ELSEVIER

Brain Research Bulletin 70 (2006) 49–54

BRAIN RESEARCH BULLETIN

www.elsevier.com/locate/brainresbull

# Information coding capacity of cerebellar parallel fibers

Chi-Ming Huang [a,*], Jennifer A. Titus Pirtle [a], Yu-Ping Wang [c], Rosa H. Huang [b]

[a] *School of Biological Sciences, University of Missouri-Kansas City, Kansas City, MO 64110, USA*
[b] *School of Medicine, University of Missouri-Kansas City, Kansas City, MO 64108, USA*
[c] *School of Interdisciplinary Computing and Engineering, University of Missouri-Kansas City, Kansas City, MO 64110, USA*

Received 20 June 2005; received in revised form 11 October 2005; accepted 17 January 2006
Available online 10 February 2006

**Abstract**

Understanding synaptic connectivity is a prerequisite to gaining insight on how the central nervous system processes information. Cerebellar parallel fibers make an impressive number of synapses with the Purkinje cells. These synapses are the major structural elements of a large information processing system. The objective of the present report is to describe a method to estimate the coding capacity of this information processing system. We propose to derive the coding capacity from the linear distribution pattern of synaptic varicosities along parallel fibers in a manner consistent with Shannon's information theory formalism. The coding capacity of an average parallel fiber synapse is $S = -\kappa \Sigma P_{l(i)} \ln P_{l(i)}$, where $\kappa = 1/\ln 2$, $P_{l(i)}$ is the probability of observing a particular inter-varicosital distance $l(i)$, and ln is the natural logarithm to the base $e$. In the cerebellar parallel fibers of the mouse, and in a number of other unmyelinated axonal systems, the distribution pattern of $P_{l(i)}$ as a function of $l(i)$ is exponential-like. According to information theory, the exponential-like distribution pattern suggests that information transmission in these axonal synaptic systems is operating at near-optimal coding capacity. This optimization in information coding may be the result of a stochastic-like process regulating the formation or elimination of parallel fiber synapses during development and maturation. In the adult nervous system, neuroplasticity-mediated synaptic remodeling may also regulate the coding capacity of axonal synapses via a similar stochastic-like process. The conceptual framework herein may be applicable to other axonal systems in the nervous system.


*Keywords:* Cerebellum; Entropy; Granule cells; Purkinje cells; Synapse

## 1. Introduction

A major function of the central nervous system is to process information. Understanding synaptic connectivity is a prerequisite to gaining insight on how the central nervous system processes information. In the present report, we have used the cerebellar parallel fibers as our model system. Within the cerebellar cortex, the input stage involves the granule cells. The main synaptic targets of granule cells are the Purkinje cells, which are the sole output neurons of the cerebellar cortex. Each granule cell issues a single ascending axon toward the molecular layer. There, the axon bifurcates into a parallel fiber running at a right angle through the dendritic planes of several hundred Purkinje cells. Distal to the bifurcation point, each parallel fiber is therefore a non-branching, one-dimensional, serial device. About 95% of the parallel fiber synapses innervate the Purkinje cells while the remaining 5% contact inhibitory interneurons [16]. As a first-order approximation, the function of a parallel fiber is to synaptically activate hundreds of Purkinje cells in a serial manner. As an action potential travels along a parallel fiber, the selective activation of Purkinje cells as well as the timing of their activation is determined by the number and the spacing of these synapses (pfPc synapses) along the parallel fiber. Much of the information required for Purkinje cell activation, therefore, is contained in the distances between adjacent pfPc synapses.

A near one-to-one correlation exists between pfPc synapses, which are visible at the magnification level of electron microscopy, and the pfPc synaptic varicosities, which are visible with the more convenient light microscopy. Many investigators have therefore drawn inferences about pfPc synapses from observations on the pfPc synaptic varicosities [12,16,17,24]. In particular, Palay and Palay [16] as well as Shepherd et al. [24] have examined the linear distribution pattern of inter-varicosital distances of pfPc synaptic varicosities. They have reported that the probability of finding an inter-varicosital distance of $l(i)$ is inversely and exponentially proportional to $l(i)$. This exponen-

* Corresponding author. Tel.: +1 816 235 2582; fax: +1 816 235 5595.
*E-mail address:* huangc@umkc.edu (C.-M. Huang).

0361-9230/$ – see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.brainresbull.2006.01.007

tial distribution pattern of axonal synaptic varicosities has been observed in many other types of unmyelinated axonal systems besides the cerebellar parallel fibers [4,23,24]. The functional significance for this near universal exponential pattern is unclear, although Shepherd et al. [24] have suggested that this exponential distribution may hold clues about the mechanisms of synaptogenesis and development as well as about certain neuroplasticity models invoking synaptic neogenesis. Moreover, the relationship between the distribution pattern of pfPc synaptic varicosities and information coding capacity of the pfPc synaptic network has not yet been examined. Here we propose to derive the information coding capacity from the linear distribution pattern of pfPc synaptic varicosities along the parallel fibers according to Shannon's information theory formalism [2,21]. We also propose that the exponential-like distribution pattern of axonal synaptic varicosities is an indication that the transmission of information is being carried out in the most efficient manner.

## 2. Materials and methods

Eight mice (C57BL/6J, 3-month-old) were sacrificed and perfused intracardially with 4% formalin and 1% glutaraldehyde. The protocol was executed according to the guidelines of the National Institutes of Health, and with the approval of the University of Missouri Institutional Animal Care and Use Committee. The cerebella remained in the partially opened skull and post-fixed for 24–48 h before being removed. Each cerebellum was placed in 3% $K_2Cr_2O_7$ and 1% $OsO_4$ in $DH_2O$ for 1 day, then washed and impregnated with 1% $AgNO_3$ for 1–2 days following the rapid Golgi procedure described in Scheibel and Scheibel [19]. Coronal sections (75 μm thick) were cut. This method has been used by Palay and Palay [16] as well as Hellwig et al. [4] in their studies on axonal synaptic varicosities.

In optimally stained sections, the rapid Golgi method revealed only a small fraction of randomly selected parallel fibers darkly stained against a pale background. Images of individual parallel fibers and their synaptic varicosities were traced on a light microscope fitted with a *camera lucida*. Parallel fiber synaptic varicosities were identified according to criteria of Palay and Palay [16] and Pitchipornchai et al. [17]. We then measured inter-varicosital distances along the parallel fibers and constructed frequency histograms. In the present study, we have focused on parallel fibers in the cerebellar hemispheres.

## 3. Results

### 3.1. General morphological features of parallel fiber varicosities

Parallel fibers and their synaptic varicosities generally assume a beads-on-a-string appearance (Fig. 1). There does not appear to be any obvious pattern of regularity in the placement of these axonal varicosities. We have sought the optimal sample size for the efficient acquisition of inter-varicosital distance data by plotting the mean and the standard deviation of inter-varicosital distances as we collect more inter-varicosital distances along parallel fibers within a small region of the cerebellar hemispheres in one mouse (Fig. 2). Other than the first data point ($n = 10$), both the mean inter-varicosital distance and the standard deviation are not sensitive to the sample size. Little advantage could be gained by adopting a sample size beyond 100. This is consistent with West [26] who has also recommended $n = 100$–$200$ as a sample size in stereology counting in order to reduce experimental error to within acceptable margins

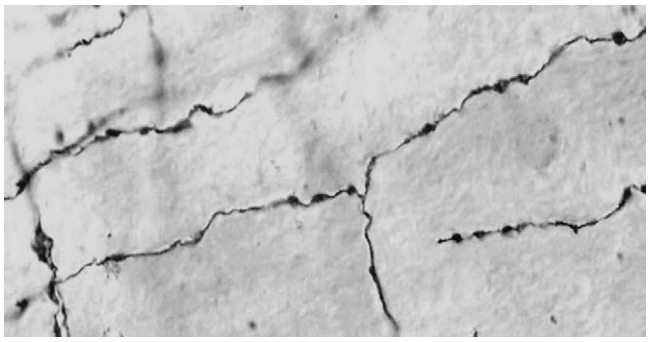

Fig. 1. A photomicrograph of a parallel fiber with synaptic varicosities. Parallel fibers and their synaptic varicosities generally assume a beads-on-a-string appearance. The placement of the synaptic varicosities (and thus the inter-varicosital distances) appears to be random.

(∼15%). For constructing inter-varicosital distance histograms in the present study, we have adopted a sample size of $n = 500$ just to be on the safe side. In addition, values of all standard deviations in Fig. 2 are large—nearly equal to the values of the mean. This is a mathematical property of the exponential distribution pattern [25] and is probably related to the information-coding role of these inter-varicosital distances (see text after Eq. (2)).

### 3.2. Information coding capacity of parallel fiber synapses

When an action potential propagates along a parallel fiber, pfPc synapses along the parallel fiber are activated sequentially, delayed only by the time it takes for the action potential to travel between adjacent synapses. Therefore, each parallel fiber together with its synapses is analogous to a telegraphic code. The code contains a series of symbols, the symbols themselves being the distances between adjacent synapses. For messages made of symbols, Shannon [21] has defined a quantity $S$, the amount of information in bits coded by an average symbol, as

$$S = -\kappa \sum P_{l(i)} \ln P_{l(i)} \quad (1)$$

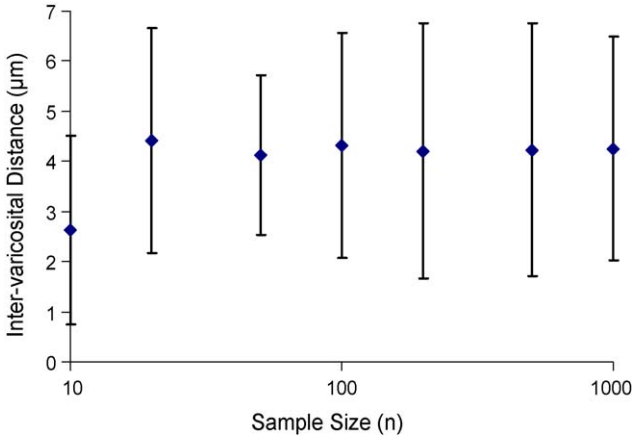


Fig. 2. The relationship between sample size and the mean as well as the standard deviation of inter-varicosital distances. Both the mean inter-varicosital distance and the standard deviation are not sensitive to the sample size.

Table 1
Mean and standard error of $P(i)$ as a function of $l(i)$

| $l(i)$ (μm) | Mean of $P(i)$ | Standard error of $P(i)$ | Stand error/mean |
|---|---|---|---|
| 1–2.5 | 0.116 | 0.0124 | 0.108 |
| 2.5–4 | 0.414 | 0.0413 | 0.100 |
| 4–5.5 | 0.261 | 0.0429 | 0.164 |
| 5.5–7 | 0.137 | 0.0259 | 0.188 |
| 7–8.5 | 0.0458 | 0.0151 | 0.330 |
| 8.5–10 | 0.0199 | 0.00511 | 0.257 |
| 10–12.5 | 0.00598 | 0.00214 | 0.358 |

Values of the mean $P(i)$ in column 2 are taken from Fig. 3A. $P(i)$ decreases faster than the standard error (column 3, derived from five histograms), hence causing the standard error to "get worse" with increasing values of $l(i)$.

where $\kappa = 1/\ln 2$, $P_{l(i)}$ is the probability of observing a particular symbol or inter-varicosital distance $l(i)$, and ln is the natural logarithm.

From images of the type shown in Fig. 1, we can tabulate $l(i)$ and derive $P_{l(i)}$ from a histogram of $l(i)$ Fig. 3A, $P(i)$ is $P_{l(i)}$. The histogram is constructed from a bin-to-bin average of five individual histograms. To avoid cluttering Fig. 3, we have listed the standard error of the mean height of each bin as well as the ratio of the standard error and the mean in Table 1. First, these standard errors reflect the variations in bin height of individual histograms and are not to be identified with the standard deviations in Fig. 2. The magnitudes of the standard errors are ~15% for the first four bins which account for ~93% of all of the data on inter-varicosital distances (Table 1). Unlike the standard deviations in Fig. 2, the magnitude of the standard error decreases in larger samples. Second, the [standard error/mean] ratio increases predictably with $l(i)$ as the sample size becomes smaller for bins at larger values of $l(i)$.

The data in Fig. 3A serves as the basis to compute $S$, the average information coding capacity in bits per symbol (per inter-varicosital distance or per pfPc synaptic varicosity since the number of synaptic varicosities is the same as the number of inter-varicosital distances for all practical purposes). We first define a quantity $s(i)$ as the coding capacity, or the amount of information coded by a single symbol of inter-varicosital distance $l(i)$ with a probability of $P_{l(i)}$. Thus, $s(i)$ is the amount of information conveyed every time this symbol is used.

$$s(i) = -\kappa \ln P_{l(i)} \quad (2)$$

The value of $s(i)$ increases as $l(i)$ increases (Fig. 3B). This is because of the negative sign in Eq. (2), and also because $P_{l(i)}$ decreases exponentially with $l(i)$ (Fig. 3A). The coding capacity of a symbol (e.g. inter-varicosital distance) is inversely related to the predictability of the symbol—the more we can predict what the next symbol is, the less uncertainty, the less information. A symbol or an inter-varicosital distance that occurs at a probability of 1% conveys more information [$-\kappa \ln P_{l(i)} = 3.4$ bits] than a symbol that occurs at a probability of 50% [$-\kappa \ln P_{l(i)} = 1$ bit].

$P_{l(i)}$ becomes smaller with larger values of $l(i)$ (Fig. 3A). Therefore, most of the information transmitted by a parallel fiber is coded by synapses with shorter inter-varicosital distances. However, because [$-\kappa \ln P_{l(i)}$] or $s(i)$ increases sharply with $l(i)$ (Fig. 3B), a small fraction of synapses with long inter-

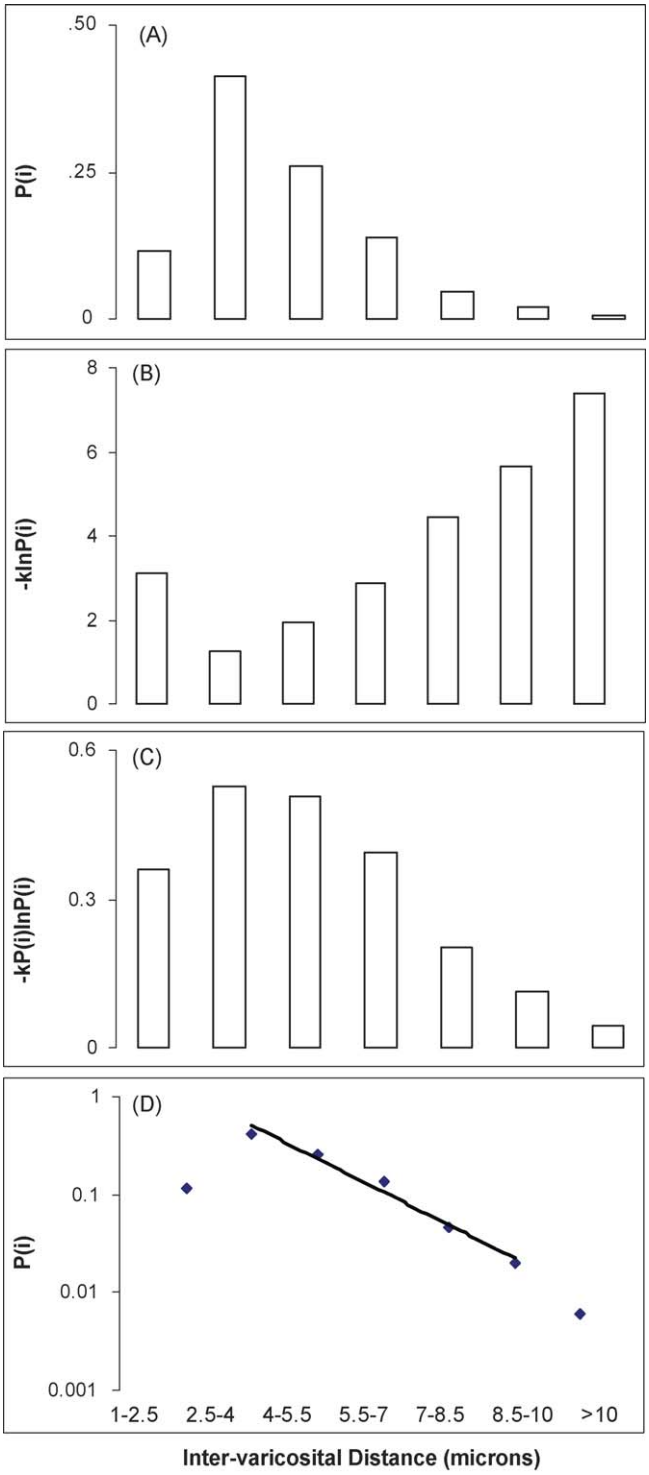


Fig. 3. (A) Histogram of inter-varicosital distance from parallel fibers in the mouse cerebellar hemisphere. The histogram is constructed from a bin-to-bin average of five individual histograms, each constructed from a total of 500 inter-varicosital distances. $Y$-axis is $P_{l(i)}$, the probability of occurrence of inter-varicosital distance $l(i)$, $X$-axis is $l(i)$. (B) $Y$-axis is $s(i)$ in bits or $[-\kappa \ln P_{l(i)}]$ where $\kappa = 1/\ln 2$, $X$-axis is $l(i)$; this shows that the amount of information coded per synapse increases at longer inter-varicosital distances. (C) $Y$-axis is $S(i)$ in bits or $[-\kappa P_{l(i)} \ln P_{l(i)}]$, $X$-axis is $l(i)$; this shows that, in a parallel fiber, most of the information is coded by synapses with short inter-varicosital distances. (D) $Y$-axis is $P_{l(i)}$, the same data as in (A) but plotted in logarithmic scale, $X$-axis is $l(i)$; this shows that $P_{l(i)}$ follows an exponential-like distribution pattern fitting a straight line.

varicosital distances becomes responsible for coding more than their fair share of the information. For example, 21% of the pfPc inter-varicosital distances are longer than 5.5 μm (Fig. 3A). These 21% of the inter-varicosital distances code about 35% of the information (Fig. 3C). This leaves the remaining 65% of the information coded by 79% of the pfPc synaptic varicosities (those with inter-varicosital distances shorter than 5.5 μm).

We next define a quantity $S(i)$ as

$$S(i) = P_{l(i)} \times s(i) \quad (3)$$

where $S(i)$ denotes the contribution of symbol $l(i)$ to the information coded by an average symbol. Clearly, if a certain symbol $l(i)$ is used more often, that symbol will figure more prominently in computing the coding capacity of an average symbol. In Eq. (3), $S(i)$ is therefore weighed by $P_{l(i)}$, the probability of symbol $l(i)$. In Fig. 3C, $S(i)$ is plotted as a function of $l(i)$ as a well-behaved converging series. This convergence is due to the smaller weighing factors $P_{l(i)}$ for larger values of $l(i)$ (Fig. 3A).

The amount of information coded by a synaptic varicosity averaged over all $l(i)$ can now be calculated by summing all values of $S(i)$ from Eq. (3), or

$$S = \sum S(i), \text{ or} \quad (4)$$

$$S = 2.15$$

The unit of $S$ is bits of information per synaptic varicosity (or per inter-varicosital distance). Eqs. (1)–(4) can be used to estimate the coding capacity of other axonal systems from the distribution pattern of inter-varicosital distances.

The bin width in Fig. 3 is 1.5 μm. Had we chosen a bin width half as wide (0.75 μm), the calculated $S$ would be larger since extra information is provided by splitting one symbol into two. The value of $S$ is therefore dependent upon the bin width of the inter-varicosital distance histogram. The correct choice of the bin width, however, cannot be determined from Fig. 3 and must be derived from insights into the biological significance of the inter-varicosital distances. The key issue is the number of distinct and biologically significant classes of inter-varicosital distances. The answer is unlikely to be one or two, as the coding capacity will be too small. In the case of *one*, the coding capacity is zero! It is also unlikely to be three or four. This is because the distribution pattern of inter-varicosital distance is exponential (Fig. 3D). In an exponential pattern with only three or four bins, most of the information will be coded in the first one or two bins as the number of inter-varicosital distances in the other two to three bins will be small. At a minimum, the axonal system must be able to distinguish five to six different classes of inter-varicosital distances. In Fig. 3, most of the inter-varicosital distances are less than 10 μm. We therefore picked a bin width of 1.5 μm, which divides 10 μm into almost seven bins.

In addition, we can define **S** as the information coding capacity for a parallel fiber. This can be calculated from,

$$\mathbf{S} = S \times N \quad (5)$$

where $N$ is the number of synaptic varicosities (or more precisely, the total number of inter-varicosital distances) in the entire parallel fiber. Since there are $\sim 10^3$ varicosities per parallel fiber [9], **S** is $\sim 2 \times 10^3$ bits per parallel fiber.

We can also define $\mathbf{S}_{\text{cerebellum}}$ as the information coding capacity for a given region or even the entire cerebellum. This can be calculated from,

$$\mathbf{S}_{\text{cerebellum}} = \mathbf{S} \times \mathbf{N} \quad (6)$$

where **N** is the number of parallel fibers in that region or the whole cerebellum. Since there are $\sim 10^{11}$ granule cells in the human cerebellum [9], $\mathbf{S}_{\text{cerebellum}}$ should be of the order of $2 \times 10^{14}$ bits. Note that this is based on the bin width of six to seven distinct classes of inter-varicosital distance, which, as stated above, should be a low estimate. The estimate of $2 \times 10^{14}$ bits is therefore also likely to be low. For the purpose of comparison, an average English word carries just a little over 10 bits of information (11.7 to be exact [22]). The information coding capacity of the cerebellar parallel fiber system is therefore equivalent to the information coded by about $2 \times 10^{13}$ English words (e.g. 20 million books, each with about 1 million words).

### 3.3. The exponential distribution pattern

Although the placement of synaptic varicosities along parallel fibers appears to be random (Fig. 1), the frequency histogram of inter-varicosital distance is suggestive of an exponential tail (Fig. 3A). We have re-plotted the data in Fig. 3A in a semi-logarithmic graph. A straight line can be fitted through the data points (Fig. 3D), indicating an exponential distribution pattern [4,24].

## 4. Discussion

### 4.1. The relationship between pfPc synaptic varicosities and pfPc synapses

Palay and Palay [16] have characterized parallel fiber synaptic varicosities in the rat with light microscopy (LM) and electron microscopy (EM). They have shown that the density of varicosities along parallel fibers is statistically indistinguishable from the density of synapses (although some varicosities have multiple pfPc synapses, some have none). Recent EM data have confirmed that the majority (~90%) of the parallel fiber varicosities contain a single synapse [3,17,27]. Therefore, the correlation between parallel fiber synapses (must be studied with EM) and synaptic varicosities (can be studied with LM) is good. An interesting issue is the use of the number of synapses as a structural marker for the functional capacity of a neural network. The information coding capacity of an axon, however, is a non-linear function of the number of synapses [Eqs. (1) and (2)]. Doubling the number of synapses along an axon will increase but fall short of doubling the coding capacity of the axon.

### 4.2. Coding capacity in other axonal systems

We have estimated the coding capacity of axons from four other studies [3,4,12,16] (Table 2). Values of the coding capac-

Table 2
Coding capacity of various unmyelinated axonal systems

| Synaptic density ($\mu m^{-1}$) | Coding capacity $S$ per synapse (bits) | Coding capacity (bits/$\mu$m) | Source |
|---|---|---|---|
| 0.171 | 2.81 | 0.481 | Rat parallel fibers at 3 month [4] |
| 0.138 | 3.38 | 0.466 | Rat parallel fibers at 9 month [4] |
| 0.090 | 4.01 | 0.361 | Rat parallel fibers at 23 month [4] |
| 0.235 | 2.15 | 0.505 | Mouse parallel fibers at 3 month |
| 0.278–0.4 | 1.70 | 0.473–0.68 | Rat parallel fibers [16] |
| 0.222 | 2.70 | 0.599 | Rat cortical axons [5] |
| 0.0952 | 4.14 | 0.394 | Rat corticostrial axons [12] |

Values of synaptic density (column 1) are taken from the relevant references cited. Values for the coding capacity per synapse (column 2) are first calculated [using Eqs. (1)–(4)] directly from the distribution histograms of inter-varicosital distances available from the references. A correction is then made such that all coding capacities are based on an 1.5 $\mu$m bin width for the histograms of inter-varicosital distances. For example, if the cited reference uses a 1 $\mu$m bin width, then the correction is done by adding results from Eqs. (1) to (4) to a factor of $[\ln(1/1.5)/\ln(2)] = -0.585$ bits [5,12]. If the cited reference uses a 4 $\mu$m bin width, then the correction is done by adding results from Eqs. (1) to (4) to a factor of $[\ln(4/1.5)/\ln(2)] = 1.42$ bits. Values for the coding capacity per $\mu$m of axon (column 3) are calculated as a product of columns 1 and 2.

ity per synapse are between 1.7 and 4.1 bits. In rat parallel fibers, age-related changes can be significant between 3 and 23 months. As the number of synapses decreases and the mean inter-varicosital distance increases, the coding capacity per synapse actually increases. As a result, the coding capacity per micron of parallel fiber exhibits little or no changes between 3 and 23 months in the rat cerebellum. But, if there is a significant shortening of parallel fibers with age, then the coding capacity per parallel fiber will decrease. This happens to be the case as the mean length of the parallel fiber at 23 months is approximately one-third of the value at 3 months [5].

### 4.3. Coding capacity and information processing

The information coding capacity **S** is a structural parameter determined by the spatial distribution pattern of the pfPc synaptic varicosities. For example, consider the anterior lobe which receives its primary afferents from muscle spindles in the weight-bearing lower extremities. From Eqs. (1) to (6), we can calculate the coding capacity, $\mathbf{S}_{\text{anterior lobe}}$ ($\mathbf{S}_{\text{ant}}$). In this way, $\mathbf{S}_{\text{ant}}$ has not been calculated based on the amount of information conveyed by the muscle spindles, which we can call $\mathbf{S}_{\text{spinocerebellar tract}}$ ($\mathbf{S}_{\text{sct}}$) [6,20]. Rather, $\mathbf{S}_{\text{ant}}$ concerns a capacity for directing the selective activation of Purkinje cells upon receiving the afferent input information, $\mathbf{S}_{\text{sct}}$. A superior pfPc synaptic system in the anterior lobe (with a large $\mathbf{S}_{\text{ant}}$) coupled with a superior proprioceptive afferent system (with a large $\mathbf{S}_{\text{sct}}$) will clearly produce excellent anti-gravity defenses. When $\mathbf{S}_{\text{ant}} \ll \mathbf{S}_{\text{sct}}$, insufficient coding capacity leaves much of the afferent information from the muscle spindles wasted. When $\mathbf{S}_{\text{ant}} \gg \mathbf{S}_{\text{sct}}$, over-design of the pfPc synaptic circuitry may have its own shortcomings. Ultimately, $\mathbf{S}_{\text{ant}}$ together with $\mathbf{S}_{\text{sct}}$, determines how the cerebellum responds to the muscle spindle input. Mechanisms must exist to ensure that $\mathbf{S}_{\text{ant}}$ and $\mathbf{S}_{\text{sct}}$ are suitably and continuously matched. Such matching may involve the cerebellar neuroplasticity mechanisms proposed by Marr [14] and recently reviewed by Ito [10,11].

Interestingly, Isope and Barbour [8] have recently reported that a significant portion of the parallel fiber synapses were silent. More recently, Mittmann et al. [15] as well as Marcaggi and Attwell [13] have suggested that the parallel fibers may also play a modulating role in reference to the excitatory actions provided by the ascending axons of the granule cells to Purkinje cells. There may be significant and dynamic changes in the function of individual pfPc synapses [8,10,11,14] as a function of experience or aging. The relationship between this modulatory function, neuroplasticity, and information processing is, however, unclear at the present time [1]. These findings can complicate the computation of coding capacity. They, however, should not fundamentally invalidate the concept of coding capacity in cerebellar parallel fibers. In particular, the coding capacity, like $\mathbf{S}_{\text{ant}}$, can be defined as an anatomic measure, whereas the actual information, like $\mathbf{S}_{\text{sct}}$, must be derived from physiology [6,20].

### 4.4. Exponential distribution, efficient coding, and entropy

Sending a coded message (e.g. a telegraph) is constrained both by the number of available symbols (e.g. $N$ as the 26 letters in the English alphabet) and the total message length (e.g. $L$ as the total number of letters in a telegraph). The boundary conditions are:

$$N = \sum n(i) \tag{7}$$

$$L = \sum n(i)l(i) \tag{8}$$

where $n(i)$ is the number of symbols of the length $l(i)$. The task of searching for the most efficient coding recipe can be reduced to searching mathematically for the best way of partitioning the total message length $L$ into $N$ segments such that the average message contains the most information while observing the boundary conditions [Eqs. (7) and (8)]. Brillouin [2] has shown mathematically that the most efficient coding is attained when the probability of $l(i)$ [or $P_{l(i)}$] assumes an exponential relationship with $l(i)$. Interestingly, this is also seen in experimental data from the inter-varicosital distances in parallel fibers [16,24], axons of cortical pyramidal cells [4], and unmyelinated axons of the hippocampus and others [23,24]. Therefore, optimization of information transmission may already be wide-spread.

Elsewhere in nature, exponential-like distribution pattern also exists (e.g. the decay of radioactive particles). Partitioning a message of length $L$ into $N$ segments is analogous to partitioning

total energy $E$ among $N$ ideal gas molecules in thermodynamics, where the boundary conditions are,

$$N = \sum n(i) \tag{9}$$

$$E = \sum n(i)e(i) \tag{10}$$

These boundary conditions are similar in form to Eqs. (7) and (8). In thermodynamics, the resultant Boltzmann distribution (exponential) leads to a maximization of entropy. In coding information, an exponential distribution leads to a maximization of information. Indeed, information can be measured in entropy units and vice versa [2]. We have reported the similarities in the formalism of information theory and thermodynamics including the link between information and entropy [7,18]. Perhaps biological information processing in axonal systems can be understood with synapses as basic structural elements on the one hand and coding capacity or entropy as a measure of function on the other.

## Acknowledgements

This work was supported by NIH grant AA13322 and the National Alliance for Autism Research (NAAR). We thank Amy Roychowdhury for the preparation of figures.

## References

[1] R. Apps, M. Garwicz, Anatomical and physiological foundations of cerebellar information processing, Nature Rev. Neurosci. 6 (2005) 297–311.
[2] L. Brillouin, Science and Information Theory, second ed., Academic Press, New York, 1962.
[3] K.D. Federmeier, J.A. Kleim, W.T. Greenough, Learning-induced multiple synapse formation in rat cerebellar cortex, Neurosci. Lett. 332 (2002) 180–184.
[4] B. Hellwig, A. Schuz, A. Aertsen, Synapses on axon collaterals of pyramidal cells are spaced at random intervals: a Golgi study in the mouse cerebral cortex, Biol. Cybern. 71 (1994) 1–12.
[5] C. Huang, N. Brown, R. Huang, Age-related changes in the cerebellum: parallel fibers, Brain Res. 840 (1999) 148–152.
[6] C. Huang, R. Burkard, Frequency sensitivities of auditory neurons in the cerebellum of the cat, Brain Res. 371 (1986) 101–108.
[7] C. Huang, H. Reiss, Transmission time and bandwidth in pulse code modulation: application of the statistical thermodynamic formalism, J. Stat. Phys. 3 (1971) 261–289.
[8] P. Isope, B. Barbour, Properties of unitary granule cell-Purkinje cell synapses in adult rat cerebellar slices, J. Neurosci. 22 (2002) 9668–9678.
[9] M. Ito, The Cerebellum and Neural Control, Raven Press, New York, 1984.
[10] M. Ito, Cerebellar long-term depression: characterization, signal transduction, and functional roles, Physiol. Rev. 81 (2001) 1143–1195.
[11] M. Ito, The molecular organization of cerebellar long-term depression, Nat. Rev. Neurosci. 3 (2002) 896–902.
[12] A.E. Kincaid, T. Zheng, C.J. Wilson, Connectivity and convergence of single corticostriatal axons, J. Neurosci. 18 (1998) 4722–4731.
[13] P. Marcaggi, D. Attwell, Endocannabinoid signaling depends on the spatial pattern of synapse activation, Nat. Neurosci. 8 (2005) 776–781.
[14] D. Marr, A theory of cerebellar cortex, J. Physiol. 202 (1969) 437–470.
[15] W. Mittmann, U. Koch, M. Hausser, Feed-forward inhibition shapes the spike output of cerebellar Purkinje cells, J. Physiol. 563 (2005) 369–378.
[16] S.L. Palay, V. Palay, The Cerebellar Cortex: Cytology and Organization, Springer-Verlag, New York, 1974.
[17] C. Pitchipornchai, J.A. Rawson, S. Rees, Morphology of parallel fibers in the cerebellar cortex of the rat: an experimental light and electron microscopic study with biocytin, J. Comp. Neurol. 342 (1994) 206–220.
[18] H. Reiss, C. Huang, Statistical thermodynamic formalism in the solution of information theory problems, J. Stat. Phys. 3 (1971) 191–216.
[19] M.E. Scheibel, A.B. Scheibel, The methods of Golgi, in: R.T. Robertson (Ed.), Neuroanatomical Research Techniques, Academic Press, New York, 1978, pp. 89–114.
[20] G.M. Shambes, J.M. Gibson, W. Welker, Fractured somatotopy in granule cell tactile areas of rat cerebellar hemispheres revealed by micromapping, Brain Behav. Evol. 15 (1978) 94–140.
[21] C.E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (1948) 379–423.
[22] C.E. Shannon, Prediction and entropy of printed English, Bell Syst. Tech. J. 30 (1951) 50–64.
[23] G.M. Shepherd, K.M. Harris, Three-dimensional structure and composition of CA3-CA1 axons in rat hippocampal slices: implications for presynaptic connectivity and compartmentalization, J. Neurosci. 18 (1998) 8300–8310.
[24] G.M.G. Shepherd, M. Raastad, P. Andersen, General and variable features of varicosity spacing along unmyelinated axons in the hippocampus and cerebellum, Proc. Natl. Acad. Sci. USA 99 (2002) 6340–6345.
[25] S. Wassertheil-Smoller, Biostatistics and Epidemiology, second ed., Springer-Verlag, New York, 1995.
[26] M.J. West, Stereological methods for estimating the total number of neurons and synapses: issues of precision and bias, Trends Neurosci. 22 (1999) 51–61.
[27] M.A. Xu-Friedman, K.M. Harris, W.G. Regehr, Three-dimensional comparison of ultrastructural characteristics at depressing and facilitating synapses onto cerebellar Purkinje cells, J. Neurosci. 21 (2001) 6666–6680.