Stony Brook University

An Optimization Theory of Brain Structures

A Dissertation Presented

by

Quan Wen

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Physics

Stony Brook University

May 2007

Stony Brook University

The Graduate School

Quan Wen

We, the dissertation committee for the above candidate for the Doctor of Philosophy degree, hereby recommend acceptance of this dissertation.

Dmitri B. Chklovskii
Cold Spring Harbor Laboratory
Dissertation Director

Konstantin K. Likharev
Professor, Department of Physics, Stony Brook University
Chairman of Dissertation

Chris Jacobsen
Professor, Department of Physics, Stony Brook University

Lorne Mendell
Professor, Department of Neurobiology and Behavior, Stony Brook University
Outside Member

This dissertation is accepted by the Graduate School.

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

An Optimization Theory of Brain Structures

by

Quan Wen

Doctor of Philosophy

in

Physics

Stony Brook University

2007

Over hundreds of millions of years, the nervous systems have evolved to maximize their functionality and reduce the cost associated with building and maintenance. In this thesis, we use optimization theory to explain various aspects of brain structures. Starting from the cost-benefit analysis of axons, we first explain why some long-range axons in the brain are myelinated and some others are not. Second, we develop a theory explaining why long-range axons and local arbors are segregated spatially, which gives rise to two distinguished tissues – the white and gray matter. Third, we quantify and minimize the cost in the dendrite and

explain several salient features of Purkinje dendrites in the cerebellum. Finally, to quantify the functional roles of dendrites, we posit that the dendritic arbors not only minimize their cost, but also maximize the available connectivity patterns. We develop a statistical theory of neuronal shape and predict several scaling relationships that can be directly tested from the anatomical data.

To my family

Contents

List of Figures

Acknowledgements

I am grateful to my thesis advisor, Dmitri Chklovskii. If I had not learned the trade from him, I would have never been able to learn it. During these years, he has offered so many invaluable things: from big pictures to small details, from inspirations to precisions. He persuaded me to believe that theory will play an important role in neurobiology, but more importantly he taught me how to build such a theory.

I thank members and former members in out lab. In particular, during the last couple of months, I have benefited a lot from discussing with Armen Stypanyants, now an assistant professor in Northeastern university. Interactions with Yuri Mishchenko, Sen Song and Beth Chen have also influenced this work in many different ways.

I also thank Guy Elston at the university of Queenland in Australia, who kindly provided his tremendous dendritic arbor dataset.

I dedicate this thesis to my wife, Junchao. We have gone through all the happiness and bitterness together in the past five years. As the pivot of my life, she continually bring laugh and optimism, which makes my work much easier. I also dedicate this thesis to my parents and my grandmother. Their visit from China brings new elements and joy to this family. Their love and

encouragement are always the sources of motivation.

Chapter 1

Introduction

Studying the structure of an organ helps unravel the mystery of its function. More than one hundred years ago, by applying the histological staining technique to study the anatomy of nervous systems, Santiago Ramón y Cajal, one of the founders of modern neurobiology, formulated the "neuron doctrine." He proposed that the brain is made up of billions of separate neurons, and these cells are interconnected with each other by two polarized components, axons and dendrites. The function of dendrites and axons is to conduct electrical signals from post-synaptic terminals to the integration site, which often is the cell body, and from the integration site to the pre-synaptic terminals, respectively.

One hundred years later, neurobiologists, armed with advanced tools, are embracing a great opportunity to dissect the complex structure of the neuronal circuitry. For instance, cell-labeling methods such as those based on biocytin or green fluorescent protein (GFP) make it possible to describe neuronal arbors in a comprehensive and quantitative way. Reconstructing the wiring diagram of the brain circuitry may become an attainable goal due to the development

of electron microscopy. However, a quantitative theory of brain anatomy does not exist. Such a theory should formulate the principles of brain design and establish relationships between structures and functions.

Insight in this direction was also contributed by Cajal, as he wrote [1] "After the many shapes assumed by neurons, we are now in a position to ask whether this diversity ... has been left to chance and is insignificant, or whether it is tightly regulated and provides an advantage to the organism. ... we realized that all of the various conformations of the neuron and its various components are simply morphological adaptations governed by laws of conservation for time, space, and material."

Echoing Cajal's thoughts, we propose that evolution has "tinkered" with brain design to maximize its functionality while minimizing the cost associated with building and maintaining it. This hypothesis can be formulated by an optimization approach. Specifically, brain functionality must benefit from high synaptic connectivity and large plasticity, because synaptic connections are central to information processing as well as learning and memory, which is thought to manifest itself in synaptic modifications. However, increasing connectivity requires adding wiring to the network, which comes at a cost. The cost of wiring could be accrued by various things, including conduction delays, signal attenuation, and wiring volume.

We will elaborate upon these points in the following chapters of the thesis and explain various aspects of the brain structures. In summary, we obtained two major results. First, based on the hypothesis that the brain maximizes the synaptic connectivity while minimizing conduction delays in axons and dendrites, we developed an optimization theory explaining why the brain is

segregated into gray and white matter, a ubiquitous feature in the vertebrates. The theory provides a possible explanation for the structure of various brain regions such as cerebral cortex, neostriatum, and spinal cord. Moreover, we derive a formula for the cortical thickness that can be directly tested from the anatomical data. Second, by analyzing a large dataset of pyramidal cells, we observe scaling in the dendritic arbors. Among other things, we find that the dependence of dendritic arbor span on the total dendritic length can be fitted by a power law. Next, we turn to a theoretical explanation based on the assumption that a dendritic arbor maximizes the number of different connectivity patterns while minimizing its own cost. Our theory is able to calculate the critical exponent in the scaling law and explain the variations of arbor shape among different cell types.

Chapter 2

To Myelinate or Not To Myelinate

2.1 Background

In vertebrate brains, axons are often myelinated, *i.e.*, wrapped in a thin myelin sheath. Myelinated and non-myelinated axons co-exist in many structures, such as the corpus callosum. This observation leads to a question: why does nature myelinate some axons but not others? Understanding this question may shed light on the principles of brain design. About half a century ago, Rushton [3] proposed an answer, which can be stated as follows (see also Waxman and Bennett [2]): The crucial difference between a myelinated and a non-myelinated axon is in how the conduction velocity scales with axon diameter, as shown in Figure 2.1 (redrawn from Waxman and Bennett [2]). In a myelinated axon, conduction velocity (CV), v_m, increases linearly with the

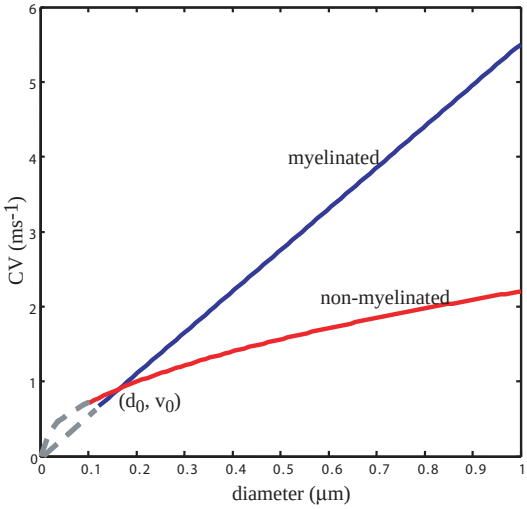

Figure 2.1: Conduction velocity as a function of the axon diameter for myelinated and non-myelinated axons, modified from Waxman and Bennett [2]. Axon diameters much smaller 0.1 micron are not observed experimentally and plots in this regime are represented by dashed lines. At d_0, myelinated and non-myelinated axons have the same CV, v_0.

axon diameter [3, 4] as:

$$v_m = BD, \tag{2.1}$$

where B is the proportionality constant with dimension m/s·μm^{-1}, and D is the axon diameter including the myelin sheath. In a non-myelinated axon, CV, v_n, increases with the square root of axon diameter, d_a[5, 3, 4, 6]:

$$v_n = b\sqrt{d_a}, \tag{2.2}$$

where b is also a proportionality constant with the dimension m/s·μm$^{-1/2}$. Because axon diameters smaller than 0.1 micron have rarely been observed experimentally, Equations 2.1-2.2 are valid only for relatively large axon diameters (solid lines in Figure 2.1).

According to Equations 2.1-2.2, there is a diameter, d_0, at which the CVs, v_0, in a myelinated and a non-myelinated axon are the same (Figure 2.1):

$$v_0 = BD = b\sqrt{d_a} = \frac{b^2}{B}. \tag{2.3}$$

Rushton [3] proposed that, for a given CV, nature picks an axon with a smaller diameter. If the required CV is less than v_0, a non-myelinated axon is thinner and preferred. Otherwise, a myelinated axon is thinner and preferred (Figure 2.1).

It follows from Rushton's argument that there is a critical axon diameter d_0 ($d_0 \approx 0.2$ μm) according to Waxman and Bennett [2]). All axons that are thinner than d_0 should be non-myelinated, and those thicker than d_0 should be myelinated. This prediction does not accord with experimental data [7, 8, 9, 10]. Although myelinated axons are typically thicker than the non-myelinated ones, the diameter distributions of myelinated and non-myelinated axons overlap.

In this chapter, we resolve this old issue by assuming explicit expressions of axonal cost for both myelinated and non-myelinated axons. We point out that minimization of the axonal cost does not necessarily lead to the existence of a critical axon diameter. In particular, we consider two possible optimization scenarios. In the first scenario, we assume that the conduction delay in an axon (*i.e.*, the time taken for an action potential to travel between two end points) is set *a priori*, and that the axon cost is proportional to its volume. Depending on the relative cost per volume for a myelinated and non-myelinated axon, the distributions of axon diameters can overlap or be separated by a gap. In the

second scenario, we assume that the conduction delay is not fixed. Rather, an increase in the conduction delay invokes a finite cost. We then search for the axon design that optimizes the cost function, including volume and conduction delay. Interestingly, the latter scenario predicts that there must be a gap in the conduction velocity distribution of the axons.

2.2 Axonal cost with Fixed Conduction Delay

In this section, we formulate the following two-choice problem: given the conduction delay in an axon, should the axon be myelinated or non-myelinated? We will answer this question for a non-branching axon with fixed length by calculating the axonal cost as a function of CV. Assuming that the cost is proportional to the axonal volume [11, 12, 13], the costs of a myelinated and a non-myelinated axon per length are given by :

$$\epsilon_m = K\frac{\pi}{4}D^2, \tag{2.4}$$

$$\epsilon_n = k\frac{\pi}{4}d^2, \tag{2.5}$$

where K and k are costs per volume for myelinated and non-myelinated axons respectively. By substituting Equations 2.1-2.2, we express the axonal costs in terms of CV:

$$\epsilon_m = K\frac{\pi v_m^2}{4B^2}, \tag{2.6}$$

$$\epsilon_n = k\frac{\pi v_n^4}{4b^4}. \tag{2.7}$$

By setting $\epsilon_m = \epsilon_n$, we determine the critical CV v_c where the costs of a

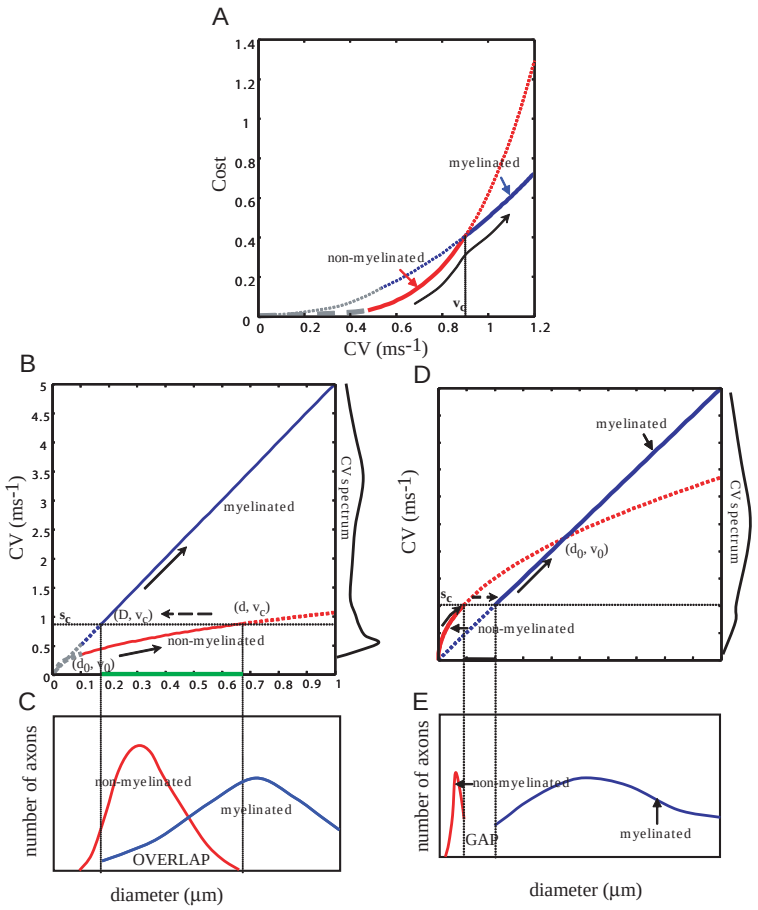

Figure 2.2: The axon diameter spectrum in scenario I. A. If CV is below the critical CV, v_c, a non-myelinated axon is less costly and preferred. Otherwise, a myelinated axon is preferred. B,C. If the myelinated axon is costlier (on a per volume basis) than the non-myelinated axon, the critical CV $v_c > v_0$, there is an overlap in the diameter distributions of myelinated and non-myelinated axons. When the ratio $K/k = 16$, the diameter distributions of myelinated and non-myelinated axons overlap in the regime 0.2 μm-0.7 μm. On the right of B, a continuous CV spectrum is assumed. D,E. If the non-myelinated axon is costlier (on a per volume basis) than the myelinated axon, $v_c < v_0$, there is a gap in the diameter spectrum. On the right of D, a continuous CV spectrum is assumed.

myelinated and a non-myelinated axon are equal (see Figure 2.2A):

$$v_c = \sqrt{\frac{K}{k}} v_0. \tag{2.8}$$

When the required CV is smaller than v_c, a non-myelinated axon should be used; otherwise a myelinated axon should be used (Figure 2.2A). The implications of this result on the spectrum of the axon diameters depend on the ratio of the cost per volume for a myelinated and a non-myelinated axon.

- $K = k$. When costs per volume of a myelinated and a non-myelinated axon are equal, we recover Rushton's result: axons with a diameter greater than the critical diameter d_0 are myelinated; below the critical diameter they are non-myelinated.

- $K > k$. When cost per volume of a myelinated axon is greater than that of a non-myelinated axon, the critical CV v_c is above the intersection point v_0 (Figure 2.2B, 2.2C). In this case we expect an overlap between the diameter distributions of myelinated and non-myelinated axons.

- $K < k$. When cost per volume of a myelinated axon is less than that of a non-myelinated axon, the critical CV v_c is below the intersection point v_0 (Figure 2.2D, 2.2E). In this case, we expect a gap in the diameter spectrum that separates the non-myelinated and myelinated axon distributions.

2.3 Axon with Fixed Conduction Delay Cost

In the previous scenario, the conduction delay is specified *a priori* for an axon. In an alternative scenario, we may assume that the conduction delay only results in a finite penalty and the total cost linearly combines the cost of volume, V_a, and the cost of conduction delay, T_a, weighted by k and α, respectively [14],

$$E_a = kV_a + \alpha T_a, \tag{2.9}$$

where the unit conduction delay cost α is now set *a priori* for a given axon length. Then, we ask a slightly different question from that in the first scenario. If conduction delay cost is proportional to α, should the axon be myelinated or non-myelinated? To answer this question, we first determine the minimal costs for a myelinated axon and a non-myelinated axon, respectively. We then compare the two costs in order to choose the less costly axon.

By rewriting the cost as a function of diameter (see Methods), we show that the volume term in the cost increases with diameter while the conduction delay term decreases with diameter. By setting the derivative of the cost function with respect to diameter to zero, we find that the optimal CV in a myelinated axon is given by

$$v_m = B^{2/3}\left(\frac{2\alpha}{\pi K}\right)^{1/3}. \tag{2.10}$$

The corresponding cost per length is given by

$$\epsilon_m = \frac{3\alpha}{2v_m} \tag{2.11}$$

In a non-myelinated axon, the optimal CV and the minimal cost per length are given by

$$v_n = b^{4/5}\Big(\frac{\alpha}{\pi k}\Big)^{1/5}, \tag{2.12}$$

$$\epsilon_n = \frac{5\alpha}{4v_n}. \tag{2.13}$$

Figure 2.3D shows the myelinated and non-myelinated axonal costs as a function of the unit conduction delay cost α. They are equal at the critical value α_0. If α less that α_0, a non-myelinated axon is preferred. Otherwise, a myelinated axon is preferred.

However, unlike in the previous scenario, minimizing the axonal costs with a finite conduction delay penalty predicts a gap in the CV spectrum (Figure 2.3). This is because the CV of a myelinated axon is different from that of a non-myelinated axon when the costs of the two types are the same (Figure 2.3C). To see this, by setting Equation 2.11 and Equation 2.13 equal to each other, we find that when the costs of a myelinated and non-myelinated axon are the same

$$v_m = \frac{6}{5}v_n. \tag{2.14}$$

The diameter spectrum still depends on the ratio K/k. Similar to the previous scenario, we find that

- $K = \frac{5}{3}k$. Axons with a diameter greater than the critical diameter d_0 are myelinated; below the critical diameter, they are non-myelinated. We note that although Rushton's result [3] is recovered in this case, the gap in the CV spectrum always exists.

- $K > \frac{5}{3}k$. The diameter distributions of non-myelinated and myelinated

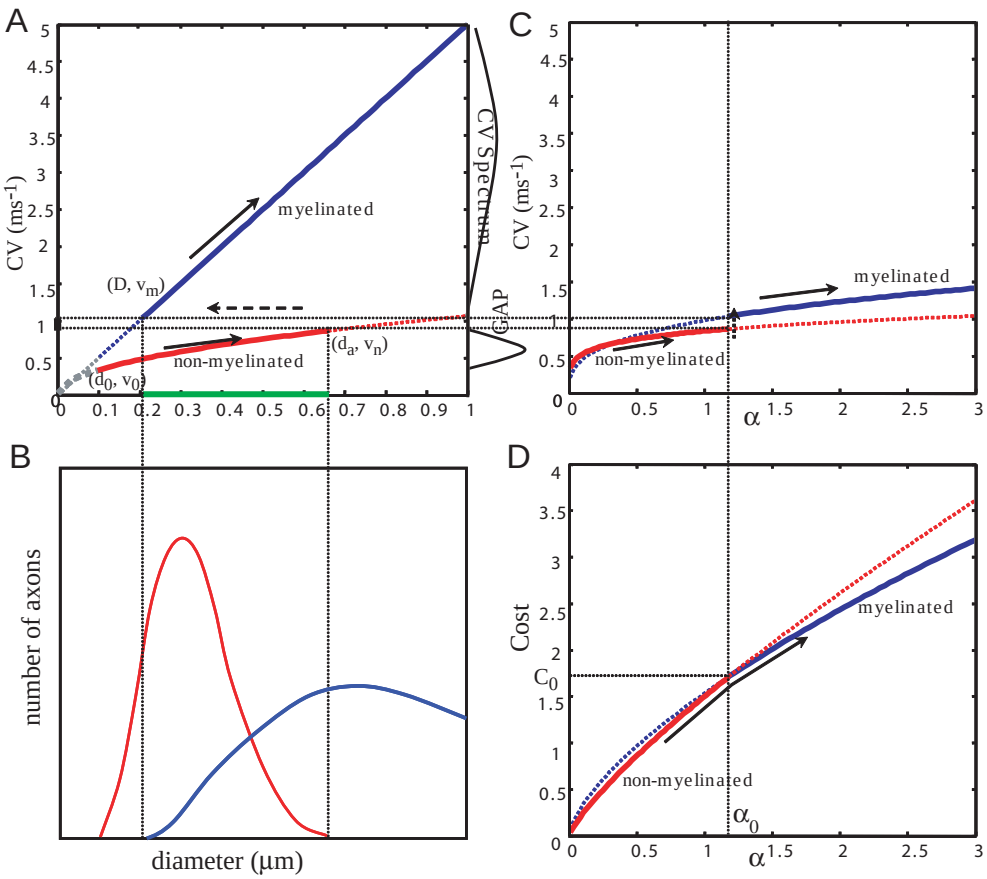

Figure 2.3: The axon diameter and CV spectra in Scenario II, provided that $K \gg k$. A,B. If $K/k = 17$, the diameter distributions of myelinated and non-myelinated axons overlap in the regime 0.2 μm-0.7 μm. On the right of A, the CV spectrum has a gap. C. CV as a function of the unit conduction delay cost α. D. Axonal costs as a function of the unit conduction delay cost α. At the critical α_0, the myelinated and non-myelinated axons have the same cost, but the CVs are different. This induces a gap in the CV spectrum when the cost of a non-myelinated axon becomes greater than that of a myelinated axon.

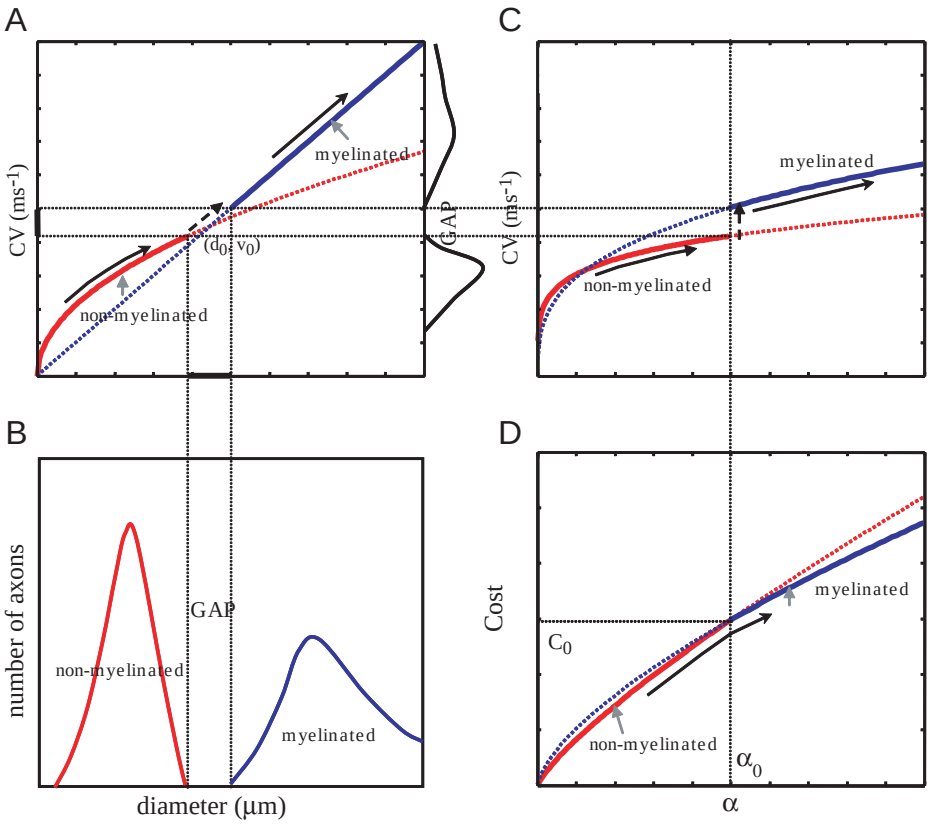

Figure 2.4: The axon diameter and CV spectra in Scenario II, provided that $K \ll k$.

axons overlap (Figure 2.3A, 2.3B).

- $K < \frac{5}{3}k$. The diameter spectrum has a gap that separates non-myelinated and myelinated axon distributions (Figure 2.4A, 2.4B).

2.4 Comparison with Experiments

Next, we compare our theory with the available experimental data from the corpus callosum [15, 7, 16, 10]. This region of the brain offers several advantages. First, myelinated and non-myelinated axons coexist in it. Second, corpus callosum axons have different diameters but similar lengths (between 100 and 130 mm in a human brain [16]). As a result, the relationship between conduction delay and conduction velocity is uniform. Third, myelinated axons, in general, have the disadvantage that they cannot make *en passant* synapses. This disadvantage cannot play a role in the corpus callosum because there are no synapses there. Unfortunately, we are unaware of simultaneous measurements of conduction velocity and diameter for individual axons, making our comparison somewhat indirect. Although measurements have been done in studies of the peripheral nervous system dating back a long time [5, 17, 18], precise measurements of myelinated and non-myelinated axons with very small diameters (0.1 - 1 micron) is technically challenging.

Waxman and Swadlow [7, 15] found that, in the visual corpus callosum, the diameter of non-myelinated axons varies between 0.08 micron and 0.6 micron, while the diameter of myelinated axons changes between 0.3 micron and 1.58 micron. They observed an overlap between the diameter distributions of non-myelinated axons and myelinated axons. The distribution of conduction

velocity in the visual callosal axons (albeit different sample from the diameter measurements) ranges from 0.3 m/s to 12.9 m/s.

If we assume a square-root relationship between conduction velocity and the diameter of a non-myelinated axon, the smallest conduction velocity 0.3 m/s corresponds to the smallest diameter 0.08 micron. Then, the scaling factor b in Equation 2.2 can be estimated as 1.06 m/s$\cdot\mu$m$^{-1/2}$. This result is consistent with recent measurements on olfactory receptor neurons [19]. In myelinated axons [2, 18], the scaling factor B in Equation 2.1 is between 4.5 and 6 m/s$\cdot\mu$m^{-1}. These scaling factors were used in Figures 2.2-2.4.

These figures show that a gap in the diameter spectrum can exist only below 0.07 micron, a value too small for non-myelinated axons to exist. Indeed, experimentally measured distributions of axon diameters exhibit an overlap between the myelinated and the non-myelinated axons [10, 7], suggesting that $K > k$. Using estimates of scaling factors, $k = 1.06$ m/s$\cdot\mu$m$^{-1/2}$ and $K = 5$ m/s$\cdot\mu$m^{-1}, we can also estimate the ratio K/k. If the overlap is approximately within the range of 0.2 - 0.7 micron, $K/k = 16$, according to the first scenario and $K/k = 17$ according to the second scenario. The two estimates are reasonably close to each other. We must emphasize the approximate nature of these estimates because k and K are evaluated with a large degree of uncertainty. For example, using values inferred by Rushton [3] for the largest non-myelinated axon in the peripheral nerve, $k = 2.09$ m/s$\cdot\mu$m$^{-1/2}$ and $K = 5$ m/s$\cdot\mu$m^{-1}, both scenarios yield $K/k = 4$. Assuming that K can range between 4.5-6 m/s$\cdot\mu$m^{-1}, our estimate of K/k is between 3 and 25.

Prediction of a gap in the CV spectrum could not be verified using available experimental data. Measurements of CV for individual axons with thin

diameter (0.1-1 micron) can be very difficult. Since the first scenario does not predict a gap, testing this prediction would help us distinguish between the two scenarios.

2.5 Discussion

The square-root relation between non-myelinated axon diameter and conduction velocity has proven to be followed roughly in real nerves, but the relationship does not exactly hold in different animals and different nerves. Generally, we could assume a power law between diameter and CV by rewriting Equation 2.2 as

$$v_n = bd_a^{\theta}, \tag{2.15}$$

where $\theta < 1$. Then, we check how Equation 2.15 affects our previous arguments. In the first scenario, the critical CV v_c in Equation 2.8 should be modified as

$$v_c = \left(\frac{K}{k}\right)^{\frac{\theta}{2(1-\theta)}} \left(\frac{b}{B}\right)^{\frac{1}{1-\theta}} B, \tag{2.16}$$

but the dependence of diameter spectrum on the ratio of unit volume cost of myelinated and non-myelinated axons does not change. In the second scenario, following the procedure discussed in the main text and the Methods, we found that there is also a gap in the CV spectrum. However, Equation 2.14 changes to

$$v_m = \frac{3v_n}{2+\theta}. \tag{2.17}$$

In respect to diameter spectrum, we have

- $K = \frac{2+\theta}{3\theta}k$. Axons with a diameter greater than the critical diameter d_0 are myelinated; below the critical diameter they are non-myelinated. We note that although Rushton's result [3] is recovered in this case, the gap in the CV spectrum always exists.

- $K > \frac{2+\theta}{3\theta}k$. The diameter distributions of non-myelinated and myelinated axons overlap.

- $K < \frac{2+\theta}{3\theta}k$. The diameter distributions of non-myelinated and myelinated axons are separated by a gap.

2.6 Methods

Given the length of an axon L_a, the unit cost of conduction delay α, the total cost in 2.9 is given by

$$E_a = k\frac{\pi}{4}d^2 L_a + \alpha\frac{L_a}{v}. \tag{2.18}$$

By substituting Equations 2.1-2.2 into Equation 2.18, the cost per length of a myelinated and a non-myelinated axon can be formulated respectively as:

$$\epsilon_m = K\frac{\pi}{4}D^2 + \frac{\alpha}{BD}. \tag{2.19}$$

$$\epsilon_n = k\frac{\pi}{4}{d_a}^2 + \frac{\alpha}{bd_a}. \tag{2.20}$$

By setting $\frac{\partial \epsilon_m}{\partial D} = 0$, $\frac{\partial \epsilon_n}{\partial d_a} = 0$, we find the optimal diameter of the two

types of axons to be

$$D = \left(\frac{2\alpha}{\pi K B}\right)^{1/3}, \tag{2.21}$$

$$\text{and } d_a = \left(\frac{\alpha}{\pi k b}\right)^{2/5}. \tag{2.22}$$

Substituting Equations 2.21-2.22 into Equations 2.1-2.2, we obtain Equation 2.10 and Equation 2.12. Substituting Equations 2.21-2.22 into Equation 2.19-2.20, we obtain Equation 2.11 and Equation 2.13. By setting Equation 2.11 and Equation 2.13 equal to each other, we find the critical α satisfies

$${\alpha_0}^{2/15} = \frac{6}{5}\left(\frac{\pi K}{2B^2}\right)^{1/3}\left(\frac{b^4}{\pi k}\right)^{1/5}. \tag{2.23}$$

The axon diameters (*i.e.*, Equations 2.21-2.22) become the same when

$$\alpha^{2/15} = \left(\frac{2}{\pi B K}\right)^{2/3}(\pi b k)^{4/5}. \tag{2.24}$$

If $\alpha_0 > \alpha$, or $K/k > 5/3$, there is an overlap in the diameter distributions of myelinated and nonmyelinated axons. Otherwise there is a gap in the axon diameter spectrum.

Chapter 3

Segregation of the Brain into Gray and White Matter: A Design Minimizing Conduction Delays

3.1 Background

A ubiquitous feature of the vertebrate brain is its segregation into white and gray matter (http://www.brainmuseum.org). Gray matter contains neuron somata, synapses, and local wiring, such as dendrites and mostly nonmyelinated axons. White matter contains long-range, and in large brains mostly myelinated, axons that implement global communication. What is the evolutionary advantage of such segregation [20]? Networks with the same local and global connectivity could be wired so that global and local connections

are finely intermixed. Since such design is not observed, and invoking an evolutionary accident as an explanation has agnostic flavor, we searched for an explanation based on the optimization approach [12, 11, 21, 22, 13], which is rooted in the evolutionary theory [23, 24, 25].

We started with the assumption that evolution "tinkered" with brain design to maximize its functionality. Brain functionality must benefit from higher synaptic connectivity because synaptic connections are central for information processing as well as learning and memory, which is thought to manifest itself in synaptic modifications [26, 27]. However, increasing connectivity requires adding wiring to the network, which comes at a cost. The cost of wiring is due to metabolic energy required for maintenance and conduction [28, 29, 30], guidance mechanisms in development [31], conduction time delays and attenuation [32, 14], and wiring volume [13].

Two pioneering studies, by Ruppin *et al.* [33] and Murre and Sturdy [34], have proposed that the segregation of white and gray matter could be a consequence of minimizing the wiring volume. They modeled the brain by a network consisting of local and global connections, which give rise to gray and white matter correspondingly. Although wiring volume minimization is an important factor in the evolution of brain design, their results remain inconclusive because predictions of the volume minimization model for the present problem are not robust and are difficult to compare with empirical observations (see Discussion).

In this chapter, we adopted the model of connectivity introduced in Ruppin *et al.* [33] and Murre and Sturdy [34], including local and global connections, but minimize the conduction delay, *i.e.*, the time that takes a signal

(such as action potential and/or graded potential) to travel from one neuron's soma to another. To see that high connectivity and short conduction delay are competing requirements, note that adding wiring to the network increases not only its volume, but also the distance between neurons. In turn, this requires longer wiring, which, for the same conduction velocity, introduces longer delays. Longer delays are detrimental because fewer computational steps can be performed within the time frame imposed on animals by the environment, making the brain a less powerful computational machine [30].

We show that the competing requirements for high connectivity and short conduction delay may lead naturally to the observed architecture of vertebrate brain as seen in mammalian neocortex and bird telencephalon. As in any other theoretical analysis we make several major assumptions. First, given that exact connectivity is not known, we characterized the inter-neuronal connectivity statistically by requiring a fixed number of connections per neuron. Second, although conduction delays are known to differ between connections, we minimized the mean conduction delay. Finally, it is likely that, in the course of evolution, minimization of conduction delay was accompanied by the increase in connectivity. However, it is not known how to quantify the benefits of increased connectivity in comparison with conduction delay increase. Therefore, we adopted a mathematically sound approach of minimizing conduction delay while keeping network connectivity fixed.

To obtain quantitative results, we used two analytical (non-numerical) tools borrowed from theoretical physics. First, most of the derivations were done using the scaling approach. In this approach, a relationship between variables takes the form of proportionality rather than equality. In other words,

numerical factors of order one are ignored. One can manipulate and combine such proportionality relationships and still get an estimate that is correct by an order of magnitude. A long history of successful applications of the scaling approach supports its validity. Second, we used a perturbation theory approach, which is helpful when the exact analytical solution to a problem is unavailable. In this approach, a simpler problem is solved exactly. Then the exact solution is modified to fit the actual problem by taking advantage of the fact that such modification is minor. Again, the long history of this approach supports its validity as long as the difference between the exactly solvable and the actual problem is characterized by a parameter that is much smaller than one.

We present our theory in Results, which are organized into seven sections. In Section 3.2, we consider competing requirements between small conduction delays and high connectivity in local circuits. We show that local conduction delay limits the size of the local network with all-to-all potential connectivity to the size of the cortical column. Section 3.3 models full brain architecture as a small-world network, which combines high local connectivity with small conduction delay. We derive a simple estimate of conduction delay in global connections as a function of the number of neurons. Section 3.4 considers spatially integrating local and global connections. We argue that mixing local and global connections substantially increase local conduction delay, while the global conduction delay may be unaffected. In Section 3.5, by minimizing local conduction delay, we derive a condition, under which white/gray matter segregation reduces conduction time delays. Section 3.6 gives a necessary condition for the segregated design to be optimal. An example of such design

is given in Section 3.7. Finally, section 3.8 restates our results in terms of the numbers of neurons, inter-neuronal connectivity, and axon diameter.

3.2 Conduction Delays Limit Size of a Highly Connected Network

We begin by considering time delay in the local circuits of neocortex because their mode of operation, thought to involve recurrent computations [35, 36], seems most sensitive to the detrimental impact of time delay. We derive a scaling relationship between local conduction delay and the number of neurons that can have all-to-all potential connectivity. By assuming that the tolerable delay is on the order of a millisecond, we show that the maximum size of such network is close to that of the cortical column.

Local cortical circuits may be viewed as a network of n neurons with all-to-all potential synaptic connectivity, meaning that axons and dendrites of most neurons come close enough to form a synapse [37, 13, 38]. In the following we do not distinguish between axons and dendrites in local circuits and refer to them as local wires. The mean conduction delay t in local circuits is given by the average path length between two connected neurons (via potential synapses), ℓ, divided by the conduction velocity, v_n:

$$t = \frac{\ell}{v_n}. \tag{3.1}$$

Experimental measurements [6, 39] and theoretical arguments [3, 40] suggest that conduction velocity, v_n, scales sub-linearly with the diameter, d, of local

wires (non-myelinated axons and dendrites):

$$v_n = bd^\theta, \tag{3.2}$$

where b is a constant coefficient and θ is a positive power smaller than one (however see [41]). By combining Equations 3.1-3.2, we arrive at the expression for the conduction delay:

$$t = \frac{\ell}{bd^\theta}. \tag{3.3}$$

Equation 3.3 may give an impression that the conduction delay decreases monotonically with wire diameter d. But this is not necessarily the case because ℓ can be a function of d. The following argument [32] shows that the conduction delay, t, as a function of wire diameter, d, has a minimum (provided $0 < \theta < 1$), which defines the optimal wire diameter. Given the branching structure of axons and dendrites and uniform distribution of neurons, ℓ can be approximated by the linear size of the network [13], which can be easily estimated in the two limiting cases. In the limit when the wire diameter approaches zero, all the non-wire components, such as synapses, are compressed together and take up the space vacated by shrinking wires. Because the volume of the network approaches the volume of the non-wire components, which is constant, the conduction delay diverges as $1/d^\theta$ according to Equation 3.3, [32].

In the opposite limit when the wire diameter is large, the network volume is determined mostly by the wiring [32]. As wires run in all directions, they have to get longer as they get thicker and the linear size of the network grows proportionally to the wire diameter. Then, according to Equation 3.3, the

conduction delay increases as $d^{1-\theta}$. Therefore, conduction delay is minimized by the optimal wire diameter, for which the non-wire occupies a fixed fraction of the neuropil volume [32] (also see Methods 3.10.1). Then the optimal volume of the network is of the same order as the non-wire volume. Assuming that non-wire consists mostly of synaptic components, such as axonal boutons and spine heads, the optimal network volume is of the same order as the total synaptic volume. Therefore, the local network volume is given by:

$$\ell^3 \sim n^2 v_s, \tag{3.4}$$

where v_s is the average synapse volume, n is the total number of neurons in the local network. (In a network with all-to-all connectivity, n is also the number of local connections made by a neuron via potential synapses). For the sake of clarity, we ignore the fact that only a fraction (0.1-0.3) of potential synapses are converted into actual [37]. Such numerical factors are ignored in the equations of the main text, but can be included straightforwardly (see Methods 3.10.1). One consequence of Equation 3.4 is that the optimal wire diameter is on the same order of magnitude as the synaptic linear size, consistent with anatomical observations [42]:

$$d \sim {v_s}^{1/3}. \tag{3.5}$$

By using Equations 3.3-3.5 and assuming $\theta = 1/2$, suggested by the cable theory [3, 40], we find that the smallest possible mean conduction delay in local networks is given by

$$t \sim n^{2/3} {v_s}^{1/6}/b. \tag{3.6}$$

As the smallest possible conduction delay grows with the number of neurons in the network, fixing conduction delay imposes a constraint on the maximum size of the network. It seems reasonable to assume that the biggest tolerable conduction delay is on the order of a millisecond, a time scale corresponding to physiological events such as the extent of an action potential and the rise-time of an excitatory postsynaptic potential [43]. This time scale could be dictated by the metabolic costs [33]. If we approximate the synaptic volume at a fraction of a cubic micrometer, and $b \sim 1$ m/s$\cdot\mu$m$^{-1/2}$ [3, 15, 19], the maximum number of neurons in the all-to-all connected network is on the order of 10^4. This corresponds to roughly the size of a cortical column, which is then the largest network that can combine all-to-all potential synaptic connectivity with tolerable conduction delay.

3.3 Small-world Network Combines High Local Connectivity with Small Conduction Delay

Human neocortex contains about 10^{10} neurons; many more than could possibly be wired in an all-to-all fashion with a physiologically tolerable conduction delay. In particular, substituting this neuron number into Equation 3.6, we find that delay would be on the order of seconds – clearly too slow. Given that the brain is too large to combine high inter-connectivity with short conduction delay [44, 45] how can it maintain high functionality? In this section, we consider the architecture of the brain as a whole and show that

much shorter global conduction delay can be achieved by sacrificing all-to-all connectivity.

Anatomical evidence suggests that the brain maintains short conduction delay by implementing sparse global inter-connectivity while preserving high local inter-connectivity [42]. Such design resembles the small-world network [46] as has been noticed by several authors [47, 48, 49, 50]. In a small-world network, a high degree of clustering (the probability of a connection between two neighbors of one neuron) is combined with a small network diameter (the average number of synapses on the shortest path connecting any two neurons). In neurobiological context this means a combination of high computational power in local circuits with fast global communication [44, 45, 42, 47, 48, 49]. Thus it is not surprising that evolution adopted this architecture when the size of the network made all-to-all connectivity impractical [51, 52, 44, 53, 47].

How fast could global connections be? Global conduction delay T in a connection of length L_a with conduction velocity v_m is given by

$$T = \frac{L_a}{S}. \tag{3.7}$$

Here and below, upper-case letters are reserved for parameters of global connections, while lower-case letters are for parameters of local connections. In big brains, global axons are mostly myelinated as would be expected, given higher demand on their conduction velocity see Chapter 1 and [3]. In myelinated axons, conduction velocity scales linearly with diameter [5, 3], D:

$$v_m = BD, \tag{3.8}$$

where B is a proportionality coefficient. Combining Equations 3.7-3.8, we find that the conduction delay is given by

$$T = \frac{L_a}{BD}. \tag{3.9}$$

Average length of global connections is given by

$$L_a \sim V^{1/3}, \tag{3.10}$$

where V is brain volume. In turn, brain volume can be estimated by adopting the following model. Based on anatomical data [42], we assume that most neurons send one global connection to another local network in the brain. Initially, we ignore the volume occupied by local connections. We denote the number of neurons in the brain as N, which can be much larger than the number of local connections (via potential synapses) per neuron, n. Global connections have length, L_a and diameter, D. Then the total volume of the brain can be approximated as

$$V \sim ND^2L_a. \tag{3.11}$$

Combining Equations 3.10-3.11, we find

$$L_a/D \sim \sqrt{N}. \tag{3.12}$$

Substituting this expression into Equation 3.9, we obtain

$$T \sim \frac{\sqrt{N}}{B}. \tag{3.13}$$

Equation 3.13 can be used to estimate conduction delay in global axons. By substituting $B \sim 5$ m/s$\cdot\mu$m^{-1} [5, 18] and the number of neurons in human neocortex, $N \sim 10^{10}$, we find that the delay is around 20 ms. Compared with the several-second delay expected in a human brain if it had all-to-all connectivity this is a significant improvement. For the mouse neocortex, by substituting $N \sim 10^7$, we find that the delay is around 0.6 ms. This is much better than the 50 ms delay expected, according to Equation 3.6, if the mouse cortex had all-to-all connectivity. As these estimates are based on the scaling approach, they are reliable only up to an order of magnitude. However, they demonstrate that sparse global connections can be much faster than a fully connected network with comparable number of neurons.

3.4 Combining Local and Global Connections Increases Conduction Delays

After considering conduction delays in local and global connections separately, now we are in a position to analyze how they are combined in the brain. Here we argue that the main difficulty in integration arises when introducing global connections into local networks.

We adopt a model combining both local and global connections proposed by Ruppin *et al.* [33] and Murre [34]. In this model, each neuron connects (via

potential synapses in our case) with n local neurons and sends a global axon to another arbitrarily chosen local network in the brain. For simplicity, we neglect specificity and assume that local connections are made with nearest n neurons located in a sphere of radius l centered on a given neuron, where ℓ is given by Equation 3.4. Although local and global connections may be highly specific [54, 55, 36, 56], this approximation is sufficient to understand brain segregation into white and gray matter.

The effect of combining local and global connections on the conduction delays can be analyzed in two steps. First, consider the effect of introducing local connections into the network of global connections. This leads to the swelling of the brain volume beyond that in Equation 3.11. Thus, global axons must be longer and Equation 3.13 gives only the lower bound for global conduction delay (see Methods 3.10.2). However the increase in the global conduction delay caused by the swelling of network can be offset via speeding up global axons by making them thicker, Equation 3.8. We show in Methods 3.10.2 that the global network can absorb local connections and preserve the required global conduction delay.

Second, introduction of global connections into local circuits increases local conduction delay and is impossible to compensate by making local connections thicker (see Methods 3.10.3). While conduction velocity depends linearly on the global myelinated axon diameter, Equation 3.8, it scales sub-linearly with the local wire diameter, Equation 3.2. Thus, the smallest possible mean local conduction delay increases when more global connections are mixed with local connections. To describe this quantitatively, we introduce the ratio of global axon volume that is finely intermixed with local connections to the ini-

tial unperturbed gray matter (*i.e.*, total local circuits) volume, λ. When λ is much smaller than one, we can argue that the initial minimum local conduction delay is only slightly affected by the penetration of global connections in the gray matter. As shown in Methods 3.10.3, the increase in local conduction delay, Δt, because of intermixing global connections and local connections is proportional to the ratio λ:

$$\frac{\Delta t}{t} \sim \lambda, \tag{3.14}$$

where t is conduction delay in unperturbed local circuits given by Equation 3.6. As before, numerical factors are neglected in the spirit of the scaling estimate.

According to our original assumption, brain functionality is maximized when conduction delay is minimized. According to Equation 3.14, the smallest possible conduction delay in local circuits is achieved when $\lambda = 0$, *i.e.*, when global and local connections are fully segregated. But full segregation does not lead to a feasible design because global connections originate and terminate on neurons in local circuits. Thus, we must find a design that spatially integrates local and global connections.

We note that minimization of local and global conduction delays are competing desiderata, as can be illustrated by varying the global axon diameter, D. Increasing D speeds up signal propagation along global connections and, therefore, reduces global conduction delay. However, thicker global axons are detrimental for local conduction delay because of an increase in λ, Equation 3.14. As the relative contributions to functionality of conduction delays in local and global connections are unknown, we searched for the optimal design

that minimizes local conduction delay as a function of D. Our analysis begins with considering small values of D, *i.e.*, $\lambda \ll 1$.

3.5 Comparison of the Homogeneous Design (HD) and Designs with Gray and White Matter Segregation

In order to determine the optimal design we need to compare local conduction delays in different designs combining gray and white matter. In general, this problem is difficult to solve analytically. However, when global connections intermixed with the gray matter take less volume than local, *i.e.* $\lambda \ll 1$, the perturbation theory approach allows us to compare local conduction delays in homogeneous design (HD), in which gray matter and white matter are finely intermixed, to designs in which gray and white matter are segregated.

In HD, local and global connections are finely and uniformly inter-mixed (Figure 3.1). Then, according to Equation 3.14, the relative conduction delay increase due to the penetration of global axons of diameter D in the gray matter is given by

$$\frac{\Delta t}{t} \sim \lambda \sim \frac{N D^2 G^{1/3}}{G}, \tag{3.15}$$

where N is the total number of neurons in the network. In this expression, we used Equation 3.11 for the volume of global connections and the fact that the average length of global axons is given by the linear size of the network, which, for small λ is given by the linear size of gray matter, $G^{1/3}$. We note that the

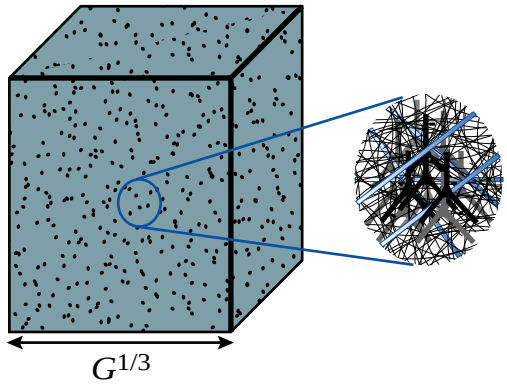

Figure 3.1: Homogeneous Design. In HD, local and global connections are uniformly and finely intermixed. Inset shows a typical local network containing local axons (thin gray lines) and dendrites (gray and black tree-like structures), and global axons (thick, light-blue lines spanning the whole circle) that perforate gray matter. When the volume of global axons is small, the linear size of the network can be approximated as $G^{1/3}$.

perturbation approach remains valid while the relative conduction increase in HD is less than one, *i.e.*, $ND^2 \ll G^{2/3}$.

Another contribution to the mean local conduction delay comes from the boundary effect. Recall that the model requires each neuron in the gray matter to establish connections with n nearest neighbors. If a neuron is far from the boundary of the gray matter, these connections can be implemented in a sphere of radius ℓ given by Equation 3.4 (Figure 3.2). However, neurons within distance ℓ of the gray matter boundary cannot find n neighbors within the sphere of the same size. Therefore, the radius of the local connections sphere must be expanded to find n nearest neighbors (Figure 3.2).

Expanding the range of local connections for neurons near the boundary increases average local conduction delay. The fraction of neurons that experience the boundary effect is proportional to the volume within distance ℓ from

the boundary. As the boundary area in HD is given by $G^{2/3}$, the fraction of affected neurons is given by $\ell G^{2/3}/G \sim \ell/G^{1/3}$, which is less than one because the linear size of the gray matter $G^{1/3} \gg \ell$. Since the relative increase in delay for each neuron in the affected volume is of order one, this expression also gives a relative increase in the average local conduction delay. As this boundary effect is determined by the external boundary, it is independent of the design and can be ignored. However, the logic of this calculation will be used in the following to estimate the effect of gray and white matter boundary on local conduction delay.

Can segregation of gray and white matter reduce local conduction delay in HD? In HD, global axons are straight and are finely intermixed with the local connections. The contribution of global axons to local conduction delays could be reduced by decreasing the length of global axonal segments within the gray matter, according to Equation 3.14. Rather than connecting neurons with a straight axon, a typical global axon would go towards the nearest white matter tract (region occupied only by global axons) and travel in the white matter until it is close to the target neuron. Then the axon would leave the white matter and traverse the gray matter towards its target (Figure 2). Such routing may increase the length of global axons but it would minimize impact on local conduction delays.

To calculate the relative local delay increase in the segregated design we estimate the relative volume of global axons in the gray matter, λ. We introduce the mean distance between a neuron and the nearest white matter tract, R, which also gives the linear size of gray matter modules (Figure 2). Then the relative volume of non-fasciculated global axons inside the gray matter in

the segregated design

$$\lambda \sim ND^2R/G. \tag{3.16}$$

Comparing Equation 3.16 with Equation 3.15 one can see that segregation may be advantageous to HD if $R \ll G^{1/3}$. In other words, introducing a sufficient number of white matter tracts in the gray matter may reduce the length of non-fasciculated global axonal segments in the gray matter and hence, local conduction delay.

Although segregation of gray and white matter may reduce local conduction delay, it has a disadvantage compared to HD in that it may induce a larger boundary effect because of the white matter tracts inside the gray matter. This effect is similar to the external boundary effect in HD but cannot be ignored because it is different for different designs. If a neuron is far from the gray and white matter interface, its local connections can be implemented in the sphere of radius ℓ (Equation 3.4, Figure 3.2). If a neuron is close to the interface, the white matter occupies part of the sphere, meaning that the local sphere radius ℓ must be expanded so that a neuron can still find its n nearest neighbors (Figure 3.2). Therefore, whether the segregated design is preferred or not depends on whether the relative local conduction delay increase through the boundary effect is much smaller than the local delay increase in HD, Equation 3.15.

To evaluate the mean local conduction delay increase through the boundary effect in the segregated design, we need to specify the geometry of the white matter tracts, as the boundary effect generally depends on the surface area of the tracts. For a typical tract that spans the whole brain (*i.e.*, has

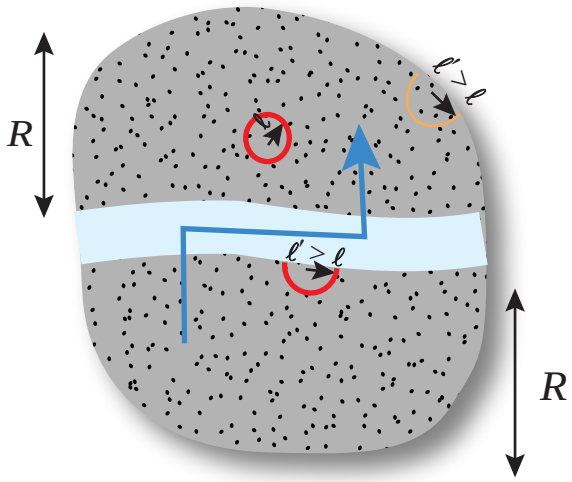

Figure 3.2: Boundary Effects in the Gray Matter. The red full circle illustrates the local connection sphere of a neuron that does not experience the boundary effect. Neurons near external boundary must inflate their local connection sphere to implement the required local connectivity, as illustrated by thin yellow semicircle. Neurons near white matter tracts penetrating gray matter must also inflate their local connection sphere to implement the required local connectivity, as illustrated by the thick red semicircle. The blue line with arrowhead shows typical routing of global axons. R is the size of gray matter modules, where global and local connections are finely intermixed.

length L), we can relate its minimal surface area A_t to its cross-sectional area, Φ:

$$A_t \sim \sqrt{\Phi} L. \tag{3.17}$$

In turn, the cross-sectional area of a tract depends on the global axon diameter D, and one may conjecture that whether the segregated designs are advantageous or not depends on D. Indeed, we can formulate the following theorem, which is valid to the first order of $ND^2/G^{2/3}$, Equation 3.15, and while our perturbation approach is valid (*i.e.*, provided $ND^2 \ll G/\ell$, as will be shown later).

Theorem 1 *In the regime $ND^2 \ll \ell^2$, local conduction delays in the optimal segregated design and HD are equivalent. In the regime $ND^2 \gg \ell^2$, there is at least one segregated design with local delays less than those in HD.*

To prove the first part of the theorem, we calculate the local conduction delay through the boundary effect in the segregated designs and compare it with HD. The length of the global tract segment inside the local sphere is ℓ. The other two dimensions of global tracts are much less than ℓ (Figure 3.3A), as the minimal boundary effect is achieved by the minimal surface area in Equation 3.17. Since the total cross-sectional area of the global tracts is $ND^2 \ll \ell^2$, each tract's cross-sectional area, Φ_i, is much less than the cross-sectional area of the local connection sphere (Figure 3.3A). Inclusion of such tract into a local sphere increases its radius to $\sqrt{\ell^2 + \Phi_i}$. Then, the relative increase in the local conduction delay for neurons in that sphere is $(\sqrt{\ell^2 + \Phi_i} - \ell)/\ell \simeq \Phi_i/\ell^2 \ll 1$.

Now we add up conduction delays contributed by all the tracts to neurons

in affected spheres. As the number of spheres affected by one tract is given by L/ℓ, the fraction of neurons experiencing the boundary effect induced by one tract is given by $ell^2 L/G$ and the relative local conduction delay increase is given by $(\Phi_i/\ell^2)\ell^2 L/G \sim \Phi_i/G^{2/3}$. The total relative increase in local delay is the sum of the boundary effects induced by different tracts,

$$\Delta t/t \sim \sum \Phi_i/G^{2/3} \sim ND^2/G^{2/3}. \tag{3.18}$$

Notice that even if there are multiple tracts within the local connection sphere (*i.e.*, the sphere radius can be larger than ℓ), the above result is still correct.

By comparing local conduction delay increase for segregated designs, Equation 3.18, with that for HD (Equation 3.15), one can see that they are the same. Therefore, when $ND^2 \ll \ell^2$, the optimal segregated designs and HD are equivalent to the first order of $ND^2/G^{2/3}$.

To prove the second part of the theorem (the $ND^2 \gg \ell^2$ regime), we specify a segregated design with smaller local delays than that in HD. In such design, global axons belong to M ($M \gg 1$) tracts with cross-sectional area $\Phi \gg \ell^2$ each and length $L \sim G^{1/3}$. The distance between two tracts is much larger than ℓ. Then, the total affected neuropil volume through the boundary effect is the product of the total surface area of the tracts, $M\Phi^{1/2}G^{1/3}$, and ℓ. For a typical neuron within the affected volume, a fraction of its local connection sphere with volume $\sim \ell^3$ is occupied by the white matter tract, as illustrated in Figure 3.3B. To implement the required local connectivity, the local sphere radius ℓ should expand by a numerical factor of order one.

Next, we add up relative local delay increase induced by all global tracts

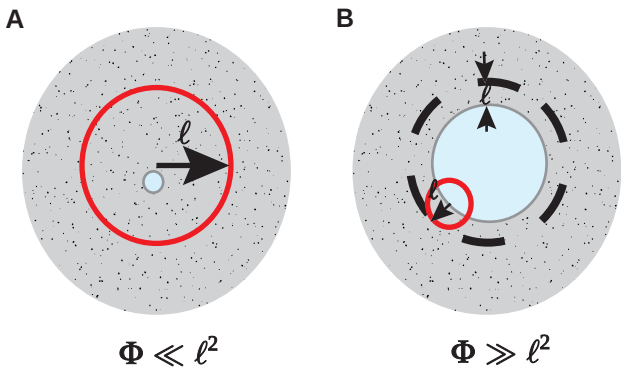

Figure 3.3: Boundary effect induced by white matter tracts with different cross-sectional areas. A. In the case $\Phi \ll \ell^2$, two dimensions of the white matter tracts (shown in white) can be much smaller than ℓ. The red circle illustrates local connection sphere of a neuron. B. In the case $\Phi \gg \ell^2$, neurons within distance ℓ from the white matter tract experience the boundary effect.

affecting all the neurons in a volume, given by $\ell M \Phi^{1/2} G^{1/3}/G$. As the total cross-sectional area $M\Phi \sim ND^2$, the relative local delay increase becomes

$$\frac{\Delta t}{t} \sim \frac{ND^2 \ell}{\Phi^{1/2} G^{2/3}}. \tag{3.19}$$

By comparing relative conduction delay in segregated design, Equation 3.19 with that in HD, Equation 3.15, one can see that because $\Phi \gg \ell^2$ as specified, segregated design is advantageous in the regime $ND^2 \ll G^{2/3}$.

Although in the regime $ND^2 \gg G^{2/3}$ we do not have a closed form expression for the local conduction delay in HD, we can still show that it has longer conduction delays than the segregated design. We show in Methods 3.10.3 that local conduction delay in HD is a monotonically increasing function of λ, and hence a monotonically increasing function of ND^2. Thus, the relative delay increase in HD exceeds one when $ND^2 \gg G^{2/3}$. However, in the regime

$ND^2 \gg G^{2/3}$, the relative local delay increase in a segregated design can still be much smaller than one. To prove this, we note that in a segregated design, the local conduction delay increase because of the non-fasciculated global axons intermixed with gray matter *i.e.*,$\lambda \sim ND^2R/G$ (Equation 3.16), can be much smaller than one, if $R \ll G^{1/3}$.

In addition, the relative local delay increase through the boundary effect can also be much smaller than one. To see this, we specify the tracts in such a way that the total surface area of the white matter tracts is the surface area of the gray matter G/R. Then, using analysis similar to that illustrated in Figure 3.3B, the relative local delay increase through the boundary effect is given by $\frac{\ell G}{RG}$, which can be much smaller than one if $\ell/R \ll 1$. We note that $\lambda \ll 1$ and $R \gg \ell$ could both be satisfied if $ND^2 \ll G/\ell$. Thus, when $ND^2 \gg G^{2/3}$ and $ND^2 \ll G/\ell$, there is at least one segregated design with local delay less than that in HD.

Having considered both the $ND^2 \ll G^{2/3}$ regime and $ND^2 \gg G^{2/3}$ regime, we have proven the second part the theorem.

3.6 Optimality Condition for Segregated Designs

In the previous section, we showed that in the regime $ND^2 \gg \ell^2$ there is at least one segregated design with local conduction delay shorter than that in HD. However, we did not specify which design is the optimal one. In this section, we give a necessary condition for a segregated design to be optimal in

the regime $ND^2 \gg \ell^2$ and if $ND^2 \ll G/\ell$.

As the advantage of segregation becomes apparent when the total cross-section of global axons $ND^2 \sim \ell^2$, it is natural to expect that a similar condition defines the optimal gray matter module size R_0, which minimizes local conduction delays. In other words, the number of neurons in the gray matter module is such that the total cross-sectional area of their global axons is given by ℓ^2. As the number of neurons in the sphere of radius R_0 is ℓ^2/D^2 and the number of neurons in the sphere of radius ℓ is n, we have

$$R_0 \sim \left(\frac{\ell^2}{nD^2}\right)^{1/3} \ell. \tag{3.20}$$

Then, we can formulate the following theorem:

Theorem 2 *In the regime $ND^2 \gg \ell^2$ and $ND^2 \ll G/\ell$, the minimum local conduction delay is achieved by the segregated design with the gray matter module containing ℓ^2/D^2 neurons.*

To prove this theorem, we consider designs with gray matter module size smaller and greater than R_0 and show that they have a local conduction delay greater than that of the design with module size R_0.

In the case $R_0 \ll R$, by applying Theorem 1 to any module, one can see that converting that module from HD to segregated designs can reduce local conduction delay. For example, fasciculating global axons within that module into multiple tracts would reduce local conduction delay.

In the other case, if modules with size R_0 contain only global axons from the neurons inside the module, by applying Theorem 1 one can see that any

optimal segregated designs containing modules with size $R \ll R_0$ is equivalent to designs containing modules with size R_0.

Moreover, if the tracts inside the module of size R_0 contain external global axons (*i.e.*, global axons that do not belong to the neurons inside the module with size R_0 and/or do not innervate the neurons inside the module), converting segregated designs with module size $R \ll R_0$ to designs with module size R_0 reduces the local conduction delay. This happens because merging all the tracts within the module of size R_0 into one reduces the boundary effect. To see this, note that the minimal surface area of the big tract inside the module with size R_0 is on the order of $(\sum \Phi_i)^{1/2} R_0 \ll \sum {\Phi_i}^{1/2} R_0$, where Φ_i is the mean cross sectional area of a small tract containing external global axons and $\sum {\Phi_i}^{1/2} R_0$ is the total surface area of the smaller tracts inside the module with size R_0. Even if the tracts run in different directions, most of the tracts can be merged together at the scale R_0 because the typical length of a tract is much greater than that and a small curvature would not affect the total length by an order of magnitude.

Taken together, by considering the two possible cases, we have proven that the minimum conduction delay in segregated designs is achieved with module size R_0. Such designs may be further classified by the relative dimensions of the gray matter. The total boundary area between gray and white matter (*i.e.*, the total surface area of the white matter tracts), A, could satisfy either $A \sim G/R_0$ or $A \ll G/R_0$. As the local conduction delay through the boundary effect grows with A, the latter design has the shorter delay. In the following, we call segregated designs satisfying $A \ll G/R_0$ the perforated design (PD).

3.7 Branching Pipe Design – An Example of Perforated Design

In the previous section, we have shown that in the optimal segregated designs, the size of the module, in which global and local connections are finely intermixed, is given by R_0. However, Theorem 2 does not specify other dimensions of the segregated design, such as the total surface area of the white matter tracts. In this section, by considering a specific example, which we name branching pipe design, we show that the condition $A \ll G/R_0$ can be satisfied in the regime where our perturbation approach is valid. In other words, we prove that PD exists in the regime $ND^2 \ll G/\ell$.

We specify the branching pipe design as follows (Figure 3.4). Global axons belong to several cylindrical white matter pipes perforating the gray matter. Higher-order branches split off lower-order pipes at regular intervals. Different order branches have different lengths and different pipe diameter. The length of the zeroth-order branches (*i.e.*, the main pipes) is given by the linear size of the brain. The length of k+1st order branches is given by the inter-pipe distance among the kth order branches, forming a space filling structure. The inter-pipe distance among the finest branches is given by R_0 in Equation 3.20 (Figure 3.4).

Although we can calculate the total surface area of the branching pipes for any given order k (as discussed in Methods 3.10.4), for simplicity we present the main results from the branching pipe design in which only first-order branches exist. We minimize the total surface area of such branching pipes and the local conduction delay by searching for the optimal length and the diameter

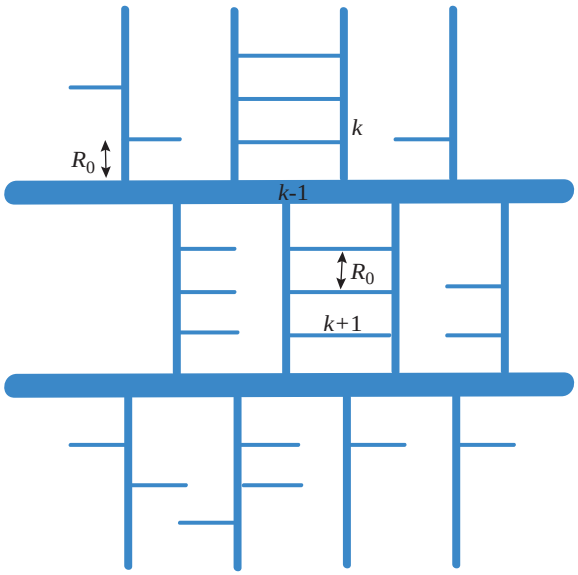

Figure 3.4: Branching pipe design. Schematic illustration of branching pipe design with three orders of branches. The distance between kth order branches determines the length of the $k+1$st-order branches. The distance between highest-order branches is given by R_0.

of the first-order branches and the optimal diameter of zeroth-order branches.

We find that the expression for the minimal total surface area of the first-order branching pipes A depends on whether the total white matter volume is greater than the total gray matter volume or not. In the regime $\ell^2 \ll ND^2 \ll G^{2/3}$, the gray matter occupies most of the brain volume, and A is calculated in Methods 3.10.4:

$$A \sim \left(\frac{ND^2}{\ell^2}\right)^{1/18} \lambda^{1/2} \frac{G}{R_0}. \tag{3.21}$$

In turn, ℓ can be found by substituting $G \sim (N/n)\ell^3$ and optimal $R \sim R_0$ (Equation 3.20) into Equation 3.16:

$$\lambda \sim \left(\frac{nD^2}{\ell^2}\right)^{2/3}. \tag{3.22}$$

Then the minimal local conduction delay is given by

$$\Delta t/t \sim \left(\frac{ND^2}{\ell^2}\right)^{1/18} \lambda \sim \left(\frac{n}{N}\right)^{2/3} \left(\frac{ND^2}{\ell^2}\right)^{13/18}, \tag{3.23}$$

This dependence of $\Delta t/t$ on ND^2 is plotted on log-log scale in Figure 3.5 (represented by thick blue line).

In the regime $G^{2/3} \ll ND^2 \ll G/\ell$, white matter occupies most of the volume and the specified segregated design has a different appearance: the gray matter is confined to a thin sheet. Sheet thickness is given by the length of highest-order branches. Then, the minimal surface area of the branching pipes (as calculated in Methods 3.10.4) is given by

$$A \sim \lambda^{1/3} G/R_0. \tag{3.24}$$

In this regime, the minimal local conduction delay is given by (Figure 3.5)

$$\Delta t/t \sim \lambda^{5/6} \sim \left(\frac{nD^2}{\ell^2}\right)^{5/9}. \tag{3.25}$$

As $\lambda \ll 1$ is equivalent to $ND^2 \ll G/\ell$ (to see this, substitute $G \sim (N/n)\ell^3$ into $ND^2 \ll G/\ell$ and compare it with Equation 3.22), we show that for such a branching pipe design, $A \ll G/R_0$ in the regime where our perturbation approach is valid. In other words, we verify the existence of PD in the regime $ND^2 \ll G/\ell$.

We note that when λ is approaching one, according to Equation 3.20 and Equation 3.22, ${R_0}^2 \sim \ell^2 \sim nD^2$, meaning that the total surface area of the gray matter with size ℓ is taken up by the global axons. Therefore, when $\lambda \to 1$, we must have $A \sim G/R_0 \sim G/\ell \sim ND^2$. This can also be seen from the expressions for A in the branching pipe design, *i.e.*, Equations 3.21 and 3.24. Moreover, $\lambda \sim 1$ (*i.e.*, $ND^2 \sim G/\ell$) is when our perturbation approach to calculating the local conduction delay in PD breaks down (Figure 3.5).

When $ND^2 \gg G/\ell$, *i.e.*, $\lambda \gg 1$, we may consider clusters with discrete spatial arrangement and each cluster has n neurons to implement local connectivity. In this case, we can estimate the lower limit of the cluster size, given by $n^{1/2}D$, assuming that cluster volume is filled by tightly packed global axons. Because of local connections, the actual cluster size must be even greater. Alternatively, clusters may abut each other to form a sheet and the sheet thickness could be much smaller than ℓ. In this case, however, we cannot determine the necessary conditions for the design to be optimal. Fortunately, existing anatomical data suggest that actual brains are not even close to the

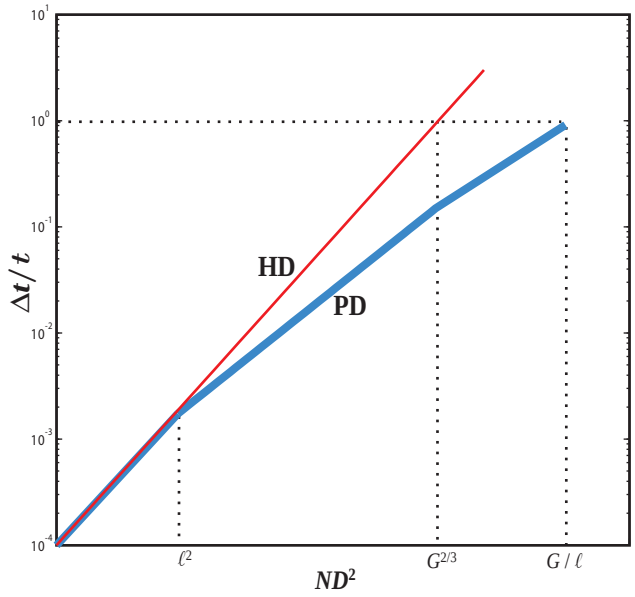

Figure 3.5: Local conduction delay as a function of global axon diameter in HD and PD. Local conduction delay is calculated for specific values $\ell = 0.5$ mm, $N = 10^8$, and $G = 10^3$ mm^3 and plotted in log-log coordinates. Thin red line, local conduction delay in HD; thick blue line, local conduction delay in PD. Delay in PD is calculated for the branching pipe design containing only first-order branches.

regime where $\lambda \gg 1$, as will be shown later.

3.8 Phase Diagram of Optimal Designs

In previous sections we derived conditions under which various designs are optimal in terms of minimizing conduction delays. Specifically, HD is optimal if $ND^2 \ll \ell^2$ and PD is optimal if $ND^2 \gg \ell^2$ and $\lambda \ll 1$. Next, we illustrate these results on a phase diagram (Figure 6) in terms of basic network parameters, such as the local wire diameter d, the number of local connections

(via potential synapses) per neuron n, global axon diameter D and the total number of neurons in the brain N. To obtain the phase diagram, in the first order perturbation theory, we substitute the expression for l, Eqs. (4)-(5), into $ND^2 \gg \ell^2$, and find that PD is optimal when $(N/n)^{1/2}\frac{D}{n^{1/6}d} \gg 1$. In the linear-log space of Figure 3.6, this expression corresponds to the regime above the thick green line.

Next, we estimate where perturbation theory fails by setting λ to one. By substituting Equations 3.4-3.5 into the expression for λ (Equation 3.22) we find that λ can be rewritten as

$$\lambda \sim \frac{D^{4/3}}{n^{2/9}d^{4/3}}. \tag{3.26}$$

Then condition $\lambda \sim 1$ is equivalent to $n^{1/6}d/D \sim 1$, corresponding to the thin red line in Figure 6.

3.9 Discussion

We have shown that the segregation of the brain into gray and white matter may be a natural consequence of minimizing conduction delay in a highly inter-connected neuronal network. We related the optimal brain design to the basic parameters of the network such as the numbers of neurons and connections between them, as well as wire diameters. Although we do not know whether competing desiderata of short time delay and high interconnectivity were crucial factors driving evolution of vertebrate brains, our theory makes testable predictions. Below, we compare these predictions with

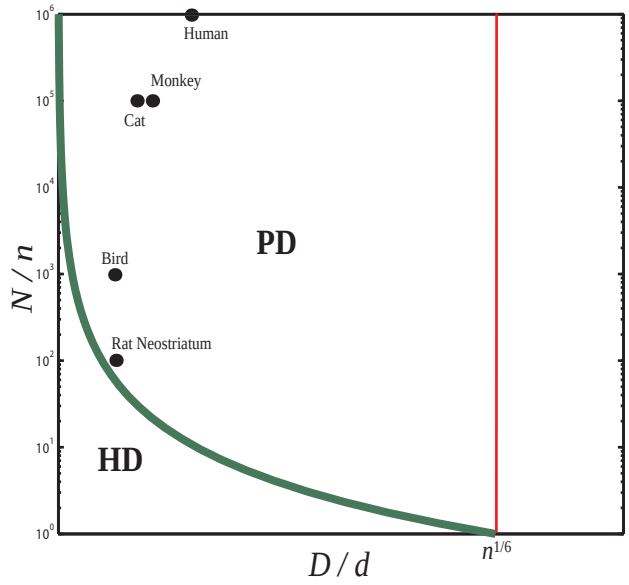

Figure 3.6: Phase diagram of optimal designs. In this phase diagram, we show parameter regimes in which HD or PD are optimal in terms of the global axon diameter D, local wire diameter d, total neuron number N, and the number of local connections per neuron n. We assume $n = 10^4$ and $d = 1$ μm for all empirical data points. Values of D in mammalian brains are from S. S.-H Wang (personal communication) and [57], and values of N in the neocortex are from [53]. Value of N in rat neostriatum is from [58]. For birds, we assume $N = 10^7$.

known anatomical facts.

3.9.1 Scaling Estimate of the Cortical Thickness

As fasciculated fibers are usually not observed in neocortical gray matter (according to Nissl and myelin stains), we identify cortical thickness with gray matter module size, R. Our prediction for the optimal module size R_0 (Equation 3.20) can be rewritten by using Equations 3.4-3.5 as

$$R_0 \sim n^{7/9} d^{5/3} / D^{2/3}. \tag{3.27}$$

Using $n \sim 10^4$ [42, 36], $d \sim 1$ μm [42] and $D \sim 1$ μm[42] (also measured in the corpus callosum of macaque monkey, Wang S.S.-H, personal communication), we predict cortical thickness $R_0 \sim 1$ mm. This estimate agrees well with the existing anatomical data [59, 60, 51] despite being derived using scaling. By substituting these values into Equation 3.26, we find that λ is smaller than one, justifying our perturbation theory approach.

Next, we apply our results to the allometric scaling relationship between cortical thickness, R_0, and brain volume, V. We assume that n and D both increase with brain size [47, 48, 25] according to the following power laws: $n \sim V^{1/3}$ [53, 47, 48, 25], $D \sim V^{1/6}$ (see Methods 3.10.5). Then, by using Equation 3.27 and the constancy of the optimal local wire diameter d across different species [42], we predict that $R_0 \sim V^{4/27}$. This prediction agrees well with the empirically obtained power law relationship (with exponent 1/9) between cortical thickness and brain volume [59, 60, 51, 61, 47]. Thus, our theory explains why the cortical thickness changes little, while brain volume

varies by several orders of magnitude between different species.

Two previous studies [61, 47] also discussed the nature of the scaling law between cortical thickness and brain volume. One study relies on the assumption that the number of neurons in a module of the neocortex is constant. The volume of the module might be cubic R_0. Because the neuronal density may scale inversely as the cubic root of brain volume (see Methods 3.10.5), R_0 should scale as one-ninth of brain volume to ensure that the number of neurons in a module is independent of brain volume. The other study relies on the assumption that the number of such modules scales as two third of the total gray matter volume. Hence, the volume of the module scales as one third of the gray matter volume. As the total cortical gray matter volume may scale linearly with the brain volume (see Methods 3.10.5), the size of the module scales as one ninth of the brain volume. In this chapter, we take a different approach by deriving the expression for the cortical thickness based on the optimization principle. However, we obtain a scaling exponent close to but not exactly equal to one ninth.

3.9.2 Comparison of the Cortical Structure and PD

Neocortex has a sheet-like appearance, and the total area of the gray and white matter boundary is given by $A \sim G/R_0$, where G is the total gray matter volume. According to our theory, such design is optimal when l becomes close to one, which may be the case in big brains. Cortical convolutions may correspond to the geometry expected in the pipe design. However, when $\lambda \ll 1$, our theory predicts that the optimal design satisfies $A \ll G/R_0$. This prediction

does not seem to be consistent with empirical observations from small brains, such as the smooth and sheet-like mouse cortex. It would be interesting to see if different requirements on connectivity or other developmental and/or functional constraints could resolve such discrepancy.

3.9.3 Comparison of Mammalian Neostriatum and PD

Neostriatum is named for its striated appearance (in nissl- and myelin-stained material [62, 63]) caused by axons of neostriatal neurons gathering into fiber fascicles and perforating the gray matter [64]. Areas with higher cell density, or lower global fiber density (myelin-poor region [62, 63]) are called striosomes or patches [65, 66, 67]. As such structure resembles PD, we identify patch size with R_0, Equation 3.27. In a typical rodent (rat or mouse) neostriatum, each principal neuron may locally contact thousands other neurons [64]. Taking $n \sim 10^3$, $d \sim 1$ μm [42] and $D \sim 0.6$ μm [57], we estimate that $R_0 \sim 300$ μm. This estimate agrees well with existing anatomical data [68]. In addition, we may estimate the average axonal fascicle size. Given the total number of neurons in the rat neostriatum is about 10^6 [58], we find that the fascicle diameter is of the same order of ℓ, approximately 100 μm, see Equation 3.58 in Methods 3.10.4. This estimate agrees well with fascicle size [63] (see also, http://www.hms.harvard.edu/research/brain/atlas.html).

3.9.4 Comparison of the Avian Telencephalon and PD

Bird brains also exhibit segregation into gray and white matter and may resemble PD. Distinct fiber fascicles have been identified to connect different

brain regions (see http://avianbrain.org/boundaries.html), such as the connections from HVC to RA in songbirds, which are presumably myelinated axons [69]. Interestingly, unlike the mammals that have a large cortex on the top of other brain structures, the white matter fascicles of the bird could be scattered throughout the whole forebrain. However, more experimental measurements would be desirable, such as measuring large-scale myelin distribution in serial sections of bird telecephalons.

3.9.5 Comparison of the Spinal Cord and PD

While the inner core of the spinal cord contains gray matter, the outer shell contains the white matter consisting of long axons from spinal and cortical neurons [64]. According to our theory, such organization is optimal if the inner core diameter is on the same order of R_0. To see if this is the case, note that a principal (motor) neuron in the spinal cord have a very large arbor span [70, 64] and may receive $10^5 - 10^6$ potential connections. Given $n \sim 10^5$, $d \sim 1$ μm and $D \sim 1$ μm, we find $R_0 \sim 8$ mm according to Equation 3.27, which is on the same order of the inner core diameter [64].

3.9.6 Related Work

Our work builds upon several insights from recent studies. In particular, the idea of minimizing conduction delay has been used to explain why axons and dendrites take certain fraction of the neuropil [32]. The main result in that paper is further extended in this study to show that local conduction delay has to increase after mixing gray and white matter (see Methods 3.10.1-

3.10.3). Also, in our model local circuits are approximated by the network with all-to-all connectivity, which relies on the concept of potential synapses [37]. Adopting this model allowed us to derive explicit results for the total length of local connections (see Methods 3.10.1 and [13]).

We benefited from several previous studies of anatomical and functional connectivity between different cortical areas. These studies helped conceptualize network connectivity by revealing many interesting features of the network [71, 72, 73, 74, 75, 76, 77], such as hierarchal [78], clustering [79] and small-world properties [50, 80], which helped to generate new models to address functional specialization and integration [81, 82, 83, 84, 85, 86].

We adopted (with the potential synapse caveat) the connectivity model used by Ruppin *et al.* [33] and Murre and Sturdy (Murre and Sturdy 1995). These authors applied the wiring optimization approach to explain the segregation of white and gray matter in the brain. Given a network with local and global connections, they searched for design having minimum total wiring volume. They attempted to show that segregated cortex-like design has smaller volume than homogeneous structure by using different approaches.

Murre and Sturdy [34] used the scaling approach to calculate network volume for several network connectivity patterns and layouts. We verified their calculation of the interior (homogeneous) structure volume. However, their calculation of the external (cortex-like) structure volume does not seem to be self-consistent. The volume of axons in the external structure was calculated by using the expression that was unjustifiably adapted from the internal structure calculation, thus undermining their conclusion.

Ruppin *et al.* [33] did not rely on scaling arguments and calculated the

volume of brain structures given their geometric characteristics, under reasonable assumptions of connectivity parameters. These authors showed that segregation of the network into the inner core organization, which has an inner core of gray matter surrounded by white matter, does not lead to volume efficiency compared to homogeneous structure. They also showed that the external sheet (cortex-like) structure has a smaller volume than the inner core organization. However, this does not prove that the cortex-like structure has a smaller volume than the homogeneous structure, a conclusion relying on a fine balance of numerical factors.

We analyze the advantages of gray and white matter segregation from the conduction delay perspective. Our results complement previous studies in some respects but differ in many others. Here, we summarize several novel points. First, we show that the segregation of white and gray matter is consistent with minimizing conduction delay. Second, we determine the maximum number of neurons in the all-to-all connected network with reasonable conduction delay and show that local cortical networks are close to that limit. Third, we propose a possible explanation for the thickness of the neocortex, which varies surprisingly little among mammalian species. Unlike Murre and Sturdy [34] who suggested that cortical thickness is determined by the maximum density of incoming and outgoing global axons (condition indicated by the thin red line in our Figure 3.6), we argue that in most brains it is the result of minimizing local conduction delay. Fourth, our theory is based on the scaling approach and yields a phase diagram of optimal designs for a wide range of parameters. This allows us to apply the theory to several different structures other than the neocortex. Derived scaling relationships can be tested by future

experimental measurements.

3.9.7 Wiring Volume and Conduction Delay Minimization

As features of brain design have been explained by minimizing both the total volume and the conduction time delay, it is natural to wonder how these approaches relate to each other. In general, the evolutionary cost is likely to include both the volume and the time delay. Hopefully, such unified framework will emerge eventually. In the meanwhile, since the exact form of the cost function is not known, we sought to construct theories to explain features of brain architecture based on the simplest possible assumptions. Next, we proposed how time delay and volume can be related based on the current theory.

In our model, conduction delay in local circuits is minimal when the local wire diameter is at its optimal value, which corresponds to an optimum gray matter volume. For details, see section one in the Results. The local conduction delay increases when the local wire diameter d is smaller than the optimum value. In this case, volume cost and conduction delay cost are competing requirements. In the opposite case, when the local wire diameter is thicker than the optimal value, invoking additional conduction delay cost is accompanied by additional volume cost. Therefore, as long as the gray matter volume is greater than its optimal volume, *e.g.*, because of intermixing global axons with gray matter, we may associate the additional conduction delay cost with the volume cost, named the effective volume cost.

However, in the white matter, the relationship between volume and delay is different. Increasing white matter volume by making the global axon diameter thicker does not increase the global conduction delay (see Methods 3.10.2). Thus, the effective volume cost of white matter is just the tissue cost. From this perspective, we propose that gray matter has a greater effective volume cost than white matter. This may have several biological implications:

- Initial segments of axons originating from pyramidal neurons head straight towards (and are perpendicular to) the boundary between the white and gray matter. Once axons cross the white/gray matter border, they change direction. Although such design may increase the length of global axons, it largely reduces the effective volume cost of gray matter because the volume of global axons in the gray matter is minimal.

- Another implication of differential effective volume cost in the gray and white matter is that the global axons in gray matter may be thinner than in white matter. Such variation in diameter could preserve short conduction delays in local and global connection. Of course, global axons cannot be made infinitesimally small without sacrificing global conduction delay. Further exploration of this effect would require more experimental measurements of diameter changes at the white/gray matter border.

- In abutting topographically organized cortical sensory areas, the maps are mirror reflections of each other relative to the border of the areas. The purpose of such organization remains unclear because inter-area connections in the white matter do not benefit from this organization. In particular, placing two cortical areas next to each other (without

mirror reflection) would not increase the length of inter-area connections in the white matter. However, according to our theory, neurons close to the border would be at a disadvantage because their local connections would have to reach further to find appropriate targets. Mirror reflecting maps relative to the inter-area border would eliminate a discontinuity in a map and place neurons with similar receptive fields closer to each other. Such arrangement would benefit intra-cortical connections.

3.10 Methods

3.10.1 Minimization of Conduction Delay in a Local Network with Branching Axon and Dendrite Design

Here we revisit the analysis from [32] using more specific information about the network. Consider wiring up a local network of n neurons with all-to-all potential connectivity. The mean conduction delay in local circuits is given by

$$t = \frac{\ell}{s} \sim \frac{v^{1/3}}{bd^{\theta}}, \tag{3.28}$$

where d is the local wire diameter and $v^{1/3}$, the linear size of the local network, approximates the average path length between two potentially connect two neurons. We assume a sub-linear relationship between local wire diameter and conduction velocity and b is a proportionality coefficient. From Equation 3.28, we want to find the minimal local conduction delay and the corresponding

optimal local network volume. Therefore, we have to eliminate wire diameter d from the previous equation and rewrite it as a function of local network volume. To get this expression, we first notice that the total volume of the local network is given by

$$v = n\chi d^2 + v_n, \tag{3.29}$$

where v_n is the non-wire volume, which is assumed to be a constant, and χ is the total wire length per neuron. Second, for an all-to-all potentially connected network, by applying the branching axon and dendrite design [13], we also have

$$\frac{\chi^2 d}{v} \sim 1. \tag{3.30}$$

This expression is derived as follows [13]. First, the local network volume, v, is divided into cubes of volume, d^3, *i.e.*, into v/d^3 voxels. Then, the number of potential contacts between an axon and a dendrite is given by the number of voxels that contain them both. Each axon occupies χ/d voxels, the same number as a dendrite. The fraction of voxels containing the axon is $\frac{\chi/d}{v/d^3}$, the same as the fraction containing the dendrite. Then, the fraction of voxels containing both the axon and the dendrite is the product of the two fractions, $\chi^2 d^4/v^2$. By multiplying this fraction by the total number of voxels, we find the number of voxels containing axon and dendrite, $\chi^2 d/v$. Then, the condition for having at least one potential contact is given by Equation 3.30. Combining Equation 3.29 with Equation 3.30 and excluding χ yields

$$d \sim \frac{(v - v_n)^{2/3}}{n^{2/3} v^{1/3}}. \tag{3.31}$$

By combining Equation 3.28 and Equation 3.31, we obtain

$$t \sim \frac{n^{2\theta/3} v^{(1+\theta)/3}}{b(v - v_n)^{2\theta/3}}. \tag{3.32}$$

In Equation 3.32, by setting the first derivative of v to zero, we find the optimal network volume, or gray matter volume, should be

$$v \sim \frac{1-\theta}{1+\theta} v_n. \tag{3.33}$$

The minimal local conduction delay is given by

$$t \sim \left(\frac{1+\theta}{1-\theta}\right)^{(1+\theta)/3} \left(\frac{1-\theta}{2\theta}\right)^{2\theta/3} n^{2\theta/3} {v_n}^{(1-\theta)/3}/b. \tag{3.34}$$

We assume that non-wire consists mostly of synaptic components, such as axonal boutons and spine heads. In addition, only a fraction, f (0.1-0.3), of potential synapses are actual synapses [37]. Therefore, the non-wire volume can be estimated as

$$v_n \sim f n^2 v_s, \tag{3.35}$$

where v_s is a single synapse volume. Assuming that $\theta = 0.5$ from classical cable theory and substituting it into Equations 3.34-3.35, we find the minimal local conduction delay is proportional to

$$t \sim f^{1/6} n^{2/3} {v_s}^{1/6}/b. \tag{3.36}$$

For simplicity, after neglecting f, this expression is used in Equation 3.6 in the main text. Furthermore, the optimal wire diameter can also be calculated

by combining Equations 3.31, 3.33 and 3.35, which gives

$$d \sim f^{1/3} {v_s}^{1/3}. \tag{3.37}$$

After neglecting f, this expression also appears in Equation 3.5 in the main text.

3.10.2 Global Conduction Delay Can be Preserved After Inter-mixing Gray and White Matter

After introducing the local connections (gray matter) into the global connections, the total network volume swells and Equation 3.11 of the main text changes to

$$V \sim ND^2 L_a + G, \tag{3.38}$$

where G is the total gray matter volume. After substituting $L_a \sim V^{1/3}$, $D \sim \frac{L_a}{BT}$ (Equations 3.10 and 3.9 in the main text) into Equation 3.38, the expression for V can be rewritten as

$$V \sim \frac{G}{1 - N/(B^2T^2)}. \tag{3.39}$$

After substituting Equation 3.39 into $D \sim \frac{L_a}{BT} \sim D \sim \frac{V^{1/3}}{BT}$, we find the global axon diameter is given by

$$D \sim \frac{G^{1/3}}{[1 - N/(B^2T^2)]^{1/3} BT}. \tag{3.40}$$

Therefore, as long as $T > N^{1/2}/B$, we can find the corresponding global axon diameter D.

3.10.3 Local Conduction Delay Increases After Intermixing Gray and White Matter

Consider again the network described in Methods 3.10.1 with n neurons and all-to-all potential connectivity. After white matter perforates the neuropil, its volume inside gray matter can be expressed by $v\lambda$, where v is the unperturbed optimal local gray matter volume given by Equation 3.33, and λ is a positive dimensionless parameter. After such perturbation, the volume of the local network (Equation 3.29) changes to

$$v' = n\chi d^2 + v_n + v\lambda. \tag{3.41}$$

Second, for an all-to-all potentially connected network, by applying the branching axon and dendrite design [13], Equation 3.30 changes to

$$\chi^2 d/v' \sim 1. \tag{3.42}$$

By combining Equations 3.28, 3.41, 3.42 and excluding χ and d, we can express the local conduction delay as a function of the total local network volume v':

$$t' \sim \frac{n^{2\theta/3} v'^{(1+\theta)/3}}{b[v' - v_n - v\lambda]^{2\theta/3}}. \tag{3.43}$$

Equation 3.43 shows that t' is a monotonically increasing function of λ,

and we recover the expression for t in Equation 3.32 as $\lambda = 0$. Moreover, when $\lambda \ll 1$, the local network is still close to the unperturbed optimal state, *i.e.*, $v' \simeq v$, and we can expand Equation 3.43 to the first order of λ, which yields

$$t' \sim \frac{n^{2\theta/3} v^{(1+\theta)/3}}{b(v - v_n)^{2\theta/3}} \left(1 + \frac{2\theta}{3} \frac{v\lambda}{v - v_n}\right). \tag{3.44}$$

After combining Equation 3.44 with Equations 3.32-3.33, we obtain the expression for local conduction delay from the perturbation theory,

$$t' \sim t[1 + (1 + \theta)\lambda/3], \tag{3.45}$$

or

$$\Delta t/t \sim (1 + \theta)\lambda/3. \tag{3.46}$$

After neglecting the numerical coefficient in the spirit of scaling estimate, the last expression also appears in Equation 3.14 in the main text.

3.10.4 Local Conduction Delay and Surface Area in the Branching Pipe Design

We will address stepwise the process by which we developed this design; first, we present general considerations; second, we develop the first-order branching design; and third, we describe the nonbranching pipes design.

First, to calculate the local conduction delay in the branching pipes, we-consider a general model in which the white matter pipes have total J branching orders. A branch at order k ($0 \leq k \leq J$) has length L_k and pipe diameter

P_k. The total number of kth order branches within the neuropil with linear size L_k is given by M_k. Then, we can evaluate the relative local conduction delay increase through the boundary effect introduced by the kth order branches. The affected neuropil volume through the boundary effect is given by the product of total pipe surface area, $M_k P_k L_k$, and distance ℓ. This means that the ratio of the affected volume to the total gray matter volume, or the relative local conduction delay increase is given by

$$\Delta t/t \sim \ell M_k P_k / {L_k}^2. \tag{3.47}$$

However, Equation 3.47 does not tell what the total local conduction delay is, as different branching orders can have different branching length and diameter.

To examine this further, we assume that the branching structure has a space-filling feature. In particular, the length of the main branch L_0 is given by the linear size of the network, $G^{1/3}$, and the length of $k+1$st order branch is given by the inter-pipe distance among the kth order branches. For the terminal branches $k = J$, the inter-pipe distance between them is given by R_0, Equation 3.20.

If the length of the $k+1$st order branches is much larger than the diameter of the kth order branches, *i.e.*, $L_{k+1} \gg P_k$, the inter-pipe distance between kth order branches is given by ${L_k}/{M_k}^{1/2}$. Thus, we have

$$L_k^2 \sim L_{k+1}^2 M_k, 0 \leq k \leq J, L_{J+1} \sim R_0, \tag{3.48}$$

where L_{J+1} is the inter-pipe distance among the terminal branches, given by

R_0 (Equation 3.20). By denoting N_k as the number of neurons in the neuropil with linear size L_k, N_k and N_{k+1} also have the following relationship

$$N_k \sim N_{k+1} M_k^{3/2}, 0 \leq k \leq J, N_{J+1} \sim \ell^2/D^2, \tag{3.49}$$

according to Equation 3.48, where N_{J+1} is the total neuron number in the neuropil with linear size R_0. In addition, as the pipes with length L_k contains the global axons from the neurons inside the neuropil with linear size L_k, we should also have

$$M_k P_k^2 \sim N_k D^2. \tag{3.50}$$

By substituting Equations 3.48-3.50 into Equation 3.47, we find that

$$\Delta t_k/t \sim \frac{\ell N_{J+1}^{1/2} D M_k^{1/4}}{L_{J+1}^2 \prod_{i=k+1}^{J} M_i^{1/4}} \sim \frac{\lambda M_k^{1/4}}{\prod_{i=k+1}^{J} M_i^{1/4}},\ \Delta t_J/t \sim \lambda M_J^{1/4}, \tag{3.51}$$

where $\ell N_{J+1}^{1/2} D/L_{J+1}^2 \sim \ell^2/R_0^2 \sim \lambda$ because according to Theorem 2, ℓ^2 is the total cross sectional area of the global axons inside the module with size R_0. Then, the total local conduction delay increase through the boundary effect is given by

$$\Delta t/t \sim \sum_{k=0}^{J} \Delta t_k/t \sim \lambda \frac{\sum_{k=0}^{J-1} M_k^{1/4}}{\prod_{i=k+1}^{J} M_i^{1/4}} + \lambda M_J^{1/4}. \tag{3.52}$$

This expression can be minimized as a function of M_k. As a result, we obtain

$$M_0 \sim M_1^2. \tag{3.53}$$

For $J > 1$, we also have

$$M_k \sim \frac{1}{16} M_{k+1}^2, 1 \leq k \leq J-1. \tag{3.54}$$

Given the total number of neurons in the gray matter $N = N_0$ and the total branching orders J, by substituting Equations 3.53-3.54 into Equation 3.49, we can also obtain M_k explicitly. Next, by using Equations 3.48-3.50, we can find the optimal branching length and diameter for different branching orders.

Second, we consider a simple branching model in which only the first order branches exist. In this case, $J = 1$ and by substituting Equation 3.53 into Equation 3.49, we obtain

$$M_1 \sim \left(\frac{ND^2}{\ell^2}\right)^{2/9}. \tag{3.55}$$

By substituting Equations 3.55 and 3.53 into Equation 3.52, the relative local conduction delay increase through the boundary effect is given by

$$\Delta t/t \sim \left(\frac{ND^2}{\ell^2}\right)^{1/18} \lambda, \tag{3.56}$$

where we neglect the numerical factor of the order of one in the spirit of scaling estimate. The total local conduction delay increase is the sum of Equation

3.56 and the expression for relative local conduction delay increase due to intermixing non-fasiculated global axonal segments and gray matter, *i.e.*, λ. However, for the scaling estimate, the second term could be ignored and we obtain Equation 3.23 in the main text.

Next, we calculate the total surface area of the branching pipes A. According to Equation 3.56 and $\Delta t/t \sim \ell A/G$, we then obtain the total surface area of the branching pipes

$$A \sim \left(\frac{ND^2}{\ell^2}\right)^{1/18} \lambda(R_0/\ell)(G/R_0) \sim \left(\frac{ND^2}{\ell^2}\right)^{1/18} \lambda^{1/2}(G/R_0), \tag{3.57}$$

where the last expression uses the relationship $\ell^2/R_0^2 \sim \lambda$. This expression also appears in Equation 3.21 in the main text.

In addition, we can also estimate the diameter and length of the first-order branching pipes. P_1 can be obtained by combining Equations 3.49-3.50, which yields

$$P_1 \sim M_1^{1/4}\ell \sim \left(\frac{ND^2}{\ell^2}\right)^{1/18} \ell. \tag{3.58}$$

According to Equation 3.48, L_1 is given by

$$L_1 \sim M_1^{1/2} R_0 \sim \left(\frac{ND^2}{\ell^2}\right)^{1/9} R_0. \tag{3.59}$$

In the previous analysis, we assume that the length of the first order branches is much larger than the diameter of the main branches, *i.e.*, $L_1 \gg P_0$, which allows us to use Equation 3.48. This assumption holds when the total white matter volume is much smaller than the gray matter volume, *i.e.*, $ND^2 \ll G^{2/3}$.

In the opposite regime, however, $L_1 \ll P_0$ must hold, as the volume of the main branching pipe is much larger than the gray matter volume surrounding it. To see this, note that the volume of the main branching pipe is given by $P_0^2 L_0$, where L_0 is the length of the main branch and the volume of the gray matter surrounding an individual pipe is given by $(P_0 + L_1)^2 L_0 - P_0^2 L_0$. Then, it is easy to check that if the gray matter volume is much larger than the white matter pipe volume, we have $L_1 \gg P_0$, while in the opposite case we have $L_1 \gg P_0$. Geometrically, when $ND^2 \gg G^{2/3}$, the gray matter resembles a sheet and the sheet thickness is given by the length of the first order branches.

As the pipe design exhibits a different configuration when $ND^2 \gg G^{2/3}$, we expect that the expressions for the total surface area of the pipes and the minimal local conduction delay are different from what we derived above. In this case, the total surface area of the main branching pipes is equal to the surface area of the gray matter sheet G/L_1, and the relative local conduction delay increase through the boundary effect of the main branches is given by $\Delta t_0/t \sim \frac{\ell G}{L_1 G} \sim \ell/L_1$.

To calculate the boundary effect induced by the terminal branches, we assume that $R_0 \gg P_1$, where P_1 is the diameter of the terminal branches. This condition allows us to use Equations 3.48-3.50. Later, we will confirm that $R_0 \gg P1$ holds. Then, $L_1 \sim M_1^{1/2} R_0$, $P_1 \sim M_1^{1/4} \ell$, and the relative delay increase due to the terminal branches is given by $\Delta t_1/t \sim \ell P_1 M_1 / {L_1}^2 \sim \ell M_1^{1/4}$. By adding up the local delay from the main and the first order branches, we find that in the regime $ND^2 \gg G^{2/3}$, the total local conduction delay increase

is given by

$$\frac{\Delta t}{t} \sim \frac{\Delta t_0}{t} + \frac{\Delta t_1}{t} \sim \frac{\ell}{M_1^{1/2} R_0} + \lambda M_1^{1/4} \sim \frac{\lambda^{1/2}}{M_1^{1/2}} + \lambda M_1^{1/4}. \tag{3.60}$$

Minimizing this expression as a function of M_1, we obtain $M_1 \sim \lambda^{-2/3}$, and $\Delta t/t \sim \lambda^{5/6}$, as appeared in Equation 3.25.

Next, we calculate the total surface area of the pipes A. As $\Delta t/t \sim \ell A/G \sim \lambda^{5/6}$, we then obtain the total surface area of the branching pipes

$$A \sim \lambda^{5/6}(R_0/\ell)(G/R_0) \sim \lambda^{1/3}(G/R_0), \tag{3.61}$$

as appeared in Equation 3.24 in the main text.

Finally, to check whether $R_0 \gg P_1$, we note that $P_1 \sim M_1^{1/4}\ell$. Then, $R_0 \gg P_1$ requires $R_0 \gg \lambda^{-1/6}\ell$, as $M_1 \sim \lambda^{-2/3}$. In turn, this requires $\lambda \ll 1$ as $\ell/R_0 \sim \lambda^{1/2}$. Thus, $R_0 \gg P_1$ iff $\lambda \ll 1$. This condition should always be satisfied for the PD.

Third, non-branching pipe model corresponds to $J = 0$. It does not belong to the PD because $A \ll G/R_0$ does not always hold in such design when $\lambda \ll 1$, *i.e.*, $ND^2 \ll G/\ell$. To see this, we note that in the regime $ND^2 \gg G^{2/3}$, $A \sim G/R_0$ must hold in the non-branching pipe model, because the pipe diameter P_0 is much larger than R_0. In other words, when $G/\ell \gg ND^2 \gg G^{2/3}$, the gray matter in the non-branching pipe model resembles a sheet with thickness R_0.

3.10.5 Scaling of the Mammalian Neocortex

The theoretical framework developed in this chapter allows us to derive several scaling laws for the neocortex. Provided our perturbation theory is valid, the total neocortical volume G should be proportional to the total non-wire volume. Assuming that non-wire contains mostly by synapses, we have

$$G \sim Nnv_s, \tag{3.62}$$

First, from Equation 3.62, we find that the synaptic density, ρ_s, is a constant, since $\rho_s \sim Nn/G \sim 1/v_s$, where the average synapse volume v_s is assumed to be a constant in different cortical areas and across different species. The prediction of constant synapse density is supported by experimental observations [87, 88, 42, 48], from a small number of taxa so far, and was used as a starting point to derive scaling laws of the mammalian brains in several theoretical papers [51, 47].

Second, we find the neuronal density $\rho \sim N/G \sim \frac{N}{Nnv_s} \sim 1/n$. Since ρ scales inversely as the cubic root of total brain volume V across different mammalian species [89, 48] ($\rho \sim V^{-1/3}$) and the cortical volume is loosely proportional to the brain volume [90] ($G \sim V$), we find $n \sim V^{1/3}$, $N \sim V^{2/3}$, and $n \sim N^{1/2}$. We note that Braintenberg [42, 53] has previously proposed the square-root relationship between n and N. He assumed that the cerebral cortex could be divided into $N^{1/2}$ compartments and each compartment contains $N^{1/2}$ neurons. The local connectivity within a compartment is almost all-to-all and every compartment is connected to every other one by a global axon.

Third, we find that the global axon diameter D scales as $V^{1/6}$. To see this, we note that the total white matter volume W is given by $ND^2V^{1/3}$, where the average length of global axons in the white matter is assumed to be proportional to the brain size, $V^{1/3}$. Since $N \sim V^{2/3}$, and it has also been reported that $W \sim V^{4/3}$ across different mammalian species [91, 90, 22, 47, 92], we find $D \sim V^{1/6}$. This is consistent with recent measurements from corpus collusum, which indicates that the average diameter of global axons scales monotonically with the brain size [48]. Then, using $n \sim V^{1/3}$, $D \sim V^{1/6}$ and Equation (27), we obtain $R_0 \sim V^{4/27}$, an expression from Discussion 3.9.1.

Chapter 4

A Cost-benefit Analysis of Neuronal Morphology

4.1 background

Structure-function relationships have long played an important role in biology. In addition to describing the bewildering variety of axonal and dendritic arbor shapes, Cajal has inferred that the function of dendrites and axons was to conduct electrical signals from post-synaptic terminals to the integration site, which often is the cell body, and from the integration site to the pre-synaptic terminals, respectively [1]. Moreover, he speculated qualitatively that the structure of axons and dendrites minimizes their cost for given functional constraints [1]. Yet, a quantitative theory of axonal and dendritic shape and dimensions, which could provide valuable insights into to their function, is still missing [93, 94, 95, 96].

In some cases, axonal dimensions can be trivially explained by the re-

quirement of making specific connections [64]. In particular, global axons projecting over long distances have to be long enough to reach from the pre-synaptic cell body to the post-synaptic target(s). For example, an axon of a pyramidal neuron projecting from one cortical area to another must be as long as the distance between those areas. Even local axons sometimes make specific connections, which determine their shape. For example, the shape of the climbing fiber in the cerebellum must match the shape of the Purkinje cell dendrites so that the axonal arbor can make numerous contacts with a single dendrite. Although dendrites do not implement long-range projections (gustatory neurons being an exception), they may have specific local targets. For example, apical dendrites of particular pyramidal neurons arborize only in certain cortical layers.

In many cases, however, the requirement to make specific connections does not fully determine arbor dimensions. Consider, for example, local axonal and basal dendritic arbors of two pyramidal neurons belonging to the same cortical column (*i.e.*, within a few hundred microns). Even without correlations in the layout of axonal and dendritic branches [56], these overlapping arbors make a potential synapse with each other (*i.e.*, come within a spine length of each other, allowing them to form an actual synapse by growing a spine) [38, 97]. Then, what determines the topology and dimensions of dendritic and axonal branches of pyramidal neurons?

Another example is the Purkinje dendrite, which establishes potential synapses with most parallel fibers (granule cell axons) that course through its volume. There is no specificity in the topology and dimensions of the dendritic arbor. Similarly, parallel fiber can contact any of the Purkinje dendrites it

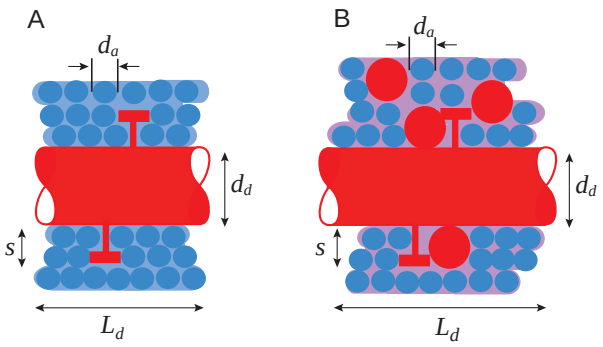

Figure 4.1: Schematic illustration of two designs of the neuropil micro-architecture. Axons (blue) make potential synapses with a dendritic segment (red) if they pass through a region within a spine length s of that segment. We call this region a spine-reach zone and compare two designs with the same number of potential synapses. In (A) spine-reach zone contains axons but excludes dendrites of other neurons. d_a - axon diameter, d_d - dendritic diameter, L_d - dendritic length. In (B), dendrites from various neurons interpenetrate each other's spine-reach zone. As a result, they add to the excluded volume of axons and increase the total length of dendrites L_d.

trespasses. What determines the topology and dimensions of these axons and dendrites?

Previously, we have shown that the total dendritic length is determined by the requirement to form a given number of potential synapses [13]. All axons that could potentially synapse on a dendrite must pass within a spine length of a dendrite, Figure 4.1A. Because of volume exclusion by these axons, the dendritic length is approximately given by the number of potential axons times the axonal cross-sectional area divided by the spine length. Similar argument applies to the total length of axons, which is given by the number of potential dendritic targets times the dendrite cross-sectional area divided by the spine length. Therefore, the total length of axons and dendrites is minimized for given potential connectivity.

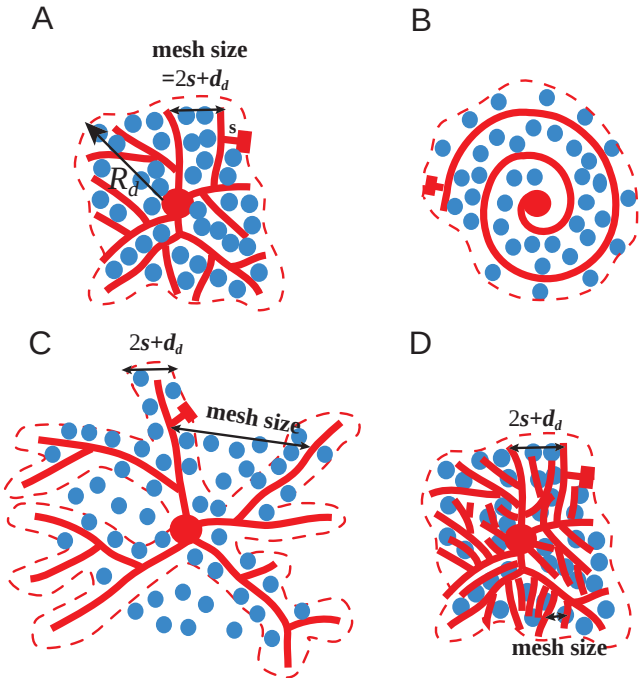

Figure 4.2: Schematic illustration of four dendritic arbor designs. Dendrites shown in red could be either planar or 2D projections of 3D arbors onto the plane perpendicular to axons (blue). We consider four designs with the same number of potential axonal targets. Dashed line indicates the spine-reach zone of an arbor. s - spine length, d_d - dendritic diameter, R_d - arbor span. (A) Compact branching arbor. A compact arbor makes on average one potential synapse with each axon (blue) passing through the arbor. The mesh size is defined as the arbor span area divided by the total dendritic length. The mesh size of a planar compact arbor is $2s + d_d$. The mesh size of a 3D compact arbor differs from the 2D arbor by a numerical factor. (B) Compact non-branching arbor. This arbor has the same total dendritic length and the same mesh size as the compact branching arbor but greater path length. (C) Sparse branching arbor. A sparse branching arbor cannot make potential synapse with each axon passing through the arbor because the area of spine-reach zone is smaller than arbor span area. In other words, the mesh size of a sparse arbor is much larger than $2s + d_d$. (D) Dense branching arbor. A dense branching arbor makes more than one potential synapses with each axon passing through the arbor. The mesh size of a dense arbor is much smaller than $2s + d_d$.

The above argument accounts for the total length of dendrites and axons but does not explain why dendrites and axons branch. For example, consider two alternative designs for a dendritic arbor, Figure 4.2A and Figure 4.2B modified from [34]. For simplicity we assume that these arbors are planar and they must receive inputs from a bundle of axons running perpendicular to the dendritic plane (like parallel fibers in the cerebellum). Both arbors have the same total length and the same number of potential axons. Therefore, their total wire length cost and potential connectivity are the same. Then, why has not the arbor in Figure 4.2B been observed?

One possible answer is that branching plays a computational role in axons [98, 99, 100, 101], and, in particular, dendrites [40, 102, 103, 104]. For example, specific interactions between excitatory and inhibitory inputs on different dendritic branches and their combinations may be used to construct logical gates [105]. Having a large number of dendritic branches may also increase the information storage capacity of a neuron, presumably due to the nonlinear integration of synaptic inputs from different branching units [106]. However, a quantitative theory of dendritic branching based on computational requirements does not exist.

In this chapter, we sought to explain the branching structure of dendrites by optimizing not just the total wiring length, but a combination of the total wiring volume and the cost of signal transmission between the soma and synapses [107, 108]. Previously, a combination of volume and signal transmission cost has been used to explain other features of brain design [32, 14, 109]. We quantify the cost of signal transmission by signal attenuation along a dendrite. Optimization is performed under the neuronal connectivity constraint,

by fixing the potential convergence factor, which is the number of neurons making potential synapses with a dendrite.

The chapter is organized as follows. First, we argue that a compact and branching dendrite minimizes a combination of volume and attenuation cost. Second, we show that the optimal dendrite is planar, perpendicular to the orientation of axons, and does not admit other dendrites within its spine length. This design is adopted by the Purkinje dendrites in the cerebellum. Third, we point out that sparseness of cortical and hippocampal dendrites cannot be explained solely by minimizing cost.

4.2 Quantitative Expression for the Cost of Dendrites

Optimization-based explanation of dendritic shape requires a quantitative expression for the cost of dendrites. We adopt the cost function, used previously for long-range axons [14], which is the sum of dendritic volume, V_d, and signal attenuation, T_d,

$$E_d = V_d + \alpha T_d, \tag{4.1}$$

where α is an unknown constant, which will be determined later by comparing our predictions with anatomical data (see Discussion). In this section, we recap the argument that led to this cost function and express volume and attenuation in terms of dendritic dimensions. The resulting expression will be minimized in subsequent sections.

Based on previous experimental and theoretical results suggesting that neuropil volume is a costly resource [13, 110, 34, 111, 112, 12, 11, 113, 114, 32], we use volume to quantify the cost of dendrites. If we model a dendritic segment as a solid cylinder, the volume of a dendrite can be straightforwardly expressed in terms of its total length, L_d, and its mean diameter, d_d,

$$V_d = \frac{\pi}{4} L_d d_d^2. \tag{4.2}$$

Here and below, if we calculate a quantity exactly in a certain model by including all the numerical factors, we use the equal sign (*e.g.*, $A = B$). If we ignore all the numerical coefficients, we use the symbol $\sim$ (*e.g.*, $A \sim B$).

Although the volume of dendrites can be minimized by making dendrites thinner, this would negatively impact their function: thinner dendrites attenuate synaptic signals more because of greater leakage current. The loss of signal strength is captured by the attenuation cost T_d proportional to the fractional attenuation of the signal.

To derive the expression for the attenuation cost T_d, first, we assume that the mean path length from a synapse to the soma, ℓ, is much smaller than the space constant ξ. This condition is probably satisfied for local dendritic arbors such as Purkinje dendrites [115] and basal pyramidal dendrites [116, 117]. In this case, the fractional attenuation of the voltage is given by ℓ/ξ. Second, by taking into account the fact that a dendrite integrates many inputs, we posit that the attenuation cost is the sum of fractional attenuation from different synapses. Assuming that the number of actual synapses converging

on a dendrite is proportional to the potential convergence factor C_p, we have

$$T_d \sim C_p \frac{\ell}{b d_d^{\theta}}, \tag{4.3}$$

where we substitute the relation $\xi = b d_d^{\theta}$ and b is a constant coefficient. In the following, we assume $\theta = 0.5$ as used in deriving the cable equation [118].

Although active conductances, which are present along a dendrite, could compensate for passive attenuation, they must invoke metabolic costs of their own. To the first approximation, these costs are commensurate with the passive attenuation.

Combining Equations 4.2 and 4.3 we can re-write Equation 4.1 as:

$$E_d \sim d_d^2 L_d + \alpha C_p \frac{\ell}{b d_d^{1/2}}. \tag{4.4}$$

4.3 Minimization of Volume and Attenuation Yields Compact Centripetal Dendrite

According to Equation 4.4, for given convergence, the cost of dendrites depends on three variables: mean path length, ℓ, total length of dendrites, L_d, and mean dendritic diameter, d_d. These variables are not independent. In this section, by sequentially re-expressing these variables, we reduce the cost expression to the function of a single variable, dendritic diameter. This function will be minimized to find the optimal dendritic diameter.

First, according to Equation 4.4, to minimize the cost, the typical path should be as short as possible. But the mean path length ℓ cannot be smaller

than the linear size or span of the arbor, R_d, at least by order of magnitude:

$$\ell \sim R_d. \tag{4.5}$$

To minimize the path length, and satisfy Equation 4.5, each segment of the dendrite must be directed towards the integration site, which is often the cell body. We call such arbor design centripetal. If the total dendritic length is greater than the dendritic arbor span, the centripetal arbor must branch, Figure 2A. Therefore, branching of dendrites is a trivial consequence of minimizing mean path length.

Using Equation 4.5, we can rewrite the cost of a centripetal dendrite as a function of arbor size R_d:

$$E_d \sim d_d^2 L_d + \alpha C_p \frac{R_d}{b d_d^{1/2}}. \tag{4.6}$$

The cost can be minimized by reducing arbor size, R_d, under constraint of the potential convergence factor C_p. To satisfy this constraint, an arbor must span an area containing C_p axons. If dendritic arbor is planar and axons run orthogonally to it, the minimum arbor span area

$$\pi R_d^2 = \frac{\pi}{4} C d_a^2 + d_d L_d, \tag{4.7}$$

where $\pi/4 C_p d_a^2$ is the total axonal cross-sectional area and $d_d L_d$ is the area occupied by the dendritic branches themselves. Equation 4.7 still holds if dendritic arbor is not planar but 3D, provided that numerical factors of order one are ignored.

By substituting Equation 4.7 into Equation 4.6, we find

$$E_d \sim d_d^2 L_d + \alpha C_p \frac{(C_p d_a^2 + d_d L_d)^{1/2}}{b d_d^{1/2}}. \tag{4.8}$$

Note that reducing L_d minimizes cost of dendrites. The minimum dendritic length is again determined by the potential connectivity constraint: C_p axons must fit within a spine length s of a dendrite, which we call the spine-reach zone, Figure 4.1A. Then, for a planar dendritic arbor, we set the area of the spine reach zone equal to the total axonal cross-section [13] and find that the minimum total dendritic length is:

$$L_d = \frac{\pi}{4} C_p \frac{d_a^2}{2s}, \tag{4.9}$$

where factor 2 reflects the fact that the spine reach zone exists on both sides of a dendrite. Equation 4.9 also holds for a 3D dendritic arbor provided numerical factors of order one are ignored.

Next, we calculate the arbor mesh size – a parameter that quantifies the sparseness of an arbor - by dividing arbor area by the total length, R_d^2/L. For the optimal arbor, by combining Equations 4.7 and 4.9, we find

$$\frac{\pi R_d^2}{L_d} = 2s + d_d. \tag{4.10}$$

Equation 4.10 also holds for a 3D dendrite provided that numerical factors of order one are ignored.We call a planar or 3D arbor satisfying Equation 4.10 compact, Figure 4.2A. One property of a compact arbor is that it forms on average one potential synapse with each axon passing through the arbor.

Compact branching arbor is less costly than other branching arbors with the same potential convergence. Consider a sparse arbor, the mesh size in which is much larger than $2s + d_d$, Figure 4.2C, and which does not form potential synapses with every axon passing through the arbor, Figure 4.2C. A compact arbor is less costly because it has smaller span than a sparse branching arbor.

Compact branching arbor is also advantageous to a dense branching arbor, in which the arbor mesh size is much smaller than $2s + d_d$, Figure 4.2D. A dense branching arbor can form more than one potential synapse with each axon passing through the arbor, Figure 4.2D. Given the same number of axons forming potential synapses with the dendrite, such design makes the total dendritic length greater than that in the compact arbor.

Finally, we express the cost of dendrites in terms of dendritic diameter. By substituting Equation 4.9 into Equation 4.7, we find that the minimum arbor size is given by

$$R_d = \frac{1}{2} C_p^{1/2} d_a (1 + \frac{d_d}{2s})^{1/2}. \tag{4.11}$$

By substituting Equations 4.11 and 4.9 into Equation 4.6, E_d is given by

$$E_d \sim \frac{C_p d_a^2 d_d^2}{s} + \alpha \frac{C_p^{3/2} d_a}{b s^{1/2}} (1 + s/d_d)^{1/2}. \tag{4.12}$$

Now we can search for the optimal dendritic diameter that minimizes the cost of a dendrite, E_d. According to Equation 4.12, the volume and attenuation have opposite dependence on the dendritic diameter and the competition between them determines the optimal diameter. By setting $\partial E_d / \partial d_d = 0$, we

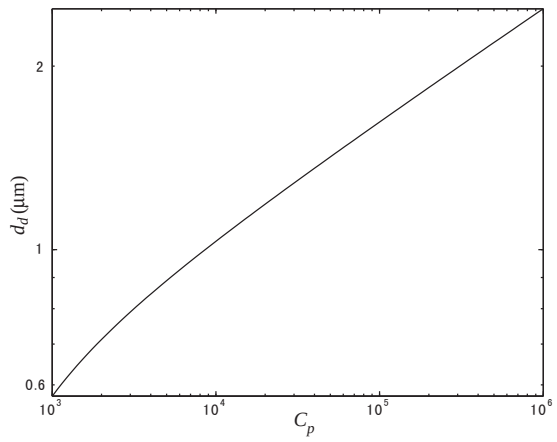

Figure 4.3: Optimal dendritic diameter as a function of the potential convergence factor. dendritic diameter is calculated for specific values of parameters ($d_a = 0.2$ μm [119], $s = 1.4$ μm [120], and $\alpha/b = 3 \times 10^{-3}$ μm$^{2.5}$) and is plotted in the log-log coordinates.

find that the optimal dendritic diameter satisfies the following relationship

$$d_d^6 + s d_d^5 \sim (\alpha/b)^2 s^3 C_p / d_a^2. \tag{4.13}$$

According to Equation 4.13, we plot numerically the optimal dendritic diameter as a function of the convergence factor C_p in Figure 4.3. Note that the explicit expression for d_d can be easily obtained in two limiting cases: $d_d \ll s$ and $d_d \gg s$. When $d_d \ll s$, $d_d \sim [\alpha s/(b d_a)]^{2/5} C_p^{1/5}$. In the opposite limit, $d_d \sim \left[\alpha s^{3/2}/(b d_a)\right]^{1/3} C_p^{1/6}$.

So far, we have argued that the cost of planar and 3D arbors was the same up to numerical factors of order one. When these numerical factors are included, a planar dendrite has a smaller cost than a 3D dendrite. To see this,

note that a planar arbor can be viewed as a 2D projection of the 3D arbor onto the plane orthogonal to the axons and the projection is always shorter than the original. Then, both the minimum path length, ℓ, and the minimum total dendritic length, L_d, in a planar arbor are shorter than those in a 3D arbor. According to Equation 4.4, the cost of dendrites increases with ℓ and L_d for any given d_d, and a planar arbor is advantageous to a non-planar arbor.

Moreover, the minimum L_d is achieved by a spatial organization of the neuropil, in which adjacent dendrites from different neurons are excluded from each other's spine-reach zone, Figure 4.1A. If dendrites penetrate each other's spine-reach zones, Figure 4.1B, they would add to the excluded volume of axons, Equation 4.9, and increase the total dendritic length.

4.4 Discussion

In this chapter, we sought to understand the branching structure of the neurons as a trade off between functional benefits and physical costs. For a given potential connectivity of the neuronal network, we minimized the combination of volume and attenuation cost in dendrites. Next, we compare our predictions with the experimental data.

4.4.1 Purkinje Dendritic Arbor is Compact, Centripetal and Planar

In order to prove that Purkinje dendritic arbors are compact, we show that the dendrite forms on average one potential synapse with each axon passing

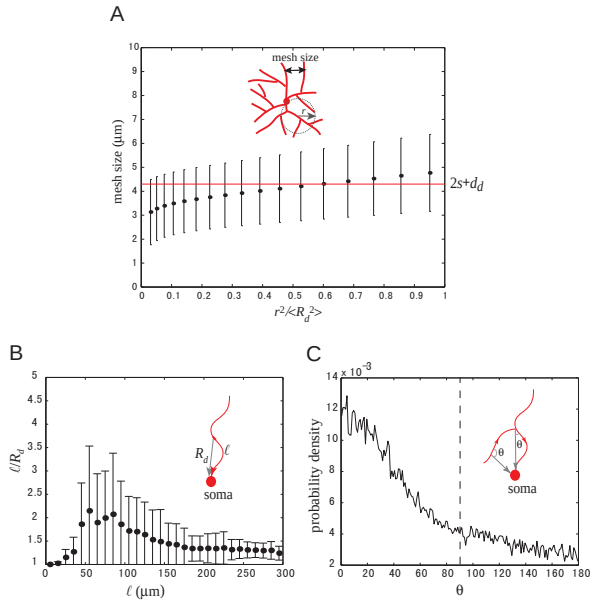

Figure 4.4: Purkinje dendrites are compact and centripetal. (A) Arbor mesh size of Purkinje dendrites. The red line is the predicted mesh size for a compact arbor, where $s = 1.4$ μm [120], $d_d = 1.5$ μm [120] are used. To measure the mesh size, first, we project the cells onto the plane perpendicular to the parallel fibers (sagittal plane). Second, we calculate the mean square radius of the arbor $\langle R_d^2 \rangle = \left(\sum_{i=1}^{N} R_i^2 \right) / N$, which is defined as the mean square distance from the dendritic segment to the centroid of the dendrite. Third, centered on a dendritic segment whose distance from the centroid of the arbor is less than $\sqrt{\langle R_d^2 \rangle}$, we draw circles and calculate the mesh size by dividing the area of the circle by the total dendritic length within the circle. Fourth, we plot the mean mesh size, which is averaged over different centers of the circles and different cells, as a function of the area of the circle r^2 normalized by the mean square radius $\langle R_d^2 \rangle$. (B) Ratio of the path length a dendritic segment to the soma, ℓ, to the euclidian distance from the dendritic segment to the soma, R_d. The ratio is close to one for a given path length. (C) The probability distribution of the orientation angle θ between the vector of the signal flow along a dendritic segment and the vector pointing centripetally from the segment to the soma. 70 percent of the angles are less than ninety degree. The above analysis is performed on ten digitally reconstructed Purkinje dendrites [121, 115] (publicly available from http://neuromorpho.org). All error bars are standard deviations.

through the arbor. According to Equation 4.10, this requirement is achieved when the mesh size of a planar arbor is $2s + d_d$.

We measure the mesh size of different parts of a dendrite by calculating the area of a local tree divided by its total length. In Figure 4.4A, we plot the mesh size as a function of the area of a local tree, where error bars are standard deviations (see figure captions for detailed explanation). The mesh size is reasonably close to the theoretical prediction (red line in Figure 4.4A), where we use $s = 1.4$ μm [120], $d_d = 1.5$ μm [120]. Therefore, our measurement confirms that Purkinje dendrites are compact.

To demonstrate that Purkinje dendrites are centripetal, first, we calculate the ratio of the path length from a dendritic segment to the soma, ℓ, to the euclidian distance from the dendritic segment to the soma R_d. Data presented in Figure 4.4B show that this ratio is close to one. Second, we plot the probability distribution of the orientation angle θ between the vector of the signal flow along a dendritic segment and the vector pointing centripetally from the dendritic segment to the soma (Figure 4.4C). We find that 70 percent of the angles are less than 90 degree. Taken together, these measurements suggest that the Purkinje dendrites are centripetal.

Moreover, we suggest that the characteristic planar shape of Purkinje dendrites can be explained as a result of minimizing the cost of dendrites. It would be interesting to see if the same principle also applies to other types of planar dendrites in the brain.

According to Figure 4.3 and by using the relationship $L \sim C_p$ in Equation 4.9, we also predict that the mean dendritic diameter increases with the total dendritic length. We hope that this prediction can be tested quantitatively

for Purkinje dendrites.

We can determine the unknown constant α in the cost function, Equation 4.1, by substituting the anatomical and physiological parameters of Purkinje dendrites into the expression for d_d in Equation 4.13. Using $C_p = 10^5$ [120], $s = 1.4$ μm [120], $d_d = 1.5$ μm [120] $d_a = 0.2$ μm [119] and $b = \sqrt{\frac{R_m}{2R_i}} = 2 \times 10^3$ $\mu m^{1/2}$ [115] (R_m is the specific membrane resistance and R_i is the intracellular resistitivity), we find $\alpha \sim 1$ μm^3.

4.4.2 Micro-architecture of the Cerebellum Molecular Layer

According to the Results section, to minimize the cost of dendrites, adjacent dendrites should be mutually exclusive, Figure 4.1A. We verify this prediction by estimating whether interval η between potential synapses along a parallel fiber in the molecular layer satisfies $\eta = 2s + d_d$. We note that a similar calculation has been performed previously by Napper and Harvey [119].

The interval between potential synapses along a parallel fiber $\eta = L_a/D_p$, where L_a is the length of a parallel fiber and D_p is the potential divergence. D_p can be calculated as follows. Because Purkinje dendrites are compact, an axon can potentially connect with all the dendrites in the volume $L_a wh$, where w is the width of the dendritic arbor and h is the height of the arbor. Therefore, we have

$$D_p = \rho L_a wh, \tag{4.14}$$

where ρ is the neuronal density. Because Purkinje cell bodies are arranged in monotone layer, we may rewrite Equation 4.14 as a function of the neuronal

density per unit area σ,

$$D_p = \sigma L_a w. \tag{4.15}$$

As a result, the interval of potential synapses on an axon η is given by

$$\eta = L_a / D_p = \frac{1}{\sigma w}. \tag{4.16}$$

By substituting the values from the rat cerebellum [119] $\sigma = 1018$ mm^{-2}, $w = 250$ μm, we obtain $\eta = 4$ μm. Recalling that $s = 1.4$ μm [120], $d_d = 1.5$ μm [120], we find that the relation $\eta = 2s + d_d$ is satisfied and adjacent Purkinje dendrites are on average excluded from each other's spine-reach zone. We hope that in the future this calculation will be verified directly by electron microscopic reconstructions.

4.4.3 Arbor Shape in the Neocortex and Hippocampus

According to the theory, the optimal dendritic shape is a compact arbor. By substituting $d_d = 0.9$ μm [42], $s = 2$ μm [42] into Equation 4.10 we find the mesh size of a pyramidal neuron dendrite should be several microns. However, in reality, pyramidal dendrites are much sparser [42]: the mesh size $\sim 20 - 30$ μm for cortical layer II-III basal dendrites of pyramidal cells [122]. Therefore, the sparseness of cortical pyramidal dendrites cannot be understood within the current framework.

How can such a discrepancy between the theory and the experiments be resolved? This question will be answered in the next chapter based on the hypothesis dendritic arbors not only minimize their cost but also maximize

the number of achievable connectivity patterns. We will show that sparse dendritic arbors can achieve more connectivity patterns than compact arbors. We postulate that the sparseness of pyramidal dendrites may be a consequence of great structual plasticity in the neocortex.

Chapter 5

A Statistical Theory of Dendritic Morphology

5.1 background

In his seminal work [1], Santiago Ramón y Cajal not only described the bewildering variety of neuronal shapes, but also attempted to relate neuronal structure and function. He proposed the law of polarization, stating that the function of dendrites is to conduct signals from synapses to the soma, while the function of axons is to conduct signals from the soma to synapses. Although some exceptions to this are now known, notably the dendritic backpropagation, the law mostly passes the test of time. Cajal also proposed that neuronal shape maximizes functionality while minimizing the use of available resources, such as space time and energy. However, because of a lack of anatomical data and quantitative formulation, the hypothesis has been difficult to test until recently.

In the last two decades, proliferation of reliable staining methods and improvements in imaging and computing infrastructures have led to the growth of neuronal shape datasets accompanied by various attempts to quantify arbor shape. One important tool in characterizing arbor shape has been the Sholl plot, which yields the number of branches intersecting spheres with incremental radii [123, 107, 124]. Another important observation was the statistical scaling of arbor shape [125, 126, 127, 128, 129, 130, 131].

In this chapter, by analyzing several datasets of neuronal shape, we confirm and expand the observation of scaling in dendritic arbors. We demonstrate that the Sholl plot can be fitted by an analytic function which is approximated by a power law at small distances. This implies that a fragment of an arbor, in a statistical sense, looks like a scaled down version of the whole arbor. In other words, the arbor is self-similar. We also report scaling of the arbor span with the total dendritic length over a wide range of arbor sizes. Most importantly, we show that the two observations of scaling are related, implying that arbors are not only self-similar but are built according to common rules. In addition, we measure the tortuosity and the branching of the arbors.

What is the theoretical explanation of scaling and self-similarity? As arbor shape is an outcome of neural development, developmental models could provide such an explanation [132, 133, 134, 135]. However, because of the complexity of development, a systematic theoretical framework explaining neuronal arbor shape does not exist. Here we pursue a different approach, in which brain structure is explained as a result of evolutionary fitness optimization that maximizes brain functionality, while minimizing costs associated with building and maintenance.

Much progress in explaining neuron structure based on fitness optimization has been made by minimizing cost, rather than maximizing functionality. This is because costs have a clear physical origin and are easier to quantify. In particular, building on the qualitative suggestions of Cajal that neuronal shape minimizes costs, we have been able to explain many brain structures as discussed in the previous chapters.

In this chapter, we propose a theory for the shape of dendritic arbors by quantifying not only their costs but also their functionality. As the principal role of dendrites is to make synaptic connections with axons, we quantify their functionality by the number of achievable connectivity patterns. Thus we maximize the connectivity repertoire while minimizing cost. We can, with all significant limitations in mind, make some calculations based on this principle and arrive at the results which are imperfect in the sense that critical exponents are not quite in agreement with the data, but the very fact of scaling is definitely recovered.

Maximization of the connectivity repertoire and minimization of the cost is closely related to the maximum information storage capacity approach that can explain the properties of cortical synapses [136]. Here, we take this approach to the next logical step by considering all the possible trajectories of dendritic branches through the neuropil. We calculate the total number of available connectivity patterns by taking advantage of powerful methods from statistical physics developed for branching polymers [137, 138].

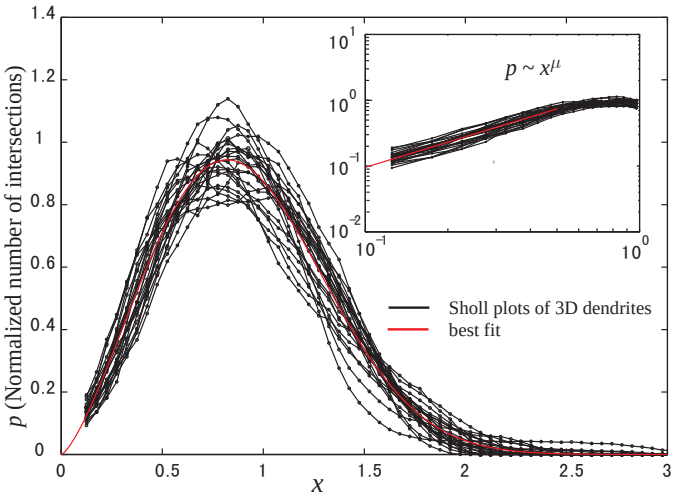

Figure 5.1: Generalized Sholl plot of 3D pyramidal basal dendritic arbors. For each dendritic arbor, the number of branches intersecting a sphere is plotted as a function of the sphere radius. The sphere radius was normalized by the dendritic arbor span and the number of intersections was also normalized so that the total dendritic length is equal to unity (see Methods for detailed description). The Sholl plots for different dendritic arbors are combined and best fitted by the function $p = 2.196x^{1.375}\exp(-0.92x^{2.398})$. The inset shows the head of the Sholl plots in log-log coordinates, which can be fitted by a power law with the exponent $\mu = 1.273 \pm 0.081$.

5.2 Scaling Laws of Dendrites

First, we revisit the Sholl plot [123] that has been the main tool for dendritic shape analysis in the last fifty years. A 3D Sholl plot gives the number of branches intersecting spheres of different radii. We generalized the original scheme so that the center of the sphere does not have to be the soma, but can also be other dendritic segments (see Methods for a detailed description). We combine generalized Sholl plots for different pyramidal basal dendritic arbors by normalizing the sphere radii by the dendritic arbor span and the total dendritic length to one. The plot can be fitted nicely with a function of the form $\sim x^{\mu}\exp(-x^{\delta})$ (Figure 5.1). The head of the Sholl plot can be fitted by a power law with exponent $\mu = 1.273 \pm 0.081$ as demonstrated on the log-log scale (Figure 5.1). This implies that, in a statistical sense, a fragment of an arbor looks like a scaled down version of the whole arbor, *i.e.*, the arbor is self-similar.

Another important relationship, seemingly unrelated to the Sholl plot, is the dependence of dendritic span on the total dendritic length. Because existing 3D datasets cover only a limited range, we turn to a collection of basal pyramidal dendrites drawn on 2D from various cortical areas. Using the 2D dataset we show that dendritic arbor span scales as a power of total dendritic length and the exponent $\nu = 0.445 \pm 0.010$ (Figure 5.2).

If arbors are built according to common rules, the statistics inside one cell and across cells must follow a scaling relation. Details of the derivation

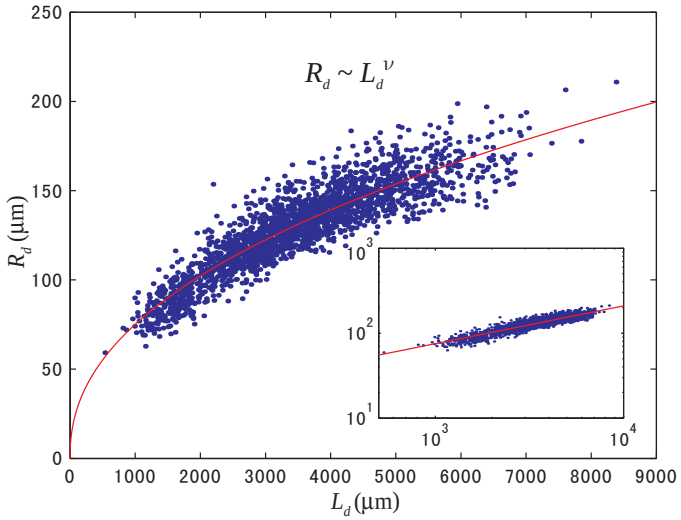

Figure 5.2: Dependence of the dendritic arbor span on the total dendritic length. The red curve is the power law fit with exponent $\nu = 0.445 \pm 0.010$. Inset is the same plot in the log-log coordinates.

are presented in the Methods, where we show that

$$\mu = 1/\nu - 1. \tag{5.1}$$

The measured exponents satisfy approximately Equation 5.1.

Therefore, neurons seem to obey a scaling law in the sense that a part of a big cell looks statistically similar to an entire smaller neuron. This is a non-trivial result as many biological objects scale differently: the bigger neuron could be just a scaled up version of the smaller neuron.

To verify that our 2D and 3D datasets are consistent, we compare 2D Sholl plots from the two datasets and show that they accord with each other (Figure 5.3).

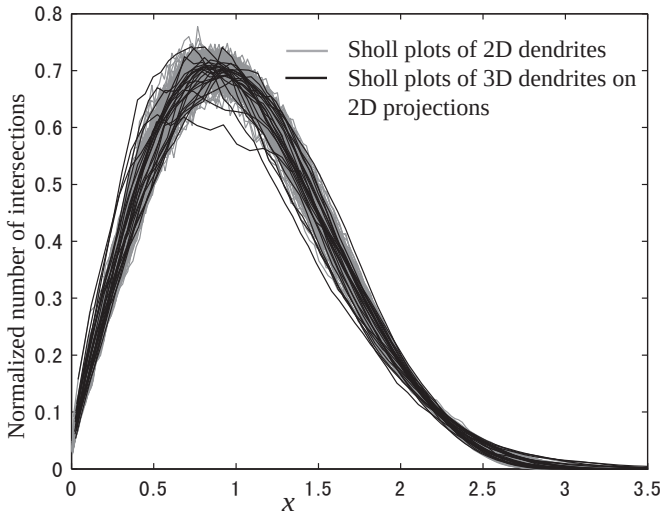

Figure 5.3: Comparison of the generalized Sholl plots between two datasets. Gray curves are generalized Sholl plots of basal dendritic arbors drawn on the 2D plane parallel to the cortical surface. Black curves are generalized Sholl plots of 3D basal dendritic arbors projecting onto the plane parallel to the cortical surface. The generalized Sholl plots of the two datasets are consistent with each other.

5.3 Maximizing connectivity repertoire under the wiring cost constraint explains scaling law

In this section, we propose that the scaling laws could follow from maximizing functionality of dendrites while minimizing their cost. We will summarize the key steps in formulating the theory. The detailed derivations are presented in the Methods.

First, we quantify the cost of dendrites. Because dendrites take up valuable space they must invoke costs proportional to their total length. Another contribution to the cost of dendrites is the attenuation of synaptic currents from synapses to soma, which increases with longer path length and smaller dendritic diameter. By calculating the optimal dendritic diameter and substituting it into the cost expression, we find that the cost can be approximated as a product of total dendritic length and path length.

Second, we consider dendritic functionality assuming that the principal role of dendritic arbors is to implement connections with axons of other neurons. Then the functionality of dendrites can be quantified by the number of possible connectivity patterns that could be achieved, or in other words, a connectivity repertoire.

How can we estimate the number of different connectivity patterns of dendritic arbors? Our calculation relies on the following assumptions.

- Individual neuron acts as a single unit. Thus an actual connection represents a pre- post-synaptic neuron pair irrespective of the number of

actual synapses made between them.

- Appropriate axonal inputs of a dendritic arbor are randomly distributed in space. This assumption is in contrast to the specious alternative, which states that the appropriate axonal inputs form specific spatial patterns and determine the dendritic arbor shape. Such hypothesis is unrealistic because synapses are made every half micron on a dendrite and an entire arbor receives $\sim 10^4$ synapses [42]. It is unlikely that all the appropriate inputs of a dendritic arbor happen to align into lines and arrange into a tree-like spatial pattern. Even if this could be achieved for a single arbor, it is impossible to arrange so many inputs specifically and simultaneously for thousands of neurons in the neuropil.

- A dendritic arbor is sparse so that when an axonal arbor field overlap with the dendritic arbor field, the probability of establishing a potential synapse is small (*i.e.*, axonal and dendritic segments come within a spine length of each other, allowing them to form an actual synapse by growing a spine). This assumption implies that the majority of potentially pre-synaptic axons establish a single potential synapse with the dendrite, which has been tested experimentally for excitatory cortical neurons. For example, in cat V1 the fraction of potentially pre- post-synaptic neuron pairs connected through two potential synapses is estimated to be less than 10 percent (based on data from Stepanyants et. al. [139]). Higher order potential connections occur even less frequently and will be disregarded.

- Dendritic arbors are characterized as flexible chains jointed by rigid seg-

ments, and the length of a rigid segment is much longer than the inter-synapse interval on a dendrite. This assumption should be tested experimentally in the future, for example, through measuring the stiffness of dendritic segments of neurons in culture.

We then estimate the connectivity repertoire by calculating the number of different potential synapses combinations through counting the number of dendritic arbor conformations (Figure 5.4A) and by enumerating the number of connectivity patterns through choosing actual connections out of potential synapses (Figure 5.4B). The strength of this approach is that the number of arbor conformations can be calculated by using the formulism from the statistical physics of polymers.

When choosing actual connections out of potential synapses we must take into account the possibility that even under the assumption that dendritic arbors are sparse, some axons can still establish two potential synapses with a dendritic arbor (double-hits, Figure 5.4C). These additional potential synapses do not expand the repertoire of different connectivity patterns and must be subtracted from the number of combinations.

To maximize the connectivity repertoire and thereby to reduce the number of "double-hits," a dendritic arbor favors longer dendritic branches and bigger arbor span. To see this qualitatively, let us consider two 3D dendritic arbors projected on a 2D plane (Figure 5.5A,B). Both arbors have the same total dendritic length but arbor B has a smaller arbor span than arbor A. Straight axons shooting in a direction perpendicular to the 2D plane have a bigger chance to double-hit the smaller arbor, because there are more over-

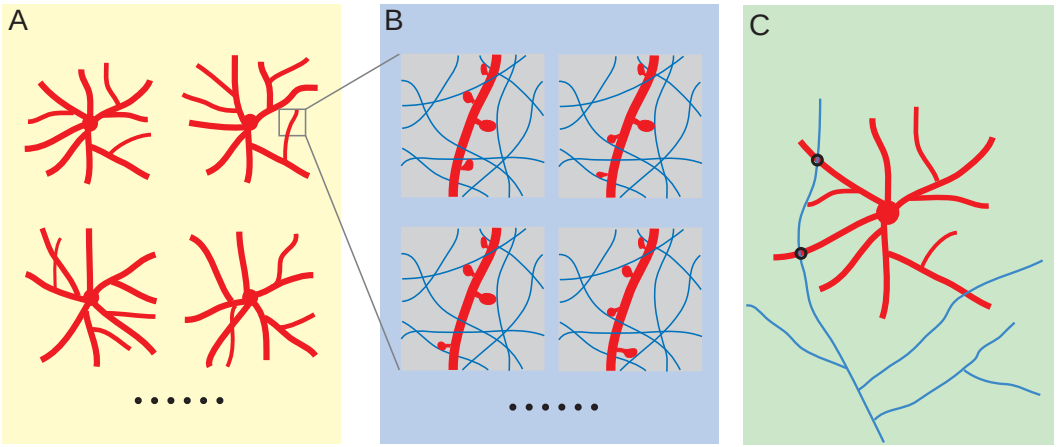

Figure 5.4: Schematic illustrations of major steps in estimating the connectivity repertoire of dendritic arbors. Red – dendrites, blue – axons. (A) Calculation of different arbor conformations. (B) Calculation of different connectivity patterns by choosing actual connections out of potential synapses. (C) Subtraction of "double-hits" due to the fact that some axonal arbors can establish two potential synapses with the dendritic arbor, as shown by the black circles.

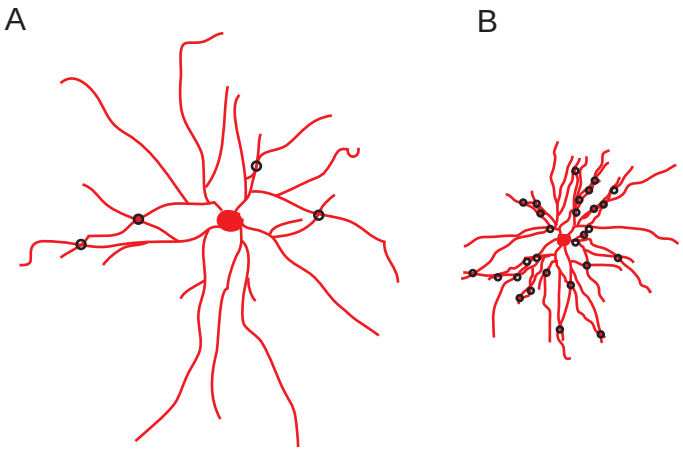

Figure 5.5: Number of axons establishing two potential synapses with a dendritic arbor ("double-hits") depends on the dendritic arbor span. In this schematic illustration, two 3D dendritic arbors are projected onto a 2D plane. Both arbors have the same total dendritic length but arbor B has a smaller arbor span than arbor A. Solid blue circles represent straight axons passing through the dendritic arbors in a direction perpendicular to the 2D plane. Axons have a bigger chance to double-hit arbor B because dendritic branches in a smaller arbor overlap more substantially, as shown by the black circles.

laps of dendritic branches on the 2D projection (black circles in Figure 5.5). Therefore, subtracting the "double-hits" can be seen effectively as a repulsive force to expand the arbor size.

To calculate the reduction of connecitivity patterns due to "double-hits," we rely on the density-density correlation function of an axonal arbor, $g(r)$, which is defined as the length density of the axonal arbor at a distance r from a segment of the same axonal arbor. $g(r)$ was measured for pyramidal axonal arbors, as shown in Figure 5.6. At short distances, this plot can be fitted by a power law with exponent $\gamma = 1.25 \pm 0.05$.

Finally, by maximizing the connectivity repertoire while minimizing the

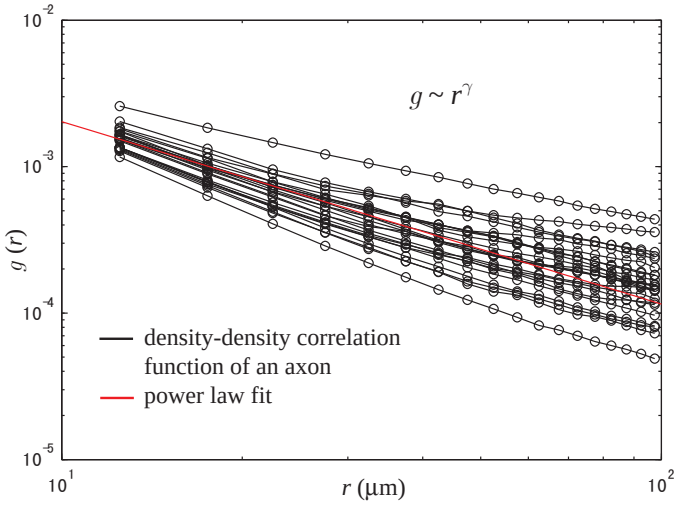

Figure 5.6: Density-density correlation function of axonal arbors $g(r)$. $g(r)$ is defined as the length density of axonal segments at a distance r from the segment of the same axonal arbor. The head of $g(r)$ from different axons are combined and fitted by a power law with exponent $\gamma = 1.25 \pm 0.05$.

dendritic cost, we find the power law dependence of dendritic arbor span on the total dendritic length with exponent $\nu = \frac{1}{1+\gamma}$. Although the exact values of exponents must be verified, experimental measurements seem consistent with theoretical predictions.

As a result of minimizing the dendritic cost, we find that the optimal dendritic arbor shape is centripetal so that each segment of the dendrite is directed towards the soma. Thus, the path length from a dendritic segment to the soma is approximately equal to the euclidean distance between the dendritic segment and the soma. We confirm this prediction by plotting the tortuosity of basal pyramidal dendrites, which is defined as the ratio of the path length to the euclidean distance. Data in Figure 5.7 shows that the tortuosity is close to one.

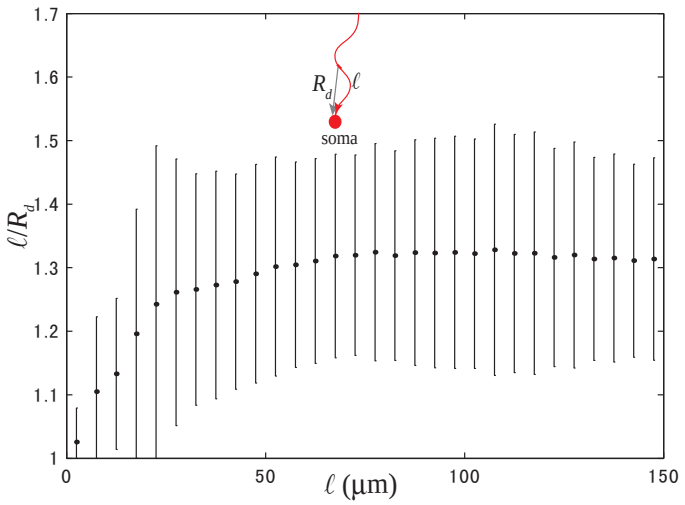

Figure 5.7: Tortuosity of basal pyramidal dendrites. We plot the tortuosity, which is defined as the ratio of the path length from a dendritic segment to the soma to the euclidian distance between the dendritic segment and the soma. The tortuosity is close to one for given path length. Error bars are standard deviations. The measurement was performed by using the 3D dataset.

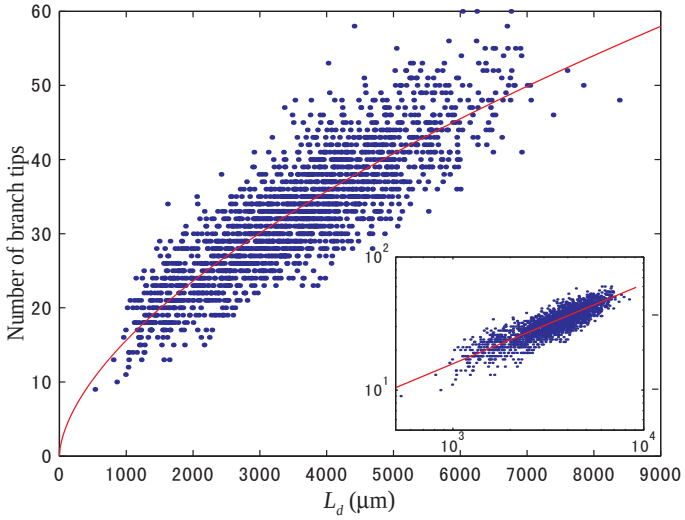

Figure 5.8: Dependence of the number of branch tips on the total dendritic length. The red curve is the power law fit with exponent 0.6 ± 0.02. Inset is the same plot in the log-log coordinates. The measurement was performed by using the 2D dataset.

In addition, the total number of dendritic branches of a centripetal arbor N_b can be approximated as $N_b \approx L_d/R_d \sim L_d^{1-\nu} \sim L_d^{0.55}$, where L_d is the total dendritic length. To test this prediction, we counted the number of branch tips of dendritic arbors and find it scales as a power law of the total dendritic length with exponent 0.6 ± 0.02 (Figure 5.8). Thus, the experimental meansurement seems close to the theoretical predictions.

5.4 Methods

5.4.1 Reconstruction of Neuronal Arbors

The 3D dataset consists of pyramidal neurons obtained from different layers of the primary visual cortex of adult cats. The neurons were reconstructed and digitized by other experimental groups [140, 141, 142]. Detailed methodology has been described elsewhere [140, 141, 142]. Briefly, cells were labeled with biocytin and biotinylated dexran amine *in vivo* and drawn using a computerized 3D reconstruction system, Neurolucida (MicroBrightField, Cochester, VT) from multiple tissue sections. All reconstructions were corrected for tissue shrinkage in three dimensions.

The 2D dataset consists of images of layer II-III pyramidal basal dendrites obtained from different cortical areas of the primates, which are kindly provided by Elston [143, 144, 122, 145, 146, 147, 148]. Cells were labeled with Lucifer-yellow *in vitro* and drawn in a 2D plane parallel to the cortical surface with the aid of camera lucida attached to a Zeiss microscope. All cells were corrected for tissue shrinkage.

5.4.2 Curve Fitting and Error Analysis

All the plots are fitted by using the nonlinear least squares method from the MATLAB curve-fitting toolbox (MathWorks, Natick, MA). The errors of the fitting parameters are within 95 percent confidence bounds.

5.4.3 Generalized Sholl Plot of Dendritic Arbors

We quantify the dendritic arbor shape by generalizing Sholl's original scheme [123], and the detailed analysis is described as follows. A sphere with radius r was centered on one dendritic segment and the number of branches intersecting the sphere was counted. Unlike the original Sholl plot in which the center of the sphere is the soma, we repeat the procedure by using different dendritic segments as the centers.

Because spheres near the exterior boundary of the arbor intersect fewer branches than they normally do, we restricted our measurement to the interior part of the arbor by requiring that the distance between the origin of a sphere and the centroid of the arbor is less than one half of the dendritic arbor span R_d, where R_d^2 is defined as the mean square distance from any dendritic segment to any other dendritic segment, *i.e.*, $R_d^2 = (1/K^2)\sum_{i=1}^{K}\sum_{j=1}^{K}|\mathbf{u_i} - \mathbf{u_j}|^2$. Here $\mathbf{u_i}$ and $\mathbf{u_j}$ are Cartisian coordinates of the center mass of the dendritic segments.

Then, for a given r, the mean number of branches intersecting the spheres ζ was found by averaging over different origins of the spheres.

To combine the generalized Sholl plots that show the dependence of ζ on r for different dendritic arbors, we normalized the sphere radius r to $x = r/R_d$

and the total dendritic length to unity by scaling the number of intersections ζ to $p = \zeta R/L_d$, where L_d is the total dendritic length before normalization.

The generalized Sholl plots after normalization excluded sphere radii smaller than 0.1. At that scale, spheres do not intersect multiple dendritic branches, which leads to a constant number of intersections.

By combining the basal dendrites in the 3D dataset, we find that the generalized Sholl plots can be nicely fitted by the function of the form

$$p = b_1 x^{\mu} \exp(-b_2 x^{\delta}), \tag{5.2}$$

The coefficients b_1 and b_2 are chosen so that the total dendritic length is normalized to unity:

$$\int_0^{\infty} p(x)dx = 1, \tag{5.3}$$

and the second moment that defines R_d^2 is also normalized to unity

$$\int_0^{\infty} p(x)x^2 dx = 1. \tag{5.4}$$

In addition, the head of the Sholl plot (sphere radii x ranging from 0.1 to 0.5) is fitted separately by a power law function ($p \sim x^{\mu}$).

The generalized 3D Sholl plot can be easily extended to the 2D Sholl plot by counting the number of branches intersecting the circles rather than spheres.

5.4.4 Analysis of 2D Dendritic Arbors

To analyze the 2D dataset, we converted the raw images of dendritic arbors into skeletonized binary images, where dendritic segments are represented by white pixels with value 1. To measure the arbor span R_d, we calculate the mean square distance from any white pixels to any others, *i.e.*, $R_d^2 = (1/K^2)\sum_{i=1}^{K}\sum_{j=1}^{K}|\mathbf{u_i} - \mathbf{u_j}|^2$,where $\mathbf{u_i}$ and $\mathbf{u_j}$ are Cartisian coordinates of pixel 1s. The total dendritic length was calculated by counting the number of white pixels multiplying the pixel size. To calculate the total number of branch tips, we counted the number of white pixels whose seven out of eight neighboring pixels are 0. We made generalized Sholl plots for a 2D dataset by following the same procedure as in the 3D Sholl plot, except for using circles rather than spheres to intersect the arbors. All programming was done in MATLAB (MathWorks, Natick, MA).

5.4.5 Scaling Relationship of Self-similar Dendritic Arbors

We have shown that the head of the Sholl plot can be fitted by a power law with exponent μ, and the dependence of arbor span R_d on the total dendritic length L_d can be fitted by a power law with exponent ν. In this section, we show that these two exponents satisfy Equation 5.1 in self-similar dendritic arbors.

Consider a sphere with radius $r \ll R_d$ centered on a dendritic segment. In the sphere, we have dendritic branches with total length l. If part of a dendritic arbor shows the same statistical properties as the entire arbor, we

have $r \sim l^{\nu}$. Then the number of branches intersecting the sphere ζ is given by the dendritic length within the spherical shell of thickness Δr at r divided by Δr. By taking the limit $\Delta r \to 0$, we have

$$\zeta = \frac{dl(r)}{dr} \sim r^{1/\nu - 1} \sim r^{\mu}. \tag{5.5}$$

Thus, the relationship between ν and μ satisfies

$$\mu = \frac{1}{\nu} - 1. \tag{5.6}$$

5.4.6 Quantitative Expression for the Cost of a Dendritic Arbor

Based on the results presented in Chapter 4, we postulate that the cost of a dendritic arbor is the sum of the dendritic volume, V_d, and the signal attenuation, T_d,

$$E_d = V_d + \alpha T_d, \tag{5.7}$$

where α is a constant coefficient.

If we model a dendritic segment as a solid cylinder, the volume of a dendritic arbor can be straightforwardly expressed in terms of its total length, L_d, and its mean diameter, d_d,

$$V_d \sim L_d d_d^2. \tag{5.8}$$

The attenuation cost T_d may be quantified as the sum of fractional at-

tenuation of the voltage signal from synapses to the soma. Assuming that the mean path length from a synapse to the soma, ℓ, is much smaller than the space constant, we have

$$T_d \sim N\frac{\ell}{d_d^\theta}, \tag{5.9}$$

where θ is a positive power and C is the number of actual connections converging on a dendritic arbor. Combining Equations 5.8 and 5.9 we can re-write Equation 5.7 as:

$$E_d \sim d_d^2 L_d + \alpha C\frac{\ell}{d_d^\theta}. \tag{5.10}$$

According to Equation 5.10, the volume and attenuation have opposite dependence on the dendritic diameter and the competition between them determines the optimal diameter. By setting $\partial E_d/\partial d_d = 0$, we find that the optimal dendritic diameter is given by

$$d_d \sim \left(\frac{\ell C}{L_d}\right)^{\frac{1}{2+\theta}}. \tag{5.11}$$

By substituting Equation 5.11 into Equation 5.10, E_d is given by

$$E_d \sim \ell^{\frac{2}{2+\theta}} C^{\frac{1}{2+\theta}} L_d^{\left(1-\frac{1}{2+\theta}\right)}. \tag{5.12}$$

If we assume $C \sim L_d$ as suggested by anatomical data [126], Equation 5.12 can be rewritten as

$$E_d \sim \ell^{\frac{2}{2+\theta}} L_d. \tag{5.13}$$

If $\theta \ll 1$, the dendritic cost may be approximated as

$$E_d \sim \ell L_d. \tag{5.14}$$

Moreover, to minimize E_d in Equation 5.14, the typical path should be as short as possible. However the path length ℓ cannot be smaller than the arbor span, R_d. To estimate by order of magnitude, we have

$$\ell \sim R_d. \tag{5.15}$$

To satisfy Equation 5.15, each segment of the dendrite must be directed towards the soma. We call such arbor design centripetal.

5.4.7 Calculation of the Number of Possible Connectivity Patterns of Sparse Centripetal Arbors

In this section, we will estimate the total number of actual connectivity patterns Ω of centripetal dendritic arbors. An actual connection represents a pre- post-synaptic neuron pair irrespective of the number of actual synapses made between the neurons. For our calculation, we assume that the axonal inputs are randomly distributed in space and the dendritic arbor is sparse so that when an axonal arbor field overlap with the dendritic arbor field, the probability P of establishing a potential synapse is small (*i.e.*, axonal and dendritic segments come within a spine length of each other, allowing them to form an actual synapse by growing a spine).

Under the above assumptions, Ω can be calculated as follows. First, we

will estimate different potential synapses combinations through estimating the total number of distinct arbor conformations Ω_p. To calculate arbor conformations, we model dendritic arbors as flexible chains that are jointed by rigid segments. Second, we show that when the length of a rigid segment is much longer than the inter-synapse interval on a dendrite, the number of common connectivity patterns that can be achieved by two distinct arbor conformations is small. Then, we could calculate Ω as the product of Ω_p, and the number of actual connectivity patterns that can be achieved by a given arbor conformation, Ω_a. Third, we enumerate Ω_a by choosing actual connections out of the potential synapses established between axonal and the dendritic arbors.

Different arbor conformations Ω_p

To calculate the number of conformations of branching and centripetal dendritic arbors Ω_p, we characterize the arbor with total length L_d as a continuous and flexible chain, which is composed of L_d/a rigid segments and each has length a. Thus, our problem is equivalent to calculating the entropy of a fully stretched and branching polymer with persistent length a.

The entropy of fully stretched branching polymers depends on the polymer span R_d and the average path length between one segment and the other ℓ [149]. Here, we may omit the detailed derivation and refer the final expression [138, 149]:

$$\log \Omega_p(R_d, \ell) = S_0(L_d, a) + \frac{\ell}{a} \log(1 - \frac{R_d}{\ell}), \tag{5.16}$$

where S_0 is the total entropy of randomly branching polymers with the leading term $S_0 \approx L_d/a$.

Then, Given the number of actual connectivity patterns that can be achieved by an arbor, Ω_a, the logarithm of the total number of achievable connectivity patterns Ω can be expressed as

$$\log \Omega = \log \Omega_p + \log \Omega_a, \tag{5.17}$$

To justify Equation 5.17, we should take into account the fact that two different arbors can probably have the same actual connectivity patterns and show that the number of common connectivity patterns is much smaller than Ω. As an axon has probability P to establish a potential synapse with a dendritic arbor, the chance that two distinct arbors receive the same C connections is P^C. Then, a given connectivity pattern can be achieved by $P^C\Omega_p$ distinct arbors. Thus the total number of common connectivity patterns is $\Omega P^C \Omega_p$, which is a small correction and can be ignored if

$$P^C \Omega_p \ll 1. \tag{5.18}$$

By substituting Equation 5.16 into Equation 5.18, we find that $\log(P^C\Omega_p)$ $\sim L_d/a - C \ll 0$ provided that $L_d/C \ll a$. Therefore, Equation 5.17 is a valid approximation when the inter-synapse interval is smaller than the persistent length a. This assumption should be tested experimentally.

Different connectivity patterns for a given arbor conformation Ω_a

For a given dendritic arbor conformation, we calculate the number of connectivity patterns Ω_a by choosing C actual connections out of C_p potential

synapses established between the dendritic arbor and the axons. The result would simply be the binomial coefficient $\binom{C_p}{C}$ provided that each potential axonal arbor made just one potential synapse with the dendritic arbor. However, even under the assumption of sparse arbors, a small number of axons, N_2, will establish two potential synapses with the dendritic arbor (*i.e.*, "double-hits"). If we Assume that a neuron acts as a unit, Ω_a is reduced to

$$\Omega_a = \binom{C_p - N_2}{C}. \tag{5.19}$$

Because in the case of sparse arbors, $C_p \gg N_2 \gg 1$, the logarithm of Ω_a can be approximated as

$$\log \Omega_a \approx \log \binom{C_p}{C} + N_2 \log(1 - f), \tag{5.20}$$

where $f = C/C_p$ is called the filling fraction. The above equation shows that the "double-hits" from N_2 axons lead to a negetive correction.

For cortical excitatory neurons, Stepanyants et. al [37] has estimated the filling fraction and found it much smaller than one. Then, using the relation $\log(1 - f) \approx -f$, we can simplify Equation 5.20 further as

$$\log \Omega_a \approx \log \binom{C_p}{C} - N_2 \frac{C}{C_p}, \tag{5.21}$$

Number of axons with double-hits, N_2

To estimate the number of axons with double-hits, N_2, first, we will estimate the number of axons establishing potential synapses with two dendritic

segments with length a by calculating the number of axons establishing potential synapses with one segment and the probablity of those axons establishing potential synapses with the other dendritic segment.

As all the potential axons should fit within the spine length of a dendritic segment, the number of axons establishing potential synapses with one segment is given by the spine-reach zone area of the segment sa times the length density of axons ρ_a, where ρ_a is defined as the total length of axonal arbors within a chunk of neuropil divided by the neuropil volume.

Consider two dendritic segments i and j separated by a distance $r_{ij} \gg s$ and an axonal arbor that has established a potential synapse with dendritic segment i. The probability of establishing potential synapses with both dendritic segments is the product of the spine-reach zone area of the dendritic segment j, sa, and the length density of the axonal arbor at distance r from the axonal segment that establishes potential synapse with dendritic segment i, $g(r_{ij})$ (*i.e.*, density-density correlation function).

As a result, the total number of axons establishing potential synapses with both segments is given by $s^2a^2\rho_a g(r_{ij})$.

Next, by summing the number of axons establishing potential synapses with all pairs of dendritic segments of the arbor, we have

$$N_2 = \frac{1}{2}\sum_{i=1}^{\frac{L_d}{a}}\sum_{j=1}^{\frac{L_d}{a}} s^2a^2\rho_a g(r_{ij}) = \frac{1}{2}\rho_a s^2 L_d^2 \left\langle g(r_{ij})\right\rangle , \tag{5.22}$$

where $\langle g(r_{ij})\rangle$ averages the density-density correlation function for different r_{ij} .

Moreover, in the spirit of mean field approach, we make an essential

approximation by replacing $\langle g(r_{ij})\rangle$ with $g(\langle r_{ij}\rangle) \sim g(R_d)$, where all the correlations between dendritic segments are ignored. Thus, the number of axons with double-hits may be estimated as

$$N_2 \approx s^2 L_d^2 \rho_a g(R_d). \tag{5.23}$$

Substituting Equation 5.23 into Equation 5.21, we obtain

$$\log \Omega_a \approx \log \binom{C_p}{C} - \frac{\rho_a s^2 L_d^2 C}{C_p} g(R_d), \tag{5.24}$$

Because the total number of potential synapses of a dendritic arbor C_p is set by the spine-reach zone area of the arbor, sL_d, times the length density of axons ρ_a, [13, 37], Equation 5.24 can be rewritten as

$$\log \Omega_a \approx \log \binom{C_p}{C} - sL_d C g(R_d). \tag{5.25}$$

Now, by substituting Equations 5.16 and 5.25 into Equation 5.17 and leaving out all the constant terms independent of the arbor span R_d, we arrive at the full expression for the total number of achievable connectivity patterns Ω,

$$\log \Omega \sim \frac{\ell}{a} \log(1 - R_d/\ell) - sL_d C g(R_d). \tag{5.26}$$

We note that subtraction of the connectivity patterns due to double-hits can be viewed effectively as a repulsive force, and the second term $-sL_d C g(R_d)$ is analog to the Flory-like term in calculating the entropy of real polymer chains in good solvents.

5.4.8 Scaling Law of Dendritic Arbors

In this section, based on the hypothesis that dendrites maximize the available connectivity patterns Ω while minimizing the cost E_d, we derive the power law between the arbor span R_d and the total dendritic length L_d. We perform such optimization by minimizing the following objective function,

$$F = \beta E - \log \Omega, \tag{5.27}$$

where β is an unknown constant coefficient reflecting the relative contribution of the dendritic cost to the objective function.

By substituting Equations 5.14, 5.26 into Equation 5.27, we have

$$F \sim \beta \ell L_d - \frac{\ell}{a} \log(1 - \frac{R_d}{\ell}) + s L_d C g(R_d). \tag{5.28}$$

We will find the optimal arbor span by sequentially minimizing F as a function of ℓ and R_d. After setting $\partial F / \partial \ell = 0$, we find that

$$\beta L_d a \sim \log(1 - R_d/\ell) + \frac{1}{\ell/R_d - 1}. \tag{5.29}$$

Since $\ell \approx R_d$, the first term on the right side of Equation 5.29 is much smaller than the second term and can be ignored. Thus we have

$$\ell/R_d \sim 1 + \frac{1}{\beta L_d a}. \tag{5.30}$$

According to Equation 5.30, in order for this approach to be self-consistent, we must have $\beta L_d a \gg 1$, which suggests that the dendritic arbor is fully

stretched when the contribution of dendritic cost in the objection function F is sufficiently high.

Next, substituting Equations 5.29 and 5.30 into Equation 5.28 and eliminating ℓ, we have

$$F \sim \beta L_d R_d + s L_d C g(R_d). \tag{5.31}$$

At short distance, $g(R_d)$ can be approximated as a power law and Equation 5.31 becomes

$$F \sim \beta L_d R_d + \frac{s L_d C}{R_d^{\gamma}}, \tag{5.32}$$

where γ is an exponent that can be measured from axonal data. Equation 5.32 has a minimum at which the arbor span R_d is optimal. Omitting all the coefficients and assuming $C \sim L_d$ as suggested by anatomical data, we find

$$R_d \sim L_d^{\frac{1}{1+\gamma}}. \tag{5.33}$$

Using the value of $\gamma = 1.25 \pm 0.05$ as measured from axonal arbor of pyramidal cells, we find that $\nu = \frac{1}{1+\gamma} = 0.44$, which is in agreement with experimental measurement (Figure 5.2).

Chapter 6

Conclusion

In this thesis, by applying the constrained optimization approach, I have successfully explained various aspects of the brain structures. The work is based on one paper published in PloS computational biology (Chapter 3), one paper that has been submitted to Journal of neurophysiology (Chapter 4) and two manuscripts in preparation (Chapter 2 and Chapter 5).

The theory, however, is still far from being complete because we do not know all the fundamental constraints of the system, and what is being optimized under what constraint is also problem-dependent. How do you know that certain cost has been optimized? We do not have a simple answer. Based on how the principles are identified in the physical sciences, we sought to formulate an optimization theory that has fewer assumptions and applies to more cases ("Occam's Razor"). In several cases, when experiments reveal discrepancies, they lead to the discovery of new principles. For example, we have shown that minimizing the dendritic cost alone cannot explain the sparseness of pyramidal dendrites (Chapter 4). Such discrepancy forces us to find

another principle: the dendritic arbor shape maximizes the total number of connectivity patterns.

The optimization approach is straightforward in the sense that the theory bypasses the highly complicated developmental processes of the brain. Yet this is also its weakness because ultimately we want to know how the brain structures are formed through the genes. It would be interesting to know, for example, what kind of molecules can act like a repulsive force among dendritic branches so that the final form of an arbor is sparse.

Many puzzles in the brain structures remain unanswered. The following are a couple of problems that may inspire future research.

- What determines the shape of an axonal arbor? Axonal arbors of pyramidal neurons are much larger and look more irregular than dendritic arbors. Can we use the same principles to describe the axons?

- In many cases, when two neurons connect with each other, multiple synapses are made. What are the advantages of such a design?

- Why are cell body positions different in different brain regions? In the neocortex, cell bodies are scattered throughout the gray matter. In the hippocampus and cerebellum, cell bodies are condensed into a single layer.

With all these caveats in our mind, I am still very confident that our approach is on the right track because we are able to explain such a broad spectrum of anatomical features with relatively few assumptions. If the goal of this chapter is to summarize the thesis with one sentence, I would like to

ask a question reflecting a theoretical physicist's idealism. "Can we have a unified principle to describe the brain structures?"

Bibliography

[1] Santiago Ramón y Cajal. *Histology of the nervous system of man and vertebrates*. History of neuroscience ; no. 6. Oxford University Press, New York, 1995.

[2] S. G. Waxman and M. V. Bennett. Relative conduction velocities of small myelinated and non-myelinated fibres in the central nervous system. *Nat New Biol*, 238(85):217–9, 1972.

[3] W. A. Rushton. A theory of the effects of fibre size in medullated nerve. *J Physiol*, 115(1):101–22, 1951.

[4] Theodore Holmes Bullock and G. Adrian Horridge. *Structure and function in the nervous systems of invertebrates*. W. H. Freeman, San Francisco,, 1965.

[5] JB Hursh. Conduction velocity and diameter of nerve fibers. *Amer J Physiol*, 127:131–139, 1939.

[6] A. L. Hodgkin. A note on conduction velocity. *J Physiol*, 125(1):221–4, 1954.

[7] S. G. Waxman and H. A. Swadlow. Ultrastructure of visual callosal axons in the rabbit. *Exp Neurol*, 53(1):115–27, 1976.

[8] P. Franson and C. Hildebrand. Postnatal growth of nerve fibres in the pyramidal tract of the rabbit. *Neurobiology*, 5(1):8–22, 1975.

[9] S Remahl and C. Hildebrand. Changing relation between onset of myelination and axon diameter range in developing feline white matter. *J Neurl Sci*, 54:33–45, 1982.

[10] Jennifer R Shultz and et al. Speed limits in mammalian brains: scaling constraints from biophysics. in preparing.

[11] C. Cherniak. Local optimization of neuron arbors. *Biol Cybern*, 66(6):503–10, 1992.

[12] G. Mitchison. Neuronal branching patterns and the economy of cortical wiring. *Proc R Soc Lond B Biol Sci*, 245(1313):151–8, 1991.

[13] D. B. Chklovskii. Synaptic connectivity and neuronal morphology: two sides of the same coin. *Neuron*, 43(5):609–17, 2004.

[14] D. B. Chklovskii and A. Stepanyants. Power-law for axon diameters at branch point. *BMC Neurosci*, 4(1):18, 2003.

[15] H. A. Swadlow and S. G. Waxman. Variations in conduction velocity and excitability following single and multiple impulses of visual callosal axons in the rabbit. *Exp Neurol*, 53(1):128–50, 1976.

[16] F. Aboitiz, A. B. Scheibel, R. S. Fisher, and E. Zaidel. Fiber composition of the human corpus callosum. *Brain Res*, 598(1-2):143–53, 1992.

[17] HS Gasser. Unmedullated fibers originating in dorsal root ganglia. *J Gen Physiol*, 33:277–297, 1950.

[18] I. A. Boyd and K. U. Kalu. Scaling factor relating conduction velocity and diameter for myelinated afferent nerve fibres in the cat hind limb. *J Physiol*, 289:277–97, 1979.

[19] E. R. Griff, C. A. Greer, F. Margolis, M. Ennis, and M. T. Shipley. Ultrastructural characteristics and conduction velocity of olfactory receptor neuron axons in the olfactory marker protein-null mouse. *Brain Res*, 866(1-2):227–36, 2000.

[20] R.D. Wright. some mechanical factors in the evolution of the central nervous system. *Journal of Anatomy*, 69:86–88, 1934.

[21] D. C. Van Essen. A tension-based theory of morphogenesis and compact wiring in the central nervous system. *Nature*, 385(6614):313–8, 1997.

[22] K. Zhang and T. J. Sejnowski. A universal scaling law between gray matter and white matter of cerebral cortex. *Proc Natl Acad Sci U S A*, 97(10):5621–6, 2000.

[23] Harry J. Jerison. *Evolution of the brain and intelligence.* Academic Press, New York,, 1973.

[24] John Morgan Allman. *Evolving brains.* Scientific American Library series, no. 68. Scientifc American Library : Distributed by W.H. Freeman and Co., New York, 1999.

[25] G.F Striedter. *principles of brain evolution*. Sinauer Associates, Sunderland, MA, 2005.

[26] D. O. Hebb. *The organization of behavior; a neuropsychological theory*. Wiley book in clinical psychology. Wiley, New York,, 1949.

[27] Larry R. Squire and Eric R. Kandel. *Memory : from mind to molecules*. Scientific American Library paperback. Scientific American Library, New York, 2000.

[28] W. B. Levy and R. A. Baxter. Energy efficient neural codes. *Neural Comput*, 8(3):531–43, 1996.

[29] D. Attwell and S. B. Laughlin. An energy budget for signaling in the grey matter of the brain. *J Cereb Blood Flow Metab*, 21(10):1133–45, 2001.

[30] S. B. Laughlin and T. J. Sejnowski. Communication in neuronal networks. *Science*, 301(5641):1870–4, 2003.

[31] B. J. Dickson, H. Cline, F. Polleux, and A. Ghosh. Making connections. meeting: axon guidance and neural plasticity. *EMBO Rep*, 2(3):182–6, 2001.

[32] D. B. Chklovskii, T. Schikorski, and C. F. Stevens. Wiring optimization in cortical circuits. *Neuron*, 34(3):341–7, 2002.

[33] E. Ruppin, E. L. Schwartz, and Y. Yeshurun. Examining the volume efficiency of the cortical architecture in a multi-processor network model. *Biol Cybern*, 70(1):89–94, 1993.

[34] J. M. Murre and D. P. Sturdy. The connectivity of the brain: multi-level quantitative analysis. *Biol Cybern*, 73(6):529–45, 1995.

[35] R. J. Douglas, C. Koch, M. Mahowald, K. A. Martin, and H. H. Suarez. Recurrent excitation in neocortical circuits. *Science*, 269(5226):981–5, 1995.

[36] T. Binzegger, R. J. Douglas, and K. A. Martin. A quantitative map of the circuit of cat primary visual cortex. *J Neurosci*, 24(39):8441–53, 2004.

[37] A. Stepanyants, P. R. Hof, and D. B. Chklovskii. Geometry and structural plasticity of synaptic connectivity. *Neuron*, 34(2):275–88, 2002.

[38] N. Kalisman, G. Silberberg, and H. Markram. The neocortical microcircuit as a tabula rasa. *Proc Natl Acad Sci U S A*, 102(3):880–5, 2005.

[39] J.M. Ritchie. *The axon : structure, function, and pathophysiology, Physiology of axons. pp. 68-96.* Oxford University Press, New York, 1995.

[40] Christof Koch. *Biophysics of computation : information processing in single neurons.* Computational neuroscience. Oxford University Press, New York, 1999.

[41] B. Hoffmeister, W. Janig, and S. J. Lisney. A proposed relationship between circumference and conduction velocity of unmyelinated axons from normal and regenerated cat hindlimb cutaneous nerves. *Neuroscience*, 42(2):603–11, 1991.

[42] V. Braitenberg and A. Schuz. *Cortex: statistics and geometry of neuronal connectivity*. Springer, Berlin New York, 1998.

[43] D. Attwell and A. Gibb. Neuroenergetics and the kinetic design of excitatory synapses. *Nat Rev Neurosci*, 6(11):841–9, 2005.

[44] J. L. Ringo. Neuronal interconnection as a function of brain size. *Brain Behav Evol*, 38(1):1–6, 1991.

[45] J. L. Ringo, R. W. Doty, S. Demeter, and P. Y. Simard. Time is of the essence: a conjecture that hemispheric specialization arises from interhemispheric conduction delay. *Cereb Cortex*, 4(4):331–43, 1994.

[46] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–2, 1998.

[47] M. A. Changizi. Principles underlying mammalian neocortical scaling. *Biol Cybern*, 84(3):207–15, 2001.

[48] K. H. Harrison, P. R. Hof, and S. S. Wang. Scaling laws in the mammalian neocortex: does form provide clues to function? *J Neurocytol*, 31(3-5):289–98, 2002.

[49] J. Karbowski. How does connectivity between cortical areas depend on brain size? implications for efficient computation. *J Comput Neurosci*, 15(3):347–56, 2003.

[50] O. Sporns and R. Kotter. Motifs in brain networks. *PLoS Biol*, 2(11):e369, 2004.

[51] C. F. Stevens. How cortical interconnectedness varies with network size. *Neural computation*, 1:473–479, 1989.

[52] T Deacon. Rethinking mammalian brain evolution. *Amer Zool*, 30:629–705, 1990.

[53] V. Braitenberg. Brain size and number of neurons: an exercise in synthetic neuroanatomy. *J Comput Neurosci*, 10(1):71–7, 2001.

[54] D. J. Felleman and D. C. Van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex*, 1(1):1–47, 1991.

[55] S. T. Carmichael and J. L. Price. Connectional networks within the orbital and medial prefrontal cortex of macaque monkeys. *J Comp Neurol*, 371(2):179–207, 1996.

[56] A. Stepanyants, G. Tamas, and D. B. Chklovskii. Class-specific features of neuronal wiring. *Neuron*, 43(2):251–9, 2004.

[57] G. Partadiredja, R. Miller, and D. E. Oorschot. The number, size, and type of axons in rat subcortical white matter on left and right sides: a stereological, ultrastructural study. *J Neurocytol*, 32(9):1165–79, 2003.

[58] D. E. Oorschot. Total number of neurons in the neostriatal, pallidal, subthalamic, and substantia nigral nuclei of the rat basal ganglia: a stereological study using the cavalieri and optical disector methods. *J Comp Neurol*, 366(4):580–99, 1996.

[59] J. W. Prothero and J. W. Sundsten. Folding of the cerebral cortex in mammals. a scaling model. *Brain Behav Evol*, 24(2-3):152–67, 1984.

[60] M. A. Hofman. Size and shape of the cerebral cortex in mammals. i. the cortical surface. *Brain Behav Evol*, 27(1):28–40, 1985.

[61] J. Prothero. Cortical scaling in mammals: a repeating units model. *J Hirnforsch*, 38(2):195–207, 1997.

[62] P. S. Goldman-Rakic. Cytoarchitectonic heterogeneity of the primate neostriatum: subdivision into island and matrix cellular compartments. *J Comp Neurol*, 205(4):398–413, 1982.

[63] M. Herkenham, S. M. Edley, and J. Stuart. Cell clusters in the nucleus accumbens of the rat, and the mosaic relationship of opiate receptors, acetylcholinesterase and subcortical afferent terminations. *Neuroscience*, 11(3):561–93, 1984.

[64] Gordon M. Shepherd. *The synaptic organization of the brain*. Oxford University Press, New York, 4th edition, 1998.

[65] A. M. Graybiel, Jr. Ragsdale, C. W., E. S. Yoneoka, and R. P. Elde. An immunohistochemical study of enkephalins and other neuropeptides in the striatum of the cat with evidence that the opiate peptides are arranged to form mosaic patterns in register with the striosomal compartments visible by acetylcholinesterase staining. *Neuroscience*, 6(3):377–97, 1981.

[66] M. Herkenham and C. B. Pert. Mosaic distribution of opiate receptors, parafascicular projections and acetylcholinesterase in rat striatum. *Nature*, 291(5814):415–8, 1981.

[67] C. R. Gerfen. The neostriatal mosaic: striatal patch-matrix organization is related to cortical lamination. *Science*, 246(4928):385–8, 1989.

[68] L. L. Brown, S. M. Feldman, D. M. Smith, J. R. Cavanaugh, R. F. Ackermann, and A. M. Graybiel. Differential metabolic activity in the striosome and matrix compartments of the rat striatum during natural behaviors. *J Neurosci*, 22(1):305–14, 2002.

[69] G. E. Vates, B. M. Broome, C. V. Mello, and F. Nottebohm. Auditory pathways of caudal telencephalon and their relation to the song system of adult male zebra finches. *J Comp Neurol*, 366(4):613–42, 1996.

[70] S. Cullheim, J. W. Fleshman, L. L. Glenn, and R. E. Burke. Membrane area and dendritic structure in type-identified triceps surae alpha motoneurons. *J Comp Neurol*, 255(1):68–81, 1987.

[71] K. J. Friston, C. D. Frith, P. F. Liddle, and R. S. Frackowiak. Functional connectivity: the principal-component analysis of large (pet) data sets. *J Cereb Blood Flow Metab*, 13(1):5–14, 1993.

[72] M. P. Young, J. W. Scannell, M. A. O'Neill, C. C. Hilgetag, G. Burns, and C. Blakemore. Non-metric multidimensional scaling in the analysis of neuroanatomical connection data and the organization of the primate cortical visual system. *Philos Trans R Soc Lond B Biol Sci*, 348(1325):281–308, 1995.

[73] C. C. Hilgetag, G. A. Burns, M. A. O'Neill, J. W. Scannell, and M. P. Young. Anatomical connectivity defines the organization of clusters of

cortical areas in the macaque monkey and the cat. *Philos Trans R Soc Lond B Biol Sci*, 355(1393):91–110, 2000.

[74] O. Sporns, G. Tononi, and G. M. Edelman. Connectivity and complexity: the relationship between neuroanatomy and brain dynamics. *Neural Netw*, 13(8-9):909–22, 2000.

[75] K. E. Stephan, C. C. Hilgetag, G. A. Burns, M. A. O'Neill, M. P. Young, and R. Kotter. Computational analysis of functional connectivity between areas of primate cerebral cortex. *Philos Trans R Soc Lond B Biol Sci*, 355(1393):111–26, 2000.

[76] R. Kotter, K. E. Stephan, N. Palomero-Gallagher, S. Geyer, A. Schleicher, and K. Zilles. Multimodal characterisation of cortical areas by multivariate analyses of receptor binding and connectivity data. *Anat Embryol (Berl)*, 204(4):333–50, 2001.

[77] K. E. Stephan, L. Kamper, A. Bozkurt, G. A. Burns, M. P. Young, and R. Kotter. Advanced database methodology for the collation of connectivity data on the macaque brain (cocomac). *Philos Trans R Soc Lond B Biol Sci*, 356(1412):1159–86, 2001.

[78] C. C. Hilgetag and S. Grant. Uniformity, specificity and variability of corticocortical connectivity. *Philos Trans R Soc Lond B Biol Sci*, 355(1393):7–20, 2000.

[79] C. C. Hilgetag, M. A. O'Neill, and M. P. Young. Hierarchical organization of macaque and cat cortical sensory systems explored with a novel

network processor. *Philos Trans R Soc Lond B Biol Sci*, 355(1393):71–89, 2000.

[80] O. Sporns and J. D. Zwi. The small world of the cerebral cortex. *Neuroinformatics*, 2(2):145–62, 2004.

[81] G. Tononi, O. Sporns, and G. M. Edelman. A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proc Natl Acad Sci U S A*, 91(11):5033–7, 1994.

[82] K. Friston. Functional integration and inference in the brain. *Prog Neurobiol*, 68(2):113–43, 2002.

[83] R. E. Passingham, K. E. Stephan, and R. Kotter. The anatomical basis of functional localization in the cortex. *Nat Rev Neurosci*, 3(8):606–16, 2002.

[84] R. Kotter and K. E. Stephan. Network participation indices: characterizing component roles for information processing in neural networks. *Neural Netw*, 16(9):1261–75, 2003.

[85] G. Tononi and O. Sporns. Measuring information integration. *BMC Neurosci*, 4(1):31, 2003.

[86] W. D. Penny, K. E. Stephan, A. Mechelli, and K. J. Friston. Modelling functional integration: a comparison of structural equation and dynamic causal models. *Neuroimage*, 23 Suppl 1:S264–74, 2004.

[87] B. G. Cragg. The density of synapses and neurones in the motor and visual areas of the cerebral cortex. *J Anat*, 101(4):639–54, 1967.

[88] A. Schuz and G. P. Demianenko. Constancy and variability in cortical structure. a study on synapses and dendritic spines in hedgehog and monkey. *J Hirnforsch*, 36(1):113–22, 1995.

[89] DB Tower. Structural and functional organization of mammalian cerebral cortex: the correlation of neuronal density with brain size. *Journal of Comparative Neurology*, 101:19–52, 1954.

[90] M. A. Hofman. On the evolution and geometry of the brain in mammals. *Prog Neurobiol*, 32(2):137–58, 1989.

[91] M. A. Hofman. Size and shape of the cerebral cortex in mammals. ii. the cortical volume. *Brain Behav Evol*, 32(1):17–26, 1988.

[92] E. C. Bush and J. M. Allman. The scaling of white matter to gray matter in cerebellum and neocortex. *Brain Behav Evol*, 61(1):1–5, 2003.

[93] Y. N. Jan and L. Y. Jan. The control of dendrite development. *Neuron*, 40(2):229–42, 2003.

[94] K. L. Whitford, P. Dijkhuizen, F. Polleux, and A. Ghosh. Molecular control of cortical dendrite development. *Annu Rev Neurosci*, 25:127–49, 2002.

[95] E. K. Scott and L. Luo. How do dendrites take their shape? *Nat Neurosci*, 4(4):359–65, 2001.

[96] H. T. Cline. Dendritic arbor development and synaptogenesis. *Curr Opin Neurobiol*, 11(1):118–26, 2001.

[97] A. Stepanyants and D. B. Chklovskii. Neurogeometry and potential synaptic connectivity. *Trends Neurosci*, 28(7):387–94, 2005.

[98] D. Debanne. Information processing in the axon. *Nat Rev Neurosci*, 5(4):304–16, 2004.

[99] Y. Manor, C. Koch, and I. Segev. Effect of geometrical irregularities on propagation delay in axonal trees. *Biophys J*, 60(6):1424–37, 1991.

[100] Y. Grossman, I. Parnas, and M. E. Spira. Differential conduction block in branches of a bifurcating axon. *J Physiol*, 295:283–305, 1979.

[101] Y. Grossman, I. Parnas, and M. E. Spira. Mechanisms involved in differential conduction of potentials at high frequency in a branching axon. *J Physiol*, 295:307–22, 1979.

[102] Z. F. Mainen and T. J. Sejnowski. Influence of dendritic structure on firing pattern in model neocortical neurons. *Nature*, 382(6589):363–6, 1996.

[103] H. Agmon-Snir, C. E. Carr, and J. Rinzel. The role of dendrites in auditory coincidence detection. *Nature*, 393(6682):268–72, 1998.

[104] M. London and M. Hausser. Dendritic computation. *Annu Rev Neurosci*, 28:503–32, 2005.

[105] C. Koch, T. Poggio, and V. Torre. Retinal ganglion cells: a functional interpretation of dendritic morphology. *Philos Trans R Soc Lond B Biol Sci*, 298(1090):227–63, 1982.

[106] P. Poirazi and B. W. Mel. Impact of active dendrites and structural plasticity on the memory capacity of neural tissue. *Neuron*, 29(3):779–96, 2001.

[107] R. E. Burke, W. B. Marks, and B. Ulfhake. A parsimonious description of motoneuron dendritic morphology using computer simulation. *J Neurosci*, 12(6):2403–16, 1992.

[108] H. Cuntz, A. Borst, and I. Segev. Optimization principles of dendritic structure. *Preprint*.

[109] Q. Wen and D. B. Chklovskii. Segregation of the brain into gray and white matter: a design minimizing conduction delays. *PLoS Comput Biol*, 1(7):e78, 2005.

[110] A. A. Koulakov and D. B. Chklovskii. Orientation preference patterns in mammalian visual cortex: a wire length minimization approach. *Neuron*, 29(2):519–27, 2001.

[111] O. Shefi, A. Harel, D.B. Chklovskii, E. Ben Jacob, and A. Ayali. Biophysical constraints on neuronal branching. *Neurocomputing*, 58-60:487–495, 2004.

[112] V. A. Klyachko and C. F. Stevens. Connectivity optimization and the positioning of cortical areas. *Proc Natl Acad Sci U S A*, 100(13):7937–41, 2003.

[113] C. Cherniak. Neural component placement. *Trends Neurosci*, 18(12):522–7, 1995.

[114] D. B. Chklovskii. Optimal sizes of dendritic and axonal arbors in a topographic projection. *J Neurophysiol*, 83(4):2113–9., 2000.

[115] A. Roth and M. Hausser. Compartmental models of rat cerebellar purkinje cells based on simultaneous somatic and dendritic patch-clamp recordings. *J Physiol*, 535(Pt 2):445–72, 2001.

[116] G. Major, A. U. Larkman, P. Jonas, B. Sakmann, and J. J. Jack. Detailed passive cable models of whole-cell recorded ca3 pyramidal neurons in rat hippocampal slices. *J Neurosci*, 14(8):4613–38, 1994.

[117] A. U. Larkman, G. Major, K. J. Stratford, and J. J. Jack. Dendritic morphology of pyramidal neurones of the visual cortex of the rat. iv: Electrical geometry. *J Comp Neurol*, 323(2):137–52, 1992.

[118] W. Rall. Branching dendritic trees and motoneuron membrane resistivity. *Exp Neurol*, 1:491–527, 1959.

[119] R. M. Napper and R. J. Harvey. Quantitative study of the purkinje cell dendritic spines in the rat cerebellum. *J Comp Neurol*, 274(2):158–67, 1988.

[120] R. M. Napper and R. J. Harvey. Number of parallel fiber synapses on an individual purkinje cell in the cerebellum of the rat. *J Comp Neurol*, 274(2):168–77, 1988.

[121] M. Rapp, I. Segev, and Y. Yarom. Physiology, morphology and detailed passive models of guinea-pig cerebellar purkinje cells. *J Physiol*, 474(1):101–18, 1994.

[122] G. N. Elston, R. Tweedale, and M. G. Rosa. Cellular heterogeneity in cerebral cortex: a study of the morphology of pyramidal neurones in visual areas of the marmoset monkey. *J Comp Neurol*, 415(1):33–51, 1999.

[123] D. A. Sholl. Dendritic organization in the neurons of the visual and motor cortices of the cat. *J Anat*, 87(4):387–406, 1953.

[124] A. U. Larkman. Dendritic morphology of pyramidal neurones of the visual cortex of the rat: I. branching patterns. *J Comp Neurol*, 306(2):307–19, 1991.

[125] A. U. Larkman. Dendritic morphology of pyramidal neurones of the visual cortex of the rat: Ii. parameter correlations. *J Comp Neurol*, 306(2):320–31, 1991.

[126] A. U. Larkman. Dendritic morphology of pyramidal neurones of the visual cortex of the rat: Iii. spine distributions. *J Comp Neurol*, 306(2):332–43, 1991.

[127] P. Rothnie, D. Kabaso, P. R. Hof, B. I. Henry, and S. L. Wearne. Functionally relevant measures of spatial complexity in neuronal dendritic arbors. *J Theor Biol*, 238(3):505–26, 2006.

[128] T. Binzegger, R. J. Douglas, and K. A. Martin. Axons in cat visual cortex are topologically self-similar. *Cereb Cortex*, 15(2):152–65, 2005.

[129] N. T. Milosevic and D. Ristanovic. Fractality of dendritic arborization of spinal cord neurons. *Neurosci Lett*, 396(3):172–6, 2006.

[130] G. N. Elston and B. Zietsch. Fractal analysis as a tool for studying specialization in neuronal structure: The study of the evolution of the primate cerebral cortex and human intellect. *Advances in Complex Systems*, 8(2-3):217–227, 2005.

[131] F. Caserta, H. E. Stanley, W. D. Eldred, G. Daccord, R. E. Hausman, and J. Nittmann. Physical mechanisms underlying neurite outgrowth: A quantitative analysis of neuronal shape. *Physical Review Letters*, 64(1):95–98, 1990.

[132] R. S. Nowakowski, N. L. Hayes, and M. D. Egger. Competitive interactions during dendritic growth: a simple stochastic growth algorithm. *Brain Res*, 576(1):152–6, 1992.

[133] J. van Pelt, A. E. Dityatev, and H. B. Uylings. Natural variability in the number of dendritic segments: model-based inferences about branching during neurite outgrowth. *J Comp Neurol*, 387(3):325–40, 1997.

[134] J. van Pelt and H. B. Uylings. Branching rates and growth functions in the outgrowth of dendritic branching patterns. *Network*, 13(3):261–81, 2002.

[135] A. V. Samsonovich and G. A. Ascoli. Statistical determinants of dendritic morphology in hippocampal pyramidal neurons: A hidden markov model. *Hippocampus*, 15(2):166–83, 2005.

[136] L. R. Varshney, P. J. Sjostrom, and D. B. Chklovskii. Optimal information storage in noisy synapses under resource constraints. *Neuron*, 52(3):409–23, 2006.

[137] Pierre-Gilles de Gennes. *Scaling concepts in polymer physics*. Cornell University Press, Ithaca, N.Y., 1979.

[138] A. I. Fe U. Grosberg and A. R. Khokhlov. *Statistical physics of macromolecules*. AIP series in polymers and complex materials. AIP Press, New York, 1994.

[139] A. Stepanyants, J.A. Hirsch, L.M. Martinez, Z.F. Kisvárday, A.S. Ferecskó, and D.B. Chklovskii. Local potential connectivity in cat primary visual cortex. *Cerebral Cortex*, in print, 2007.

[140] J. A. Hirsch, C. A. Gallagher, J. M. Alonso, and L. M. Martinez. Ascending projections of simple and complex cells in layer 6 of the cat striate cortex. *J Neurosci*, 18(19):8086–94, 1998.

[141] J. A. Hirsch, J. M. Alonso, R. C. Reid, and L. M. Martinez. Synaptic integration in striate cortical simple cells. *J Neurosci*, 18(22):9517–28, 1998.

[142] Z. F. Kisvarday and U. T. Eysel. Cellular organization of reciprocal patchy networks in layer iii of cat visual cortex (area 17). *Neuroscience*, 46(2):275–86, 1992.

[143] G. N. Elston and M. G. Rosa. The occipitoparietal pathway of the macaque monkey: comparison of pyramidal cell morphology in layer iii of functionally related cortical visual areas. *Cereb Cortex*, 7(5):432–52, 1997.

[144] G. N. Elston and M. G. Rosa. Morphological variation of layer iii pyramidal neurones in the occipitotemporal pathway of the macaque monkey visual cortex. *Cereb Cortex*, 8(3):278–94, 1998.

[145] G. N. Elston and M. G. Rosa. Pyramidal cells, patches, and cortical columns: a comparative study of infragranular neurons in teo, te, and the superior temporal polysensory area of the macaque monkey. *J Neurosci*, 20(24):RC117, 2000.

[146] G. N. Elston, A. Elston, J. H. Kaas, and V. Casagrande. Regional specialization in pyramidal cell structure in the visual cortex of the galago: an intracellular injection study of striate and extrastriate areas with comparative notes on new world and old world monkeys. *Brain Behav Evol*, 66(1):10–21, 2005.

[147] G. N. Elston, R. Benavides-Piccione, A. Elston, P. Manger, and J. DeFelipe. Pyramidal cell specialization in the occipitotemporal cortex of the vervet monkey. *Neuroreport*, 16(9):967–70, 2005.

[148] G. N. Elston, R. Benavides-Piccione, A. Elston, J. DeFelipe, and P. Manger. Pyramidal cell specialization in the occipitotemporal cortex of the chacma baboon (papio ursinus). *Exp Brain Res*, 167(4):496–503, 2005.

[149] A. M. Gutin, A. Y. Grosberg, and E. I. Shakhnovich. Polymers with annealed and quenched branchings belong to different universality classes. *Macromolecules*, 26(6):1293–1295, 1993.

Notation and symbols

A: Total surface area of the white matter tracts, 42

A_t: Minimal surface area of a white matter tract, 37

B: Proportionality coefficient between conduction velocity and global axon diameter, 28

B: Proportionality constant between myelinated axon diameter and conduction velocity, 5

C: Number of actual connections converging on a dendritic arbor, 108

C_p: Number of potential synapses converging on a dendritic arbor, 79, 112

D: Global axon diameter, 27

D: Myelinated axon diameter, 5

D_p: Potential divergence of an axonal arbor, 87

E_a: Cost of an axon, 10

E_d: dendritic cost, 77, 108

F: Free energy, 115

V_d: Dendritic volume, 77, 107

Ω: Total number of actual connectivity patterns, 109

Ω_a: Number of actual connectivity patterns that can be achieved by a given arbor conformation, 110

Ω_p: Total number of distinct arbor conformations, 110

Φ: Cross-sectional area of a white matter tract, 37

α: Unit attenuation cost, 77, 108

α: Unit conduction delay cost, 10

β: Relative contribution of dendritic cost, 115

ℓ: Average path length between two connected neurons (via potential synapses), 23

ℓ: Mean path length from a synapse to the soma, 78, 108

ϵ_m: Cost of a myelinated axon per length, 7

ϵ_n:Cost of a non-myelinated axon per length, 7

η: Potential interbouton interval, 87

γ: Scaling exponent in the density-density correlation function of an axonal arbor, 100

λ: Ratio of global axon volume that is finely intermixed with local connections to the initial unperturbed gray matter (*i.e.*, total local circuits) volume, 31

μ: Scaling exponent in the Sholl plot, 94

ν: Critical exponent between dendritic arbor span and total dendritic length, 94

ρ_a: length density of axons, 113

θ: Scaling exponent, 79, 108

θ: scaling exponent between conduction velocity and axon diameter, 16

θ: scaling exponent between conduction velocity and local wire diameter, 24

ξ: space constant, 78

ζ: Mean number of branches intersecting the spheres in the Sholl plot, 105

a: Persistent length, 110

b: Proportionality constant between local wire diameter and conduction velocity, 24

b: Proportionality constant between non-myelinated axon diameter and conduction velocity, 5

d: Local wire diameter, 23

d_0: critical diameter, 6

d_a: Non-myelinated axon diameter, 5

d_d: Dendritic diameter, 78, 108

f: filling fraction, 112

$g(r)$: Density-density correlation function of an axonal arbor, 100

k: Cost per volume for a non-myelinated axon, 7

n: Total number of neurons in the local network, 25

p: Normalized number of intersections, 105

s: spine length, 81

t: Local conduction delay, 23

v_0: Conduction velocity when $v_m = v_n$, 6

v_c: Critical conduction velocity, 8

v_m: Conduction velocity of a global axon, 27

v_m: Conduction velocity of a myelinated axon, 4

v_n: Conduction velocity of a non-myelinated axon, 5

v_n: Local conduction velocity, 23

v_s: Average synapse volume, 25

x: Normalized sphere radius in the Sholl plot, 105