Neuron, Vol. 43, 745–757, September 2, 2004, Copyright ©2004 by Cell Press

Optimal Information Storage and the Distribution of Synaptic Weights: Perceptron versus Purkinje Cell

Nicolas Brunel,[1] Vincent Hakim,[2] Philippe Isope,[3] Jean-Pierre Nadal,[2] and Boris Barbour[4,*]
[1]Neurophysique et Physiologie
Université René Descartes
45 rue des Saints Pères
75270 Paris Cedex 06
France
[2]Laboratoire de Physique Statistique
Ecole Normale Supérieure
24 rue Lhomond
75231 Paris Cedex 05
France
[3]Department of Psychiatry
Kinsmen Lab
4N11-2255 Wesbrook Mall
Vancouver, British Columbia V6T 1Z3
Canada
[4]Laboratoire de Neurobiologie
Ecole Normale Supérieure
46 rue d'Ulm
75230 Paris Cedex 05
France

Summary

It is widely believed that synaptic modifications underlie learning and memory. However, few studies have examined what can be deduced about the learning process from the distribution of synaptic weights. We analyze the perceptron, a prototypical feedforward neural network, and obtain the optimal synaptic weight distribution for a perceptron with excitatory synapses. It contains more than 50% silent synapses, and this fraction increases with storage reliability: silent synapses are therefore a necessary byproduct of optimizing learning and reliability. Exploiting the classical analogy between the perceptron and the cerebellar Purkinje cell, we fitted the optimal weight distribution to that measured for granule cell-Purkinje cell synapses. The two distributions agreed well, suggesting that the Purkinje cell can learn up to 5 kilobytes of information, in the form of 40,000 input-output associations.

Introduction

Distributed changes of synaptic efficacy are thought to underlie much of the learning and memory occurring in the brain (Hebb, 1949). The distribution of synaptic weights should thus be related to what has been learned and the manner of its learning. However, few studies have attempted to exploit this link, despite measurements of synaptic weight distributions becoming available for an increasing number of connections (Sayer et al., 1990; Mason et al., 1991; Markram et al., 1997; Sjöström et al., 2001; Isope and Barbour, 2002; Holmgren et al., 2003). One obvious difficulty is the absence of a theoretical framework in which to interpret synaptic weight distributions. Recent theoretical activity in this direction has investigated synaptic weight distributions arising from the application of specific rules of spike timing-dependent synaptic plasticity (van Rossum et al., 2000; Song et al., 2000; Rubin et al., 2001; Cateau and Fukai, 2003; Gütig et al., 2003). In these studies, no learning task was defined.

Here, we explore the complementary approach of defining a learning task and deducing the resulting distribution of synaptic weights. The network we study is the classical perceptron, a single-layer feedforward network (Rosenblatt, 1962; Minsky and Papert, 1988), with excitatory weights. The task we assign to this network consists of learning the largest possible number of random input/output associations given a particular reliability level. Analytical techniques from the statistical mechanics of disordered systems (Mézard et al., 1987) make it possible to calculate the maximal capacity of such a network (Gardner, 1988; Engel and van den Broeck, 2001). This maximal capacity depends upon a number of network parameters but, crucially, is independent of the precise learning rule used (inasmuch as the rule is able to attain this maximal capacity). Using similar techniques, we investigated the distribution of synaptic weights associated with maximal storage capacity in a perceptron with continuous nonnegative synaptic weights.

This optimal distribution of synaptic weights contained a majority of silent synapses and decayed strictly monotonically (as a function of synaptic weight). The analysis identified two unsuspected computational roles for silent synapses—they are a necessary byproduct of optimal learning and of making storage reliable. The distribution resembled the recently reported distribution of synaptic weights for the cerebellar synapses between granule cells and Purkinje cells (Isope and Barbour, 2002), for which a majority of undetectable synapses has been reported (Ekerot and Jörntell, 2001; Isope and Barbour, 2002). This was particularly intriguing, because the perceptron has long been considered analogous to the cerebellar Purkinje cell. We therefore proceeded to a quantitative comparison of the model with the experimental data for the Purkinje cell. This showed that the model could fully explain the unusual experimental distribution. Constraint of model parameters by fitting the model to the data allowed us to estimate high-level properties of Purkinje cell function, such as its information storage capacity.

This study therefore offers a quantitative comparison of a theory of synaptic weights and experimental data, illustrating the insight into learning processes that can be gained by studying the distribution of synaptic weights.

Results

The Purkinje Cell Perceptron

The analogy between the cerebellum and the perceptron was highlighted in the pioneering work of Marr (1969)

*Correspondence: barbour@ens.fr

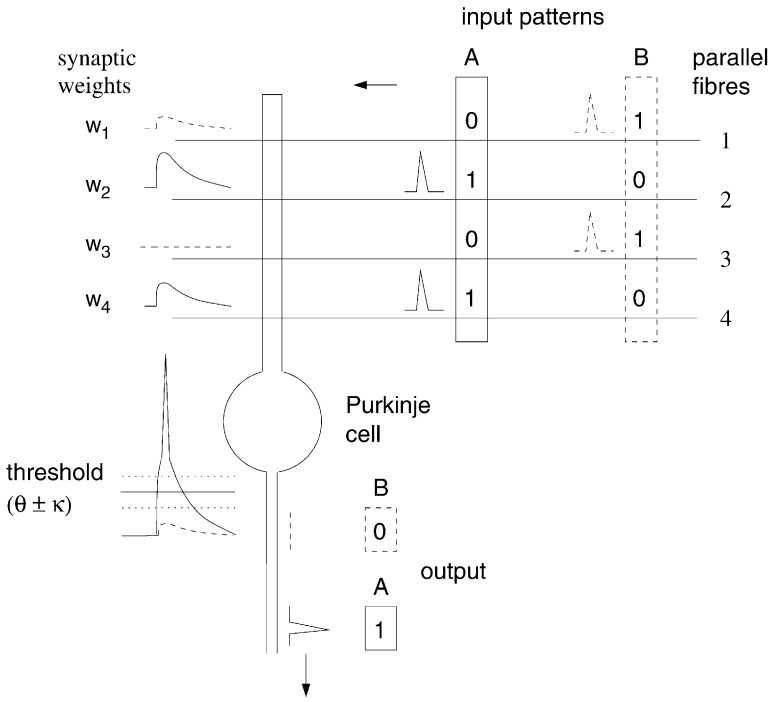

Figure 1. The Purkinje Cell Perceptron

A binary input pattern represents action potentials in a subset of the $N = 150{,}000$ granule cell axons (parallel fibers) contacting the Purkinje cell. The mean fraction of granule cells active in an input pattern is f. Each synapse has its own modifiable weight (peak somatic depolarization), w. The output is also binary and the fraction of input patterns eliciting an action potential is f'. We assume that the linear sum of the weights of active synapses is compared to the threshold (θ). The threshold stability parameter, κ, is associated with threshold. Storage requires responses to sum to values outside $\theta \pm \kappa$.

and Albus (1971). The central element of the perceptron is an association neuron that receives multiple modifiable synaptic inputs. It must learn to generate a stereotyped output in response to each given *input pattern* (set of active inputs). These *associations* are learned (stored) by modifying the synaptic weights during training. Both the input patterns and outputs to be associated are imposed upon the perceptron. Analogously, the cerebellar cortex learns to associate sensory and contextual inputs with motor control outputs dictated by the needs of the ongoing motor behavior. This learning involves plasticity of the granule cell-Purkinje cell synapses (Ito and Kano, 1982; Ito et al., 1982; De Zeeuw et al., 1998), establishing the analogy between the perceptron's association neuron and the Purkinje cell.

The Purkinje cell-perceptron model used for our calculations was as follows (Figure 1A). Activity in granule cell (parallel fiber) inputs and the Purkinje cell output was assumed to be binary (zero or one action potentials). Each input pattern involved activity in a random subset of the $N = 150{,}000$ granule cells (Harvey and Napper, 1991); on average, a fraction f were active. A fraction f' of the associated outputs, also chosen at random, required emission of an action potential. We defined the synaptic weights (w_i) as the peak somatic depolarization, since the axon hillock is thought to be the site of action potential initiation (Stuart and Häusser, 1994). Because the synapses are excitatory, we assumed the weights to be positive (or zero). The linear sum of the weights of active inputs was compared to the threshold, θ. In order for an association to be considered reliably stored, its input pattern had to sum to a mean greater than $\theta + \kappa$ if an action potential was to be output or less than $\theta - \kappa$ if no action potential was required, where κ is the *threshold stability constant*. As discussed in more detail below, κ introduces the notion of resistance to noise, or reliability, into the model, since noise must exceed κ in order to cause erroneous threshold crossings.

Experimental Testing of Model Assumptions

Our model of the Purkinje cell as a perceptron assumes that both inputs and output are binary and that synaptic inputs sum linearly on average (i.e., the summation is allowed to be noisy). We investigated the plausibility of this mode of synaptic integration using whole-cell voltage recordings from Purkinje cells in cerebellar slices. We elicited parallel fiber inputs using extracellular stimulation in the granule cell layer (Figure 2A). The site of stimulation was chosen to lie in the middle of the beam of connected granule cells, the beam having been delimited by prior mapping with the extracellular stimulation. In accordance with the classic recordings of Eccles and coworkers (Eccles et al., 1967), stimulation of granule cells evoked a biphasic response (Figure 2B): an excitatory phase corresponding to monosynaptic parallel fiber inputs, promptly truncated by disynaptic inhibition mediated by molecular layer interneurons, which are themselves driven by parallel fibers. The effect of inhibition is clearly demonstrated by the application of a $GABA_A$ antagonist (picrotoxin, 100 μM), which unmasks a slowly decaying monophasic EPSP. The duration of the net depolarization (in the presence of inhibition) was brief: for $\sim$1 mV depolarizations, full-width half-maximum (FWHM, width at half-maximum amplitude) = 1.8 $\pm$ 0.5 ms (n = 5). These depolarizations are likely to involve inputs from at least 15 granule cells (Isope and Barbour, 2002). The briefness of the net depolarization is to be compared with the membrane time constant of 52 $\pm$ 24 ms (n = 4; measured in the absence of picrotoxin), which is often considered to represent the time of effective integration (Figure 2C).

We next examined how the excitatory and inhibitory

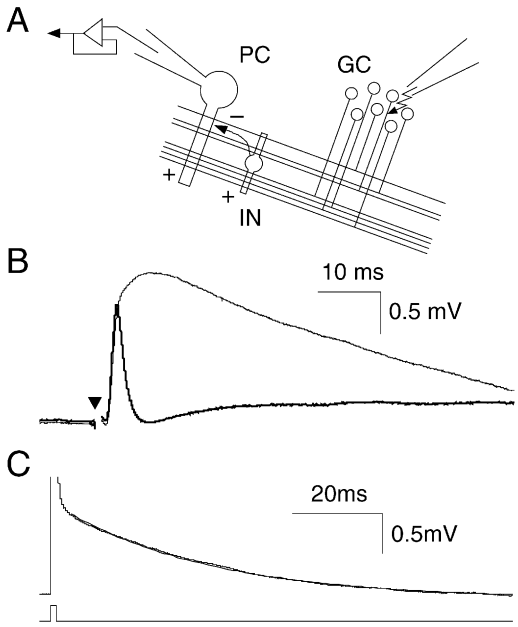

Figure 2. Feedforward Inhibition Limits Excitation to a Millisecond Window

(A) Diagram depicting voltage recording of a Purkinje cell in a cerebellar slice, stimulation of a group of on-beam granule cells, and the molecular layer interneurons they drive.

(B) Voltage responses of a Purkinje cell to extracellular stimulation (triangle) of granule cells. The heavy line shows a very brief net depolarization (FWHM = 1.4 ms) that is subthreshold. The light line shows the response in the presence of picrotoxin (100 μM), a $GABA_A$ receptor antagonist. Averages of ~100 sweeps.

(C) Relaxation of the voltage response to a brief current injection, revealing the membrane time constant. The superimposed exponential fit has $\tau = 39$ ms.

inputs were recruited, by recording synaptic responses over a range of stimulus intensities (Figure 3A). The specimen traces suggest that inhibitory inputs are recruited concomitantly with even the weakest of parallel fiber inputs, and this was confirmed by measuring peak depolarization and hyperpolarization (Figure 3B) as a function of stimulus intensity.

The form of the biphasic response to granule cell stimulation strongly suggests that only coincident inputs will be able to sum effectively. This was tested using interaction of two independent granule cell inputs. Figure 4A shows the family of responses obtained as the interval between the stimulations of the two granule cell inputs was varied. As expected, only coincident inputs were able to sum effectively. This is quantified in the graph of Figure 4B, which shows that the FWHM for summation was 2.4 ms (n = 4).

Finally, we investigated the nature of the summation when the inputs were coincident. The dashed line in Figure 4B indicates the prediction for linear summation of the inputs with synchronous stimulation. When the stimuli were simultaneous, the resultant response was 95% ± 9% (n = 5) of this linear prediction.

In summary, in the presence of feedforward inhibition, only coincident granule cell inputs can excite a Purkinje cell effectively. The window of net excitation is then sufficiently brief that it is unlikely that either Purkinje cells or granule cells (Chadderton et al., 2004) would emit more than one action potential during that time. This supports the assumption of binary inputs and output. Coincident inputs sum reasonably linearly, and this supports the linearity approximation in the perceptron model.

The Optimal Synaptic Weight Distribution

Since the work of Marr (1969) and Albus (1971), statistical physics has provided a framework for analyzing the perceptron (Engel and van den Broeck, 2001) and, in particular, its maximum storage capacity, under quite general conditions (Gardner, 1988; Gutfreund and Stein, 1990). Using a generalization of the approach of Kohler and Widmaier (1991), we calculated the synaptic weight distribution of the perceptron at this *critical* (or *maximal*) capacity α_c (see the Supplemental Data [http://www.neuron.org/cgi/content/full/43/5/745/DC1]), defined as the number of input-output associations learned per synapse (i.e., normalized by $N = 150{,}000$). Although the derivation is involved, the family of optimal weight distributions obtained is surprisingly simple. A specimen solution is shown in Figure 5A. The distribution decays monotonically from a maximum at zero weight. It in fact represents part of a Gaussian curve (the mean of the Gaussian is negative). To this Gaussian is added a "spike" (Dirac delta function) of zero weight—silent—synapses.

The distribution is specified by two composite parameters. The first is the average synaptic weight,

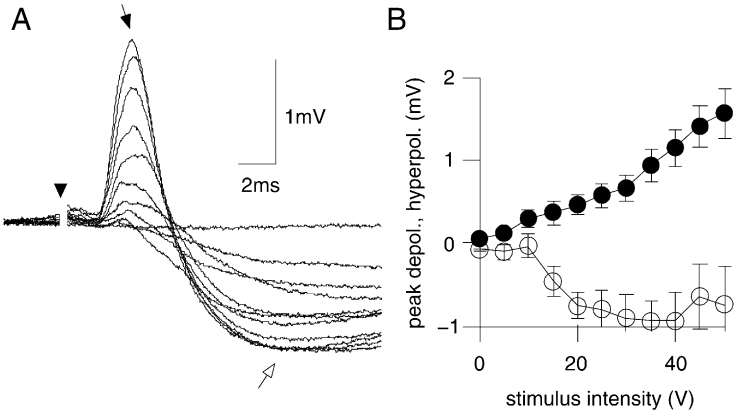

Figure 3. Inhibitory Interneurons Are Recruited by Subthreshold Excitation

(A) Family of responses recorded in a Purkinje cell to granule cells stimulations (triangle) of increasing intensity, illustrating the concomitant recruitment of excitation and inhibition.

(B) Plot of peak depolarization (filled circles, filled arrow panel [A]) and subsequent peak hyperpolarization (open circles, open arrow panel [A]) as a function of stimulus intensity. Means ± SDs of n = 5 cells.

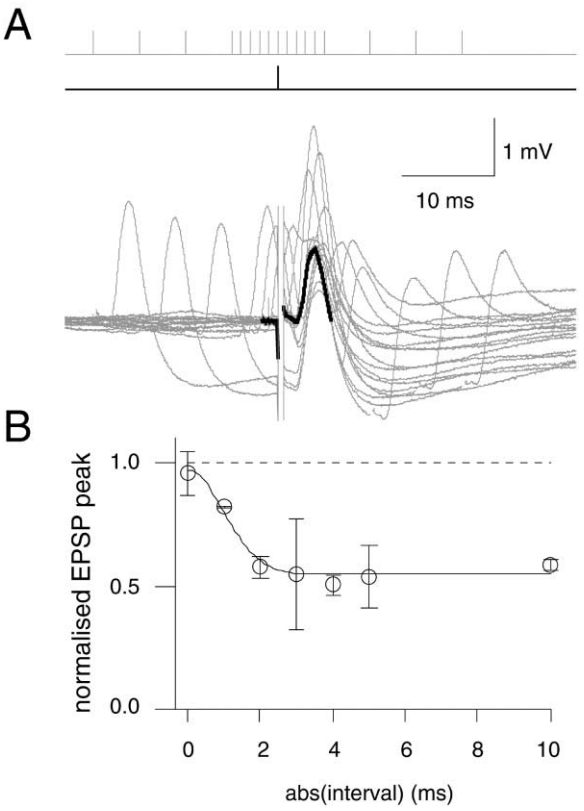

Figure 4. Coincidence Detection and Linear Summation

(A) Family of curves showing summation of two independent inputs (independent because of different stimulation locations in the granule cell layer) as a function of interval between stimuli. The black curve shows the average response to the fixed-time stimulus in isolation. The gray traces indicate the summed responses. The timings of the fixed (black lines) and swept (gray lines) stimuli are indicated above the voltage records. Averages of 20 sweeps.

(B) Summary graph pooling five such experiments. The plot shows the mean (±SD) maximum EPSP amplitude (anywhere in the 40 ms window), normalized to the prediction for linear summation of responses to coincident stimuli (dashed line), as a function of the absolute interval between stimuli. n = 5–10; some data points were unavailable because of the stimulus artifacts. The smooth curve is a Gaussian fit: baseline = 0.54, peak = 0.96, mean = 0, FWHM = 2.4 ms.

$$\overline{w} = \frac{\theta}{f N}, \quad (1)$$

which simply normalizes the distribution along the weight axis. The second is the fraction of silent synapses. This fraction of silent synapses is mostly controlled by the dimensionless *reliability parameter*,

$$\rho = \frac{\kappa}{\overline{w}\sqrt{f N(1-f)}} = \frac{\kappa\sqrt{f N}}{\theta\sqrt{1-f}}, \quad (2)$$

but also depends, very weakly, on the fraction f' of outputs for which the Purkinje cell is active (Equation 26 in the Supplemental Data [http://www.neuron.org/cgi/content/full/43/5/745/DC1]). The numerator of the reliability parameter, Equation 2, is κ, while the denominator,

$$\overline{w}\sqrt{f N(1-f)},$$

gives the scale of the fluctuations arising through the random activation of a fraction f of N synapses with average weight $\overline{w}$. When $\rho = 0$, the distribution is a hemi-Gaussian and 50% of synapses are silent (Figure 5B). As ρ increases, the Gaussian underlying the optimal distribution moves leftward and broadens, and a greater fraction of synapses becomes silent (inset, Figure 5B).

We recall that κ represents the robustness of storage in the face of various sources of noise, such as errors of input patterns, synaptic variability, input jitter, membrane potential noise, and background synaptic activity. Given the (unknown) distribution of noise, the probability of erroneous output could be computed from κ (see Figure 5C). Storage will be very reliable if the standard deviation (SD) of the noise is much smaller than κ, while it will be unreliable if the SD of the noise is of the same order or larger than κ. Through the influence of κ on ρ, we can see how increasing the reliability of storage requires a further increase in the fraction of silent synapses (Figure 5B).

We conclude that a majority of silent synapses is *required* to attain maximal storage capacity. We stress that no ad hoc assumptions about the shape of this distribution were made in its derivation—the shape emerges from the requirement for optimality and the constraint that weights should be nonnegative.

Some insight into how silent synapses arise can be gained from simulated learning with the same model parameters, using an activity-dependent error correction learning rule (see the Supplemental Data [http://www.neuron.org/cgi/content/full/43/5/745/DC1]). Many patterns produce erroneous output action potentials, and the active synapses must therefore be depressed. Repeated depression causes many individual synapses to strike the constraint at zero weight, leading to the accumulation of silent synapses. The simulation was also carried out for a technical reason: to check the possible influence of a number of approximations in the derivation that are only exact for large numbers of synapses (which seems reasonable in the case of the Purkinje cell). The distribution resulting from this simulation is in good agreement with the analytical solution, as shown in Figure 5A, indicating the validity of the assumptions made in the analytical derivation.

Subcritical Distributions

We have so far considered the optimal weight distribution which applies at critical capacity. However, there are several ways in which the Purkinje cell might not have attained critical capacity. These include not having completed training or having learned fewer associations than possible. A characteristic feature of the corresponding *subcritical* weight distributions is that relief of the optimality constraint allows the silent synapses to assume small but nonzero weights. We calculated representative subcritical distributions for the case of learning fewer associations than the maximum number possible, and these are illustrated in Figure 5D. When few patterns are stored (α close to zero), there are few constraints on the synaptic distribution, except that the average synaptic weight must be equal to the threshold θ divided by the mean number of active inputs in an association fN. This constraint leads to an exponential distribution

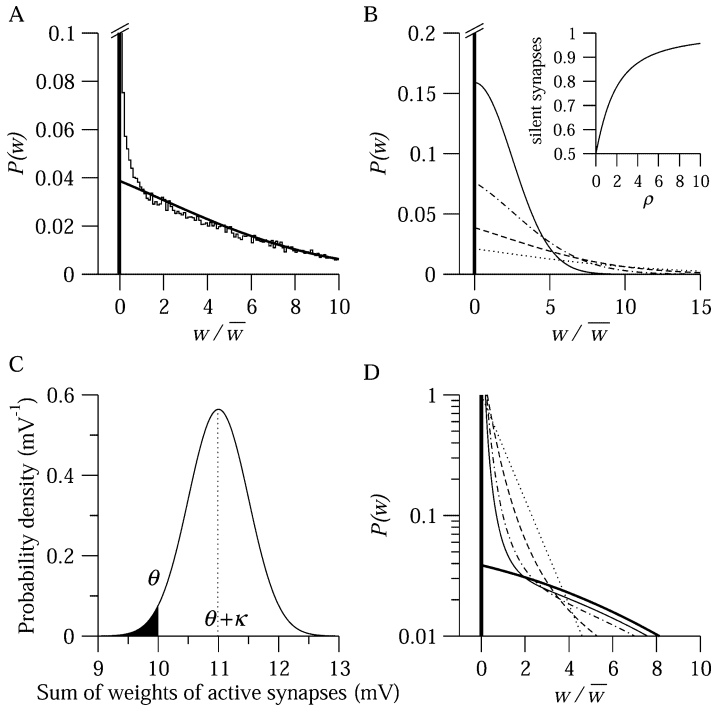

Figure 5. The Perceptron Synaptic Weight Distribution Contains Many Silent Synapses

(A) The perceptron distribution of synaptic weights at critical capacity is a partial Gaussian with a Dirac delta of silent synapses. The smooth heavy curve is an analytical solution. The light histogram shows the results of a simulation.

(B) The weight distribution at critical capacity is determined by two composite parameters: the mean weight (including silent synapses) $\overline{w}$, which normalizes the distribution along the weights axis; and the reliability parameter ρ. Increasing ρ shifts the underlying Gaussian leftward and broadens it (solid, $\rho = 0$; dot-dashed, $\rho = 1$; dashed, $\rho = 2$; dotted, $\rho = 3$), thereby increasing the fraction of silent synapses (inset).

(C) Relation between threshold stability parameter, κ, and storage reliability. A hypothetical probability density function for an input is plotted with respect to θ and κ. The filled area of the distribution tail indicates the probability of erroneous output when an action potential is the correct output (ϵ_1).

(D) A family of subcritical weight distributions for $f' = 0.25$ and $\rho = 2.1$. The curves shown are for numbers of associations per synapse $\alpha = \alpha_c = 0.33$ (thick solid line), $0.9\alpha_c$ (thin solid line), $0.8\alpha_c$ (dot-dashed line), $0.5\alpha_c$ (dashed), and α small (dotted).

of weights with average θ/fN. As more associations are learned, additional constraints upon the distribution are introduced, causing it to broaden. Since the synaptic weights are constrained to be nonnegative, more and more synapses assume low weights. At critical capacity, the distribution is fully constrained, and these synapses become "silent."

Granule Cell-Purkinje Cell Synaptic Weights Conform to the Optimal Distribution

Having derived and analyzed the perceptron weight distribution, we compared it to that measured experimentally for granule cell-Purkinje cell synapses. The experimental distribution was extracted from Isope and Barbour (2002). That paper established a discrepancy between the observed probability of detecting a connection electrically ($\sim$10%) and the predicted probability of a synaptic connection existing ($\sim$50%). The latter prediction was based upon the detailed stereology of Napper and Harvey (1988a). It was deduced that $\sim$80% of parallel fiber-Purkinje cell synapses did not generate detectable responses. See the Experimental Procedures for a description of the experimental distribution and of the fitting procedures. We fitted the optimal distribution (i.e., at critical capacity) to the experimental distribution using a maximum likelihood method, fixing $N = 150{,}000$ and $f' = 0.37$ (see below), the latter having very little influence upon the fit. The resulting best fit is shown superimposed upon the experimental weight distribution in Figure 6A. The parameter values of the best fit were the following: the reliability parameter was $\rho = 2.0 \pm 0.3$ ($\pm$ standard error estimate) and the average synaptic strength (including silent synapses) $\overline{w} = 0.015 \pm 0.002$ mV. We examined the sensitivity of these parameters to systematic errors in the measurement of synaptic weights or the fraction of undetected synapses (Figures 6B and 6C). It is noteworthy that ρ is insensitive to systematic errors of weight measurement.

We used the fit to constrain the values of the three remaining model parameters (Equations 1 and 2): the fraction of active inputs f, threshold θ, and stability parameter κ. Since the fit provides two values and there are three unknowns, we plot the possible combinations between these three parameters in Figure 6D. If we assume that threshold is $\sim$10 mV, relative to the resting potential (Isope and Barbour, 2002), we obtain $f = 0.0044 \pm 0.0006$ and $\kappa = 0.78 \pm 0.08$ mV. The effects of systematic errors of weight measurement upon these values can be predicted using Equations 1 and 2. Thus, f is proportional to $1/\overline{w}$, while κ is less sensitive to this error, being proportional to $\sqrt{1/\overline{w}}$ (for small f).

We also fitted a subcritical distribution to the experimental distribution. It turned out that the best fit was very similar to the critical distribution, suggesting that the Purkinje cell operates near critical capacity. The parameter values obtained were $\alpha = (0.97 \pm 0.03)\alpha_c$, $\overline{w} = 0.016 \pm 0.002$ mV, $\kappa = 0.80 \pm 0.07$, and $f = 0.0042 \pm 0.0005$. The fit also provided an estimate of $f' = 0.37 \pm 0.05$. The subcritical fit was carried out for various fractions of silent synapses in order to test the influence of systematic errors in this parameter. The resulting best-fit values are illustrated in Table 1.

Deducing the Storage Capacity of the Purkinje Cell

The number of distinct input-output associations that are learned is of fundamental importance to Purkinje cell operation, since this is related to the number of

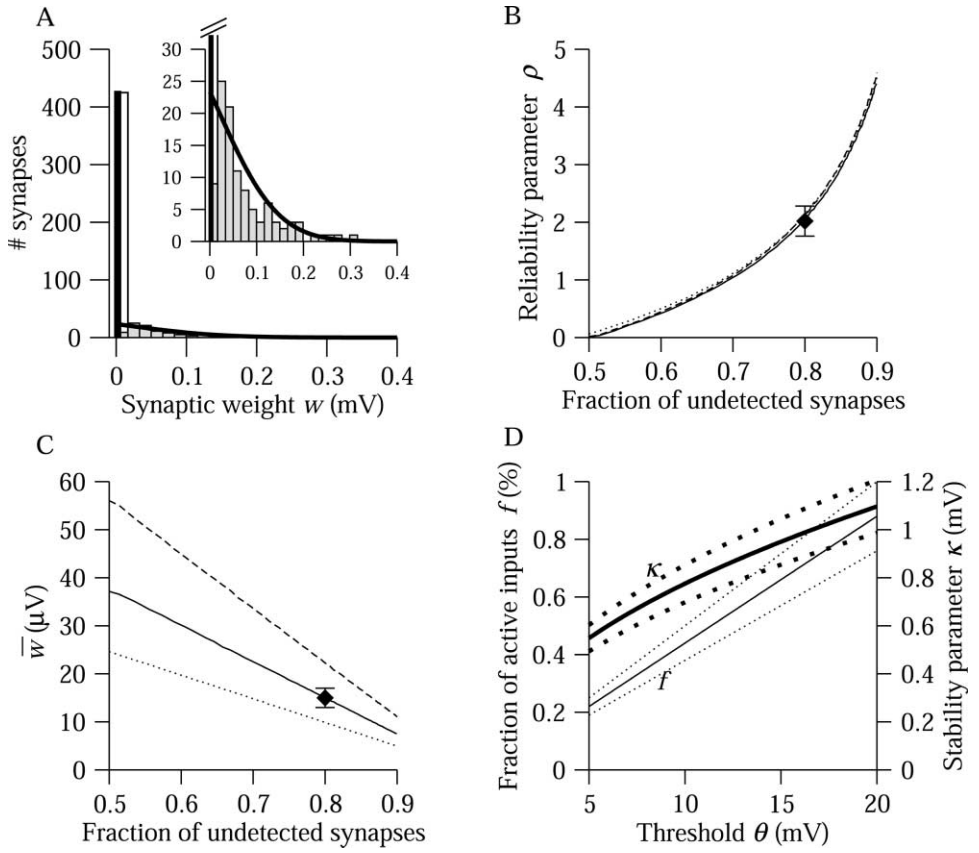

Figure 6. Comparison of Theoretical and Empirical Weight Distributions

(A) Histogram of unitary granule cell-Purkinje cell synaptic weights extracted from the data of Isope and Barbour (2002). Gray bars, n = 104 detected connections; open bar, estimated 80% undetected connections (estimated detection threshold of 16.6 μV). The solid line and bar shows the fit of the perceptron model at critical capacity to the data. The fit yields (for $f' = 0.37$) $\alpha = \alpha_c = 0.3$. Inset shows the detail of the detected connections.

(B) Best-fit value for ρ (black lozenge), bootstrap standard error estimate (error bars), and curves showing the sensitivity of this parameter to the fraction of undetectable synapses. The different curves examine the potential effect of systematic mismeasurement of the synaptic weights: overestimation by a factor of 1.5 for the dashed line, underestimation by a factor of 1.5 for the dotted line. As expected, ρ depends only upon the shape of the weight distribution.

(C) Similar representation of best-fit value, standard error, and sensitivity of $\bar{w}$.

(D) Plot of κ and f as a function of θ, determined from the best-fit values of ρ and $\bar{w}$ (solid lines). The dotted lines indicate the bootstrap standard error estimates.

different motor behaviors to which it can contribute. It is well known from theoretical studies that the storage capacity of a perceptron is of the order of 1 bit/synapse (Cover, 1965; Gardner, 1988). Both Marr (1969) and Albus (1971) offered estimates of the storage capacity of the Purkinje cell. Marr suggested greater than 200 associations, assuming binary synapses and a nonoptimal learning rule. Albus suggested 200,000, calculating the capacity according to Cover (1965), which is equivalent to the special case of $\kappa = \rho = 0$, $f = f' = 0.5$ and allowing synapses to have negative weights. Using our optimum fit parameters, we reexamined the storage capacity of the Purkinje cell (Figure 7). The number of associations per synapse, α_c, decreases as either ρ (Figure 7A) or f' (Figure 7B) increases. The actual number that can be stored depends upon the fraction of active

Table 1. Parameter Sensitivity to Systematic Errors in the Fraction of Undetected Synapses

Undetected Synapses	ρ	$\bar{w}$ (μV)	κ (mV)	f	f'	α	α_c	α/α_c
0.50	0.55	0.038	0.36	0.0017	0.26	0.46	0.70	0.66
0.60	0.59	0.033	0.33	0.0020	0.42	0.51	0.62	0.83
0.70	1.39	0.022	0.65	0.0031	0.49	0.31	0.34	0.89
0.80	2.00	0.016	0.80	0.0042	0.37	0.26	0.27	0.97
0.90	2.53	0.010	0.81	0.0064	0.25	0.24	0.24	0.99

Parameter values resulting from fitting the subcritical distribution assuming different values for the fraction of undetected synapses.

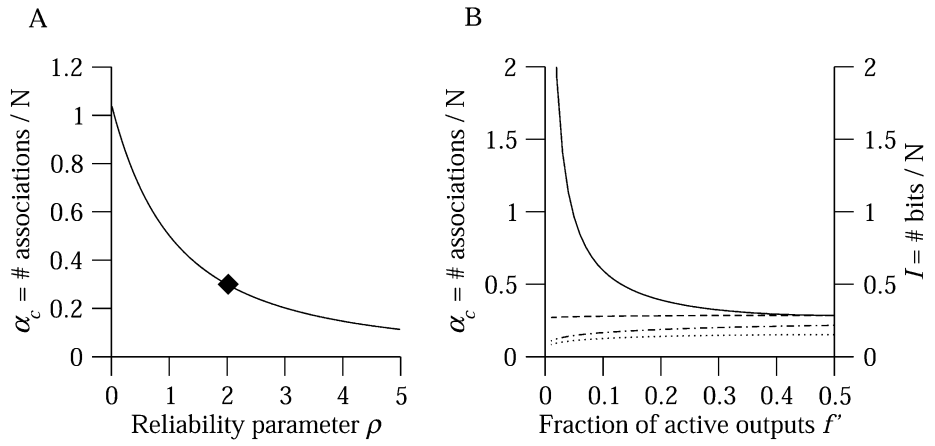

Figure 7. Storage Capacity of the Purkinje Cell

(A) The maximum number of associations stored, α_c, depends upon the value of the reliability parameter ρ (here $f' = 0.37$). The black lozenge indicates the value of ρ given by the best (critical) fit of the experimental data (see Figure 6), corresponding to a maximal capacity of $\alpha_c = 0.3$. (B) Dependence of the Purkinje cell storage capacity upon the fraction of active outputs f'. The capacity can be expressed in terms of the number of associations, α_c (solid line, left axis), or information, I (right axis): dashed line, $\epsilon_1 = \epsilon_2 = 0$; dot-dashed, $\epsilon_1 = 0$, $\epsilon_2 = 0.1$; dotted, $\epsilon_1 = \epsilon_2 = 0.1$, $\rho = 2$.

outputs, f'; using the best-fit parameters ($\rho = 2$, $f' = 0.37$, $\alpha = 0.26$), we calculate that the Purkinje cell stores 40,000 different associations.

Storage capacity can also be represented as information stored per synapse. If $p = \alpha_c N$ patterns are stored with error rates ϵ_1 among the $f'p$ patterns with a positive (action potential) output and ϵ_2 among the $(1 - f')p$ patterns with a null output, the total number of errors being $\epsilon_1 f' p + \epsilon_2(1 - f')p$, the amount of information in bits per synapse is (see, for example, Nadal and Toulouse, 1990)

$$I = \alpha_c\{s[(1 - \epsilon_1)f' + \epsilon_2(1 - f')] - f's[\epsilon_1] - (1 - f')s[\epsilon_2]\} \quad (3)$$

where $s[f] \equiv -f \ln_2 f - (1 - f) \ln_2 (1 - f)$. Note that $I = \alpha_c s[f']$ in the error-free case. As a function of ρ, the information is proportional to α_c, but, in contrast to α_c, the information depends only weakly upon f' (Figure 7B). Applying again our best-fit parameters, we estimate that a Purkinje cell stores the equivalent of up to 0.25 bits/synapse or a total of 5 kilobytes of information, though significantly reduced information capacity would result from probabilistic storage (Figure 7B).

Distribution of Compound EPSP Amplitudes Elicited by Input Patterns

Our theory generates a simple prediction: in order to operate reliably in the presence of noise, compound EPSPs elicited by learned patterns should be either significantly above threshold (when the pattern should elicit an action potential) or significantly below it (in the opposite case). We calculated the distribution of compound EPSPs elicited by input patterns (see the Supplemental Data [http://www.neuron.org/cgi/content/full/43/5/745/DC1]). The distribution is shown in Figure 8. As expected, the distribution in the absence of noise (equivalent to the distribution of average amplitudes) is bimodal: a fraction $(1 - f')$ of patterns elicit compound EPSPs smaller than $\theta - \kappa$, while the remaining patterns elicit compound EPSPs greater than $\theta + \kappa$. There are two delta functions at $\theta - \kappa$ and $\theta + \kappa$, and Gaussian tails to the left of $\theta - \kappa$ and to the right of $\theta + \kappa$. When noise is added to the compound EPSPs, the peaks at $\theta \pm \kappa$ broaden, but the distribution remains bimodal while the noise level is smaller than κ. For large noise levels, the distribution becomes unimodal, indicating a large fraction of errors.

Discussion

A Theory of Synaptic Weights

By analyzing a perceptron with excitatory synapses at critical capacity, we have developed a theory of synaptic weights according to which their optimal distribution is a Gaussian tail added to a delta function of silent synapses. This shape of distribution is quite different from those reported in other modeling studies, which mostly examined the result of applying spike timing-dependent plasticity rules to model neural networks (van Rossum et al., 2000; Song et al., 2000; Rubin et al., 2001; Cateau and Fukai, 2003; Gütig et al., 2003). These studies generated two basic types of weight distribution. The first of these is a bimodal distribution where weights cluster at the lower and upper bounds upon synaptic weights (the bounds being necessary to prevent divergence of the weights). The second type, obtained with modified rules, is a unimodal distribution with a positive mode.

The perceptron optimal distribution fits well that reported for granule cell-Purkinje cell synapses. How much support can be derived for the theory from this fit? And what specific predictions does our theory generate?

Two central features of the perceptron weight distribution do not depend upon particular parameter values: (1) the requirement for a majority of silent synapses

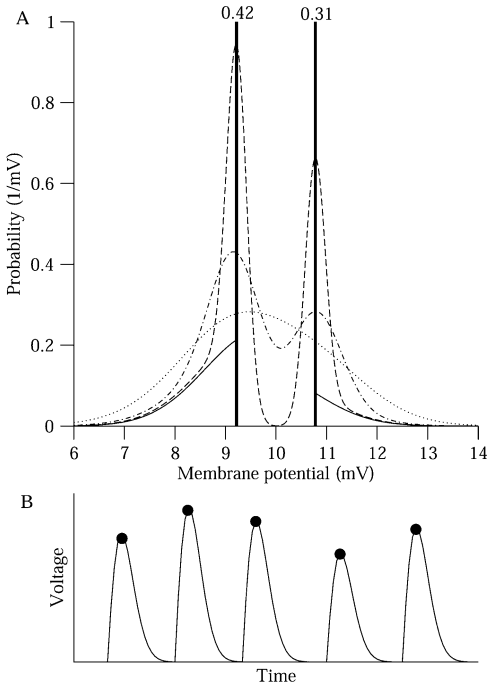

Figure 8. Distribution of Compound EPSP Amplitudes Induced by Input Patterns

(A) Several distributions of compound EPSP amplitudes corresponding to different noise levels (with standard deviation σ) are shown: solid lines, $\sigma = 0$; dashed line, $\sigma = 0.2$ mV; dot-dashed line, $\sigma = 0.5$ mV; dotted line, $\sigma = 1$ mV. In the case $\sigma = 0$ delta functions (marked by vertical bars) are present at $\theta - \kappa$ and $\theta + \kappa$; the weight of the corresponding delta functions is indicated on top of the corresponding vertical bars. The distribution was calculated using the best fits parameters for the model at critical capacity ($f = 0.0044$, $\kappa = 0.78$ mV, $\theta = 10$ mV, $N = 150{,}000$, and $f' = 0.37$).

(B) Sketch of how the distribution of compound EPSP amplitudes could be obtained from in vivo recordings when action potentials are blocked, during execution of learned movements. One expects a series of depolarizing peaks interleaved with inhibitory episodes. The amplitudes of the peaks (filled circles) should be distributed according to the distribution shown in (A). Note that the depolarizations need not be periodic.

stems directly from the condition of optimality, with only the exact fraction between 50% and 100% depending upon the parameter values; (2) over positive weights, the distribution decays monotonically in all situations. Thus, if either a majority of synapses are not silent or the distribution is nonmonotonic, our conclusions would be invalidated. Conversely, since the perceptron optimal distribution is currently the only theoretical distribution possessing these properties, satisfying these conditions would provide strong support for our theory.

The first of these conditions is supported by the reported presence of a large majority of silent granule cell-Purkinje cell synapses (Isope and Barbour, 2002). Although the evidence is indirect, it emerges from a comparison between exhaustive stereology studies (Harvey and Napper, 1988, 1991; Napper and Harvey, 1988a, 1988b) and electrophysiology designed to exploit them (Isope and Barbour, 2002). Moreover, it should be noted that control experiments (Isope and Barbour, 2002) ruled out numerous possible artifacts, including shunting or filtering of synaptic responses, sectioning of axons, or action potential propagation block. Furthermore, experimental conditions limited synaptic plasticity and damage to slices. The existence of a large majority of silent granule cell-Purkinje cell synapses has also been deduced from recordings in vivo (Ekerot and Jörntell, 2001; Jörntell and Ekerot, 2002). The second condition, of monotonicity, is directly verified down to the detection threshold (Figure 6A).

By fitting our model to the experimental data, we obtained values for the model parameters. These may therefore be considered additional quantitative predictions of the model. This essentially concerns the parameters describing the pre- and postsynaptic activity levels ($f = 0.4\%$ and $f' = 37\%$, see also Table 1). Their values arise in different ways from the model. The low value of f is determined by the simple relation of Equation 1, reflecting the fundamental requirement that the Purkinje cell, despite the huge number of excitatory synapses it receives, should spend some time below threshold. In contrast, the value for f' was obtained from the subcritical fit, which depends upon all of the model parameters.

The activity parameters would be best studied in the framework of our predictions for the distribution of the compound EPSP amplitudes elicited by input patterns (Figure 8). Two interesting features of this distribution are that (1) it is bimodal (unless noise renders operation completely unreliable); and (2) that even input patterns that do not generate an action potential all elicit quite large depolarizations. These predictions could be tested by in vivo intracellular recordings in which action potential emission was blocked using an internally acting local anesthetic. In such a recording, during learned motor behaviors, we would predict the appearance of a sequence of brief depolarizations interspersed by inhibitory phases (as sketched in Figure 8B). The peaks of the depolarizations should distribute bimodally according to Figure 8 (depending upon the noise levels, which could be measured directly). The value of f' would be given by the fraction of these peaks that exceed the level of threshold, while the value for f could be evaluated by dividing the mean peak depolarization by $\overline{w}$.

It is interesting to ask whether it will be possible to test our theory by modulating fundamental quantities such as the amount of information learned by a Purkinje cell. In particular, it is tempting to investigate the weight distribution in young animals, i.e., before they have learned much. If the only difference was the number of associations learned, we would predict lower numbers of silent synapses in the young than in the adult. However, an alternative scenario would predict just the opposite: younger animals might have learned fewer associations, but very reliably (see below). In this scenario, the animal would follow the curve of α_c versus ρ of Figure 7A and gradually move toward the left as learning progresses. The fraction of silent synapses would then *decrease* with learning. Finally, such measurements could be confounded by developmental processes and synaptogenesis, which occur postnatally during the acquisition of motor coordination. Of particular concern is the probability that newly formed synapses would be silent, as occurs in other brain regions.

Relaxing Model Assumptions

Excitatory inputs to the Purkinje cell combine to drive it to threshold. The assumption of linear summation in our model is a simple representation of this process. The experiments of Figures 2–4 support this approximation. However, the model is able to accommodate a number of departures from this ideal behavior.

The basic features of the optimal distribution do not depend upon summation being exactly linear. We present in the Supplemental Data (http://www.neuron.org/cgi/content/full/43/5/745/DC1) a scenario of nonlinear summation which leads to a nonlinear transformation of the weight axis of the distribution. The distribution retains its main qualitative features (majority of silent synapses and monotonic decay of positive weights), but it is no longer a Gaussian at positive weights.

A more extreme nonlinearity, in which action potentials are initiated within dendritic compartments, would require a change in our perceptron model. One approach to modeling such a situation would be to represent the cell as a two-layer neural network (Poirazi et al., 2003; Polsky et al., 2004). The optimal weight distribution for such a network is significantly more difficult to obtain and remains to be calculated. However, we expect that in this case optimizing information storage would also lead to a majority of silent synapses, since each computational unit in the input layer (e.g., dendrite) would be similar to the one-layer perceptron with positive weights modeled here.

Several processes can be represented at least in part as a noise source, thereby influencing κ. Examples of this would be membrane potential noise (even spontaneous activity: Häusser and Clark, 1997), jitter, or errors within learned input patterns. Other processes would effectively scale the synaptic weights or the threshold. This is the case for several background effects of inhibition:

- Shunting inhibition could perform a divisive reduction of excitatory inputs (decrease synaptic weights w_i).
- Hyperpolarizing inhibition could perform a subtractive reduction of excitatory inputs (increase θ).
- Inhibition could increase noise (increase κ or decrease reliability).

In each case, the optimal weight distribution would still contain a majority of silent synapses and would still decay monotonically.

Paired-pulse facilitation of inputs could in part be represented by scaling the synaptic weights and κ. If granule cell inputs are correlated rather than random as modeled here (but see Schweighofer et al., 2000; Coenen et al., 2001), then the critical capacity α_c remains unchanged, but the information storage I is reduced (Monasson, 1992).

The Role of Inhibition

Until recently, it was generally considered that the time scale of synaptic integration was set by the membrane time constant, which in slice preparations was often in the range 10–100 ms. Nevertheless, it was also recognized that this time constant would be reduced by synaptic activity (Bernander et al., 1991) and that this effect may be very significant under in vivo conditions (Destexhe and Paré, 1999).

Although feedforward inhibition is hardly a new discovery (Fatt and Katz, 1953; Eccles et al., 1967), two recent works (Pouille and Scanziani, 2001; Wehr and Zador, 2003) have now clearly established the importance of the excitation-inhibition sequence and the fact that it limits synaptic integration to a window that can be one or two orders of magnitude shorter than the membrane time constant. We show that a similar mechanism operates in the Purkinje cell and, in addition, that coincident inputs sum linearly. These results suggest a simple mode of synaptic integration that allows the Purkinje cell to subserve a precise timing function and also to perform many independent calculations in a short time (Isope et al., 2002).

Albus (1971) suggested that granule cell-interneuron synapses also store information, and this notion is supported by recent work on plasticity at this synapse (Jörntell and Ekerot, 2002, 2003). If synaptic weights in the perceptron can be negative as well as positive, the optimal weight distribution would be a complete Gaussian with a positive mean, and the storage capacity would be doubled (Amit, 1989; Nadal, 1990). This raises the question of whether addition of interneuron IPSPs to granule cell EPSPs would allow effectively unconstrained weights in order to exploit this extra capacity. In fact, the measured excitatory weight distribution is incompatible with this hypothesis for the function of the interneuronal plasticity. Basically, because a majority of excitatory (granule cell-Purkinje cell) synapses are silent, the distribution that would result from adding inhibitory weights to the excitatory weights could not be Gaussian *and* have a positive mean, irrespective of the distribution of the inhibitory weights and of how they are added to the excitatory weights. Moreover, the summation aspect of this hypothesis would be difficult to implement, because excitatory and inhibitory inputs are not simultaneous (which would have made summing them easy and unavoidable), but temporally disjoint, with inhibition appearing to succeed and dominate excitation (Figures 2–4).

Consequences for Cerebellar Operation

The analytical nature of our theory allows us to identify two novel roles for silent synapses: (1) a majority of silent synapses is a necessary byproduct of optimal learning; (2) additional silent synapses are required for storage reliability. These roles for silent synapses contrast with the current consensus that they are a stage of synaptogenesis or development (Durand et al., 1996; Isaac et al., 1997; Rumpel et al., 1998; Petralia et al., 1999) and could explain their continued presence in the adult (Nusser et al., 1998; Takumi et al., 1999; Kharazia and Weinberg, 1999; Ekerot and Jörntell, 2001; Isope and Barbour, 2002). That so many silent synapses are maintained in adult life may appear paradoxical (since "silent synapses do nothing"). However, if a synapse exists, whether it is silent or not will clearly alter the postsynaptic neuron's behavior, so in this sense it is "doing something." Once a synapse has been silenced, it is true that it could be removed without altering circuit

behavior, *but only if no new learning were required*. New learning might require a silent synapse to adopt a nonzero weight. This condition provides a plausible explanation for the retention of silent synapses that is reported in the adult cerebellum (Isope and Barbour, 2002; Ekerot and Jörntell, 2001). Thus, if the ~80% of synapses thought to be silent were eliminated, the storage capacity for learning new associations or modifying existing associations would be reduced in proportion (i.e., to ~20% of the original capacity). Many studies of behavioral models of cerebellar learning have confirmed that the cerebellar circuitry remains plastic in the adult. About 50% of parallel fibers traverse the Purkinje cell dendritic tree without synapsing. The importance of this kind of "potential synapse" for plasticity has recently been emphasized in another context (Stepanyants et al., 2002). Here, it is possible that these unconnected fibers represent synapses that existed initially, were silenced, and were then eliminated (though we are unaware of any evidence for such elimination). In this case, N and the fraction of silent synapses should be increased in the fits. For $N = 300{,}000$ and 90% silent synapses, we obtain $f = 0.0043$ and $\kappa = 0.83$.

Fitting the subcritical distribution to the experimental one led to the conclusion that the Purkinje cell operates very close to critical capacity. We interpret this result as follows. It seems unlikely that a Purkinje cell implements a fixed κ or has the fixed capacity it would entail. Instead, we believe that, for a given number of associations, the Purkinje cell optimizes their learning by minimizing errors due to noise (see Figure 5C). By extending the approach of Gardner and Derrida (1988) and Griniasty and Gutfreund (1991) (see the Supplemental Data [http://www.neuron.org/cgi/content/full/43/5/745/DC1]), it can be shown that minimizing errors for a given number of associations in the presence of noise leads to an identical optimal distribution of synaptic weights. However, the threshold stability parameter κ is then replaced by a function of the noise SD (σ) and the error probability (ϵ) (for instance, requiring a small fraction of errors for a noise of significant amplitude is roughly equivalent to having a large κ). Through ρ, there is a fundamental trade-off between capacity and reliability (Figures 5 and 7). Minimizing errors would therefore provide a flexible optimization scheme which could store a few associations reliably yet accommodate many associations with a graceful degradation in reliability.

Although noise sources in vivo remain to be characterized, the value we obtain for κ (0.8 mV) appears low for very reliable storage. For example, 95% reliability ($\epsilon = 0.05$) would require the combined noise sources to have $\sigma \leq 0.4$ mV. It seems reasonable to assume that noise sources collectively exceed this level, so we are led to conclude that Purkinje cells store associations in a probabilistic manner. There is evidence for such "population coding" by Purkinje cells (Thier et al., 2000). It would still be possible to recover a reliable output from probabilistic Purkinje cells by averaging together the outputs from several cells. The obvious location for such averaging would be the deep cerebellar nuclei.

Our theory supports the long-standing predictions that granule cell coding is sparse (estimates were about 1% average activity: Marr, 1969; Albus, 1971; Schweighofer et al., 2000; Coenen et al., 2001), though with the caveat that it applies only to the fraction of granule cells contributing to synchronous inputs. The benefit of the sparse coding is to increase storage capacity, through its effect on the reliability parameter ρ: for a given stability constant κ, Equation 2 shows that decreasing f decreases ρ, which in turn increases α_c (Figure 7A). The existence of an optimal distribution requiring a majority of silent synapses is, however, independent of the granule cell activity levels and is not specific to sparse coding. In contrast, the high value for f' appears to confer little benefit in terms of information storage; this behavior may thus be imposed by the motor output requirements. One such requirement might be rapid modulation of firing rate in both directions, something that a high basal frequency makes possible.

Generalization of the Theory

The perceptron can be considered the "prototypical" learning machine. Many of the analytical results obtained for it can be generalized or extended to more complicated neural network topologies. The present optimal weight distribution derives from such basic assumptions (excitatory synapses, optimality) that it is reasonable to consider whether it might apply to other excitatory connections. How do our two key predictions of a monotonic weight distribution and a majority of silent synapses fit measurements for other synapses?

Synaptic weight distributions for recurrent connections between neocortical pyramidal cells are the best studied; the distribution of Holmgren et al. (2003) shows a monotonic decrease as a function of synaptic weight, as predicted here. Other published distributions appear to have a maximum at positive weight (Markram et al., 1997; Sjöström et al., 2001), in apparent contradiction with our optimal distribution. However, experimental noise can convert a monotonic distribution to one with a maximum at positive weight, since the noise prevents detection of very weak inputs, causing them to be underrepresented.

How many silent synapses are there at interpyramidal cell connections? The observed probability of detecting a connection (~10%: Sayer et al., 1990; Mason et al., 1991; Markram et al., 1997; Sjöström et al., 2001; Holmgren et al., 2003) certainly leaves room for a large proportion of undetected connections, and a least some silent synapses are thought to exist in both hippocampus and neocortex (Liao et al., 1995; Isaac et al., 1995; Nusser et al., 1998; Takumi et al., 1999; Petralia et al., 1999; Montgomery et al., 2001; Kharazia and Weinberg, 1999). A suggestive result is that of Lübke et al. (1996), who demonstrated that 80% of pyramidal cells form apparently nonfunctional autapses (synapses upon themselves). It is difficult to rationalize why pyramidal cells should be eight times more likely to contact themselves than their nearest neighbors. Our theory can reconcile these discordant observations: the local pyramidal cell network may be approximately fully connected, but with a majority of those connections being silent.

In summary, it is possible that learning at other synapses is optimized in a manner related to our theory's predictions. However, testing this hypothesis satisfactorily will require better quantification of the numbers of silent and weak synapses, further investigation of the

mechanisms of synaptic integration, extension of our model to more complex neural network topologies, and an understanding of what and how pyramidal cells learn.

Experimental Procedures

Recording Techniques

Protocols conformed to national and National Institutes of Health guidelines on animal experimentation. Rats about 20d were deeply anesthetized (ketamine/xylazine $\geq$ 75/10 mg kg^{-1} i.p.) and decapitated. The cerebellum was rapidly dissected into a low-sodium solution (230 mM sucrose, 26 mM $NaHCO_3$, 3 mM KCl, 0.8 mM $CaCl_2$, 8 mM $MgCl_2$, 1.25 mM NaH_2PO_4, 25 mM D-glucose, 1 mM ketamine-HCl; 0°C–5°C, bubbled with 95% O_2/5% CO_2). Transverse slices (450 μm) were prepared in the same solution. The low-NaCl and ketamine helped protect interneurons during slicing. Slice maintenance and recording were carried out in the following solution: 125 mM NaCl, 3 mM KCl, 26 mM $NaHCO_3$, 1.25 mM NaH_2PO_4, 2 mM $CaCl_2$, 2 mM $MgCl_2$, 25 mM D-glucose, bubbled with 95% O_2/5% CO_2. For recording, slices were held in a chamber allowing solution flow above and below the slice (Isope and Barbour, 2002). Patch-clamp recording of Purkinje cells used the following pipette solution: 150 mM K-gluconate, 4 mM NaCl, 10 mM HEPES, 10 mM Mg-ATP, adjusted to pH 7.3 with KOH at 300 mOsm. Recordings were corrected for a junction potential of 10 mV. Recording temperature was 32°C. Optimal capacitance neutralization was employed, and cells were maintained near −65 mV with 100–400 pA of hyperpolarizing current. Monopolar stimulation in the granule cell layer was carried out using a saline-filled patch pipette connected to an isolated stimulator. Granule cell-layer stimulation was preferred to molecular-layer stimulation of parallel fibers, because the latter activates unphysiological dense bundles, while the former evokes a more diffuse input. Results are reported as mean $\pm$ SD, and stimulus artifacts have been blanked in the figures.

The Experimental Weight Distribution

The empirical distribution of granule cell-Purkinje cell synaptic weights was obtained from paired recordings in Isope and Barbour (2002). In that work, granule cells were stimulated in two locations: 300–500 μm from the Purkinje cell plane (in the parallel fiber direction) or close to the Purkinje cell. The distant granule cells made parallel fiber contacts, while some proximal granule cells may have made synapses via "ascending axons." Because the weight distribution for proximal granule cells was similar to that for distal granule cells, we combined weights for all recorded granule cells (n = 104 detected connections), excluding only three very strong connections almost certainly made by ascending axons (such connections compose only a tiny minority of granule cell-Purkinje cell connections). Since proximal granule cells have a slightly higher apparent connection probability, we took 80% (rather than 85%) as our estimate of the fraction of undetected ($\leq$2 pA) synapses for this composite distribution; the sensitivity of our results to the precise value of this fraction was tested (Results). We converted EPSC amplitudes to somatic depolarizations using a conversion factor of 8.3 μV/pA.

Fitting Procedures

We fitted the theoretical weight distribution to the experimental one using a maximum likelihood method followed by a bootstrap analysis (Efron and Tibshirani, 1993) to estimate standard errors (10,000 bootstrap samples). The predicted weight distribution at critical capacity is part of a Gaussian with a delta of silent synapses (see the Supplemental Data, Equation 34 [http://www.neuron.org/cgi/content/full/43/5/745/DC1]). It is described by two parameters, W_s and B, the latter controlling the fraction of silent synapses. The aim was to find the optimum values for these parameters. In calculating the maximum likelihood, undetected synapses and those below the detection threshold T = 16.6 μV (2 pA) were assigned a uniform probability density.

We also fitted the data with subcritical distributions (Supplemental Data, Equation 31 [http://www.neuron.org/cgi/content/full/43/5/745/DC1]). The goal was to find the maximum of the likelihood function in a space of four parameters: κ, f', α, and $\overline{w}$. The optimization procedure was performed using the downhill simplex method (Press et al., 1992). Synapses below the detection threshold were treated as described above. Bootstrap analysis was limited to estimating the parameter standard errors from 200 bootstrap samples because finding numerically the best fit parameters of each bootstrap sample took several hours of computing time.

Acknowledgments

This work was supported by the CNRS, the Ecole Normale Supérieure, and the FET unit of the European Commission (grant no. IST-2001-35271). We thank Philippe Ascher, David Attwell, Clément Léna, Walter Senn, and anonymous referees for suggesting significant improvements to the manuscript.

Received: February 3, 2004
Revised: July 21, 2004
Accepted: August 16, 2004
Published: September 1, 2004

References

Albus, J.S. (1971). A theory of cerebellar function. Math. Biosci. *10*, 26–51.

Amit, D. (1989). Modeling Brain Function (Cambridge: Cambridge University Press).

Bernander, O., Douglas, R.J., Martin, K.A., and Koch, C. (1991). Synaptic background activity influences spatiotemporal integration in single pyramidal cells. Proc. Natl. Acad. Sci. USA *88*, 11569–11573.

Cateau, H., and Fukai, T. (2003). A stochastic method to predict the consequence of arbitrary forms of spike-timing-dependent plasticity. Neural Comput. *15*, 597–620.

Chadderton, P., Margrie, T.W., and Häusser, M. (2004). Integration of quanta in cerebellar granule cells during sensory processing. Nature *428*, 856–860.

Coenen, O.J.-M.D., Arnold, M.P., Sejnowski, T.J., and Jabri, M.A. (2001). Parallel fiber coding in the cerebellum for life-long learning. Auton. Robots *11*, 291–297.

Cover, T.M. (1965). Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. IEEE Trans. Electron. Comput. *14*, 326.

De Zeeuw, C.I., Hansel, C., Bian, F., Koekkoek, S.K., van Alphen, A.M., Linden, D.J., and Oberdick, J. (1998). Expression of a protein kinase C inhibitor in Purkinje cells blocks cerebellar LTD and adaptation of the vestibulo-ocular reflex. Neuron *20*, 495–508.

Destexhe, A., and Paré, D. (1999). Impact of network activity on the integrative properties of neocortical pyramidal neurons in vivo. J. Neurophysiol. *81*, 1531–1547.

Durand, G.M., Kovalchuk, Y., and Konnerth, A. (1996). Long-term potentiation and functional synapse induction in developing hippocampus. Nature *381*, 71–75.

Eccles, J.C., Ito, M., and Szentágothai, J. (1967). The Cerebellum as a Neuronal Machine (New York: Springer-Verlag).

Efron, B., and Tibshirani, R.J. (1993). An Introduction to the Bootstrap (New York: Chapman Hall).

Ekerot, C.F., and Jörntell, H. (2001). Parallel fibre receptive fields of Purkinje cells and interneurons are climbing fibre-specific. Eur. J. Neurosci. *13*, 1303–1310.

Engel, A., and van den Broeck, C. (2001). Statistical Mechanics of Learning (Cambridge: Cambridge University Press).

Fatt, P., and Katz, B. (1953). The effect of inhibitory nerve impulses on a crustacean muscle fibre. J. Physiol. *121*, 374–389.

Gardner, E. (1988). The phase space of interactions in neural network models. Journal of Physics A: Mathematical and General *21*, 257–270.

Gardner, E., and Derrida, B. (1988). Optimal storage properties of neural network models. Journal of Physics A: Mathematical and General *21*, 271–284.

Griniasty, M., and Gutfreund, H. (1991). Learning and retrieval in

attractor neural networks above saturation. Journal of Physics A: Mathematical and General *24*, 715–734.

Gutfreund, H., and Stein, Y. (1990). Capacity of neural networks with discrete synaptic couplings. Journal of Physics A: Mathematical and General *23*, 2613–2630.

Gütig, R., Aharonov, R., Rotter, S., and Sompolinsky, H. (2003). Learning input correlations through nonlinear temporally asymmetric Hebbian plasticity. J. Neurosci. *23*, 3697–3714.

Harvey, R.J., and Napper, R.M. (1988). Quantitative study of granule and Purkinje cells in the cerebellar cortex of the rat. J. Comp. Neurol. *274*, 151–157.

Harvey, R.J., and Napper, R.M. (1991). Quantitative studies on the mammalian cerebellum. Prog. Neurobiol. *36*, 437–463.

Häusser, M., and Clark, B.A. (1997). Tonic synaptic inhibition modulates neuronal output pattern and spatiotemporal synaptic integration. Neuron *19*, 665–678.

Hebb, D.O. (1949). Organization of Behavior (New York: Wiley).

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.

Isaac, J.T., Nicoll, R.A., and Malenka, R.C. (1995). Evidence for silent synapses: implications for the expression of LTP. Neuron *15*, 427–434.

Isaac, J.T., Crair, M.C., Nicoll, R.A., and Malenka, R.C. (1997). Silent synapses during development of thalamocortical inputs. Neuron *18*, 269–280.

Isope, P., and Barbour, B. (2002). Properties of unitary granule cell → Purkinje cell synapses in adult rat cerebellar slices. J. Neurosci. *22*, 9668–9678.

Isope, P., Dieudonné, S., and Barbour, B. (2002). Temporal organization of activity in the cerebellar cortex: a manifesto for synchrony. Ann. N Y Acad. Sci. *978*, 164–174.

Ito, M., and Kano, M. (1982). Long-lasting depression of parallel fiber-Purkinje cell transmission induced by conjunctive stimulation of parallel fibers and climbing fibers in the cerebellar cortex. Neurosci. Lett. *33*, 253–258.

Ito, M., Sakurai, M., and Tongroach, P. (1982). Climbing fibre induced depression of both mossy fibre responsiveness and glutamate sensitivity of cerebellar Purkinje cells. J. Physiol. *324*, 113–134.

Jörntell, H., and Ekerot, C.F. (2002). Reciprocal bidirectional plasticity of parallel fiber receptive fields in cerebellar Purkinje cells and their afferent interneurons. Neuron *34*, 797–806.

Jörntell, H., and Ekerot, C.F. (2003). Receptive field plasticity profoundly alters the cutaneous parallel fiber synaptic input to cerebellar interneurons in vivo. J. Neurosci. *23*, 9620–9631.

Kharazia, V.N., and Weinberg, R.J. (1999). Immunogold localization of AMPA and NMDA receptors in somatic sensory cortex of albino rat. J. Comp. Neurol. *412*, 292–302.

Kohler, H.M., and Widmaier, D. (1991). Sign-constrained linear learning and diluting in neural networks. Journal of Physics A: Mathematical and General *24*, L495–L502.

Liao, D., Hessler, N.A., and Malinow, R. (1995). Activation of postsynaptically silent synapses during pairing-induced LTP in CA1 region of hippocampal slice. Nature *375*, 400–404.

Lübke, J., Markram, H., Frotscher, M., and Sakmann, B. (1996). Frequency and dendritic distribution of autapses established by layer 5 pyramidal neurons in the developing rat neocortex: comparison with synaptic innervation of adjacent neurons of the same class. J. Neurosci. *16*, 3209–3218.

Markram, H., Lübke, J., Frotscher, M., Roth, A., and Sakmann, B. (1997). Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. J. Physiol. *500*, 409–440.

Marr, D. (1969). A theory of cerebellar cortex. J. Physiol. *202*, 437–470.

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.

Mézard, M., Parisi, G., and Virasoro, M.A. (1987). Spin Glass Theory and Beyond (Singapore: World Scientific).

Minsky, M., and Papert, S. (1988). Perceptrons: An Introduction to Computational Geometry, Expanded Edition (Cambridge, MA: MIT Press).

Monasson, R. (1992). Properties of neural networks storing spatially correlated patterns. Journal of Physics A: Mathematical and General *25*, 3701–3720.

Montgomery, J.M., Pavlidis, P., and Madison, D.V. (2001). Pair recordings reveal all-silent synaptic connections and the postsynaptic expression of long-term potentiation. Neuron *29*, 691–701.

Nadal, J.-P. (1990). On the storage capacity with sign-constrained synaptic couplings. Network: Computation in Neural Systems *1*, 463–466.

Nadal, J.-P., and Toulouse, G. (1990). Information storage in sparsely-coded memory nets. Network: Computation in Neural Systems *1*, 61–74.

Napper, R.M., and Harvey, R.J. (1988a). Number of parallel fiber synapses on an individual Purkinje cell in the cerebellum of the rat. J. Comp. Neurol. *274*, 168–177.

Napper, R.M., and Harvey, R.J. (1988b). Quantitative study of the Purkinje cell dendritic spines in the rat cerebellum. J. Comp. Neurol. *274*, 158–167.

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.

Petralia, R.S., Esteban, J.A., Wang, Y.X., Partridge, J.G., Zhao, H.M., Wenthold, R.J., and Malinow, R. (1999). Selective acquisition of AMPA receptors over postnatal development suggests a molecular basis for silent synapses. Nat. Neurosci. *2*, 31–36.

Poirazi, P., Brannon, T., and Mel, B.W. (2003). Pyramidal neuron as two-layer neural network. Neuron *37*, 989–999.

Polsky, A., Mel, B.W., and Schiller, J. (2004). Computational subunits in thin dendrites of pyramidal cells. Nat. Neurosci. *7*, 621–627.

Pouille, F., and Scanziani, M. (2001). Enforcement of temporal fidelity in pyramidal cells by feed-forward somatic inhibition. Science *293*, 325–331.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P. (1992). Numerical Recipes in C (Cambridge: Cambridge University Press).

Rosenblatt, F. (1962). Principles of Neurodynamics (New York: Spartan Books).

Rubin, J., Lee, D.D., and Sompolinsky, H. (2001). Equilibrium properties of temporally asymmetric hebbian plasticity. Phys. Rev. Lett. *86*, 364–367.

Rumpel, S., Hatt, H., and Gottmann, K. (1998). Silent synapses in the developing rat visual cortex: evidence for postsynaptic expression of synaptic plasticity. J. Neurosci. *18*, 8863–8874.

Sayer, R.L., Friedlander, M.J., and Redman, S.J. (1990). The time course and amplitude of EPSPs evoked at synapses between pairs of hippocampal CA3/CA1 neurons in the hippocampal slice. J. Neurosci. *10*, 826–836.

Schweighofer, N., Doya, K., and Lay, F. (2000). Unsupervised learning of granule cell sparse codes enhances cerebellar adaptive control. Neuroscience *103*, 35–50.

Sjöström, P.J., Turrigiano, G.G., and Nelson, S.B. (2001). Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. Neuron *32*, 1149–1164.

Song, S., Miller, K.D., and Abbott, L.F. (2000). Competitive hebbian learning through spike-time-dependent synaptic plasticity. Nat. Neurosci. *3*, 919–926.

Stepanyants, A., Hof, P.R., and Chklovskii, D.B. (2002). Geometry and structural plasticity of synaptic connectivity. Neuron *34*, 275–288.

Stuart, G., and Häusser, M. (1994). Initiation and spread of sodium action potentials in cerebellar Purkinje cells. Neuron *13*, 703–712.

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.

Thier, P., Dicke, P.W., Haas, R., and Shabtai, B. (2000). Encoding of movement time by populations of cerebellar Purkinje cells. Nature *405*, 72–76.

van Rossum, M.C.W., Bi, G.-Q., Nelson, S.B., and Turrigiano, G.G. (2000). Stable Hebbian learning from spike timing-dependent plasticity. J. Neurosci. *20*, 8812–8821.

Wehr, M., and Zador, A.M. (2003). Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. Nature *426*, 442–446.