Chapter 7

Organized Complexity: The Network Modeling Problem

§ 1. Considerations in General Model Order Reduction

When one compares the anatomy and physiology literature with the bulk of the literature on neural network theory, one quickly notices what appears to be a glaring contradiction. Biological neural networks characteristically are much more heterogeneous than the models used by neural network theorists. The model networks common in neural network theory tend to be composed of only one or a few different types of neuron models and tend to be very homogeneous. The discrepancy is so obvious it is understandable if one is led to question whether the studies carried out by network theorists have anything at all to do with real neural networks.

The neuron *in vivo* exists in an extraordinarily complex environment in the central nervous system. Most neurons receive synaptic input connections from thousands of other neurons, and in turn project outputs to hundreds or thousands of other neurons. Synapses are distributed over complex dendritic arbors, across the cell body, and even along the axon. Furthermore, a presynaptic cell frequently will make multiple synaptic connections to the same target neuron. Central systems exhibit random-looking background firing activities, upon which is overlaid substantially higher levels of action potential firing activities in particular regions correlated to psychophysical phenomena associated with sensory reception, sensorimotor actions, and higher cognitive phenomena. For a person who prefers a neat, clean, easy-to-understand "picture" of what the typical operating environment of a neuron looks like, all this adds up to as pretty a mess as one is ever likely to encounter. And it is within the context of this environment where the neural network theorist must do his or her work.

Even if neuroscience knew everything there was to know about the individual neuron, which we do not, the complexity of the neuronal environment guarantees that network-level modeling will perforce need to employ numerous approximations, simplifications, and outright guesses in grappling to produce any theory of the neural network function. What hope, then, could there possibly be for the researcher to discover anything about the workings of biological neural networks with enough confidence to claim an understanding of the brain in these murky waters between the level of the individual neuron and the level of neural maps?

Elsewhere in science, systems are studied at two extremes. The first deals with small numbers of variables and fairly reliable quantitative models as the basis. We can call this the *small number*

extreme, and here theoretical sciences, such as mechanics in physics and engineering, work very well. At the other extreme are systems involving on the order of Avogadro's number of constituents and for which statistical methods, such as in thermodynamics and statistical mechanics, work quite well. We can call this the *large number extreme*. But in the hierarchy of modeling levels in neuroscience, the neural network level is neither. System theorist Gerald Weinberg called this the regime of *medium number systems*:

> For systems between the small and the large number extremes, there is an essential failure of the two classical methods. On the one hand, the Square Law of Computation says that we cannot solve medium systems by analysis, while on the other hand the Square Root of *N* Law warns us not to expect too much from averages [WEIN: 19-20].

The two "laws" of which Weinberg speaks in this quote are names for describing the calculations used by science at the small number and large number extremes. He is on the whole rather pessimistic about the ability of science to cope with most of the systems in nature because most systems are medium number systems. Yet try to cope with them we must or else we may as well go home. The questions we face, then, are: What do we try? and How do we determine the confidence with which we may justifiably hold our theories and hypotheses to be true?

None too surprisingly, computational neuroscience's approach to the medium number systems with which it must deal is a blending of the approaches used for small and large number systems. Inasmuch as system theory prides itself on being a mathematical science, you might find it a bit surprising to learn that this blending approach is largely (but not wholly) nonmathematical. To appreciate the method computational neuroscience employs, it is helpful to understand a bit more about the "laws" of which Weinberg speaks in the quote above.

§ 1.1 Weinberg's "Laws"

Weinberg's "Square Law of Computation" refers to small number systems. The example *par excellence* of this is classical mechanics.

> Consider first the equations needed to describe the most general system of only two objects. We must first describe how each object behaves by itself – the "isolated" behavior. We must also consider how the behavior of each body affects that of the other – the "interaction." Finally we must consider how things will behave if neither of the bodies is present – the "field" equation. Altogether, the most general two body system requires four equations: two "isolated" equations, one "interaction" equation, and one "field" equation.
>
> As the number of bodies increases, there remains but a single "field" equation, and only one "isolated" equation per body. The number of "interaction" equations, however, grows magnificently, with the result that for n bodies we would need 2^n relationships! . . . Experience has shown that *unless some simplifications can be made*, the amount of computation involved increases at least as fast as the square of the number of equations. This we call the "Square Law of Computation." Thus, if we double the number of equations, we shall have to find a computer four times as powerful to solve them in the same amount of time. Naturally, the time often goes up faster than this – particularly if some technical difficulty arises, such as a decrease in the

precision of results [WEIN: 7].

Weinberg uses Newton's analysis of planetary orbits to illustrate this point. If no simplifications at all are made in computing the orbits of the planets in the solar system, the number of equations requiring solution is on the order of $10^{30,000}$, clearly an impossible number. Through a series of simplifications, Newton reduced the number of equations to about 10.

> At this point, Newton stopped simplifying and solved the equations analytically. He had actually made numerous other simplifications, such as his consideration of each of the solar bodies as point masses. In each of these cases, he and his contemporaries were generally more aware of – and more concerned about – the simplifying assumptions than are many present-day physics professors who lecture about Newton's calculations. Students, consequently, find it hard to understand why Newton's calculation of planetary orbits is ranked as one of the highest achievements of the human mind.
>
> But the general systems thinker understands. He understands because it is his chosen task to understand the simplifying assumptions of a science – in Wigner's words, those "objects of interest" and "well-defined conditions" *that delimit the domain of application and magnify its power of prediction*. He wants to go right to the beginning of the process by which a scientist forms his models of the world, and to follow that process just as far as it will help him in suggesting useful models for other sciences.
>
> Why is the general systems thinker interested in the simplifications of science – in the science of simplifications? For exactly the same reason as Newton was. The systems theorist knows that the Square Law of Computation puts a limit on the power of any computing device. . . Newton was a genius, but not because of the superior computing power of his brain. Newton's genius was, on the contrary, his ability to simplify, idealize, and streamline the world so that it became, in some measure, tractable to the brains of perfectly ordinary men. By studying the methods of simplification that have succeeded and failed in the past, we hope to make the progress of human knowledge a little less dependent on genius [WEIN: 11-12].

We have seen examples of this simplification process in chapter 6, both with Wilson's approximation models and with Izhikevich's and Rulkov's mimic models. In Wilson's case the simplification comes about because he found it was possible to ignore some objects (particular voltage-gated ionotropic channels in his case) by making phenomenological corrections to the models of other objects so that the overall result was a numerically accurate approximation to the behavior explained at a more mechanistic level by Hodgkin-Huxley models. Izhikevich and Rulkov take this to an even greater extreme by requiring their models to accurately *mimic* Hodgkin-Huxley-like behaviors without having to approximate the mechanisms that underlie it.

At the other extreme of scale we come upon large number systems. The scientific example *par excellence* for the treatment of large number systems is statistical mechanics.

> Newton's achievement was in describing the behavior of a system of perhaps 10^5 objects, of which he found 10 of interest. By the nineteenth century, however, physicists wanted to tackle other systems, simple little systems such as the molecules in a bottle of air.
>
> The molecules in a bottle of air differ from the solar system in several ways. First of all, there are not 10^5 of them, but 10^{23}. Second, the nineteenth century physicists were not interested in just 10 of the molecules, but in all of them. Third, had they been interested in only 10, they would have had to study all 10^{23}, since the molecules were pretty much identical in mass and were, furthermore, in close interaction.
>
> These nineteenth century physicists already knew from Newton that they only had to consider

> pair relations, but this merely reduced the number of equations from about $2^{10^{23}}$ to 10^{46}. Although this is undoubtedly a substantial reduction, the prospects of further reduction of that 10^{46} looked rather grim. After a few fruitless tries at the job, these physicists must have felt much like the fox in Aesop's fable who just could not quite reach the grapes. We know they must have felt that way because they solved their problem the same way the fox did: They decided they did not really want to know about the individual molecules anyway.
>
> Actually, of course, the matter was not entirely one of sour molecules. We might more realistically describe the position of these physicists (such as Gibbs, Boltzmann, and Maxwell) by saying that they were *lucky* not to be interested in things for which they could not solve their equations. They had inherited a set of observed laws (such as Boyle's law) about the behavior of certain measurable properties of gases (such as pressure, temperature, and volume). They believed that gases were made of molecules, but they had to bridge the gap between that belief and the observed properties of gases. They bridged that gap by postulating that the interesting measurements were a few *average* properties of the molecules, rather than the exact properties of any one molecule.
>
> Since the number of different average properties was small, this simplification brought down the amount of computation in one fell swoop. Furthermore, the precision of prediction that was obtained for the averages was excellent, because the number of molecules was very, very large, and therefore the so-called "Law of Large Numbers" could be invoked. What this says, in essence, is that the larger the population, the more likely we are to *observe* values that are close to the *predicted* average values [WEIN: 13-14].

In information theory this same "essence" is known as the "asymptotic equipartition property" of information. It is from this fortunate behavior of systems comprised of very large numbers of constituent objects that one comes to what Nobel laureate Erwin Schrödinger dubbed the "Square Root of *N* Law." Weinberg cites Schrödinger:

> If I tell you that a certain gas under certain conditions of pressure and temperature has a certain density, and if I expressed this by saying that within a certain volume (of a size relevant for some experiment) there are under these conditions just *n* molecules of the gas, then you might be sure that if you could test my statement in a particular moment in time, you would find it inaccurate, the departure being of the order of $\sqrt{n}$. Hence if the number $n = 100$, you would find a departure of about 10, thus relative error = 10%. But if $n = 1$ million, you would be likely to find a departure of about 1000, thus a relative error = 0.1 %. Now, roughly speaking, this statistical law is quite general. The laws of physics and physical chemistry are inaccurate with a probable relative error on the order of $1/\sqrt{n}$, where *n* is the number of molecules that co-operate to bring about the law – to produce its validity within such regions of space or time (or both) that matter, for some consideration or for some experiment.
>
> You see from this again that an organism must have a comparatively gross structure in order to enjoy the benefit of fairly accurate laws, both for its internal life and for its interplay with the external world. For otherwise the number of co-operating particles would be too small, the "law" too inaccurate. The particularly exigent demand is the square root. For though a million is a reasonably large number, an accuracy of just 1 in 1000 is not overwhelmingly good, if a thing claims the dignity of being a "Law of Nature" [SCHR].

To put a number on this "square root of *n*" effect, the measured correlation coefficient in a typical well-controlled experiment between pressure and 1/volume for Boyle's gas law is about 0.9999918. This is one of the best experimental agreements found in science[1] and it illustrates

[1] Why is this accuracy not much better still, as one might expect when Avogadro's number of constituents is involved? The explanation lies with the fact that *other* factors, such as the accuracy of the measurement

what Schrödinger was driving at by saying "an accuracy of just 1 in 1000 is not overwhelmingly good." The accuracy of Boyle's gas law is almost three orders of magnitude better than this.

§ 1.2 Region I and Region II Systems

Correlation coefficients for predictions made in other sciences are not so good as this. For example, an economist might be thrilled if his theory returned a correlation of about 0.7 when its predictions were stacked up against observable or measurable data. But few true-blue physicists would be willing to regard this as an overwhelming triumph. Their standard of judgment rests on theories such as Boyle's law, and perhaps this has something to do with why economics and psychology are dubbed "social sciences" while physics and chemistry reserve for themselves the title of "hard science" or even "exact science." The difference, of course, is that physics gets to operate in either the regime of small numbers or that of large numbers, whereas economics and psychology are condemned to work in the regime of medium numbers. Weinberg classifies types of systems with respect to methods in three groups. Region I he calls the region of ***organized simplicity*** (or the "machine" region). It enjoys the benefits of simplification allowing it to operate well within the limitations imposed by the "Square Law of Computation." Region II he calls the region of ***unorganized complexity*** (or "aggregates" region), which benefits from the "Square Root of *N* Law." Region III he calls the region of ***organized complexity*** (or "systems" region, an appellation that more or less reveals where his interests in system theory lie). Region III is precisely the regime of what he calls medium number systems.

The attainment of scientifically successful theories in both Region I and Region II rests upon the same factor, namely achieving simplification of the models of the systems being studied. In both cases, this is brought about by reducing the number of interaction equations that must be considered. The difference between these regions lies with how they go about achieving this reduction.

Region I systems achieve simplification by being able to ignore factors that have no significant short-term quantitative effect on the phenomena being modeled. For example, the Hodgkin-Huxley model need not take into account the processes within the nucleus of the cell that produce proteins, nor does it need to take into account the various metabolic processes involved in cell respiration. Both of these have long-term effects on cell behavior. Intracellular levels of ATP (one of the products of cell respiration) do affect membrane channel function, but over the short run the ATP levels in the cell can be approximated as relatively constant and thus

---

equipment and knowledge of the precise volume of the gas container, are not as good as the predictive power of Boyle's law.

this factor is absorbed into the parameters that set the cell's resting potential and channel function. This is organized simplicity in the model, because of which the number of equations required for the model can be made small and tractable. It is the basis also for approximation models, in which one knows of certain unmodeled factors that *do* have a direct bearing on the function of the system, but which can be accounted for adequately enough for the purpose of the modeler by "corrections" made to the numerical value other parameters.

Electric circuit models in general are Region I models because they are able to ignore certain effects, such as radiation, and treat the Maxwell equations of electromagnetic theory using "lumped-element" models (resistors, capacitors, inductors, voltage sources, and current sources). In this way, circuit models convert Maxwell's partial differential equations with their boundary conditions into ordinary differential equations with initial conditions. It is only when the wavelengths of the electromagnetic waves (which are the actual phenomena underlying circuit behavior) are comparable to the linear dimensions of the circuit that modelers must take spatial distribution of these waves into account. The electrical engineer does this accounting by introducing "mutual inductances" and "transmission lines" into his model. A circuit model is said to be a "quasi-static approximation" to Maxwell's equations.

Organized simplicity is also the basis for mimic models. In this case, the mimic model is possible because the modeler only cares about stimulus-response ("input-output") behavior and can safely presume that whatever physical factors underlie this behavior are time-invariant enough he need not be concerned with their specific details. Almost all abstract models used in control system theory and communication system design are of this type. The control system or communication system designer need not be specifically concerned with the physical models of the constituent parts of his system because he can properly account for the behavior of the system merely by modeling the information-bearing signals[2] present and the transformations effected on these signals by the "functional blocks" in his system model. The Izhikevich and Rulkov models are examples of mimic models that suppress physical detail and focus on the signal transformation properties of the system being modeled.

In the case of Region II systems, model order reduction (simplification of the model) is made possible because the gross effects of interactions among a huge number of variables tend to produce measurable results ("observables" in the language of system theory) that cluster very tightly around a very small numerical range of values. The "pressure" of a gas in a container is an example of this. The center of this range is called the "expected" or "average" value of the observable. What is key to the success of Region II models is that the deviations of the observable

[2] These are made up of input signals, output signals, and signals representing "state variables."

from its expected value are very small and, equally important, have a predictable range into which observations of these deviations will fall. In other words, the deviations themselves are observables distinct from the expected value and can be given model expressions of their own.

For example, Ohm's law states that the voltage across a resistor is proportional to the current through it. The proportionality constant is called the resistance. Macroscopically, Ohm's law is as exact as any fundamental law of physics. But *microscopically*, the conduction of electrons through the resistor (which constitutes the macroscopic current) is subject to a large number of irregularities, such as reflection of electrons at crystal boundaries within the resistor, that cause tiny fluctuations around the average current and produce tiny fluctuations in the voltage across the resistor. These fluctuations are called thermal ***noise***. More generally, "noise" is the term used to describe *any* fluctuation of a model variable from its expected ("mean") value. Noise is itself treated as a distinct Region II variable in the model, and here what is important is that *the equations describing the noise variables are distinct from the "element law" describing the mean values of the current* (or voltage, or pressure). "Noise" is treated as just another factor contained in the overall model. A model containing one or more "noise" variables is called a ***stochastic model*** and the system it describes is called a ***stochastic system***.

Stochastic variables are said to be "random variables." Weinberg comments,

> The concept of "randomness" is most important for systems thinking, though randomness often leads to properties quite contrary to our intuition. We do not have such a problem in understanding the success of mechanics, for although "simplicity" will prove to be as slippery a concept as "randomness," to a first approximation we were able to use the number of objects as a measure of complexity – the complement of simplicity.
>
> Intuitively, *randomness is the property that makes statistical calculations come out right*. Although this definition is patently circular, it does help us to understand the scope of statistical methods. Consider a typical statistical problem. There is a flu epidemic and we want to know how it will spread through the population so that we may plan for the distribution of a vaccine. If every person is just as likely to get the disease as any other, we can calculate the expected number of cases and the effect of vaccination strategies with great precision. If, on the other hand, there is some sort of nonrandomness in the population, our simple calculations will begin to deviate from the experienced epidemic [WEIN: 17-18].

Weinberg's "definition" of "randomness" is not so much circular as it is backwards. *If* a statistical treatment of a modeling problem gives answers that agree with experiment and observation, *then* we describe the system as one with "random" variables. The word "random" is basically an adjective in the sense that the "blue" in "blue sky" is an adjective. Its use denotes that fact that we are *unable to predict* the precise value of the observable measured in any one trial but we are able to predict the various ***statistical measures*** of the observables (mean, variance, correlation, etc.) to a specific degree of precision and with a specific degree of confidence. In formal mathematics, statistical measures are described by ***probability functions***. A probability function is an

abstraction. One never has a directly experienced encounter with "a probability." A statistic, on the other hand, *is* an observable. Probabilities are ghostly entities of the world of mathematics; statistics belong to the world of direct empirical experience. ***A statistical model is always a mimic model***. In it, the probability function plays the same role that the abstract variables of the Rulkov model play in mimicking neuron input-output behaviors. Just as we do not assign physiological significance to the parameters and variables of the Rulkov model, so also we do not assign any causal signification to "probability" variables in a statistical model.

Region I systems are systems said to be accurately described "on the basis of physical laws." One could say that these "physical laws" along with particular constraints *constitute the model*. If we choose to suppress the details of some of the physical variables by making an abstract model, this is a matter of practical convenience or practical necessitation under Weinberg's "Square Law of Computation." Scientists are often fond of saying of their reduced models that "in principle" we could put the abstracted variables back in and get the same result if only we were willing to undertake the labors of calculation. In fact, such "in principle" confirmation is rarely carried out, and often it is the case that the "in principle" calculation might require many lifetimes to actually accomplish. When a physicist says, "Physics explains all of chemistry," he doesn't mean physicists have actually predicted or explained every single chemical phenomenon "from first principles." What he really means is, "I'll be flabbergasted if anyone ever discovers one single case where physics predicts one thing and chemistry behaves otherwise."

Region II systems are systems said to be accurately described "on the basis of laws of probability." Now, the "laws of probability" are entirely mathematical inventions and they take nothing more from physical experience than the incentive to have developed these laws in the first place. They derive their wide-ranging scope of applicability and their honored place in the arsenal of scientific methods from their ***object independence***, i.e. from their success in being applied to a great many physical circumstances irrespective of the "physical basis" for the observables they are used to describe. For Region II systems the "laws of probability" along with particular constraints constitute the model. The one and only thing that "justifies" the use of a statistical Region II model is success. If the model successfully describes and predicts the phenomenon to which it is being applied, that phenomenon is *said to be* a Region II system. If it does not, then the phenomenon is said to be "not sufficiently unorganized in its complexity" for a statistical approach to be successful to a satisfactory degree. Science is pragmatic. If a model works, we use it. If it does not, "the phenomenon is an area of active research."

But what is meant by saying a system is one of "unorganized complexity"? Again, the meaning of this phrase is pragmatic. In order to meet the practical dictates of the "Square Law of

Computation" the parameters and variables in a Region II model must be sufficiently uncoupled, when given mathematical expression, for the equations of the model to be practically solvable. In this sense, Schrödinger's "Square Root of *N* Law" is a consequence of probability theory, and *conformity to the expectations* of this "law" is the telltale indicator that the condition of "unorganized complexity" is being met with in the object of the system. Again, it is the fact that deviations from expected values in a Region II model are small enough *to make the predicted averages useful* from whence Region II models draw their power and fecundity.

Now, this rather Platonic state of affairs is a source of great philosophical discomfort for many people, including many scientists. Scientists like "causal explanations" when they can get them, but the "laws of probability" utterly lack any symptom of "causal connection." They are "non-deterministic." This brings us to the subject of the role of statistical mechanics in Region II models. Physicist-turned-philosopher Henry Margenau wrote,

> The systems of thermodynamics are ordinary objects: solids, liquids and gases with definite boundaries. The observables of interest are somewhat more remote from direct perception than the visual properties on which mechanics concentrates attention. Temperature, pressure, and entropy lack the intuitive immediacy of positions and velocities. They are bound to Nature by more extended and more complex correspondences and lead to concepts that are more abstract. All this makes the problem of explanation in thermodynamics rather unique and gives it some features which form a bridge with quantum mechanics.
>
> In thermodynamics proper, observables are connected by what are sometimes called *empirical* relations, with the term empirical understood in a very limited and specific way. The "laws of motion" in this science are primitive equations combining the observables themselves. In a certain sense, the laws do not say "why" bodies behave thermodynamically as they do. While this connotes no defect of the methods of thermodynamics as a science, it nevertheless raises a question as to the possibility of other modes of explanation, of theories that "go behind" the phenomenologic structure of thermodynamics and its minimal assumptions. . . Now, clearly, the bodies of thermodynamics are also systems of mechanics, and it is indicated that one should inquire whether the laws of mechanics can produce, or at least simulate, the equations of thermodynamics.
>
> But, as in all transferences of a theory into a domain other than its native own, one meets here with an initial obstacle. Though the systems are the same, the observables of thermodynamics are quite different from those of mechanics. By no stretch of the imagination can a mass point or a rigid body be said to have a temperature or an entropy or to exert a pressure. A merger of the two theories therefore requires first of all a reinterpretation of the observables of one in terms of the observables of the other. . . The mechanical reinterpretation of thermodynamic observables and with it the reduction of all thermodynamic equations are performed in statistical mechanics. Neither reinterpretation of observables nor reduction of equations is possible by means of the proper tools of mechanics alone; additional constructs, not germane to mechanics and thermodynamics are needed to attain to these ends. . . Statistical mechanics is therefore a discipline in its own right, related to mechanics but operating with certain extra notions peculiar to itself [MARG: 268-269].

Margenau points out that an observable such as "gas pressure" is "possessed" by a thermodynamic system (that is, pressure *is* a primitive object in thermodynamics), but it is merely "latent" in statistical mechanics. This means that there is no one variable or parameter object in a statistical mechanics model that corresponds to "pressure." Rather, "pressure" becomes a defined

quantity through a mathematical transformation that converts "statistical mechanics objects" into a "thermodynamic object." While statistical mechanics does not actually restore "determinism" to the theoretical model, it does at least present the illusion of causality, that there is "some reason" for the system behaving the way that it does. It gives us a way to impose *rationalist principles* on what is otherwise merely an *empiricalism* of observable relationships. (This, too, is something the physicist means when he says quantum statistical mechanics "explains all of chemistry").

In this sense, statistical mechanics is a discipline for building a bridge between the "laws of mechanics" and the "laws of probability." But, Margenau warns us, it is vitally important that one always be fully aware of when he is reasoning about the objects proper to one science (e.g. temperature and thermodynamics) vs. the objects proper to the other (e.g. velocity distributions and statistical mechanics).

> Statistical mechanics is a link between thermodynamics and mechanics; it succeeds *almost* in reducing one to the other. The reason for its partial failure is in the need it has for introducing probabilities, quantities unknown among the concepts of mechanics. True, dynamical principles tolerate the use of probabilities in certain special problems, where the initial state of a mechanical system is not completely known. But even there they offer no help in generating probabilities; these must first be introduced by considerations of a nondynamical sort, for the competence of the laws of dynamics is limited to the *transformation* of one set of probabilities into another. . . Probability is a foreign element in mechanics; it does not evolve naturally in the application of these principles and must be injected into them from without if it is to function in mechanical description. Hence the failure of statistical mechanics in effecting a complete reduction of thermodynamics to mechanics [MARG: 280-281].

§ 1.3 Region III Systems

The objects with which neural network theorists must deal fall into the class of Region III systems, i.e. systems exhibiting "organized complexity." This means that these systems are too complex for successful treatment by a tractable number of deterministic equations, but they display too little "randomness" (they are "too organized") for successful *and tractable* application of stochastic principles. These are the systems that theoretical physicists and pure mathematicians alike banish to the provinces as "the concern of engineers and others interested in applications" of "fundamental" scientific and mathematical principles.

Not surprisingly, and befitting a situation that falls in between the two extremes represented by Region I and Region II systems, the methods employed in computational neuroscience are a blending of "mechanical" and "stochastic" methods. Why? Because no one has yet come up with another way of approaching the problem. The situation here is not unlike what the physicist does when he uses a "semi-classical" model of the electron to try to describe phenomena such as the ferromagnetism of bulk iron (another Region III system). Lacking tractability at the level of relativistic quantum mechanics (which is viewed as the "fundamental level of explanation" for

ferromagnetism), but too entirely foreign to classical mechanics to be successfully treated by that approach, "applied" physicists and materials scientists seek a middle ground of explanation. So, too, it is at present with the computational neuroscience of neural networks between the regimes of biological neuroscience on the one end and psychophysical neuroscience on the other.

Region II models are successful mainly because the overwhelmingly large number of factors thought to be "in play" (by statistical mechanics) largely cancel each other out and keep the variances from predicted mean values small enough and independent enough to be treated simply as "noise." But this "separation of variables" does not work too well in Region III. Here the variances are usually large and usually cannot be neatly separated out for treatment independently of the "average behaviors" of the system. The challenges imposed by organized complexity make the computational neuroscience of neural networks a most difficult area of research, and, at the same time, the most ripe for breakthrough discoveries in brain theory.

The much-sought-after "trick" to Region III model development in neuroscience is simple enough to state but, so far, difficult to accomplish. It is simply this: To find appropriate observables (the objects of the system) sufficiently simple in description to avoid the limitations of the "Square Law of Computation" and, at the same time, sufficiently reliable in predictive power that the statistical variances in parameters do not carry the model results far away from the actual behavior of the physiological object. We will be looking at a few examples of this a bit later on. All the while, the model must be fecund enough to return useful and accurate predictions of the emergent properties of neural networks. We must be prepared to endure the discomfort that usually attends increasingly abstract concepts of a mathematical nature, while at the same time not becoming so complacent about these concepts that we lose all touch with observation, experiment, and consequential interpretation.

Even if this scientific task was already an achievement of neuroscience – which it is not – it would be poor pedagogy to present the answer first before explaining the backdrop that makes it the answer. But since we are not yet in possession of a fully developed discipline for dealing with organized complexity, it seems best to begin with a kind of scouting expedition to explore the landscape into which we now begin to journey and to see the difficulties with which we must deal concretely. So, all the while bearing in mind what it is we seek to achieve with model order reduction in the region of organized complexity, let us begin with a brief examination of the environment in which the neuron operates.

§ 2. The Neuron's Contribution to Organized Complexity

We begin with examining how the neuron itself contributes to the complexity of the

environment in which it operates. In the previous chapters we have seen the model development for describing the behavior of the individual neuron. Neuroscience has been able to produce neuron models with relatively few descriptive parameters, these being tied to neuron physiology either directly or through a fairly modest and short series of abstractions. If these model parameters were found to be tightly distributed in their numerical values, we would be able to feel confident in taking a first important step on the path to a successful small-variance "Square Root of *N* Law" treatment of the neural network. Unfortunately, this is not the case.

Of the many thousands of different distinct biological species of neurons, we have hard data on the physiology of only a relative few. Even here the difficulties that attend identification of specific neuron species in the course of laboratory investigations lead to a practical situation where our available data tends to describe neurons in categories – such as "pyramidal cells of the neocortex" – rather than finer subdivisions of neuron classification. On the other hand, the data also shows that measurable differences for parameters affecting signal processing fall into ranges and this, along with the classification of RS, FS, IB, etc. signaling types, suggests that many of the functional differences among neurons in a network can be accounted for by appropriate selection of neuron types and parameter ranges.

Many of our hard numbers for neuron parameters are obtained from neocortical neurons. It is instructive to review the range data that has been obtained. Let us begin with parametric data for the three primary synapses involved in data flow signal processing by neurons. Table I summarizes review findings for AMPA, NMDA, and $GABA_A$ data for neocortical neurons [SEGE1]. Parameters g_{max} and T_{pk} are as defined in chapter 3 for the $g^{(\beta)}$ conductance equation. The space constant and electrotonic length parameters are dendrite parameters explained below.

Table I: Synaptic Parameters

Synapse type	g_{max} (nS)	T_{pk} (ms)	Space constant λ (mm)	electrotonic length (l/λ)
AMPA	0.1-0.3	0.3-1.0	0.2-1.0	0.2-2.0
NMDA	0.05-0.5	5.0-50	0.2-1.0	0.2-2.0
$GABA_A$	0.4-1.0	0.2-1.2	–	–

Table II: Neuron capacitance ranges

Neuron type	C_m (nF) minimum	typical	maximum
RS	0.09	0.51	1.97
IB	0.27	0.53	0.99
FS	0.06	0.21	0.84

The parametric values for synaptic conductance illustrate the considerable range of values observed in neocortical neurons. As previously noted in chapter 3, neurons in other regions of the brain can exhibit different maximum conductance values, such as in the cerebellar granule cells reported in [GABB]. One can generally expect smaller synapses to exhibit lower values of g_{max} because small postsynaptic densities contain fewer channel pores. However, synaptic area by itself is not a compete predictor of g_{max} because different species of membrane-spanning channel pores show a wide range in the basic conductance parameter g_p (HILL4: 401).

Most (but not all) dendrites lack the voltage-gated channels needed to produce an action potential. Thus, unlike the axon, the dendrite conveys signals passively from the synapse to the cell body [SEGE2]. Experimentally, it is observed that membrane potential is greatly attenuated and dispersed during this process. This phenomenon can be modeled using what is known as the cable model for dendrites, originally developed by Rall in the early 1960s. The net effect of the filtering action performed on EPSPs by the passive dendrite is captured by two dendrite parameters, the space constant (λ) and the electrotonic length ($L = l/\lambda$), where l is the length of the dendrite in mm. Although a cable model for a typical dendritic arbor rather quickly becomes quite complicated [SEGE1], a useful rule of thumb for approximating the attenuation of the EPSP for a synapse located a distance x from the soma is provided by [KOCH]

$$\frac{V_m(x/\lambda)}{V_m(x=0)} = \frac{\cosh(L-X)}{\cosh(L)}$$

where $X = x/\lambda$. However, even though the membrane voltage is greatly attenuated in traveling from a distal synapse to the cell body, most of the charge injected into the cytoplasm at the synapse does make its way to the cell body with very little leakage loss because the conductance within the cytoplasm is much greater than across the membrane wall in the dendrite [SEGE2]. Thus the attenuation in peak EPSC amplitude at the soma is accompanied by an increase in T_{pk} and a broadening of the current waveform which can often be well-enough approximated by replacing the expression for $G_{syn}(t)$ by a $g^{(\alpha)}$-function expression.

Physiologically, one would have every justification for modeling synaptic inputs by different conductance models based on the distance of the synapse from the soma. However, it is obvious that doing so increases by two the number of difference equations per neuron for each additional synapse conductance model. Thus, following this tactic carries the model more quickly into conflict with Weinberg's "Square Law of Computation." Another tactic is needed.

Neurons differ greatly in size, thus in surface area. This results in a rather large range of membrane capacitance values, C_m. As we saw in chapter 6, the ratio C/C_m is a scaling factor in

Wilson's models (where C = 1 nF was the value Wilson used in obtaining his VGC approximations). This ratio *must* be applied in Wilson's models to set the numerical values for the ligand-gated channel conductances given a physiological value for g_{max}. Table II summarizes ranges for C_m obtained from data presented in [McCO2]. In laboratory studies, measurement data is often biased in favor of larger neurons because these are the more likely to be impaled by the microprobes used by the experimenter.

Measurements of C_m are rarely direct. Instead, physiologists tend to measure membrane resistance, R_N, and time constants, τ_m, and compute the implied membrane capacitance as $C_m = \tau_m/R_N$. There is nothing wrong with this, but unfortunately most physiological studies do not report the correspondences between measurements of R_N and τ_m, thus making the minimum and maximum range values for C_m somewhat problematic. One can reasonably guess larger values of R_N implicate smaller-sized neurons, and therefore smaller values of τ_m (and vice versa for smaller values of R_N). This is because the highest density of membrane channels in the cell body tends to be concentrated most heavily in the trigger zone of the neuron, whereas C_m reflects the entire surface area of the soma. This assumption was made in arriving at the minimum and maximum values in Table II. As a rule-of-thumb, however, this assumption is certainly open to challenge and only more precise experimental reports will resolve the issue. The typical values in Table II are less problematic. What can be said in favor of the minimum-maximum estimates in Table II is that these values seem to be consistent with range values reported by those researchers who do explicitly report C_m values.

A Region-II-like approach to modeling the "typical" neuron in a neural network would attempt to arrive at "expected values" for the parameters listed in Tables I and II, and to use these to model all neurons of a given type (say RS-type) in the network. There are two principal objections to this. First, it is postulated by most experts that different regions of dendrite carry out different types of signal processing function, and if so this would argue against treating every dendritic synapse as a "typical" synapse. Over the past quarter-century, tantalizing evidence has mounted suggesting that the dendritic arbor should be regarded as a basic "computational unit" of the neuron, in contrast to an older view that made the neuron itself the basic unit of neuronal signal processing [MEL]. The question is by no means settled, but it is too important to dismiss.

The second argument against a classical Region-II-like model is the range of variances seen in synaptic parameters. (Presumably the range of variances in C_m is also important, but this has been much less well-characterized). Studies have shown that variations in EPSCs even within a single generic type of cell (e.g. Layer 5 pyramidal cells) show a very broad and rather Poisson-like distribution of values [SMET], [KOCH:404-405], a finding that suggests a similar distribution

about a Region-II-like "typical" synaptic conductance parameter. Variances in Region-II models used in other sciences can often either be ignored completely, because the variations about the expected value are small, or treated simply by adding a noise term to the model. But in the case of neural networks, which can and do respond very, very differently for relatively small changes in parametric values, these large variances become potentially crucial factors. To some extent, this source of variation is accounted for by distributions assigned to different neurons' synaptic weight vector, W, (see chapter 3), but a detailed quantitative study of how well this tactic matches up against physiological models is lacking at the present time. Neural network theorists generally rely on W to capture all the "main effects" of this variance, but the justification here is practical and historical rather than experimentally-based on a re-examination of this presupposition. Advances in our knowledge of neural physiology and advances in computing power over the past decade do make it possible to undertake such a re-examination.

§ 3. Anatomical Contribution to Organized Complexity

Suppose someone were to give you a job sorting pennies. You are presented with one million pennies, upon each of which is scratched one of an unknown number of symbols. Your job is to sort them into bins according to these symbols. To make matters more interesting, the pennies come embedded in dirt clods, with an unknown number of them in each clod, so you can't even see the individual pennies without extracting them from the dirt clods first. Does this sound fun? Figuring out how neurons are interconnected in the brain is something like this only harder.

We are not even close to possessing a "wiring diagram" for any of the various regions of the brain, and there is no prospect in sight that we will possess such a "wiring diagram" any time in the near future. But without one, how can we hope to model the collective behavior of neurons in a functional biological neural network? In the mid-1950s it was already known that the brain could be parcellated into various regions according to differences in local gross arrangements of cell bodies and myelinated fibers. There was also evidence in hand that specific functions are localized, or largely localized, to specific anatomical regions. It was also clear that the fine details of neuronal interconnections were staggeringly complicated.

§ 3.1 Random Neural Networks

It is none too surprising, therefore, that some of the earliest ventures into neural network research were carried out by using Region II statistical methods that had worked so well for other vastly complex systems. Perhaps the first of these attempts was made by Beurle in 1956, who introduced the idea of the *randomly connected* neural network [BEUR]. There was a spurt of

activity investigating this idea through the 1950s and 1960s, although experimental neurophysiologists lost interest in this after awhile. Significant papers on random network models continued to appear into the 1980s, and even today one occasionally still comes along.

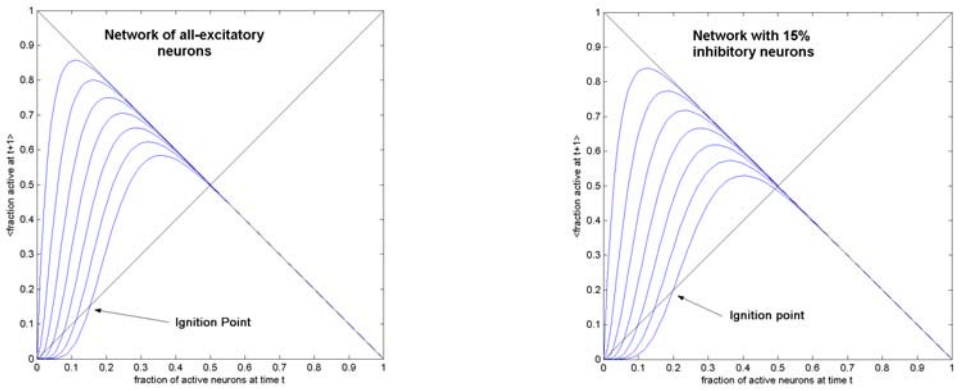
Aside from the fact that a randomly-connected neural network can perform no useful computations, these studies did turn up some interesting mathematical properties of random networks. These same properties also appear to rule out any possibility that random network models describe any sort of physiological reality in regard to brain function. Probably the most important finding has been that random network models do not seem to be capable of sustaining any low but non-zero level of neuronal activity.

Models using various probabilistic assumptions, neuron models of varying degree of biological plausibility, and a wide range of assumptions on how to represent neuron activity have been employed in randomly-connected network models. Abeles has reviewed a number of these in [ABEL1, chapter 5]. All have arrived, each in their own way, at more or less the same conclusion. What this conclusion is can be amply illustrated by considering the random network model of Anninos et al. [ANNI].

> In this paper we present the mathematical formalism, the methods used and some numerical results pertaining to a model of cerebral functioning which was discussed in the preceding paper . . . and will be referred to as I. The assumptions made there were, in brief, twofold: the structure of the neural net may be approximated by sets of discrete populations of randomly interconnected neurons; these were given the term *netlets*. The netlets are coupled to one another in a way which was described as *randomness-in-the-small* and *structure-in-the-large*. The second assumption concerns the appropriate description of neural dynamics. The statement was made in I that the spatial and temporal microstructure of activity may be disregarded. The dynamical variables considered significant in this model are the levels of activity, i.e. the fractional numbers of neurons firing in each netlet. With these two sets of assumptions the dynamics of some of the simpler prototypes of neural nets may be computed [ANNI].

Anninos et al. were able to obtain a difference equation (equation 6 of [ANNI]) that described the expected fraction of neurons in the network that would fire at iteration step $n + 1$ given that some fraction α_n were firing at iteration step n. A model of this sort is called a Markov process model, and is often called a renewal process. A stochastic model of this sort arises naturally from the assumption that all the information a neuron has of its past is summarized in the present value of its membrane potential. This assumption is consistent with our physiological models.

The details of the Anninos equation are not particularly important for our present purposes (the interested student can refer to [ANNI]) other than to note that the probabilistic expression came out to be a Poisson distribution. Nelken later showed in general, under conditions that should pertain to large random neural networks, that the statistics obtained for the excitatory and inhibitory signaling processes converge to those of the Poisson process even if synaptic inputs are weakly correlated [NELK1].



(A) (B)

Figure 7.1: Activity levels at time step $t + 1$ given activity level at time t predicted from the Anninos function for randomly-connected neural networks. (A) Random network in which all neurons are excitatory. (B) Random network in which 15% of all neurons are inhibitory. The families of curves (left to right) are for increasing firing threshold levels. Points where the activity curves cross the unit slope line ($\alpha_{t+1} = \alpha_t$) are fixed points. Ignition points are points of unstable equilibrium. Above the ignition point the stable fixed-point solutions predict nearly 50% of all neurons will be firing at any given time. Below the ignition point the stable fixed-point solution is the zero-firing rate solution. The average number of synapses per neuron was 30 for both excitatory and inhibitory synapses, and thresholds range from 1 to 7 (left to right).

Solutions for the Anninos function are shown in Figure 7.1 for two different example cases.[3] Figure 7.1A depicts a random network made up of all excitatory neurons. Figure 7.1B depicts a case where 15% of the neurons in the network are inhibitory (which corresponds to the percentage found in neocortex). Let α_t be the fractional activation, i.e. the fraction of the neurons in the network firing at time index t. Let $F(\alpha_t)$ be the Anninos function giving the expected value of α_{t+1} at the next time step. Families of curves for F are shown in Figure 7.1 for different values of average firing threshold for the neurons. On the average, then, $\alpha_{t+1} = F(\alpha_t)$ at each time step.

The network dynamics reach a fixed point solution if, for some t, $\alpha_{t+1} = \alpha_t$. Thus, fixed points are identified by the intersection of the $F(\alpha_t)$ curve with the line $\alpha_{t+1} = \alpha_t$ shown in the graph. There are two types of fixed point solutions we must consider. At a *stable* fixed point, if there is a small perturbation ε made to fixed point α_t, the system will return to α_t in a finite time, i.e.,

$$\alpha_{t+1} = F(\alpha_t + \varepsilon),\ \alpha_{t+2} = F(\alpha_{t+1}),\ \ldots,\ \alpha_{t+n} \rightarrow \alpha_{t+n-1} = \alpha_t$$

for some finite n. In contrast, at an *unstable* fixed point, the system response to a perturbation evolves as

$$\alpha_{t+1} = F(\alpha_t + \varepsilon),\ \alpha_{t+2} = F(\alpha_{t+1}),\ \ldots,\ \alpha_{t+n} \rightarrow \alpha_{t+n-1} \neq \alpha_t.$$

An unstable fixed point is called an ***ignition point*** and separates two different stable fixed points

[3] These curves were generated using a MATLAB® script written by T. Trappenberg [TRAP: 299].

for the Anninos function. Ignition points are identified by arrows in Figures 7.1.

As the average firing threshold is increased, $F(\alpha_t)$ moves down and to the right in Figure 7.1. For a sufficiently large average threshold, $F(\alpha_t)$ lies entirely below the $\alpha_{t+1} = \alpha_t$ line except at the point $\alpha_t = 0$. This ***zero activity*** solution is always a stable fixed point for $F(\alpha_t)$. If the average firing threshold is low enough such that an ignition point exists, then $F(\alpha_t)$ has two stable fixed points, one at $\alpha_t = 0$ and another at $\alpha_t \cong 0.5$. Any stimulation of the network that places α_t above the ignition point for some t will therefore carry the network off to a steady-state condition in which approximately half the neurons are firing at every time step. Conversely, if α_t is below the ignition point for some t, the network is expected to evolve to a steady-state in which *no* neuron is firing. *Neither case is representative of actual brain behavior*. We might liken the zero activity solution to "brain death." Metaphorically speaking, the randomly-connected-network model predicts either brain death or an epileptic episode as the only stable states of brain function. The actual biological situation is one in which average neural network activity level persists at some low *but non-zero* fraction of the neurons in the network.

Anninos et al. were also able to derive $F(\alpha_t)$ expressions for the case of a constant, non-zero level of average stimulus activity applied to the network. In these cases a low, non-zero steady state was possible. However, this merely begs the question. How does the network stimulating the modeled network obtain *its* biologically-realistic average firing activity level?

One might argue that perhaps the neuron models used by Anninos et al. were too simple and this might have led to the unrealistic results predicted from the model. This possibility has been considered by other investigators, whose findings have been reviewed by Abeles [ABEL1, chapter 5]. Findings obtained from different randomly-connected network models have been amazing consistent: The only two stable states of expected fractional activity are the zero activity and the $\alpha_t \cong 0.5$ "epileptic" activity levels. Abeles concludes,

> For a large network of excitatory and inhibitory neurons with small EPSPs it is very difficult, if not impossible, to attain steady ongoing activity at low firing rates. The problem of attaining a low level of ongoing activity is due to the steep slope of the input-output curve of the excitatory population at low firing rates and to the extrasynaptic delay required before the inhibitory cells can counteract random fluctuations in activity. If it is at all possible to attain stable ongoing activity at low levels, it is likely that the range of fluctuations around which the system can be stabilized will be very limited, that the inhibitory neurons will have to switch from no activity to maximal activity over a narrow range of excitatory firing rates, and that the inhibitory neurons will also have to exert strong inhibitory effects on themselves [ABEL1: 167-168].

The conditions that have to be presumed to obtain mathematical descriptions such as the Anninos function are precisely those which Weinberg called *unorganized complexity*. Region II models can all be expected to exhibit similar probabilistic functions (Poisson-like distributions and renewal process dynamics), and so the implication of Abeles' conclusion is rather obvious.

The basic nature of biological neural networks is not one of unorganized but, rather, *organized* complexity. They are Region III systems. The behaviors Abeles describes that must, minimally, be exhibited by the inhibitory neurons in the network are conditions that somehow the organization of the neural network must produce. This is something no known scheme for a randomly-connected network can achieve.

Although the Region II modeling approaches to network organization fail to produce biologically-realistic network behavior, it is a mistake to conclude this research is without value. On the contrary, the results obtained from randomly-connected network models provide an extremely valuable result. This research has told us what the organization of the nervous system *is not*. The crucial shortcomings of the model lie with fundamental mathematical properties inherent in this type of network "un-organization" and it is not likely these fundamental mathematical issues can be fixed by tinkering around the edges of this or that probabilistic assumption (while maintaining probabilistic characteristics that *would have to apply* to real neurons).

The sort of presumptive changes that would be needed to overcome these mathematical problems inherent in the randomly-connected neuron model also provide us with a glimpse of anatomical consequences that would likely have to attend the process of fixing the model. Here we find some very tantalizing correspondences with anatomical facts. Abeles goes on to note,

> The delicate balance between the internal excitation and inhibition in large networks sheds light on some of the anatomical detail of the cortex. The need to switch from no inhibition to fast inhibition explains why only few inhibitory neurons having very strong effects are found in the cortex. The need to have the inhibitory feedback with as short a delay as possible explains why there is no use for extracortical inhibitory feedback (conduction times to such subcortical nuclei and back are too long). This requirement for a short delay also explains why the axons of most inhibitory neurons are thicker than the average cortical axon (fast conduction times), why they distribute their effects mostly in their own vicinity (short conducting distances), and why their postsynaptic targets are concentrated on cell bodies and the proximal dendrites (shorter electrotonic delays to the soma and faster rise time of the EPSP).
>
> The synaptic arrangement in which excitatory neurons receive a mixture of excitatory and inhibitory synapses on their somata (and large dendritic trunks), whereas the inhibitory neurons receive excitation only on more remote dendritic sites, also contributes to fast inhibitory feedback and more sluggish positive excitatory feedback. The effective mutual inhibition that is exerted by the inhibitory neurons on each other also seems necessary for attaining stability.
>
> The conjecture that the inhibitory feedback can stabilize the ongoing activity only over a narrow range of firing rates sheds light on the arrangement of thalamic inputs to the cortex. These inputs bring strong excitatory effects to the cortex (particularly to sensory areas). The thalamic inputs spread their terminals in cortical layer IV, where they make synaptic connections with all the neural elements that pass through the layer. . . Layer IV is particularly rich with inhibitory neurons. In this manner, the input that increases the activity in the cortical excitatory neurons concomitantly activates the inhibitory neurons. This is a feed-forward activation of the inhibitory neurons that anticipates the increased excitatory activity and alleviates the problems that the extra delay of inhibitory feedback might cause [ABEL1: 168].

Anatomists had, of course, documented these synaptic and projection features of neocortex

long before Abeles drew out the conclusions stated above. (And it is appropriate to note here that excitatory synaptic connections to the somata of excitatory neocortical neurons is very, very rare). But knowing how the neocortex is generally arranged is not the same as having an understanding of the functional significance of this arrangement. This is one thing the randomly-connected network models seem to have provided to us. Naturally, conclusions and imputations such as those Abeles draws above are easier to make in hindsight; the mathematical results did not predict the anatomical arrangements seen in the laboratory. Still, the incompatibility of the model with what is known of cortical arrangement, and the concordance between what goes wrong with the model and an anatomical structure which at least grossly agrees with what is needed to eliminate the models' undesirable mathematical features, does serve to tell us we must look elsewhere for a method of model order reduction.

§ 3.2 Crystalline Neural Network Models

It is a mistake to think, as Weinberg's pessimistic appraisal might seem to imply, that Region III models have never been successfully developed in any branch of science. The example *par excellence* of a successful Region III model is provided by solid state physics and its models of materials composed with crystalline structure.

A ***crystal*** is defined to be "a solid whose regular array of particles has definite polyhedral faces meeting at definite angles and showing certain symmetry characteristics." The symmetry characteristics of a crystal are what make possible the simplification and solution of the partial differential equations that describe the physics of crystalline materials. More particularly, crystal structure gives rise to ***periodic boundary conditions***, and these are the key to obtaining solutions describing the material. Crystalline solids have ***organized complexity***, and the organization is sufficiently regular to permit successful model order reduction. Of course, perfect *single* crystal materials are fairly rare in nature, and most are carefully fabricated in the laboratory or in factories whose end product is such a material. Most crystalline materials are ***polycrystalline***, which merely means they are composed of many crystals (often called ***grains***) separated by what are generally called ***grain boundaries***. Exact solution of the underlying physics equations is usually not possible, but approximate solutions are often attainable through perturbation methods. These solutions typically are accurate to about 70 to 80% in predicting material properties (or, to put it another way, quantitative prediction errors are typically less than 20 to 30% compared to measured values). This is better than having nothing at all.

Now, obviously, biological neural networks are not crystals. They could not really even properly be called "solids" unless one wishes to regard a gel-like material as "solid." It is true, of

course, that neural anatomy does show particular geometrical features (such as tissue texturing, columnar and "blob" structures, etc. [WHIT]) faintly reminiscent of crystal structure. These geometrical and morphological features are part of what helps the anatomist classify different parts of the brain and spinal cord. But these geometrical features carry no mathematical imputation for possible mathematical simplifications of the sort enjoyed by solid state physics.

But, as any good mathematician worth her salt would tell us, physical scientists often do not think "in sufficient generality." Most of us would be indignant if a mathematician were to tell us we do not know what "distance" is. (And, of course, most mathematicians are too polite to put it so bluntly). We think: Slap a ruler down between two points, read off the calibrated markings, and *that* is "distance." The mathematician would tell us, "That is merely a special case example of a metric function in a particular metric space." Mathematicians long ago came up with a more generalized definition of "distance" and so today we have Euclidean distance, Hamming distance, and an endless repertoire of other "metric functions" that bear no names most of us would recognize.

Is it possible, then, to come up with more generalized ideas for things like "crystal" that might do for our ability to describe neural networks what the crystalline symmetry properties do for solid state physics? And might such a more generalized idea make it possible to achieve model order reduction for neural networks? Put another way, can we generalize the idea of a "crystal" in a way useful to neuroscience? There is reason to think the answer to this question is "yes"; furthermore, although rarely so called in the corpus of neural network literature, one can argue that this is already being done. A critic might argue it is being done piecemeal, but piecemeal or not the important idea is that this is how many types of models in neural network theory can be viewed. We shall call Region III models of this sort ***crystalline neural network models***. By analogy, the randomly-connected network model described in the previous section could be called a "glass" model. (A "glass" is an amorphous solid, the geometric opposite of a crystal).

System theory boldly claims to be "the science of systems in general." Since almost the very definition of "science" is that it is an organized doctrine of knowledge[4], a somewhat tongue-in-cheek way of stating the "first principle" of system theory is, "Everything is the same, only different." What this is meant to convey is: By some proper way of looking at any object, there is a way to describe it such that its description is mathematically homomorphic with the ways in which we describe other objects. A less glib description might be to say that system theory relies

[4] Immanuel Kant, the great 18th century philosopher, defined "science" as "a doctrine constituting a system in accordance with the principle of a disciplined whole of knowledge." His epistemological definition of "system" was "the unity of various knowledge under one Idea." Although nowadays it is mostly forgotten, Kant was the first to draw a clean distinction between "science" and "natural philosophy."

on the fact that the same equations have the same solutions. It is just a question of, as a system theorist might put it, "getting things to look the same." How shall we look at the notion of a "crystal" so that this more generalized notion can pertain to neural networks?

§ 3.2.1 How Crystalline Structure Reduces Physics' Models. Compared to the solid state theory of crystals, the solid state physics of glasses is extremely formidable and nowhere near as universally successful. This is not to say there is no theory of glasses at all. There is, and it is a field in which very dedicated specialists labor. However, it is not, by any honest reckoning, as generally well-developed and successful as the solid state physics of crystalline materials. What does crystalline structure do for physicists? How does it make possible the substantial model order reduction that is solid state theory?

Put vernacularly, the basic answer is that regular crystalline structure makes possible a "divide and conquer" approach to modeling the solid. The atoms in a crystalline solid are arranged in a lattice made up of a symmetrical arrangement of repeated three-dimensional geometric unit cells, each of which contains a small number of atoms. The unit cells divide the solid into spaces of equal volume with no space excluded. Each corner of a unit cell is called a "lattice point." Every lattice point has identical surroundings with every other point. There are only fourteen possible space-filling networks of lattice points, and the arrangement has translational periodicity. This periodicity ensures that the Schrödinger equation, which describes the sum of kinetic and potential energies in the crystal, has only periodic solutions.

Even so, an exact solution for the Schrödinger equation in a solid cannot be obtained by any presently known method of analysis. Physicists therefore resort to approximate solution methods. For solids with ionic or covalent bonds, the approximation method is called the linear combination of atomic orbitals (LCAO), which uses the known solutions of the Schrödinger equation for the hydrogen atom as "basis functions" for constructing approximate solutions. For simple metals, the approximation uses what are known as plane wave methods. Using approximation methods, the chemical bond can be worked out for any solid [HARR].

> The Solid State Table of the Elements, folded into the book near the back cover, exemplifies the united view of electronic structure which is sought, and its relation to the properties of solids. The table contains the parameters needed to calculate nearly any property of any solid, using a hand-held calculator; these are parameters such as the LCAO matrix elements and pseudopotential core radii, in terms of which elementary descriptions of the electronic structure can be given. The approach used throughout the book has been to simplify the description of the electronic structure of solids enough that not only electronic states but also the entire range of properties of those solids can be calculated. This is always possible; the only questions are: how difficult is the calculation, and how accurate are the results? . . . [The] simplified approaches explained in this book, although they give only tolerable descriptions of the bands, can easily be applied to the entire range of dielectric, transport, and bonding properties of imperfect as well as

> perfect solids. In most cases they give analytic forms for the results which are easily evaluated with a hand-held calculator [HARR: xiii-xiv].

It is no doubt obvious that any model which can be computed using a hand-held calculator must be quite simple in its mathematical form. In effect, this modeling schema permits the physicist to divide up the entire very complicated solid into a small number of interacting subsystems, each of which has a Region-I model, and then describe the entire solid by computing the interactions among them. Harrison's "Solid State Table of the Elements" provides the factors needed to compute these interactions, incorporating the phenomenological "correction terms" that compensate for the approximate nature of the solution.

The problem of organized complexity in a crystalline solid is thus attacked by taking the following steps. ***First***, a prototype system that has a Region-I model is selected to serve as the basis for analysis. In the case of the crystalline solid, this prototype system is the hydrogen atom with its known analytical solutions for the possible orbitals its electron can occupy. ***Next***, other more complicated, but still relatively simple, systems are considered in terms of the basis prototype. In the case of solid state physics, these are the other atoms. It is found that only certain orbitals, known as the valence orbitals, need be considered in modeling the solid. (This is where the s-, p-, d-, and f-orbital terminology, familiar from freshman chemistry, come from). The other orbitals, those forming what are called the "closed shells" of the atom, can be ignored. ***Third***, the arrangement of the constituent systems (the particular atoms in the lattice) is developed. This arrangement differs for different materials, which gives the solid to be modeled its specific organizational character. The solid is said to have a "hexagonal close pack" or a "face-centered cubic" or etc. structure.

Fourth, the knowledge of the structure (hexagonal close pack, or whatever) is combined with the basis representation (the particular valence orbitals involved) to *qualitatively* determine how to treat the interactions among the atoms in the lattice. This is where model ideas such as the "sp^3 hybrid orbital," familiar from freshman chemistry, are obtained. From this treatment a new basis set for representing the solid is obtained from the original basis functions. In chemistry this is referred to as the "molecular orbital." Molecular orbitals are obtained as modified versions of the original atomic orbitals (in the case of the LCAO method). This step is an important one that deserves further comment. The basis set for describing the solid is not the same as the set used to describe the isolated atom. It cannot be, because a molecule is not just a pair of atoms in proximity. There are interactions between them that alter the orbital solutions, and these interactions are responsible for the emergent properties of the molecule – that is, the things that are characteristic of a molecule that are not characteristic of any of its constituents considered in

isolation.

Fifth, schematic parametric expressions and specific quantitative parameter values are found that provide, with acceptable accuracy, the *quantitative* mathematical description of the interactions. It is necessary to find such expressions because the modeling technique employs approximations in lieu of unobtainable exact solutions of the Schrödinger equation for molecular orbitals. In Harrison's book [HARR], he shows that only a surprisingly small handful of mathematical expressions are needed to describe any solid. These expressions are *independent of the particular atoms that go into the solid* (although which expressions one uses does differ between, say, a solid with covalent bonding vs. one with metallic bonding). In our terminology, we say these equations constitute a ***modeling schema***. The equations contain unspecified parameters, the values of which do depend on the particular atoms present in the solid. His "Solid State Table of the Elements" provides the specific values for each of these parameters for every atom in the periodic table of elements. These values were obtained from experimental studies conducted over many years in Harrison's laboratory at Stanford.

Sixth, a mathematical description by which the molecular orbital functions are combined in a single system is found. This equation reflects *the unit crystalline cell*. In the case of solids, the mathematical form most often used is not the Schrödinger equation in the form set down by Schrödinger but, rather, the mathematically equivalent "linear algebraic" form developed by Heisenberg (and known, appropriately, as the Heisenberg matrix formulation). In solid state physics this form is called a "Hamiltonian." It expresses the sum of kinetic and potential energies of the system, just as the Schrödinger equation does, but does so in a mathematical form more efficient for the expression of the system of model equations.

This modeling schema is one of the most successful ever developed for Region-III systems. Within a few years of the publication of Harrison's book, researchers in places such as the Stanford Research Institute were using this schema to develop new semiconductor devices, many of which have played very important roles in advancing microchip technology in the past quarter century. Of the many important features contained in this modeling schema, one of the most useful and appreciated benefits is the fact that the schema itself provides a way of visualizing what is going on inside a solid, so that the researcher does not have to rely solely upon abstract mathematics. This goes directly to the heart of, as Harrison put it, "learning the physics of the system (or 'learning the chemistry of the system,' if one is of that background)."

§ 3.2.2 The Notion of Crystalline Neural Networks. We will call any neural network model that can be developed by a sequence of analysis steps analogous to the six steps explained

above a "crystalline neural network" model. That such an approach is even possible for neural networks is perhaps not self-evident, and so some discussion of this is necessary. What features of biological neural networks are there which can inform the approach and establish its feasibility and, more importantly, its biological significance? After all, anyone can put together an arrangement of unit netlets in a "step-and-repeat" crystal-like geometry. This is merely a mathematical exercise. The network model used in [RULK1] is an example of an interconnection of Rulkov neurons organized in just such an arrangement. This is not the point. The important facts are: (1) there are orders of magnitude more neuron types than there are elements in the periodic table; (2) there is no single "fundamental law of physics," on par with the Schrödinger equation, for describing neurons, much less neural networks; (3) qualitatively, the interactions and connections between neurons in a biological network are many times more numerous than those presented to the solid state physicist; (4) parametric *variances*, such as in synaptic conductances and membrane capacitances, are far, far greater than their parametric counterparts for the atoms in a crystalline solid; and (5) there are no comparable experimental techniques, on par with methods such as the x-ray diffraction analysis used by material scientists and physicists, for examining neural network organization and structure. Why, then, should one entertain any hope at all that any kind of *useful* methodological approach, analogous to that used for crystalline solids, even exists? It must be admitted at the outset that the organized complexity of a biological neural network is far more complex than that of a mere crystalline solid.

On the other hand, neural networks are far, far less sensitive to minute variations in "atomic" constituents (neurons, in our case) than are many kinds of solids. For example, addition of trace amounts of boron or phosphorus to a silicon lattice profoundly alters the electrical properties of the entire solid (without altering its metallurgical properties significantly); this is how transistors and diodes are made. The crystal-wide impact of relatively minute amounts of dopants is due to the *global* nature of the Schrödinger equation. Neural networks, by contrast, have very few factors that exert a global effect on the network; those which do are for the most part due to readily identifiable constituents of the biological structure (e.g. the glial syncytium). Whereas inhomogeneities such as grain boundaries affect the entire solid[5], inhomogeneities in neuronal structures have only local effects on the form of the modeling equations for interactions. This localized vs. globalized property difference is one thing we have going in our favor.

Another thing we have going for us is that nature shows a tendency to repeat certain structural

[5] The magnetic coercivity (an important parameter for materials such as iron) can be three to four orders of magnitude less in many polycrystalline iron materials than it is for small single-crystal iron samples. The very nature of the magnetic behaviors of these two materials is not only quantitatively but also qualitatively different. Part of these differences has to do with grain size, but a large part of it is due to grain boundaries.

"themes" in neuronal organization. One of the most studied of these "themes" is the ***functional column organization*** of the mammalian cerebral cortex.

> The notion that the cerebral cortex of different species is built according to some common, basic plan is woven throughout the following treatise. Substantial evidence in support of this thesis has surfaced regularly since the outset of anatomical studies of cortical organization . . . and has been confirmed by the many experimental approaches that have been used to elucidate different functional aspects of the cerebral cortex. For instance, direct electrical stimulation of the brain or recording of evoked potentials from it have established that the cerebral cortex can be parcellated into functional areas whose relative positions are similar in all mammalian species[.] Subsequently, recordings of single-unit activity have confirmed the existence of functional columns in various areas of the cortex in the mouse, rat, cat, monkey, and man . . . and have provided evidence that cortical neurons in a variety of species display similar functional properties [WHIT: 1-2].

If neurons play the role of "atoms" in a neural network model, then functional columns are a prime candidate for the role of lattice analogue – perhaps even "unit cells," but at least structures composed of unit cells (so-called ***minicolumns***). If this idea proves successful, then it brings down the degree of organized complexity by a gigantic amount – from the complexity of a system composed of hundreds of thousands or millions of neurons to one with "only" on the order of hundreds or thousands of neurons. This is, clearly, still a formidable problem, but one which is now coming within range of the computational capacity of today's extremely powerful computer technology (overcoming the "Square Law of Computation" problem).

This idea of a "basic plan architecture" for neural networks is not unique to cerebral cortex. Another much studied brain region, the cerebellum, displays its own highly regular layout within the cerebellar cortex. In detail it is quite different from the organization of the cerebral cortex, but the point is that it does apparently have the sort of "lattice-framework" regularity and "functional circuit" organization the modeler can exploit [LLIN]. The thalamus, a subcortical structure through which passes almost all the peripheral information coming into the neocortex, has its own brand of regular organization, referred to by anatomists as "nuclei" rather than "columns" [SHER]. As more is learned about the various brain regions, there is a rising hope, almost amounting to expectation, that parsimony of organization schemata in the central nervous system is more the rule than the exception.

There is yet another advantage we have working for us. In recent years hard evidence has been found that firing patterns of neurons within closely connected ***cell groups*** are highly, highly correlated [STER], [FRIE1-2], [ECKH1-2], [GRAY2], [KREI]. Stimulating action potentials from just a few members of the neuronal cell group leads to the firing of action potentials by all or nearly all members of the cell group. What this means is that the signals from the different neurons in a cell group are highly ***synchronized*** in time. It is also thought likely that subgroups within the cell group – and perhaps sometimes even the entire group – all project to the same

target areas (other cell groups). When this is the case, the actions of the entire group, which might consist of hundreds or more neurons, can be represented by just one or a few members of the group. Put another way, a few model cells might proxy for the entire population of cells in the group. This, of course, is clearly analogous to the "unit cell" of solid state physics.

Aside from the physiological evidence supporting the idea of the cell group, there is also a very practical reason, in evolutionary terms, for why the central nervous system would be organized in terms of cell groups. Neuron cells die, with more or less regular death rates, and, unlike other cells in the body, they are not replaced. The statistics are fairly grim. If the central nervous system were not organized with a great deal of redundancy in the signal processing functions carried out by its neurons, no human being would be likely to live past childhood.

Beyond the biology of the central nervous system, there are mathematical grounds as well in support of the cell group hypothesis. Perhaps the earliest mathematical arguments for this were put forth by Christof von der Malsburg in 1981. Malsburg's theory, regarded as radical at the time, is known as the "correlation theory of brain function" [MALS2]. Today it has become widely accepted, at least in part, by many theorists. It has received further support from neurological studies, carried out by Damasio and his co-workers, for explaining a number of findings on the effects of various forms of brain damage caused by injury, stroke, tumor, and surgery [DAMA1-2].

When one combines the two notions of local functional organization and highly synchronized firing behaviors from cell groups, we arrive at what is often termed the idea of the ***functional microcircuit***. Although functional microcircuits might vary parametrically from one to the next, the implication of this modeling approach is simple and important: Models of neural networks can be approached by representing the behaviors of large numbers of neurons by a relatively small number of proxies, locally interconnected to form neural processing units, and linked to other neural processing units by synaptic pathways. The steps necessary for constructing such a model are: (1) identifying the membership of the neuron population making up "kernels" within a subunit; (2) identifying the key features of neuronal signaling by which information is conveyed within the subunits; (3) constructing the ***local architecture and topology*** of the different subunits; (4) determining the connectivity forms for interactions between subunits making up the functional minicolumn, and the connectivity forms between different processing units; (5) determining the quantitative mathematical descriptions for these interconnects; and (6) studying the gross behavior of the network so constructed to simplify and reduce its mathematical description as much as possible without significantly altering the behavior of the network as a whole. These are the steps closely analogous to those presented earlier in describing how the solid-state physicist

accomplishes his modeling task, and so we can see that this indeed constitutes what has here been called a crystalline neural network model. A specific example will serve to further clarify and illustrate these general ideas.

§ 4. Networks in the Neocortex

In neuroscience the neocortex is often accorded a status somewhat like that of a rock star. It is thought to be the principal brain region for all the "higher" cognitive functions in mammals. As such, its study has always attracted much attention. Along with the hippocampus and the cerebellum, the general organization of the neocortex is one of the best understood of the major functional areas of the brain.

Figure 7.2 illustrates the general column layout, interconnectivity, and afferent pathways of the neocortex as it is presently understood [DOUG1, 3], [WHIT]. Although the major interconnect pathways shown in the figure are well documented, the internal circuitry connecting

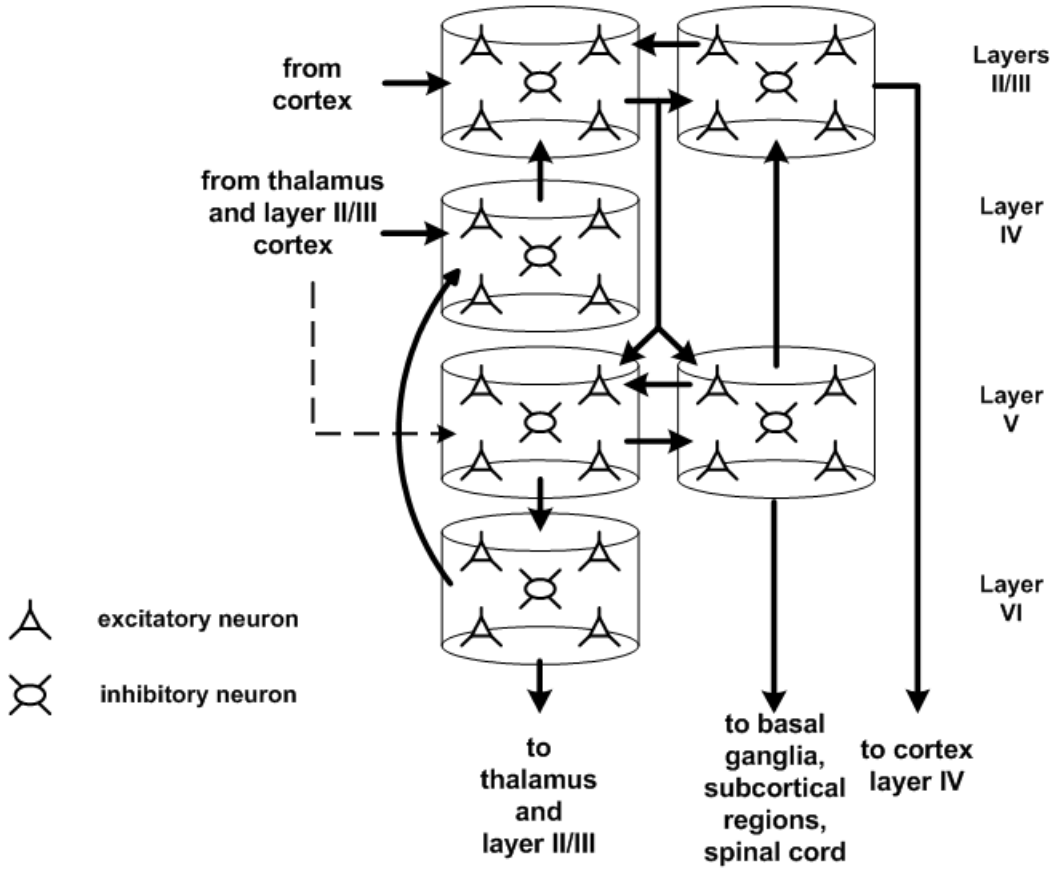


Figure 7.2: General column layout, interconnectivity, and principal afferent and efferent pathways of the neocortex. Efferent projections at the bottom of the figure are by means of the white matter below layer VI. Lateral (left-right) connections shown also continue reciprocally into and from adjacent cortical columns. Afferents at the left side of the figure primarily enter via white matter axons.

neuron-to-neuron within a column is much less well understood in detail, especially in regard to the population of inhibitory neurons. Anatomists have uncovered some general themes that appear to operate at the circuit level in neocortex. These can be summarized in six statements, which we will call ***White's rules*** [WHIT: 82, 157-158]:

1. Every neuron within the target region of a projection receives input from the projection.
2. Different dendrites of a single neuron form similar synaptic patterns; that is, the numbers, types, proportions, and spatial distribution of synapses is similar, provided the dendrites are exposed to similar synaptic inputs.
3. Neuronal types receive characteristic patterns of synaptic connections; the actual numbers, proportions, and spatial distribution of the synapses formed by each neuronal type occur within a range of values.
4. The receptive field properties of every cortical neuron are shaped by the spatial and temporal integration of inputs from a variety of excitatory and inhibitory sources. Inputs from a single source cannot be the sole determinants of the receptive field properties of cortical neurons.
5. Only a fraction of the synaptic inputs to a cortical neuron are activated at one time. Therefore, the receptive field properties of cortical columns are transitory and are determined by the cortical circuitry active at a given time.
6. Excitatory and inhibitory synaptic interactions between cortical neurons preferentially link neurons situated in close proximity to one another, and these interactions typically link neurons having similar receptive field properties. Synaptic interactions between closely spaced neurons, having similar receptive field properties, provide a basis for the similarity of receptive field properties of neurons within a functional column.

In addition to these six rules, we also have three ***White's corollaries***:

1. Axon terminals from any extrinsic or intrinsic source synapse onto every morphological or physiological neuronal type within their terminal projection field. In practice, this means that a pathway will form synapses with every element in its target region capable of forming the type of synapse normally made by the pathway (i.e., excitatory or inhibitory).
2. Axonal pathways form similar synaptic patterns onto all the dendrites of a single neuron, provided the dendrites occur within the target region of the axonal pathway.
3. Different extrinsic and intrinsic synaptic pathways form specific proportions of their synapses with different postsynaptic elements (spines vs. dendritic shafts, one cell type vs. another).

Numerous studies have shown that cortical neurons sharing similar receptive fields are arranged vertically in columns [WHIT: 109-112]. Thus, the structure illustrated in Figure 7.2 can be provisionally regarded as the prototype network structure for a functional column. However, a certain degree of caution is merited in considerations involving functional columns. Studies have

demonstrated that functional columns fall into two broad classes: (1) functional columns with a well-defined anatomical structure; and (2) functional columns that lack clearly defined anatomical substrates [WHIT: 199-201]. Anatomical structures corresponding to the first class bear such names as barrel cortex, slabs, and blobs. Columns of class (2) appear to be of a more transitory and dynamical nature, as suggested by Rule 5 above. In this context, we may also observe that the lateral connections (left-right) depicted in figure 7.2 can allow neurons to "overlap" such that a particular neuron might at one moment in time "belong" to one functional column, and at a different moment in time belong to another column, depending on its receptive field and the active stimuli it is receiving. Thus, class (2) functional columns can be said to exhibit "dynamic links" and support what Malsburg calls a "dynamic link architecture" (DLA) [MALS3].

This dynamical characteristic of functional columns makes them structures that are a bit too complex to serve as a basic "unit cell" in a crystalline network model. Ideally, the "unit cell" (a netlet) should have stable and well-definable neuronal membership and not be subject to being "divided up" such that at any given time part of it "belongs to" one functional column while another part of it "belongs to" a different functional column. Rather, it would be computationally preferable, as a practical matter, if the unit cell as a whole participated entirely in one or another functional column at any given time. Whether or not such an ideal structure actually occurs in neural anatomy is not known at present, but Rule 6 implies that the idea of such an ideal unit cell should be at the least a useful and realistic approximation to actual cortical behavior. Such a netlet is called, by different authors, either a ***functional microcircuit*** or a ***canonical microcircuit*** [CONNb3]. Functional netlets are Region-I structures exhibiting organized simplicity.

One of the earliest computational models of a functional microcircuit for neocortex was published by Douglas and Martin in 1991 [DOUG2]. The primary requirement a functional microcircuit model must meet is that it provide a functionally realistic input-output characteristic. When we are working at levels fairly close to the physiological level (involving only one or a few steps in model order reduction in making abstraction from the physiological neuron model level), the neuron models comprising the functional microcircuit should present input-output responses that are good approximations to measurable physiological membrane responses. This is the level Douglas and Martin worked at in their 1991 paper. For neurons they used an abstract, multi-compartment neuron model. The dynamics of this model were very simplified, as was practically required by the computational horsepower available in 1991, but were "tuned" to reproduce their measured membrane responses with sufficient accuracy for their purposes. One neuron was regarded as representing a population of neurons of a given type. This is a simplification justified

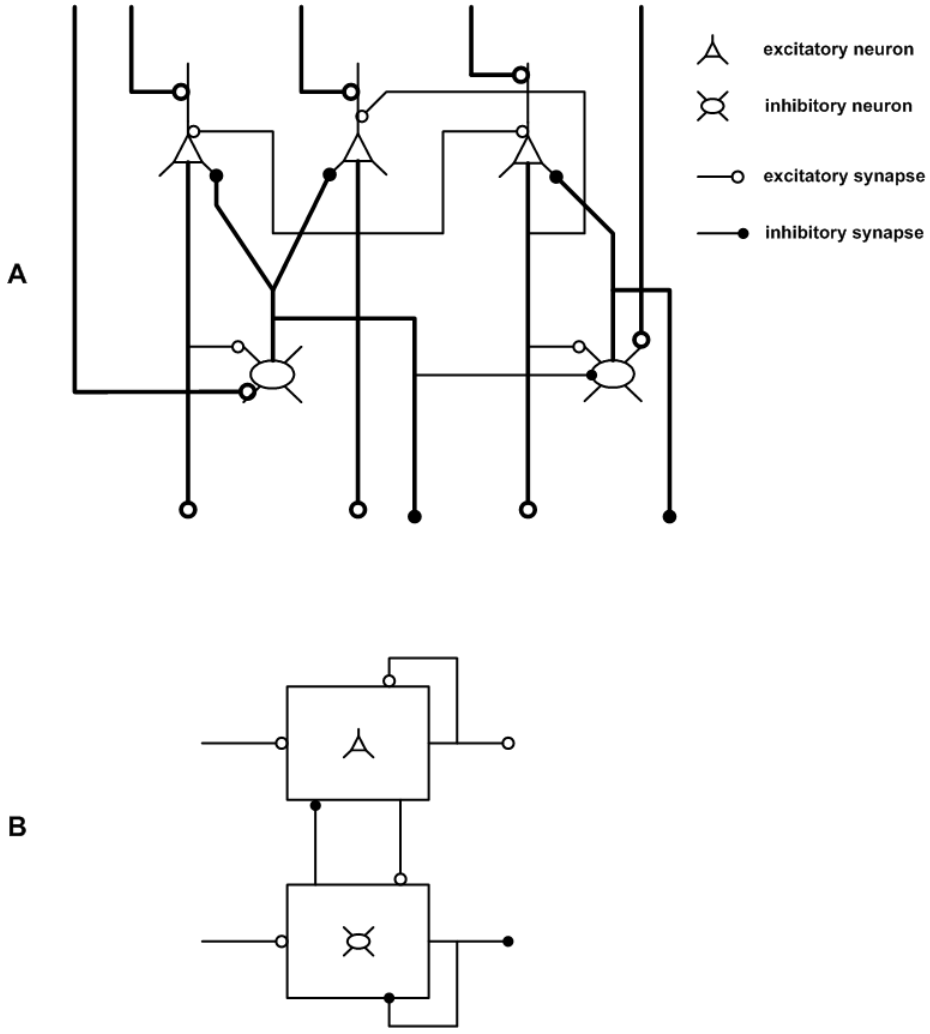


Figure 7.3: Canonical netlet of Douglas and Martin used to construct functional microcircuit models of neocortex. (A) Recurrent two-layer netlet of excitatory and inhibitory neurons. (B) Schematic representation of the microcircuit of (A).

by the implications of White's rules.

Douglas' and Martin's basic schema for neuron-level canonical microcircuits is illustrated in Figure 7.3. It is a recurrent, two-layer schema in which excitatory and inhibitory neurons are segregated by layer. Figure 7.3(B) is a schematic representation of the netlet shown in (A). The microcircuit incorporates a connection feature commonly found in neuron-level circuits, namely lateral connection among neurons in the same layer (note the center excitatory neuron in the excitatory layer; it projects to both its neighboring excitatory neurons; likewise, the left-most inhibitory neuron in the second layer projects to its inhibitory neighbor). It also reflects another feature generally regarded as being typical of neuronal organization, namely feedback from the inhibitory neurons in the second layer to their source neurons in the first layer. Unfortunately, the authors were somewhat equivocal in discussing the details of their microcircuits inasmuch as they

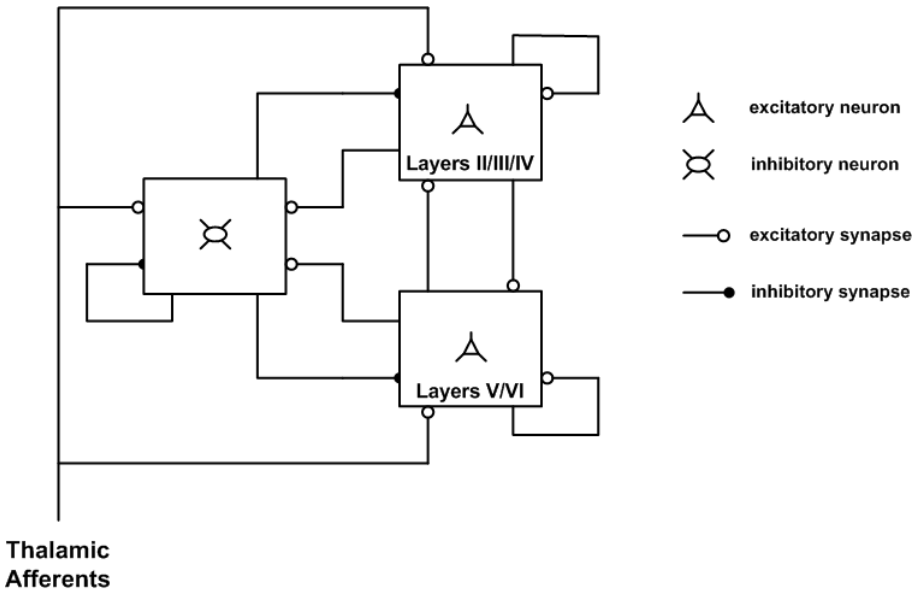


Figure 7.4: Functional microcircuit minicolumn model of Douglas and Martin.

did not make altogether clear whether the circuit of Figure 7.3(A) was used exactly as depicted in the construction of the larger functional microcircuit models they went on to present. (This is a common if unfortunate tendency in much of the literature on network models).

Figure 7.4 illustrates their functional minicolumn model of neocortex. Again, Douglas and Martin did not provide enough detailed description of the specific neuron microcircuits within each block in the diagram to permit independent reconstruction and verification of their results. It is also important to note that the canonical minicolumn structure represented in the figure was tuned for giving proper responses to thalamic input afferents and did not deal with cortico-cortical afferents. What this means is that Figure 7.4 is a starting point rather than an ending point for the construction of unit cell netlets in a general functional microcircuit model of neocortex. Still, despite these shortcomings, the Douglas-Martin model represents an important step in the theory of cortical neural networks. Figures 7.3 and 7.4 provide us with important illustrations of the basic considerations and approach to be taken in a crystalline network model. One might call [DOUG2] an important early essay in the craft of Region-III neural network modeling, this craft not yet being full-grown even in the present day.

Douglas and Martin were able to use their modeling schema to capture an important known property of the neocortical functional column, namely its ability to function as a ***spatial filter*** for thalamic inputs. It is known that the neocortex responds to inputs from the thalamus with a preferred spatial direction. The functional circuit model capturing this effect is shown in Figure 7.5. The model depicts two links in a horizontal "chain" of connections between adjacent areas in the cortex. This chain has a "preferred direction" for receiving successive thalamic stimulation. In

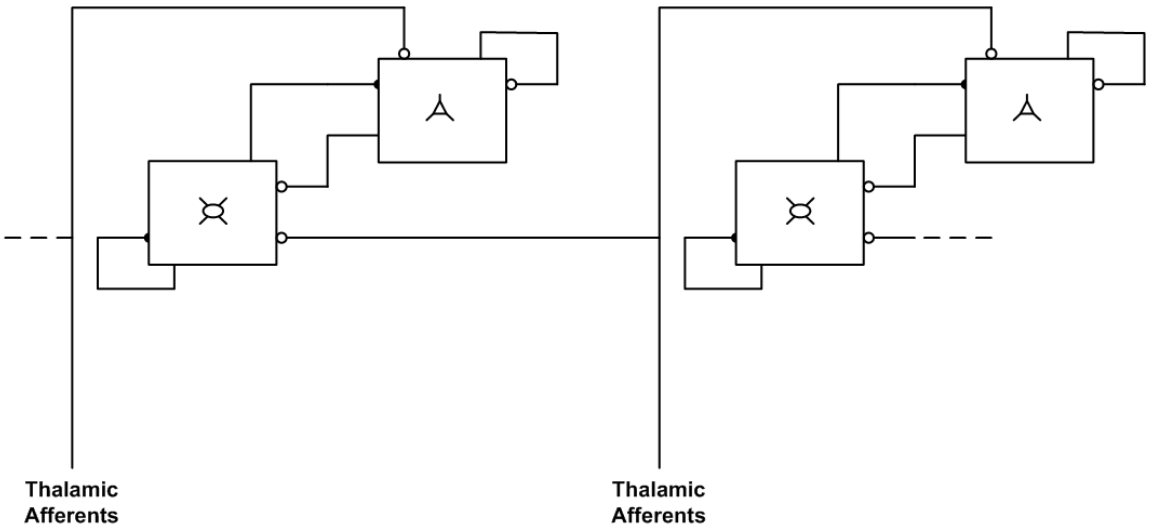


Figure 7.5: Douglas-Martin spatial filter structure for modeling the preferred direction of sequences of cortical afferents from the thalamus. The preferred direction is left-to-right. Two links in a crystal-like chain of adjacent cortical areas are depicted in the figure.

the case where the left-most thalamic input arrives before the right-most input, the excitatory neurons in the left-hand link are activated, their maximal activity being restricted by their interaction with the local population of inhibitory cells. Subsequent thalamic excitation applied to the next link excites that region's excitatory cells and increases the inhibition in the left-hand circuit. However, if the direction in which thalamic afferents arrive is reversed, the excitation of the inhibitory population in the left-most mini-column prevents the response of the excitatory population to the thalamic afferent. (Note how the excitatory populations are self-exciting; when the inhibitory population activates before the excitatory population has had a chance to do so, the nonlinear interaction is sufficient to block the excitation of the excitatory population).

[DOUG2] does not claim to present a complete model of a cortical functional column. This is evident by examining the differences between figures 7.4 and 7.5, and by comparing both to the general layout of the neocortex in figure 7.2. Clearly there are plenty of opportunities remaining for computational neuroscience research directed at neocortex. But [DOUG2] does serve to provide us with a "case study" of the general approach to dealing with the organized complexity of Region-III systems, and, within the limitations of the purposes for which the Douglas-Martin model was constructed, it is as successful a model as any that have been proposed thus far for network-level modeling of neocortex.

§ 5. Rungs of the Ladder: Model Order Reduction vs. Scientific Reduction

Our discussions up to this point have all maintained a relatively close proximity to physiology. This is natural for the method of pedagogy presented in this text, in which we have begun with

the biological roots of computational neuroscience and have been building upon these roots in all the chapters thus far. The phenomena to be modeled and explained thus far are all phenomena uncovered in the biologist's laboratory, and our models have been slowly climbing up toward increasing levels of abstraction as we advance from neural mechanism to individual neuron behavior to, now, the behavior of small neural netlets. Each step we take makes abstraction from the mechanistic world of the anatomist and physiologist, and we do this in the service of a very real need to simplify the amount of detail needed in obtaining computationally tractable models of ever more complex systems of the central nervous system. This ascending of the ladder of model representations in the hierarchy of neuroscience is called model order reduction.

But there is another starting place for computational neuroscience research, and this starting place takes its point of reference and its scientific data from a different science altogether, namely psychology. In this endeavor, neuroscience seeks to explain extraordinarily complex phenomena in the behavior of the living animal. The behavioral phenomena of most interest are perceptual, cognitive, emotional, motivational, and social. This is the province of ***mind*** rather than brain – a province that stands, philosophically as well as scientifically, at as great a distance from the mechanistic world of biology as any gulf between any two topics in science ever gets. Ideas such as perception, emotion, cognition, or motivation play no part whatsoever in our models of the neuron. Nor is it scientific for us to adopt a happy thought and expect these ***mental phenomena*** to suddenly appear, as if by elfin magic, when we collect enough neurons together in a network. William James, the father of American psychology, wrote,

> Already, in discussing the localization of functions in the brain, I had said that consciousness accompanies the stream of innervation through that organ and varies in quality with the character of the currents, being mainly of things seen if the occipital lobes are much involved, of things heard if the action is focalized in the temporal lobes, etc., etc.; and I had added that a vague formula like this was as much as one could safely venture on in the actual state of physiology. . . The consciousness, which is itself an integral thing and not made of parts, 'corresponds' to the entire activity of the brain, whatever that may be, at the moment. This is a way of expressing the relation of mind and brain from which I shall not depart during the remainder of the book, because it expresses the bare phenomenal fact with no hypothesis, and is exposed to no such logical objections as we have found to cling to the theory of ideas in combination.
>
> Nevertheless, this formula which is so unobjectionable if taken vaguely, positivistically, or scientifically, as a mere empirical law of concomitance between our thoughts and our brain, tumbles to pieces entirely if we assume to represent anything more intimate or ultimate by it. The ultimate of ultimate problems, of course, in the study of the relations of thought and brain, is to understand why and how such disparate things are connected at all. But before that problem is solved (if it ever is solved) there is a less ultimate problem which must first be settled. Before the connection of thought and brain can be explained, it must at least be *stated* in an elementary form; and there are great difficulties about so stating it. To state it in elementary form one must reduce it to its lowest terms and know which mental fact and which cerebral fact are, so to speak, in immediate juxtaposition. We must find the minimal mental fact whose being reposes directly on a brain-fact; and we must similarly find the minimal brain-event which will have a mental counterpart at all [JAME: vol. 1: 176-177].

The art and science of connecting a "mental fact" and a "brain-fact" is called the psycho-physical study of neuroscience. The close of the twentieth century saw the delivery of important new instruments for exploring this field – such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) among others – and these instruments have greatly empowered neuroscience's ability to construct ***map models*** and ***network systems*** of the brain in correlation with psychological phenomena of mind, e.g. [BULL], [LEVId]. Researchers engaged in making this psychophysical "mind-brain connection" are engaged in climbing *down* the ladder of the neuroscience hierarchy. They are engaged in ***scientific reduction***.

Today a great gulf still lies between the psychophysical phenomena addressed by ART-map network system models and the biophysical phenomena addressed by neural network models. We will know we have bridged this gap only when James' statement regarding "the less ultimate problem" can be made with scientific rigor. Metaphorically, the researchers engaged in these two disparate wings of the science are working to construct a transcontinental railroad of sorts. The job will be done when the tracks running in from the east meet those running in from the west and the "golden spike" is finally driven.

But, of course, to drive this golden spike the tracks must meet in one place. To switch metaphors, it is not enough to have rungs on a ladder of representational hierarchy. The rungs must be connected by the rails of the ladder. Each higher rung in model order reduction has to be tied to the rung immediately below it; each lower rung in scientific reduction has to be tied to the rung immediately above it. The ultimate goal toward which all of neuroscience labors is to come at last to the final rung somewhere in the middle, where the rails connect in both directions.

Now, a map (as we use that term in this book) is a network of neural networks. A network system is a network of maps. Given our previous discussion of the Region-III issues confronting the modeling of a neural network, it is no doubt immediately evident how much the issue of organized complexity is likely to be in play at the levels of maps and network systems. We can therefore rightly anticipate that as we progress in this textbook, we will be obliged to introduce increasingly more abstract models to be able to deal with the greater and greater span of phenomena represented at successively higher rungs of the reductionist ladder. The common theme throughout, from functional microcircuits to ART-map network systems, is the same: Dealing effectively with the organized complexity of the system we have chosen to study. We take our next step in this progression up the ladder in chapter 8.

Exercises

1. Most introductory textbooks on economics, available from any good university library, recognize and define four distinct types of economic markets: (1) perfect competition; (2)

monopolistic competition; (3) oligopoly; and (4) monopoly. Write a short essay explaining which of the systems regions (I, II, or III) each of these falls under and what factors lead you to make your classification.

2. Professor of History and Political Science Charles A. Beard, writing in 1929, said, "No science of politics is possible; or if possible, desirable." Beard cited such factors as the uniqueness of each political situation, the fact that different people hold different political views (and do so more or less unpredictably), and that controlled experiments in political science are not possible. Beard's view is opposed by many present day political scientists who distinguish between "traditional political science" (prior to 1950) and "behavioral political science" (post 1950). If it is true, as Beard claimed, that political science cannot be a science, then it is a subject that cannot be treated by system theory. Write an essay speculating on how systems theory might be applied to political science, which regime of systems (I, II, or III) such a theory might fall into, and what factors lead you to categorize the system-theoretic treatment of political science in the way you do.

3. Many physicists intermingle the terms "thermodynamics" and "statistical mechanics" in everyday conversation to such a degree that non-physicists are often taken aback to hear that thermodynamics and statistical mechanics are distinct sciences. Using the definition of "system" given in chapter 1, explain how and why a system theorist regards statistical mechanics and thermodynamics as distinct sciences. Hint: pay attention to the *objects* of these sciences. How do you think thermodynamics and statistical mechanics stand in relationship to the actual physical phenomena these disciplines seek to explain? Represent your explanation in the form of a diagram.

4. For this exercise, you will need a cup of hot coffee and a half-dozen small, well-frozen ice cubes. The cup should be about $2^{7/8}$ inches in diameter and about $3^{1/2}$ inches deep. The ice cubes should be no larger than about 5/8ths of an inch along the sides (to leave plenty of room for them to float around freely in the coffee cup). Being careful not to splash the coffee out of the cup, drop an ice cube in the center of the cup. Observe and carefully record everything that happens from the time you release the ice cube until the time it melts completely. Replicate the experiment 5 more times, replenishing the coffee as needed to keep it from cooling off too much. From your observations, what phenomena would a model of this system be required to explain? What are the likely object variables that must go into such a model? What system region (I, II, or III) will this model likely fall into?

5. A common approach in neural netlet modeling is to represent all the neurons of a given type (e.g. RS-type pyramidal cells) using an "average neuron" model. Modify your Wilson RS-type neuron model program you wrote in chapter 6 to add AMPA synaptic inputs modeled by the $g^{(\beta)}$-function. Use $\tau_1 = 0.16$ ms and $\tau_2 = 3.00$ ms for your time constant parameters. Run simulations for different models using the smallest and largest values for g_{max} given in Table I and the minimum, typical, and maximum values for C_m given in Table II (a total of six different simulations). Apply synaptic inputs weighted for 1, 10, 20, 30, and 40 synapses receiving simultaneous stimulation arriving at $t = 10$, 60, 110, 160, and 210 ms, respectively, for these 5 different levels of synaptic input (one simulation run for each of the six neuron parameter sets). Judging from your simulation results, what kind of "average neuron" do you think might be able to represent all RS-type neurons in the network? How would you model it? If you do not think one single "average neuron" could realistically be used in a network model, what are your modeling alternatives? Comment on the effect these alternatives have in regard to the Square Law of Computation.