On the Foundation of Neural Computation

Peng Yang

Abstract

This paper demonstrates that a large network of neurons mindlessly following Feynman's "sum of all paths" rule can perform spectrum analysis, synthesize arbitrary signals, and approximate complex functions. It predicts that neural circuits are the basic storage and processing units, and the signals are transformed and stored in the frequency domain representation.

It shows the detailed neural mechanism for read-write memory, pattern recognition, reinforcement learning, and the development of abstract concept and symbolic language. It deduces the unique behavior of the human mind, e.g. spontaneous pattern recognition without backpropagation, no separate training and inference phases, one-shot learning, the learned behavior is reversible and mirrors the probability distribution of inputs, etc. It makes quantified predictions on the behavior of collective minds, e.g. Zipfian distribution of word frequency and market capitalization of companies, fractal nature of market returns over a specific time window, etc.

It proposes a proof-of-concept implementation of neural computers, e.g. physical devices performing parallel communication and Fourier transform in constant time. It provides a simple mechanism of how billions of mindless neurons fine-tune and develop a complex structure like the human brain. It discusses the driving force behind evolution and how it is related to the Principle of Least Action.

Finally, it explores the nature of consciousness, including the supposition of minds, free will, self-awareness, and the limitations of the human mind. It offers an intuitive interpretation of the basic concepts of quantum mechanics through the lens of human behavior.

Introduction

The inner workings of the brain is an irresistible intellectual challenge. Recent progress in machine learning demonstrates that deep neural networks are capable of approximating complicated functions. However, the current approach has two distinct phases: training and inference. It requires humans, "god neurons", in the loop to synchronize global coordination, i.e. building the networks, defining a cost function, and performing the trick of backpropagation. It's hard to believe that's how a biological brain actually works.

email: yangpeng1973@gmail.com

A scientific brain theory shall have a clear physical model and mathematical basis. It should provide satisfactory answers to the following questions: How are sensory inputs processed and analyzed? How are the outputs to motor neurons synthesized? What's the mechanism for read-write memory? How is the information encoded and stored in memory? How do different parts of a brain communicate? How can a brain recognize both short and long-term patterns simultaneously? How does a brain learn from feedback? How does a brain develop abstract concepts and symbolic language? Ultimately, what's the nature of consciousness?

Another important question to be addressed is the development of biological brains. The human genome contains less than 1 gigabyte information. The ion channels, ATPases, microtubules, and organelles inside neurons are not very different from those inside body cells. Most of them are built from the same set of genes that are common and essential to the survival and reproduction of all cells. A human brain consists of ~100 billion mindless neurons[4]. From a pure information point of view, the genome cannot have many bits dedicated to these neurons. How can these neurons fine-tune a network as complex as a human brain to function properly without detailed instructions from the genome? How can nature evolve such a complex structure without supernatural divine intervention?

To avoid making unfounded assumptions, we shall not require neurons to do no more than what innate objects can. The theory to be developed here is based on one simple hypothesis: The propagation of signals over a neural network follows the same rule as the propagation of particle waves over physical space.

(1.0) $$\psi(x_{k+1}, t+\varepsilon) = \int exp[i \cdot (2\pi/h) \cdot S(x_{k+1}, x_k)]\ \psi(x_k, t)\ dx_k/A$$

The rule is based on Feynman's *A Space-Time Approach of Non-Relativistic Quantum Mechanics*[1] (Eq 18). It states that the probability amplitude of observing a particle at point x_{k+1} and time $t+\varepsilon$, where ε is an infinitesimal time interval, is the sum of amplitudes contributed by all paths $x_k \rightarrow x_{k+1}$ at an earlier time t. The paths contribute equally in magnitude, but the phase of their contribution is the classical action in the units of $h/2\pi$ (h is Planck constant). The action $S(x_{k+1}, x_k)$ is a physical quantity defined by the time integral of the Lagrangian taken along the path and has the dimensions of $[energy] * [time]$. A is a normalization factor of choice to ensure the conservation of probability.

Feynman's "sum of all paths" approach is mathematically equivalent to the more usual formulation of quantum mechanics. We postulate that the identical rule is applicable to the propagation of signals over a network of neurons. Namely, the probability amplitude of observing a firing event at neuron x_{k+1} at time $t+\varepsilon$ is the sum of amplitudes contributed by all upstream neurons x_k at a previous time t. The upstream neurons contribute to the amplitude equally in magnitude but the phase of their contribution is determined by the action $[energy] * [time]$ for the signals to travel along the corresponding paths.

Some will, and rightly so, question the postulate that the propagation of neural signals follows Feynman's rule. The challenge is, a brain consists of neurons and neurons are made of mindless molecules. We will have to propound something to bridge the gap between innate objects and the consciousness we all experience. If neurons don't follow the first principle of physics in its simplest form, what laws do they follow? Life on Earth occupies a tiny niche of space-time in a vast span of the universe governed by such a simple rule. It would be too provincial for us, the observer, to make up new rules custom-made for us, the observed.

The resonance of natural frequency is a universal phenomenon observed on all physical objects. For example, when an opera singer sings in an opera house, every molecule in the air, every hair cell inside ears, and every string on the musical instruments can and will tune into some specific frequencies. When an object vibrates, it emits energy. When it resonates, it retains the energy from the vibration. The resonance of natural frequency is a superb communication protocol because objects can simultaneously talk to each other over the same network. From Feynman's rule, we will derive that it's how different parts of a brain communicate.

A burst of neuron firings propagating over a neural network is mathematically equivalent to a particle, a drop of energy, tracing a path through the configuration space. Since the propagation of neural signals follows the same Feynman's rule as the propagation of particles, the evolution of a neural network shall follow the same principle of least action. It is a peculiar behavior of nature that the motion of objects in a physical system always obeys the principle of least action[2]. There is no evidence suggesting that the motion of living objects is exempted from the principle.

These are the 3 main concepts we will use to deduce the behavior of neural computation. In the jargon of machine learning, a neural network (brain) is a neural computer: Feynman's rule is the primitive of its machine code; The resonance of natural frequency is the communication protocol; The principle of least action defines the cost function that nature always insists on optimizing.

The Formulation of Neural Network

A biological neuron has 3 main parts: a soma (cell body), an axon, and many dendrites. A neuron receives signals from its upstream neurons via synapses on its dendrites. It's estimated that average neurons have ~10,000 synapses and some like Purkinje cells have up to 200,000. Synapses can be either excitatory or inhibitory. Some signals come in as a burst of neuron firings while others are sporadic firings. When a neuron gets activated, the neural firing is an **all-or-none** event. A single neuron firing travelling down a specific axon doesn't vary in magnitude and shape. The same signal reaches all presynaptic axon terminals equally and triggers the release of neurotransmitters on these terminals. The neurotransmitters are then bound onto presynaptic receptors on dendrites, passing the signal to downstream neurons.

In our discussion, a neural network will be represented by a directed graph where a node represents a neuron and an edge represents a direct connection between two neurons. The direction of the edge corresponds to the direction of signal propagation.

The State of Neurons

For an intuitive understanding, a neural network is analogous to a network of tributaries and bayous with lots of anabranches and braided channels. The firing of sensory neurons corresponds to the raining in some upstream regions. The total number of neuron firings corresponds to the total volume of water dropped in the regions. The intensity of the rain may vary over the time. There may be simultaneous rainstorms in different regions. Neurons firings propagating over a neural network are similar to water traveling down a complicated river system. The water coming from some upstream regions may cause some floodings (intensive neuron firings) in certain downstream regions. To understand the behavior of a neural network is to understand how different rains cause flooding in different regions of a complicated river system.

To model the behavior of a neural network, we represent the state of a neuron x at time t by a complex number called probability amplitude (1.1). The absolute square of the amplitude corresponds to the probability of observing a neuron firing (1.2).

(1.1) $\psi(x, t) : \ a\ complex\ number$

(1.2) $P(x, t) = |\psi(x, t)|^2 \quad P \in [0, 1]$

The amplitude cannot be observed directly but the probability density (intensity) of neuron firings can be measured objectively. For a given time period of $[t_0, t_1]$, the probability density of neuron firings in a specific region, namely, the strength of a neural signal, is proportional to the energy of neural firings in the region.

(1.3) $$P(x) = \int_{t_0}^{t_1} |\psi(x, t)|^2 dt$$

The amplitude and equations are necessary to build a mathematical basis for our derivation but they're not required to understand the derived behavior of a neural network. Whenever you find some math or physics concepts are difficult to comprehend, it helps to use the analogy of rain and flooding in a complicated river system. A single neuron firing is equivalent to a raindrop. For those who tend to see the particle side of neural signals, i.e. a neuron firing is a discrete event, remind yourself that a drop of water made of discrete water molecules and yet you may feel perfectly natural to formulate the propagation of water molecules as ripples and waves. If you still cannot wrap your head around some equations or physics concepts, just put them aside for a moment and continue the reading. Everything will eventually become as natural and intuitive as what you experience in your everyday life.

The polar form of a complex number has a magnitude and a phase. The probability amplitude is just a mathematical trick to capture two important pieces of information about neuron firings with

one number. In our rain analogy, the square of the magnitude corresponds to the area of a cross section of a river at a specific point. The integral (1.3) is the equivalent of the water in a specific region of the river for the time period of $[t_0, t_1]$. The phase of the amplitude carries the critical timing information of the raindrops. If there is a big rainstorm in one region or several small but simultaneous rains in different regions, it may cause a major flood (intensive neuron firings) in a downstream branch. However, if it's not a rainstorm but sporadic dribblings or non-overlapping rains in different regions, it won't be a major flooding even if the total volume of the water passing the downstream branch is the same. The phase is required to properly compute how the raindrops from different regions affect the water level of downstream regions.

There is no way to know the origin of a drop of water in downstream rivers. There is also no way to pinpoint the exact cause of a downstream neuron firing. Therefore, the state of a neuron is a combination (superposition) of all possible base states (1.4). $\psi_n(\pm, t)$ represents the contribution of neural signal n. In our rain analogy, $\psi_n(\pm, t)$ corresponds to the probability amplitude of the precipitation at a specific region n. The coefficient c_n corresponds to the contribution of the amount of water from region n.

(1.4) $$\psi(x, t) = \sum_{n=0}^{\infty} c_n \psi_n(\pm, t)$$

By definition, if we observe a single neuron firing at a specific location n, the absolute square of $\psi_n(\pm, t)$ should be 1 at the specific time t. In our rain and river analogy, it's the equivalent of observing a drop of water (molecule) at a specific location n.

(1.5) $$P(\pm, t) = \left|\psi_n(\pm, t)\right|^2 = 1$$

Since it takes time and energy for a neural signal to propagate, ψ should be a function of both time and energy. We choose to represent the unit of probability amplitude by a pair of arrows on the unit circle rotating at the opposite directions. Such a representation allows us to evolve the phase for each neuron firing (raindrop) when it propagates over a complicated network .

(1.6) $$\psi(\pm, t) = e^{\pm i(2\pi/h)Et} \qquad h : Planck\ constant,\ E :\ energy,\ t :\ time$$

Let E_0 be a finite amount of energy, ε the unit of time, t a point at the time. A pair of base units are usually represented by an angular frequency n in the units of $\omega_0 = (2\pi/h)E_0\varepsilon$. If the total energy under $\psi(x, t)$ is normalized to 1, the energy represented by c_n gives the contribution of the specific signal n.

(1.7) $$\psi(x, t) = \sum_{n=0}^{\infty} c_n e^{\pm in\omega_0 t} \qquad t, n \in N,\ c_n \in R,\ \omega_0 = (2\pi/h)E_0\varepsilon$$

Critics may point out that the state representation is fuzzy about when, why, and how a neuron gets excited. The observation is fair and accurate. The decision using probability is intentional because it is impossible to predict when a neuron is going to fire with absolute certainty. The questions of why and how are irrelevant because downstream neurons cannot tell why and how their upstream neurons get excited either. By choosing such a representation, we refuse to answer unanswerable or irrelevant questions.

However, our representation is not a philosophical construct of voodoo science. A neuron firing is an observable event. The intensity of neuron firings at a specific location for a given time period can be measured objectively. The probability interpretation of the amplitude grounds the state of neurons to physical reality and makes whatever theory predicting the behavior of a neural network scientifically accountable.

In the following discussion, we may use "signal", "excited", "activated", or "neuron firing" synonymously. The propagation of a signal over a neural network means the propagation of a state (amplitude) from one neuron to another. The intensity of neuron firing is synonymous to the energy of neuron firing or the strength of a signal.

The Rule of Signal Propagation

In the state representation, we focus on the **all-or-none** events observable between soma and presynaptic axon terminals. However, lots of critical signal processing happens between the binding of neurotransmitters on postsynaptic receptors and the opening of ion channels on soma (activation). We have to address how the signals are integrated and processed because it's the essence of a neural network.

Suppose that the time is measured in the units of ε. Let $\psi(x_a,\ t)$ be the state of upstream neurons $x_a \in \{x_a : x_a \to x_b\}$ at t, and $\psi(x_b,\ t+1)$ the state of downstream neuron x_b at $t+1$. We postulate that propagation of neural signals follow Feynman's rule (1.0), which can be expressed in the following discrete form:

(1.8) $$\psi(x_b, t+1) = \sum_{x_a \in A} e^{i(2\pi/h)\, S(a\to b)}\, \psi(x_a,\ t) = \sum_{x_a \in A} e^{i\omega_0 n_{a\to b}}\, \psi(x_a,\ t) \qquad A = \{x_a : x_a \to x_b\}$$

where $S(a \to b)$ is the action $[energy] * [time]$ required for a signal to propagate from x_a to x_b. Let us set aside how to accurately measure $S(a \to b)$ for a moment. In principle, we should all agree that: 1) It takes $[energy] * [time]$ $S(a \to b)$ for a signal to propagate from x_a to x_b and $S(a \to b)$ is a physical quantity that exists in reality. 2) $S(a \to b)$ is uniquely determined by the very nature of connection $a \to b$.

The action (phase shift) of path $S(x_0,\ ...,\ x_k)$ is the sum of the action of all edges along the path (1.9). A path is often denoted by tuple $(\alpha,\ \tau)$ where α is the absolute phase shift and τ the absolute time delay of the path. The phase contribution of a path is periodical. Because the Planck Constant is such a miniscule quantity. It's more convenient to express the phase shift as angular frequency in the units of $\omega_0 = (2\pi/h)E_0\varepsilon$ where E_0 is the average energy of signal propagation per unit of time ε. So, we can also denote a path by $(n,\ t)$ where n is the angular frequency of the path and t the time required for signal propagation along the path.

(1.9) $$S(x_0,\ x_k) = \sum_{i=0}^{k-1} S(x_i,\ x_{i+1}) = \sum_{i=0}^{k-1} n_{i,\,i+1}\omega_0 = n_{0,\,k}\ \omega_0$$

The absolute refractory period of biological neurons is 1~5ms. It sets the temporal resolution (max sampling rate) of a biological neural network to 200~1,000 cycles per second. For a given temporal resolution of ε, all frequencies n_i are indistinguishable if $n_i \% N = n_0$ where $N = 1/\varepsilon$ and $n_0 \in [-N, N]$. We can choose $n = n_0$ to represent all n_i for $n_i \% N = n_0$.

The interpretation of Feynman's rule matches the reality of neural signal propagation. The neuron firing is an **all-or-none** event. All upstream signals reach downstream neurons equally in magnitude and differ only in phase (timing). If multiple excitatory signals arrive within a very short period of time, they interfere constructively and increase the probability of downstream neurons getting excited. If upstream signals arrive sporadically, downstream neurons may never get excited regardless of the total number of signals received. The formula can model inhibitory connections as well. The signals arriving at inhibitory connections are out of phase signals. They destructively interfere with excitatory signals.

The Fundamentals of Neural Computation

Imagine that we map the human brain into a directed graph and assign each neuron a unique number x. For a given unit of time ε and an average energy consumption of E_0 per unit of time, we label all edges with (n, t) and initialize the graph as a blank slate by setting all $\psi(*, t_0) = 0$.

Let us suppose that we record the firing of all sensory neuron x_s for time interval $[t_{i-1}, t_i)$ as $\psi(x_s, t_i)$. We can simulate the brain by advancing the clock one tick at a time. For each iteration, we first copy the state of sensory neurons from the recording, and then update the state of other neurons by following the rule of signal propagation (1.0, 1.8). It's technically infeasible to map the whole human brain yet, but we can deduce some of its behavior by thought experiments.

Read-write Memory

If we traverse all paths originated from a specific neuron, we should find some paths looping back and forming neural circuits. The number of neural circuits is astronomical for a neural network with ~100 billion neurons and average ~10,000 connections per neuron.

Let $C_x^{n,t}$ denote a neural circuit made of path (n, t) and anchored at neuron x. To simplify our analysis, let's also assume that the path of $C_x^{n,t}$ made of n neurons evenly connected by unit edge $(1, 1)$.

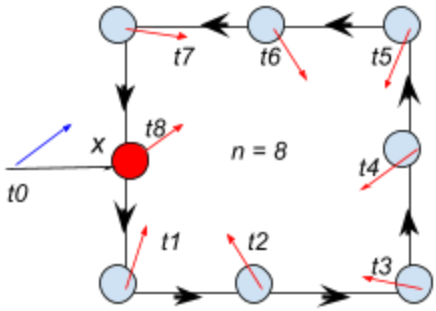

Fig 2.1 Neural Circuit Cx

According to the rule of neural signal propagation (1.8), when an impulse signal $\psi(x\,,\,t_0)$ arriving at neuron x and time t_0, shown as blue arrow in Fig 2.1, $\psi(x\,,\,t_0)$ will propagate around $C_x^{n,t}$ and shift its phase by ω_0 per iteration, shown as red arrows in Fig 2.1. If $C_{x_1}^{n,t}$ receive no more signals, $\psi(x\,,\,t_0)$ repeats itself at x with a phase shift of $n * \omega_0$ for every n iterations. All neurons on $C_x^{n,t}$ will see the same signal $\psi(x\,,\,t_0)$ and differ only in phase. The level of neuron firing around $C_x^{n,t}$ is proportional to the signal strength of $\psi(x,\,t_0)$. The sustained neuron firing around $C_x^{n,t}$ is a persistent state of $C_x^{n,t}$, which can be modified by new signals arriving at $C_x^{n,t}$. So, the neural circuits of a neural network effectively function as its read-write storage units.

Unlike the persistent storage in digital computers, the state of neural circuits must be maintained by sustained neuron firing. If the power (blood) supply within a region is cut off even for a brief moment, a brain will suffer permanent memory loss. After the power supply is restored, the circuits may be resurrected and used to store new information but the old memories are gone. The behavior can be observed in patients who have suffered but recovered from temporary damage in certain regions of their brain. If some old muscle memories, as basic as walking, get lost, the patients have to spend a long time on relearning how to walk. If the memories were persisted as bits in certain molecular and/or biological structures, the memories should be easily restored after the power supply and functionality is restored, similar to the rebooting of digital computers. This is one key difference between neural circuit-based memory and materialized storage mechanisms.

If the information were encoded as bits and stored in neurons, as suggested by the Hopfield network model, the storage capacity of ~100 billion neurons would be in the order of $O(n/2log_2(n) \simeq 2GB$ [10], barely enough for a full length HD movie. With an average branching factor of 10,000, the number of unique paths starting from a neuron is in the order of 10^{400} after 100 connections. Even if a small fraction of the paths loop back and form neural circuits, the storage capacity of a densely connected neural network like the human brain could easily surpass the total capacity of all digital media combined in the word.

Natural Frequency

Neural circuits are hardwired to their upstream neurons. They receive continuous signals not a single impulse. It's important to understand how neural circuits process continuous signals, what information is stored in different circuits, and how the information is encoded.

Let $\phi(x,\ t)$ denote the state of $C_x^{n,\Delta}$ anchored at neuron x and time t, and $\psi(x,\ t)$ be the upstream signals arriving at neuron x and time t. To simplify our analysis, we measure time t and Δ in the units of ε and treat $\psi(x,\ t)$ as a matrix of discrete series. The behavior of continuous signal processing can be derived by letting ε approach 0.

$$\begin{array}{llll}
\psi(x,\ 1+0\Delta)\ ,\ \psi(x,\ 2+0\Delta) & \ldots\ \psi(x,\ c+0\Delta) & \ldots\ \psi(x,\ \Delta+0\Delta) \\
\psi(x,\ 1+1\Delta)\ ,\ \psi(x,\ 2+1\Delta) & \ldots\ \psi(x,\ c+1\Delta) & \ldots\ \psi(x,\ \Delta+1\Delta) \\
\psi(x,\ 1+2\Delta)\ ,\ \psi(x,\ 2+2\Delta) & \ldots\ \psi(x,\ c+2\Delta) & \ldots\ \psi(x,\ \Delta+2\Delta) \\
\ldots \quad \ldots & \ldots & \ldots \\
\psi(x,\ 1+r\Delta)\ ,\ \psi(x,\ 2+r\Delta) & \ldots\ \psi(x,\ c+r\Delta) & \ldots\ \psi(x,\ \Delta+r\Delta) \\
\ldots \quad \ldots & \ldots & \ldots
\end{array}$$

Each column of the matrix is a slice of uniformly spaced samples of the continuous signal $\psi(x,\ t)$. The first N samples represent a period of $T = N\Delta$. The same slice (column) of samples always propagate back to neuron x with a phase shift of $n\omega_0(N-k)$ where $n\omega_0$ is the phase shift per cycle. They are then integrated with the next sample (row) of the same slice (column).

2.1 $$\phi\ (x,\ c+N\Delta) = \sum_{k=0}^{N-1} e^{in\omega_0(N-k)}\psi(x,\ c+k\Delta)$$

Reset the starting time of each slice $c = 0$:

2.2 $$\phi\ (x,\ N\Delta) = \sum_{k=0}^{N-1} e^{in\omega_0(N-k)}\psi(x,\ k\Delta)$$

Move the common term out of the sum:

2.3 $$\phi\ (x,\ N\Delta) = e^{in\omega_0 N}\sum_{k=0}^{N-1} e^{-in\omega_0 k}\psi(x,\ k\Delta)$$

Let us redefine Δ as the new unit of the time:

2.4 $$\phi\ (x,\ N) = e^{i(n\omega_0 T/\Delta)N}\sum_{k=0}^{N-1} e^{-i(n\omega_0 T/\Delta)k}\psi(x,\ k)$$

The sum is the Fourier transform of the input signal $\psi(x,\ t)$ at angular frequency $n_c = (n\omega_0 T/\Delta)$:

2.5 $$\phi\ (x,\ N) = e^{i(n\omega_0 T/\Delta)N}\ \hat{\psi}(x,\ n\omega_0 T/\Delta)\ =\ e^{in_c\omega_0 N}\ \hat{\psi}(x,\ n_c\omega_0)$$

Substituting N with the familiar variable t, we arrive at $\phi(x,\ t)$ as a function of time t, angular frequency $\omega_c = n_c\omega_0$, and input signal $\psi(x,\ t)$:

2.6 $$\phi\ (x,\ t) = e^{in_c\omega_0 t}\ \hat{\psi}(x,\ n_c\omega_0)\ = e^{i\omega_c t}\ \hat{\psi}(x,\ \omega_c)$$

$\phi(x,\ t)$ is the state of a neural circuit, namely, the signal circling around the circuit. Its magnitude $\hat{\psi}(x,\ \omega_c)$ corresponds to the signal strength of $\psi(x,\ t)$ at angular frequency ω_c. The frequency of a neural circuit is proportional to $E_c = nE_0/\Delta$, the energy per unit of time ε required for maintaining the state of the circuit (2.7). Since ω_c and E_c are unique properties of a neural circuit, we call ω_c and E_c the circuit's natural frequency and energy level, respectively. In the following discussion, we may denote a neural circuit as $C_x^{n_c}$ or simply C_x^n if there is no ambiguity.

2.7 $$n_c\omega_0 = (nT/\Delta) * (2\pi/h)E_0\varepsilon = (2\pi/h)\ (nE_0/\Delta)(T/\varepsilon)\ = (2\pi/h)\ E_c(T/\varepsilon)$$

In a neural network, neurons just sum up all signals from their immediate upstream neighbors and pass the result to immediate downstream neighbours. Neurons do not maintain a persistent state. Neural circuits perform continuous integration and maintain a running sum. The natural frequency of a circuit determines what information is stored in the circuit. The strength of the signals stored in neural circuits is the energy retained from the resonating frequency component of the signals passing through the circuit.

At the most fundamental level, nature always represents physical beings as frequencies: a perfect pitch of middle C is 261.63 Hz and a photon of red ~430 THz. When a red photon arrives at a cone cell in our eyes, to sense it means to figure out its frequency and to absorb its vibration (energy). The word of "red" or "红色", the sound of "red" or "红色", the color of blood or sun, the sensation of red, and the "redness" (quale) of red all have to be processed by some neural circuits in our brain. Hence, each red-related concept can be mapped to a number, the natural frequency of the circuits excited by the concept. To say our brain understands something is to say it finds out some relationship among these numbers.

Neural circuits are analogous to strings with both ends fixed on a musical instrument. Their length and tension are determined by the action and time for a signal circling around the circuit. Neural circuits (strings) resonate well with the vibrations (signals) matching their natural frequencies. They keep vibrating at their natural frequencies even after the external forces (signals) are gone, which is the essence of the memory of neural networks. The dendrites and axons attached to neural circuits are the antennas receiving and transmitting the vibrations (signals).

DTFT and Signal Synthesis

A brain has to deal with situations more complicated than integrating all incoming signals and remembering a running sum. In a life and death struggle, it's vital for both predator and prey to focus on the signals at the moment and ignoring everything else. A brain must also be able to filter out short-term signals and have a chance to learn lessons by recognizing long-term patterns. Otherwise, a busy life becomes a wasted life. How could the same brain accomplish such conflicting tasks simultaneously?

Instead of wiring input signal $\psi(x,t)$ directly to C_x^*, a brain may first relay $\psi(x,\ t)$ on a path of equally spaced neurons and have them forward $\psi(x,\ t)$ to C_x^* with a certain time delay and phase

shift, as shown in Fig 2.2. Such circuits are called DTFT circuits because they perform Discrete-Time Fourier Transform (DTFT) over N samples of $\psi(x,t)$ for a period of $T = N\tau$ (2.8). A brain can have many DTFT circuits running parallel spectrum analysis. DTFT circuits allow a brain to process incoming signals at different temporal resolutions and simultaneously recognize patterns over different time windows.

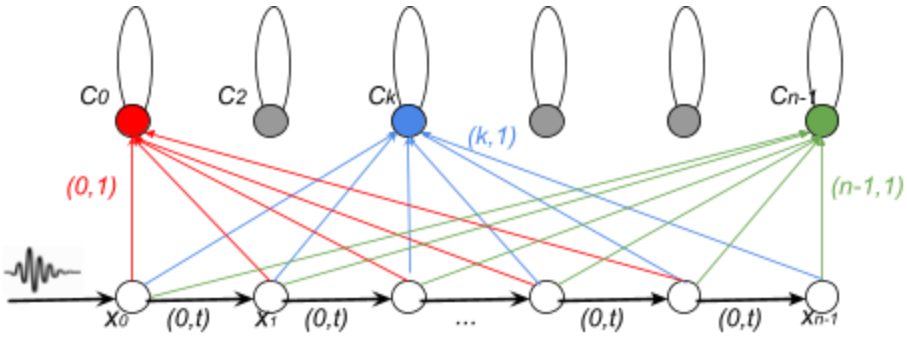

Fig 2.2 DTFT Circuits for Spectrum Analysis

2.8 $$\psi(x_k,\ t+T) = \sum_{n=0}^{N-1} e^{ik\omega_\tau n\tau}\psi(x,\ t+n\tau)\ = \hat{\psi}(x,\ k\omega_\tau)\ e^{i(\theta_k+n_k\omega_0 t)}$$

The complexity of neural computation can be measured in the amount of action consumed. The computation complexity of DTFT neural circuits is $O(N)$ in the units of action. Since the action consumed equals to $[energy] * [time]$, the performance of neural circuits is a tradeoff between energy and time. Given $O(N)$ of energy, all neural computation can be executed in a constant time $O(1)$. It's not a feat achievable by electrical circuits or digital computers.

Inverting the connections in DTFT circuits creates inverse DFT circuits, as shown in Fig 2.3. The inverse DFT circuits can retrieve the frequency domain representation of signals stored in neural circuits and transform them back to the time domain representation.

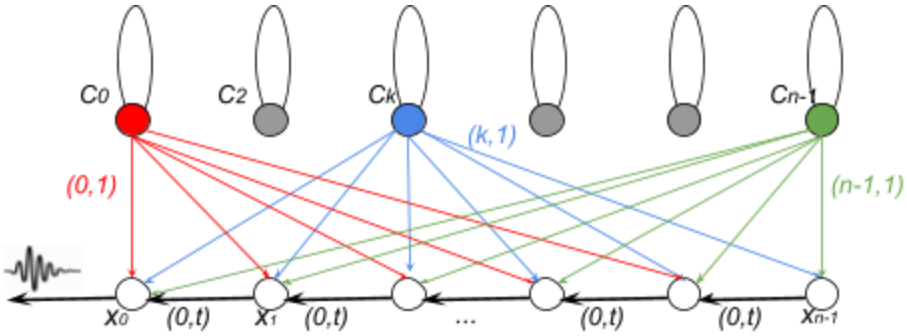

Fig 2.3 Inverse DFT Circuits for Signal Synthesis

Most sensory neurons are parts of specialized DTFT devices that transform external physical signals to their internal representation of neural signals. For example, hair cells inside ears or

cone cells on the retina respond to only certain frequencies of sound or light. Therefore, each sensory neuron represents a unique frequency of some external signals. The intensity of neuron firings at sensory neurons encodes the strength of external signals at a specific frequency and passes them to the central nervous system (CNS) for further processing.

Motor neurons connected to actuators are inverse DFT circuits transforming the frequency signals stored in neural circuits back to the time domain representation. For example, when people learn to speak a language, they build muscle memories on how to utter different words of the language. When they speak, an army of inverse DFT circuits retrieve the signals from muscle memories and synthesize the precise impulses for the muscle fibers on their vocal cords to pronounce the words. Speaking fluently requires a massive operation of information retrieval and signal synthesis.

Frequency Response

As discussed in the section of Natural Frequency, circuit $C_x^{n_c}$ maintain its state of $\phi(x,\ t)$ by circling the signal around.

2.9 $\phi\ (x,\ t) = e^{i\omega_c t}\ \hat{\psi}(x,\ \omega_c)$

The neurons on the circuit usually have their axons connected to downstream neurons. Therefore, $\phi\ (x,\ t)$ is also an input signal to downstream neurons for further processing. The frequency response of a circuit is the output behavior of the circuit in response to an external signal $\psi_\omega(x,\ t)$ as a function of frequency ω. The frequency response of $C_x^{n_c}$ can be characterized by the transfer function $H(\omega)$, which is defined as the ratio of output signal $\psi_{out}(\omega)$ to input signal $\psi_{in}(\omega)$.

2.9 $H(\omega) = \psi_{out}(\omega)/\psi_{in}(\omega)$

Let $\psi_{\omega,\,in}\,(x,\ t)$ denote a generic input signal whose frequency is ω and magnitude c_ω (2.10). According to the rule of signal propagation, the input signal and the state of circuit $C_x^{n_c}$ will be combined at neuron x and then propagate around the circuit. The output of $C_x^{n_c}$ will be the sum of $\psi_{\omega,\,in}\,(x,\ t)$ and $\phi\ (x,\ t)$ with a time shift of τ (2.11).

2.10 $\psi_{\omega,\,in}(x,\ t) = c_\omega e^{i\omega t} \quad c_\omega \in R$

2.11 $\psi_{\omega,\,out}(x,\ t+\tau) = \psi_{\omega,\,in}(x,\ t)\ + \phi(x,t)$

Let $c_0 = \hat{\psi}(x,\ \omega_c)$ denote the signal strength stored in $C_x^{n_c}$. We can derive the transfer function $H(\omega)$ by substituting $\psi_{out}(\omega)$ and $\psi_{in}(\omega)$ in (2.10) and (2.11) :

2.12 $H(\omega) = \frac{\hat{\psi}(x,\,\omega_c)\ e^{i\omega_c t} + c_\omega e^{i\omega t}}{c_\omega e^{i\omega t}}$

$$= \frac{c_0\, e^{i\omega_c t} + c_\omega e^{i\omega t}}{c_\omega e^{i\omega t}}$$
$$= (c_\omega + c_0/c_\omega)\, e^{i(\omega_c - \omega)t}$$

If an input frequency ω matches a circuit's natural frequency ω_c, namely, $\omega_c - \omega = 0$, $H(\omega)$ is a constant. The output signal $\psi_{out}(\omega)$ is the same as input $\psi_{in}(\omega)$ and its magnitude is amplified by a factor of $A = c_\omega + c_0/c_\omega$. A neural circuit will greatly amplify a tiny signal matching its natural frequency ($\omega_c = \omega$ and $c_\omega \ll c_0$). If ω differs from ω_c, $H(\omega)$ oscillates at a frequency of $\omega_c - \omega$. The output signal $\psi_{out}(\omega)$ is a modulation of input signal $\psi_{in}(\omega)$ and circuit state $\phi(x, t)$.

The most prominent feature of a neural circuit is the shark resonant peak at its natural frequency. When a resonant signal passes through, all neurons on the circuit seem to work together to amplify the signal for downstream neurons. They also retain some of its energy and keep a memory of the shared experience. When a non-resonant signal passes through, all neurons seem to be indifferent. The passing signal leaves little trace on the collective memory of the circuit. It's an impressive feature for a storage device, but the phenomenon of resonance is common for all physical objects.

Let us take a look at our previous example of muscle memories. No matter whether a subject is speaking or not, the circuits encoding the muscle memory for all words should already be wired to the muscle fibers. Why don't we hear people constantly utter random words?

Let C^*_{nlg} denote a NLG circuit which includes all neural circuits anchored at a cluster of neurons nlg. The cluster nlg denotes all motor neurons directly connected to all muscle fibers involved in uttering (or writing) the words. Let C^{word}_{nlg} denote a word circuit. C^{word}_{nlg} includes a subset of C^*_{nlg} circuits synthesizing the specific signal for $word$ and delivering it at nlg neurons. For example, C^{cat}_{nlg} and C^{dog}_{nlg} represent the word circuits responsible for generating the muscle impulse for cat and dog, respectively. Different word circuits, e.g. C^{cat}_{nlg}, C^{dog}_{nlg}, C^{pet}_{nlg}, may share some common circuits. The individual circuits of C^{cat}_{nlg} may have different frequencies, but all individual circuits of C^{cat}_{nlg} share a cat-specific resonating frequency.

The NLG circuit C^*_{nlg} behaves like an array of pendulums with variable lengths and masses hanging on the same rod. Without the presence of external forces, all pendulums will swing quite randomly. If we add up the energy from the pendulums found at a specific point, i.e. at neural cluster nlg, and normalize the sum against the total energy of all pendulums, i.e. the energy circling sound all circuits of C^*_{nlg}, it amounts to a tiny number. However, if we twist and turn the rod at a specific frequency, i.e. cat-triggering frequency, all pendulums resonating with the frequency will synchronize. The signal converged at a specific point (motor neurons at nlg) will be unequivocally the specific muscle impulse for pronounce cat. Those not resonating with cat won't respond strongly and their signals at nlg will be weak like random background noises.

Wave Packet

We deduce the state of a circuit $\phi(x,\ t)$ by taking a slice of evenly spaced samples of $\psi(x,\ t)$ at a specific neuron x (2.13). However, there are many slices of $\psi(x,\ \omega_c)$ circling around the circuit. All neurons of the same circuit also pick up frequency component ω_c from their input signals and integrate them to the circuit. As a consequence, many copies of similar signals will circle around the same circuit. How would these signals interact?

2.13 $\phi\ (x,\ t) = e^{i\omega_c t}\ \hat{\psi}(x,\ \omega_c)$

Suppose that the signals travel at a constant speed c on a specific axon. $\phi(x,\ t)$ shall take the general form of $f(x - ct)$. Let wavenumber k and magnitude c_k be two independent variables. By taking the general form of $f(x - ct)$, we could guess $\phi(x,\ t)$ as the function of both time t and location x (2.14). Comparing (2.13) with (2.14), we can deduce $c_k = \hat{\psi}(x,\ \omega_c)$.

2.14 $\phi_k(x,\ t) = c_k e^{ik(x-ct)} = c_k e^{i(kx-ckt)} = c_k e^{i(kx-\omega(k)\ t)}$

The speed c on a specific axon is constant but c around the circuit may vary. Let us represent the time component of $\phi_k(x,\ t)$ by its natural frequency $\omega(k)$, where $\omega(k)$ means ω is a function of k but their exact relationship is unknown yet. The dimensionless quantity kx and ωt are equivalent to the phase shift of the signals circling around. We can describe a circuit either by its temporal frequency ω, which is the number of cycles per unit time, or its wavenumber k, which is the number of cycles per unit distance.

Let us first look at an ideal case where all slices of signals have the same k and ω. They circle around at a constant speed ($c = \omega/k$). According to the superposition of signals, $\phi(x,\ t)$ is a plain wave combining the magnitude of all individual signals (2.15). The probability density $|\phi|^2 = c_{all}^2$ is a constant and doesn't change over position x or time t. Basically, it predicts that sporadic neuron firings evenly spread around the circuit. The prediction is a bit disappointing, but it's not that far away from the truth. The ideal case represents a world with absolute certainty. A world with absolute certainty is an uneventful world. Nothing special is observed around the circuits.

2.15 $\phi\ (x,\ t) = \sum c_k e^{i(k\ x-\omega\ t)} = c_{all} e^{i(k\ x-\omega\ t)} \qquad c\ \in R$

Neural circuits are biological apparatus full of uncertainty. Both natural frequency ω and wavenumber k cannot be absolutely equal around a circuit. Realistically, both may follow some kinds of probability distribution around their average values ω_0 and k_0. To simplify our analysis, let us assume the wavenumber k is normally distributed around a mean of k_0 and a standard deviation of $\sigma_k = \sigma\sqrt{2}$.

2.16 $A(k) = \frac{1}{2\sigma\sqrt{\pi}}\ e^{-(k-k_0)^2/2\sigma_k^2} = \frac{1}{2\sigma\sqrt{\pi}}\ e^{-(k-k_0)^2/4\sigma^2}$

The state of a circuit can be expressed the integral of all slices of signals circling around the circuit using the probability density function $A(k)$ defined in (2.16).

2.17 $$\phi(x,\ t\ ) = \int_{-\infty}^{+\infty} A(k)\ e^{i[kx-\omega(k)t\ ]}dk$$

Let us focus on the observable parts of a neural circuit, namely, the axon of a specific neuron. Assuming signals travel at a constant speed c on the specific axon, where the simple relationship of $\omega(k) = ck$ still holds.

2.18 $$\phi(x,\ t\ ) = \frac{1}{2\sigma\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-(k-k_0)^2/4\sigma^2}\ e^{ik(x-ct)}dk$$

Substitute $u = x - ct$:

2.19 $$\phi(x,\ t\ ) = \frac{1}{2\sigma\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-(k-k_0)^2/4\sigma^2}\ e^{iku}dk$$

Move all terms without k out of the integral:

2.20 $$\phi(x,\ t\ ) = \frac{1}{2\sigma\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-[(k-k_0)^2\ -4iku\sigma^2]\ /4\sigma^2}\ dk$$

$$\phi(x,\ t\ ) = \frac{1}{2\sigma\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-[k^2-2k(k_0+2iu\sigma^2)\ +\ (k_0+2iu\sigma^2)^2\ +\ k_0^2\ -(k_0+2iu\sigma^2)^2]\ /4\sigma^2}\ dk$$

$$\phi(x,\ t\ ) = \frac{1}{2\sigma\sqrt{\pi}}\ e^{-[\ k_0^2\ -(k_0+2iu\sigma^2)^2]/4\sigma^2} \int_{-\infty}^{+\infty} e^{-[k-(k_0+2iu\sigma^2)]^2\ /4\sigma^2}\ dk$$

Rearrange different terms:

2.21 $$\phi(x,\ t\ ) = \frac{1}{2\sigma\sqrt{\pi}}\ e^{-(k_0^2\ -k_0^2\ -4k_0iu\sigma^2\ +\ 4u^2\sigma^4)/4\sigma^2} \int_{-\infty}^{+\infty} e^{-[k-(k_0+2iu\sigma^2)]^2\ /4\sigma^2}\ dk$$

$$\phi(x,\ t\ ) = e^{-u^2\sigma^2}\ e^{-iu}\ \frac{1}{2\sigma\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-[k-(k_0+2iu\sigma^2)]^2\ /4\sigma^2}\ dk$$

Replace Gaussian integral of the last two terms:

2.22 $$\phi(x,\ t\ ) = e^{-u^2\sigma^2}\ e^{-iu}$$

Substituting back $u = x - ct$, we have $\phi\ (x,\ t)$ as a function of position x and time t:

(2.23) $$\phi(x,\ t\ ) = e^{\ -(x-ct)^2\sigma^2}\ e^{-i(x-ct)}$$

How do we interpret the solution? $\phi\ (x,\ t)$ has a magnitude and a phase. The phase causes the real part to oscillate, which is not very interesting. Let us look at the magnitude by taking a snapshot of $\phi\ (x,\ t)$ at $t_0 = 0$.

(2.24) $$\phi(x,\ t_0 = 0) = e^{\ -x^2\sigma^2}\ e^{-ix}$$

By definition, the probability of neuron firings is the absolute square of $\phi\ (x,\ t)$. If we define $\sigma_x = 1/2\sigma$ and choose a normalization factor A so that the total probability sums up to to 1, we get the probability distribution of neuron firings in x.

(2.25) $P(x,\ t_0 = 0)\ dx\ = A\ \left|\phi(x,\ t_0)\right|^2 dx = Ae^{-x^2/2\sigma_x^2} = \frac{1}{\sigma_x\sqrt{2\pi}}\ e^{\ -x^2/2\sigma_x^2}\ dx$

It is a bell shaped wave packet traveling at a group velocity $c = \omega_0/k_0$. A wave packet is a burst of concentrated neuron firings. The packet of neuron firing is confined in a region of space-time, proportional to the signal's wavelength $\lambda_0 = 2\pi/k_0$ and period $T_0 = 2\pi/\omega_0$.

The travelling wave packets are observable. A group of synchronized wave packets traveling along a bundle of similar neural circuits consume so much energy that they should be directly observed via brain imaging. The largest wave packet in a brain is often where the attention is. Some specialized circuits, i.e. the SA node, can maintain non-dispersive wave packets and generate reliable heartbeat signals. Most wave packets eventually get dispersed because ω and k is not perfectly linear. The formation and dispersion of wave packets are the neural basis for many biological rhythms. When two wave packets collide, the two signals get entangled and scattered over the network. It is the neural basis for combining different ideas and generating new ones.

The wave packets propagating over a neural network are analogous to the particles travelling in physical space. They carry some quanta of energy ω_0 and momentum k_0 from source circuits. Wave packets are formed in source circuits and emitted via the axons attached to the source circuits. When wave packets propagate over the network, downstream circuits sense the signals via their dendrites and may absorb and retain some of their energy.

The axons of a neural network offer us a window to observe the signals ($k_0,\ \omega_0$) carried by these traveling wave packets. But, multiple wave packets could travel on the same axon and interfere with each other. The uncertainty in the wavenumber σ_k and the uncertainty in the width of wave packet σ_x follow the same Heisenberg uncertainty principle (2.25). A single neuron firing ($\sigma_x \to 0$) doesn't reveal the precise wavenumver of a signal ($\sigma_p \to \infty$). A neuron firing constantly ($\sigma_x \to \infty$) reveal a signal's precise wavenumber ($\sigma_p \to 0$) but it conveys little information.

(2.26) $\sigma_k\sigma_x = (\sigma\sqrt{2})\ (1/2\sigma)\ = 1/\sqrt{2}$

In the example of NLG circuit, a resonating signal for cat triggers the formation of tiny wave packets in individual circuits of C_{nlg}^{cat}. These wave packets converge at neuron cluster nlg and deliver the precise signal for cat at a precise moment. In order for a subject to speak fast and fluently, all word circuits have to be synchronized. If the wave packets of different words arrive at cluster nlg with some time overlap, the subject will stutter. It's why people suffering a stroke or other brain disorders may speak slowly, have pauses, or utter repeated sounds.

The natural frequency of a circuit ω_0 is determined by the energy level (effort) E_c required for maintaining its state. Suppose that the word circuit taking the least effort to maintain has a time period of $T_0 = 2\pi/\omega_0$. To guarantee that the combined wave packets of different words never

collide at cluster nlg, all word circuits must have a time period of $T_r = r * T_0 \quad r \in \{1, 2, 3, ...\}$ where r is the ranking of the effort taken to deliver the corresponding word.

Since the time period is a property of the circuits and it's not affected by the triggering signals, the occurrence of different words should match how often the signals circle around the corresponding word circuits. If we collect the words spoken or written by lots of people, we should expect that the frequency of each word will be inversely proportional to its rank of effort $f_r = 1/r$ where $f_1 = 1/T_0$, as shown in Fig 2.4.

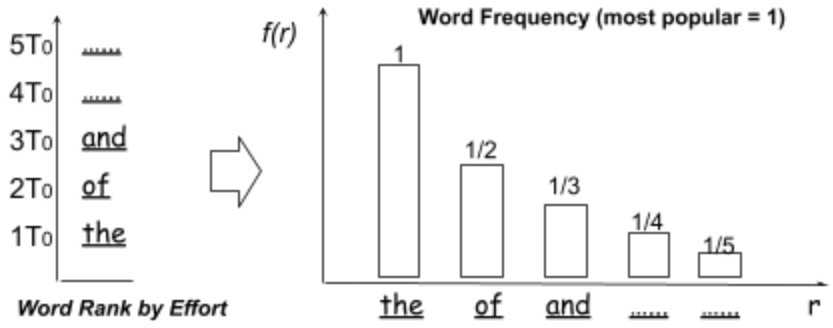

Fig 2.4 Zipf's Law on Word Rank-Frequency Distribution

Most people think they have absolute control over what words they speak, write, or even think. But, the distribution of word rank-frequency observed in many natural languages does follow the commonly known Zipf's law[9]. The phenomenon results from how our brain is wired and operates. It is another example that nature always obeys the principle of least action (effort). The Zipfian distribution is a special power law distribution observed across many natural systems. Based on our analysis, any systems that consist of components with different frequencies and have to share a common resource (output channel) may generate such distributions.

Time Translation

Suppose that, according to an external clock, sensory neuron a and b receive external signal $\psi(a, t_0)$ and $\psi(b, t_0)$ simultaneously. $\psi(a, t_0)$ reaches circuit C_x^* and C_y^* via path $(0, t_{a\to x})$ and $(0, t_{a\to y})$, respectively. $\psi(b, t_0)$ reaches circuit C_x^* and C_y^* via path $(0, t_{b\to x})$ and $(0, t_{b\to y})$, respectively. If $t_{a\to x} < t_{b\to x}$ and $t_{a\to y} > t_{b\to y}$, C_x^* will process the signals as if the event at neuron a happens before the event at b while C_y^* has every reason to believe that event b happens before a. For any real number Δ, $h(t) = f(t - \Delta)$ if $\hat{h}(\omega) = e^{-i2\pi\omega\Delta}\hat{f}(\omega)$. So, the time translation can also be caused by the difference of the phase shifts between two signals in their frequency representation.

The time translation has profound impacts on our experience with reality. The same sequence of events in one place may generate different sequences of signals in different parts of a brain. They may come into different conclusions on what the reality is. The signals arriving at a specific neural

circuit are treated and processed equally. A downstream circuit has no way to discern whether a signal is “real” or not. Therefore, for some parts of a brain, vivid dreams could be as “real” as the actual experience triggered by external sensory inputs. A subject could feel “real” pain when watching another person’s fingers get pricked. By retelling stories of past experience, a brain can relive, rewrite, and reinforce some memories. If the signal comes back via inhibitory connections, it can attenuate or erase an existing memory.

In Pavlov’s experiment, dogs begin to salivate before the food is placed in front of them. Some may interpret that the dogs “learn”, “anticipate”, or “predict” that Pavlov’s assistant would bring them the food. From the perspective of neural signal processing, the sound of the assistant's foodstep, the sight and smell of the food, and other correlated events are just sequences of signals propagating through the dog's neural circuits. When the signals arrive at the neurons connecting to salivary glands, they cause the secretion of saliva. The strength of the signals are accumulated and stored in neural circuits. The conditioning of Pavlov’s dogs could be understood as the shortening of the response time from the onset of a sequence of multiple signals, i.e. the sound of foodstep.

Einstein enlightens us that the simultaneity shall not be taken for granted. The inner workings of a brain gives us the insight that the reality is not absolute, either. It takes $[energy] * [time]$ for the information to propagate. Different parts of a brain are inevitably presented with different versions of reality while interacting with each other. The multiverse picture of reality is true not just for neurons in a brain but also for all brains in the world. Both neurons and brains are computing nodes on a super-neural network. The information propagates over the network heterogeneously. We may share the same reality but we always get our own version of the reality at various times.

The Computation of Neural Network

In essence, a neural network is a function transforming inputs from sensory neurons $\psi(s,\ t_s)$ to outputs to motor neurons $\psi(m,\ t_m)$.

(2.27) $$\sum_{m \in M} \psi(m,\ t_m) = NN(\sum_{s \in S} \psi(s,\ t_s)) \qquad t_m > t_s$$

Let us assume that neurons may attenuate input $\psi(s,\ t_s)$ when desensitized (2.28) and a network may have a built-in bias (2.29).

(2.28) $$\psi(m,\ t_m) = c_r \psi(s,\ t_s) \qquad c_r \in [0,\ 1]$$

(2.29) $$c_0 = a\ bias\ (constant) \qquad c_0 \in C$$

The primitive for building a neural network is ultimately given by Feynman’s rule (1.0). A path $(a,\ \tau)$ can transform input $\psi(s,\ t_s)$ to output $\psi(m,\ t_m)$ with a phase shift of a and a time shift of τ.

(2.30) $$\psi(m,\ t_m) = e^{i2\pi a}\ \psi(s,\ t_s + \tau)$$

A path $(1/2,\ \tau)$ negates input $\psi(s,\ t_s)$.

(2.31) $\psi(m,\ t_m) = e^{i2\pi(1/2)}\,\psi(s,\ t_s+\tau) = -\psi(s,\ t_s+\tau)$

A path $(1/4,\ \tau)$ rotates input $\psi(s,\ t_s)$ by $\pi/2$.

(2.32) $\psi(m,\ t_m) = e^{i2\pi\,(1/4)}\,\psi(s,\ t_s+\tau) = i\psi(s,\ t_s+\tau)$

Connecting $\psi(s,\ t_s)$ to the same output with n identical paths $(0,\ \tau)$ amplifies the signal by a factor of n.

(2.33) $\psi(m,\ t_m) = \sum_n e^{i2\pi 0}\psi(s,\ t_s+\tau) = n\psi(s,\ t_s+\tau) \quad n \in N$

Combining (2.28 ~ 2.33), $\psi(s,\ t_s)$ can be scaled, rotated and translated arbitrarily.

(2.34) $\psi(m,\ t_m) = c_0 + c_1\,\psi(s,\ t_s) \qquad c_0\ ,\ c_1 \in C$

Connecting multiple inputs with path $(0,\ \tau)$ adds up (modulate) multiple inputs.

(2.35) $\psi(m,\ t_m) = \sum_{s\in S} e^{i2\pi 0}\,\psi(s,\ t_s+\tau) = \sum_{s\in S}\psi(s,\ t_s\ +\ \tau)$

Negating subtrahend b and then combining with minuend a performs subtraction.

(2.36) $\psi(m,\ t_m) = \psi(a,\ t_s+\tau) - \psi(a,\ t_s+\tau)$

Circuit $C_s^{0,\tau}$ computes the integral of $\psi(s,\ t_s)$ by sampling it at an evenly spaced time interval τ.

(2.37) $\psi(m,\ t_m) = \int_0^{t_m}\psi(s,\ \tau)\,d\tau$

DTFT and inverse DFT circuits transform $\psi(s,\ t_s)$ between its time domain representation to its frequency domain representation. The frequency domain representation itself is a function of time. It may be used as an input for downstream processing.

(2.38) $\hat{\psi}(\omega,\ t_m) \leftrightarrow \psi(s,\ t_s)$

To compute the n-th derivative of t $\psi(t)$. First use DTFT to convert $\psi(t)$ to its frequency domain representation $\hat{\psi}(\omega)$, use (2.34) to add coefficient $(2\pi\omega i)^n$, and then transform the signal back to its time domain representation.

(2.39) $\psi(t) \rightarrow \hat{\psi}(\omega) \rightarrow (2\pi\omega i)^n\hat{\psi}(\omega) \rightarrow \psi^n(t)$

A differentiable input can be transformed into its frequency domain representation and stored in an array of neural circuits, as shown in the 1st column of matrix (2.40). All changes of the signal strength of each frequency can be captured by the n-th derivatives of the frequency, as

represented by rows in (2.40). The components of any input signals can be decomposed as a sum of Taylor series and implemented by a neural network as a $f \times n + 1$ matrix.

(2.40)
$$\begin{matrix} \psi(s_1) & \psi'(s_1) & \psi''(s_1) & \psi'''(s_1) & ... & \psi^{(n)}(s_1) \\ \psi(s_2) & \psi'(s_2) & \psi''(s_2) & \psi'''(s_2) & ... & \psi^{(n)}(s_2) \\ \psi(s_3) & \psi'(s_3) & \psi''(s_3) & \psi'''(s_3) & ... & \psi^{(n)}(s_3) \\ ... & ... & ... & ... & ... & ... \\ \psi(s_f) & \psi'(s_f) & \psi''(s_f) & \psi'''(s_f) & ... & \psi^{(n)}(s_f) \end{matrix}$$

If an output has any relationship with any components of input signals, the output can be approximated as a sum of Talyor series.

(2.41)
$$\psi(m,\, t_m) \,=\, \sum_{f,\, n=0}^{\infty} NN_{f,n}[\, \psi^{(n)}(f,\, n,\, t_s)]$$

A sufficiently large network can build a hierarchy of subnets of different temporal resolutions and approximate the transform function between two arbitrary differentiable signals.

What if the target output is not differentiable and has no relationship with any inputs?

According to the Central Limit Theorem, a sum of random signals becomes normally distributed as more and more of the random signals are added together. We can feed the output of a circuit back to its input and create non-dispersive Gaussian wave packets circling around the circuit. Integrating the Gaussian signals creates the Cumulative Distribution Function (CDF). CDF can be used to approximate step functions. A continuous stream of Gaussian packets from the same circuit is a reliable heartbeat signal. Integrating the heartbeat signal over multiple time periods creates a linear function. Modulating a CDF signal and a linear signal creates a Gaussian Error Linear Unit (GELU). GELU can be used to approximate a Rectified Linear Unit (ReLU). More functions can be built using these basic building blocks. Within a certain limit, an analytic function can also be approximated by neural circuits.

Here is another fascinating example. Let C_ζ^t be a cluster of N circuits anchored at neuron ζ, $\omega_n = ln(1/n)$ the frequency of the n-th circuit, $c_n = e^{ln(1/n)/2}$ the coefficients linking the n-th circuit to anchor ζ (2.42). As time goes by, neuron ζ will reach an absolute tranquility at some special moments of t where $\zeta\,(1/2 + it) = 0$. If C_ζ^t is wired in a brain, the mind is destined to wander along the critical line of Riemann's zeta function and cannot help appreciating prime numbers.

(2.42)
$$\psi(x,\, t) \,=\, \sum_{n=1}^{N} c_n e^{i\omega_n t} \,=\, \sum_{n=1}^{N} e^{ln(1/n)/2}\, e^{i\cdot ln(1/n)\cdot t} \,=\, \sum_{n=1}^{N} (1/n)^{\,(1/2\,+\,it)} \,=\, \zeta\,(1/2 + it)$$

We need only ~42 decimals of π to compute the circumference of the visible universe to the accuracy of a proton. Yet, we can comprehend a perfect circle far beyond the limitations of physical reality. It's probably due to how our brain is wired.

Dirac insists that the laws of nature should be expressed in beautiful equations. Simple analytic equations have great appeal because they can be represented by simple circuits and resonate well with similar ones in the brains of our peers. If our brain were a digital computer, all functions could be implemented as a lookup table. We may be more susceptible to arbitrary laws of physics made of lots of independent variables (dimensions).

The Development of Neural Computer

The digital computer is approaching its physical limits. One solution is to use photons for computing because photons are faster and allow a higher bandwidth than the electrons used in digital computers. Most research projects in photonic computing focus on replacing electronic transistors with optical equivalents. The goal of the field is not to change the paradigm of Turing machines but to make them faster.

Another solution is the paradigm shift of computing. To differentiate our approach to the current quantum computing, we'll call it neural computing. Any networks of connected points where the propagation of signals (energy) follows Feynman's rule (1.0) is a neural computer. The key to build a neural computer is to connect neurons with desirable paths (α, τ). A biological brain is a neural computer made of connected neurons. We can connect the optical fibers to make optical neural computers. We can also simulate neural computers in software, but the simulation running on digital computers will be slow and hard to scale, as we will discuss later.

Fourier Analysis

Before we discuss the development of neural computers, we must verify that our model of neural computation is in touch with physical reality. Since the propagation of photons follows Feynman's rule, we should be able to use optical fibers to implement DTFT circuits. If we encode a signal as a beam of light with different colors and strengths, the circuits shall output the frequency domain representation of the input signal. The inverse DFT circuits shall transform the frequency domain representation back to the original signal.

The key wiring of DTFT circuit (Fig 2.2) is that all paths from input x_0 to a specific output C_k share the same phase shift but differ in time. We can implement such a circuit using two types of optical fibers made of different materials, as shown in Figure 3.1.

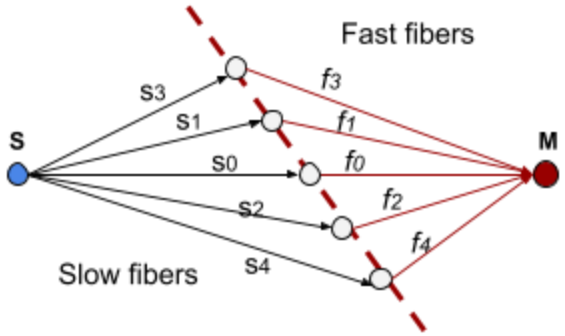

Fig 3.1 DTFT Circuit by Optical Fibers

Let us assume that light travels slow in optical fiber s_i with a shorter wavelength λ_s and fast in f_i with a longer wavelength λ_f. First, we connect two points S and M with a straight path by merging s_0 and f_0 in the middle. When we add new connections between S and M, we make sure that the total wavenumbers along the paths are constant:

$$s_{2i} = s_0 + i * \lambda_s \text{ and } f_{2i} = f_0 - i * \lambda_f$$
$$s_{2i+1} = s_0 - i * \lambda_s \text{ and } f_{2i+1} = f_0 + i * \lambda_f$$

Imagine that we add millions of such connections and merge s_i and f_i along the dash line. Next we add millions of new output neurons along the dotted line, as represented by R, G, and B in Fig 3.2. We connect the output neurons to the merge points along the dash line using fast fibers. According to our model of neural computation, if we shine a signal encoded as a beam of light of mixed colors and strengths at S, we will see the frequency domain representation of the signal along the dotted line. The color and intensity of light observed at R, G, and B correspond to the frequency components of the original light signal.

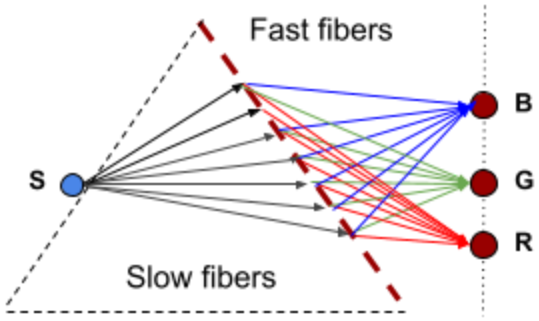

Fig 3.2 Melt Optical DTFT into Prism

Since all slow fibers are made of the same material (glass) and they're densely packed within the triangle area, we can just melt them into a single block of glass. We can play the same trick on fast fibers or simply replace them with a new transparent material called air. It transforms our optical implementation of DTFT neural circuits into a prism. So, prisms are the cheapest, fastest,

and most reliable neural computers. We can all order them online for a few dollars and verify that they do perform spectrum analysis in constant time.

As a proof of concept, we can build a speech recognition circuit employing the same parallel communication protocol in neural computers. Suppose that we encode our information in a sound wave, i.e. giving each word a unique sound. We can connect strings of different lengths and tensions to two frames, as shown in Fig 3.3. For each word, we record the vibration of all strings and store the mapping of vibrations to words in a lookup table. When the sound of a word is emitted at S, all strings of speech recognition circuit will process the signal simultaneously and result in the output of the word at M. Reversing the circuit creates a Text-To-Speech (TTS) circuit. To pick up very subtle vibrations, we can replace the frames with an echo chamber of similar shape and fill the chamber with some fluids. If so, our speech recognition circuit essentially becomes the same structure that nature has been developing in our ears for billions of years.

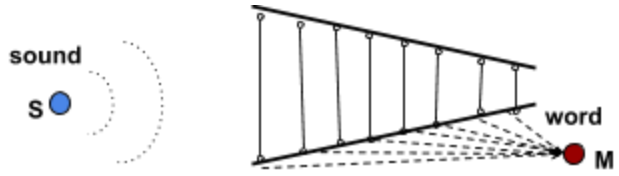

Fig 3.3 Speech Recognition Circuit

It's nice to see the physical implementation of our neural circuits in nature. They run the expected neural computation. Next we'll see how nature builds more complicated neural computers through trial and error.

Attenuation and Amplification

The simplest neural network is a single sensory neuron S directly connected to an actuator M. As an organism grows in size, so grows its neural network. Some intermediate neurons are added to relay signals from sensory neurons to actuators. If the intermediate neurons don't change the shape or magnitude of the signals, we usually ignore intermediate neurons and draw a direct path between input S and output M (Motor neuron). We may annotate the path by its phase and time shift (α, τ), as shown in Fig 3.4.

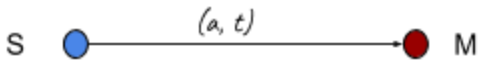

Fig 3.4 The Simplest Neural Network

All neurons have an absolute refractory period during which the sodium channels are inactivated and remain so until hyperpolarization occurs. During the absolute refractory period (1~5ms), a second stimulus (no matter how strong) will not excite the neurons. The refractory period of a neuron determines its maximum sampling rate of input signals. Neurons usually have a builtin

mechanism to attenuate inputs. For instance, photoreceptor cells can adjust the sensitivity to detect a few photons in the dark and quickly adjust to cope with the bombardment of millions of photons under the sunlight. The attenuation allows the downstream networks to operate under their max firing rates so that the input signals are not distorted to square signals. The attenuation is represented by an attenuating triangle pointing to the direction of signal propagation, as shown in Fig 3.5. The attenuation may be annotated by an attenuation factor $a \in (0, 1)$.

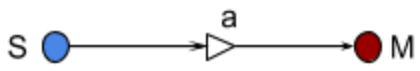

Fig 3.5 Attenuation

A network of neurons is capable of amplifying signals with high fidelity and low latency. For instance, the signal from a dancing hair cell[8] can be fanned out to millions of muscle cells to control the whole body movement. It allows people to enjoy the illusion that they're synchronizing their movement to music. Signals can be amplified by focusing similar inputs on a single neuron or fanning out an output to an array of actuators (Fig 3.6 a). The amplification is represented by an amplifying triangle in the direction of signal propagation and annotated by amplification factor A (Fig 3.6 b).

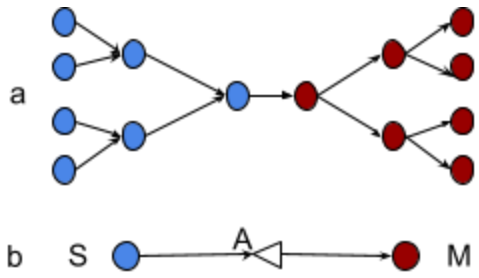

Fig 3.6 Amplification

Attenuation and amplification change the magnitude of signals, namely, the energy carried by the signals. They don't change the shape and frequency of signals, where the information is encoded. Because of attenuation and amplification, the signal processing of a neural network can operate at very low power and largely maintain the linearity. Assuming inputs and outputs are attenuated and amplified properly, a processing unit of a neural computer can be represented by a simple block diagram (Fig 3.7).

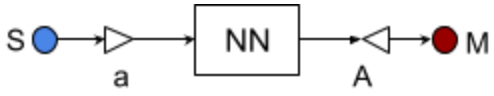

Fig 3.7 A subnet with properly sized I/O

Memory and Recall

If a connection between input S and output M is crucial to the survival of a species, e.g. salivating on the smell, sight, or taste of the food, a direct neural path between S and M is usually preserved during the evolution of the species. Those who cannot simply go extinct.

As a neural network evolves, some neurons on the crucial path $S \rightarrow M$ may make random connections to the nearby dendrites and form numerous neural circuits. As we discussed before, neural circuits integrate incoming signals and keep a running sum of them. The information stored in each circuit is determined by the natural frequency of the circuit. Some have a very short time period while others keep track of longer rhythms like daily patterns.

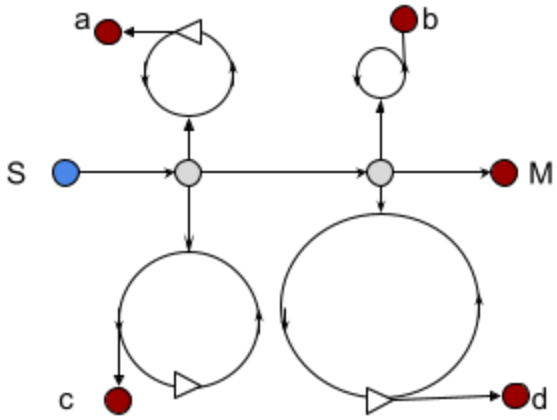

Fig 3.8 The Formation of Neural Circuits

The signals circling around various circuits may become inputs to other downstream processing, as shown in Fig 3.8 a, b c, and d. If a subject (dog) is fed regularly everyday, the signals on daily circuits will accumulate and become stronger and stronger. The subject may feel hunger at the regular feeding time even if there are no feeding signals found on path $S \rightarrow M$ yet.

The intensity of neuron firings around a circuit is limited by the total power of neurons on the circuit. The signals will get attenuated over time and the weight of the contribution of ancient signals will diminish. Therefore, what neural circuits actually compute and store is a running sum of incoming signals with a decay factor. The decay factor determines the time window of the running sum.

For a densely connected large network like the human brain, i.e. ~100 billions neurons and ~10,000 synapses per neuron, there may be enough capacity for the network to track the stats of all processed signals. The wave packets travel at group velocity not at the speed of light. If we have an optical implementation of new circuits, as shown in Fig 3.8, we may see wave packets, bright spots of light, form and circle around. The intensity of light along the circle is proportional to

the accumulated strength of input S. If we attach such circuits to every interesting point in an artificial neural network, we can inspect the accumulated signals by measuring the intensity of the light of the corresponding circuits.

Pattern Recognition

Suppose that S_0 is the detection of the smell or sight of the food, M_0 the secretion of saliva, and S_i the detection of a random event, e.g. the ring of a bell, the sound of footsteps, or just being in Pavlov's lab. Let's also assume that neuron y on a side circuit attached to path $S_0 \rightarrow M_0$ make a random connection back to neuron z on path $S_0 \rightarrow M_0$, and neuron v on path $S_i \rightarrow M_i$ also make a random connection to the same side circuit, as shown in Fig 3.9. These connections are all created by chance. The connections are pretty weak and the signals are heavily attenuated because there are not many neurons powering the signal propagation along the side circuit.

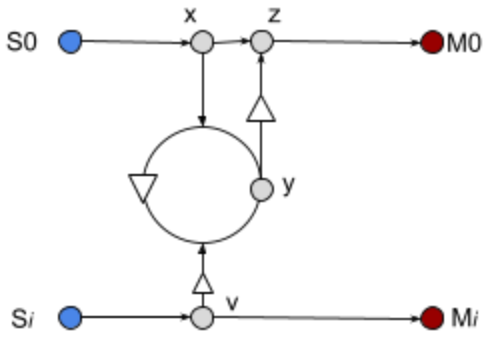

Fig 3.9 Pavlov Circuit

When S_0 arrives, it propagates along the main path $S_0 \rightarrow M_0$ and takes care of the salivation business as usual. A tiny fraction of the energy carried by S_0 is siphoned off the main path at neuron x and propagates along the circuit. The energy is further diverted at various neurons on the circuit. When it propagates back to neuron z via neuron y later, it's too weak to cause any ripples along path $z \rightarrow M_0$. The same happens to other signal S_i. A bit of the energy from S_i may be diverted toward the circuit , but it shouldn't affect the main business along the established path $S_i \rightarrow M_i$. It's too weak to cause any ripples to other paths like $S_0 \rightarrow M_0$. So far, such side circuits look like some useless but benign random creation.

It's business as usual with one exception. If the events represented by S_0 and S_i happens within a small time window, the energy diverted from different signals to the circuit will add up. The constructive interference may form a sizable wave packet traveling along the circuit. When the signal arrives at neuron z, it will be stronger than those triggered by either S_0 or S_i alone. The signal may be still too weak to cause the salivation at M_0 for the first time. If the pattern repeats a few times, the energy (strength) circling around the circuit will get accumulated. Soon enough, S_i alone can put a signal over $y \rightarrow z \rightarrow M_0$ strong enough to cause the salivation at M_0.

The training/conditioning doesn't have to be on consecutive days or within a few minutes. It just needs to happen on a regular basis (frequency) so that the combined signals can be accumulated by some side circuits. For a densely connected network, there will be so many side circuits between the more established paths. They will capture the subtle constructive interference among different signals and create new signal transduction paths. The manifestation of such new connections is the behavior of pattern recognition.

A large network of neural circuits are capable of learning lots of new patterns spontaneously. The underlying mechanism suggests that neural learning behaves very differently from the current approach of machine learning. First, the learning doesn't require backpropagation or distinct learning and inference phases. If the signal is not heavily attenuated along the side circuits, the network can learn a pattern with a few examples (one-shot learning). If so, the superficial correlation will be easily overwritten by another new correlation. This is probably what happens in the head of those who watch the financial market and keep finding new patterns. If the correlation among different signals disappears, the learned behavior may fade away because of attenuation. If the correlation persists, the activity may stimulate the formulation of similar connections along the new path. The connection can be strengthened to a level that the new path $S_i \rightarrow M_0$ becomes second nature.

The signal circling around the Pavlov circuit, as shown in Figure 3.6, represents a temporal correlation M_0 between two signals S_0 and S_i, namely, $M_0 = S_0 + S_i$. The strength of the correlation, namely, how often the correlation happens, is captured by the strength of the signal circling around the circuit. The signal persists in the circuit even when the events of S_0, S_i, and M_0 are not present. Therefore, the circuit is the physical embodiment of the concept of the correlation $M_0 = S_0 + S_i$. Since S_0 and S_i also result from the processing of some upstream circuits and the strength of the accumulated S_0 and S_i are stored in some circuits, we can replace the detailed Pavlov circuit, as shown in Figure 3.6, with a simplified concept activation circuit, as shown in Fig 3.10.

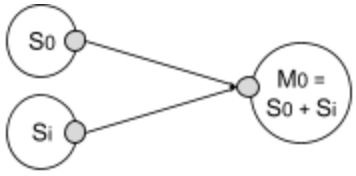

Fig 3.10 Concept Activation Circuit

The concept extracted by neural circuits can be both temporal and spatial correlation between two signals. In fact, it can be any mathematically significant relationship of an arbitrary combination of any number of signals. For example, when the visual signals propagate from retina to CNS, the neighboring neurons, represented by a 3×3 matrix (Fig 3.11), may make random connections to the same neuron on the next layer.

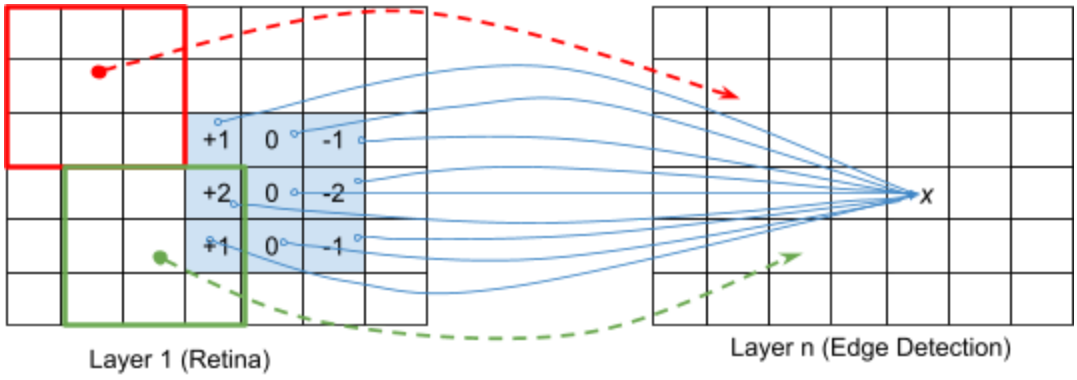

Fig 3.11 Spatial Relationship Detection

As discussed before (2.34), upstream signals may be translated, scaled, and/or rotated when reaching downstream neurons. The coefficients of the connection between two layers can be represented by a matrix, often called a kernel in digital image processing. For instance, the kernel (solid square in Fig 3.11) will excite neuron x if and only if there is an edge in the region of the solid square. For all practical purposes, the neural circuit C_x^* anchored at neuron x are fully conscience about one specific thing happening in the little world of 3×3 square on a retina. The life of C_x^* is centered around the perpetual questions:

> *Am I seeing an edge in my little world right now?*
> *How often have I seen such an edge in the last second, minute, hour ...?*

A large neural network making lots of random connections will inevitably develop a variety of kernels for detecting different features. If the network is deep enough, the features extracted at a higher layer can be invariant to the translation, scaling, and/or rotation of the images received on a retina.

Symbolic Language

When we hear or see a word, many circuits will get excited in different regions of our brain. Imagine that we invent a non-invasive technology that can monitor every circuit in the specific regions in charge of Natural Language Understanding (NLU) and Generation (NLG). We can identify the excited circuits when a specific word is spoken. We can measure the strength of the signal circling around the circuit even when the word is not spoken.

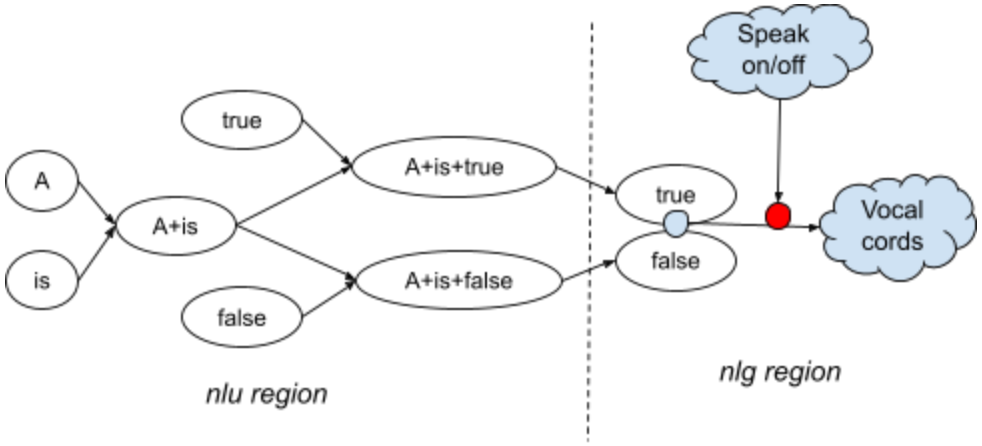

Fig 3.12 NLU/NLG circuits

Let us see what happens if the subject hears a sentence like: $A\ is\ true$. The NLU circuit C^{A}_{nlu}, C^{is}_{nlu}, and C^{tue}_{nlu} will get excited sequentially. The energy stored in C^{X}_{nlu} is proportional to the total number of neuron firing triggered by a specific signal X. The signal strength in C^{A}_{nlu}, C^{is}_{nlu}, and C^{true}_{nlu} corresponds to how often the word A, is, and $true$ are processed in the NLU region, respectively. The word doesn't have to come from the same sentence. For Instance, the signal in C^{A}_{nlu} will get strengthened when the subject hears, speaks, or even thinks sentences like $A\ is\ false$, $A\ is\ funny$, $A\ loves\ B$, etc. Just as discussed in the example of Pavlov's dogs, C^{A+is}_{nlu} and all downstream circuits such as $C^{A+is+true}_{nlu}$ and $C^{A+is+false}_{nlu}$ will get strengthened equally if the bigram $A\ is$ is heard, spoken, or thought by a subject.

Suppose that $C^{A+is+true}_{nlu}$ and $C^{A+is+false}_{nlu}$ are connected to C^{true}_{nlg} and C^{false}_{nlg} in the NLG region. Let's see how a subject would complete his sentence after speaking $A\ is\ ...$ or answer $is\ A\ true\ or\ false$. The prefix will excite C^{A}_{nlu}, C^{is}_{nlu}, and C^{A+is}_{nlu} equally. The wave packet of the combined signals will arrive at both $C^{A+is+true}_{nlu}$ and $C^{A+is+false}_{nlu}$ around the same time. Because the upstream signal is the same, the probability of $true$ or $false$ getting excited is proportional to the energy (signal strength) in $C^{A+is+true}_{nlu}$ or $C^{A+is+false}_{nlu}$, respectively. If the subject has heard or thought about $A\ is\ true$ a lot, he's very likely to complete the sentence by saying $A\ is\ true$. If the signal strength in $C^{A+is+true}_{nlu}$ and $C^{A+is+false}_{nlu}$ are close, the probability of the subject saying either $A\ is\ true$ or $A\ is\ false$ is close, too.

Therefore, the energy (signal strength) in $C^{A+is+true}_{nlu}$ and $C^{A+is+false}_{nlu}$ represent a subject's belief on the statement $A\ is\ true$ and $A\ is\ false$, respectively. All thoughts ever coming across our mind are not only stored in neural circuits but also weighed by them. It's how we can speak fluently without long pauses between words. If the difference between alternative thoughts is small, we may hesitate when the circuits in our brain are waiting for some subtle signals from more convoluted paths to tip the balance.

A machine learning expert immediately cries out that it is an absurd idea. For a language with $\sim 100,000$ common words like English, a subject would need $100,000^{10} = 10^{50}$ circuits to hold all possible thoughts consisting of 10 words. It's more than all atoms on our planet. Even if it takes only a nano-second to search a path, it will take longer than the age of our universe to complete a sentence of 10 words.

Fortunately, our brain is not a digital computer but a neural computer. The storage problem is solved by sharing neurons and paths among circuits. A neuron has an average of $10,000$ synapses. If we start with a neuron, we could find $10,000^{100} = 10^{400}$ unique paths after following 100 connections. Even if a tiny fraction of these paths form valid circuits for storing different thoughts, it has the capacity for a lot of thoughts. It's true that grasping a language like English requires a huge storage capacity. That's why dogs and cats haven't started to talk yet, but it's obviously not out of reach for a brain like ours.

The searching problem is even easier. The processing of a neural computer is parallel signal propagation over the network. It's very different from a Turning machine, which encodes the information in bits and processes them by flipping the bits sequentially. The parallelism achieved by modern multi-core processors is nothing compared with how fast all molecules in an opera house can process and respond to the vibration originated from a singer's vocal cords. Once the wiring is done in a neural computer, the time complexity is always $O(1)$. The physical size of our brain is in the order of 0.1m. The speed of signal propagation in CNS is estimated in the order of ~10 m/s (Grey matter largely consists of unmyelinated neurons) . So, the predicted speed of human thoughts is a few words per second. It sets an up limit on how many words per minute people can speak, write, type, comprehend, and think.

Reinforcement Learning

Suppose that a kid knows 3 words, $Rock\,(R)$, $Paper\,(P)$, and $Scissors\,(S)$, but the kid is never told the rule of Rock-Paper-Scissors. There are no strong correlation among R, P, and S in everyday conversations. The signal strength in C_{nlu}^{R+P}, C_{nlu}^{R+S}, C_{nlu}^{S+P} are equally weak. When someone says $Rock\,(R)$ to the kid, C_{nlu}^{R} gets excited first, and then C_{nlu}^{R+P} and C_{nlu}^{R+S}. Though both C_{nlu}^{R+P} and C_{nlu}^{R+S} are connected to downstream C_{nlg}^{RPS}, the kid won't know what to say. If asked to choose $Paper\,(P)$ or $Scissors\,(S)$ without other hints, he probably picks a random one because the signals in C_{nlu}^{R+P} and C_{nlu}^{R+S} are equally weak.

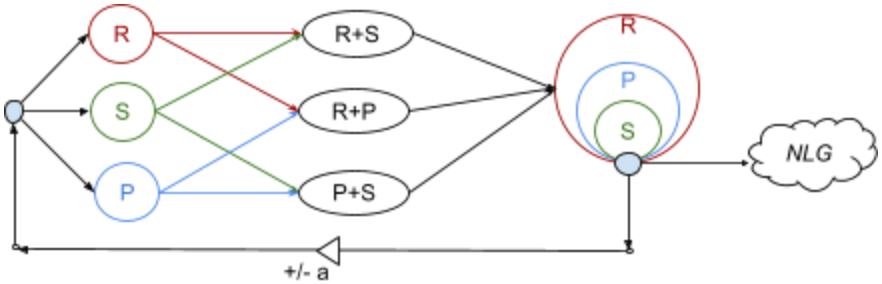

Fig 3.13 RL Circuit for Rock-Paper-Scissors

If the kid gets a reward or a punishment after giving a random response, his brain will recognize the correlation between his response and the feedback. For instance, if he says $Paper\ (P)$ after hearing $Rock\ (R)$, the same signal will propagate back to C_{nlu}^{P}, and then C_{nlu}^{P+R} and C_{nlu}^{P+S}. If it is a reward, the feedback signal is positive and strengthens the signals in C_{nlu}^{P+R} and C_{nlu}^{P+S}. If a punishment, the signal is negative, out of phase therefore inhibitory, and weakens the signals in C_{nlu}^{P+R} and C_{nlu}^{P+S} (Fig 3.13).

As long as there is a feedback mechanism, the feedback will tip the balance between different choices. Reinforcement learning is handled by very basic circuits. The feedback doesn't have to be external rewards or punishments. It can come from other parts of the brain. A dog can learn to pick the right letters after enough training. It doesn't require fully developed language capability. Both the kid and the dog will learn how to respond properly even if they cannot articulate the rules of the game in a symbolic language.

What if the kid plays the game with a computer which throws a biased dice to decide R, P, or S? Let's start with a simple example: the computer always says R, which excites C_{nlu}^{R+S} and C_{nlu}^{R+P} equally. The feedback signal tips the balance and the kid will respond with P more often than S. It further strengthens the signal in C_{nlu}^{R+S} and C_{nlu}^{S+P}. Since the computer never says S or P, the signal in C_{nlu}^{S+P} won't affect the result. For each round, the energy in C_{nlu}^{R+S} and C_{nlu}^{P+S} increases by 1 unit while C_{nlu}^{R+P} by 2+ units. If we normalize the probability among these 3 circuits, the probability of C_{nlu}^{R+P}, namely, answering P after computer's R, increase monotonically while the other two decrease monotonically.

Let us assume that the computer says R half of the time, and S the other half. After the computer says R and the kid responds with P, C_{nlu}^{R+S}, C_{nlu}^{R+P}, and C_{nlu}^{P+S} will get updated by 1, 2, and 1, respectively. Since C_{nlu}^{R+S} and C_{nlu}^{P+S} are strengthened equally, it won't affect the choice when the computer says S. The feedback signal, no matter how weak it is, will train the kid to respond with R. Each round starting with the computer's S and the kid's R will update C_{nlu}^{R+S}, C_{nlu}^{R+P}, and C_{nlu}^{P+S} by 2, 1, and 1, respectively. If there are equal rounds started with R and S, C_{nlu}^{R+S}, C_{nlu}^{R+P}, and C_{nlu}^{P+S} will get updated by 3, 3, and 2. After the normalization, the probability distribution of the kid's responses tries to mirror the probability distribution of the computer's choices.

We can analyze a more complicated probability distribution of various inputs. The conclusion should hold that, for a neural network with a feedback mechanism, the probability distribution of outputs always tries to mirror the probability distribution of its inputs. The strength of the coupling is determined by the strength of the feedback mechanism. Most people may find it hard to believe that their behavior (the outputs of their brain) is merely a mirror of the inputs. As more and more human activities are migrated online, the collective behavior of billions of minds can be observed and measured with little disturbance. It should be possible to prove the mirroring scientifically.

The Future and Beyond

Suppose that nature hands you a neural computer and informs you that you're now in charge of its future development. You cannot find any manuals. You don't even know why it's built in the first place. How are you going to further its development? Well, to envision the future, we have to understand the past.

Let us imagine an alien scientist studying how human society works from a distant galaxy. He first points his telescope to the U.S. He diligently traces financial transactions and telecommunication packets. He concludes that Wall Street is essential to financial activities and Silicon Valley to Internet services. Being a good scientist, he drops some nuclear bombs onto these places and confirms that all financial activities and internet services are severely disrupted before publishing his findings in the scientific journal Nature or Science. His conclusion is not entirely wrong but he misses the essence of human society. If he first points his telescope to China, he will find out that financial activities and internet services are organized in very different ways. The development of biological brains is like the development of human society. It's a chain of historical accidents. Doing simple archaeology doesn't lead us to a true understanding of the underlying mechanism.

You may wonder what's the essence of human society. If the alien has the superpower to uproot a large number of random population from the Earth and place them on another inhabitable planet, he will find out that individuals spontaneously interact with each other. They start to pursue lots of interesting or nonsensical activities. Sooner or later, complex behavior emerges. They build activity centers similar to Wall Street or Silicon Valley as if there were an invisible hand guiding them. Similar process happens in the evolution of biological brains.

Based on our model of neural computation, the behavior of neural networks solely results from individual neurons interacting with their immediate neighbors. The key to neural computation is the connections (a, τ) between neurons. But, it's still a mind-boggling task to fine-tune an astronomical number of connections to produce functional brains with a relatively low failure rate. How can 100 billion mindless neurons accomplish such a task without divine intervention?

First, it's worth pointing out that the assignment of functionalities among different parts of a brain is a product of evolution. The high level wiring instructions among different regions of a brain are encoded in the genome and guided by biomarkers during its early development. Such guidances are essential to the instantiation of a specific brain, but they are not essential to the computation

to be performed. If we would wire an array of infrared or sound sensors to the visual cortex of a blinded brain in its early development, the brain could have developed the capabilities of detecting objects via infrared vision or echolocation.

Let us assume that an axon is guided to a target region by biomarkers. What neurons should it connect to? The answer is simple. It doesn't really matter. No neurons are special in a small region. The axon can just connect to any neurons in the neighborhood and explore all options. After an axon makes its initial connections, how does it figure out what connections to be strengthened and what connections to be pruned? The answer is important because the number and strength of connections are essential to the nature of neural computation to be carried out.

The natural frequency of neural circuits is proportional to the energy level $E_c = nE_0T/\Delta$ required to maintain the state of the circuits. The energy to maintain the action potential across the membrane on the axon is relatively stable. Most variations come from the membrane of dendrites. The action potential across the membrane on dendrites reflects the activity and usefulness of the connection. It's the initial trigger of a signal transduction pathway that regulates gene expression and leads to the formation or elimination of similar synapses. Therefore, the fine-tuning at local level follows the simple evolution process. It starts with random connections. The winners get resources to build more similar connections. The losers get starved and wither away.

Unlike digital computers, a biological brain is a living organism. The development of a brain is a continuous evolution of the network. A question remains, how does the network know whether it's done a good job globally? For a large and densely connected neural network like the brain of Pavlov's dogs, it's plausible that there exist some neural paths between the circuits detecting the sound of the assistant's footsteps and the circuits causing salivation. But, why does the evolution of the network lead to the shortening of the response time and not the opposite? What's the force driving evolution? How is the direction of evolution determined?

As mentioned in the introduction, a signal propagating from one circuit to another can be modeled as a drop of energy propagating between two points in a configuration space. The fine-tuning of a neural network during its development is the process of exploring the paths in the configuration space. The energy circling around the newly established paths lowers the cost $[energy] * [time]$ for the propagation of future signals along the paths. If the cost of maintaining the paths is amortized, most signals (energy) seem to propagate along the least action paths. It's analogous to spending resources on building bridges and widening roads when a new route is found. If the demand keeps increasing, the newly established route becomes a major highway and will transport most future cargos. If not, the route becomes less maintained and the resources will be recycled to somewhere else. In the end, most energy of the system, i.e. signals, cargos, particles, and so on, seem to flow through the paths with the least cost, defined as $[energy] * [time]$.

For the physicists who have the luxury to observe the evolution of a system from a great distance, they find the principle of least action and derive simple equations governing the motion of objects. For the observers who have to live within the observed, they find it hard to formulate the motion of

life in a simple equation because they're distracted by much nuanced interactions. The best description so far is Darwin's "evolution by natural selection", which describes the same physical process as the principle of least action. In summary, evolution by natural selection is the process of exploring the paths in the configuration space. The principle of least action is the natural (destined) result of such an exploration and may be viewed as the driving force of evolution.

There are many laws of physics. Some are more fundamental than others. The two overarching principles that underlie most laws of physics are the principle of equivalence and the principle of least action. The equivalence principle can be simply stated that the fundamental laws of the universe shall not change because of the observers. This leads to the general idea of relativity and symmetry. So follow all conservation laws and equations of motions. Continuing this line of thought, the motion of life shall be governed by the same principles of physics. This helps to unify our understanding of the motion of all objects, including the evolution of living objects on Earth.

The evolution of species is the process of exploring genetic configuration space. The flow of energy from prey to predator is a form of signal propagation between two points (genomes) in the genetic space. The population of a species corresponds to the energy accumulated at a specific point in the genetic space. Nature explores the genetic space by tinkering one mutation at a time. Because of the principle of least action, the system is expected to move towards a configuration where most energy flows through the least action paths, resulting in survival of the fittest.

After billions of years, the evolution of species locates a point in the genetic space where most energy is devoted to the development of biological brains. The emergence of the human brain results in a network of connected brains, namely, human society. The evolution of human society is the process of exploring the configuration space of ideas. Money and power are the proxy of the resources (energy) that a specific idea commands. The competing ideas are embodied by individuals, companies, religions, or other entities. The least action principle is the invisible hand guiding the resource allocation among competing ideas.

We're all part of a neural computer that's been evolving for billions of years. We receive and react to the inputs from our neighbours. Our reactions in turn become the inputs to our neighbours. This is how the neural computer was built in the past and will be developed in the future.

The Nature of Consciousness

Consciousness is everything that we experience as humans. The discourses on consciousness often follows a familiar pattern:

> *What is consciousness?*
> *Consciousness is blah blah blah …*
> *What is blah then?*
> *Blah is foo …*

What is foo?
Foo is bar …
… …

To avoid chasing our tails like dogs, we need to anchor our discussion to concrete physical reality.

Develop Abstract Concept

Let C_{nlg}^{cat} denote the word circuit responsible for uttering the English word cat. The word circuit C_{nlg}^{cat} is anchored at nlg, a cluster of neurons directly connected to muscle fibers on vocal cords. We may not agree what cat is, but we have a precise definition of what it means by pronouncing the English word cat. It's a ~400ms wide neural impulse that's very different from C_{nlg}^{dog}, $C_{nlg}^{猫}$ (the Chinese word for cat), or C_{nlg}^{meow} (making a cat sound).

Now, let us move to the sensory side and define what it mean by seeing a cat. we have to start with photoreceptors C_{cone}^{x} and C_{rod}^{y}. These photoreceptors have no idea of what cat is, but they do have an identity x or y and knows the concept of being hit by a photon. The signals encoded by C_{cone}^{x} or C_{rod}^{y} are very simple. They have a magnitude of either 1 or 0. They also have a phase $e^{i\omega t}$, the key difference between the signals of C_{cone}^{x} or C_{rod}^{y} and the pixels of digital images. The phase $e^{i\omega t}$ is like a clock carried by the signals and it tells the rest of the brain whether the two photons hitting C_{cone}^{x} and C_{rod}^{y} are originated from the same object or not.

When the signals travel down the network, the neighboring axons may be connected to the same neurons. Let C_{edge}^{z} denote the circuits connected by the neighboring neurons originated from a small region on the retina. The axons from C_{cone}^{x} or C_{rod}^{y} may be connected to C_{edge}^{z} via a variety of paths (α, τ). Given enough connections, it can be proven that some C_{edge}^{z} will fire if and only if an edge is detected within a specific region on the retina for a small time interval Δt. So, we can say that C_{edge}^{z} develops the concept of an edge in the specific region. The concept can be precisely represented by a signal lasting a time period of Δt. The edge signals also carry a clock $e^{i\omega t}$ so that the rest of the brain can tell whether different edges are from the same object or not.

As the signals from C_{cone}^{x}, C_{rod}^{y}, and C_{edge}^{z} travel down the network, the signals are recombined in various ways. Some circuits will get excited if and only if some high level features are present. Namely, these circuits develop and store the concepts for the corresponding features. When the signals of different features reach a layer deep down in the network, a neuron may get lighted up if and only if a cat shows up in the front of the eyes. We call the neuron cat neuron and denote all contributing circuits as C_{image}^{cat}. The cat neuron is only activated at the sight of cat, but the individual circuits of C_{image}^{cat} store all cat features. If we connect an inverse IDF circuit to the cat neuron on C_{image}^{cat}, we will get a precise signal of what a cat looks like. It's an average of all cat

images that the network's seen. The signals of cat neurons may get scaled, rotated, and translated. They are recombined and processed by other downstream layers. A circuit may get excited regardless of the size, orientation, and location of cat images. We call the circuit $C^{cat'}_{image}$, which develops the abstract concept of how a cat looks like.

If we zoom in the cat neuron on C^{cat}_{image}, we find it's not a single neuron but a cluster of neurons. A subset of neurons on C^{poodle}_{image} are activated more when poodles are present while another subset on $C^{chihuahua}_{image}$ responds strongly at the sight of chihuahuas. Therefore, C^{cat}_{image} actually represents the abstract concept of different cat breeds. If we look at neighboring circuits of C^{cat}_{image}, we may find C^{dog}_{image} that represents the concept of what an average dog looks like. C^{dog}_{image} may share some common circuits (features) as C^{cat}_{image}, but the anchor neuron of C^{dog}_{image} gets activated if and only if dogs are present.

If we follow the nerves from the hair cells in the ears, we find that C^{cat}_{sound} develops the concept of what a cat sound like. If we do the same tracing when a word is spoken or shown, we can identify circuit C^{cat}_{word} and $C^{猫}_{word}$ that are activated for the word cat or 猫. If we follow the axons originated from C^{cat}_{image}, C^{cat}_{sound}, C^{cat}_{word}, and $C^{猫}_{word}$ multi-layers down the network, we find an interesting circuit C^{cat}_{nlu} which is activated when a subject see a cat image, hear a cat sound, or the word cat or 猫 is spoken or shown. So, circuit C^{cat}_{nlu} develops some truly abstract concept about cat and the capability of using symbolic language. If we continue our tracing, we find out that C^{cat}_{nlu} is also connected to C^{cat}_{nlg}. When the anchoring neuron of C^{cat}_{nlu} is activated, we see a specific signal traveling down to C^{cat}_{nlg}. Suddenly, we hear the subject utters the magic word cat.

The behavior will emerge spontaneously if a neural network is sufficiently large and deep. No one has to design it. You might ask, why haven't I heard a dog or a chimpanzee utters the magic word cat yet? Because their brains don't have enough space for more layers. The genetic differences between humans and our closest relatives are very small. Most of the genetic differences are useless random mutations. The magic mutation(s) that makes us human is most likely a mutation that removes the constraint(s) on our brain capacity. It doesn't really give us superpower. It just sets free the neural network that's been trapped inside our skull for millions of years.

Some philosophers like to talk about the concept of "qualia", defined as individual instances of subjective and conscious experience. The commonly used example is the "redness" of red. Is the redness experienced in your mind the same as what's experienced in my mind when we both stare at a red block? A unhinged debate will quickly turn into a game of words. We can avoid it by having an operational definition of a subjective and conscious experience.

Imagine that we attach a probe to every neuron in a brain and record the state of the whole brain $\psi(NN, t)$. An individual's experience in response to a sensory signal $\psi(s, t_0)$ is the difference

$\Delta\psi(NN)$ between the new state of the brain after incorporating $\psi(s, t_0)$ and the state of the brain if it evolves without $\psi(s, t_0)$. We can limit $\Delta\psi(NN)$ in a specific region of the brain and in a finite time window after t_0. Using this definition, an individual's experience should be similar but not identical in response to the same signal. For instance, if we keep flashing the same red block to a subject's eye, some regions of the brain will have very similar responses, but other parts will get desensitized or even annoyed. The subjective experience also depends on the accumulated state of $\psi(NN, t)$ up to t_0 and other sensory signals received during the same period. For instance, the redness experienced by a subject while enjoying a sunset could be quite different from those experienced when strapped onto a MRI machine in a lab.

An interesting problem is how to compare the experience between two different brains, NN_a and NN_b. Let us assign each neuron a layer number, defined as the smallest number of the connections from any sensory neurons to the neuron. For example, all photoreceptors on the retina are on layer 0, those directly to them are on layer 1, and so no. We can use the layer numbers to break cycles and sort NN_a and NN_b topologically. After NN_a and NN_b are sorted, we can map similar neurons between two graphs and do a pairwise comparison of $\Delta\psi(NN_a)$ and $\Delta\psi(NN_b)$. The signal propagation along the main paths should be similar but there will be subtle difference due to the different wirings between NN_a and NN_b. Generally speaking, the concepts (features) experienced on lower layers should be very close and those on higher layers may diverge more.

The circuits at a much higher layer like those in the human brain correspond to abstract concepts used in a symbolic language. The circuits, the embodiment of such concepts, are defined by their connections to other circuits. It's why philosophers inevitably end with playing the game of words. A symbolic language is a network of interconnected symbols (words). It is a living and evolving organism, just like a network of neurons. The concepts abstracted by neural circuits will develop a life of its own. For instance, $F = ma$ and "*evolution by natural selection*" survive long after the atoms of the circuits conceiving them are returned to the Earth.

To Be, or Not To Be

Even the founders of quantum mechanics are puzzled by the apparent paradox of superposition. In a thought experiment devised by Schrödinger, a cat is simultaneously alive and dead inside a box. The fate of the cat can be decided by a few elementary particles and become either alive or dead when the box is opened. According to our model, the state of the mind is the superposition of different states. So, the mind should behave like Schrödinger's cat locked in a skull. To be, or not to be, that will be the perpetual question for the mind.

There is a wide wide spectrum of different minds in terms of decision making. Some suffer analysis paralysis and can never make up their minds. Others are very stubborn and never change their minds. Is there a way to go beyond anecdotal evidence and prove the superposition of our minds? We could recruit volunteers and do some experiments in a lab. Such an approach

might be considered as cargo cult science by Feynman. If a subject is aware of being observed, the state of his mind is already entangled with the observer. We wouldn't get any meaningful results. To prove that the state of the mind is undecided until a decision or action is observed, we need many undisturbed minds. We must demonstrate that we can alter the state of their minds at will with a few photons (the equivalent of opening Schrödinger's box).

Fortunately, such experiments are conducted at scale in the age of the Internet. Whenever someone is online, some pCTR models somewhere predict the Click-Through-Rate (pCTR) for all impressions that the user would have. They decide the placement of ads, posts, videos, and so on. This is done on billions of users trillions of times every single day. These pCTR models are very accurate in predicting the state of the user's minds, e.g. whether they will click specific ads, how long they will watch a video, how likely they will engage with a post or a recommendation, and so on. Most users are unconscious of it and their minds are largely undisturbed.

Let us do a thought experiment. Suppose that the pCTR model predicts the probability of a user clicking an ads is very high. By monitoring the user's mouse movement, we're pretty sure that the click will happen in half a second. If we had the superpower to freeze time and peek into the user's brain, we would see a wave packet forming and being sent to the muscle fibers to execute the action. The state of the user's mind would be leaked out of the box imminently. Imagine what would happen if we flash a few photons to the user's eyes and reveal the current state of his mind:

> *We're 99.99% sure you're gonna click this ads in half a second, sucker!*

Most likely, the user would cancel the click. If we keep playing the trick on the same user, the outcome becomes very unpredictable again. To click, or not to click, that's indeed the perpetual question in everyone's mind.

The tech companies wouldn't run our thought experiment. They do run similar mind-altering experiments at scale to improve ads-clicking and user engagement. Such experiments usually have carefully designed control and accurately measure how effective different tricks are. Some may attest that they never click ads or their mind cannot be altered so easily, but the evidence is overwhelmly the opposite. People's minds are wide open on what to click, what to watch, what to purchase, where to visit, and even who to date and marry. There is a significant difference between what people think they will do and what they actually do.

Next time when Schrödinger is pondering the fate of his cat, we can assure him that the universe behaves no more weird than the human mind. No one, including a subject himself, can tell the exact state of a mind until it's revealed to the outside world.

Free Will and Awareness

When people talk about free will, sometimes what they really mean is the unpredictability of a mind. As just discussed, the uncertainty of a mind is due to the superposition of different states, at

the most fundamental level, the probabilistic nature of the universe. The “will” of an unpredictable mind is not really “free” because the mind itself doesn’t really know its exact state. It doesn’t have any control over what’s going to happen next. In this section, we will discuss the phenomenon that a mind, at least our human mind, can take voluntary actions. It’s aware of its surroundings. It can introspect its own behavior and make sense of the world.

Most people can dance or move their body in sync with music no matter whether they have dance talents or not. Dancing to music is a complicated and supermassive operation at cell level. A subject has to “understand” the music, “feel” the rhythm, “coordinate” the movement of trillions of voluntary muscle cells in different parts of the body while “enjoy” the process. No one would doubt that it is a conscious act carried out by the subject’s free will. We can have a better understanding of “free will” or “consciousness" by nailing down who has done what during sync-dancing.

Fig 4.1 Signal Processing of Dancing to Music

From an alien physicist’s point of view, there are some sound waves originating from a point in space and some interesting body movements are observed at another (Fig 4.1). There seems to be a correlation between these two events. The alien physicist zooms in a bit. He finds out that the sound wave is converted to neural signal $\psi(s,\ t_s)$ by the hair cells inside the subject’s ears, and the body movements are caused by neural signal $\psi(m,\ t_m)$ sent to various muscle cells of the body.

To his surprise, the movements of the body just mimic the movements of hair cells[8]. There are a few hundred milli-seconds of time shift between $\psi(s,\ t_s)$ and $\psi(m,\ t_m)$ due to the time required for the signals to propagate from hair cells to muscle cells. A few minor waveforms are added onto $\psi(s,\ t_s)$ by various circuits along signal propagation paths. Besides the modulation, the shape of $\psi(m,\ t_m)$ is almost identical to $\psi(s,\ t_s)$. When $\psi(s,\ t_s)$ originated from hair cells is really strong, some of its energy overspill to side circuits along the paths, causing the smile on the subject’s face and/or the sing-along from the subject’s vocal cords. The main task carried by most neurons is to amplify $\psi(s,\ t_s)$ and relay them to trillions of muscle cells. The wiring of these neurons is not being done on the fly by any “conscious” parts of the brain. Where and how does “free will” get involved in the operation?

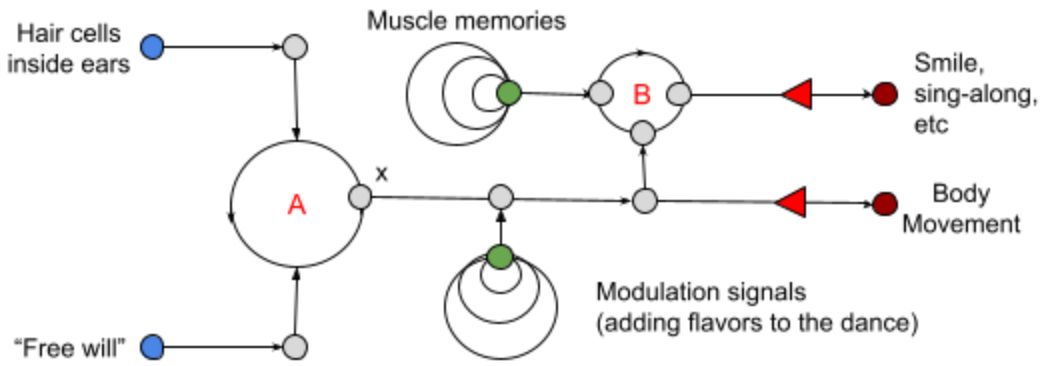

Fig 4.2 Sync-Dancing Circuit

The alien scientist identifies a critical circuit A connecting hair cells to muscle cells (Fig4.2). Usually, the signals from hair cells are very weak when they reach neuron x on circuit A. They won't cause any body movements even if the paths connecting x to muscle cells can greatly amplify any signals arriving at x. When a subject is awake, there is generally a high energy level (neuron firing) maintained in certain regions of the brain. Some random wave trains constantly form and direct the subject's attention to different areas. If a wave train happens to circle around circuit A when the signals from hair cells arrive, the modulated signals at x will be strong enough to be amplified millions of times and reach trillions of muscles. It's a phenomenon that we're all familiar with. The sound of a radio broadcaster won't travel too far no matter how loud he tries to shout. But, if the signal of his sound wave is added onto a train of radio waves, the high energy radio signal can carry the sound of the broadcaster many miles and reach numerous listeners.

We could call the specific region "conscious layer" and the wave train turning on/off various actions "free will". But, there is nothing special about them. The "conscious layer" is made of ordinary neural circuits. The "free will" behaves exactly like the propagation of other signals. Both have no idea of how various responses are actually carried out. The circuit A is not special, either. Many similar circuits are situated in different regions of the brain, as shown in Figure 4.2 B. They make their own decisions on whether and how to react to different inputs. There is not a single special command and control center in the brain.

The inner workings of a brain is the same as how human society works in the US. As the impacts of COVID-19 propagates through different communities, individuals react to inputs they receive. Those who hold political offices or run major corporations may have some powers to make certain decisions, but they generally have no ideas how the decisions are carried out. Most of the time, they're as powerless as others to control or even know how others would react. Because individuals process different versions of the reality, they may form different opinions on what's going on and how to react. One half of the nation is always at war with the other. Some may expect that such a chaotic system would have collapsed without centralized coordinations. Miraculously, most individuals and businesses adopted and carried on. The fragile healthcare system managed to hold up and deliver two vaccines. It's messy and far from "ideal", but it's just how neural computation works.

We might call the chaotic nature of the system the “consciousness” of the society, and individualism “free will”. But, it’s wrong to pinpoint any specific individuals or regions as the “will” or “consciousness” of the society. For the same reason, it’s futile to look for the center of “consciousness” in our brain and keep debating “free will”. For those who cannot let go the idea that they have “free will”, try to publish a paper or even write a private diary without using the, of, or and in defiance of Zipf’s distribution. You will find the words coming out of your “free will” are like charged particles passing through a magnetic field. Your “free will” will get bent according to each particle’s charges. If your writings are long and diverse enough, they inevitably follow a certain distribution.

One aspect of our consciousness is the state of being awake and aware of our surroundings. The capability of understanding and reacting sensory inputs is critical to the survival of any species. Awareness alone is not a unique human trait. For instance, when a dog wakes up, it must be aware and able to process the signals from its surroundings. The wake/sleep cycle is related to the rise and fall of energy levels in different regions of a brain. It behaves exactly like the daily tides caused by gravitational pull. When a subject’s awake, the energy (attention) is pulled toward the regions processing external sensory inputs. When a subject falls asleep, the energy is dispersed to other regions. The nightly flood opens up paths unavailable during the day. A subject may experience dreams, a world largely independent of direct sensory inputs.

The unique thing about our awareness is its tendency to offer a continuous narrative of what’s going on in the world. When sensory inputs propagate through the network, they excite neural circuits on lower layers as well those on higher layers. The activation of neural circuits on the layers handling symbolic language is the stream of thoughts and reasonings going through our mind. It just cannot stop inspecting its own behavior and making sense of the world. If the Pavlov’s dogs were not constrained, they might go and greet the assistants on hearing their footsteps. A human mind watching it would give a personified description of what’s going in the dog’s “mind”.

Can we prove that sense-making and decision-making are two separate processes happening inside the same brain? We can potty train a kid or a dog. After a while, they will do the right thing when they feel the urge to pee. The kid may communicate his intent before taking the action. If we ask what’s happening, he may explain why and how he behaves in a certain way. The dog cannot make sense in the form of symbolic language. Nevertheless, it does make the right decision and carry out the action without any problem. It suggests that sense-making is not a prerequisite for decision-making and execution.

The human brain consists of two hemispheres connected by a thick band of neural fibers known as the corpus callosum. Some patients who suffered from epilepsy had undergone surgery that severed the corpus callosum. The two hemispheres in these split brain patients cannot communicate with each other. In a study done by Michael S. Gazzaniga, etc[7], two hemispheres of the split brain patients were given different inputs and asked to make certain choices. The

study kept the inputs given to one hemisphere away from the other half. They find out that each hemisphere can process its inputs and make correct decisions. When asked why the other half made certain choices, one hemisphere just made up a story even if it didn't have any information about the other half's decision making. This shows that the sense-making in a brain can happen independent of the actual decision-making. We shall not place too much trust on the narratives provided by our own brain.

Collective Minds

Each hemisphere of a brain can make its own decisions and have its own thoughts. If two hemispheres are connected, they become a bigger neural computer. It's expected that, the more connections are added between subnets, the more tightly coupled they become. Eventually, it behaves like one coherent entity and has its own life. Connecting lots of brains should create a super neural computer. It's actually a working-in-progress pursued by the human race in the past few hundreds of years. The development of the ultimate super neural network including all people on the planet has been greatly accelerated by the invention of the Internet, but it's still in its infancy. Here we will look into a more developed subnet, the financial market, and examine the behavior of collective minds.

Our understanding of quantum mechanics comes from some alien physicists who study physical systems from a great distance. Every time when they examine a subject, they have to destroy it, the equivalent of hijacking and killing the subject in order to check their gender or net worth when studying the financial market. Let us first review a few phenomena that alien physicists might deem as quantum weirdness but are considered as common sense to the general public.

A big public company could have tens of thousands ongoing projects simultaneously. The value of the company is the sum of all present and future projects. When you buy one share of a company, the return of your share is the superposition of the return of all projects carried out by the company. When you sell your share, you realize only one possibility at a specific moment. The alien physicists would use fancy terms such as "superposition" and "quantum decoherence", but they are simple concepts in the world of investing. What economists get wrong is that they add up the expected values of all projects using real probability instead of probability amplitude. Their models would miss the critical timing and interference between different projects.

Suppose that our alien physicists are equipped with a special detector and they can observe the gender (spin) of a subject (particle) and an obnoxious number on the bottom line of their balance sheet. There are no particular interesting patterns between genders and the numbers except one mysterious phenomenon. If they follow a pair of subjects coming out of a special place, they will find one male and one female. The sum of the two subject's bottom line numbers are entangled in a mysterious way. If they destroy one subject, the other always gets the sum. The alien physicists would call the mysterious phenomenon "quantum entanglement", but even a layman can explain to Einstein that there is no "spooky action at a distance" and the special place is just a wedding

chapel. The financial market is quantum mechanics. It can teach both our alien physicists and the general public the inner workings of our mind (and universe).

When a subject buys a stock, it has to boil down to the activation of neural cluster C^X_{buy}, where X represents a stock symbol and buy the neurons carrying out the buy action. As long as the subject is a participant in the market, his or her C^X_{buy} will synchronize with everyone's C^X_{buy}, just like C^{word}_{nlg} responsible for outputting words in a language system. How often a stock gets bought depends on how often the stock comes across the buy region in a subject's mind. The market cap of different companies is ultimately determined by how often their stocks excite the collective minds. Therefore, the market caps of different companies should also follow similar Zipf's distribution, just as the word frequency of various language systems.

Buying and selling are carried out by separate circuits: C^X_{buy} and C^X_{sell}. We can model the buying and selling behavior of the collective minds by two probability amplitudes: $\psi(buy,\ t)$ and $\psi(sell,\ t)$. Let T_0 denote the minimum transaction interval, nT_0 a trading period. The absolute square of the coefficient b_n and s_n corresponds to the money (energy) devoted to the trading period nT_0. For instance, an individual who sets aside some money for day trading and others for passive investing in retirement accounts has n_dT_0 and n_rT_0 in the order of days and years (or even decades), respectively. An individual's b_n doesn't always match s_n, especially for long-term investment. Each individual has its own set of b_n and s_n, and a market of collective minds just sums up all individual b_n and s_n (initial phases omitted for simplicity).

(4.1) $$\psi(buy,\ t)\ = \sum_{n=1}^{\infty} b_n e^{i2\pi/(nT_0)}$$

(4.2) $$\psi(sell,\ t)\ = \sum_{n=1}^{\infty} s_n e^{i2\pi/(nT_0)}$$

The buying and selling pressure for a given period of $\Delta = n_\Delta T_0$ is defined as the total probability (energy) of buying and selling during the period, as shown in (4.3) and (4.4). The velocity of market price movement is caused by the imbalance of buying and selling pressure (4.5).

(4.3) $$P_\Delta(buy,\ t) = \int_t^{t+\Delta} |\psi(buy,\ \tau)|^2 d\tau$$

(4.4) $$P_\Delta(sell,\ t) = \int_t^{t+\Delta} |\psi(sell,\ \tau)|^2 d\tau$$

(4.5) $$V_\Delta(t) = P_\Delta(buy,\ t)\ /\ P_\Delta(sell,\ t)$$

Suppose that the period of $\Delta = n_\Delta T_0$ equals one day. Let's look at how different components of the market affect the daily price movements. For $n \ll n_\Delta$, namely, the money traded more often than once per day, the contribution of their probability amplitudes are actually the same. It may

feel counterintuitive, but it's actually true if you think carefully. The HFT traders who are in and out the market millions of times a day have a big impact on price volatility in a short time window, but they have little impact on the daily price movements because they're already in the baseline. For $n \gg n_\Delta$, namely, the money with a time horizon much longer than one day, their contribution drifts slowly but doesn't fluctuate a lot. In a first-order approximation, they add a constant term to the velocity of price movement $V_\Delta(t)$ and doesn't cause big acceleration of acceleration of $V_\Delta(t)$. Most fluctuations in daily price movements $V_\Delta(t)$ are mainly caused by the money in and out of the market within a time period in the vicinity of $\Delta = n_\Delta T_0$. The fluctuations are largely periodical and may be mistakenly modeled as random distribution. However, the cycles of long term money may coincide with those of short term money once in a while. The constructive/destructive interference may cause huge price dislocations. Some in Wall Street may treat such events as outliers but it's an expected behavior of a network of collective minds.

The above analysis should hold true for different $\Delta = n_\Delta T_0$ as long as n_Δ doesn't approach the extreme ends of the spectrum ($n_\Delta \to 1$ or $n_\Delta \to lifetime$). In other words, the price movements of the market are fractal in nature. An observer won't be able to tell whether a chart of normalized price movements is hourly, daily, weekly, monthly, or yearly as long as the distribution of b_n and s_n are largely smooth for the time window. The total energy under the curve of $\psi(buy,\ t)$ and $\psi(sell,\ t)$ should be the same. However, the energy distribution under $\psi(sell,\ t)$ will be more concentrated on the shorter end than $\psi(buy,\ t)$ because the human brain responds to negative feedback faster and more acutely than positive feedback. As a consequence, a market correction usually happens in a much shorter time window than the corresponding market rally. Benoit Mandelbrot noticed both phenomena[5] . They result from a network of collective minds running neural computation.

The evolution of financial markets can be modeled as the energy (money, wealth) tracing all possible paths in the configuration space of a network of collective minds. The principle of least action dictates that most of the energy will travel through the least action (classical) path(s). Therefore, it will be very hard to find the money deviating from the classical path(s), namely, the overall market. It's analogous to the phenomenon of a beam of light traveling through physical space. If we zoom out, we find almost all photons travel along the classical path(s), namely, the paths minimizing action (time). If we zoom in to the order of the wavelength of the photons, we will find lots of photons around the classical path. They are exploring alternative paths just like genetic mutations in the evolution of species. If we could follow individual photons, we would find some of them explore arbitrary and less travelled paths. The same process happens in the financial market and leads to the destined paths. Those who examine the behavior of collective minds see the result of the principle of least action and propound the Efficient Market Hypothesis (EMH)[6] . Those who focus on the behavior of individual minds see the evolution process and come up with the theory of behavioral finance. They are the two aspects of the same physical and computational process.

Gödel's Incompleteness

The neural computation running inside our brain imposes some inherent limitations on our capability of thinking and reasoning. Let us look at an example called Liar's Paradox. If you're given the following two statements, someone asks you "*Is A true or false?*". What would be your answer?

$A:$ *This statement is true*
$B:$ *The other statement is false*

To understand why our brain cannot resolve such a simple question, we need to examine how the energy flows through relevant circuits. On hearing or reading "*A: This statement is true*", the train of thoughts (neuron firings) will excite circuit C^{A}_{nlu}, C^{A+is}_{nlu}, and $C^{A+is+true}_{nlu}$ sequentially. Such a statement leaves some energy circling around these circuits. If we immediately ask the brain "*Is A true or false?*" , the question will excite C^{A}_{nlu}, C^{A+is}_{nlu}, $C^{A+is+true}_{nlu}$, and $C^{A+is+false}_{nlu}$. Because there are some extra energy lingering around $C^{A+is+true}_{nlu}$ deposited by the previous statement (A), the brain will favor $C^{A+is+true}_{nlu}$ over $C^{A+is+false}_{nlu}$. However, if the brain is also given "*B: The other statement is true*", the train of thoughts will travel through $C^{The\ other}_{nlu}$, C^{A}_{nlu}, C^{A+is}_{nlu}, and $C^{A+is+false}_{nlu}$. If we ask the brain "*Is A true or false?*" after statement B, the energy level on $C^{A+is+true}_{nlu}$ over $C^{A+is+false}_{nlu}$ will be the same and the brain will find it hard to make a choice, as shown in Fig 5.0.

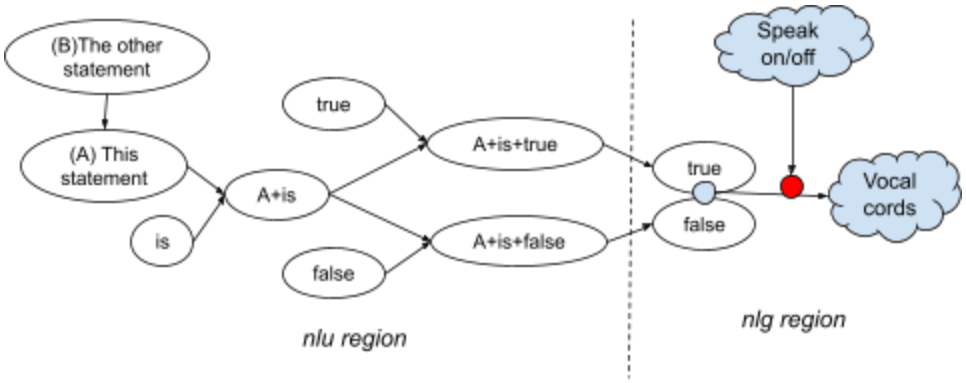

Fig 5.0 Liar's Paradox Circuit

Kurt Gödel uses similar approaches to demonstrate the inherent limitations of every formal axiomatic system. Now we know that the limitations are actually due to how our brain operates. Let us imagine an alternate universe where the fundamental laws of the universe are almost identical to ours. In the alternate universe, the signal propagation follows the same Feynman's rule. The probabilistic interpretation of quantum mechanics is also the same with one additional *Recency Rule*: the newly updated circuits always take precedence in case of a tie. The Gödels in the alternate universe would never discover anything like Liar's Paradox.

Even in our universe, the brain can easily resolve the issue if the energy from some circuits helps to tip the balance between $C_{nlu}^{A+is+true}$ and $C_{nlu}^{A+is+false}$. For instance, if a brain is told "*Statement A (or B) takes precedence in case of a tie*", it will have no difficulty in picking an answer. You may think that we're just replacing Gödel's beautiful logical argument with a clumsy idea of probability and energy in physics. But, even for a logician/mathematician, the reasoning is still a physical process that requires both energy and time. We can prove it by running a quick experiment on our own mind. Please read the following statements line by line, and **immediately** answer the question at the end!

> *this statement is false*
> *B is the other statement*
> *both statements could be true*
> *A is this statement*
> *the other statement is true*
> *is A true or false*

If pressed to give an answer in a few seconds, most people would say "*A (this statement) is false*". However, if they're allowed to think long and hard, some circuits in their brain start to deposit energy to $C_{nlu}^{A+is+true}$. After a while, they will end up with the same dilemma as Liar's Paradox.

In our analysis so far, it seems that we can draw a connection between two arbitrary concepts (words). You can prove that's not how our brain actually operates. Please open a book, point to a random word, close your eyes, and see what words come into your mind. You may know tens of thousands of words, but only a few will come into your mind. There may be some physical connections between two random circuits on the same layer. Only a few resonate well with each other.

If two signals with similar frequencies travel over the same network, the combined signal may generate an oscillation with a slowly pulsating intensity. The frequency of the oscillation differs from the original signals and may excite a circuit for a very different concept. For instance, if you think about a pair of words, e.g. male and female, up and down, buy and sell, and so on, you're thinking of something that resonates among all pairs but not among individual words. The creation of new concepts are not limited to a pair of words or antonyms. It could be any patterns resulting from the modulation of a sequence of related signals. Such phenomena can be observed across different languages.

The resonating requirement suggests that not all concepts (words) are created equal. Some words, e.g. "the", "and", "of", etc, are mainly fillers modulating with other words so that the combined signal can resonate (connect) with the next phrases. Other words, e.g. "god", "love", "dream", etc, are like small prime numbers. They can resonate well with many circuits and have

great healing power. As more concepts are abstracted on the same layer, it requires more and bigger circuits. The subject has to think long (time) and hard (energy) when involving such more abstract concepts. One solution is to move the higher level of abstractions to a different layer (a specialized region) so that they can be handled by smaller circuits. Either way, the capacity of a brain puts an ultimate physical limitation on what a mind can imagine.

The Meaning of Life

In the pursuit of a basic understanding of my own consciousness, I've destroyed everything that I once held dearly in my life: emotion, dream, love, free will, and self-awareness. It feels too cold to live a life full of illusions. When the temperature drops outside, I see water molecules cuddling together and forming these beautiful snowflakes. I see the colors of the rainbow reflected by their ephemeral existence. I think to myself what a beautiful sight! It's a miracle for a snowflake to even contemplate the meaning of life.

After a day of mundane life, it's become a pastime for me to watch Feynman's lectures online. His path integral formulation of quantum mechanics gives me an intuitive understanding of the inner workings of the universe. Feynman once remarked after reflecting on the intellectual journey of developing his theory:

So what happened to the old theory that I fell in love with as a youth? Well, I would say it's become an old lady that has very little attractive left in her and the young today will not have their hearts pound anymore when they look at her. But, we can say the best we can for any old woman, that she has been a very good mother and she has given birth to some very good children.

Well, at least one young today looks at the old lady. His heart starts to pound and his mind cannot help but ponder the secret of life and nature. Hope the sparklings from one tiny snowflake add a bit of color to your understanding of life and nature, too.

References

1. Feynman, Richard. P. (1948). Space-Time Approach to Non-Relativistic Quantum Mechanics. *Reviews of Modern Physics*. **20** (2): 367–387.
2. Feynman, Richard. P. (2005) [1942/1948]. Brown, L. M (ed.). *Feynman's Thesis — A New Approach to Quantum Theory*. World Scientific.
3. Dirac, Paul A. M. (1933). The Lagrangian in Quantum Mechanics. *Physikalische Zeitschrift der Sowjetunion*. **3**: 64–72.
4. Herculano-Houzel, S (November 2009). The human brain in numbers: a linearly scaled-up primate brain. *Frontiers in Human Neuroscience*. **3**: 31.

5. Mandelbrot, Benoît B. (2004). *The (Mis)Behavior of Markets: A Fractal View of Risk, Ruin, and Reward*. New York: Basic Books.
6. Fama, Eugene. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance*. **25** (2): 383–417.
7. Gazzaniga, Michael S. (1998). The Split Brain Revisited. *Scientific American*, July 1998.
8. Ashmore, Jonathan. (1987). Dancing Outer Hair Cell. *Ear We Go*, BBC on Aug 13, 1987.
9. Piantadosi, Steven (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychon Bull Rev*. **21** (5): 1112–1130.
10. Folli, M. Leonetti, and G. Ruocco. (2017) On the maximum storage capacity of the Hopfield model. *Frontiers in Computational Neuroscience*, 10(144).