesian Decoding

The Decoding Procedure

$Step\ 1$: $\hat{x}_t$ $(When\ t = 1,\ Maximum\ Likelihood)$

$Step\ 2$: **Gaussian Prior:** $P(x) = e^{-(x-\hat{x}_t)^2/2\tau_t^2} / (\sqrt{2\pi}\,\tau_t)$

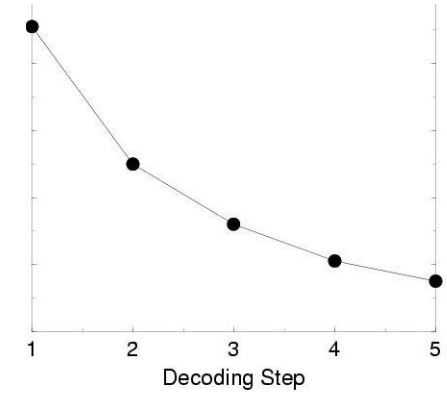
$Step\ 3$: $\hat{x}_{t+1}$ $(Maximum\ a\ Posterior)$

$Step\ 4$: **Repeat** $Step\ 2$

The optimal result

$$\Omega_t^2 = \frac{1}{t}\Omega_1^2$$

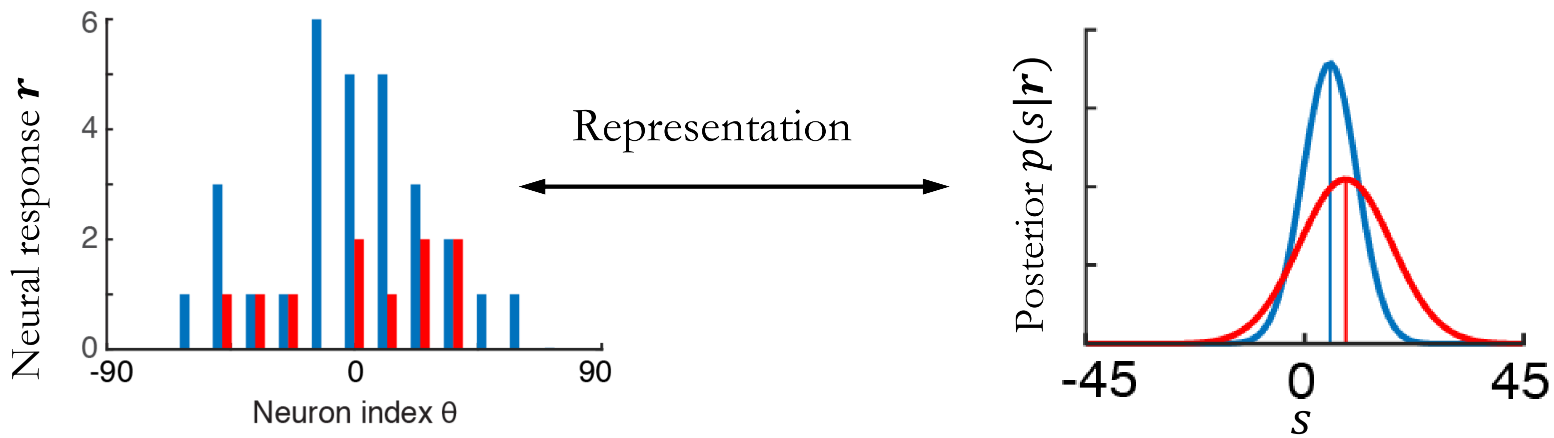
$$\tau_t^2 = \frac{1}{t} \cdot \frac{1}{-\nabla\nabla \ln P\,(\mathbf{r}|x)}$$



Wu et al., NIPS*2001

Uncertainty representation by population code

- Probabilistic neural population code
  - The posterior of a 1D stimulus feature is represented by N neurons.
  - A snapshot of $\boldsymbol{r}$ represents a whole posterior.

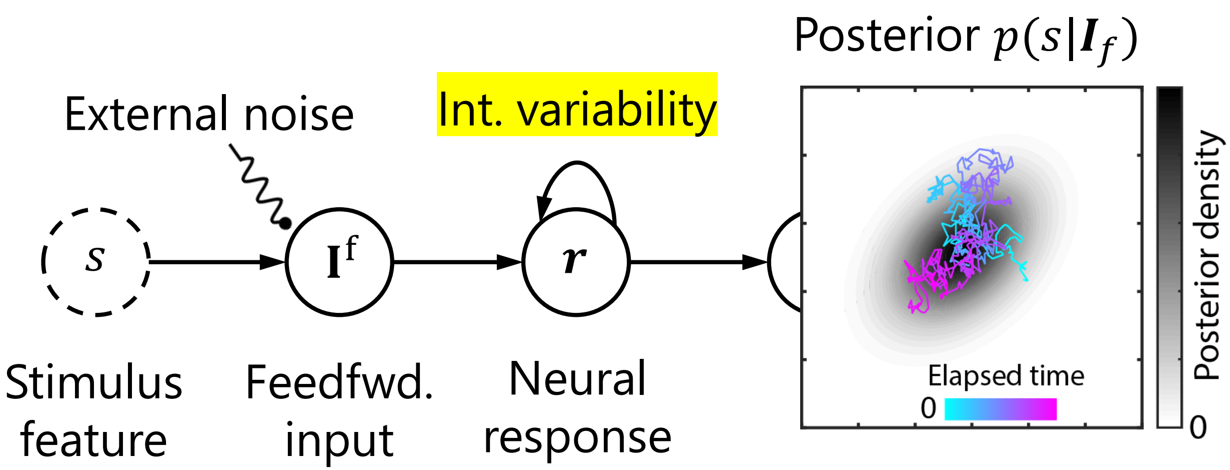


(Ma, *Nat. Neurosci.*, 2006)

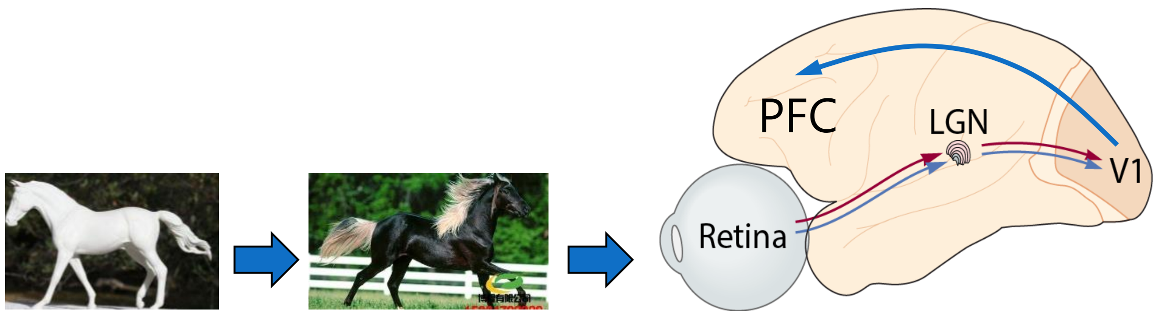
Uncertainty representation by sampling

Langevin sampling of the posterior

$$\tau \frac{ds}{dt} = \frac{1}{2}\frac{d\ln p(s|x)}{ds} + \sqrt{\tau}\xi(t)$$



What is really decoded?



References

1. S. Wu and S. Amari. (2002). Neural Implementation of Bayesian Inference in Population Codes. Advances in Neural Information Processing Systems 14 (NIPS*2001).
2. S. Wu, H. Nakahaara, N. Murata and S. Amari. (2000). Population Decoding Based on an Unfaithful Model. Advances in Neural Information Processing Systems 12 (NIPS*1999), pp.192-198, 2000.
3. S. Wu, H. Nakahara and S. Amari. (2001). Population Coding with Correlation and an Unfaithful Model. Neural Computation, v.13, p.775-798.