INFORMATION THEORETIC ANALYSIS OF A BIOLOGICAL SIGNAL TRANSDUCTION SYSTEM

SHU WANG

Submitted in partial fulfillment of the requirements for the degree of

Master of Science

Department of Biology

CASE WESTERN RESERVE UNIVERSITY

August 2018

CASE WESTERN RESERVE UNIVERSITY

SCHOOL OF GRADUATE STUDIES

We hereby approve the thesis/dissertation of

SHU WANG

candidate for Master of Science degree

Committee Chair

Dr. Jean Burns

Committee Member

Dr. Peter Thomas

Committee Member

Dr. Jessica Fox

Committee Member

Dr. Alan Levine

Date of Defense

03/27/2018

*We also certify that written approval has been obtained

for any proprietary material contained therein.

Table of Contents

Figures

Tables

Information Theoretic Analysis of A Biological Signal Transduction System

Abstract

by

SHU WANG

Signal transduction is a process by which living cells detect and respond to changes in their environment, including signaling, reception, transduction, and response reactions. Information theory as introduced by Shannon uses mutual information, the difference between entropy of output and entropy of output given input, to estimate uncertainty in message transmission processes. Neurotransmission is a typical signal transduction process in which the neurotransmitter works as a signaling molecule, carrying information from the source to an associated receptor. Viewing the input and output signals as point processes, we apply information theory to analyze its communication capabilities. This thesis provides a detailed glutamate simulation procedure and its mutual information analysis in MATLAB, and describes how information theory applies to a neurotransmission simulation.

Chapter 1 --- Introduction of Information Theory and Signal Transduction

1.1 Signal Transduction in Biological Systems

A cell is highly responsive to specific chemicals in its environment. For examples, hormones are chemical signals that tell a cell to respond to a change in conditions, or molecules in food communicate taste and smell through their interaction with specialized sensory cells.[1] From the detection of biochemical molecules to bodily reactions to stimuli, it is all completed by signal transduction, the process by which a chemical or physical signal is transmitted through a cell as a series of molecular events, resulting in a cellular response.[2]

Berg described signal transduction that normally occurs in biological systems.[1] When the signal molecules are released after stimuli generation, such as hormones or other biochemical molecules, particular proteins located at the cell surface detect (or "receive") chemical signals. When signal molecules bind to particular proteins (receptors) on cell membranes, the association activates a conformational change in the proteins, which is transmitted to the cytoplasmic domain or part of the receptor molecule. This transduction further starts a series of chain reactions causing cellular response, such as a change in gene expression, ion permeability, enzyme activity or protein three-dimensional structure, which ultimately affects the metabolism of the cell or organism.[3] Meanwhile, depending on the efficiency of the nodes, a signal can be amplified so that one signal molecule can generate responses involving hundreds of millions of molecules.[4] In this thesis, I essentially focus on two aspects to simplify the investigation, which are the signaling, and reception processes. We don't address

response, amplification and feedback here.

This complicated system is composed of several components. First of all, in order to respond to changes in their biological environment, cells must be able to receive and process signals generated outside of their borders. On one hand, some cells are sensitive to mechanical stimuli, such as mechanotransduction, which can be triggered by calcium-dependent cell adhesion molecules in the service of hearing, touch, proprioception and balance.[5] On the other hand, most cell signals are chemically mediated processes, which makes biochemical molecules, things like growth factors, hormone, neurotransmitters, to work as ligand to develop their biochemical effects locally or travel over long distances.[6]

The second component is biochemical molecules' associated receptors, that play an essential role in interaction with ligand in target cells. Based on the range of locations over which the signal spreads, receptors can be divided into two categories, intracellular receptors, which are found inside of the cell, such as that in the cytoplasm or nucleus, and cell surface receptors, which are located on the plasma membrane. Both of them induce a change in the conformation of the inside part of the receptor, a process called "receptor activation".[7] The receptor that most studies concentrate on is the latter type, such as G protein-coupled receptor and ion channel linked receptor.

The last important component is the second messengers. The most familiar examples of second messenger molecules include cAMP, cGAMP, Ca^{2+}, and inositol triphosphate. An essential feature of second messenger signaling systems is that

second messengers can link downstream reactions to multi-cyclic kinase cascades in order to amplify the strength of the original signal.[8] This is the reason why an association of ligand with its one type receptor could evoke multiple downstream responses in gene expression and protein production in different target cells and tissues. The three components described above, signal molecules (ligands), receptors (first messengers), and second messengers form the main constituents of signal transduction processes.

1.2 Neurotransmission

In biological systems, neurotransmission is one typical type of signal transduction, in which neurotransmitters act as the ligand species. Neurotransmission is a communication between neurons that is accomplished by the movement of chemicals or electrical signals across synapses. According to the stimuli types, it can be divided into two groups, which are electrical neurotransmission, communication occurring between two neurons at electrical synapses (or gap junctions), and chemical neurotransmission, communication occurring at chemical synapses.[9] Since the structure of the latter one is similar to that of signal transduction, I decided to use it as my simulation model in this thesis.

The chemical synaptic neurotransmission is composed of three main parts, presynaptic neuron, neuron cleft, and postsynaptic neuron (***Figure 1***). The presynaptic neuron manufactures neurotransmitters, and stores them in presynaptic vesicles. When the presynaptic neuron generates an action potential by stimuli, the action potential reaches the axon terminal, which causes a calcium current and subsequently causes a vesicle to release neurotransmitter into the synaptic cleft. The synaptic cleft is a small gap separating the presynaptic neurons and postsynaptic neurons, and filled with extracellular fluid. After its release, the neurotransmitter diffuses across the cleft and binds to receptors located on the postsynaptic membrane. If the activated receptors trigger an inward (depolarizing) current, the signal from the synapse may contribute to generating a postsynaptic action potential in the postsynaptic neuron. If the activated receptors trigger an outward (hyperpolarizing) current, or open a

shunting conductance, the signal from the synapse may suppress action potential generation. Additionally, the neurotransmitter could be either destroyed enzymatically, or absorbed by the terminal from which it comes as *via* a reaction.[10,11]

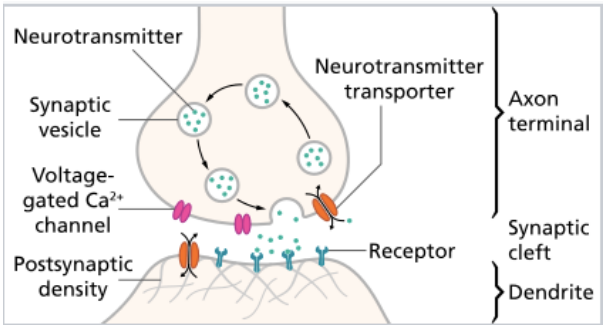

Figure 1 Principles of neurotransmission. The presynaptic neuron (on the top) synthesizes and releases neurotransmitter, which associates with and activates receptors at the postsynaptic neuron membrane (at the bottom).[12] (This figure is credited by Thomas Splettstoesser on www.scistyle.com.)

In the neurotransmission process we consider, the neurotransmitter acts as the ligand, the first component involved in chemical reactions and triggering downstream response. Normally, the neurotransmitter could be classified by molecule type. For an example, amino acid neurotransmitters include glutamate (Glu), aspartate, γ-aminobutyric acid (GABA).[13] They are all packed in the presynaptic vesicles before releasing. Nitric oxide (NO) and carbon monoxide (CO) are common gasotransmitters.[14] Some monoamines and peptides, such as dopamine (DA) and adenosine triphosphate (ATP), can work as neurotransmitters as well. In biological function, some neurotransmitters have excitatory affect on postsynaptic neuron activities, while some of them work as inhibitor.

The other component we consider is the neurotransmitter's associated receptors located on the postsynaptic membrane. According to the molecule structure, they can be classified into ligand-gate ion channels (ionotropic receptors) and G protein-coupled receptor (metabotropic receptors).[15] The main difference between them is their mechanism in interaction with the neurotransmitter. Ionotropic receptors

have an ion channel pore. When it is activated, it makes ion channels located on the postsynaptic membrane switch from a closed to an open state, to allow Na^+, K^+, or Cl^- to flow.[16] In contrast, metabotropic receptors need G-proteins to be involved, to provide a link with ion channels and to trigger a series of second messenger reactions in order to make ion channels open and generate ion currents.[15]

1.3 Information Theory

Information theory studies the transmission, processing, extraction, and utilization of information. It was originally developed by Claude E. Shannon in 1948 in a landmark paper entitled “A Mathematical Theory of Communication”. In this paper, Shannon conceived “information” as a set of possible messages. The goal was to send these messages over a noisy channel, and then to have the receiver reconstruct the message with low probability of error in spite of the channel noise.[17] The significant process was examining the rate of information transported through communication system.

In Shannon’s paper, he described a schematic diagram of a general communication system. ***Figure 2*** shows that the communication system consists of essentially five parts, which are the information source, transmitter, channel, receiver, and destination. The information source is the place where the “sender” produces a message or sequence of messages to be transmitted to the receiving terminal. Any system capable of taking on multiple states can be a source of information, and that state can be mathematically represented as a random variable that can take on multiple values. A transmitter is a mediator to operate on the message in some way to produce a signal suitable for transmission over the channel. After the transmitter transports messages generated by the source, the messages are transmitted from transmitter to receiver to go through the channel. At the end of channel, the receiver performs the inverse operation of that done by the transmitter, and reconstructs the messages from the signal. Finally, the messages are detected as output at the destination. A channel is

"noisy" if the output of the channel (received signal) is only determined probabilistically by the channel input. If a random variable X represents the input and another variable Y represents the output, then the conditional probability $P(Y|X)$ describes the (noisy) signal transmission process.

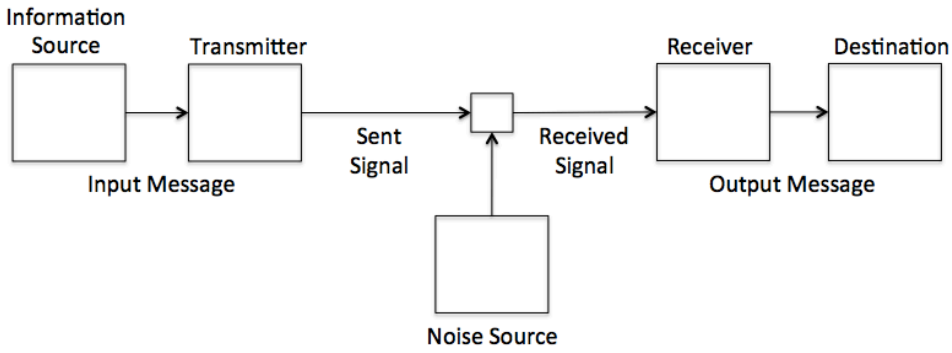

Figure 2 Schematic diagram of a general communication system by Shannon, including 5 essential components: information source (message production), transmitter (an operator to transport messages), channel (a medium for message transportation), receiver (receiving messages), and destination (output message detection). [17] (The figure was redrawn from reference [17]. The copyright was reserved @1948 by The Bell System Technical Journal.)

The main purpose of information theory is calculating information utilization in a message transmission process. Shannon used an important concept, entropy, to represent the amount of uncertainty involved in the value of a random variable or the outcome of a random process. Named after Boltzmann's H-theorem, Shannon defined the entropy H of a discrete random variable X, which can take on the values $\{x_1, ..., x_n\}$ with the respective probabilities $p_1, p_2, \ldots, p_n$, as showed in ***equation (1)***

$$H(X) = -\sum_{i=1}^{n} p_i log_b p_i \qquad (1)$$

where b is the base of logarithm used. By the usual convention, he defined entropy using a base 2 logarithm so that the entropy was measured in bits. A message carrying a single bit of information may be thought of as a signal that is sufficient to answer

exactly one yes/no question posed so that the outcomes “yes” and “no” occur with equal 50/50 probability.[18]

However, in communication systems, the messages are transmitted from one place to another. Because the input is considered to be a random variable, the output whose value depends on the input is also a random variable. Information theory calculates if the output messages allow the input messages sent through the channel to be fully or partially determined. Mutual information is used to quantify this in terms of the amount of information that the value of one random variable contains about the value of another random variable. If X defines an input random variable taking on the values of $\{x_1, \ldots, x_n\}$ and Y defines an output random variable taking the values from $\{y_1, \ldots, y_n\}$, respectively, then the mutual information in communication system $I(Y;X)$ is defined as ***equation (2)***.

$$I(Y;X) = H(X) - H(X|Y) \qquad (2)$$

where $H(X)$ designates the entropy of input discrete random variable, while $H(X|Y)$ defines the entropy of input events given the output events, also called conditional entropy. The conditional entropy is the entropy of the conditional distribution of X, given a particular value of Y, averaged over the various values of Y. Since entropy is valued in bits, the mutual information is measured in bits as well.

By using the same method, the following formula, ***equation (3)*** is also available. In conclusion, the mutual information between output and input in a communication system is the difference of entropy of input variable and the conditional entropy of input event given output event, meanwhile, it also equals to the difference between

entropy of output events and the conditional entropy of output events given input events. The mutual information represents the information communicated from source to receiver.

$$I(Y;X) = H(Y) - H(Y|X) \quad (3).$$

Shannon's article introduced a mathematical framework for information theory. The original purpose lay in the field of telecommunications engineering, namely, to quantify the amount of information transmitted through a noisy engineered communication channel, such as telephone or telegraph wires.[18] While the physical sciences and engineering fields provided many applications for information theory, it has found relatively fewer applications in the life sciences. One significant application of information theory to the life sciences was Henry Quastler's research. Quastler pointed out that the DNA in a mammalian cell had an information capacity of roughly 2×10^{11} bits. He claimed that this observation implied that the level of complexity or organization of an organism could be quantified in units of bits.[19]

Despite the initial enthusiasm, however, many attempts at applying information theory within the biological science were ultimately considered to be disappointing. For example, Johnson criticized the careless application of information theoretic methods.[20] Johnson argued that information theoretic analysis lacked two key features, namely the omission of qualitative factors and non-applicability to open systems. "Qualitative factors" refer to conditions in which some bits (of information) are more significant for an organism's survival than others; Shannon's purely quantitative formulation treats all bits as having equal importance. Moreover,

Shannon's construction applies to closed systems while biological systems are thermodynamically open, interacting with their environment in a way that can interfere with strict information theoretic inequalities.

Nevertheless, the application of information theoretic analysis to biological signaling systems has seen a growing number of successes in recent years. Principled application of Shannon's framework has provided new insights into information processing in areas as diverse as neural networks[21], evolutionary genetics, and signal transduction networks. For examples, Tkacik's work showed that the mutual information between bicoid and hunchback concentrations was helpful to distinguish the anterior-posterior embryonic axis in the development of the fruit fly, *Drosphila melanogaster*.[22] As another example, in Cheong and Rhee's article, "Information Transduction Capacity of Noisy Biochemical Signaling Networks", the authors concluded that tumor necrosis factor (TNF) signal transduction had a *bottleneck*, limiting the information gained via downstream multiple integrated pathways. This conclusion about the topology of the signaling network was made possible by quantitatively measuring the mutual information between the input signal (TNF concentration) and response levels (NF-*k*B and ATF-2 expression) in individually resolved cells.[23]

1.4 Information Theory Applied to Signal Transduction

Information theory is a field at the intersection of mathematics, statistics, and computer science. It also finds application in other areas, including statistical inference, neurobiology, and model selection in statistics.[24] Taking Johnson's critique into account, a successful information theoretic analysis in a biological system should address qualitative as well as quantitative factors, within a thermodynamically closed system. As my goal in this thesis is to explore the application of information theoretic analysis to signaling pathways, I sought examples in which a complicated signaling pathway could be treated with a simplified representation: a small volume, whose size (area and length), signal molecule concentration, signal molecule/response receptor kinetic scheme composed of association and disassociation rate, decay rate should have been measured experimentally. Each of these parameters potentially could affect the results of the analysis. ***Table 1*** provides parameters of several typical signaling molecules from different biological systems. For each of these systems the literature provides a range of typical length scales over which signaling occurs, as well as receptor binding and unbinding kinetics. On one hand, in some cases the relevant parameter ranges can be quite broad, for instance cytokines in the immune system operate on length scales ranging from inter-capillary distances to whole body length scales. On the other hand some signaling molecules are similar to those shown in the table, for example acetylcholine has a similar molecular size as glutamate. Although the methods I develop here should apply to any signaling system, in order to pursue a concrete

example we focus on the glutamate/NMDA receptor in mammalian synapses. If my simulation of the glutamate signaling pathway is successful and proves biologically meaningful, it should provide a framework for analyzing any existing signal system as long as its parameters could be examined and collected.

Table 1 Geometry and kinetics for four molecular signaling pathways.**

Molecule	**Glutamate**	**GABA**	**IL-1**	**IL-8**
Receptor	NMDA receptor	$GABA_A$ receptors	$IL1R_a$	CXCR1
Travel distance	$20nm$	$20nm$	$10nm{\sim}1m$	$10nm{\sim}1m$
Binding rate $(\mu M^{-1}ms^{-1})$	$5{\times}10^{-3}$	$2.1{\times}10^{-2}$	$7.17{\times}10^{-4}$	$2.73{\times}10^{-4}$
Unbinding rate (ms^{-1})	$6.7{\times}10^{-3}$	$5.1{\times}10^{-1}$	$2.08{\times}10^{-8}$	$2.35{\times}10^{-6}$

**: Glutamate parameters are obtained from reference [25], [26]. GABA parameters are obtained from reference [27]. IL-1 parameters are derived from reference [28]. And IL-8 parameters are derived from reference [29]. Additionally, travel distance refers to the typical distance separating the source and receptor.

As an example of the framework of communication system described in the theory, the neurotransmission systems enjoy a similar structure. The neurotransmitter works as the information or message carrier, produced by the presynaptic neuron, and detected by the postsynaptic neuron, which are the information source and receiver, respectively. The synaptic cleft between presynaptic and postsynaptic membrane provides the physical channel for neurotransmitter-based communication. Therefore it is not unreasonable to investigate a neurotransmission system using quantitative

simulations together with information theory.

In this thesis, I will focus on one particular neurotransmitter transduction process, the glutamate signaling pathway, to which I will apply an information theoretic analysis. Glutamate is the most abundant neurotransmitter in the vertebrate nervous system. It is used by every major excitatory function in the vertebrate brain, accounting in total for well over 90% of the synaptic connections in the human brain.[30] It has three classes of biochemical receptors, AMPA receptors, NMDA receptors, and metabotropic glutamate receptors. The first two types are both ionotropic receptors. On the one hand, AMPA receptors are specialized for fast excitation, which results in producing excitatory electrical responses in their targets in a fraction of a millisecond after being stimulated. On the other hand, NMDA receptors differ from AMPA receptors in being permeable to Ca^{2+} when they are activated, which results in Ca^{2+} current flow through the membrane. Their properties make them particularly important for learning and memory.[31] For the third type of glutamate receptor, the metabotropic receptor, its biological function is mediated by a G-coupled protein and acts through second messenger systems to create slow, sustained effects on their targets. Since the mechanism of interaction with ionotropic receptors is simpler, and activated NMDA receptors can produce postsynaptic potentials caused by Ca^{2+} current flowing through the membrane, I decide to use NMDA receptors as glutamate receptors in the reception process.

To match with the information theory framework described in ***Figure 2***, the ***Figure 3*** gives a similar diagram showing the glutamate simulation process. The information

source comes from sequences of action potentials generated in the presynaptic neuron, which activates release of glutamate molecules from vesicles that diffuse across the synaptic cleft and bind with postsynaptic NMDA receptors. The interaction makes ion channels located on the postsynaptic membrane switch from a closed to an open state, resulting in an excitatory current flowing across the postsynaptic membrane. As a starting point for modeling the synapse as a communications channel, we take the incoming action potential arrival times to be samples of a Poisson process. At the same time, we consider the output of the channel to be the sequence of channel opening times, which is another point process (not necessarily Poissonian). We may think of the "destination" of the incoming "message" to be action potentials in the postsynaptic cell, but we do not include postsynaptic membrane excitability in our model or analysis. For our purposes the communications channel transduces the input timing sequence of presynaptic action potentials into the output timing sequence of NMDA channel opening events.

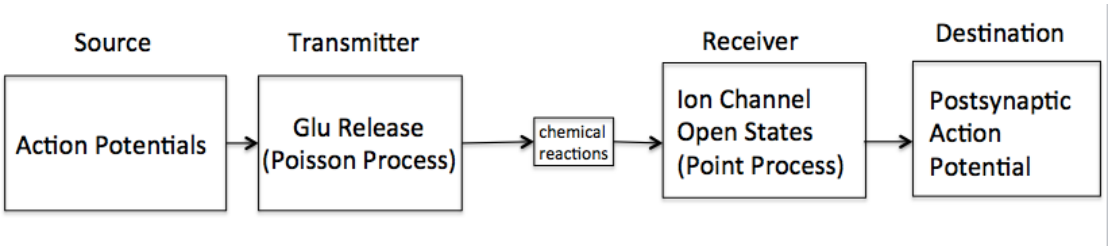

Figure 3 Description of glutamate simulation process applied in our information theory framework.

The process to transmit messages from transmitter to receiver is chemical reactions between glutamate and its NMDA receptors. Based on Lester's glutamate and NMDA receptor binding model, ***Figure 4*** is redrawn to show the kinetic scheme of glutamate and NMDA receptor association and dissociation processes.[26] It describes that one

NMDA receptor can bind with two glutamate molecules at most, which causes there to be three states of receptors existing on the postsynaptic membrane, unbound receptors (R), singly-bound receptors (GluR) and doubly-bound receptors (Glu_2R). During normal functioning, the ion channels switch to open states only when doubly-bound receptors transfer its structure to Glu_2R^*, but in this thesis, it is assumed that once doubly-bound receptors are produced, the ion channels open and activate Ca^{2+} current. So the kinetic scheme is simplified to two main steps: glutamate interacts with unbound receptors generating singly-bound receptors with kinetic rate $2k_{on}$, and singly-bound receptors associate with another glutamate producing doubly-bound receptors with kinetic rate k_{on}, and their backward reactions in kinetic rate k_{off} and $2k_{off}$, respectively.

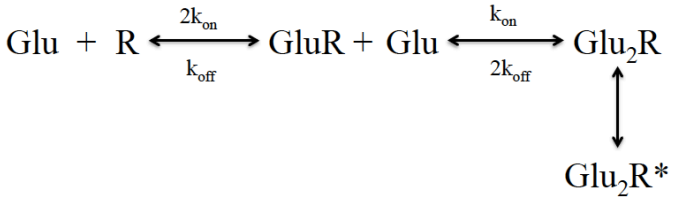

Figure 4 Glutamate and ionotropic receptors association and dissociation kinetic model. It provides the following description: "Glu, R, GluR, Glu_2R, Glu_2R^* represents glutamate, unbound state receptor, singly bound state receptor, doubly bound state receptor (ion channel closed), doubly bound state receptor (ion channel open), and desensitized state receptor, respectively. k_{on} and k_{off} are binding and unbinding rates of chemical reactions. There are also transfer rates between doubly bound state receptors, both in closed and open ion channel. "[32] Our version of the model omits doubly bound state receptors in open ion channel, where the transfer rates are both set to zero. (Redrawn from reference [26].)

After the simulation process is established, based on Shannon's work, I collected molecule numbers, and calculated the entropy of output and the entropy of output given input. In this simulation process, the input signals, action potential in

presynaptic neuron, are generated in Poisson random distribution, and the output messages, ion channels open states on postsynaptic membrane, are estimated as a point distribution. Although ***equation (1)*** provides a method to measure entropy of a discrete random variable sequence, McFadden proposed another method to estimate entropy especially for point processes, which is used to calculate entropy of output and entropy of output given input in glutamate simulation process. He derived a mathematical formula for the entropy of point process given by observing the rate $\beta(t)$ over a period of time length T, and proved that the rate of entropy over time satisfied ***equation (4)***. Hence the entropy over time length and the instantaneous entropy in time series satisfied ***equation (5)***.[33] Here $\beta(t)$ is event occurrence rate of time *t*. For a standard, unmodulated Poisson process, $\beta(t)$ is constant.

$$\frac{dH}{dt} = \beta(t)(1 - \log_2 \beta(t)) \quad (4)$$

$$H = \int_0^T \beta(t)(1 - \log_2 \beta(t))dt \quad (5)$$

Based on McFadden's method, the difference between two entropies (conditioned and unconditioned), which we can interpret in terms of how many messages could be transmitted in this process, with a given uncertainty of message transmission. Meanwhile, it can also be used in estimating input information by observing output events.

Chapter 2 Glutamate Simulation Procedure in MATLAB

2.1 Geometry in Simulation

Based on the conceptual model in ***Figure 3***, I constructed a numerical simulation of glutamate signal transduction as follows. A volume whose area was 100μm^2 and length was 10μm was established (***Figure 5***). In this volume, glutamate molecules were released from the left side at constant release rate λ, which was set to be 0.01 per millisecond. The times of glutamate release (triggered by incoming action potentials) served as the input messages. At the time of each glutamate release event, the mean glutamate molecule number released at one time was set up to be 3000. The exact number of glutamate released was a random variable, chosen using a Poisson distribution with the given mean. Once released, it was assumed that all glutamate molecules instantaneously spread throughout the whole volume and could bind with its NMDA receptors. NMDA receptors were located on the right side of the synaptic volume with a density of 600μm^{-2} , with kinetic rate, k_{on} and k_{off}, respectively for the binding and unbinding reactions. Binding on each of two receptor sites was assumed to be independent (no cooperativity or allosteric interactions) with identical kinetics. On the receiver side, the times at which any receptor entered the doubly-bound state (Glu_2R) was defined to be the output sequence. I took this to be the output sequence because only when the receptor was in this form, did the ion channels switch to the open state. At the same time, glutamate was eliminated by an enzyme or uptake system with a *per capita* decay rate γ, set to be 0.02 per molecule per millisecond. The whole process was set up to run for an interval of 3000

milliseconds, using a fixed simulation time step of 1 millisecond. All basic parameters set up for the simulation and their values are listed in ***Table 2***.

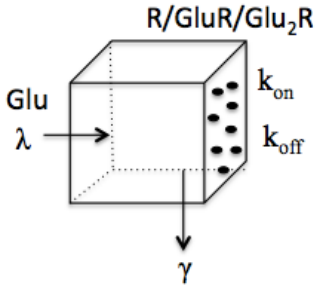

Figure 5 A model established for glutamate simulation process. Glutamate is released from the left side at constant rate λ, which provides the input to the information channel. Diffusing glutamate interacts with NMDA receptors on the right side with kinetic rates k_{on} and k_{off}, generating three receptor states, R, GluR, and Glu_2R. The times at which receptors enter the third state (Glu_2R) work as output information. During whole process, diffusing glutamate decay at *per capita* rate γ.

Table 2 All parameters and their values in the glutamate signaling simulation.[32]

Parameter	**Symbol**	**Value**
Area	area	$100\mu m^2$
Length	length	$10\mu m$
Mean Glu release number	N_G	3000
Receptor density	density	600 per μm^2
Release rate	λ	0.01 per msec
Decay rate	γ	0.02 per molecule per msec
Total time	t	3000 msec
Time step	dt	1 msec
Binding rate	k_{on}	5×10^{-3} per msec per μmolar
Unbinding rate	k_{off}	6.7×10^{-3} per msec

2.2 Initial Conditions Setting Up

In the glutamate simulation process, the release process, decay process, and chemical reactions involving receptors mentioned previously were divided into individual reaction in ***Table 3*** step by step. There were five reactions in total. The release process was defined as reaction number 0 with constant rate λ, where glutamate molecules were produced following arrival of action potentials at the presynaptic membrane. Reactions number 1 and number 2 were generation and degradation of singly-bound receptor (GluR) with rate $2k_{on}$ and k_{off}, while reactions number 3 and number 4 were production and dissociation of doubly-bound receptor (Glu_2R) with rate k_{on} and $2k_{off}$, respectively. At last, reaction 5 was considered as glutamate decay process in rate of γ. In the model, this process occurred only for free glutamate, not while glutamate was bound to the receptor.

Table 3 Chemical reactions in simulation and their kinetic rates

Chemical Reaction	Kinetic Rate
0: $0 \rightarrow Glu$	λ
1: $Glu + R \rightarrow GluR$	$2k_{on}$
2: $GluR \rightarrow Glu + R$	k_{off}
3: $GluR + R \rightarrow Glu_2R$	k_{on}
4: $Glu_2R \rightarrow GluR + R$	$2k_{off}$
5: $Glu \rightarrow 0$	γ

As ***Table 3*** shows, there were four types of molecules involving in the simulation, which were glutamate (Glu), unbound receptors (R), singly-bound receptors (GluR), and doubly-bound receptors (Glu_2R). Simulations began by choosing initial conditions for each molecule type. However, all molecules did not just begin with the fixed parameters set up in ***Table 2***, such as glutamate started with 3000 molecules,

unbound receptors begun with 60000 molecules, while singly-bound receptors and doubly-bound receptors were 0 in initial condition. In biological aspect, all chemical reactions will arrive at dynamic equilibrium after a sufficiently long time, where the average rate of each molecule's concentration change is 0, meaning each molecule's number in production and degradation are same in steady state. Since the initial condition should fit this steady state, the main idea was to set up initial conditions for each type of molecule by actually finding out the chemical reactions' steady states.

In dynamic equilibrium, the rate of each molecule type's concentration change was shown in ***equation (6)-equation (9)***, where the concentration of each molecule was expressed, in units of moles per liter, as [Glu], [R], [GluR], and [Glu_2R].

$$\frac{d[Glu]}{dt} = \lambda N_G - 2k_{on}[Glu][R] + k_{off}[GluR] - k_{on}[Glu][GluR] + 2k_{off}[Glu_2R] - \gamma[Glu] \quad (6)$$

$$\frac{d[R]}{dt} = -2k_{on}[Glu][R] + k_{off}[GluR] \quad (7)$$

$$\frac{d[GluR]}{dt} = 2k_{on}[Glu][R] - k_{off}[GluR] - k_{on}[Glu][GluR] + 2k_{off}[Glu_2R] \quad (8)$$

$$\frac{d[Glu_2R]}{dt} = k_{on}[Glu][GluR] - 2k_{off}[Glu_2R] \quad (9)$$

By setting all formulas equal to 0, we solve for the four equilibrium concentrations. The solution are listed in ***equation (10)-equation (13)***, where the concentration of glutamate in steady state could be solved directly in the first equation, and the last three equations were the fractions, $f_{[R]} + f_{[GluR]} + f_{[Glu_2R]} \equiv 1$, of each receptor state among the total receptor population set up in simulation geometry.

$$[Glu] = \frac{\lambda}{\gamma}N_G = \frac{0.01\ \text{msec}^{-1}}{0.02\ \text{msec}^{-1}} \times 3000\ \text{molecules} = 1500\ \text{molecules} \quad (10)$$

$$f_{[R]} = \frac{{k_{off}}^2}{{k_{off}}^2 + 2k_{on} \times k_{off}[Glu] + {k_{on}}^2[Glu]^2} = 0.9965 \quad (11)$$

$$f_{[GluR]} = \frac{2k_{on} \times k_{off}[Glu]}{{k_{off}}^2 + 2k_{on} \times k_{off}[Glu] + {k_{on}}^2[Glu]^2} = 0.0035 \quad (12)$$

$$f_{[Glu_2R]} = \frac{{k_{on}}^2[Glu]^2}{{k_{off}}^2 + 2k_{on} \times k_{off}[Glu] + {k_{on}}^2[Glu]^2} = 3.1139 \times 10^{-6} \quad (13)$$

Since the glutamate initial condition was expected to show a Poisson random distribution (*poissonrnd,* in Matlab), and the assignment of each receptor state among total unbound receptors located on receiver side was naturally represented as a multinomial random distribution (*mnrnd,* in Matlab), the initial conditions of the four types molecules could be set up in **command (1)** and **command (2)**.

$Glu_{initial} = poissrnd(Glu_{steady\ state})$ command (1)

$Receptors_{initial} = mnrnd(R_{total}, [f_{[R]}, f_{[GluR]}, f_{[Glu_2R]}])$ command (2)

2.3 Molecules Number Collection

To calculate message utilization in glutamate signal transduction, each type of molecule's abundance should be collected in the simulation. It was an easier way to collect molecule counts by tracking the number of each chemical reaction's occurrence events in the simulation. Based on all chemical reactions shown in ***Table 3***, the production and degradation of each molecule could be expressed in ***Table 4***. Each row, called the "stoichiometry vector" for the corresponding reaction, was used to described variation of each involved molecule when there was one reaction occurrence. For an example, in reaction 1, there was one glutamate associating with one unbound receptor for each singly-bound receptor generation, so the effect of reaction 1 was expressed as -1 both in glutamate and unbound receptor, and +1 in singly-bound receptor, where minus represents reactants, plus represents products.

Table 4 Number change for each type of molecule in each chemical reaction

Reaction	Glu	R	GluR	Glu_2R
0	1	0	0	0
1	-1	-1	1	0
2	1	1	-1	0
3	-1	0	-1	1
4	1	0	1	-1
5	-1	0	0	0

Table 4 described each chemical reaction's occurrence rate in the simulation. For the association reactions, the rate of reaction events was not the kinetic rate shown in ***Table 2***. Considering the unit of chemical reaction's occurrence rate should be reaction event number per millisecond, the units in reaction number 1 and number 3 should be converted from micromolar per second to molecules per msec, by dividing

out a factor of $602\times Volume$, where $Volume$ is the volume of the synaptic cleft expressed in cubic micrometers (cubic microns) by cancelling micromolar unit provided by k_{on} in ***equation (14)***.

$$1\mu M = \frac{\frac{\#\ of\ molecule}{6.02\times10^{23}}mol}{Volume(\mu m^3)} = \frac{\frac{\#\ of\ molecule}{6.02\times10^{23}}mol}{Volume\times10^{-18}(m^3)} = \frac{\#\ of\ molecule}{6.02\times10^{5}\times Volume}\frac{mol}{m^3} = \frac{\#\ of\ molecule}{Volume\times602}micromolar \qquad (14)$$

Then the expected number of reaction occurrences in a small (1 msec) time step was calculated by multiplying each reactant's number and the reaction occurrence rate. While the release process was modeled by a Poisson random distribution ($poissrnd$, in Matlab), all other chemical reactions were modeled using the appropriate binomial random distribution ($binornd$, in Matlab). The reaction occurrence number in time series could be expressed in **command (3)-command (8)**.

Table 5 Occurrence rate of each chemical reaction

Chemical Reaction	Occurrence Rate
0: 0 → Glu	λ
1: Glu + R → GluR	$2k_{on}/(Volume\times602)$
2: GluR → Glu + R	k_{off}
3: GluR + R → Glu_2R	$k_{on}/(Volume\times602\ )$
4: Glu_2R → GluR + R	$2k_{off}$
5: Glu → 0	γ

$rxn\_0 = poissrnd(N_{Glu} * \lambda * dt)$ command (3)

$rxn\_1 = binornd(n\_Glu * n\_R, ((2 * kon * dt)/Volume)/602)$ command (4)

$rxn\_2 = binornd(n\_GluR, koff * dt)$ command (5)

$rxn\_3 = binornd(n\_GluR * n\_Glu, ((kon * dt)/Volume)/602)$ command (6)

$rxn\_4 = binornd(n\_Glu2R, 2 * koff * dt)$ command (7)

$rxn\_5 = binornd(n\_Glu, \gamma\ * dt)$ command (8)

From each reaction occurrence number, it can collect each molecule number in simulation based on ***Table 4*** in ***equation (15)-equation (18)***.

$$N_{Glu} = N_{rxn_0} - N_{rxn_1} + N_{rxn_2} - N_{rxn_3} + N_{rxn_4} - N_{rxn_5} \quad (15)$$

$$N_R = -N_{rxn_1} + N_{rxn_2} \quad (16)$$

$$N_{GluR} = N_{rxn_1} - N_{rxn_2} - N_{rxn_3} + N_{rxn_4} \quad (17)$$

$$N_{Glu_2R} = N_{rxn_3} - N_{rxn_4} \quad (18)$$

Chapter 3 Simulation Results and Analysis

3.1 Molecules Number Curve

Figure 6 shows four types of molecule number time series in one simulation trial. There was one spike at the time when there was one glutamate release reaction occurrence (top panel). After glutamate was released, the unbound receptors decreased (second panel) while singly-bound receptors increased in molecule number (third panel), which was the same as conditions in chemical reactions. But the time points of their increase trend and decrease trend had a time lag compared with that of glutamate molecule number release times. This was consistent with biological perspective that it should take target cells a little time to detect signal ligands and activate interaction of ligands with receptors. Although unbound receptors and singly-bound receptors molecule number both had smooth curves, small fluctuations from time step to time step could be seen.

Compared with the curves for unbound receptor and singly-bound receptor, there were only few events observed in the doubly-bound receptors. This means that, for the parameters used in the simulations, not all glutamate release actions could generate the same number of doubly-bound receptor occurrence events ultimately. For example, in the first panel of ***Figure 6***, there were 25 glutamate release events (blue stars), while there were only 7 double-binding events in the fourth panel (red stars). After the first event seen in the fourth panel, there was really long time vacant approximately 250msec to 1600msec, when there were several release events visible in the first panel, which meant in this time period, the target cells could not transfer

messages carried by signal ligands and make responses. These effects illustrated one way in which information was lost during signal transduction. Information was also lost due to variability in the lag time between release and double-binding events, when those events did occur. The goal of an information theoretic analysis was to quantify the loss of information (or the incomplete gain of information) during the signal transduction process. So the next step was calculating mutual information in the simulation process. Heuristically, information provided an idea of how many messages were transmitted and the uncertainty of information transmission.

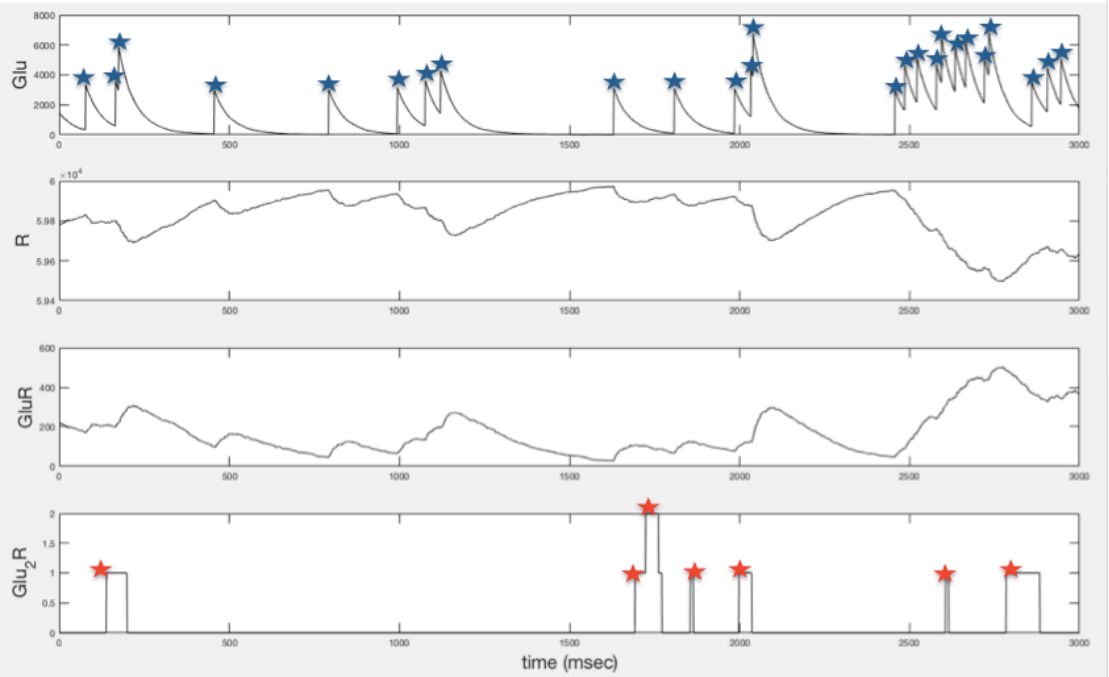

Figure 6 Four types of molecule number curves in one simulation trial. In the panel from top to bottom, there was molecule number of glutamate (Glu), unbound receptor (GluR), singly-bound receptor (GluR) and doubly-bound receptor (Glu_2R). Blue star: release occurrence events in glutamate release reaction. The sequence of times of these glutamate release events provided the input to the information channel. Red star: double-binding occurrence events in doubly-bound receptor generation reaction. The times at which double-binding events occurred conveyed information about the input. And durations of bound interval and time of unbinding did not convey information about the input. Thus the output of the channel was taken to be the time sequence of double-binding events.

3.2 Entropy Calculation

3.2.1 Multiple Input Events Datasets

Before doing entropy analysis work, multiple input event simulations were established. Chapter 2 described the simulation process for one single input trial. In order to collect an ensemble of input and output data, I made input events have 20 different trials, and output events have 10 trials, which meant there were 20 different input event sequences (glutamate release reaction sequences), and 10 different output event sequences (doubly-bound receptor generation sequences) for each input event sequence. If X was defined as an input message, the glutamate molecule number in time series, which took on the values $\{X_{1,1,i}\ ,\ X_{1,2,i}\ ,\ X_{1,3,i}\ ,\ \ldots\ X_{j,k,i}\}$, then Y was considered as output message, the doubly-bound receptors molecule number in time series, which derived from $\{Y_{1,1,i}\ ,\ Y_{1,2,i}\ ,\ Y_{1,3,i}\ ,\ \ldots\ Y_{j,k,i}\}$, where j=1,2,3,…, n_{input}=20, indexed the input simulation trials, k=1,2,3,…, n_{output}=10, indexed the output simulation trials, and i=0,1,2,3,…,3001, was the index of the time series, each increment in i representing a single time step of duration 1 msec.

The difference between the input event sequences was the release reaction occurrence times. As ***Figure 7*** showed, in first 5 input event sequences of 20 input simulation trials, each spike represented one occurrence event in glutamate release process. Since glutamate release process was modeled with a Poisson random distribution, the average spike number of release reaction was 30, but the system could not guarantee the release events to be consistent in each trial. So it showed that there were more than 30 spikes in some input simulation trials, while some of them

had fewer than 30 spikes. Meanwhile, in a fixed output simulation trial, the release reaction time should be totally the same, which implied that in biological system, the action potentials from presynaptic membrane was generated at same time point, but the glutamate concentrations released out of vesicles were different as input information. As previously mentioned, the glutamate initial condition was modeled with a Poisson random distribution, so glutamate released molecule number changed in variation with the mean of it derived from solution in ***equation (10)***.

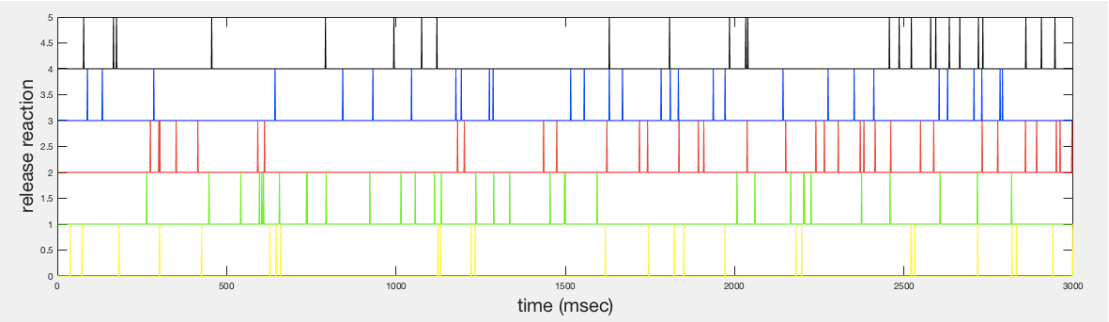

Figure 7 Release reaction events in time series of first 5 input simulation trials among total 20 input simulation trials. From the top to bottom are release reaction events in input simulation trial #1 to #5 colored in black, blue, red, green and yellow lines, respectively. Each spike represented one glutamate release reaction occurrence.

3.2.2 Entropy Analysis

Based on McFadden's method of measuring the entropy of a random point process in ***equation (4)-equation (5)***, the entropy of the output event time series, and the entropy of the output given an input event time series, can be estimated from multiple simulation trials by the following equations.

Since the output messages only came from transition from singly-bound receptor to doubly-bound receptor, the output event occurrence rate, $\beta(t)$, was defined to be the occurrence rate of chemical reaction number 3. So $\beta(t)$ could be estimated in ***equation (19)*** and ***equation (20)***.

$$\beta(t) = \frac{k_{on} \times N_{Glu}(t) \times N_{GluR}(t)}{Volume \times 602} \quad (19)$$

$$\beta_{j,k,i} = \frac{k_{on} \times N_{Glu_{j,k,i}} \times N_{GluR_{j,k,i}}}{Volume \times 602} \quad (20)$$

By fixing input simulation trial, the average $\beta(t)$ of all output simulation trials in time series was defined as $\beta_{j,i}$ in ***equation (21)***.

$$\beta_{j,i} = \frac{1}{noutput} \sum_{k=1}^{noutput} \frac{k_{on} \times N_{Glu_{j,k,i}} \times N_{GluR_{j,k,i}}}{Volume \times 602} \quad (21)$$

On one hand, we may define β_i as the average $\beta_{j,i}$ of all input simulation trials in time series in ***equation (22)***. By applying β_i to ***equation (5)***, the ***equation (23)*** showed the entropy of output $H(Y)$.

$$\beta_i = \frac{1}{ninput} \sum_{j=1}^{ninput} \beta_{j,i} \quad (22)$$

$$H(Y) = \int_0^t \beta_i (1 - \log_2 \beta_i)\, dt \quad (23)$$

On the other hand, in a fixed input event sequence, $\beta_{j,i}$ was applied to calculate the entropy of output given a fixed input signal $H\left(Y \middle| X_j\right)_j$ from ***equation (24)***, and the average of $H\left(Y \middle| X_j\right)_j$ of all input simulation trials in ***equation (25)*** was defined

as the entropy of output given input $H(Y|X)$.

$$H\left(Y\middle|X_j\right)_j = \int_0^t \beta_{j,i}\,(1-\log_2\beta_{j,i})dt \tag{24}$$

$$H(Y|X) = \frac{1}{ninput}\sum_{j=1}^{ninput} H\left(Y\middle|X_j\right)_j \tag{25}$$

Applying the above ideas to the MATLAB simulation, the results showed that the entropy of output event sequence was approximately 90 bits, while the entropy of output given input event sequence was approximately 82 bits.

3.2.3 Mutual Information

As ***equation (3)*** shows, the mutual information is defined as the difference between the entropy of the output, and the entropy of the output given the input. So the mutual information in this glutamate simulation process was measured to be approximately 8 bits. How can we interpret this outcome from a biological perspective?

While the average mutual information is calculated from ***equation (3)***, the rate of accumulation of mutual information can be estimated using the same method, namely, the difference of change rate of entropy of output and change rate of entropy of output given input (***equation (26)***). And the mutual information in the time series could be estimated by the integral formula (***equation (27)***).

$$\frac{dI(Y;X)}{dt} = \frac{dH(Y)}{dt} - \frac{dH(Y|X)}{dt} \tag{26}$$

$$I(Y;X)(t) = \int_0^t \left(\frac{dH(Y)}{dt} - \frac{dH(Y|X)}{dt}\right) dt \tag{27}$$

Figure 8 describes the change rate of entropy of output (red curve) and change rate of entropy of output given input (black curve), estimated from simulation time series in top panel, and change rate of mutual information in time series in bottom panel. It showed that the spike occurrences both in entropy change rate and mutual information change rate were totally consistent. Besides, all curves were not smooth, but had fluctuation due to molecule number fluctuation in the signal transduction process at the single-molecule level.

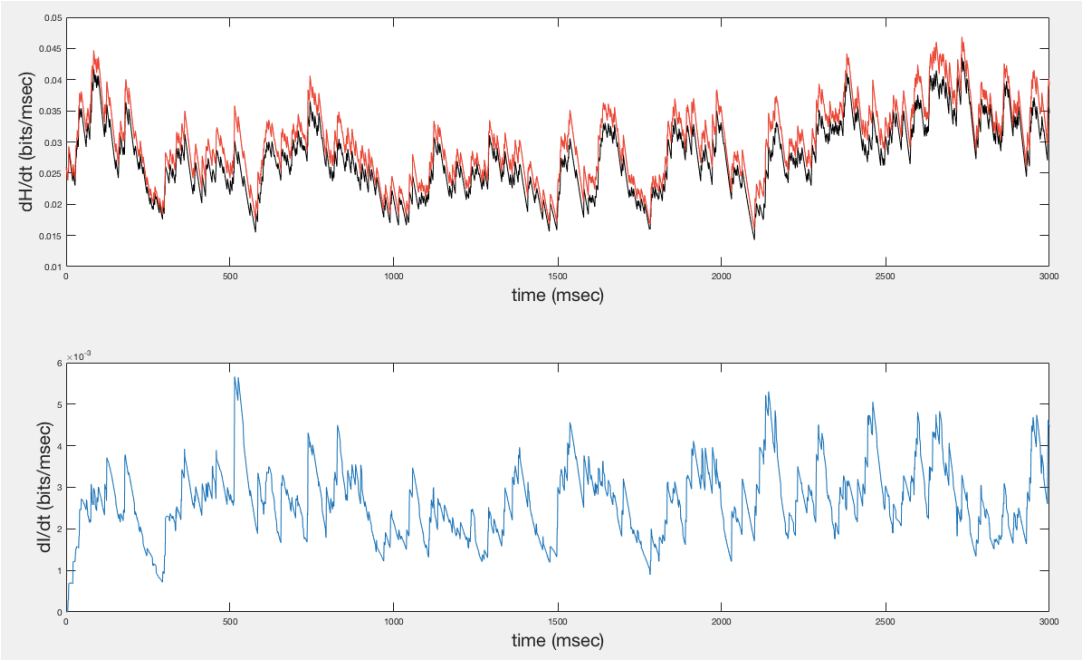

Figure 8 Plots of change rate of entropy and change rate of mutual information in time series. Top panel: change rate of entropy of output in red curve, and change rate of entropy of output given input in black curve. Bottom panel: change rate of mutual information in blue curve. All curves are in bits per msec.

Observing the bottom panel in ***Figure 8***, it showed that the change rate of mutual information varied significantly during the simulation time series. This was caused by the y-axis unit measurement in 0.001 bit/msec. However, if estimating accumulated mutual information from ***equation (27)***, the ***Figure 10*** showed that the curve was close to a smooth and straight line, which implied that the accumulation of mutual information was approximately constant along the time series.

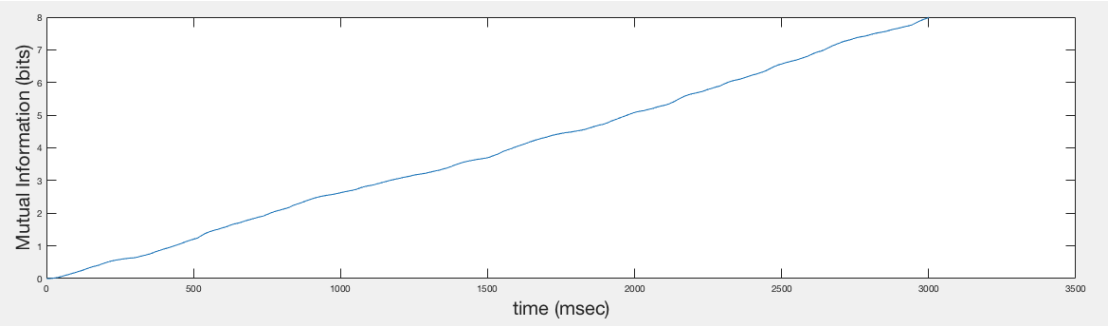

Figure 10 Accumulated mutual information in time series estimated from ***equation (27)***. The curve was in bits.

Since the total mutual information of glutamate simulation process was 8 bits, the average change rate of mutual information was calculated to be 0.0027 bits/msec (***equation (28)***). Thus, the time length for 1 bit information accumulation was around 370 msec, and the total simulation time could be divided into 8 time intervals when each interval was presented to have 1 bit information.

$$average\left(\frac{dI}{dt}\right) = \frac{I(Y;X)}{total\ time} = \frac{8\ bits}{3000\ msec} = 0.0027\ bits/msec \quad (28)$$

$$t_{1\ bit} = \frac{1\ bit}{averate_{\frac{dI}{dt}}} = \frac{1\ bit}{0.0027\ bits/msec} \approx 370\ msec \quad (29)$$

By analyzing input events and output events of all simulation trials in each time interval, ***Table 6*** suggested that there were, on average, 4 input events in sender and typically 1~2 output events in receiver, which presented 1 bit information in transmission as well. After total time accumulation, the information transmitted in the process was 8 bits while there were average 30 input events and 13 output events.

Table 6 The average input events and output events in each time interval.

Interval	1	2	3	4
	1~370msec	371~740msec	741~1110msec	1111~1480msec
#(Input)	4	4	3	3
#(Output)	2	1	1	2
Interval	5	6	7	8
	1481~1850msec	1851~2220msec	2221~2590msec	2591~3000msec
#(Input)	4	4	4	4
#(Output)	1	1	2	2

One usually thinks of 1 bit of information as estimating the outcome of a flip of an evenly weighted coin.[18] It could also be considered as any Yes/No question with a 50/50 outcome probability. For example, a question about the input signal that could possibly be answered by observing the output signal. Specifically, in the glutamate

simulation, 1 bit of information could be interpreted as follows: there was equivalent probability when 1~2 output events were consistent with 4 input events. So 8 bits information meant that 8 individual event sequences where each of it included 4 input events and 1~2 output events with its occurrence probability was 0.5. Here 4 release reaction was only an average of input events in each interval, but what about intervals in which there were fewer than 4 or more than 4 input events?

Since input information generated as input events is generated according to the Poisson distribution, the probability of event occurrence number is measured as ***equation (30)***, where k is input event occurrence number, and λ is glutamate release rate in source.

$$P(x = k) = \frac{\lambda^k e^{-\lambda}}{k!} \qquad (30)$$

If there were 3 release events working as input signal, there were 4 possible outcomes in output, which were 0 output event, 1 output event, 2 output events and 3 output events. When the possibly occurred outcome events were fewer than 3, it was estimated that the time interval when there were average 3 input events was around 367msec from the solution of ***equation (31)***, where μ equaled to λt, and λ was glutamate release rate. By dividing total time length into several intervals with each of them on average 367 msec in length, the average outcome event count was 1~2 while there were average 3 input events, which implied that it was evenly likely whether 1~2 output events are matched with 3 input events or not (***Table 7***). The same method can be used to estimate outcome uncertainty when there are more than 4 input events.

$$P(x \leq 3) = \mu^0 e^{-\mu} + \frac{\mu^1 e^{-\mu}}{1!} + \frac{\mu^2 e^{-\mu}}{2!} + \frac{\mu^3 e^{-\mu}}{3!} = \frac{1}{2} \qquad (31)$$

Table 7 The average input events and output events with time interval was 367msec.

Interval	**1**	**2**	**3**
#(Input)	4	3	3
#(Output)	2	1	1
Interval	4	5	6
#(Input)	2	1	1
#(Output)	3	4	3
Interval	7	8	9
#(Input)	4	4	1
#(Output)	2	2	0

From ***Table 6*** and ***Table 7***, it suggested that when the total time length was divided into 8 intervals where each of it include 370 milliseconds, the output events were around 1~2 consistent with average 4 input events. However, when the total time length was separated by 367 milliseconds into 9 intervals, where the output had fewer than 2 events, closer to 1 or fewer event matching with average 3 input events. This situation could be interpreted as a Yes/No question that the answer to the input events by estimation output events, which meant if there were 1~2 output events, higher proportion in 2 events, it was more confident to say that there were average 4 input events in sender, while if there were fewer than 2 output events, higher possibility in 1 or 0 events, it more obviously indicated that there were average 3 input events to match with it. So this is a method to estimate input signal situation by observing output events. Although the mutual information and the mutual information rate were abstract quantities in glutamate simulation in this article, this example helped build a bridge to our biological intuition.

Discussion

The main purpose of this thesis was to apply information theoretic analysis to the glutamate/NMDA signaling pathway. Consequently, the choice of model parameters for the simulation was of critical importance. Although constants in the paper were taken from reference [32], different efforts at quantitative modeling of glutamate/NMDA signaling have been made, using a range of different parameters. Another direction for future work would be to study the sensitivity of the mutual information rate to changes in parameters such as the assumed volume of the synaptic cleft, the glutamate decay rate, the association and dissociation rates. Some of these changes could be anticipated, for example, decreasing the synaptic volume, while keeping the mean numbers and frequency of glutamate release constant, would lead to higher glutamate concentrations inside the cleft, pushing the receptor population towards saturation of the doubly bound state. If the doubly bound state accounts for more than half the receptor population, the double-binding events are no longer approximately independent of one another, and our estimate of the output entropy rates based on McFadden's formula for the Poisson process no longer applies. It is not clear a priori whether increasing or decreasing the volume would tend to increase or decrease the mutual information rate, pointing to an interesting avenue for future investigation. Also, with smaller volumes, reaction rates increase, and we may have to decrease the simulation time step significantly, or explore other simulation methods.

Meanwhile, in addition to synaptic cleft volume and other geometric parameters, the kinetic rates in the glutamate/NMDA receptor chemical reaction scheme also play

an important role. In a normal ligand/receptor kinetic scheme, such as that depicted in ***Figure 4***, a glutamate/NMDA receptor has two binding sites, leading to three functionally distinct receptor states: an unbound state, a singly bound state, and a doubly bound state. The kinetic rates for the binding reactions are identified as k_1 and k_2 in the unbound/singly bound forward reaction and the singly/doubly bound forward reaction, respectively, while the dissociation rates are represented as k_{-1} and k_{-2}. On one hand, if the reactions occurring at the two binding sites are independent of one another, which is to say the system exhibits no cooperativity, then the kinetic rate, k_1 should be twice k_2, and k_{-2} should be twice k_{-1}. On the other hand, if the glutamate/NMDA receptor exhibits cooperativity, this would imply that binding a glutamate molecule at first NMDA receptor binding site makes it easier for a second glutamate molecule to bind at the second NMDA receptor binding site. This type of cooperativity is familiar from the well hemoglobin/oxygen binding reactions: when an oxygen molecule binds to one of hemoglobin's four binding sites, the affinity of another oxygen to three remaining binding sites increases[26]. For the glutamate/NMDA system, cooperativity would mean, the kinetic rate k_2 should be greater than $\frac{k_1}{2}$, and k_{-1} should be greater than $\frac{k_{-2}}{2}$.

In this article, we used kinetic scheme provided by Lester and Jahr[26] (***Figure 4***). In this scheme, the binding/unbinding reactions at the two NMDA receptor binding sites are presumed to be independent and not cooperative, so the kinetic rate $k_1 = 2k_2$, $k_2 = k_{on}$, and $k_{-2} = 2k_{-1}$, $k_{-1} = k_{off}$ (***Table 2***)[25]. However, in a more realistic model, it could well be that the chemical reactions in two binding sites

should be cooperative, like that in hemoglobin/oxygen binding events. Studying the effect of cooperativity on the mutual information between the input signal and the receptor provides another direction for future work. In this connection, it is interesting to note that in Thomas and Eckford's article, "Shannon capacity of signal transduction for multiple independent receptors", the authors showed that introducing non-independence of binding (cooperativity) reduced the mutual information of a communications system comprising multiple binding sites.[34] Nevertheless, cooperativity could confer other biologically relevant benefits even at the cost of reduced information transmission.

In the simulation process described in this article, there were 200 individual simulation samples, including 20 input simulation trials and 10 simulation trials matched to each of them. The simulation results showed that in a single simulation trial of an input event sequence, all molecule number curves were smooth, without obvious fluctuation, with 1msec time step. However, if the time step was shortened to be 0.1 msec or 0.05 msec, the fluctuations could be more obvious.

My analysis of the mutual information suggests that there were approximately 8 bits information of this simulation process, which meant that there was 8 bits of information transmitted in the simulation, over a 3 second simulation interval. Based on the observation that the rate of accumulation of mutual information in the time series was close to constant, each bit of information could be interpreted as an estimation of whether average 1~2 output events were consistent with, on average, 4 input events. The above conclusion derived from the simulation in this article with

around 200 individual samples. Future studies can focus on generating and analyzing larger sample datasets, for instance using 50 input simulation trials when each of them has 20 output simulation trials, to confirm the curves reprsenting molecular event counts and molecule abundances over time and the rate of change of mutual information.

Generally, one bit information is the minimum required to distinguish a binary signal with each outcome having equal probability. Thus, no matter how much the total information is in the simulation, it can be divided into several individual 1 bit sections such that each of them can determine the choice of two equally likely outcomes. However, the main question of the topic is that what the relationship is between communication and signaling in biological systems and mutual information? To answer this question, based on the 1-bit information interpretation, I could apply it other signal transduction system, such as the immune system. If I could simulate a particular drug/vaccine interaction in immune system where drug concentration is the input and gene or protein expression level in a response cell is the output, I could measure mutual information transduced through this channel. In this situation, the amount of mutual information can represent whether the input signal, drug concentration is high or low. In this way, the combination of information theoretic analysis and other study fields has the potential for broad application.

Appendix

All coding files and datasets are uploaded into Google Drive. Here is an overview of the several folders.

(https://drive.google.com/open?id=1wlaLfEhG-U6ZOHfmbl_ponmN2b10Vy_S)

1. File folder (parameters_setting_up) has one Matlab coding file named (Parameters.m) and one dataset file (initial_condition.mat), including all basic glutamate simulation geometry parameters and their values. And by using these parameters, to set up all four type molecules' initial conditions in individual simulation trial.
2. File folder (input_simulation_samples) and (simulation_results) include individual simulation trial's coding file and dataset. Since there are 20 input simulation trials, there are 20 separate coding files and datasets for trial #1 to trial #20. In each coding file, it runs one fixed input event sequence for 10 trials to get 10 output event sequences, which results in 200 samples in total.
3. File folder (results_analysis) has one coding file (mutual_information.m) for entropy calculation, mutual information calculation, and mutual information interpretation from datasets in file folder (simulation_results).

References:

[1] Berg JM, Tymoczko JL, Stryer L. (2002) Biochemistry. 5th edition. *New York: W H Freeman;* Chapter 15, Signal-Transduction Pathways: An Introduction to Information Metabolism.

[2] Bradshaw, Ralph A. Dennis, Edward A. (2010). Handbook of Cell Signaling (2nd ed.). *Amsterdam, Netherlands: Academic Press.*

[3] Gilbert SF. (2000) Developmental Biology. 6th edition. *Sunderland (MA): Sinauer Associates; Cell Surface Receptors and Their Signal Transduction Pathways.*

[4] Reece, Jane; Campbell, Neil (2002). Biology. *San Francisco: Benjamin Cummings.*

[5] Ingber, D. E. (2006). Cellular mechanotransduction: putting all the pieces together again. T*he FASEB Journal.* 20 (7): 811–827.

[6] Xing Du, Yi Li, Yuan-Ling Xia, Shi-Meng Ai, Jing Liang, Peng Sang, Xing-Lai Ji, Shu-Qun Liu. (2016). Insights into Protein-Ligand Interactions: Mechanisms, Models, and Methods. *International Journal of Molecular Sciences.* (17):144.

[7] Lester A. Rubenstein, Randy J. Zauhar, Richard G. Lanzara. (2006). Molecular Dynamics of a Biolophysical Model for b2-adrenergic and G protein-coupled receptor activation. *Journal of Molecular Graphics and Modeling* (25) 396-409

[8] Purves D, Augustine GJ, Fitzpatrick D, et al., editors. (2001). Neuroscience. 2nd edition. *Sunderland (MA): Sinauer Associates Second Messengers.*

[9] Robert Stufflebeam. (2008). Neurons, Synapses, Action Potentials, and Neurotransmission.

[10] Kolb, Bryan; Whishaw, Ian Q. (2003). Fundamentals of Human Neuropsychology (5th ed.). Worth. pp. 102–104.

[11] Michael R. Markham. (2008). Synaptic Transmission: Figures and Animation.

[12] credit by Thomas Splettstoesser on www.scistyle.com

[13] Robert Sapolsky (2005). Biology and Human Behavior: The Neurological Origins of Individuality, 2nd edition.

[14] Dr. John Andrews. International Symposium on Nitric Oxide. MaRS.

[15] Purves D, Augustine GJ, Fitzpatrick D, et al., editors. (2001) Neuroscience. 2nd edition. *Sunderland (MA): Sinauer Associates; Two Families of Postsynaptic Receptors.*

[16] HUGO Gene Nomenclature Committee. "Gene Family: Ligand gated ion channels".

[17] Claude E. Shannon. (1948) A Mathematical Theory of Communication. *Bell Systems Technical*

Journal.

[18] Alex Rhee, Raymond Cheong, Andre Levchenke. (2012) The application of information theory to biochemical signaling systems. *Phys. Biol.* Vol. 9(4). doi: 10.1088/1478-3975/9/4/045011.

[19] H., Quastler, Ed. (1953) Information theory in Biology. University of Illinois Press, Urbana

[20] Johnson, Horton A. (1970) Information theory in biology after 18 years. *Science.* 168.3939: 1545-1550.

[21] Borst A, Theunissen FE. (1999) Information theory and neural coding. *Nat Neurosci.* 2(11): 947-957.

[22] Tkacik G, Callan CG, Bialek W. (2008) Information capacity of genetic regulatory elements. *Physical Review E.* 80(3):031920.

[23] Raymond Cheong, Alex Rhee, et al. (2011) Information transduction capacity of noisy biochemical signaling networks. *Science.* 334(6054): 354-358.

[24] F. Rieke; D. Warland; R Ruyter van Steveninck; W Bialek (1997). Spikes: Exploring the Neural Code. *The MIT press.* ISBN 978-0262681087.

[25] Clements, J. D. and G. L. Westbrook. (1991) Activation kinetics reveal the number of glutamate and glycine binding sites on the N-methl-D-aspartate receptor. *Neuron*. 7: 605-613.

[26] Lester, R. A. J., and C. E. Jahr. (1992) NMDA channel behavior depends on agonist affinity. *J. Jeurosci.* 12: 635-643.

[27] Haas, K. E., and Macdonald, R. L. (1999) GABAA receptor subunit gamma2 and delta subtypes confer unique kinetic properties on recombinant GABAA receptor currents in mouse fibroblasts. *J. Physiol.* 514:27-45.

[28] David J. Dripps, Barbara J. Brandhuber, Robert C. Thompson, Stephen P. Eisenberg. (1991) Interleukin-1 (IL-1) receptor antagonist binds to the 80-kDa IL-1 receptor but does not initiate IL-1 signal transduction. *The Journal of Biological Chemistry.* 266(16): 10331-10336

[29] Waldemar Gonsiorek, Xuedong Fan, David Hesk, James Fossetta. (2006) Pharmacological Characterization of Sch527123, a potent allosteric CXCR1/CXCR2 antagonist. *The Journal of Pharmacology and Experimental Therapeutics.* 322(2): 477-488

[30] Meldrum BS (2000). "Glutamate as a neurotransmitter in the brain: review of physiology and pathology". *The Journal of Nutrition*.

[31] McEntee WJ, Crook TH (1993). Glutamate: its role in learning, memory, and the aging brain. *Psychopharmacology.* 111(4): 391–401.

[32] William R. Holmes. (1995) Modeling the Effect of Glutamate Diffusion and Uptake on NMDA

and Non-NMDA Receptor Saturation. *Biophysical Journal.* 69:1734-1747.

[33] J. A. McFadden. (1965) The Entropy of a Point Porcess. *J. Soc. Indust. Appl. Math.* 13(3):988-994.

[34] Peter J. Thomas, Andrew W. Eckford. (2016) Shannon capacity of signal transduction for multiple independent receptors. *2016 IEEE International Symposium on Information Theory.* 1804-1808.