

Information Storage Capacity

Lav R. Varshney

The storage of information in biological memory relies on changes in neuronal circuits, termed plasticity. Synaptic contributions to plasticity, which are an important component, may be divided into changes in existing synapses, and changes in interneuronal connectivity through formation and elimination of synapses. Interneuronal connectivity changes may be further divided into contributions associated with dendritic spines without major remodeling of dendritic or axonal arbors, and contributions associated with major arbor remodeling [1].

An important step in understanding brain function is to quantify the information storage capacity of a neural system. Investigating an information theoretic way of defining the information capacity of a storage device, one finds different definitions in the literature. One way of defining the information capacity is to treat the recorder channel as an ordinary communications channel, *e.g.* an additive white Gaussian noise channel, and to consider the information capacity as the ordinary channel capacity [2]. Another way of defining the information capacity of a memory is as the logarithm of the number of cases it can distinguish, without explicit consideration of noise [3]. The difference between these two definitions is analogous to the difference between the definitions of Shannon entropy [4] and the Hartley information measure [5], thus we will refer to these two definitions as Shannon capacity and Hartley capacity, respectively. The quantity that we call Hartley capacity is closely related to MacKay's concepts of structural and metrical information [6]-[7].

The three classes of synaptic plasticity occur on distinct time scales, synaptic strength change within a minute, dendritic spine formation on the order of tens of minutes, and branch remodeling on the order of days [1]. For a basic look at the overall storage capacity we consider the information storage capacity arising from the three classes of synaptic plasticity separately. First let us consider the information capacity associated with formation and elimination of dendritic spines, following the analysis of [1].

In [1], Stepanyants *et al.* estimate the number of different interneuronal connectivity patterns that can be formed through dendritic spine formation/elimination without major arbor remodeling and calculate the Hartley capacity. They estimate the number of potential synapses, N_p , through a geometrical analysis of anatomy and the number of actual synapses, N , through anatomical measurement. These two values may be used to combinatorially determine the Hartley capacity. In particular the Hartley capacity is

$$I = \log \binom{N_p}{N} \approx \log \left(\left[(1-f)^{(1-f)} f^f \right]^{-N_p} \right) = N_p h_b(f), \quad (1)$$

where the filling fraction is defined as $f = N/N_p$, and $h_b(\cdot)$ is the binary entropy function. The approximation in (1) assumes that N_p is large and that f is between 0 and 1, which it clearly must be. In particular, applying Stirling's approximation

$$x! \approx \sqrt{2\pi x} \left(\frac{x}{e}\right)^x = \sqrt{2\pi x} x^{(x+1/2)} e^{-x}, \quad (2)$$

to the definition of the binomial coefficient

$$\binom{N_p}{N} = \binom{N_p}{fN_p} = \frac{N_p!}{(fN_p)! [(1-f)N_p]!}, \quad (3)$$

yields

$$\binom{N_P}{N} \approx 2^{N_P h_b(f)}, \quad (4)$$

which, after taking the base 2 logarithm yields the desired result. It should be noted that the approximation is an upper bound on the true Hartley capacity [8]. It is easily seen that for fixed N_P , the Hartley capacity is maximized when f is $\frac{1}{2}$ due to the binary entropy function in (1) or more directly from the property of the binomial coefficient.

If one could increase N_P without bound while keeping the filling fraction strictly greater than zero (*i.e.* having at least one synapse), then one could achieve infinite Hartley capacity. Since the brain is restricted in physical volume due to morphological constraints, if one increased N_P without bound, there would be an asymptotically dense set of potential synapses. Clearly in this asymptotically dense regime, the ability to distinguish different connectivity patterns would be lost and the Hartley capacity would no longer be applicable. In the terminology of MacKay [7], the logon-capacity (amount of structural information per unit volume) of neural tissue is finite due to the indistinguishability that eventually must occur. Some sort of noise or uncertainty relationship fundamentally limits distinguishability and thus capacity; therefore, MacKay's logon-capacity notwithstanding, some Shannon capacity measurement seems more natural. It should be noted that Shannon capacity is equal to the highest rate at which information can be stored with arbitrarily low probability of error, by the channel coding theorem [4], however in neural signaling [9] (and presumably in neural information storage), high rates of error may be tolerated and may in fact be beneficial from an energy standpoint [10].

Characterization of noise as a function of N_P density is required to calculate Shannon capacity. The maximum N_P density possible while maintaining distinguishability is required to calculate Hartley capacity. It was shown in [1] that the value of N_P for a given type of neural tissue may be calculated geometrically from the values of dendritic length per neuron, spine length, interbouton interval, and synaptic density. The question remains, however, regarding the noise.

It was demonstrated that there is a linear relationship between postsynaptic density size and release probability. In particular, postsynaptic density area is equal to the active zone area and the active zone area (μm^2) $\times 5.75$ gives release probability [11]. If postsynaptic density area is directly related to signal to noise ratio, this linear relationship to release probability (which may be considered as one minus the error probability) might imply a fading-type channel model as in wireless communications. It is well known that the detection error probability decays exponentially in SNR in the AWGN channel while it (outage probability in the information theoretic analysis) decays only inversely with the SNR in the fading channel [12, p. 69; p. 218].

Let us now turn to the information storage capacity of the synaptic weight process. We will assume that the information capacity of each existing synapse is independent and equal. Furthermore, we will model the storage channel as an additive white Gaussian noise (AWGN) communication channel, deferring to the central limit theorem. Manwani and Koch also use the white Gaussian noise assumption for synaptic signaling [9]. The Shannon capacity of an AWGN channel is perhaps the most celebrated result in all of information theory

$$C_s = \frac{1}{2} \log_2 \left(1 + \frac{P}{P_N} \right), \quad (5)$$

where P is the average power constraint and P_N is the channel noise variance. If the AWGN channel is used as a discrete channel, the capacity is reduced. For example, if the continuous

AWGN channel is converted into the discrete binary symmetric channel (BSC), then the capacity is

$$C_s = 1 - h_b \left[\Phi \left(\sqrt{\frac{P}{P_N}} \right) \right], \quad (6)$$

where $\Phi(\cdot)$ is the cumulative normal function

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{t^2}{2}\right\} dt. \quad (7)$$

Again it should be reiterated that the Shannon capacity is equal to the highest rate at which information can be stored with arbitrarily low probability of error. Though it is also known, by the strong converse of the noisy channel coding theorem, that the probability of error goes exponentially to 1 for rates greater than capacity [13, p. 207]. Thus one may reasonably consider a Hartley capacity that is slightly greater than the Shannon capacity given in (5) or (6).

The information storage capacity is then simply the product of the synapse capacity and the number of synapses

$$I = NC_s. \quad (8)$$

Information measures, by definition, are additive, so the total information storage capacity arising from interneuronal connectivity and from synaptic weights is the sum of the individual information capacities of these individual processes

$$I_{tot} = N_p h_b \left(\frac{N}{N_p} \right) + \frac{N}{2} \log_2 \left(1 + \frac{P}{P_N} \right). \quad (9)$$

Let us assume that there is a maximum volume, V^0 , that the neural tissue may take, and based on the maximum potential synapse density that allows differentiability of connectivity patterns, D_p^0 potential synapses per unit volume, there is a maximum N_p denoted as $N_p^0 = V^0 D_p^0$. The maximum potential synapse density will be a function of some noise in the system, however let us not be concerned with the origin of this limit. In order to maximize total information capacity, as given in (9), the number of potential synapses should take the value N_p^0 .

“From numerous examples, it is qualitatively clear that the reduction of noise is accomplished through resource consumption in neurobiology, as it is in electronics” [14, p. 1634], however we must establish a quantitative relationship. Now let us assume that the synapse signal to noise ratio is strictly a function of the signal amplitude, i.e. the excitatory postsynaptic potential (EPSP) amplitude. This implies that the noise level is a fixed level independent of the signal level. So increasing the signal level increases the signal to noise ratio. This assumption seems to go against experimental evidence of a weak dependence between EPSP amplitude and standard deviation that is given in [15, p. 25], however, this relationship is most likely caused by some sort of variable quantal amplitude as observed in [9, p. 7], rather than by noise. This assumption may need to be revised. The variable quantal amplitude may also be fruitfully modeled as a type of fading [12]. It seems reasonable that the EPSP amplitude can be greater on larger synapses and so let us assume a linearly proportional relationship between synapse size and EPSP amplitude

$$P = \beta V_s, \quad (10)$$

where V_s is the volume of the synapse and β is some proportionality constant.

Further, let us assume that the number of synapses, N , is less than half of the value of potential synapses, N_P , i.e. that the filling fraction is less than one half. Clearly the total volume of the synapses must be less than the volume constraint

$$NV_S \leq V^0, \quad (11)$$

where V_S is the volume of a synapse. Let us assume that one can specify a maximum fraction of the total volume that the synapses can take, g . Then the maximum volume that the synapses can occupy is gV^0 . Assume that the synapses do occupy this maximum volume, then V_S may be specified as gV^0/N . So now the question of maximizing information storage capacity can be framed as

$$\arg \max_{N < \min\left(\frac{D_P^0 V^0}{2}, \frac{N}{g}\right)} I_{tot} = \arg \max_{N < \min\left(\frac{D_P^0 V^0}{2}, \frac{N}{g}\right)} D_P^0 V^0 h_b \left(\frac{N}{D_P^0 V^0} \right) + \frac{N}{2} \log_2 \left(1 + \frac{\beta g / V^0 N}{P_N} \right). \quad (12)$$

This can be simplified as

$$= \arg \max_{N < \frac{D_P^0 V^0}{2}} D_P^0 V^0 h_b \left(\frac{N}{D_P^0 V^0} \right) + \frac{N}{2} \log_2 \left(1 + \left(\frac{\beta g}{P_N V^0} \right) \left(\frac{1}{N} \right) \right). \quad (13)$$

In the interneuronal connectivity term (the first term), it is clearly best to have N as large as possible. In the synaptic weight term (the second term), the factor of N outpaces the factor of N^{-1} inside the logarithm, and so in this term it is also best to have N as large as possible. When the $\beta g / P_N V^0$ term is much less than one, then the increase in information capacity that results from increased N diminishes quite a bit after a certain point. So mathematically speaking, the optimizing N is $N_P/2$.

If we look at the synaptic weight term as a binary storage device rather than a continuous-valued storage device, then the information capacity will be

$$\begin{aligned} I_{tot} &= N_P h_b \left(\frac{N}{N_P} \right) + N - N h_b \left[\Phi \left(\sqrt{\frac{P}{P_N}} \right) \right] = D_P^0 V^0 h_b \left(\frac{N}{D_P^0 V^0} \right) + N - N h_b \left[\Phi \left(\sqrt{\frac{\beta g / V^0 N}{P_N}} \right) \right] \\ &= D_P^0 V^0 h_b \left(\frac{N}{D_P^0 V^0} \right) + N - N h_b \left[\Phi \left(\sqrt{\left(\frac{\beta g}{V^0 P_N} \right) \left(\frac{1}{N} \right)} \right) \right] \end{aligned} \quad (14)$$

Again, the first term is maximized with maximal N . The second term is likewise maximized with maximal N . The third term will be maximized when the argument of the binary entropy function is $1/2$. This occurs when the argument of the cumulative normal function is zero. For a finite positive value of $\beta g / P_N V^0$, N would need to be infinite for this to happen, and since $N \leq N_P/2$, this will not happen. In any case, the greater N is, the smaller the value of the cumulative normal function argument is, and so the smaller the value of the cumulative normal function, and so the greater the value of the binary entropy function, and so greater total storage. So in this case also, the best N is $N_P/2$.

References

- [1] A. Stepanyants, P.R. Hof and D.B. Chklovskii, "Geometry and Structural Plasticity of Synaptic Connectivity," *Neuron*, vol. 34, pp. 275-288, Apr. 2002.
- [2] D.F. Eldridge, "A Special Application of Information Theory to Recording Systems," *IEEE Trans. Audio*, vol. 11, pp. 3-6, Jan. 1963.

- [3] Y.S. Abu-Mostafa and J.-M. St. Jacques, "Information Capacity of the Hopfield Model," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 461-464, Jul. 1985.
- [4] C.E. Shannon, "A Mathematical Theory of Communication," *Bell Sys. Tech. J.*, vol. 27, pp. 379-423 and 623-656, Jul. and Oct. 1948.
- [5] R.V.L. Hartley, "Transmission of Information," *Bell Sys. Tech. J.*, vol. 7, pp. 535-563, Jul. 1928.
- [6] D.M. MacKay, "Quantal Aspects of Scientific Information," *Phil. Mag.*, ser. 7, vol. 41, pp. 289-311, Mar. 1950.
- [7] D.M. MacKay, "Quantal Aspects of Scientific Information," presented at *London Symposium on Information Theory*, Sep. 1950, printed as *IRE PGIT-1*, Feb. 1953.
- [8] W. Feller, *An Introduction to Probability Theory and Its Applications*, New York: Wiley, 1957.
- [9] A. Manwani and C. Koch, "Detecting and Estimating Signals over Noisy and Unreliable Synapses: Information-Theoretic Analysis," *Neural Comput.*, vol. 13, pp. 1-33, 2000.
- [10] W.B. Levy and R.A. Baxter, "Energy-Efficient Neuronal Computation via Quantal Synaptic Failure," *J. Neuroscience*, vol. 22, pp. 4746-4755, Jun. 2002.
- [11] T. Schikorski and C.F. Stevens, "Quantitative Ultrastructural Analysis of Hippocampal Excitatory Synapses," *J. Neurosci.*, vol. 17, pp. 5858-5867, Aug. 1997.
- [12] D. Tse and P. Vishwanath, *Fundamentals of Wireless Communication*, Cambridge: Cambridge University Press, 2005. [Online] Available: <http://degas.eecs.berkeley.edu/~dtse/book.pdf>.
- [13] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, New York: John Wiley & Sons, Inc., 1991.
- [14] R. Sarpeshkar, "Analog Versus Digital: Extrapolating from Electronics to Neurobiology," *Neural Comput.*, vol. 10, pp. 1601-1638, 1998.
- [15] S. Song, P.J. Sjöström, M. Reigl, S. Nelson and D.B. Chklovskii, "Highly Non-Random Features of Synaptic Connectivity in Local Cortical Circuits," unpublished.