Information Theory and Neuroscience I: Optimal Storage in Noisy Synapses

John Z. Sun and Da Wang

Massachusetts Institute of Technology

October 14, 2009

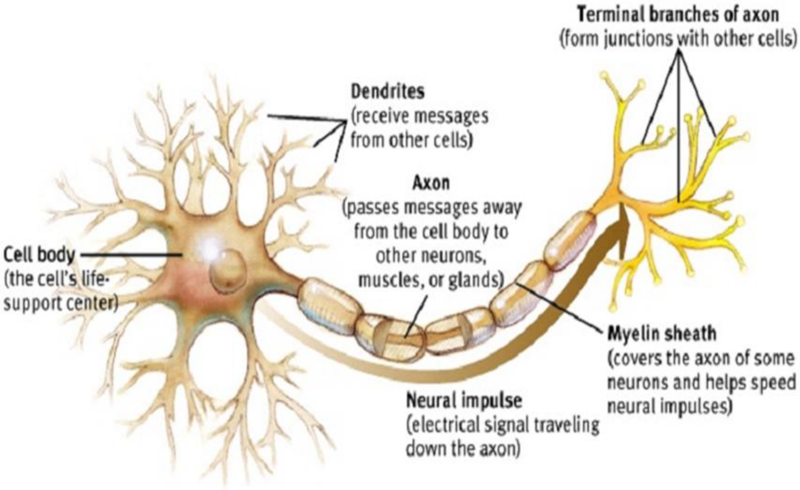
Outline

- An Introduction to Neuroscience
- Information Theory
- Questions about Synapses
- Theoretical Framework
- Investigations
  - Optimal Storage Capacity Per Unit Volume
  - Optimal Distribution of Synaptic Weights
  - Synaptic Cost Function from Its Weight Distribution
- Summary and Remarks

The Brain

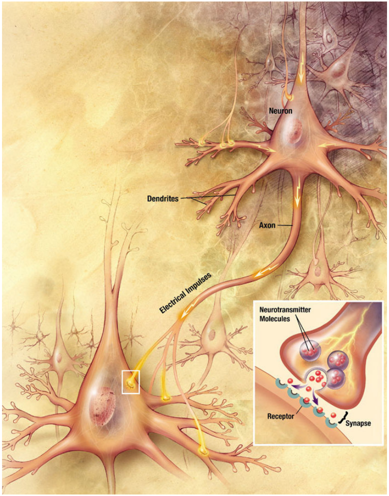
- Controls human behavior
- Is an extremely powerful computer, with “switches” being neurons, connected by “wires” called synapses
- Humans have tens of billions of neurons
- Has fan-in and fan-out on order of 10^4

Neurons



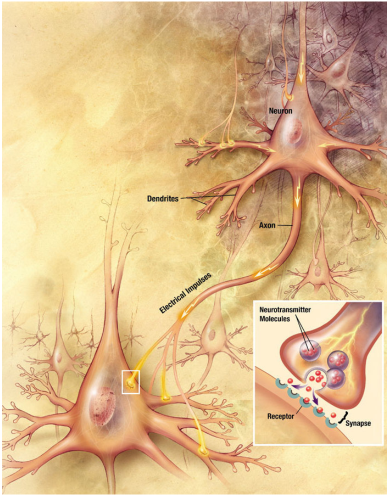
Terminal branches of axon
(form junctions with other cells)
Dendrites
(receive messages
from other cells)
Axon
(passes messages away
from the cell body to
other neurons,
muscles, or glands)
Cell body
(the cell's life-
support center)
Myelin sheath
(covers the axon of some
neurons and helps speed
neural impulses)
Neural impulse
(electrical signal traveling
down the axon)

Communication through Synapses



- Neurons have different objectives and behaviors
  - Propagate information to other parts of the body
  - Store memory
  - Filter stimuli

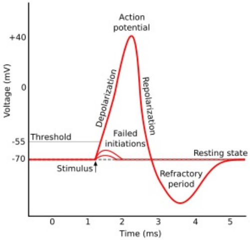
Communication through Synapses



- Modeled in first-order approximation as classical integrate and fire (CIF)
  - Importance of each synapse is quantified in synaptic weight
  - Synaptic weight (or efficacy) is defined as the average EPSP amplitude over trials
  - Experiments show that increasing EPSP increases SNR
  - Experiments show that synaptic weight is proportional to synaptic volume

Communication through Synapses

- Neurons send and receive pulses to communicate
- These are called spike trains or action potentials
- Several ideas on how information is sent
  - Temporal codes
  - Rate codes
  - Spatial-temporal codes
- Usually approximated as digital, slotted, random transmissions



Information Theory

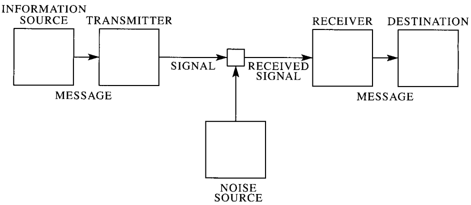


Fig. 1 — Schematic diagram of a general communication system.

[Shannon (1948)]

- Deals with asymptotic performance of information transfer
- Elegant theory for probabilistic sources and channels
- Introduced concepts of entropy, capacity and mutual information

Information Theory

- Assume input X and output Y to the channel
- The channel is then described by $P(Y|X)$
- Want to find $P(X)$ to maximize $C = I(X;Y)$
- Extensions include adding cost constraint $b(x)$ on input
- In general, very difficult optimization problem

The Theory of Everything?

"Information theory has, in the last few years, become something of a scientific bandwagon... Although this wave of popularity is certainly pleasant and exciting for those of us working in the field, it carries at the same time an element of danger." [Shannon (1956)]

"There is a constructive alternative for the author of this paper. If he is willing to give up larceny for a life of honest toil, he can find a competent psychologist and spend several years at intensive mutual education, leading to productive joint research." [Elias (1958)]

Information Theory and Neuroscience

- Capacity for specific neuronal models
  - [MacKay & McCulloh (1952)], [Stein, French & Holden (1972)], [Eckhorn & Pöpel (1975)]
- Using mutual information for sensory coding
  - [Lewicki (2002)], [Smith & Lewicki (2006)], [Chechik *et al* (2006)]
- Finding optimal $P(X)$ given cost constraints
  - [Varshney, Sjöström & Chklovskii (2006)], [Berger & Levy (2009)]
- Surveys on IT in Neuroscience
  - [Rieke *et al* (1996)], [Borst & Theunissen (1999)], [Johnson (2003)], [HST.722J]

Questions about Synapses

Varshney et. al. attempt to answer the following questions about synapses:

- Why are typical central synapses noisy?
  - Arriving spikes often fail to evoke EPSP.
  - EPSP amplitude varies from trial to trial.
  - Synaptic unreliability detrimental for transmission.
- Why does the distribution of synaptic strengths display a notable tail of strong connections?
  - Most synaptic efficacies are weak (mean EPSP $<$ 1mV).
  - But a notable tail of stronger synaptic strengths (mean EPSP $>$ 20mV) are observed.
- Why is synaptic connectivity sparse?
  - Connectivity sparse at both local and global levels.
  - Increased connectivity would enhance system functionality.

Theoretical Framework

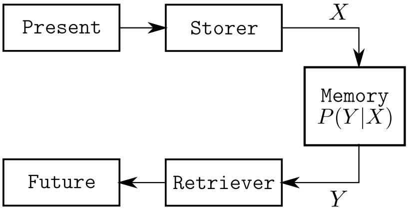
Develop a theoretical framework based on the roles of synapses as mechanism of **information storage**.

Optimization Approach

- Evolution by natural selection favors genotypes of high fitness.
- Qualities that affect fitness tend to improve through evolution.
- Mathematical analysis of fitness optimization seems reasonable as an approach to understand organisms.

Memory as a Channel

Memory channel:



Input distribution $P(X)$:

- distribution of **synaptic weights** across synapses.
- $X = 0$: zero-weight connections.

Channel $P(Y|X)$ determined by the noises:

- storage noise
- *in situ* noise
- read-out noise

Remarks

- Each synapse is a channel usage.
- Channel usage are separated in space, rather than time.

Optimization Formulation

Maximizing information storage of synapses under **resource constraints.**

Storage capacity

$$C = \max_{P_X} I(X;Y)$$

where

$$\text{SNR} = E\left[\frac{X^2}{\text{Noise}}\right]$$

Constraints

- Minimize volume usage.
  - Cost function $b(x)$ is determined by synapse volume V.
  - Known: synaptic weight $X \propto$ volume V
- No specific assumptions regarding the neural network.
- No specific assumptions on error control coding.

Synaptic Weight and Synaptic Volume

Experimental studies show synaptic weight (efficacy) X correlate with synapse volume $V(X)$ positively:

$$\frac{V(X)}{V_N} = \left(\frac{X}{A_N}\right)^{\alpha}$$

where

- A_N: noise amplitude.
- $V(X)$: volume of synapse with weight X.
  - $V(X = x)$: may write as V_x.
- V_N: synapse normalization constant volume (volume of synapse with SNR $= 1$).
- α: undecided, maybe $\alpha \leq 1$.

Cost Function

- In addition to the volume of synapses $V(X)$ itself, there is a cost of accessory volume $V_0 > 0$ that supports a synapse.
  - There are no free symbols.
- Cost function:

$$b(X) = V(X) + V_0$$

Average (expected) cost:

$$B \triangleq E[b(X)] = E[V(X)] + V_0$$

Used $\langle \cdot \rangle$ for average/expectation:

$$\langle V \rangle = E[V(X)]$$
$$\langle b \rangle = B = \langle V \rangle + V_0$$

Optimal Storage Capacity Per Unit Volume

Given the mathematical model of neuron memory:

- Perform optimization to generate predictions and explanations.
- Optimize information storage capacity per unit volume, *i.e.*, capacity per unit cost.
  - Since there is no zero-cost symbol ($V_0 > 0$), cannot apply the analytical results in [Verdu, 1990 & 2002].

Special Case: AWGN

Let $\alpha = 2$:

$$\text{SNR} = E\left[\frac{X^2}{A_N^2}\right] = \frac{\langle V \rangle}{V_N}$$

Let the channel be AWGN:

$$C_{\text{AWGN}} = \frac{1}{2}\log\left(1 + \text{SNR}\right)$$

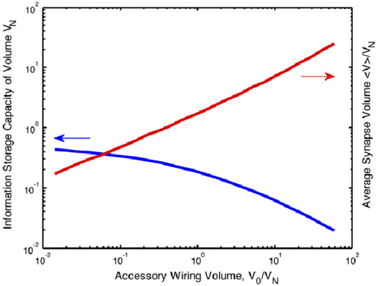
Capacity per unit volume:

$$\hat{C} = \frac{C}{B} = \frac{C}{\langle V \rangle + V_0} = \frac{1}{2(\langle V \rangle + V_0)} \log\left(1 + \frac{\langle V \rangle}{V_N}\right)$$

Optimizing $\langle V \rangle$ depends on accessory volume V_0.

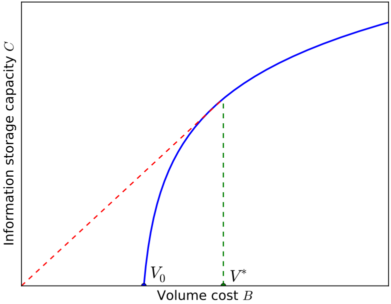
When $V_0 \ll V_N$, capacity is optimized by $\langle V \rangle = \sqrt{V_0 V_N}$.

Exact dependence $\rightarrow$



Determining the optimal B

- V_0 is small, so $\langle V \rangle$ should be such that SNR < 1.
- V_0 is not infinitesimal, so $\langle V \rangle$ should be such that SNR is not infinitesimal.
- Similar conclusion holds for general channels, by concavity of $C(B)$.



Answer to the First Question

Optimization Principle 1

To optimize information storage capacity per unit volume of neural tissue, synapses should be **small** and **noisy** on average.

Several experimental studies show similar results:

- Average SNR is less than one, but not infinitesimal.

Noisy synapses maximize information storage capacity!

Establishing Synaptic Efficacy Distribution

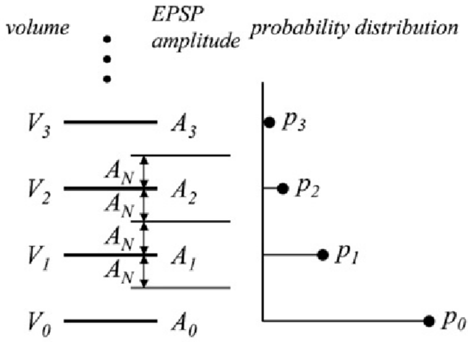
- Want a full distributional characterization of the synaptic weights: the capacity achieving input distribution.
- AWGN assumption is not precise.
  - non-negativity
  - It is unlikely that $\alpha = 2$.
- A full input-output characterization of the synaptic memory channel is not available.
  - Look at **approximations**!

ε-Capacity Cost Approximation

To yield a exactly solvable optimization problem:

- Assume noise is approximately bounded (by ε).
- Partitioning input space so that $\mathcal{X} = \{0, 1, \cdots, \}$.

Convert the memory channel to a **zero-error, discrete input** channel. [Kolmogrov & Tihomirov, 1959; Root, 1968]



Maximum Entropy Method

Channel is noiseless $\Rightarrow$ the input distribution with maximum entropy achieves capacity.
Need to maximize entropy $H(X)$ per average volume $\bar{V} = \langle V \rangle + V_0$:

$$\begin{aligned} \text{maximize} \quad & I_{\text{synapse}} = H(X) \\ \text{sub. to} \quad & \sum_x P(x)V(x) = \bar{V} \end{aligned} \quad \Rightarrow \quad P_{opt}(x) = \exp(-\beta V(x))$$

β satisfies $\sum_x P_{opt}(x) = 1$.

The storage capacity per unit volume:

$$\frac{I_{\text{synapse}}}{\bar{V}} = \frac{\beta \sum_x V(x) \exp\left(-\beta V(x)\right)}{\sum_x V(x) \exp\left(-\beta V(x)\right)} = \beta$$

Equidistant Assumption & Filling Fraction

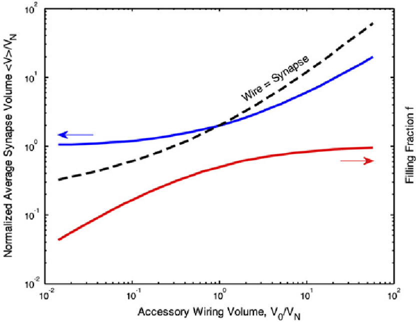
Motivated by experimental observations, assume synaptic state volume is distributed equidistantly:

$$V(i) = V_0 + 2iV_N, \quad i = 0, 1, 2, \cdots$$

$\Rightarrow$

$$\langle V \rangle_{i>0} = 2V_N \exp(\beta V_0)$$
$$f = 1 - \exp(-\beta V_0)$$

where f is the fraction of synapses in states $X > 0$.



Optimal achieved when V_0 is as small as possible:

$\Rightarrow \; f \ll 1$

$\Rightarrow \; \langle V \rangle_{i>0} \approx 2V_N$

Sparse Synaptic Connectivity

Optimization Principle 2

To optimize information storage capacity per unit volume, the filling fraction should be **small**. Small filling fraction is equivalent to **sparse synaptic connectivity**.

- The filling fraction is between 0.1 and 0.3 for various brain regions in various mammals (Stepanyants *et al.*, 2002).
- Using experimentally determined values of V_0, the predicted and actual filling fractions are quite close to each other.

Distribution of Volume and Efficacy

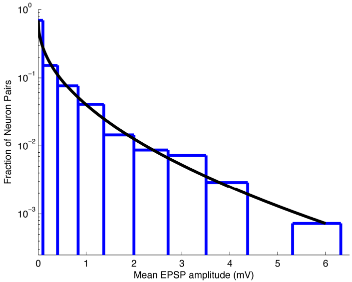
- The capacity achieving volume distribution:

$$P(x) = \exp(-\beta V(x))$$

- Combining with relationship between volume cost and efficacy yields the capacity achieving input distribution.

$$P(x) = \exp\left(-\beta V_N \left(\frac{x}{A_N}\right)^{\alpha}\right)$$

- Stretched exponential with the stretching factor determined by α.



Theoretical fit of experimental data:

$$\alpha = 0.79$$

Optimization Principle 3

To optimize information storage capacity per unit volume, although there will be many absent synapses and **numerous small synapses**, there will also be **some large synapses**, with synaptic efficacy distribution like a **stretched exponential.**

Inferring the Input Cost Function

– Previous result was based on the ε-capacity cost approximation.
– Alternative: inverse information theory approach.
– What is the cost function $b(\cdot)$ for which the measured $P(X)$ and $P(Y|X)$ are optimal?

$$C(B) = \max_{P(X):E[b(X)]\leq B} I(P(X); P(Y|X))$$

Choose cost function $b(\cdot)$ so a fixed $P(X)$ is the optimal solution.

$\Rightarrow$ Under proper conditions [Gastpar, 2003]:

$$b(x) = vD(P_{Y|X}(\cdot|x)\|P_Y(\cdot)) + v_0$$

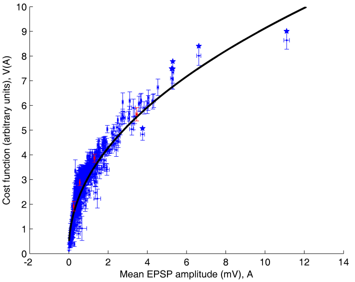
where $v > 0$ and v_0 is an arbitrary constant.

Inverse Information Theory Approach

AWGN channel:

$$V(x) \sim x^2$$

($\sim$: equality up to affine transformation.)



Discrete noiseless channel:

$$V(x) \sim -\log P(x)$$

– Obtain $P(Y|x)$ and $P(Y)$ from experiment data.
– Calculate $V(x)$ accordingly.
– Power law fit of of $V(x)$ gives:

$$\alpha = 0.77$$

Agrees with the previous approach!

Remarks on Modeling

- Optimality in biological systems
  - Does evolution find optimal design through optimization?
  - Causality versus correlation
- Caveats in modeling systems
  - Follow scientific method
  - Create testable hypotheses and rigorous experiments
  - Can only be validated, never proven (falsifiability)

Discussion

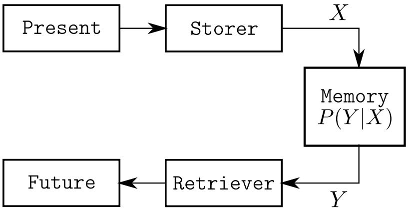
"The strength of the information theory approach is that it provides an upper bound on information storage... without considering how memory is stored, retrieved, coded, or decoded. Regardless of whether error-correcting codes are used or not, the capacity achieving input distribution must be used for optimal performance." [Varshney, Sjöström & Chklovskii (2006)]

Summary

- Developed a theory of information storage in the brain using an information theoretic optimization
  - Optimal distribution of synaptic weights
  - Volume and structure from its weight distribution
    - No restrictive model assumptions
- Validates (unintuitive) observations from experiments
- Predicts relationships on uninvestigated aspects of synapses
  - α as power-law relationship between synapse volume and efficacy.
  - β as parameter in synapse volume distribution.

Issues

Memory channel:



Memory as a Channel:

- Is each synapse a channel use?
- How does this relate to long-term memory

Issues

The model holds only when memory is the main function of the synapse

- For information transmission, synapses may be large and reliable
- This is our discussion for next time...

T. Berger & W. Levy, “Information Transfer by Energy-efficient Neurons,” 2009.