Neural Coding and Information Theory

Prasanna Natarajan
1410110298
Computer Science, B. Tech
Shiv Nadar University
pn337@snu.edu.in

Vedant Chakravarthy
1410110489
Computer Science, B. Tech
Shiv Nadar University
vc909@snu.edu.in

Abstract—This goal of this paper is to do a literature review in the intersection of the fields of neural information transfer and information theory – Neural Coding and Information Theory. This paper looks at three papers and pursues this topic with a focus on the Information Theory-related concepts.

I. INTRODUCTION

The human brain processes motor and sensory information in multiple stages. At each stage, the neural representation of the motor commands or stimulus features are manipulated. Information theory is used to measure stimulus-response models, encoding efficiency, frequency filtering and other metrics relevant to information transfer between neurons. Further, it is used to understand the predisposition of neurons in selectively or preferentially encoding information.

This paper first introduces the foundational ideas that link information theory to neural coding through [2]. It then goes on to [3] and discuss different models for the estimation of entropy in neural spike trains. The paper then reviews [4] to explore mechanisms allowing information filtering that is seen in neuronal systems and discusses in detail the paradigm of using frequency-resolved measures as opposed to measuring signals on the whole. Finally, this paper discusses and aggregates some key ideas that have been brought up in the three papers.

II. INFORMATION THEORY AND NEURAL CODING

Borst, Alexander, and Frédéric E. Theunissen

A. Introduction

Information transmission between neurons is done by spike trains (trains of action potentials) or graded membrane potential shifts. Spike trains are more frequent information transmission mechanism than graded membrane potential shifts. Information transmission is done using a neural code, a representation of information in a neural system. In this context, there are mainly three questions that arise. Broadly, they are (1) what is being encoded? (2) how is it being encoded? and (3) with what precision is it encoded? To answer these questions, scientists use stimulus-response curve where the x-axis(stimulus) and y-axis (response) answer the first two questions. The third question is answered by error bars.

It is important to measure the neural reliability[1] in order to understand combinations of the three questions. The attempt to measure neural reliability is done through the rigor offered by Information Theory. The measurement of information transfer precision is done by determining the exact probability distribution of outputs for some given input signal. Further, ideas of the maximum information transferrable in a communication channel (here, neuronal network) are relevant. The paper argues that precise quantification of the reliability of stimulus-response is crucial to understand what information is being encoded and how.

An example of the understanding of how information is transmitted in neuronal systems is the idea of information filtering. Neurons that contain uncorrelated Gaussian background noise and white input signal are seen as low-pass filters of information since they encode information about slower parts of stimulus. These calculations are made by estimating the lower bound of information rate using simulation-response coherence. Coherence can be calculated in terms of input (stimulus) and output (spike trains) in frequency domain.

Information theory helps with its quantifying ability in calculating the maximum information that could be transmitted as a function of firing rate which is then compared to actual information transfer as a measure of coding efficiency. It also helps in quantitative evaluation of a model's quality as the theoretical limits can be obtained using information theory. It has also helped researchers to determine the 'limiting spike timing precision' used in encoding.

Information theory is thus used to determine how much information about stimulus parameter values is contained in neural responses. This gives the statistical significance in terms of the variation of responses and stimuli.

B. general concepts

Probability that the neural response takes the value
$$r_i = p(r_i)$$
Probability that the stimulus condition takes the value
$$s_i = p(s_i)$$
Probability that the response is r_i given the stimulus is
$$s_i = p(r_i | s_i)$$

Information about stimulus condition sx

$$I(R, sx) = \Sigma p(r_i | s_x) \, log_2 \frac{p(r_i | s_x)}{p(r_i)}$$

Average information obtained from all stimulus conditions

$$I(R,S) = \sum\sum p(s_j)p(r_i|s_j)log_2 \frac{p(r_i|s_j)}{p(r_i)}$$

Some basic formulas

Plotting I (R, S_X) as a function of stimulus condition X allows us to replace the traditional stimulus–response curve with a stimulus–information curve that shows how well an ideal observer could discriminate between the stimulus conditions based on a single response trail

1. Calculation of Maximal rates of information transfer:

H(R) is the entropy and thus the number of bits required to uniquely represent all the response values. It also represents the maximal information that can be carried by the neuron. H(S) is the entropy and thus the number of bits required to uniquely specify all stimuli conditions. H(R|S), the entropy of R(response) given that the conditions of the S(stimuli) are known is called neural noise. H(S|R) is called stimulus equivocation. It is the entropy remaining in the stimulus given that the response is known.

The entropy of the stimulus H(S) represents the maximum information that could be encoded, from which the stimulus equivocation H(S|R) is lost and H(R|S) is gained. The information about the stimulus preserved in the neural response (mutual information) is

I(R, S) = H(S) – H(S|R)

The comparison of H(R|S) to H(S) gives some estimate of the efficiency of neural code, however, since the H(R) that is actually found in an experiment is affected by the stimuli in question, it becomes meaningful to define a maximal response entropy.

Since,

I(R, S) = H(R) – H(R|S), H(S) – H(S|R) = H(R) – H(R|S).

From this arrangement encoding efficiency of stimulus conditions can be determined. For a H(S), H(R|S) could be small in comparison to H(R): this indicates that neuronal efficiency is high, since, knowing the stimulus, there is a low uncertainty about the response. In this situation, if H(S|R) is large relative to H(S), this implies that the stimulus is being inefficiently encoded.

Bayes Theorem

$$p(s,r) = p(s|r).p(r)$$

Joint entropy of R and S

$$H(R,S) = -\sum\sum p(r_j,s_i)\, log_2 p(s_i,r_i)$$

Average information obtained from all stimulus conditions

$$H(R|S) = \sum p(s_j)\sum p(r_i|s_j)log_2 p(r_i|s_j)$$

Equivalent forms for average information:
$$I(R,S) = H(R) - H(R|S)$$
$$= H(S) - H(S|R)$$

$$= H(R) + H(S) - H(R,S)$$

A practical limitation to finding maximal values of information transfer is that this calculation depends on the stimulus conditions and the choice of the stimuli, all of which are very specific to a particular experiment and are chosen on a case-to-case basis. In an attempt to make a minimal amount of assumptions on the conditions around the parameters of the experiment, another approach for measuring information and representing information in these systems is using a lower bound on the information transfer. In this connection, the data processing inequality is relevant. Data processing inequality states that,

if $X \rightarrow Y \rightarrow Z$, then I (X; Y) ≥ I (X; Z).

That is, if X, Y and Z are in a Markov chain in that order then the given inequality holds.

Information theory can be used to determine the parameters for to represent the information being tested.

C. Information theory and dynamic (continuously time-varying) stimuli:

Since the search space for stimulus becomes huge for any neural system with memory, as they depend on the history of the stimulus. Therefore, it can be represented as a vector of parameters, describing all preceding stimulus states relevant to the response. For instance, if a certain parameter can have 7 different values, and the response depends on 6 previous states, then there are 7^7 possibilities that are to be represented. Therefore, estimating the probabilities of stimulus response is not practical.

To characterize information, there are broadly three methods. Namely, direct, upper bound and lower bound calculation. Out of these three the direct method is a theoretical approach. The mutual information I(S;R) can be calculated by using the formula H(R) - H(R|S). Due to dimensionality explosion calculating the joint probability of time varying stimulus and response is practically not feasible, but the responses of a single spiking neuron can be limited to strings of 0s and 1s if the time window that is used to divide the window is really small. Then, spike train noise is determined by repeating a dynamic stimulus and recording various responses to determine the response distribution. Though this method provides correct information measure, it has a lot of limitations, which are: number of dimensions have to be limited by the size of the binary string, it does not indicate which aspect of stimuli are best represented and it depends on the stimulus ensemble. Therefore, to overcome these other methods of calculating mutual information are used that provide upper and lower bounds.
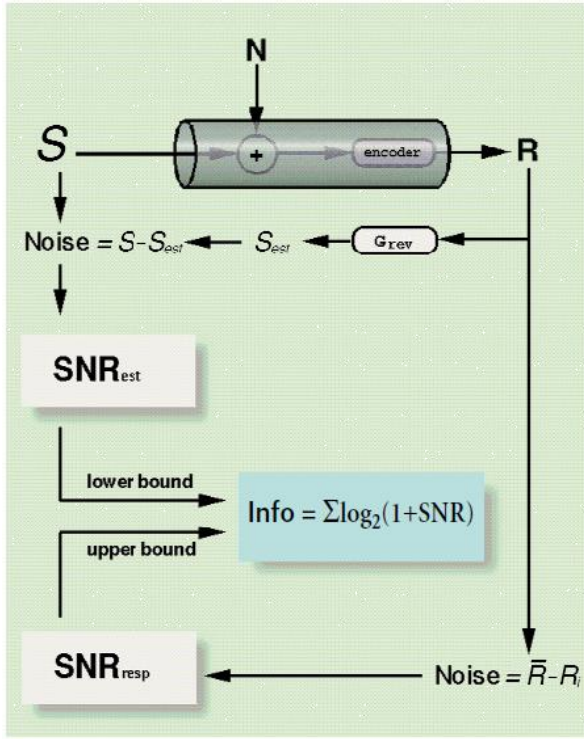
Figure1. Summary diagram for calculation of upper and lower bounds on information transfer. Top, situation where a stimulus S is corrupted by additive noise and subsequently fed through an unknown encoder to result in the response R. The lower bound is obtained with a linear reverse filter operation. The upper bound is obtained directly by comparing average and individual responses.

The diagram above explains the general process involved in estimating upper and lower bound of information transfer.

The table below summarizes the calculation methods and assumptions that are associated. All the methods described in the table below follow the general methodology that is explained above in the diagram

Method of estimation	Driving Principle	Assumptions
Direct	Separate R into deterministic and a random component by repeating S many time $I(S,R) \to I(R,R_{det})$	The temporal resolution is small enough.
Upper Bound	Same as Direct method.	$R_{det} = R_{avg}$ $N = R - R_{avg}$ $I(R,R_{det}) \to I(R,R_{avg})$ It assumes that if Noise(N) is Gaussian then $I(S,R) = \int_0^k \log(1+SNR(f))df$
Lower bound	Find best S_{est} from R. $I(S,R) \to I(S,S_{est})$	S is Gaussian. $N = S - S_{est}.$
Absolute lower	Same as lower bound. Find smallest $I(S',S_{est})$ that would give the same error as $(S - S_{est})$.	S is Gaussian.

II. ENTROPY AND INFORMATION IN NEURAL SPIKE TRAINS: PROGRESS ON THE SAMPLING PROBLEM

Nemenman, Ilya, William Bialek, and Rob de Ruyter van Steveninck

A. Introduction

Classical experimental approaches have focused on mean responses of neurons to relatively simple stimuli. With the help of information theoretic methods, modern approaches can quantify the responses to arbitrarily complex and even fully natural stimuli, taking account of both the mean response and its variability in a rigorous way, independent of detailed modeling assumptions.

The hypothesis that this paper is tries to prove is neural code adapts to the distribution of sensory inputs, optimizing the rate or efficiency of information transmission.

This can be verified with the help of measurements of entropy and information in spike trains.

A problem with such measurements in that entropy and information explicitly depends on the full distribution of neural responses and only a limited sample is provided by experiments. We need to know the distribution of responses to each stimulus in our ensemble. Also, the number of samples from the distribution is limited by the ability to repeat and same stimuli and capture the response with the required precision. Natural stimuli generate neural responses of high timing precision and thus the space of meaningful responses itself is very large.

A solution to this problem is to give up the generality of a completely model independent information theoretic approach. That is some explicit model is required to regularize learning of the underlying probability distributions from the experiments. The question is if the generality of the analysis can be maintained by introducing the gentlest of regularizations for the abstract learning problem, or if stronger assumptions about the structure of the neural code itself is to be made.

The paper proposes a simple and abstract Bayesian prior comes close to the objective, that is, to generalize Ma's idea of estimating entropy by counting coincidences to arbitrary distributions. Also, this method works well into a classically

under sampled regime for an experimentally relevant case of neurophysiological recordings.

B. An estiamtion Strategy

1. Notations

N: Sample size or Number of examples

K: possible neural responses

{p_i}: probability distribution over all K possibilities

$S = -\sum (p_i \log(p_i))$: Entropy

n_i: number of times i^{th} possibility occurred.

$f_i = n_i/N$ - Frequency of occurrence of the ith possibility.

$S_{naive} = -\sum (f_i \log(f_i))$: Naive Entropy

S_{ML}: Maximum likelihood estimator

2. Maximum likelihood estimation of entropy

S_{ML} is S_{naive} because the maximum likelihood estimate of probabilities is given by the frequencies.

It is already established that S_{ML} underestimates the entropy [3]. With good sampling rate a term $(K-1)/2*N$ should be used to correct ML estimate [3]. In practice, a lot of responses from the total possible number of responses (called bins from here on) may have truly zero probability, for example, as a result of refractoriness (the insensitivity to further immediate stimulation that develops in irritable and especially nervous tissue as a result of intense or prolonged stimulation [5]), and the samples from the distribution might not be completely independent. Then S_{ML} still deviates from the correct answer by a term ~1/N, but the coefficient is no longer known a priori.

Given these conditions the options are to heuristically verify and extrapolate the 1/N behavior from subsets of the available data [3] or still agreeing on the 1/N correction, a coefficient of correction can be calculated, which would represent the number of effective bins for some classes of distributions [3]. All of these approaches, however, work only when the sampling errors are in some sense a small perturbation.

1. S_{ML} as a limiting case of Bayesian estimation

For practical purposes, that is, when N does not tend to infinity, we need an estimator that does not have a perturbative expansion in 1/N with S_{ML} as the zeroth order term. We know that S_{ML} is a limiting case of Bayesian estimation with Dirichlet priors. Formally, we consider that the probability distributions p = {p_i} are themselves drawn from a distribution $P_\beta(p)$ of the form

$$\mathcal{P}_\beta(\mathbf{p}) = \frac{1}{Z(\beta;K)}\left[\prod_{i=1}^{K}p_i^{(\beta-1)}\right]\delta\left(\sum_{i=1}^{K}p_i - 1\right).$$

where

1. δ function enforces normalization of distribution p
2. partition function Z(β;K) normalizes the prior $P_\beta(p)$.

Maximum likelihood estimation is Bayesian estimation with this prior in the limit $\beta \to 0$, while the natural "uniform" prior is $\beta =1$.

2. Sample dependent bias in entropy estimation

While the priors are quite smooth on the space of p, the distributions drawn at random from P_β all have very similar entropies, with a variance that vanishes as K increases.

Basically, this is the origin of the sample size dependent bias in entropy estimation, therefore the paper proposes to correct the bias at its source.

Therefore, we need to construct a prior on the space of probability distributions which generates a nearly uniform distribution of entropies. This can be achieved by averaging over β because the entropy of the distribution chosen from P_β is sharply defined and monotonically dependent on the parameter β.

$$\mathcal{P}_{NSB}(\mathbf{p}) \propto \int d\beta \frac{d\bar{S}(\beta;K)}{d\beta}\mathcal{P}_\beta(\mathbf{p}).$$

Where
1. $\bar{S}$ (β;K) is the average entropyof distributions chosen from P_β[3]

3. Bayesian estimator

Given this prior, the standard Bayesian probability of observing the data n={ni} given the distribution p is

$$P(\mathbf{n}|\mathbf{p}) \propto \prod_{i=1}^{K}p_i^{n_i},$$

and therefore by using bayes theorem,

$$P(\mathbf{p}|\mathbf{n}) = P(\mathbf{n}|\mathbf{p})\mathcal{P}_{NSB}(\mathbf{p}) \cdot \frac{1}{P(\mathbf{n})},$$

where,

$$P(\mathbf{n}) = \int d\mathbf{p}\ P(\mathbf{n}|\mathbf{p})\mathcal{P}_{NSB}(\mathbf{p}),$$

is the probability of n.
The entropy of mth moment is given by

$$(S^{NSB})^m = \int d\mathbf{p}\left(-\sum_{i=1}^{K}p_i\log_2 p_i\right)^m P(\mathbf{p}|\mathbf{n}).$$

Some important observations about using this method. First, since the analyst is Bayesian, we obtain not one S_{NSB} but also its a posteriori standard deviation, $\delta*S_{NSB}$ - an error bar on our estimate. Second, for N→∞ and N/K → 0 the estimator admits asymptotic analysis. The important parameter is the number of coincidences

$$\Delta = N-K1$$

where K1 = the number of bins with non zero counts.
If there are many coincidences(Δ /N < 1) then the standard saddle point evaluation of the integrals in calculating $\bar{S}$(β;K) is possible. Third, both of the above asymptotics show that the estimation procedure relies on Δ to make its estimates[3].

Fourth, S_{NSB} is unbiased if the distribution being learned is typical in $P_\beta(p)$ for some β, that is, its rank ordered.

4. What the paper tries to accomplish?

Any reasonable estimator will converge to the right entropy value as N tends to infinity and so does the S_{NSB} which is a standard Bayesian estimator. The main concern with entropy estimation is systematic bias which causes the result to over or under estimate the information content of spike trains or the efficiency of neural code. This bias occurs when N is small and for any practical purpose. Therefore, an ideal estimator would remove this bias or at least give results within its error bars from the correct answer. As N increases the error bars should narrow, with relatively little variance of the estimate itself. When data are such that no reliable estimation is possible the variance should be large.

This paper aims to show that the NSB procedure applied to natural and nature-inspired synthetic signals come close to this ideal, over a wide range of N << K, and even N << 2^S. It also aims to prove that the procedure thus is a viable tool for experimental analysis.

C. A model problem

To test the estimator a synthetic neural spike train is used such that they represent a natural spike train for which the authors also know the output response.

5. Experimental setup

Therefore, the synthetic spike trains' intervals between successive spikes were independent and chosen from an exponential distribution with a dead time or refractory period of g = 1.8ms; the mean spike rate was r = 0.26 spikes/ms.

The authors examine the spike train in windows of duration T=15ms and discretize the response with a time resolution tow=0.5ms. Because of the refractory period each bin of size tow can contain at most one spike, and hence the neural response is a binary word with T/tow =30 letters. The space of response without considering refractories K = 230, but we know that a lot of bins have zero probability because of refractoriness, and therefore K = 216.

Theoretically, calculating the entropy using the formula for entropy of a Poisson process, the entropy turns out to be 13.57 bits.
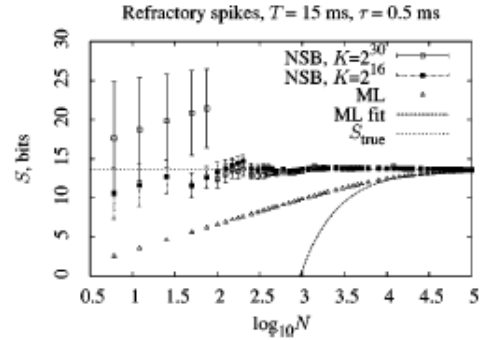
6. Results

Figure 1. Entropy estimation for a model problem.

The naïve S^{ML} represented by the dotted line (ML fit), does not do so well for smaller values of N and only estimates correctly near N=10^5. That is it only reaches its asymptotical behaviour when N>2^S.

The NSB plots, in contrast, meet the true value within the accepted error range at nearly N=100.

Even better results are obtained when the estimator is also given a smaller K value as that is true most of the time. That is, NSB converges fast when N = $2^{S/2}$.

D. Analyzing real data

For a test on real neurophysiological data, the authors used a recording from a wide field motion sensitive neuron (H1) in the visual system of the blowfly Calliphora vicina. While action potentials from H1 were recorded, the fly rotated on a stepper motor outside among the bushes, with time dependent angular velocity representative of natural flight. [3]

From the first paper, the information content of a spike train is the difference between its total entropy and the entropy of neural responses to repeated presentations of the same stimulus. The latter is substantially more difficult to estimate. It is called the noise entropy Sn since it measures response variations that are uncorrelated with the sensory input. The noise in neurons depends on the stimulus itself— there are, for example, stimuli which generate with certainty zero spikes in a given window of time—and so we write $S_{n|t}$ to mark the dependence on the time t at which we take a slice through the raster of responses.

7. Experimental setup

In this experiment the full stimulus was repeated 196 times, which actually is a relatively large number by the standards of neurophysiology. The fly makes behavioral decisions based on ,10−30 ms windows of its visual input, and under natural conditions the time resolution of the neural responses is of order 1 ms or even less, so that a meaningful analysis of neural responses must deal with binary words of length 10–30 or

more. Refractoriness limits the number of these words which can occur with non-zero probability, but nonetheless we easily reach the limit where the number of samples is substantially smaller than the number of possible responses.
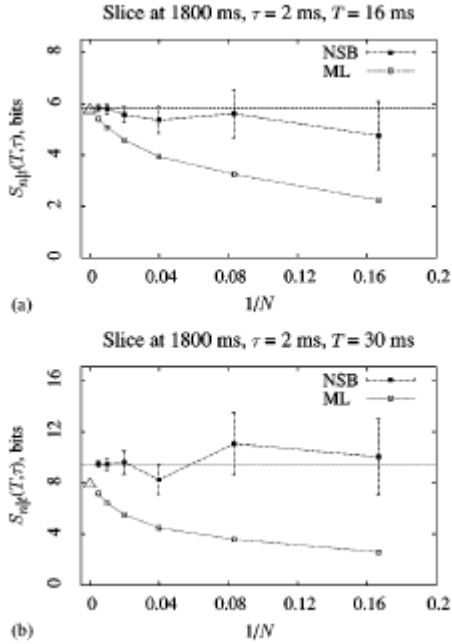
8. Results

(a)

(b)

Figure2. Slice entropy vs sample size. The graph ploted depicts a slice at 1800ms with the given T and τ values . This is done since the signal is continuous and it will be hard to intrepet a time serires data. Triangle corresponds to the value of S^{ML} extrapolated to $N\rightarrow\infty$ from the four largest values of N. First and second panels show examples of word lengths for which S^{ML} can or cannot be reliably extrapolated. S^{NSB} is stable in both cases, shows no N dependent drift, and agrees with S^{ML} where the latter is reliable.
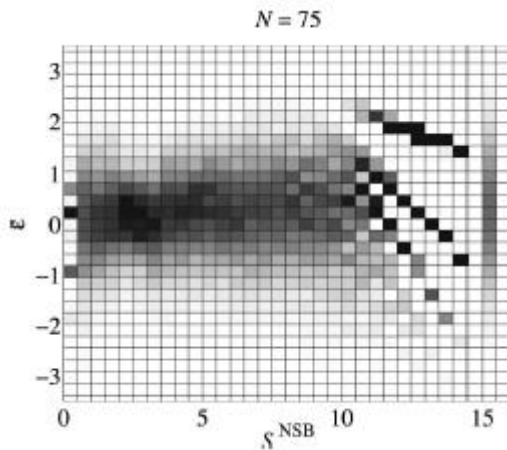

Figure3. Distribution of the normalized entropy error conditional on S^{NSB} (N_{max}) for N=75 and t =0.75 ms.

Darker patches correspond to higher probability. The band in the right part of the plot is the normal distribution around zero with the standard deviation of 1 (the standard deviation of plotted conditional distributions averaged over S^{NSB} is about 0.7, which indicates a non-Gaussian form of the posterior for small number of coincidences [3]). For values of S^{NSB} up to about 12 bits the estimator performs remarkably well.

For yet larger entropies, where the number of coincidence is just a few, the discrete nature of the estimated values is evident, and this puts a bound on reliability of S^{NSB}.

E. Discussion

The Bayesian estimator assumes a NSB prior which is uniform and also under experimentally verifies that this gives acceptable estimates of entropy. It is also verified that the prior almost completely removes the systematic bias in a synthetic setting and also in a real data with under sampling case.

The paper does not clearly provide evidence that NSB prior will work efficiently in all cases. There could be other such uniform priors which may allow successful estimation in real world problems. These priors have to be experimented with and only with those results the best prior can be chosen.

III. MECHANISMS OF INFORMATION FILTERING IN NEURAL SYSTEMS

Benjamin Lindner

A. Introduction

The way in which sensory information is encoded varies vastly across different model system. Generally, the manner of data encoding in sequential action potentials is not known. This makes room to study how data is not only encoded but also transmitted within neuronal networks.

Efforts of quantization of information contained and transmitted by spike trains have concentrated on the calculation of coherence and then estimate entropy measures based on those findings. Generally, a Gaussian stimulus is considered. As a consequence, frequency components are statistically independent.

One method is the usage of the stimulation-response coherence. It is given by –

$$C(f) = \frac{|S_{xx}(f)|^2}{S_{xx}(f)S_{ss}(f)}$$

The coherence is a squared "linear correlation coefficient between input(stimulus) and output (spike train)" in the frequency domain. Further, when a stimulus is sent, the message that can be reconstructed from the response to the stimulus is given by the mean square error term as –

$$\epsilon^2 = \int_0^\infty df S_{ss}(f)\left[1 - C(f)\right],$$

From the equation, it is clear that when C(f) is 1, the mean squared error in the linear reconstruction of the input stimulus is then 0. This implies that the stimulus can be flawlessly

reconstructed from the response. Conversely, when the C(f) term goes to 1, the mean squared error becomes equal to the variance of the signal. The coherence can thus also be related to the signal-to-noise ratio as –

$$SNR = C(f)/[1 - C(f)]$$

Relating the coherence measure to the concepts of information theory, a conservative estimate for the lower bound of the Shannon information rate, given a Gaussian stimulus and resultant spike train response is as follows –

$$R_{LB} = - \int_0^\infty df \log_2 [1 - C(f)]$$

The advantage of using a coherence measure for the calculation of information in the system is preferential to a direct attempt to calculate information transfer since coherence is a frequency-resolved method of calculating information measures. In neuronal systems, it becomes relevant to look at components of signals that are in different frequency bands and frequency resolutions allows for this.

Another method to estimate information rate is by using response-response coherence for the same stimulus. The response-response coherence is given as –

$$C_{x_1, x_2} = \frac{\langle S_{x_1, x_2} \rangle_s}{\sqrt{\langle S_{x_1, x_1} \rangle_s \langle S_{x_2, x_2} \rangle_s}},$$

S_{x_1, x_2} is the cross-spectrum of fixed-stimulus trials. An average is taken over the stimulus ensemble. The square root of this term (response-response coherence) gives an upper bound for stimulus-response coherence. This can be used to estimate an upper bound for mutual information rate.

The coherence function has been observed to take many shapes: broadband, low-pass, high-pass and band-pass, but these are usually restricted to very specific systems as well as the stimulus type that is considered. The shaping of the coherence function is an action of both the type of stimulus that is being received by a neuron as well as the action of that neuron on that stimulus. These actions are usually affected by biophysical properties of some combination of the neurons themselves, the connections between them. Information can be quantified in these systems by relating coherence to mutual information and information and since, using frequency resolution methods, the selective information rates can be observed at different frequency bands. A consequence of this is that it is possible to see the selective action of neurons on information transfer as a consequence of biophysical and chemical properties. This thus forms the basis for understanding information filtering as an action of the cell.

B. Information and Power Filtering

To differentiate between power and information and to then link them together, consider the following linear system –

$$x(t) = \int_{-\infty}^t dt' K(t - t')[s(t') + \eta(t')]$$

Here, s(t) is a time-dependent filter and η(t) is time-dependent noise. K(f) is called susceptibility. The form of this causal linear filter is determined by the amount of power that the filter allows at different frequency bands. Quantifying the cross-spectrum based on the correlations of the fluctuations of the output with the signal,

$$S_{xs} = \tilde{K}(f) S_{ss}(f)$$

Thus, for the power spectrum,

$$S_{xx} = |\tilde{K}(f)|^2 [S_{ss}(f) + S_{\eta\eta}(f)]$$

If the input stimulus as well as the noise are not temporally correlated, their spectra are constant. Thus, the cross-correlations between stimulus and output and the transmitted power will have a frequency dependence as a consequence of the filter K(f). If the filter decays the magnitude of output at higher frequencies, the effect if that of a low-pass filter. If, instead, a maximum is observed about a non-vanishing frequency, it behaves as a band-pass filter of power.

Calculating coherence gives –

$$C(f) = \frac{|\tilde{K}(f) S_{ss}(f)|^2}{|\tilde{K}(f)|^2 [S_{ss}(f) + S_{\eta\eta}(f)]}$$
$$= \frac{1}{1 + S_{\eta\eta}(f)/S_{ss}(f)},$$

Now, this gives a value that is completely unaffected by the power filter. For the case of when stimulus and noise are temporally uncorrelated, the coherence is a constant value. Thus, in this case, it behaves as a broadband transmitter of information.

For information filtering to occur, the shape of coherence function must change, and for that, the following conditions can affect the shape of the coherence function.

1. Instead of being completely independent of the frequency, noise entering the system may be correlated.
   a. A positive temporal correlation of the noise entering the system over typical time scales causes, lower frequencies, for increased noise power and at higher frequencies, diminished noise power. As a consequence of this, if a stimulus is spectrally flat with noise with these properties, the coherence increases towards higher frequencies. Such slow noise causes the high-pass filtering of information.
   b. Conversely, if the noise has a higher power at higher frequencies and lower power at lower frequencies, a low-pass filter of information is obtained.
2. The stimulus signals could have preferred power in certain frequency bands and not others or if the noise and signal are filtered by different kernels (more than one) and their peaks are not correlated.

C. Cellular Mechanisms of Information Filtering

Unlike the previously considered example, however, neural dynamics is non-linear. This non-linearity arises from the nature of neuronal response which is in the form of action potentials and spike trains.

1. Integrate-and-fire neuron

The Integrate-and-fire (IF) neuron with white noise is a biophysical model of a stochastically firing neuron. It is given by –

$$\tau_m \dot{v} = f(v) + \mu + \sqrt{2D(1-c)}\xi(t) + \sqrt{2Dc}s(t).$$

v describes the voltage across the nerve. A spike is registered when it crosses a threshold v_T which, after a period, resets voltage to a prescribed value v_R. $f(v)$ determines the subthreshold dynamics of the neuron, the mean input μ, the intrinsic noise $\xi(t)$ and the time-varying stimulus is given by $s(t)$. Both the $s(t)$ as well as the $\xi(t)$ are modelled by white Gaussian noise.

D. Intercellular and Network Mechanisms of Information Filtering

1. Synaptic Short-Term Dynamics

Neurons have short-term dynamics like synaptic facilitation and synaptic depression. Synaptic facilitation is when the synapse becomes stronger with use. Synaptic depression is the condition where the synapse becomes weaker with usage. This is also known as Short-Term Plasticity (STP).

These affect neural communication as signals are carried through stochastic synapses with these short-term dynamics.

In a stochastic Integrate-and-Fire neuron model, given a homogenous population of synapses with short-term plasticity, a model like this shows only a minimal deviation from a model containing only static synapses without short-term plasticity factored in. This is mostly attributed to the similar filtering of noise and the signal. In a realistic stochastic model, when a depressing synapse is used, the shape of the coherence function changes further. A high-pass filtering is observed on information.

In the different scenario of multi-sensory input, different synapses encode input independently and since they can each be affected by either of or both depressing and facilitating effects. Due to these different presynaptic input populations, signals that come through the facilitating neurons produce slow noise and block information (or dull it) transfer from the depressing neurons. The inputs entering though facilitating synapses are thus low-pass filtered and the inputs entering from depressing synapses are high-pass filtered as long as both signals occur at the same time.
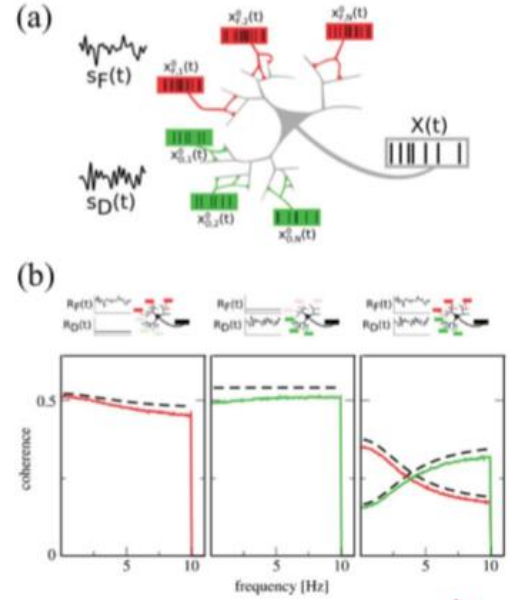

Figure M

Figure M (a) shows heterogenous synapses transmitting time-dependent signals. Figure M (b) shows how they each cause filters for each other and thus either high-pass or low-pass information filtering occurs.

#### 1. Synchronous Activity of a Neural Population

It is important to distinguish between the synchronous and asynchronous spikes since they essentially form separate information channels. Together, synchronous and asynchronous populations of neurons become multiplexed channels and encode distinct features of sensory stimuli.

It is seen that the synchronous output preferentially encodes information on fast stimulus components. Considering only the synchronous spikes, as a consequence of this preference, postsynaptic cells will suppress information about slower parts of the synapse.

E. The Coherence Function as a Frequency-Resolved Information Rate

The direct method, or the more traditional method of information rate calculation for neuronal networks are as follows. Given a neuron or a neuron model, broadband Gaussian inputs are repeatedly presented to this model and the spike trains that result from this neuron or neuron model are recorded. This spike trains are in the form of spike/no spike bins. Estimates for entropy is made on these bit sequences. Based on the difference between the unconditioned entropy and the noise entropy gives the mutual information. When the time window is taken into consideration, the information rate between the stimulus and spike train is obtained. In the end, this provides a single number corresponding to the signal.

A drawback of this approach is that, while it provides a measure for the overall information rate, neurons and neuron models have features that are not uniform across different frequencies of input due to short-term plasticity and

information filtering. Thus, in an effort to create a better measure for information rate in this context, the direct method is generalized to a frequency-resolved mutual information rate (MIR). The input Gaussian stimulus is decomposed into components that contain power only in distinct frequency bands.
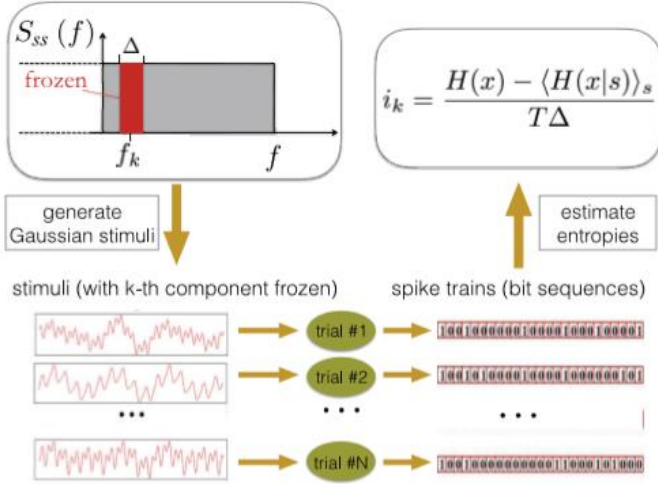
As in the figure, if only the region marked in red is considered as the signal and the rest is considered to be noise, doing the calculations through the direct method give the relevant information rate values for this specific band of frequencies.

$$i_{LB,k} = -\frac{1}{\Delta} \int_{f_k-\Delta/2}^{f_k+\Delta/2} df \; \log_2[1 - C(f)];$$

Using the estimations for the lower bound, this integrand gives a lower bound for the average over the frequency band. Further, using the bound obtained from the coherence-coherence calculations,

$$i_{RR,k} = -\int_0^\infty df \log_2\left[1 - \sqrt{C_{x_1,x_2}(f)}\right]$$

This is thus an upper bound for information rate where $C_{x1,x2}(f)$ is the spectral coherence function between the two trials for the same stimulus.

Further, the summing up of these numbers against the calculation of all the frequency bands put together can show whether there are any effects of the synergetic parts of the transmission due to which information transfer is affected by the interface between two frequency bands.

IV. CONCLUSION

Through this paper, we have summarized 3 papers in the field of information theory and neural coding. To some extent this work could be considered as a critical survey of the 3 papers from the information theory point of view.

V. REFERENCES

[1] "The reliability with which a neuron is able to create the same firing pattern when presented with the same stimulus" - Alstrøm, Preben, et al. "Reliability of neural encoding." Physica A: Statistical Mechanics and its Applications 314.1 (2002): 61-68.

[2] Borst, Alexander, and Frédéric E. Theunissen. "Information theory and neural coding." Nature neuroscience 2.11 (1999): 947-957.

[3] Nemenman, Ilya, William Bialek, and Rob de Ruyter van Steveninck. "Entropy and information in neural spike trains: Progress on the sampling problem." Physical Review E 69.5 (2004): 056111.

[4] Lindner, Benjamin. "Mechanisms of information filtering in neural systems." IEEE Transactions on Molecular, Biological and Multi-Scale Communications 2.1 (2016): 5-15.

[5] https://www.merriam-webster.com/medical/refractoriness