

Optimal Information Storage in Noisy Synapses under Resource Constraints

Viewpoint

Lav R. Varshney,^{1,2} Per Jesper Sjöström,³ and Dmitri B. Chklovskii^{2,*}

¹Department of Electrical Engineering and Computer Science

Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

²Cold Spring Harbor Laboratory

Cold Spring Harbor, New York 11724

³Wolfson Institute for Biomedical Research and

Department of Physiology

University College London

London WC1E 6BT

United Kingdom

Summary

Experimental investigations have revealed that synapses possess interesting and, in some cases, unexpected properties. We propose a theoretical framework that accounts for three 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. We also comment on the possibility that synaptic weights may vary in discrete steps. Our approach is based on maximizing information storage capacity of neural tissue under resource constraints. Based on previous experimental and theoretical work, we use volume as a limited resource and utilize the empirical relationship between volume and synaptic weight. Solutions of our constrained optimization problems are not only consistent with existing experimental measurements but also make nontrivial predictions.

Introduction

As synapses play central roles in the two principal tasks of the brain, information processing and information storage (Ramón y Cajal, 1899), their properties have been the subject of extensive experimentation. Out of many important synaptic properties revealed over the years, we focus on the following three. First, synaptic connectivity is sparse not only in the brain in general, but also in local circuits. In other words, the probability of finding a synaptic connection between a randomly chosen pair of excitatory neurons, even nearby neurons, is much less than 1 (Holmgren et al., 2003; Isope and Barbour, 2002; Markram, 1997; Markram et al., 1997; Mason et al., 1991; Song et al., 2005; Thomson and Bannister, 2003; Thomson et al., 2002). Second, typical central synapses are noisy devices. Due, for example, to probabilistic transmitter release, firing of the presynaptic neuron occasionally fails to evoke an excitatory postsynaptic potential (EPSP). Moreover, the amplitude of the EPSP varies from trial to trial (Allen and Stevens, 1994; Hessler et al., 1993; Isope and Barbour, 2002;

Mason et al., 1991; Raastad et al., 1992; Rosenmund et al., 1993; Sayer et al., 1990). Third, while the majority of synaptic weights are relatively weak (mean EPSP < 1 mV), the weight distribution is broad with a notable tail of stronger connections (Holmgren et al., 2003; Isope and Barbour, 2002; Markram et al., 1997; Mason et al., 1991; Sayer et al., 1990; Sjöström et al., 2001).

Although a unified theoretical framework capable of accounting for all of these properties does not exist, several of these properties have been addressed previously. The noisiness of typical central synapses seemed particularly puzzling since synapses act as conduits of information between neurons (Koch, 1999). Several theoretical studies have considered the impact of synaptic noise on information transmission through a synapse, generally in the context of sensory processing (Goldman, 2004; Levy and Baxter, 2002; Manwani and Koch, 2000; Zador, 1998). These papers have shown that, under some conditions or under some constraints, synaptic noisiness facilitates the efficiency of information transmission. Moreover, Laughlin et al. (1998) have pointed out that splitting information and transmitting it over several less reliable but metabolically cheaper channels reduces energy requirements. Adding information channels invokes costs associated with building and maintaining those channels (Levy and Baxter, 1996; Schreiber et al., 2002), which must also be taken into account (Sarpeshkar, 1998). In a separate line of inquiry, Brunel et al. (2004) explain the sparseness of synaptic connectivity and the distribution of synaptic weights in the cerebellum by maximizing the storage capacity of a perceptron network.

Here, we account for several properties of central synapses by developing a theoretical framework based on the role of synapses as mechanisms of information storage, rather than their dual role in transmitting information between neurons. It is widely believed that long-term memories are recorded in neuronal circuits through alteration in the strength of existing synapses (Hebb, 1949; McGaugh, 2000), through long-term potentiation (LTP) and long-term depression (LTD) (Lynch, 2004; Morris, 2003). Memories are retrieved by electrical activity of neurons that “reads out” the pattern of synaptic connectivity between them. Thus a synaptic memory system can be viewed as a communication channel from the present to the future (Figure 1). Although information storage is well recognized as a case of a general communication system (Csiszár and Körner, 1997; Eldridge, 1963; Imminck et al., 1998) and information theory has been successfully applied in neuroscience (Rieke et al., 1997), the application of information theory to the analysis of synapses as memory elements has received little attention.

Our theoretical analysis is based on maximizing information storage capacity of synapses under resource constraints. Generally, information storage capacity of a system depends on the signal-to-noise ratio (SNR); in the case of synapses, this is a ratio between average synaptic weight and average noise. It would seem that the best strategy for increased information storage

*Correspondence: mitya@cshl.org

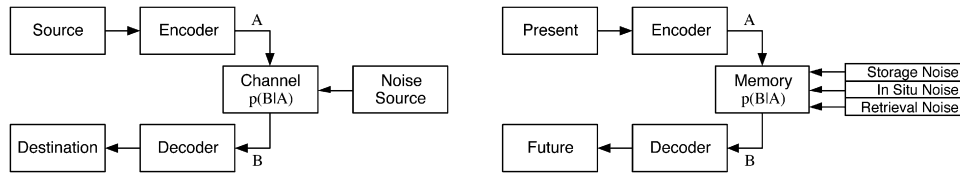


Figure 1. Information Theoretic Models of Communication and Memory

(Left) Shannon's schematic diagram of a general communication system (Shannon, 1948). Here, incoming information is denoted A , whereas the information transmitted through and distorted by the channel is labeled B . (Right) Schematic diagram of a memory system cast as a communication system. In this view, information is stored in the input variable A , while the retrieved value—which is corrupted by the noise of the system—is indicated by B . The various sources of noise 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.

capacity would be to increase the synapse SNR; however, this increase in SNR comes at a cost. For given noise, increasing the SNR requires increasing the average synaptic weight. But the weight of an individual synaptic contact is positively correlated with its volume (comprising the combined volume of the spine head and axonal bouton) (Kasai et al., 2003; Lisman and Harris, 1993; Matsuzaki et al., 2001; Murthy et al., 2001; Nusser et al., 1998; Schikorski and Stevens, 1997; Takumi et al., 1999; Tanaka et al., 2005). As the weight of a synaptic connection is composed of the weights of individual synaptic contacts and its volume is the sum of contact volume, the correlation between the weight and the volume should hold for the synaptic connection as a whole. The volume, however, is a costly resource (Cherniak et al., 1999; Chklovskii, 2004; Hsu et al., 1998; Mitchison, 1991; Ramon y Cajal, 1899). Thus, information storage capacity should be maximized under volume constraint.

Here, we cast the problem of long-term memory into the framework of information theory and deduce structural and connectivity properties of synapses that lead to optimal performance. Optimal performance is defined as the maximization of the brain's information storage capacity under constrained cost, quantified by synaptic volume. Note that we consider memory storage from a physical perspective, looking at the information storage density of neural tissue. Other than presupposing that volume is a constrained resource, our approach requires no specific assumptions regarding the network itself, such as a particular network organization or certain activity patterns. Previous work, however, examined the memory storage capacity of particular neural network models (Brunel et al., 2004; Gardner, 1987; McEliece et al., 1987; Newman, 1988; Rolls and Treves, 1998). That approach is different, since it assumes specific network designs and properties, thereby providing results that are in part due to the a priori assumptions.

The paper is organized as follows. In section I, we specify our model and formalize the empirical relationship between synaptic weight and synaptic volume. This preliminary step allows a quantitative tradeoff between storage capacity and cost. In section II, we show that synaptic connections should, on average, have small volume and consequently be noisy to maximize information storage capacity per unit volume. In section III, we determine the distribution of synaptic weights—not just the average structural property of synapses—that maximizes capacity for a particular synapse model. This optimal distribution includes many zero-weight

connections, or potential synapses, which is in accordance with experimental observations of sparse synaptic connectivity. In section IV, we use an experimentally determined distribution of synaptic weights, as well as synaptic noise, to compute the cost function for which the information storage system operates optimally. The solution to this inverse problem makes more precise the synaptic weight cost relationship specified in section I. In section V, we argue that discrete synaptic states may perform almost optimally or, in a slightly different yet naturally constrained model, better than continuous valued synaptic states. Finally, section VI compares our theoretical predictions with known experimental data and suggests further tests of the theory.

Results

I. Relationship between Synaptic Weight and Volume

We start by formulating the model of a synapse in the information storage context, which is based on existing experimental observations. Although synaptically connected pairs of cortical neurons usually share multiple synaptic contacts (Kalisman et al., 2005; Koester and Johnston, 2005; Markram et al., 1997; Silver et al., 2003), here, unless specified otherwise, we refer to these contacts collectively as a synapse. Such a definition is motivated by electrophysiological measurements, which record synaptic weight of all the contacts together.

We assume that information is stored in the synaptic weight, A , of each synapse. The weight can be obtained by averaging EPSP amplitude measured in multiple trials in response to the firing of a presynaptic neuron. Then the standard deviation of the EPSP amplitude from trial to trial in a given synapse is the noise amplitude, A_N . As one might expect from the Poisson model of synaptic release (Bekkers and Stevens, 1995; del Castillo and Katz, 1954), the noise amplitude increases sublinearly with the synaptic weight. Recent measurements suggest a power law with the exponent about 0.38 (Markram et al., 1997; Song et al., 2005) (see Figure S1 in the Supplemental Data available online).

The volume of a synapse is composed of individual synaptic contact's volume, which, in turn, correlates with the contribution of each synaptic contact to synaptic weight as suggested by the following experimental results. Anatomically, individual synaptic contact's volume correlates with many ultrastructural characteristics, such as the number and area of active zones, number

of vesicles, area of the postsynaptic density, and the number of receptors (Lisman and Harris, 1993; Murthy et al., 2001; Nusser et al., 1998; Pierce and Mendell, 1993; Schikorski and Stevens, 1997; Streichert and Sargent, 1989; Takumi et al., 1999; Tanaka et al., 2005; Yeow and Peterson, 1991). Physiologically, individual synaptic contact's volume correlates with the synaptic weight (Kasai et al., 2003; Matsuzaki et al., 2001). In fact, an increase in synaptic spine volume may sometimes accompany LTP (Matsuzaki et al., 2004; Kopec et al., 2006), whereas LTD may result in the converse volume decrease (Zhou et al., 2004).

Because contributions of individual contacts to synaptic weight may add up linearly (Cash and Yuste, 1999), the volume of a synapse correlates with the synaptic weight. Indeed, a neuron can be viewed as a single computational unit (Chklovskii et al., 2004), as there is evidence that multiple synaptic contacts within a connected pair of neurons have correlated release probability (Koester and Johnston, 2005) and that the total synaptic connection weight correlates with the number of synaptic contacts (Kalisman et al., 2005). Alternatively, the integrative compartment may be smaller—such as a single dendritic branch (Poirazi et al., 2003; Polsky et al., 2004)—and individual synaptic contacts may in addition vary their weights independently. In this case, our model would have to be modified.

As the noise amplitude, A_N , is related by a power law to the mean EPSP amplitude, A , which is strongly correlated with the synapse volume, V , we can formulate the following scaling relationship:

$$\frac{V}{V_N} = \left(\frac{A}{A_N} \right)^\alpha \quad (1.1)$$

where V_N is the volume of a synapse with an SNR of 1. Although existing experimental measurements (Kasai et al., 2003; Matsuzaki et al., 2001; Murthy et al., 2001; Schikorski and Stevens, 1997; Song et al., 2005; Takumi et al., 1999; Tanaka et al., 2005) support Equation 1.1, they are not sufficient to establish the value of the exponent, α .

II. Noisy Synapses Maximize Information Storage Capacity

In this section, we deduce optimal average synaptic weight and volume by maximizing information storage capacity per unit volume. We invoke the synaptic weight/volume relationship formulated in the previous section (Equation 1.1) with $\alpha = 2$; other cases will be considered in later sections. For $\alpha = 2$, the problem of maximizing information storage capacity in a given volume reduces to the well-studied problem of maximizing channel capacity for a given input power. When the channel contributes additive white Gaussian noise (AWGN), such a problem is exactly solvable.

In our context, information is stored in the alteration of synaptic weights and retrieved by electrical activity. Then each synapse corresponds to a channel usage with information encoded in its weight. Maximum storage capacity is achieved when synaptic weights are uncorrelated. The retrieval noise is manifested in fluctuations of EPSP from trial to trial. For concreteness, we assume here that the noise is Gaussian with a given var-

iance; we will argue at the end of this section that the conclusions hold for other noise models.

Information storage capacity per synapse (measured in nats rather than bits) is given by the expression derived by Shannon (1948) for the AWGN channel:

$$I_{\text{synapse}} = \frac{1}{2} \ln \left(1 + \left\langle \frac{A^2}{A_N^2} \right\rangle \right) \quad (2.1)$$

where

$$\left\langle \frac{A^2}{A_N^2} \right\rangle$$

is the average SNR among synapses. SNR for each synapse is defined as the square of the mean EPSP amplitude divided by the trial-to-trial variance of EPSP amplitude. (In the AWGN model, A can take negative as well as positive values.) Using Equation 1.1, we can rewrite information storage capacity in terms of volume:

$$I_{\text{synapse}} = \frac{1}{2} \ln \left(1 + \frac{\langle V \rangle}{V_N} \right) \quad (2.2)$$

where $\langle V \rangle$ is the average synapse volume.

As volume is a scarce resource, information storage capacity is likely to be optimized on a per-volume basis (see Introduction). For example, placing two or more smaller synapses (connecting different pairs of neurons) in the place of one larger synapse may increase memory capacity. Then the total storage capacity of a unit volume of neural tissue is

$$I_{\text{volume}} = \frac{I_{\text{synapse}}}{\langle V \rangle + V_0} = \frac{1}{2(\langle V \rangle + V_0)} \ln \left(1 + \frac{\langle V \rangle}{V_N} \right) \quad (2.3)$$

where V_0 is the accessory volume necessary to support a synapse. Accessory volume includes the volume of wiring (axons and dendrites), glia, and perhaps extracellular space. Information storage capacity as a function of the size of the synapse, the relationship in Equation 2.3, is plotted in Figure 2A for different values of V_0 .

Optimal storage capacity is achieved at the maximum of the I_{volume} -versus- $\langle V \rangle / V_N$ curve in Figure 2A. The maximum can be found by setting the derivative of Equation 2.3 to zero as described in the Experimental Procedures. Figure 2B shows the dependence of information storage capacity I_{volume} (peak height in Figure 2A) and optimal synaptic volume $\langle V \rangle$ (horizontal coordinate of the peak in Figure 2A) on the accessory volume V_0 . 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_N and—according to Equation 1.1—synapses should therefore be noisy.

In reality, the accessory volume may not be infinitesimal, as this would affect system functionality adversely. For example, there is a hard limit on how thin axons can be (Faisal et al., 2005). Also, reducing wiring volume may increase conduction time delays and signal attenuation (Chklovskii et al., 2002). In fact, delay and attenuation are optimized when the wiring volume is of the same order as the volume of synapses (Wen and Chklovskii, 2005), which happens when they are of the order of V_N . Then the optimal performance—in terms of jointly

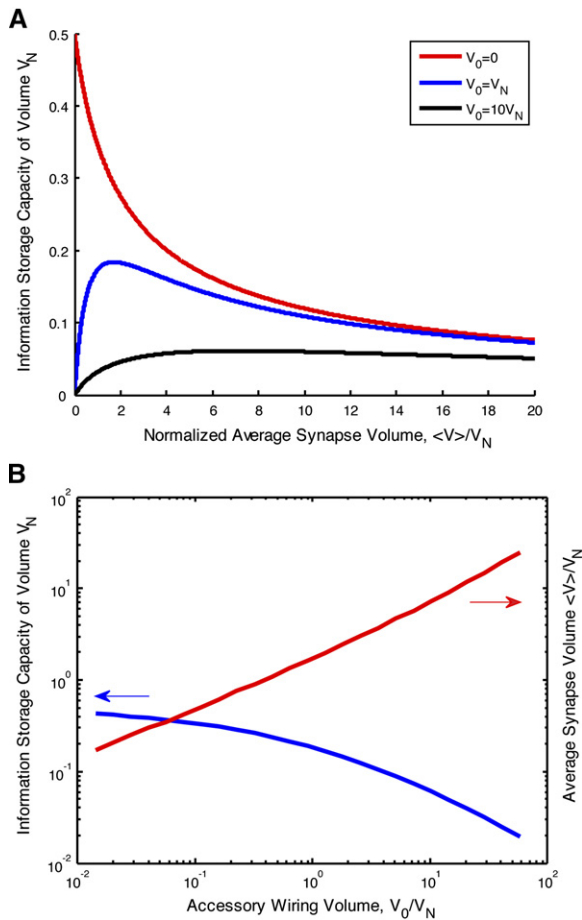


Figure 2. Results of the AWGN Model

(A) Information storage capacity per volume, V_N , of neural tissue as a function of normalized average synapse volume. The relationship between signal-to-noise ratio and volume for this plot uses accessory volume, V_0 , values of 0, 1, and 10, normalized with respect to V_N . When $V_0 = 0$, the maximum storage capacity per unit volume occurs when average synapse volume is infinitesimal. When $V_0 > 0$, the finite maximum storage capacity per unit volume occurs at some non-zero normalized synapse volume.

(B) Blue line: Maximum information storage capacity per volume V_N of neural tissue (vertical coordinate of the peak in [A]) as a function of the accessory volume V_0 . Red line: the corresponding average synapse volume (V) (horizontal coordinate of the peak in [A]).

maximizing information storage and minimizing conduction time delay and attenuation—is achieved when average synapse volume (V) is either less than or of the order of V_N . 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 that was considered above, but also for many reasonable noise and cost models. For these other models, individual synapse channel capacity, I_{synapse} , is nondecreasing and logarithmic in SNR. Thus, the inversely linear $\langle V \rangle$ term that arises from the number of synapses in the unit volume outpaces the logarithmic $\langle V \rangle$ term that arises from individual synapse capacity, and so total capacity decreases with increasing $\langle V \rangle$ for large $\langle V \rangle$.

An alternate way to see that the advantage of having greater numbers of smaller synapses extends to other

reasonable noise models is through the concavity of the capacity cost function of information theory. The capacity cost function generalizes channel capacity by imposing average cost constraints on the channel inputs. Like channel capacity, it is the maximum rate at which one can transmit information over a channel while still achieving arbitrarily small probability of error; however, now the optimization is constrained by cost. This function is nondecreasing and concave downward (McEliece, 1977; Shannon, 1959), which means that the slope (capacity/cost) is larger at lower costs. If there are no zero cost symbols, the capacity per unit cost is maximized at the average cost for which a line constrained to pass through the origin has its point of tangency to the capacity cost function. Such tangency points correspond to the location of the peaks in Figure 2A. If there is a zero cost symbol, however, then the optimum is for zero average cost ($V_0 = 0$ curve in Figure 2A). In many cases, it is difficult to find the optimum capacity per unit cost analytically (Verdú, 1990). There is, however, a numerical algorithm that can be used for such a computation (Jimbo and Kunisawa, 1979). Similar mathematical arguments have been used in the context of information transmission to show that having parallel, less reliable channels—such as synapses (Laughlin et al., 1998) and ion channels (Schreiber et al., 2002)—reduces metabolic costs.

In this section, we showed that—provided the accessory volume needed to support a synapse is small—numerous small and noisy synapses possess greater information storage capacity per unit volume than a few large and reliable synapses. This result may help explain why central synapses typically are unreliable (Allen and Stevens, 1994; Hessler et al., 1993; Isope and Barbour, 2002; Mason et al., 1991; Raastad et al., 1992; Rosenmund et al., 1993; Sayer et al., 1990).

III. Optimal Distribution of Synaptic Weights in the Discrete-States Model

Having established that synapses should be small and noisy on average, we next examine how volume and EPSP amplitude should be distributed among synapses. In the AWGN model used in section II, the capacity-achieving input distribution is also Gaussian (Shannon, 1948), and the synaptic volume is distributed exponentially. If the noise amplitude A_N is constant, synaptic weight has a Gaussian distribution, as previously suggested (Brunel et al., 2004). If, on the other hand, A_N scales as a power of A (Figure S1), the synaptic weight distribution is a stretched (or compressed) exponential. Here, exponential and Gaussian distributions are two different, special cases.

However, it is not clear whether these predictions from the AWGN model can be taken at face value. First, the AWGN channel model allows both negative and positive signals, whereas synaptic weight is positive for excitatory synapses. Second, the Gaussian noise assumption is unlikely to hold, especially if synaptic weight must be non-negative. Third, synaptic volume may not scale as the synaptic weight SNR squared.

We therefore consider a different, discrete-states model, where the cost function can be chosen arbitrarily and the synaptic weight is non-negative, but which still yields an exactly solvable optimization problem. The

reason an exact solution can be found is that the noise is treated approximately. Rather than considering a continuous distribution of synaptic weights, we 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. Then the difference in synaptic weight between adjacent states A_i and A_{i+1} is given by the two noise amplitudes, $A_N(A_i) + A_N(A_{i+1})$. From the information theoretic point of view, each state is viewed as a symbol from an alphabet characterized by a different cost (Figure 3A).

Such conversion of the noisy continuous-valued input channel into a zero-error, discrete-valued input channel is a convenient approximation (Kolmogorov and Tihomirov, 1959; Root, 1968) because the mutual information of the noiseless channel reduces to the self-information of the channel input distribution (or, equivalently, channel output distribution). By resorting to this approximation, we do not wish to imply that synaptic weights in the brain necessarily vary in discrete steps. In section VI, we will validate this approach by comparing its predictions to the predictions from a continuous channel model (section IV).

Since the self-information is identical to entropy, the maximization of information storage capacity per volume reduces to entropy maximization per volume, a standard problem from statistical physics (see Experimental Procedures). In neuroscience, such a mathematical problem has been solved in the context of neuronal communication by the spike code (Balasubramanian et al., 2001; de Polavieja, 2002, 2004). We consider a set of synaptic states, i , characterized by the EPSP amplitudes, A_i , and volume (or some other generalized cost), V_i (Figure 3A). We search for the probability distribution over synaptic states, p_i , that maximizes information storage capacity (measured in nats):

$$I_{synapse} = - \sum_i p_i \ln p_i \quad (3.1)$$

per average volume of a synapse \bar{V}

$$\bar{V} = \sum_i p_i V_i \quad (3.2)$$

Note that the average synaptic volume, \bar{V} , includes the accessory volume, V_0 , which was excluded from the definition of $\langle V \rangle$ used in the previous section.

We show in the Experimental Procedures that the probability distribution over synaptic states, p_i , that maximizes information capacity per volume is given by

$$p_i = \exp(-\beta V_i) \quad (3.3)$$

where β is defined by the condition

$$\sum_i p_i = 1$$

and

$$I_{volume} = \frac{I_{synapse}}{\bar{V}} = \frac{\beta \sum_i V_i \exp(-\beta V_i)}{\sum_i V_i \exp(-\beta V_i)} = \beta \quad (3.4)$$

Motivated by experimental observations (Kopec et al., 2006), we assume that synaptic state volume is distrib-

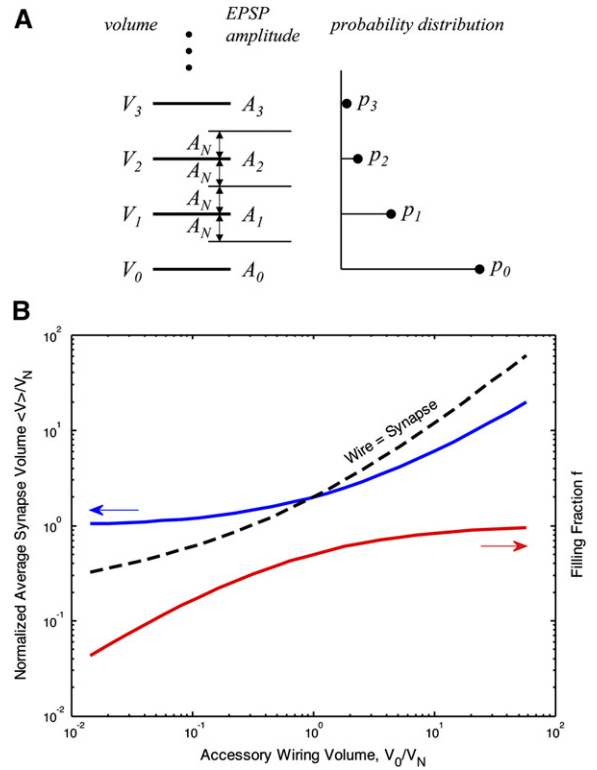


Figure 3. Discrete-States Model

(A) Synapses are modeled by a set of discrete noiseless synaptic states characterized by mean EPSP amplitude A_i and volume V_i . The difference in synaptic weight between adjacent states is $A_N(A_i) + A_N(A_{i+1})$.

(B) The average volume of actual synapses $\langle V \rangle_{i>0}$ (blue line) and the fraction of synapses in $i > 0$ states, i.e., the filling fraction, $f = (1 - p_0)$ (red line) as a function of accessory synapse volume normalized by V_N . Dashed line corresponds to the equipartition of volume into synapses and wires. The solution satisfying competing requirements of maximizing information storage capacity and minimizing conduction delays must be at $f < 0.5$.

uted equidistantly, i.e., the volume of the i th synaptic state is given by

$$V_i = V_0 + 2iV_N \quad (3.5)$$

Then the average volume per potential synapse (including accessory volume, V_0), defined as the total volume divided by the number of potential synapses (including actual ones), can be expressed analytically (see Experimental Procedures) as:

$$\bar{V} = V_0 + 2V_N \exp(\beta(V_0 - 2V_N)) \quad (3.6)$$

To allow comparison with empirical measurements (section VI), we also derive an expression for the average volume of actual synapses (see Experimental Procedures), i.e., states with $i > 0$, and excluding accessory volume, V_0 (Figure 3B):

$$\langle V \rangle_{i>0} = 2V_N \exp(\beta V_0) \quad (3.7)$$

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 bigger synapses that have higher SNR.

The ratio between the number of actual synapses and the number of potential synapses (including actual) is called the filling fraction, f . In our model the filling fraction is just the fraction of synapses in states $i > 0$ and is given by (see [Experimental Procedures](#)):

$$f = \exp(-2\beta V_N) \quad (3.8)$$

and plotted in [Figure 3B](#).

Information storage capacity per volume can be calculated using [Equations 3.4 and 3.6](#). Just as in the AWGN model, information storage capacity increases monotonically with decreasing accessory volume. Unlike the AWGN model, the growth of information storage capacity is unlimited. As information storage capacity diverges with decreasing accessory volume, V_o , optimal information storage is achieved when V_o is as small as possible. In this limit, the filling fraction, f , is much less than 1, as illustrated in [Figure 3B](#). This prediction is consistent with empirical observations of sparse connectivity. In addition, according to [Figure 3B](#) most actual synapses have volume $2V_N$, and thus have SNR of order 1 ([Equation 1.1](#)). This prediction is in agreement with the experimentally established noisiness of typical synapses.

Although local cortical circuits are sparse and typical synapses are noisy, the filling fraction is not infinitesimal. One explanation for this fact—which was discussed in the previous section—is that very small V_o affects system functionality adversely. The condition that accessory wire volume is of the order of synapse volume ([Wen and Chklovskii, 2005](#)) implies that $\langle V \rangle \sim V_o/f$. This condition is illustrated in [Figure 3B](#) by a dashed line intersecting the blue line. Then the competing desiderata of maximizing information storage and minimizing conduction delays should yield a value of V_o less than at the intersection. The corresponding filling fraction is less than half, but not infinitesimal.

By using [Equations 3.3 and 1.1](#), we can find the probability of synaptic states in terms of the EPSP amplitude,

$$p_i = \exp(-\beta(V_o + V_N(A_i/A_N)^\alpha)) \quad (3.9)$$

Such a distribution is called a stretched (or compressed) exponential and is compared with experimental data in section VI. In the continuum limit, when the probability changes smoothly between states, we can convert [Equation 3.9](#) to the probability density. Considering that there should be one synaptic state per two noise amplitudes, $2A_N$, the probability density of the EPSP distribution is given by

$$p(A) = \frac{1}{2A_N(A)} \exp(-\beta(V_o + V_N(A/A_N)^\alpha)) \quad (3.10)$$

Interestingly, the explicit consideration of noise does not alter the result, which follows from [Equation 3.3](#), that for $V_o/V_N \rightarrow 0$ optimum information storage is achieved by using mostly the $i = 0$ state, with $i = 1$ used with exponentially low frequency. If $V_o = 0$, this type of problem can be solved exactly ([Verdú, 1990, 2002](#)), and the information storage capacity is maximized when, in addition to the zero cost symbol, only one other symbol is chosen. The additional symbol is chosen to maximize the Kullback-Leibler (KL) divergence between conditional probabilities of that symbol and of the zero cost symbol divided by the cost of the

additional symbol. If $V_o > 0$, however, the problem of optimizing information storage capacity cannot be solved analytically, prompting us to pursue a reverse approach discussed in the next section.

IV. Calculation of the Synaptic Cost Function from the Distribution of Synaptic Weights

The problem of directly and analytically finding the capacity-achieving input distribution and the channel capacity for a specified cost function is often rather difficult and is only known in closed form for certain special cases. In most cases, the channel capacity and capacity-achieving input distribution are found using numerical algorithms ([Arimoto, 1972; Blahut, 1972](#)). In neuroscience, this algorithm was used in the context of optimal information transmission by the spike code ([Balasubramanian et al., 2001; de Polavieja, 2002, 2004](#)).

An alternative way to attack the optimization problem is to specify the channel noise distribution and the channel input distribution and then determine the channel input cost function for which the system is operating at capacity ([Csiszár and Körner, 1997; Gastpar et al., 2003](#)). This methodology does not seem to have been used for neuroscience investigations, other than for a brief look at sensory processing ([Gastpar, 2003](#)). Although this method inverts the problem specification, it seems reasonable if we are not sure of what the channel input cost function is (e.g., we do not know what α in [Equation 1.1](#) is). The result of the computation is a cost function, which may then be examined for relevance to the problem at hand.

As before, we consider memory as a communication channel ([Figure 1](#)). Information is stored in the input variable A , the retrieved (output) value of which is designated B . [Gastpar et al. \(2003\)](#) show that for a fixed channel input distribution $p(A)$ and channel noise $p(B|A)$, the system is optimal—in the sense of operating at capacity cost—if the cost function is of the form

$$V(A) = \nu D(p(B|A) || p(B)) + \nu_o \quad (4.1)$$

where $\nu > 0$ and ν_o are arbitrary constants. Furthermore, $D(\cdot || \cdot)$ denotes the KL divergence, which quantifies the difference between the two probability distributions. KL divergence is zero if and only if the two distributions are identical. Note that the computed cost function is optimal for any accessory volume cost.

First, we demonstrate that [Equation 4.1](#) is valid in the cases considered in previous sections. For the AWGN model considered in section II, the input distribution and the noise are Gaussian. Then the output distribution is Gaussian as well. By substituting these distributions into [Equation 4.1](#), we can calculate the synaptic cost function explicitly (see [Experimental Procedures](#)):

$$\begin{aligned} V(A) &\sim \int p(B|A) \ln \frac{p(B|A)}{p(B)} dB \\ &\sim \int B^2 \exp(-(B-A)^2) dB \sim A^2 \end{aligned} \quad (4.2)$$

where \sim implies equality up to affine transformation. We find that the cost function is quadratic in synaptic weight, as was initially assumed in section II, thus validating [Equation 4.1](#) for this case.

Another example that validates Equation 4.1 is the model of discrete noiseless synaptic states considered in section III. In this case, $p(B=A_i|A_j) = \delta_{ij}$ and—using Equation 4.1—we find that the cost function is given by the logarithm of the input (or, alternatively, output) distribution (see [Experimental Procedures](#)):

$$V(A_i) \sim \sum_j p(B=A_j|A_i) \ln \frac{p(B=A_i|A_j)}{p(B=A_j)} \sim -\ln(p(A_j)) \quad (4.3)$$

This is exactly what Equation 3.3 would predict, thus providing another validation for Equation 4.1.

Next, we use Equation 4.1 to calculate the synaptic cost function from experimentally measured distributions of synaptic weights and noise. We use the dataset from [Sjöström et al. \(2001, 2003\)](#), also analyzed in [Song et al. \(2005\)](#), where EPSPs were recorded in several consecutive trials for each of 637 synapses. To carry out this calculation, we rely on the assumption that information stored at a synapse, A , can be identified by the mean EPSP amplitude. Then, the conditional density, $p(B|A)$, is estimated for each synapse as the distribution of EPSP amplitudes across trials ([Figure 4A](#)). The marginal density, $p(B)$, is the distribution of EPSP amplitude over all trials and synapses. By substituting these distributions into Equation 4.1 we find estimates of the cost function, $V(A)$, for each synapse ([Figure 4B](#)). A power law with exponent 0.48 provides a satisfactory fit. Error bars are obtained from a bootstrapping procedure (see [Experimental Procedures](#)).

V. Discrete Synapses May Provide Optimal Information Storage

The models of synaptic information storage presented in sections II and IV might have given the impression that the optimal distribution of synaptic strength must be continuous. Indeed, section II modeled information storage in synapses by the AWGN channel with average power constraint, for which the optimal, capacity-achieving distribution is the continuous Gaussian distribution ([Shannon, 1948](#)). In addition, using the methods of section IV, one can construct numerous cost-constrained channels with continuous capacity-achieving distributions.

Here we suggest that discrete synaptic states may achieve optimal or nearly optimal information storage. First, we point out that, surprisingly, not all continuous input channels have optimal input distributions that are continuous. In particular, imposing a constraint on the maximum weight (or volume) of a single synapse may change the optimal, capacity-achieving distribution of synaptic weights from continuous form to a set of discrete values. Such a maximum amplitude constraint is quite natural from the biological point of view, because neither volume nor EPSP can be infinitely large. Note that, unlike in section III, where discreteness was an assumption used to simplify mathematical analysis, here the discrete solution emerges as a result of optimization.

For concreteness, we return to the AWGN channel model considered in section II, but now we impose a maximum weight constraint in addition to the average volume constraint that was originally imposed. The problem then reduces to the well-studied problem of finding channel capacity for a given average input power

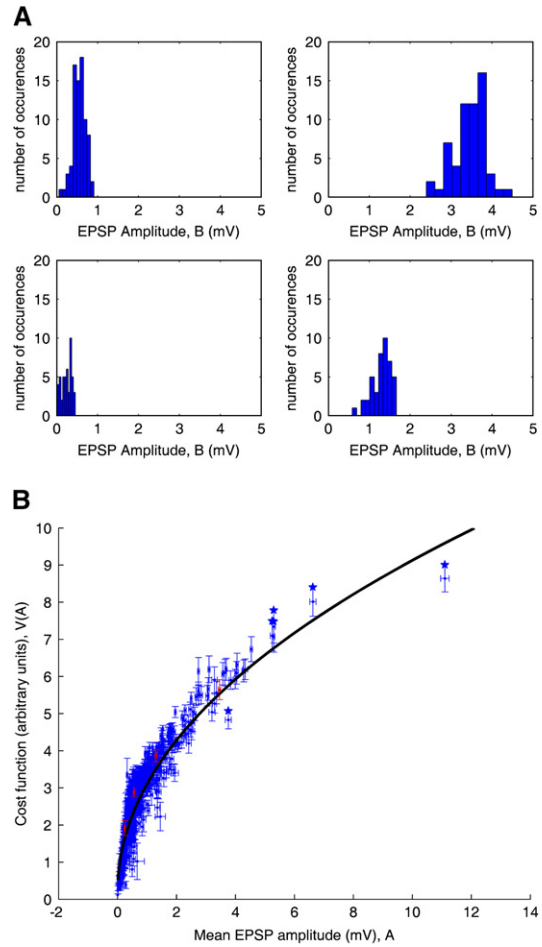


Figure 4. Synaptic Cost Function Calculated from EPSP Measurements

(A) Typical distributions of EPSP amplitude among trials for synapses characterized by different mean EPSP amplitudes.

(B) Synaptic cost function as a function of mean EPSP amplitude calculated from Equation 4.1 under assumption of optimal information storage. Each data point represents a different synapse, with those appearing in (A) highlighted in red. Horizontal error bars represent the standard error for the mean EPSP amplitude; vertical error bars represent the standard error for the KL divergence quantity in Equation 4.1. The standard error was estimated by the bootstrap procedure described in [Experimental Procedures](#). The points shown with starred upper vertical error had infinite vertical error, as estimated by the bootstrap procedure. The black line shows a least-squares power law fit with exponent 0.48.

and peak input power. For the AWGN channel, the unique optimal input distribution consists of a finite set of points. A proof of this fact is based on methods of convex optimization and mathematical analysis ([Smith, 1971](#)). Note that the Blahut-Arimoto algorithm for continuous channels is based on sampling the input space ([Blahut, 1972](#)) and cannot be used to determine whether the optimal input distribution is continuous or discrete. Consequently, an analytical proof is necessary.

Since it is known that the optimal input distribution consists of a finite number of points, one can numerically search over this sequence of finite dimensional spaces to find the locations and probabilities of these points for particular average power and peak power values. Moreover, there is a test procedure, based on

the necessity of satisfying the Karush-Kuhn-Tucker optimality conditions, to determine whether the obtained numerical solution is in fact optimal. So one can apply the numerical procedure to generate a possible solution and unmistakably recognize whether this solution is optimal (Smith, 1971). Applying Smith's optimization procedure, including both the search and the test for optimality, yields the following result for the AWGN channel. For noise power 1, symmetric peak amplitude constraint $[-1.5, 1.5]$, and input power constraint 1.125 (an SNR close to 1), the optimal input distribution consists of the zero point with large probability, and the -1.5 and 1.5 points with equal smaller probability (Smith, 1971) (see Figure S2).

The conclusion that the distribution of synaptic weights should be discrete valued holds not only for the AWGN channel with hard limits imposed on synapse size and weight, but also for other noise models. In particular, the discreteness result holds for a wide class of additive noise channels under maximum amplitude constraint (Tchamkerten, 2004). Some fading channels that have both additive and multiplicative noise and are similarly constrained (Gursoy et al., 2005) also have this discrete input property. Furthermore, channels other than AWGN with constraints on both average power and maximum amplitude have optimal input distributions that consist of a finite number of discrete points (Huang and Meyn, 2005).

A second observation is that—although there are channels that have optimal input distributions that consist of finite sets of discrete points—even channels that have continuous optimal input distributions can be used with discrete approximations of the optimal input distribution and perform nearly at capacity. It is well known that—in the average power constrained AWGN example and in the limit of small SNR—the use of an alphabet with only two symbols, $\pm(A^2)^{1/2}$, does not significantly reduce information storage capacity (Figure S3; also see Experimental Procedures). In addition, Huang and Meyn (2005) demonstrate numerically that discrete input distributions, in some cases generated by sampling the optimal continuous distribution, are only slightly suboptimal.

VI. Theoretical Predictions and Experiment

In this section, we compare theoretical predictions with known experimental data and suggest further experimental tests of the theory.

In section II, we find that, by considering an AWGN channel, information storage capacity increases as a function of accessory volume $V_0 \rightarrow 0$. When the accessory volume (V_0) is less than the volume of a synapse with unitary SNR (V_N), storage is optimized by synapses with average volume given by the geometric mean of V_0 and V_N . However, small accessory volume has a detrimental effect on the system functionality, because the conduction time delay diverges as $V_0 \rightarrow 0$. As the minimum conduction time delay is achieved when accessory volume is of the order of the synaptic volume (Chklovskii et al., 2002; Wen and Chklovskii, 2005), the competition between these requirements results in the optimal mean synaptic volume being less than or equal to V_N . This result is corroborated in the discrete-states model of section III, where optimal synaptic volume was found to be

$2V_N$. Although the noise is not explicitly represented in that model, the synapse volume is the minimum possible. These two results predict that typical synapses should be small and consequently noisy. Indeed, Isope and Barbour (2002) report an SNR of ~ 0.6 at parallel fiber synapses onto cerebellar Purkinje cells. This finding is in keeping with our prediction that the SNR should be less than 1 but not infinitesimal. More generally, our prediction is also in agreement with other experimental data showing the noisiness of typical central synapses (Allen and Stevens, 1994; Hessler et al., 1993; Isope and Barbour, 2002; Mason et al., 1991; Raastad et al., 1992; Rosenmund et al., 1993; Sayer et al., 1990).

In section III, we argue that optimal information storage requires sparseness of synaptic connectivity, and we predict a relationship between the filling fraction, f , and the relative volume occupied by synapses and wires (Equations 3.7 and 3.8). To make a quantitative comparison with empirical observations, we consider a mouse cortical column. Potential synaptic connectivity in a cortical column is all to all, meaning that axons and dendrites of any two neurons pass sufficiently close to each other that they can be connected through local synaptogenesis (Chklovskii et al., 2004; Kalisman et al., 2005; Le Be and Markram, 2006; Stepanyants and Chklovskii, 2005). According to Stepanyants et al. (2002), the fraction of potential synapses converted into actual ones in mouse cortex is ~ 0.3 ; we take this fraction to be our filling fraction, $f \sim 0.3$. By using Equation 3.8, we find that $2\beta V_N = -\ln(0.3) = 1.2$, and by using Equation EP.9, we find that $\beta V_0 = 0.36$. Then, the average volume per actual synapse is of the same order as the accessory volume per actual synapse V_0/f , in agreement with experiments (Chklovskii et al., 2002). More detailed calculation using Equation 3.7 shows that actual synapse volume should be about 40% greater than accessory volume per actual synapse. In reality, wire volume is greater than synapse volume. This may be a consequence of minimizing conduction delays as discussed in section III. Hopefully, a future optimization framework that combines conduction delays and information storage capacity will account for this discrepancy.

Does this theory apply to the global brain network beyond the cortical column? In principle, sparseness of the global network seems consistent with the high cost of wiring. However, a quantitative analysis is complicated by the fact that—for a network that does not possess potential all-to-all connectivity—the wiring cost depends not just on the numbers of synapses but also on which particular synapses are implemented. Therefore, a detailed analysis of such a network would require characterizing the cost and the information storage capacity of dendritic and axonal arbors quantitatively. This is a difficult problem, because a theory of neuronal arbors does not yet exist.

In section III, we predict that synaptic volume follows an exponential distribution with the decay constant β (Equation 3.3). This prediction can be tested experimentally by measuring the volume of spine heads and boutons in cortical neuropil. In comparing the distribution of volume, one should keep in mind that we are referring to the total volume of all synaptic contacts between two neurons (section I). In addition, if one measures the filling fraction in the same neuropil, the

test involves no fitting parameters because β can be calculated from the wiring volume and the filling fraction; from Equation EP.9 and Equation 3.8, we get that $\beta = -\log(1 - f)/V_0$. To overcome the difficulty in measuring V_N or V_0 , one can alternatively measure the experimentally accessible quantities f and $\langle V \rangle_{i > 0}$ to determine β . Then, from Equations 3.7 and 3.8, $\beta = -\log(f)/(1 - f)/\langle V \rangle_{i > 0}$. However, these predictions are only approximate, as the relative importance of maximizing information storage and minimizing conduction delays is unknown. In fact, these properties may vary depending on animal species, brain region, and animal age.

In section III, we predict the distribution of synaptic weight for arbitrary values of α (Equation 3.9), which can be compared to the experimentally observed synaptic weight distribution obtained in neocortical layer V neurons (Song et al., 2005). To perform such a comparison, we sort synaptic weights into bins $[A_i - A_N(A_i), A_i + A_N(A_i)]$ and plot a histogram (Figure S4). 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. A least-squares fit of the standard deviation of EPSP amplitude as a function of mean EPSP amplitude (Figure S1) yields a power law with exponent 0.38. Hence, $A/A_N \sim A^{0.62}$, and from Equation 3.9 we find that $\alpha = 0.49/0.62 = 0.79$.

In section IV, we established a reverse 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 4B yields a sublinear cost function with exponent ~ 0.48 (Figure 4B). Recalling that $A/A_N \sim A^{0.62}$ we find that $\alpha = 0.48/0.62 = 0.77$. This estimate is similar to that obtained using the discrete-states model, thus validating the use of that model to approximate the continuous distribution of synaptic weights (section III). The prediction of α can be tested directly by measuring the relationship between synaptic volume and weight. Such an experiment would involve jointly measuring the physical and electrophysiological properties of individual synapses. Should the relationship between synaptic weight and volume differ from that predicted in section IV, other factors may contribute to the distribution of synaptic weights.

In section V, we argue that discrete synaptic states could optimize information storage almost as well as—and under some conditions better than—synapses with continuous weights. This does not prove that synapses with discrete states are strictly optimal; it merely suggests that they could be. There is experimental evidence that changes in the weights of individual synapses are, in fact, discrete (Lisman, 2003; O'Connor et al., 2005; Petersen et al., 1998), which seems consistent with maximizing information storage. However, our model assumes the same relationship between synaptic weight and volume for all synapses, which is only approximately correct. For example, synapses that are more distant from the soma must be bigger to ensure that somatic EPSP remains the same in the face of electrotonic attenuation (Magee and Cook, 2000). Although the optimal solution is not known in this case, we speculate that even if individual synapses were to have discrete states, these states would not be the same among all the synapses. In other words, the finding that individ-

ual synapses may change in discrete steps during plasticity (Lisman, 2003; O'Connor et al., 2005; Petersen et al., 1998) would not necessarily make the overall distribution of synaptic weights discrete.

Is there any evidence of discreteness in the distribution of synaptic weights? The distribution reported in Song et al. (2005) is not monotonic; it has a maximum at around 0.2 mV. Recalling that the full distribution of synaptic states should include EPSPs of zero amplitude (absent and silent synapses), there is a gap between the zero-amplitude EPSP and the peak at around 0.2 mV (Figure S1; also see Figure 5 in Song et al., 2005), which hints that synaptic weight may be increased in discrete steps and not continuously. Finally, we note that this distribution was obtained combining data from hundreds of animals (Song et al., 2005). Even if synaptic states were discrete within one animal, such discreteness would presumably be blurred by interanimal variability.

Discussion

We have argued that maximizing information storage capacity per volume yields typical synapses that are small and hence noisy. This explains experimental observations of synaptic unreliability. From the same principle, we derived the distribution of synaptic weights and found it to be a stretched exponential, in agreement with existing measurements. This argument also explains the sparseness of synaptic connectivity. We also suggest that the discreteness of synaptic states is consistent with maximization of information storage capacity.

The strength of the information theory approach is that it provides an upper bound on information storage simply based on physical limitations, without explicitly considering how memory is stored, retrieved, coded, or decoded. For example, it is not known whether error-correcting codes are used in the brain when information is stored across numerous synapses. Regardless of whether error-correcting codes are used or not, the capacity-achieving input distribution must be used for optimal performance. Thus, our explanations and predictions stand irrespective of whether or how the brain uses error correction codes in information storage.

Nevertheless, the independence of results obtained using information theory approach on a specific implementation is also its weakness, because the impact of unknown mechanisms is difficult to assess. Although information theory provides physical limits on information storage capacity, there could be other constraints due to mechanisms of storage and read-out, as well as operation requirements on the network. Neural network models commonly assume specific mechanisms and yield information storage capacity estimates different from ours (Brunel et al., 2004; Gardner, 1987; McEliece et al., 1987; Newman, 1988; Rolls and Treves, 1998). Interestingly, Brunel et al. (2004) predict a distribution of synaptic weights similar to ours, although results such as this one may depend on the details of the neural network model at hand. Future research is needed to shed more light on the biological mechanisms that shape and constrain information storage and retrieval.

As our analysis relies on optimizing information storage capacity, it is not applicable to brain regions for which information storage is not the main task. For example, synapses associated with early sensory processing, e.g., in the retina (Laughlin et al., 1998; Sterling and Matthews, 2005), or calyx of Held (von Gersdorff and Borst, 2002), or those belonging to motoneurons (Pierce and Mendell, 1993; Yeow and Peterson, 1991) may be large and reliable. This would be consistent with optimizing information transmission. In actuality, any given brain circuit probably contributes to both information storage and information transmission. Indeed, by applying our analysis in reverse, one could infer the role of a given circuit from its structural characteristics. In particular, different cortical layers may be optimized for a different combination of storage and processing.

Our formulation of memory in the Shannon framework implicitly casts each synapse—both potential and actual—as a channel usage. The total storage capacity is therefore the number of synapses multiplied by the average synaptic storage capacity. This makes the storage capacity on the order of the number of synapses, which would correspond to an overall maximal storage capacity of several kilobits for a neocortical L5 pyramidal neuron (Braitenberg and Schüz, 1998). It is possible, however, that the synaptic information retrieval mechanism involves multiple read-out attempts from a single synapse. Since each channel usage is separated in space rather than in time, this does not increase the number of channel usages. Regardless, one may wonder what impact multiple read-out attempts would have on our analysis of information storage capacity.

It is known that the SNR increases approximately as the square root of the number of read-out trials for most forms of signal integration (Harrington, 1955), so if the information stored in each synapse was retrieved using the same number of read-out attempts, this simply introduces a fixed multiplicative constant in Equation 1.1. A fixed constant in Equation 1.1 can simply be incorporated into the V_N term, and all of our results stand. Contrarily, if the number of read-out attempts is not fixed, but varies across different synapses, then it would cast our estimate of noise into doubt. We point out, however, that multiple read-out attempts would lead to large time delays. Yet, if information is used to control dynamical systems, it is known that large delay can be disastrous (Sahai and Mitter, 2006). In addition, it is not clear how short-term plasticity caused by multiple read-out attempts would be overcome.

Other possible concerns arise from the lack of a true experimentally established input-output characterization of synaptic memory. To address this concern would require identification and description of the so-called engram—the physical embodiment of memory—which corresponds to the channel input, A (Figure 1). In addition, it would necessitate a better characterization of the noise process that determines the input-output probability distribution, $p(B|A)$. Description of the alphabet of A would furthermore settle the question, alluded to in section V, of whether synapses are discrete valued or continuously graded. In addition, we assumed in section IV that the channel input letter A is given by the arithmetic mean of EPSPs observed in several trials. Alternatives to this assumption may alter the horizontal coordinate of points in Figure 4B.

Although our analysis relies on identifying synaptic noise with retrieval—or more specifically the variability of EPSP amplitude on “read-out”—the noise may also come from other sources. The main concern is perhaps that long-term memory storage at a synapse is open to perturbations due to active learning rules and ongoing neuronal activity (Zhou and Poo, 2004), the so-called in situ noise (Figure 1). The longer the information is stored, the greater the perturbations caused by such processes (although see Abraham et al., 2002). Under generic assumptions, Amit and Fusi (1994) demonstrated that this noise restricts memory capacity significantly and even paradoxically. Fusi et al. (2005) recently proposed a solution to this paradox: the introduction of a cascade of synaptic states with different transition probabilities results in a form of metaplasticity that increases retention times in the face of ongoing activity. Presumably, other forms of metaplasticity may also help protect stored information from unwanted perturbations. In addition, the stability of physiological synaptic plasticity appears to depend critically on the details of activity patterns during and after the induction of plasticity (Zhou and Poo, 2004), suggesting that specific biological mechanisms for the protection of stored information may exist.

Our theory can be modified to include sources of noise other than retrieval. For example, if in situ noise is quantified and turns out dominant, it can be used in the calculations presented in sections II–IV. In fact, optimality of noisy synapses (section II) may be relevant to the resolution of the above paradox. In general, a better understanding of the system functionality including characterization of storage, in situ, and retrieval noise should help specify $p(B|A)$ in the future.

Finally, our contributions include not only many explanations and predictions of physical structures, but also the introduction of methods developed elsewhere to the study of memory in the brain. For example, to develop the optimization principles, we have applied information theory to the study of physical neural memory systems. Moreover, our application of an alternate formulation of the capacity cost problem to study the cost function (section IV) appears to be the first instance where this alternate characterization of channel optimality has been successfully applied to real system analysis, whether biological or human engineered. This problem inversion has wide applicability to the experimental study of information systems.

Our theory can be modified to include sources of noise other than retrieval. For example, if in situ noise is quantified and turns out dominant, it can be used in the calculations presented in sections II–IV. In fact, optimality of noisy synapses (section II) may be relevant to the resolution of the above paradox. In general, a better understanding of the system functionality including characterization of storage, in situ, and retrieval noise should help specify $p(B|A)$ in the future.

Experimental Procedures

Calculation of the Optimum Average Synapse Volume

Here we calculate analytically the optimum average synapse volume $\langle V \rangle$ that maximizes information storage capacity per volume I_{volume} for given accessory volume V_0 and normalization V_N . This problem is mathematically identical to maximizing information transmission along parallel pathways (Sarpeshkar, 1998). We take the derivative of Equation 2.3 and set it to zero to obtain

$$2 \frac{\partial I_{\text{volume}}}{\partial \langle V \rangle} = \frac{-1}{(\langle V \rangle + V_0)^2} \ln \left(1 + \frac{\langle V \rangle}{V_N} \right) + \frac{1}{\langle V \rangle + V_0} \frac{1}{\langle V \rangle + V_N} = 0 \quad (\text{EP.1})$$

This implies that the optimal $\langle V \rangle$ can be found by solving the following equation:

$$\frac{\langle V \rangle + V_0}{\langle V \rangle + V_N} = \ln \left(1 + \frac{\langle V \rangle}{V_N} \right) \quad (\text{EP.2})$$

In the limiting cases, the optimizing average volume $\langle V \rangle$ and the maximum storage capacity achieved are given by (Sarpeshkar, 1998):

i)

$$V_0 \ll V_N \Rightarrow \begin{cases} \langle V \rangle = \sqrt{V_0 V_N} \\ \max\{I_{\text{volume}}\} = \frac{1}{2V_N} \end{cases}$$

ii)

$$V_0 \gg V_N \Rightarrow \begin{cases} \langle V \rangle \ln(\langle V \rangle / V_N) = V_0 \\ \max\{I_{\text{volume}}\} = \frac{1}{2\langle V \rangle} \end{cases} \quad (\text{EP.3})$$

The exact dependence of synaptic volume and the storage capacity on the accessory volume is shown in Figure 2B.

Derivation of Optimal Probability Distribution for Discrete Zero-Error Synaptic States

Following Balasubramanian et al. (2001) and de Polavieja (2002, 2004), we first consider the problem of maximizing information storage capacity

$$I_{\text{synapse}} = - \sum_i p_i \ln p_i \quad (\text{EP.4})$$

in a given (average synaptic) volume

$$\bar{V} = \sum_i p_i V_i \quad (\text{EP.5})$$

Both the volume constraint and the normalization condition for the probabilities of synaptic weights can be included in the constrained optimization by using Lagrange multipliers. Then we need to maximize

$$I_{\text{synapse}} = - \sum_i p_i \ln p_i - \beta \left(\sum_i p_i V_i - \bar{V} \right) - \lambda \left(\sum_i p_i - 1 \right) \quad (\text{EP.6})$$

By setting the derivatives of Equation 3.3 with respect to p_i equal to zero, we find that

$$p_i = \frac{1}{Z} \exp(-\beta V_i) \quad (\text{EP.7})$$

where $Z = \sum_i \exp(-\beta V_i)$ is a normalization constant (called the partition function in statistical physics) and β is implicitly specified by the condition

$$\bar{V} = \frac{1}{Z} \sum_i V_i \exp(-\beta V_i) \quad (\text{EP.8})$$

Recall now that \bar{V} is not given, and our objective is to maximize information per unit cost. Such an optimization problem can be solved, as pointed out by Balasubramanian et al. (2001), by choosing β such that the partition function $Z = 1$, i.e.,

$$\sum_i \exp(-\beta V_i) = 1 \quad (\text{EP.9})$$

In this case, the probability expression (Equation EP.7) simplifies to

$$p_i = \exp(-\beta V_i) \quad (\text{EP.10})$$

Substituting this expression into Equation EP.6, we find that information storage capacity is given by

$$I_{\text{synapse}} = \beta \sum_i V_i \exp(-\beta V_i) \quad (\text{EP.11})$$

Combining this expression with Equation 3.2, we find that information per volume is given by

$$I_{\text{volume}} = \frac{I_{\text{synapse}}}{\bar{V}} = \frac{\beta \sum_i V_i \exp(-\beta V_i)}{\sum_i V_i \exp(-\beta V_i)} = \beta \quad (\text{EP.12})$$

which is Equation 3.4 of the main text.

Distribution over Discrete Synaptic States Equidistant in Volume Space

Sums appearing in Equations EP.9 and EP.11 can be expressed in a closed form if we assume that

$$V_i = V_0 + 2iV_N \quad (\text{EP.13})$$

Then we can rewrite the normalization condition (Equation EP.9) as

$$1 = \sum_{i=0}^{\infty} \exp(-\beta V_i) = \exp(-\beta V_0) + \exp(-\beta V_0) \times \sum_{i=1}^{\infty} \exp(-2i\beta V_N) = \exp(-\beta V_0) + \frac{\exp(-\beta(V_0 + 2V_N))}{1 - \exp(-2\beta V_N)} \quad (\text{EP.14})$$

where we used an expression for the sum of the geometric series. Multiplying both sides of this expression by the denominator, we find

$$\exp(-\beta V_0) + \exp(-2\beta V_N) = 1. \quad (\text{EP.15})$$

Average synapse volume (including accessory volume V_0) is given by

$$\bar{V} = \sum_{i=0}^{\infty} V_i \exp(-\beta V_i) = - \frac{\partial}{\partial \beta} \sum_{i=0}^{\infty} \exp(-\beta V_i) = - \frac{\partial}{\partial \beta} \frac{\exp(-\beta V_0)}{1 - \exp(-2\beta V_N)} = \frac{2V_N \exp(-2\beta V_N) + V_0 \exp(-\beta V_0)}{\exp(-\beta V_0)} = V_0 + 2V_N \exp(\beta(V_0 - 2V_N)) \quad (\text{EP.16})$$

which is Equation 3.6 of the main text. In the limiting cases, these expressions reduce to:

i)

$$V_0 \ll V_N \Rightarrow \begin{cases} \bar{V} = 2\beta V_N V_0 \\ \beta V_0 = \exp(-2\beta V_N) \end{cases}$$

ii)

$$V_0 \gg V_N \Rightarrow \begin{cases} \bar{V} = V_0 \\ 2\beta V_N = \exp(-\beta V_0) \end{cases} \quad (\text{EP.17})$$

The volume of actual synapses (excluding accessory volume V_0):

$$\begin{aligned} \langle V \rangle_{i>0} &= \sum_{i=1}^{\infty} 2iV_N \frac{\exp(-\beta V_i)}{1 - \exp(-\beta V_0)} \\ &= - \frac{\exp(-\beta V_0)}{1 - \exp(-\beta V_0)} \frac{\partial}{\partial \beta} \sum_{i=1}^{\infty} \exp(-2i\beta V_N) \\ &= - \frac{\exp(-\beta V_0)}{\exp(-2\beta V_N)} \frac{\partial}{\partial \beta} \frac{\exp(-2\beta V_N)}{1 - \exp(-2\beta V_N)} \\ &= \frac{\exp(-\beta V_0)}{\exp(-2\beta V_N)} \frac{2V_N \exp(2\beta V_N)}{(\exp(2\beta V_N) - 1)^2} = 2V_N \exp(\beta V_0) \end{aligned} \quad (\text{EP.18})$$

which is Equation 3.7 of the main text.

Finally the filling fraction:

$$f = \sum_{i=1}^{\infty} p_i = \frac{\exp(-\beta(V_0 + 2V_N))}{1 - \exp(-2\beta V_N)} = \exp(-2\beta V_N) \quad (\text{EP.19})$$

which is Equation 3.8 of the main text.

Measurement of EPSPs in Synaptic Connections of L5 Pyramidal Neurons

Our cost function computation (section IV) and experimental plots (Figures S1 and S4) are based on the dataset from Sjöström et al. (2001, 2003), where detailed methods have been previously described. This dataset was analyzed with respect to connectivity patterns and synaptic weights in Song et al. (2005). Briefly, acute visual

cortical slices were cut from rats aged P12–P20; whole-cell recording configuration was established on up to four thick-tufted neocortical layer V pyramidal neurons using a gluconate-based internal solution; connectivity was assessed using a minimum of ten traces; EPSPs were measured using a 1 ms window centered on the peak of the averaged EPSP trace. The dataset consisted of recordings from 637 connected pairs of neurons. Between 11 and 150 responses were recorded in each connected pair (repeated every 7–20 s to avoid short-term depression); the vast majority of connections admitted between 40 and 65 responses. We define the synaptic weight as the mean EPSP averaged across all responses.

Computing the Optimal Cost Function for Gaussian Input-AWGN Channel and Zero Error Channel

For the AWGN channel, let Z be the random variable that represents the independent additive noise. Then the expression in Equation 4.1, up to affine transformation, reduces to

$$\begin{aligned} D(p_{B|A}(b|a)||p_B(b)) &= D(p_Z(b-a)||p_B(b)) \\ &= -h(p_Z(b-a)) - \int p_Z(b-a) \ln p_B(b) db \end{aligned}$$

Since we are only interested in the quantity up to affine transformation, we need not calculate the first entropy term explicitly, since it is constant, call it C_1 .

$$\begin{aligned} D(p_{B|A}(b|a)||p_B(b)) &= C_1 - \int \frac{1}{A_N \sqrt{2\pi}} \exp\left(-\frac{(b-a)^2}{2A_N^2}\right) \\ &\times \left(\ln \frac{1}{\sqrt{2\pi}(\langle A^2 \rangle + A_N^2)} - \frac{b^2}{2(\langle A^2 \rangle + A_N^2)} \right) db = C_2 + C_3 \\ &\times \int \frac{1}{A_N \sqrt{2\pi}} \exp\left(-\frac{(b-a)^2}{2A_N^2}\right) b^2 db = C_2 + C_3(A_N^2 + a^2) = C_4 a^2 + C_5 \end{aligned}$$

where C_2, C_3, C_4 , and C_5 are other constants that are formed by combining terms that need not be calculated explicitly. The result is Equation 4.2 in the main text.

For the zero-error channel, the optimal cost function is

$$\begin{aligned} D(p(B=A_i|A=A_i)||p(B=A_i)) &= \sum_{i=0}^{\infty} p(B=A_i|A=A_i) \ln \frac{p(B=A_i|A=A_i)}{p(B=A_i)} \\ &= \sum_{i=0}^{\infty} \hat{\delta}_{ij} \ln \frac{\hat{\delta}_{ij}}{p(B=A_i)} \\ &= -\ln p(A=A_i) = -\ln p_i \end{aligned}$$

Calculation of the Synaptic Cost Function

In order to compute the optimal cost function according to Equation 4.1, we require the channel output distribution as well as the channel output distribution conditioned on the input. To estimate the channel output cumulative distribution function, $F(B)$, we simply use the empirical cumulative distribution function, $F_{emp}(B)$. To account for the variable number of EPSPs acquired from each synaptic connection, the step size in the empirical cumulative distribution function contributed by each data point is inversely proportional to the number of EPSPs obtained from the synapse in question. To model the effect of absent synapses, we included in the empirical distribution function a steep Gaussian distribution function with mean at zero EPSP amplitude and standard deviation of 0.1 mV (typical noise amplitude). The area under this Gaussian distribution function is given by one minus the filling fraction of 11.6% (Song et al., 2005). To estimate the channel conditional density, $p(B|A)$, we assume that all EPSPs from a given synapse correspond to the same input letter; furthermore, we make a correspondence between the mean EPSP amplitude and this input letter. For each synapse, we use a histogram with ten uniformly spaced bins to estimate the conditional density, $p_{emp}(B|A)$. Then the KL divergence is approximated by the following:

$$D(p_{emp}(B|A)||p_{emp}(B)) = \sum_{i=1}^{10} p_{emp}(B=b_i|A) \ln \frac{p_{emp}(B=b_i|A)}{F_{emp}(B=r_i) - F_{emp}(B=l_i)} \quad (\text{EP.20})$$

where $B = b_i$ implies presence in the i th histogram bin, and the right and left bin edges are denoted r_i and l_i , respectively. This KL diver-

gence is computed for each synapse, and the result is the optimal cost function (Figure 4B).

We use the bootstrap method to determine confidence intervals. The confidence interval represented by the horizontal error bars is a small sample estimate of the standard error of the mean EPSP amplitude. The estimate is the sample standard deviation of the mean EPSP amplitudes generated by sampling with replacement from the empirical distribution that were measured for each synapse. The result is based on 50 bootstrap trials. The confidence interval represented by the vertical error bars is a small sample estimate of the standard error of the KL divergence quantity in Equation EP.20. In order to do the sampling with replacement for the bootstrap procedure, randomness is introduced in two ways. First, 637 synapses are selected uniformly with replacement from the set of 637 measured synapses. These resampled synapses are used to generate the empirical cumulative distribution function, $F_{emp}(B)$. Second, for each of the measured synapses, EPSP values are selected uniformly with replacement to produce the empirical conditional probability mass functions $p_{emp}(B|a)$ for each of the 637 values of a . Finally, Equation EP.20 is used to calculate a bootstrap version of the data points in Figure 4B. This doubly stochastic resampling procedure is repeated for 50 bootstrap trials, and the estimate is the sample standard deviation of the KL divergence quantities for each of the synapses. When we resample the synapses to generate the $F_{emp}(B)$ for each of the bootstrap trials, there is a possibility that no large EPSP amplitude synapse is selected. In such a case, the KL divergence quantity will come out to be infinite since the output distribution will be zero where the conditional distribution will be non-zero, so the two distributions will not be absolutely continuous with respect to each other. More simply, $F_{emp}(B = r_i \text{ mV})$ will equal $F_{emp}(B = l_i \text{ mV})$, so the denominator inside the logarithm in Equation EP.20 will be zero and cause the entire expression to be infinite. This phenomenon of lack of distribution overlap resulting in an infinite discontinuity in the KL divergence functional causes the two infinite upper standard deviation errors represented by stars in Figure 4B. For these two points, the lower standard deviation is estimated by excluding points greater than the mean. For other points, the bootstrap distributions are approximately symmetric, so the vertical error bars are symmetrically replicated. The unweighted least-squares fit in the MATLAB curve-fitting toolbox is used to generate the fit to a function of the form νA^η with $\eta = 0.48$, Figure 4B.

AWGN Channel Capacity at Low SNR

The information capacity of the AWGN channel with binary input $\pm(A^2)^{1/2}$ is achieved by using the two inputs with equal probability. Define the variable SNR to be

$$\left\langle \frac{A^2}{A_N^2} \right\rangle$$

Perturbing the input values with Gaussian noise yields the output distribution

$$\begin{aligned} p_B(b) &= \frac{1}{2} \left[\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(b - \sqrt{\text{SNR}})^2}{2}\right\} \right] \\ &+ \frac{1}{2} \left[\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(b + \sqrt{\text{SNR}})^2}{2}\right\} \right] \end{aligned}$$

where the system has been normalized so the channel output is normalized by the noise amplitude: $b \rightarrow b/A_N$. The output distribution then has entropy

$$h(B) = - \int_{-\infty}^{\infty} p_B(b) \ln p_B(b) db$$

which can be computed numerically.

The capacity is then given by

$$I(A; B) = h(B) - h(B|A) = h(B) - \frac{1}{2} \ln(2\pi e)$$

where we have used the property that for additive noise channels, the conditional entropy is the entropy of the noise process as well as the known entropy of Gaussian random variables. The value of the mutual information is close to the value from Equation 2.1 at small SNR (Figure S3).

Supplemental Data

The Supplemental Data for this article can be found online at <http://www.neuron.org/cgi/content/full/52/3/409/DC1/>.

Acknowledgments

We are grateful to M. DeWeese, A. Koulakov, C. Machens, R. Malinow, and S.K. Mitter for commenting on the early versions of the manuscript and to A. Roth, Y. Mishchenko, M. Häusser, L. Srinivasan, and S.B. Nelson for discussions. We also thank the anonymous reviewers and the editor for helping clarify our presentation. This work was supported by the NIH Grant MH69838, the Klingenstein Foundation Award, an NSF Graduate Research Fellowship, the NSF Grant CCR-0325774, the Wellcome Trust, and an EU Marie Curie grant.

Received: May 4, 2005

Revised: April 27, 2006

Accepted: October 10, 2006

Published: November 8, 2006

References

- Abraham, W.C., Logan, B., Greenwood, J.M., and Dragunow, M. (2002). Induction and experience-dependent consolidation of stable long-term potentiation lasting months in the hippocampus. *J. Neurosci.* **22**, 9626–9634.
- Allen, C., and Stevens, C.F. (1994). An evaluation of causes for unreliability of synaptic transmission. *Proc. Natl. Acad. Sci. USA* **91**, 10380–10383.
- Amit, D.J., and Fusi, S. (1994). Learning in neural networks with material synapses. *Neural Comput.* **6**, 957–982.
- Arimoto, S. (1972). An algorithm for computing the capacity of arbitrary discrete memoryless channels. *IEEE Trans. Inform. Theory* **IT-18**, 14–20.
- Balasubramanian, V., Kimber, D., and Berry, M.J., II. (2001). Metabolically efficient information processing. *Neural Comput.* **13**, 799–815.
- Bekkers, J.M., and Stevens, C.F. (1995). Quantal analysis of EPSCs recorded from small numbers of synapses in hippocampal cultures. *J. Neurophysiol.* **73**, 1145–1156.
- Blahut, R.E. (1972). Computation of channel capacity and rate-distortion functions. *IEEE Trans. Inform. Theory* **IT-18**, 460–473.
- Braitenberg, V., and Schüz, A. (1998). *Cortex: Statistics and Geometry of Neuronal Connectivity* (Berlin, New York: Springer).
- Brunel, N., Hakim, V., Isope, P., Nadal, J.-P., and Barbour, B. (2004). Optimal information storage and the distribution of synaptic weights: perceptron versus Purkinje cell. *Neuron* **43**, 745–757.
- Cash, S., and Yuste, R. (1999). Linear summation of excitatory inputs by CA1 pyramidal neurons. *Neuron* **22**, 383–394.
- Cherniak, C., Changizi, M., and Kang, D.W. (1999). Large-scale optimization of neuron arbors. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics* **59**, 6001–6009.
- Chklovskii, D.B. (2004). Synaptic connectivity and neuronal morphology: two sides of the same coin. *Neuron* **43**, 609–617.
- Chklovskii, D.B., Schikorski, T., and Stevens, C.F. (2002). Wiring optimization in cortical circuits. *Neuron* **34**, 341–347.
- Chklovskii, D.B., Mel, B.W., and Svoboda, K. (2004). Cortical rewiring and information storage. *Nature* **431**, 782–788.
- Csiszár, I., and Körner, J. (1997). *Information Theory: Coding Theorems for Discrete Memoryless Systems* (Budapest: Akadémiai Kiadó).
- de Polavieja, G.G. (2002). Errors drive the evolution of biological signalling to costly codes. *J. Theor. Biol.* **214**, 657–664.
- de Polavieja, G.G. (2004). Reliable biological communication with realistic constraints. *Phys. Rev. E Stat. Nonlin. Soft. Matter Phys.* **70**, 061910. [10.1103/PhysRevE.70.061910](https://doi.org/10.1103/PhysRevE.70.061910).
- del Castillo, J., and Katz, B. (1954). Quantal components of the end-plate potential. *J. Physiol. (London)* **124**, 560–573.
- Eldridge, D.F. (1963). A special application of information theory to recording systems. *IEEE Trans. Audio AU-11*, 3–6.
- Faisal, A.A., White, J.A., and Laughlin, S.B. (2005). Ion-channel noise places limits on the miniaturization of the brain's wiring. *Curr. Biol.* **15**, 1143–1149.
- Fusi, S., Drew, P.J., and Abbott, L.F. (2005). Cascade models of synaptically stored memories. *Neuron* **45**, 599–611.
- Gardner, E. (1987). Maximum storage capacity in neural networks. *Europhys. Lett.* **4**, 481–485.
- Gastpar, M. (2003). To code or not to code. PhD dissertation, École Polytechnique Fédérale de Lausanne, Switzerland.
- Gastpar, M., Rimoldi, B., and Vetterli, M. (2003). To code, or not to code: lossy source-channel communication revisited. *IEEE Trans. Inform. Theory* **49**, 1147–1158.
- Goldman, M.S. (2004). Enhancement of information transmission efficiency by synaptic failures. *Neural Comput.* **16**, 1137–1162.
- Gursoy, M.C., Poor, H.V., and Verdú, S. (2005). The noncoherent Rician fading channel—part I: structure of the capacity-achieving input. *IEEE Trans. Wireless Commun.* **4**, 2193–2206.
- Harrington, J.V. (1955). An analysis of the detection of repeated signals in noise by binary integration. *IRE Trans. Inform. Theory* **1**, 1–9.
- Hebb, D.O. (1949). *The Organization of Behavior: A Neuropsychological Theory* (New York: Wiley).
- Hessler, N.A., Shirke, A.M., and Malinow, R. (1993). The probability of transmitter release at a mammalian central synapse. *Nature* **366**, 569–572.
- Holmgren, C., Harkany, T., Svennenfors, B., and Zilberter, Y. (2003). Pyramidal cell communication within local networks in layer 2/3 of rat neocortex. *J. Physiol.* **551**, 139–153.
- Hsu, A., Tsukamoto, Y., Smith, R.G., and Sterling, P. (1998). Functional architecture of primate cone and rod axons. *Vision Res.* **38**, 2539–2549.
- Huang, J., and Meyn, S.P. (2005). Characterization and computation of optimal distributions for channel coding. *IEEE Trans. Inform. Theory* **51**, 2336–2351.
- Immink, K.E.S., Siegel, P.H., and Wolf, J.K. (1998). Codes for digital recorders. *IEEE Trans. Inform. Theory* **44**, 2260–2299.
- Isope, P., and Barbour, B. (2002). Properties of unitary granule cell → Purkinje cell synapses in adult rat cerebellar slices. *J. Neurosci.* **22**, 9668–9678.
- Jimbo, M., and Kunisawa, K. (1979). An iteration method for calculating the relative capacity. *Information and Control* **43**, 216–223.
- Kalisman, N., Silberberg, G., and Markram, H. (2005). The neocortical microcircuit as a tabula rasa. *Proc. Natl. Acad. Sci. USA* **102**, 880–885.
- Kasai, H., Matsuzaki, M., Noguchi, J., Yasumatsu, N., and Nakahara, H. (2003). Structure-stability-function relationships of dendritic spines. *Trends Neurosci.* **26**, 360–368.
- Koch, C. (1999). *Biophysics of Computation: Information Processing in Single Neurons* (New York: Oxford University Press).
- Koester, H.J., and Johnston, D. (2005). Target cell-dependent normalization of transmitter release at neocortical synapses. *Science* **308**, 863–866.
- Kolmogorov, A., and Tihomirov, V. (1959). ϵ -entropy and ϵ -capacity of sets in functional spaces. *Uspekhi Matematicheskikh Nauk* **14**, 3–86.
- Kopec, C.D., Li, B., Wei, W., Boehm, J., and Malinow, R. (2006). Glutamate receptor exocytosis and spine enlargement during chemically induced long-term potentiation. *J. Neurosci.* **26**, 2000–2009.
- Laughlin, S.B., de Ruyter van Steveninck, R.R., and Anderson, J.C. (1998). The metabolic cost of neural information. *Nat Neurosci* **1**, 36–41.
- Le Be, J.V., and Markram, H. (2006). Spontaneous and evoked synaptic rewiring in the neonatal neocortex. *Proc. Natl. Acad. Sci. USA* **103**, 13214–13219.
- Levy, W.B., and Baxter, R.A. (1996). Energy efficient neural codes. *Neural Comput.* **8**, 531–543.

- Levy, W.B., and Baxter, R.A. (2002). Energy-efficient neuronal computation via quantal synaptic failures. *J. Neurosci.* *22*, 4746–4755.
- Lisman, J. (2003). Long-term potentiation: outstanding questions and attempted synthesis. *Proc. R. Soc. Lond. B Biol. Sci.* *358*, 829–842.
- Lisman, J.E., and Harris, K.M. (1993). Quantal analysis and synaptic anatomy—integrating two views of hippocampal plasticity. *Trends Neurosci.* *16*, 141–147.
- Lynch, M.A. (2004). Long-term potentiation and memory. *Physiol. Rev.* *84*, 87–136.
- Magee, J.C., and Cook, E.P. (2000). Somatic EPSP amplitude is independent of synapse location in hippocampal pyramidal neurons. *Nat. Neurosci.* *3*, 895–903.
- Manwani, A., and Koch, C. (2000). Detecting and estimating signals over noisy and unreliable synapses: information-theoretic analysis. *Neural Comput.* *13*, 1–33.
- Markram, H. (1997). A network of tufted layer 5 pyramidal neurons. *Cereb. Cortex* *7*, 523–533.
- Markram, H., Lübke, J., Frotscher, M., Roth, A., and Sakmann, B. (1997). Physiology and anatomy of synaptic connections between thick tufted pyramidal neurons in the developing rat neocortex. *J. Physiol.* *500*, 409–440.
- Mason, A., Nicoll, A., and Stratford, K. (1991). Synaptic transmission between individual pyramidal neurons of the rat visual cortex in vitro. *J. Neurosci.* *11*, 72–84.
- Matsuzaki, M., Ellis-Davies, G.C.R., Nemoto, T., Miyashita, Y., Iino, M., and Kasai, H. (2001). Dendritic spine geometry is critical for AMPA receptor expression in hippocampal CA1 pyramidal neurons. *Nat. Neurosci.* *4*, 1086–1092.
- Matsuzaki, M., Honkura, N., Ellis-Davies, G.C., and Kasai, H. (2004). Structural basis of long-term potentiation in single dendritic spines. *Nature* *429*, 761–766.
- McEliece, R. (1977). *The Theory of Information and Coding: A Mathematical Framework for Communication* (London: Addison-Wesley).
- McEliece, R.J., Posner, E.C., Rodemich, E.R., and Venkatesh, S.S. (1987). The capacity of the Hopfield associative memory. *IEEE Trans. Inform. Theory* *IT-33*, 461–482.
- McGaugh, J.L. (2000). Memory—a century of consolidation. *Science* *287*, 248–251.
- Mitchison, G. (1991). Neuronal branching patterns and the economy of cortical wiring. *Proc. R. Soc. Lond. B Biol. Sci.* *245*, 151–158.
- Morris, R.G. (2003). Long-term potentiation and memory. *Proc. R. Soc. Lond. B Biol. Sci.* *358*, 643–647.
- Murthy, V.N., Schikorski, T., Stevens, C.F., and Zhu, Y. (2001). Inactivity produces increases in neurotransmitter release and synapse size. *Neuron* *32*, 673–682.
- Newman, C. (1988). Memory capacity in neural network models: rigorous lower bounds. *Neural Netw.* *1*, 223–238.
- Nusser, Z., Lujan, R., Laube, G., Roberts, J.D., Molnar, E., and Somogyi, P. (1998). Cell type and pathway dependence of synaptic AMPA receptor number and variability in the hippocampus. *Neuron* *21*, 545–559.
- O'Connor, D.H., Wittenberg, G.M., and Wang, S.S.-H. (2005). Graded bidirectional synaptic plasticity is composed of switch-like unitary events. *Proc. Natl. Acad. Sci. USA* *102*, 9679–9684.
- Petersen, C.C., Malenka, R.C., Nicoll, R.A., and Hopfield, J.J. (1998). All-or-none potentiation at CA3-CA1 synapses. *Proc. Natl. Acad. Sci. USA* *95*, 4732–4737.
- Pierce, J.P., and Mendell, L.M. (1993). Quantitative ultrastructure of la boutons in the ventral horn: scaling and positional relationships. *J. Neurosci.* *13*, 4748–4763.
- Poirazi, P., Brannon, T., and Mel, B.W. (2003). Arithmetic of sub-threshold synaptic summation in a model CA1 pyramidal cell. *Neuron* *37*, 977–987.
- Polsky, A., Mel, B.W., and Schiller, J. (2004). Computational subunits in thin dendrites of pyramidal cells. *Nat. Neurosci.* *7*, 621–627.
- Raastad, M., Storm, J.F., and Andersen, P. (1992). Putative single quantum and single fibre excitatory postsynaptic currents show similar amplitude range and variability in rat hippocampal slices. *Eur. J. Neurosci.* *4*, 113–117.
- Ramón y Cajal, S. (1899). *Textura del Sistema Nervioso del Hombre y de los Vertebrados* (New York: Springer).
- Rieke, F., Warland, D., de Ruyter van Steveninck, R., and Bialek, W. (1997). *Spikes: Exploring the Neural Code* (Cambridge, MA: The MIT Press).
- Rolls, E., and Treves, A. (1998). *Neural Networks and Brain Function* (Cambridge: Oxford University Press).
- Root, W.L. (1968). Estimates of ϵ capacity for certain linear communication channels. *IEEE Trans. Inform. Theory* *IT-14*, 361–369.
- Rosenmund, C., Clements, J.D., and Westbrook, G.L. (1993). Non-uniform probability of glutamate release at a hippocampal synapse. *Science* *262*, 754–757.
- Sahai, A., and Mitter, S.K. (2006). The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link Part I: scalar systems. *IEEE Trans. Inform. Theory* *52*, 3369–3395.
- Sarpeshkar, R. (1998). Analog versus digital: extrapolating from electronics to neurobiology. *Neural Comput.* *10*, 1601–1638.
- Sayer, R.J., Friedlander, M.J., and Redman, S.J. (1990). The time course and amplitude of EPSPs evoked at synapses between pairs of CA3/CA1 neurons in the hippocampal slice. *J. Neurosci.* *10*, 826–836.
- Schikorski, T., and Stevens, C.F. (1997). Quantitative ultrastructural analysis of hippocampal excitatory synapses. *J. Neurosci.* *17*, 5858–5867.
- Schreiber, S., Machens, C.K., Herz, A.V.M., and Laughlin, S.B. (2002). Energy-efficient coding with discrete stochastic events. *Neural Comput.* *14*, 1323–1346.
- Shannon, C.E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* *27*, 379–423 and 623–656.
- Shannon, C.E. (1959). Coding theorems for a discrete source with a fidelity criterion. *IRE National Convention Record* *4*, 142–163.
- Silver, R.A., Lübke, J., Sakmann, B., and Feldmeyer, D. (2003). High-probability unquantal transmission at excitatory synapses in barrel cortex. *Science* *302*, 1981–1984.
- Sjöström, P.J., Turrigiano, G.G., and Nelson, S.B. (2001). Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron* *32*, 1149–1164.
- Sjöström, P.J., Turrigiano, G.G., and Nelson, S.B. (2003). Neocortical LTD via coincident activation of presynaptic NMDA and cannabinoid receptors. *Neuron* *39*, 641–654.
- Smith, J.G. (1971). The information capacity of amplitude- and variance-constrained scalar Gaussian channels. *Inf. Control* *18*, 203–219.
- Song, S., Sjöström, P.J., Reigl, M., Nelson, S., and Chklovskii, D.B. (2005). Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biol.* *3*, 0507–0519. 10.1371/journal.pbio.0030068.
- Stepanyants, A., Hof, P.R., and Chklovskii, D.B. (2002). Geometry and structural plasticity of synaptic connectivity. *Neuron* *34*, 275–288.
- Stepanyants, A., and Chklovskii, D.B. (2005). Neurogeometry and potential synaptic connectivity. *Trends Neurosci.* *28*, 387–394.
- Sterling, P., and Matthews, G. (2005). Structure and function of ribbon synapses. *Trends Neurosci.* *28*, 20–29.
- Streichert, L.C., and Sargent, P.B. (1989). Bouton ultrastructure and synaptic growth in a frog autonomic ganglion. *J. Comp. Neurol.* *281*, 159–168.
- Takumi, Y., Ramírez-León, V., Laake, P., Rinvik, E., and Ottersen, O.P. (1999). Different modes of expression of AMPA and NMDA receptors in hippocampal synapses. *Nat. Neurosci.* *2*, 618–624.
- Tanaka, J., Matsuzaki, M., Tarusawa, E., Momiyama, A., Molnar, E., Kasai, H., and Shigemoto, R. (2005). Number and density of AMPA receptors in single synapses in immature cerebellum. *J. Neurosci.* *25*, 799–807.
- Tchamkerten, A. (2004). On the discreteness of capacity-achieving distributions. *IEEE Trans. Inform. Theory* *50*, 2773–2778.

- Thomson, A.M., and Bannister, A.P. (2003). Interlaminar connections in the neocortex. *Cereb. Cortex* 13, 5–14.
- Thomson, A.M., West, D.C., Wang, Y., and Bannister, A.P. (2002). Synaptic connections and small circuits involving excitatory and inhibitory neurons in layers 2–5 of adult rat and cat neocortex: triple intracellular recordings and biocytin labelling in vitro. *Cereb. Cortex* 12, 936–953.
- Verdú, S. (1990). On channel capacity per unit cost. *IEEE Trans. Inform. Theory* 36, 1019–1030.
- Verdú, S. (2002). Spectral efficiency in the wideband regime. *IEEE Trans. Inform. Theory* 48, 1319–1343.
- von Gersdorff, H., and Borst, J.G. (2002). Short-term plasticity at the calyx of held. *Nat. Rev. Neurosci.* 3, 53–64.
- Wen, Q., and Chklovskii, D.B. (2005). Segregation of the brain into gray and white matter: a design minimizing conduction delays. *PLoS Comput. Biol.* 1, e78. 10.1371/journal.pcbi.0010078.
- Yeow, M.B., and Peterson, E.H. (1991). Active zone organization and vesicle content scale with bouton size at a vertebrate central synapse. *J. Comp. Neurol.* 307, 475–486.
- Zador, A. (1998). Impact of synaptic unreliability on the information transmitted by spiking neurons. *J. Neurophysiol.* 79, 1219–1229.
- Zhou, Q., and Poo, M.M. (2004). Reversal and consolidation of activity-induced synaptic modifications. *Trends Neurosci.* 27, 378–383.
- Zhou, Q., Homma, K.J., and Poo, M.M. (2004). Shrinkage of dendritic spines associated with long-term depression of hippocampal synapses. *Neuron* 44, 749–757.