

On Optimal Information Storage in Synapses

Lav R. Varshney
Massachusetts Institute of Technology

Dmitri B. Chklovskii
Cold Spring Harbor Laboratory

Abstract—Among the functions of the brain is information storage, which is physically implemented through changes in the strengths of synapses. Experimental investigations have revealed that synapses possess interesting and, in some cases, unexpected properties. Adopting the optimization approach to biology, we describe an information theoretic framework that accounts for several of these properties: typical central synapses are noisy, the distribution of synaptic weights among central synapses is wide, and synaptic connectivity between neurons is sparse. Our approach is based on maximizing channel capacity of neural tissue under resource constraints. We cast volume as a limited resource and utilize the empirical relationship between volume and synaptic strength. We find that capacity-achieving input distributions not only explain existing experimental measurements but also make non-trivial predictions about the physical structure of the mammalian brain. We also comment on the robustness of our optimization principles to uncertainties that are inherent in science.

I. INTRODUCTION

Investigations in electrical engineering and neuroscience have been intertwined for centuries [1] and have included attempts to use information theory to qualitatively understand neural memory [2], [3]. Previous quantitative neuroscience studies have used information theory to analyze transmission of signals between neurons through synapses, particularly for sensory processing [4]. The work we summarize here applies the information theoretic framework to a quantitative neuroscience investigation of synapses as memory elements.

It is widely believed that memories are stored in the brain and that the physical mechanism for information storage involves changes in synapses, termed synaptic plasticity. As such, we develop a theory that provides explanations for and makes predictions about the physical properties of synapses. Our approach is rooted in the optimization approach to biology [5] in which biological systems are thought to be best solutions to some optimization problem and thus seem to be “engineered” under some optimality criterion. Since evolution favors genotypes of high fitness, qualities that affect fitness tend to improve. Thus, mathematical analysis of fitness optimization seems reasonable as an approach to understand why animals are the way that they are.

Our scientific program proceeds in steps: 1) ask explicit scientific question, 2) define feasible set of structures in system model, 3) define objective fitness function, 4) determine optimization principles as hypotheses, yielding predictions, and 5) test hypotheses through experiment. By maximizing

the information storage capacity of synapses under volume constraints, we promulgate a unified theory that explains why typical synapses are noisy devices, why neuronal connectivity is sparse, and why the distribution of synaptic strength is broad. These properties and quantitative refinements thereof are all evident in extant experimental observations. We also provide non-trivial predictions, such as a quantitative prediction for the relationship between synaptic volume and synaptic strength, which has not yet been experimentally verified.

This short paper is a summary of the results contained in [6]. After briefly discussing the system model, we focus on the development of optimization principles, and finally touch on experimental verification. The reader is referred to [6], [7] for more biological details in system modeling and experimental verification, details on mathematical derivations, as well as perspectives on other results and alternate theories for synaptic structure. Let us reiterate that we omit many references to the literature that appear in [6]. We close with a discussion of the results and comments on robustness.

II. SCIENTIFIC QUESTIONS POSED

Since synapses are important for information processing and information storage in the brain [8], their properties have been the subject of extensive experimentation. We review the results of some experiments and use them to formulate our scientific questions.

Typical synapses in the central nervous system are noisy devices. Arriving pre-synaptic action potentials occasionally fail to evoke an excitatory post-synaptic potential (EPSP); furthermore, when an EPSP is evoked, the amplitude varies from trial to trial [9], [10]. This noisiness of typical central synapses has seemed puzzling because synapses act as conduits of information between neurons and in general, synaptic unreliability is detrimental for information transmission. Thus we ask:

Question 1: Why are typical central synapses noisy?

The distribution of synaptic weight across different synapses in the brain is broad. Although the majority of synaptic weights are relatively weak (mean EPSP < 1 mV), there is a notable tail of stronger connections [10]–[12], with EPSP values exceeding 20 mV having been observed. Given the pervasiveness of equality and symmetry in solutions to optimization problems, one might think that if there is some optimal value of synapse strength, then this value should be used for all synapses, hence the puzzling nature of this observation.

Work supported in part by National Science Foundation Grant CCR-0325774.

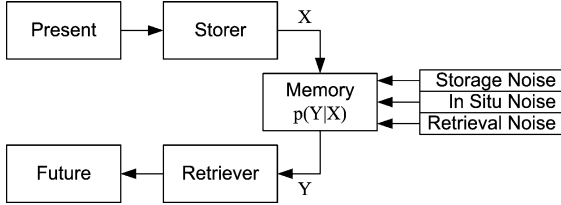


Fig. 1. Schematic of a neural memory system cast as a general communication system. The various sources of noise in the system have been explicitly notated. Storage noise refers to noise that arises in the storage process; in situ noise refers to noise that perturbs the information while it is stored; and retrieval noise refers to noise in the retrieval process.

Question 2: Why does the distribution of synaptic strengths display a notable tail of strong connections?

A third observation that has been made regarding synapses is their absence. One can experimentally create an adjacency matrix of neuron-neuron connections. This adjacency matrix is sparse in the brain. One might explain this away through geometric constraints by inferring that absent connections are between neurons that are physically distant, however the synaptic connectivity is sparse in local circuits also [11], [12]. In local circuits, neuronal axons and dendrites are close enough so that the potential connectivity is nearly all-to-all, so the cost of adding connections is not overwhelming. It seems that increased connectivity would enhance system functionality, so the sparse connectivity is puzzling.

Question 3: Why is synaptic connectivity sparse?

Having formulated the questions to be addressed, we move to the second and third steps of our scientific program: the development of a physicomathematical model of synapses as information storage elements.

III. MODEL OF SYNAPTIC MEMORY

We develop a model of synapses based on their role in long-term information storage. It is widely believed that long-term memories are recorded in neuronal circuits through alteration in the strength of existing synapses through long-term potentiation (LTP) and long-term depression (LTD) [13], [14] or through synaptogenesis and synapse elimination. Memories are retrieved from storage by chemical and electrical activity of neurons, which generates synaptic potentials determined by the pattern of synaptic connectivity between them. Thus a synaptic memory system is naturally cast into the model of a general communication system as shown in Figure 1.

We specify the model in the usual information theory way. Each synapse represents a channel usage, the channel input is the synaptic strength, and the distribution of synaptic strengths across synapses is the channel input distribution. Figure 2 illustrates the basic form that the signaling scheme can take.

There is noise associated with weight change processes, which we term the storage noise. While information is stored in the synapse, computational and short-term memory processes may act to perturb the synaptic strength; this is termed in situ noise. The retriever consists of activity to retrieve synaptic strengths, as well as some sort of decoding. There

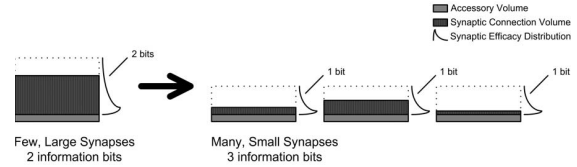


Fig. 2. Illustration of two possible volume-constrained signaling strategies for synaptic information storage. Replacement of a large synapse with several smaller ones (connecting different neurons) may increase information storage capacity. Notice that the total volume of the smaller synapses is the same as that of the larger synapse. Since synapse volume is a monotonic function of synapse strength, the realization of synapse volume also represents the strength realized by that synapse.

is noise associated with this EPSP-generating chemical and electrical activity, which we term retrieval noise.

We identify the input letters with the mean EPSP amplitude generated over numerous firings of the pre-synaptic neuron. Numerous studies have shown synaptic weight to be positively correlated with synaptic volume, e.g. [15]. Synaptic volume increases with LTP or synaptogenesis and decreases with LTD or synapse elimination [16], [17]. As stated in Ramón y Cajal's laws of economy of space, time, and matter [8], volume in the brain is a costly resource. We take our channel cost constraint to be determined by the single-letter synapse volume cost function.

In addition to the volume cost of the synapse itself, some accessory volume is needed to support a synapse. Accessory volume includes the volume of axons, dendrites, cell bodies, glia, and perhaps extracellular space. Since each potential synapse represents a channel usage, this accessory volume may be interpreted as the cost of providing discrete-space bandwidth. Unlike in other information systems, bandwidth and channel alphabet cost (power in other systems) consume the same resource and thus an interesting tradeoff between allocating resources for bandwidth or for power is presented. An alternative view is that there are no free symbols, and that the minimal achievable cost is this accessory volume, denoted V_0 .

The channel input letters are the average EPSP amplitudes, $x \in \mathcal{X}$, where $x \triangleq E[Y|x]$. A second-order noise characterization is given by the standard deviation of the EPSP readout trials, denoted as noise amplitude A_N . The relationship between synapse amplitude and noise amplitude is well-approximated by a power law with exponent about 0.38 [12], [18]. Since A_N is related by a power law to the mean EPSP amplitude, x , which is itself strongly correlated with the synapse volume, V , we formulate the following relationship:

$$\frac{V}{V_\nu} = \left(\frac{x}{A_N} \right)^\alpha, \quad (1)$$

where V_ν is a normalization constant, the volume of a synapse with SNR one. Although existing experimental measurements support (1), they are not sufficient to establish the value of the exponent, α . Determination of α would specify part of the cost, $V(x)$, of the input letter x . This establishes the physicomathematical model of synapses in terms of channel

input alphabet \mathcal{X} , the single-letter cost function $b(x) = V(x) + V_0$, and a description of the channel noise by its moment A_N .

It seems reasonable that having a large memory capacity is beneficial to an animal, so the overall objective function is to achieve channel capacity per unit cost [19]. Since coding may be very difficult to recognize experimentally without knowing the codebook and would require in vivo experiments on a single animal at a fixed time, we focus on characterizing more readily observable quantities. It is known that the capacity-achieving input distribution must be used by any code that operates at capacity [20], so as to achieve shaping gain [21].

For our problem then, the decision variable for optimization is the signaling scheme, described by the distribution of synaptic strength across synapses, $p_X(x)$ [22]. In an alternate formulation, the decision variable may be the single-letter cost function $b(x)$ [23]. Note that constraints imposed by the evolutionary process and by some biological mechanisms have been implicitly incorporated. The basic structure of the system has been assumed and the details are to be determined by the optimization theory.

IV. NOISY SYNAPSES MAXIMIZE CAPACITY

Question 1 asked why synapses are noisy on average. By the relationship between noisiness and small volume established in (1), the question can be recast as the question of why synapses are small on average. That is to say, why is the average incurred cost $B = E[b(X)]$ small. We deduce optimal average synaptic weight and volume by maximizing information storage capacity per unit volume. We want to determine where on the capacity-cost function $C(B)$ should signaling operate, finding the capacity per unit cost. As shown by Verdú [19], the capacity per unit cost is given by

$$\mathfrak{C} = \sup_{B>0} \frac{C(B)}{B} = \sup_{p_X(x)} \frac{I(X;Y)}{E[V(X) + V_0]}. \quad (2)$$

For concreteness and ease of exposition, assume that $\alpha = 2$ in the weight/volume relationship (1) and that the noise is Gaussian. The qualitative result and the general optimization principle do not depend on this assumption; at the end of this section, we argue that the conclusions hold for other noise models. Information storage capacity per synapse for the AWGN channel is given by [22]

$$C_{synapse} = \frac{1}{2} \ln \left(1 + \frac{\langle X^2 \rangle}{A_N^2} \right), \quad (3)$$

where $\langle X^2 \rangle$ is used to denote $E[X^2]$. The quantity $\frac{\langle X^2 \rangle}{A_N^2}$ is the average SNR among synapses with respect to the channel input and to the noise distributions. SNR for each synapse is defined as the square of the mean EPSP amplitude divided by the trial-to-trial variance of EPSP amplitude. Using (1), we rewrite information storage capacity in terms of volume

$$C_{synapse} = \frac{1}{2} \ln \left(1 + \frac{\langle V \rangle}{V_\nu} \right), \quad (4)$$

where $\langle V \rangle$ is the average synapse volume, $E[V(X)]$, excluding the accessory volume.

Dividing by the average cost of a synapse, the total storage capacity of a unit volume of neural tissue is

$$\mathfrak{C}_{volume} = \frac{C_{synapse}}{\langle V \rangle + V_0} = \frac{1}{2(\langle V \rangle + V_0)} \ln \left(1 + \frac{\langle V \rangle}{V_\nu} \right). \quad (5)$$

The location of the maximum of a curve drawn according to (5) gives the normalized average volume, $\langle V \rangle / V_\nu$, that provides the optimal storage capacity. By differential calculus, we find that the optimal $\langle V \rangle$ satisfies the equation

$$\frac{\langle V \rangle + V_0}{\langle V \rangle + V_\nu} = \ln \left(1 + \frac{\langle V \rangle}{V_\nu} \right). \quad (6)$$

As would be expected, maximum information storage capacity per unit volume is achieved when the accessory volume V_0 is the smallest possible. In this regime, average synapse volume $\langle V \rangle$ is much less than V_ν and according to (1), synapses should therefore be noisy. In reality, the accessory volume V_0 may not be infinitesimal as this could affect system functionality adversely. As an example, reducing wiring volume increases conduction time delays and signal attenuation [24]; in fact, delay and attenuation are optimized when the wiring volume is of the same order as the volume of synapses [25]. Since the increase in capacity achieved by reducing accessory volume below V_ν is not too large, it is reasonable to think that the average synapse volume $\langle V \rangle$ is either less than or of the order of V_ν . In either case, we arrive at the conclusion that typical synapses should be noisy, in agreement with experimental observations.

The advantage of having greater numbers of smaller synapses is valid not only for the AWGN model, but also for many reasonable noise and cost models, following from the non-decreasing and concave downward nature of the capacity-cost function [26]. Due to concavity, the slope (capacity/cost) increases at lower costs. If there are no zero-cost symbols, then the capacity per unit cost is maximized at the average cost value where a line through the origin and tangent to the capacity-cost function has its point of tangency. If there is a zero-cost symbol, then the optimum is for zero average cost. We have developed the following general optimization principle, provided that the accessory volume needed to support a synapse is small.

Optimization Principle 1: To optimize information storage capacity per unit volume of neural tissue, synapses should be small and noisy on average.

V. OPTIMAL DISTRIBUTION OF SYNAPTIC STRENGTHS

Having established that synapses should be small and noisy on average, we are interested in how volume and EPSP amplitude should be distributed among synapses. The distribution of synaptic weight addresses both Questions 2 and 3 because absent synapses are synapses with zero weight. In the AWGN model, the capacity-achieving input distribution is also Gaussian [22], however, it is not clear whether the prediction from the AWGN model can be taken at face value.

Consider an alternative noise model, where the cost function can be chosen arbitrarily and the synaptic weight is non-negative, but still yields an exactly solvable optimization problem. Treating noise approximately, consider a set of discrete synaptic states, with each state representing the range of weights in the continuous distribution that could be confused on retrieval due to noise. The difference in synaptic weight between adjacent states is twice the noise amplitude $2A_N$. We are engaging in hard sphere packing rather than soft sphere packing, hence the factor of 2. Each of these finite states is viewed as an input symbol with associated volume cost.

The conversion of the noisy continuous-valued input channel into a zero-error, discrete-valued input channel is a convenient approximation [27]. This approximation allows us to calculate the ϵ -capacity-cost, a generalization of the Kolmogorov ϵ -capacity, rather than the Shannon capacity. Since the converted channel is noiseless, the maximum entropy method may be used. By resorting to this ϵ -capacity-cost approximation, we do not wish to imply that synaptic weights in the brain must vary in discrete steps, though arguments can be made to support discrete capacity-achieving input distributions [6].

Consider a set of synaptic states, i , characterized by the EPSP amplitudes, x_i , and volume cost, b_i , which includes both the synapse volume and the accessory volume V_0 . We want p_i to maximize information storage capacity per average volume of a synapse:

$$\frac{C_{synapse}}{B} = \frac{C_{synapse}}{E_{p_i}[b(x)]} = \frac{-\sum_i p_i \ln p_i}{\sum_i p_i b_i}. \quad (7)$$

The average synaptic volume cost B includes the accessory volume, V_0 , which was excluded from the definition of $\langle V \rangle$.

It is well-known that the maximum entropy distribution takes a Gibbs form, therefore the optimal p_i is

$$p_i = \frac{1}{Z} \exp(-\lambda b_i), \quad (8)$$

where $Z = \sum_i \exp(-\lambda b_i)$ is a normalization constant and the Lagrangian quantity λ is specified implicitly. The capacity per unit cost problem is solved by choosing λ such that $Z = 1$, so (8) simplifies to $p_i = \exp(-\lambda b_i)$, where λ is defined by the condition $\sum_i p_i = 1$. This input distribution gives

$$\mathfrak{C}_{volume} = \frac{C_{synapse}}{B} = \frac{\lambda \sum b_i \exp(-\lambda b_i)}{\sum b_i \exp(-\lambda b_i)} = \lambda. \quad (9)$$

Now assume that synaptic state volume is distributed equidistantly, $b_i = V_0 + i2V_\nu$. To allow comparison with empirical measurements, we also derive an expression for the average volume of actual synapses, i.e. states with $i > 0$, and excluding accessory volume, V_0 , under the assumption.

$$\langle V \rangle_{i>0} = E_{q_i}[b - V_0], \quad (10)$$

where we have defined a distribution on states $i > 0$, q_i , which is the conditional probability of being in state i given not in state 0. This simplifies to

$$\langle V \rangle_{i>0} = 2V_\nu \exp(\lambda V_0). \quad (11)$$

The optimal average volume of actual synapses increases with the accessory volume. This result has an intuitive explanation: once the big investment in wiring (V_0) has already been made, it is advantageous to use large synapses with high SNR.

The ratio between the number of actual synapses and the number of potential synapses is called the filling fraction, f . In our model f is just the fraction of synapses in states $i > 0$ and is

$$f = \exp(-2\lambda V_\nu) = 1 - \exp(-\lambda V_0). \quad (12)$$

Just as before, information storage capacity increases monotonically with decreasing accessory volume, so optimal information storage is achieved when V_0 is as small as possible. As seen from (12), in the small V_0 limit, f is much less than one. This prediction is qualitatively consistent with empirical observations of sparse connectivity. In addition, most actual synapses have volume $2V_\nu$, and thus have SNR of order one, (1). This prediction is in agreement with the empirically established noisiness of typical synapses. Although local cortical circuits are sparse and typical synapses are noisy, the filling fraction is not infinitesimal. As discussed in the previous section, system functionality would be adversely affected by very small V_0 . The condition that accessory wire volume is of the order of synapse volume [25], implies that $V_\nu \sim 1/\lambda$, meaning that information storage capacity of volume V_ν is on the order of one nat. This sets the filling fraction at a value below 1/2 but not infinitesimal.

We can find the probability of synaptic states in terms of the EPSP amplitude,

$$p_i = \exp(-\lambda V_\nu (x_i/A_N)^\alpha). \quad (13)$$

This converts to a density as

$$p_X(x) = \frac{1}{A_N} \exp(-\lambda V_\nu (x/A_N)^\alpha), \quad (14)$$

which we compare with experiment in Section VII.

The explicit consideration of noise does not alter the result that for $V_0/V_\nu \rightarrow 0$ optimum information storage is achieved by using mostly the $i = 0$ state, with $i = 1$ used with exponentially low frequency. If $V_0 = 0$, this type of problem can be solved exactly [19], [28] and the information storage capacity is maximized when, in addition to the zero cost symbol, only one other symbol is chosen. If $V_0 > 0$, however, the problem of optimizing information storage capacity cannot be solved easily, prompting us to pursue a reverse approach discussed in the next section.

To summarize the results of this section, we state two further general optimization principles.

Optimization Principle 2: To optimize information storage capacity per unit volume, the filling fraction should be small, an exponential with exponent determined by the SNR exponent α and accessory volume V_0 . Small filling fraction is equivalent to sparse synaptic connectivity.

Optimization Principle 3: To optimize information storage capacity per unit volume, although there will be many absent synapses and numerous small synapses, there will also be

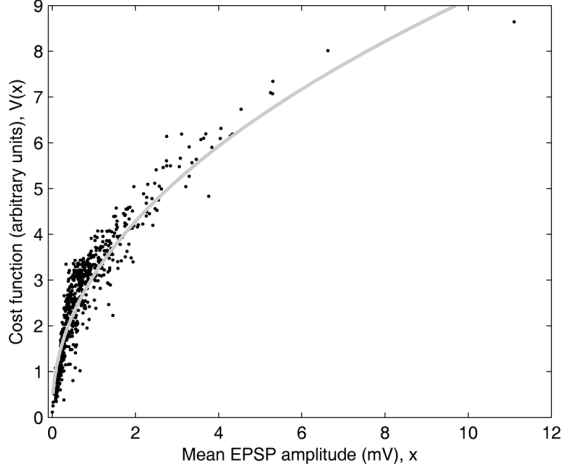


Fig. 3. Optimizing synaptic cost function calculated from EPSP measurements. Synaptic cost function as a function of mean EPSP amplitude calculated from (15) under assumption of optimal information storage. Each data point represents a different synapse. The line shows a least squares fit by a power law with exponent 0.48.

some large synapses, with synaptic efficacy distribution like a stretched exponential.

VI. CALCULATION OF THE SYNAPTIC COST FUNCTION

Previously we used an ϵ -capacity-cost approximation to obtain a capacity-achieving input distribution. To move away from this approximation, we apply optimality conditions from measure matching [23]. The channel transition probability and the channel input distribution can be measured experimentally and used to determine the single-letter cost function for which the system is operating at capacity-cost [23]. In particular, the optimizing cost function is

$$b(x) = \nu D(p_{Y|X}(\cdot|x) \| p_Y(\cdot)) + \nu_0 \quad (15)$$

for some $\nu > 0$ and ν_0 . Although this method inverts the problem specification, it seems reasonable if we are not sure of the channel input cost function (e.g. we do not know α). The resulting cost function may then be examined for relevance to the problem. Note that due to the arbitrary constants in (15), the cost function will be optimizing for any accessory volume V_0 and normalization constant V_ν .

We use the dataset from [12], where EPSPs were recorded in several consecutive trials for each of 637 synapses from the cortex of several young rats to calculate the optimizing cost function. We rely on the assumption that information stored at a synapse, x , can be identified with the mean EPSP amplitude. Then, the conditional density, $p_{Y|X}(y|X = x)$, is estimated for each synapse as the distribution of EPSP amplitude across trials. The marginal density, $p_Y(y)$, is the distribution of EPSP amplitude over all trials and synapses. By substituting these estimated distributions into (15) we get a plug-in estimator of the cost function, $V(x)$, for each synapse (Figure 3). A power law with exponent 0.48 provides a satisfactory fit.

In this section we provided a reverse characterization for Optimization Principles 2 and 3. In fact we essentially assumed

the results of Optimization Principles 2 and 3, and saw that the cost function for which these principles hold falls neatly into the power law cost function that we had initially formalized as (1).

VII. THEORETICAL PREDICTIONS AND COMPARISON TO EXPERIMENT

Now we compare the theoretical predictions with known experimental facts and also suggest further experimental tests of the theory. This is the final step in the scientific method.

By considering an AWGN channel model, we found that information storage is optimized by synapses with average volume given by the geometric mean of V_ν and V_0 . Although storage capacity increases in the limit $V_0 \rightarrow 0$, the conduction time delay grows without bound. Although delay was not explicitly considered in our performance metrics, it is certainly an important issue. Since the minimum conduction time delay is achieved when synaptic volume is of the order of the accessory volume [24], [25], the average synaptic volume should be less than or equal to V_ν . Therefore, typical synapses should be noisy, in agreement with experimental data where the SNR is found to one or less.

We argued that optimal information storage requires sparseness of synaptic connectivity and predicted a relationship between the filling fraction and the relative volume occupied by synapses and accessory volume. To make a quantitative comparison with empirical observations, we use experimental data from a mouse cortical column. Potential synaptic connectivity in a cortical column is all-to-all, so axons and dendrites of any two neurons pass sufficiently close to each other that they can be connected through local synaptogenesis. The fraction of potential synapses converted into actual synapses in a mouse cortex is ~ 0.3 . Taking $f = 0.3$, by (12), we find $2\lambda V_\nu = -\ln 0.3 = 1.2$ and we find that $\lambda V_0 = -\ln 0.7 = 0.36$. Then, the average volume of actual synapse is of the same order as the accessory volume per actual synapse V_0/f , in agreement with experiment. A more detailed calculation shows that actual synapse volume should be about 40% greater than accessory volume for an actual synapse. In reality, accessory volume is greater than synapse volume. This may be a consequence of minimizing conduction delays. Hopefully, an optimization framework combining conduction delays and information storage capacity will emerge in the future.

We predicted that synaptic volume should follow an exponential distribution with parameter λ . This prediction can be tested experimentally by measuring the volume of spine heads and boutons in cortical neuropil through imaging experiments. If the filling fraction in the same neuropil is measured, the theory can be tested in a way that involves no parameter fitting.

In Section V, we predicted the distribution of synaptic weight for arbitrary values of the volume-SNR exponent α (13), which can be compared to the experimentally observed synaptic weight distribution obtained in neocortical layer 5 neurons [12]. By performing a least squares fit of the logarithm of the EPSP distribution we find that the distribution is a

stretched exponential with exponent 0.49. Recall that a least squares fit of A_N as a function of mean EPSP amplitude yields a power law with exponent 0.38. Hence, $x/A_N \sim x^{0.62}$, and from (13), $\alpha = 0.49/0.62 = 0.79$.

In Section VI, we had established a link from the distribution of synaptic weights and noise statistics to the synaptic cost function. The best power-law fit to the points in Figure 3 yields a sublinear cost function with power law exponent ~ 0.48 . Recalling that $x/A_N \sim x^{0.62}$, we find that $\alpha = 0.48/0.62 = 0.77$. This estimate is consistent with that obtained from the discrete states model in the previous paragraph, thus validating the use of that model to approximate the continuous distribution of synaptic weights. The prediction of α can be tested directly by measuring the relationship between synaptic volume and weight through a joint imaging–electrophysiology experiment and would be a major test of our theory.

VIII. CONCLUSION

We have applied information theoretic characterizations of optimal information storage to generate a unified theory of structural characteristics of the mammalian brain. The main tool used was that the distribution of synaptic strengths should be the capacity-achieving input distribution for the neural memory channel. The predictions and explanations are of a purely informational and physical nature; specific implementation issues such as error control codes or neural networks have been abstracted. Independent of the coding mechanism, however, the predictions and explanations that we make are required for optimal performance.

One might wonder how robust the prediction of a capacity-achieving input distribution is and whether a distribution close to the optimal has good performance. Since the space of distributions is not compact and information is not a continuous functional, it is not clear that a distribution close to the capacity achieving one achieves mutual information close to capacity. Nevertheless this is true under sufficiently broad assumptions [29]. Similarly, it can be shown that if the true cost function is close to the optimizing cost function, then system performance is close to optimal [7]. Thus predictions from the optimization principles are robust to uncertainty.

ACKNOWLEDGMENT

We thank Per Jesper Sjöström, a co-author of [6], who was key in tying theory to experiment.

REFERENCES

- [1] M. Piccolino, “Animal electricity and the birth of electrophysiology: The legacy of Luigi Galvani,” *Brain Res. Bull.*, vol. 46, no. 5, pp. 381–407, July 1998.
- [2] J. A. V. Bates, “Significance of information theory to neurophysiology,” in *London Symposium on Information Theory*, Sept. 1950, printed as IRE PGIT-1, Feb. 1953.
- [3] G. A. Miller, “Human memory and the storage of information,” *IRE Trans. Inform. Theory*, vol. IT-2, no. 3, pp. 129–137, Sept. 1956.
- [4] F. Rieke, D. Warland, R. de Ruyter van Steveninck, and W. Bialek, *Spikes: Exploring the Neural Code*. Cambridge, MA: The MIT Press, 1997.
- [5] G. A. Parker and J. Maynard Smith, “Optimality theory in evolutionary biology,” *Nature*, vol. 348, no. 6296, pp. 27–33, Nov. 1990.
- [6] L. R. Varshney, P. J. Sjöström, and D. B. Chklovskii, “Optimal information storage in noisy synapses under resource constraints,” *Neuron*, vol. 52, no. 3, pp. 409–423, Nov. 2006.
- [7] L. R. Varshney, “Optimal information storage: Nonsequential sources and neural channels,” Master’s thesis, Massachusetts Institute of Technology, June 2006.
- [8] S. Ramón y Cajal, *La Textura del Sistema Nerviosa del Hombre y los Vertebrados*. Springer, 1899.
- [9] C. Allen and C. F. Stevens, “An evaluation of causes for unreliability of synaptic transmission,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 91, no. 22, pp. 10380–10383, Oct. 1994.
- [10] R. J. Sayer, M. J. Friedlander, and S. J. Redman, “The time course and amplitude of EPSPs evoked at synapses between pairs of CA3/CA1 neurons in the hippocampal slice,” *J. Neurosci.*, vol. 10, no. 3, pp. 826–836, Mar. 1990.
- [11] C. Holmgren, T. Harkany, B. Svennenfors, and Y. Zilberter, “Pyramidal cell communication within local networks in layer 2/3 of rat neocortex,” *J. Physiol.*, vol. 551, no. 1, pp. 139–153, Aug. 2003.
- [12] S. Song, P. J. Sjöström, M. Reigl, S. Nelson, and D. B. Chklovskii, “Highly nonrandom features of synaptic connectivity in local cortical circuits,” *PLoS Biol.*, vol. 3, no. 3, pp. 0507–0519, Mar. 2005.
- [13] M. A. Lynch, “Long-term potentiation and memory,” *Physiol. Rev.*, vol. 84, no. 1, pp. 87–136, Jan. 2004.
- [14] R. G. M. Morris, “Long-term potentiation and memory,” *Proc.-R. Soc. Lond., Biol. Sci.*, vol. 358, no. 1432, pp. 643–647, Apr. 2003.
- [15] M. Matsuzaki, G. C. R. Ellis-Davies, T. Nemoto, Y. Miyashita, M. Iino, and H. Kasai, “Dendritic spine geometry is critical for AMPA receptor expression in hippocampal CA1 pyramidal neurons,” *Nat. Neurosci.*, vol. 4, no. 11, pp. 1086–1092, Nov. 2001.
- [16] M. Matsuzaki, N. Honkura, G. C. R. Ellis-Davies, and H. Kasai, “Structural basis of long-term potentiation in single dendritic spines,” *Nature*, vol. 429, no. 6993, pp. 761–766, June 2004.
- [17] Q. Zhou, K. J. Homma, and M.-M. Poo, “Shrinkage of dendritic spines associated with long-term depression of hippocampal synapses,” *Neuron*, vol. 44, no. 5, pp. 749–757, Dec. 2004.
- [18] H. Markram, J. H. R. Lübke, M. Frotscher, A. Roth, and B. Sakmann, “Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex,” *J. Physiol.*, vol. 500, no. 2, pp. 409–440, Apr. 1997.
- [19] S. Verdú, “On channel capacity per unit cost,” *IEEE Trans. Inform. Theory*, vol. 36, no. 5, pp. 1019–1030, Sept. 1990.
- [20] S. Shamai (Shitz) and S. Verdú, “The empirical distribution of good codes,” *IEEE Trans. Inform. Theory*, vol. 43, no. 3, pp. 836–846, May 1997.
- [21] F. R. Kschischang and S. Pasupathy, “Optimal nonuniform signaling for Gaussian channels,” *IEEE Trans. Inform. Theory*, vol. 39, no. 3, pp. 913–929, May 1993.
- [22] C. E. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, July/Oct. 1948.
- [23] M. Gastpar, B. Rimoldi, and M. Vetterli, “To code, or not to code: Lossy source-channel communication revisited,” *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1147–1158, May 2003.
- [24] D. B. Chklovskii, T. Schikorski, and C. F. Stevens, “Wiring optimization in cortical circuits,” *Neuron*, vol. 34, no. 3, pp. 341–347, Apr. 2002.
- [25] Q. Wen and D. B. Chklovskii, “Segregation of the brain into gray and white matter: A design minimizing conduction delays,” *PLoS Comput. Biol.*, vol. 1, no. 7, pp. 0617–0630, Dec. 2005.
- [26] C. E. Shannon, “Coding theorems for a discrete source with a fidelity criterion,” in *IRE National Convention Record, Part 4*, 1959, pp. 142–163.
- [27] W. L. Root, “Estimates of ϵ capacity for certain linear communication channels,” *IEEE Trans. Inform. Theory*, vol. IT-14, no. 3, pp. 361–369, May 1968.
- [28] S. Verdú, “Spectral efficiency in the wideband regime,” *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.
- [29] R. L. Dobrushin, “Mathematical problems in the Shannon theory of optimal coding of information,” in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, J. Neyman, Ed., vol. 1. Berkeley: University of California Press, 1961, pp. 211–252.