

What do random matrices tell us about the brain?

Kanaka Rajan

Research Associate

Princeton University

krajan@princeton.edu

ABSTRACT

To make optimal use of experimental data on the anatomy and physiology of brain circuits, and to account for the effects of plasticity and modulation on behavior, we must understand the relationship between the synaptic connectivity and neuronal properties, and the activity that these properties produce at the network level. Given detailed knowledge of the properties of a complex network, can we predict what it will do? If we know how plasticity mechanisms or modulators change those properties, can we predict how the activity will change? These are difficult questions to answer. In the broadest terms, my paper explores how knowledge of synaptic connectivity and neuronal gain can be used to predict the activity produced by a network model. Along the way I address some issues in sensory and motor physiology: the role of spontaneous activity in sensory processing, memory storage and motor pattern generation using insights from the physics of random matrices.

Keywords

Random matrix theory, complex dynamics, neural networks

1. INTRODUCTION

The manner in which neurons are connected to each other in the brain has a strong influence on the dynamics that emerge from it. To approach the issue of how complex internally generated dynamics arises in a general framework, we began studying large networks of connected model neurons (a review of a few typical network architectures forms the bulk of Vogels, Rajan & Abbott, 2005).

The strengths of synapses between all pairs of neurons of a neural network are typically represented by a synaptic connectivity matrix, which contains as many elements as the square of the network size. Perhaps we can hope to determine individual synaptic strengths in the brain through sophisticated experiments someday. At present however, our resolution for experimental measurements limits us to certain statistical properties of groups of synapses. For example, we know from paired recordings in layer 5 cortical neurons that synaptic strengths may exist in a lognormal distribution (Song *et al.* 2005).

For the types of networks we build, we construct the synaptic matrix with a large fraction of 0 elements because the connection probability of biological neurons is sparse. The remaining non-zero elements are drawn randomly from a Gaussian probability distribution characterized by a mean and a variance. Provided that the synaptic strengths are not too large and there is no external input, simulating this network *in silico* results in every cell of the network firing at the same constant background rate (a trivial fixed point in the dynamics of the network).

Inducing more complex activity requires modification of the synaptic strength distribution in different ways. We studied a variety of relevant network behaviors by strengthening both

excitatory and inhibitory synapses in model networks, while maintaining the balanced state believed to exist in many neural circuits (Shadlen & Newsome, 1998 and Haider *et al.*, 2006). The key element in determining the spontaneous dynamics of networks constructed in this way is the width of the distributions of excitatory and inhibitory synaptic strengths. We will look at the many implications of this result, but first let us see why this is the relevant manipulation.

Diagonalization of a random matrix is a mathematical manipulation that yields the same number of modes as there are elements in the matrix or in our case, the number of neurons that make up the circuit under scrutiny. Each mode is usually termed an eigenvalue and is a complex number - the real part of which is related to the decay time (or the lifetime depending on convention) of a certain activity pattern and the imaginary part is proportional to the frequency of that pattern. Activation of any one of these complex modes results in a network that exhibits spontaneous oscillations at the corresponding frequency. The activation of multiple such modes results in complex persistent dynamics due to a superposition of individual frequencies in a highly nonlinear manner. Increasing the variability of the synaptic strengths across a neuronal ensemble increases the number of activated modes and therefore leads to progressively more complex dynamics. This simplified understanding could give one an intuition for how random matrix theory is a very powerful, yet underutilized tool for studying the complex emergent behavior of large networks of neurons, such as in the brain.

2. METHODS

We build a firing rate network in which the total current for a single neuron is obtained by summing the firing rates of all its presynaptic partners, weighted by the strength of the synapse between them. The description of the model closes when we define a response function that determines the firing rate of each cell as a function of the total current it receives.

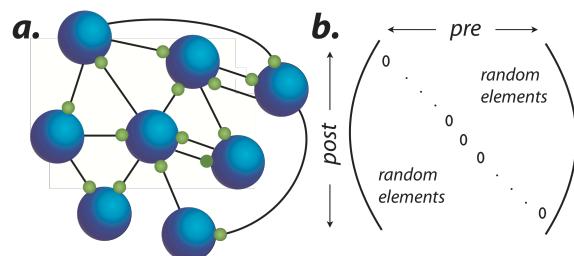


Figure 1a. Schematic of a model network of neurons parameterized by a single firing rate and in **1b**, a synaptic connectivity matrix with the individual synaptic strength values for every neuronal pair randomly assigned from a distribution with zero mean and variance of the order of the inverse of network size.

Figure 1 shows a schematic for the construction of our network of neurons and the synaptic connectivity matrix. We set the diagonal elements to 0 to rule out autapses, although this is not a necessary

condition and does not alter the results that follow. The non-zero synaptic strength values are set randomly from a normal distribution; and each row or column of the matrix has the strengths of all synapses made by a single presynaptic or a postsynaptic neuron respectively. The simplified construction of these models allows us to focus on the synaptic connectivity of the network as the quantity of interest.

3. RESULTS

We study the fluctuation of firing rates around their trivial fixed point in such networks by looking at the spectrum of eigenvalues of the synaptic connectivity matrix. To reiterate, the real part of an eigenvalue corresponds to the decay rate of the associated mode, whose frequency is given by the imaginary part of that eigenvalue. The number of modes scales with network size. A classic result in Random Matrix Theory is that all the eigenvalues of a random matrix with no additional symmetries are distributed uniformly in a circle in the complex plane, with its radius proportional to the width of the underlying distribution of elements (Girko, 1985; see also Figure 3a).

3.1 Behavior as modes of a random matrix

The radius of this circle of eigenvalues is proportional to the variance of the synaptic strength distribution. If the real part of a complex eigenvalue exceeds 1, the activated mode leads to long-lasting oscillatory behavior. Increasing the synaptic variance makes the network spontaneously active by activating multiple eigenvalues (the network is pulled back from the edge of epileptic activity by the non-linear response function we use). We can get fixed points, complicated oscillatory behavior and even chaotic dynamics from the network by progressively increasing synaptic variance (Sompolinsky *et al.*, 1988). Therefore, changing the variance of the synaptic matrix gives us a single parameter for scanning the different dynamic regimes the network is capable of exhibiting.

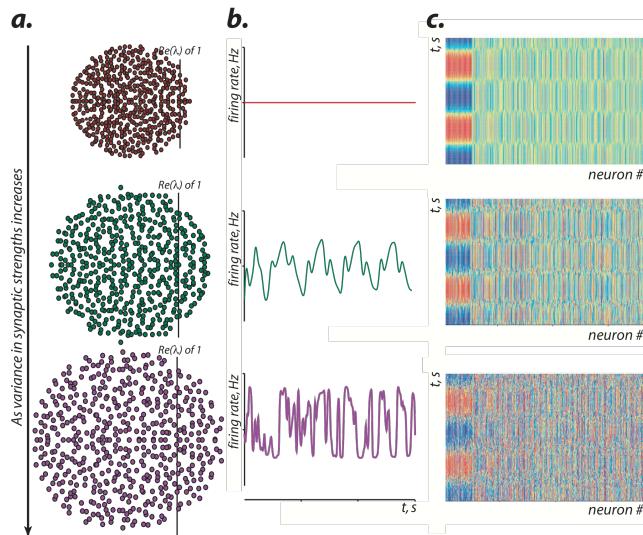


Figure 2. Effect of making the strengths of synapses more variable across the network. 2a shows the radius of eigenvalues growing and activating more modes. 2b illustrates increasing complexity in the firing rates of individual network neurons, and the same effect at the network level is seen in 2c.

Figure 2 shows the effect of changing the variance of synaptic strengths across a network of neurons. We see that as the standard deviation of synaptic strength grows, so does the radius of the circle of eigenvalues, yielding more and more activated modes with real parts >1 . The vertical line in Figure 2a shows the line of stability beyond which a mode is termed active and the related dynamics persist forever. In Figure 2b we see the firing rate of a randomly selected single neuron from a simulated network of 1000 neurons, plotted as a function of time. For very few active modes, a network is barely critical and the activity of every cell in this network is at a fixed point. This means that while every cell has a different firing rate, these rates are not time varying. The network, though autonomously active, is not terribly interesting from a dynamics perspective (top panel in Figure 2c).

As synapses become more variable, networks start to oscillate (as seen in the middle panel of Figure 2b & 2c), and finally generate chaotic activity (bottom panels of Figure 2b & 2c), both in the vernacular sense of being complex, deterministic and non-repeating, but also more formally determined by calculating the appropriate Lyapunov exponents. Indeed the activity across the network averages to 0 but the second moment, or response variability carries a trace of the encoded signal. Changing the mean synaptic strength, which is the traditional focus of experimental and theoretical neuroscience alike, does not change the distribution of eigenvalues in an interesting manner (only generates a single large fixed point) and does not yield complex temporal dynamics.

We have found that the key element for controlling the spontaneous dynamics of neural networks is the variability in the excitatory and inhibitory synaptic strengths. A dramatic prediction of this work is that plasticity mechanisms, neuro-modulatory influences and pharmaceutical manipulations that strengthen a subset of synapses while weakening others in conjunction, might be more meaningful in the generation and maintenance of stable complex dynamics in the brain.

3.2 Eigenvalues of networks of neurons

In large neural networks, elements of the synaptic connectivity matrix can be chosen randomly from appropriate distributions, making results from random matrix theory highly relevant. However, classic results on random matrices (Girko, 1985) do not apply to synaptic connectivity matrices because of the constraint that individual neurons are either excitatory or inhibitory (a property of neuronal circuits commonly known as Dale's principle).

We developed a new result in random matrix theory that applies to biologically realistic connectivity matrices with excitatory and inhibitory columns drawn from distributions with different means and equal or different variances (Rajan & Abbott, 2006). The results we obtained for biologically realistic connectivity matrices enabled us to look for the relationship between connectivity and dynamics in biologically realistic neural networks.

Figure 3 shows the result of applying Dale's principle to a naïve asymmetric random matrix. This is mathematically equivalent to insisting that every column or row of the synaptic connectivity matrix be restricted to either all positive or all negative entries. The resulting sign-wise column restriction changes the spectrum of eigenvalues obtained as a result of diagonalization, and

introduces a central zone of higher density of modes that tapers with distance from the center (Figure 3b).

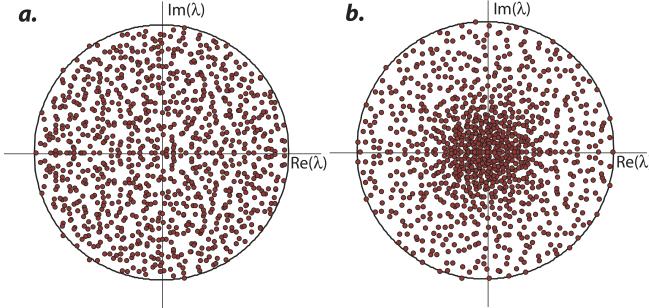


Figure 3. Girko's circle law dictates the uniform distribution of eigenvalues in a circle in the complex plane. Accordingly the radius of the circle in 3a equals the variance of the distribution of elements in the matrix. 3b shows the effect of insisting that neurons have either purely excitatory or inhibitory synapses. Note the central high-density zone of eigenvalues compared to the edges.

The functional consequence of this manipulation can be understood simply from the fact that there are fewer modes near the line of stability on the real axis. A higher number of eigenvalues need to be activated in order to produce the same complex dynamics seen from a matrix with no preferences. The relevant variance of synaptic strengths, which parameterizes the radius of these spectra, consequently needs to be larger still to observe the different dynamic regimes seen in neural networks, making the role of synaptic variability more pertinent.

3.3 Remembering with modes

A natural question while mulling brain function is to ask whether our internal thoughts have anything to do with past experience. We know that chaotic activity traverses a strange attractor embedded in the full space of network dynamics. I propose that the shape and geometry of this attractor is modified by experience and that those modifications can be read out to recover memories.

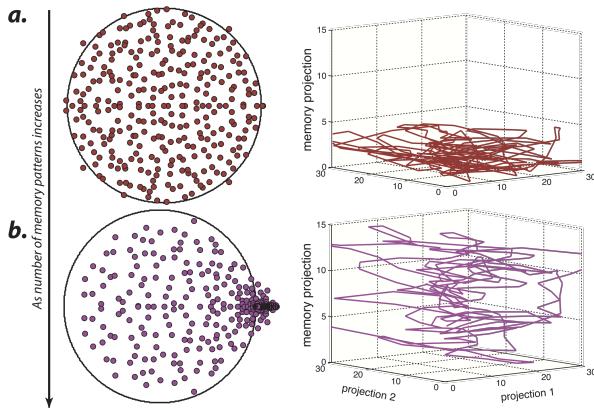


Figure 4. Building a memory model using the physics of random matrices. 4a shows the evolution of the spectrum of eigenvalues in the complex plane as a function of increasing training (red circles in 3a → purple circles in 3b). Projecting the activity of the network onto two principal components and a third memory direction shows no trace in an untrained network seen in red in the right panel of 4a, and a trace of memory - purple in the right panel of 4b.

We achieve readout of these memories by projecting the full network activity through a linear readout system. Memory storage then occurs through synaptic plasticity mechanisms that modify

network connections. We derived an optimal learning rule for making these modifications and the results of this work are shown in Figure 4.

The main result of this work is that as training increases, the spatial extent of network activity grows and occupies a growing trajectory in space, enclosed within the total space of chaotic activity. The top right panel of Figure 4 illustrates the activity of the naïve network that only exhibits spontaneous dynamics and has no memory of past occurrences. We see that the activity resides in the space described by the internal dynamics and none of it is in the memory dimension. The top left panel shows the eigenvalue spectrum of such a naïve network that shows the typical uniform density of eigenvalues near the line of stability (Figure 4a, red circles).

The bottom left panel shows a higher density of eigenvalues near the line of stability (Figure 4b, purple circles). This higher density in the eigenvalue spectrum is different from that seen in Figure 3b, in that when activated, these modes lead to fixed points rather than chaos. These fixed points are line attractors that function as memories because they are persistent, non-oscillatory and need a single firing rate to represent fully. This change occurs upon training, which is the qualitative and computational equivalent of plasticity mechanisms in the brain. The shape of the spectrum of eigenvalues changes with training (red circles in the left panel of Figure 4b) and the network activity rises to fill the cube made from the 2 internal and 1 memory dimension of network dynamics. We are currently working to understand the capacity of such networks, and the factors that influence and change this capacity (Rajan, 2010, in preparation).

3.4 More “mental” matrices

How do delays affect network dynamics and learning? Integration and transmission delays are ubiquitous in biological networks. What effects do these have on the dynamics of neural networks and what does the physics of random matrices inform us about these effects? Delays often introduce instabilities in circuits, but a mounting body of evidence in neurobiology indicates that delays can be useful (Carr & Konishi, 1988; Lisberger & Sejnowski, 1991; Unnikrishnan *et al.*, 1991). In particular, many networks become oscillatory as a result of delayed interactions, a fact which expands the class of realistic architectures capable of supporting stable oscillations. Within oscillatory circuits, we ask whether delays merely shift bifurcation points or whether they lower the frequency of the oscillations, broaden their spectrum and possibly lengthen transients (Rajan, 2009).

Introduction of delays in nonlinear systems also changes the conditions for onset of chaos. I study time-dependent dynamics in a mean-field theoretic framework for a system of nonlinear neurons governed by delay differential equations. Such a system typically exhibits a variety of oscillatory modes that eventually lead to the emergence of multiple chaotic strange attractors. We determined the number of modes that can be used to encode signals in this regime, and ask how this capacity changes as function of input strength and magnitude of delay.

We know that delays can vary in an organism over slow time scales, for instance during development. As has been analyzed for synaptic weights, a future direction is to investigate whether there could exist mechanisms for the adjustment of delays coupled with network dynamics. Adjustable delays are an additional

phenomenon through which networks could achieve a broader range of behaviors, and deriving gradient descent learning equations will help tease apart some of these issues.

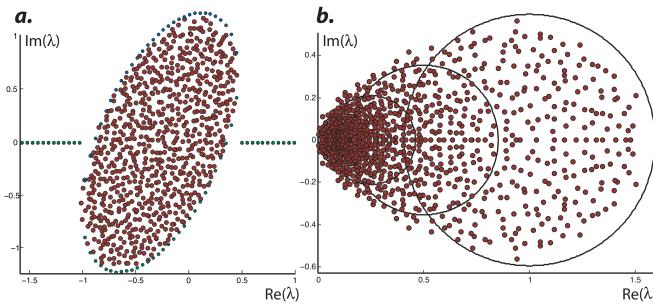


Figure 5. Introduction of delays in the synaptic connectivity matrix changes the effective stability matrix, disrupting it into a rotational ellipse of eigenvalues in the complex plane. The line of stability is still at a real part of 1, as seen in 5a, but the tilt predicts that the resulting dynamics can contain higher frequency components. In 5b, we consider multiple time constants corresponding to a diversity of neuronal subtypes in the same network.

A final direction we have taken the study of random matrices adapted to neural circuits, is to look at multi-time dynamics. We introduce a number of time constants in the basic firing rate differential question and ask how these artificial time constants interact with the slower time dynamics that emerge autonomously from the network as a result of variable synapses. To a first approximation, there is a large variety of interneurons in the brain and a large number of synapse subtypes, each with its own characteristic time constant. Most artificial neural network models ignore this diversity for mathematical tractability, but looking at ensembles of time constants much in the same way as we have been looking at matrices of synaptic strengths gives us access to the same tools. For instance, introducing about ten different time constants changes the shape of the effective eigenvalue distribution and therefore the dynamics that result from this diversity (see Figure 5b). We are currently extending this idea to questions of robustness of a network and how that scales with neuronal or synaptic diversity.

4. DISCUSSION

We have studied a variety of relevant complex behaviors by strengthening both excitatory and inhibitory synapses in model networks (Shadlen & Newsome, 1998 and Haider *et al.*, 2006). We have found that the key element in determining the spontaneous dynamics of networks is the variability in the excitatory and inhibitory synaptic strengths. As synapses become more variable, networks exhibit dynamics that increase in complexity in a direct relation to the synaptic variance, rather than the mean synaptic strength across the entire network.

This is an important step toward directing future experiments because known plasticity mechanisms, pharmacological manipulations or physiological modulators that bring about changes in synaptic variability often have confounding mixed effects, making them hard to read. On the other hand, if this is indeed the way neural circuits in the brain maintain intrinsic activity, this can lead to dramatic predictions about the efficacy of modulatory drugs and manipulations that simultaneously

strengthen and weaken subsets of synapses within the same brain region.

Sompolinsky *et al* showed 20 years ago that generic network models that are spontaneously active are chaotic. This appeared to eliminate such networks from models of anything useful, because the sensitivity of chaos to initial conditions makes their activity non-repeatable. In addition, the strong chaotic dynamics appeared to put huge constraints on a network's ability to represent stimuli. I addressed this encoding issue in a different project, the methods of which are outside the scope of this paper (details of the mean-field theoretic approach and network simulation results are in Rajan, Abbott & Sompolinsky, 2009 and 2010). Briefly however, we have found that these chaotic networks make a phase transition from a regime where the response to inputs is superposed on a complex background, to a regime where the background noise vanishes leaving behind a purely input-driven periodic response.

We determined that networks exhibiting complex pre-existing dynamics can be entrained by a strong enough input for a certain frequency range, below and above which, chaotic and periodic driven activity coexist. Indeed, the range of frequencies over which this suppression effect emerges in our theoretical study (3-7 Hz) is comparable to the range of frequencies over which many sensory systems are thought to operate.

The discovery of this phase transition has opened up the possibility of controlling chaotic activity through feedback and utilizing randomly connected model neurons to perform a range of biologically interesting tasks (learning and plasticity in networks exhibiting complex spontaneous dynamics is an ongoing project being worked on in collaboration with experimentalists in Carlos Brody's group at Princeton). This non-monotonic phase transition has been demonstrated in awake behaving animals given whole-field visual stimulation (White & Fiser, 2010). The general idea of suppression of response variability in the presence of external inputs is being studied in dorsal premotor cortices of monkeys (Churchland *et al.*, 2010) but others have been reporting the ubiquity of this phenomenon across different preparations.

In conclusion, the physics of random matrix theory has the advantages of working within the constraints of experimental biology (such as having only statistical distributions to work with) and using measurable quantities (such as correlations in response variability as obtained from experiments involving voltage-sensitive dyes or calcium measurements). I believe that this type of study yields an excellent crossover between the rich mathematical frameworks present in Theoretical Physics and the data-rich field of Experimental Neurobiology.

5. ACKNOWLEDGMENTS

The author wishes to thank William Bialek at Princeton University for providing a fruitful environment conducive to developing independent postgraduate thinking and the Swartz Foundation for a generous postdoctoral fellowship; L. F. Abbott at Columbia University and Haim Sompolinsky at Hebrew University, Israel & Harvard University for valuable guidance and collaboration. Copyright belongs to the author. Please do not reproduce in full or in part without the express permission of the author. For more in-depth treatment of the analytic results the reader is directed to Rajan, 2009.

6. REFERENCES

- [1] Abbott LF, Rajan K and Sompolinsky H, 2009 Interactions Between Intrinsic and Stimulus-Dependent Activity in Recurrent Neural Networks. In M. Ding and D. Glanzman, eds., *Neuronal Variability and Its Functional Significance*. (In Press).
- [2] Carr CE and Konishi M, 1988 Axonal Delay Lines for Time Measurement in the Owl's Brainstem. *Proc. Natl. Acad. Sci. USA* 85: 8311-8315.
- [3] Churchland MM, Yu BM, Cunningham JP, Sugrue LP, Cohen MR, Corrado GS, Newsome WT, Clark AM, Hosseini P, Scott BB, Bradley DC, Smith MA, Kohn A, Movshon JA, Armstrong KM, Moore T, Chang SW, Snyder LH, Priebe NJ, Finn IM, Ferster D, Ryu SI, Santhanam G, Sahani M and Shenoy KV, 2010 Stimulus Onset Quenches Neural Variability: a Widespread Cortical Phenomenon. *Nature Neurosci.* 13: 369-378.
- [4] Girko VL, 1985 Spectral Theory of Random Matrices. *Theory Probab. Its Appl. USSR* 29: 694.
- [5] Haider B, Duque A, Hasenstaub AR and McCormick DA, 2006 Neocortical network Activity In Vivo is Generated through a Dynamic Balance of Excitation and Inhibition. *Journal of Neurosci.* 26(17): 4535-4545, doi:10.1523/JNEUROSCI.5297-05.
- [6] Sejnowski TJ and Lisberger SG, 1991 Neural Systems for Eye Tracking. *Naval Research Reviews* 4:10-15.
- [7] Shadlen MN and Newsome WT, 1998 The Variable Discharge of Cortical Neurons: Implications for Connectivity, Computation and Information Coding. *Journal of Neurosci.* 18(10): 3870-3896.
- [8] Song S, Sjostrom PJ, Reigl SB, Nelson SB and Chklovskii DB, 2005 Highly Nonrandom Features of Synaptic Connectivity in Local Cortical Circuits. *PLoS Biol.* 3: 505.
- [9] Sompolinsky H, Crisanti A and Sommers HJ, 1988 Chaos in Random Neural Networks. *Phys. Rev. Lett.* 61: 259.
- [10] Rajan K and Abbott LF, 2006 Eigenvalue Spectra of Random Matrices for Neural Networks. *Phys. Rev. Lett.* 97: 188104.
- [11] Rajan K, Abbott LF and Sompolinsky H, 2010 Frequency-Dependent Suppression of Chaos in Recurrent Neural Networks. *Bull. of Am. Phys. Society*.
- [12] Rajan K, 2009 Spontaneous and Stimulus-driven Network Dynamics. Doctoral Dissertation, Columbia University.
- [13] Unnikrishnan KP, Hopfield JJ and Tank DW, 1991 Connected-digit Speaker-dependent Speech Recognition Using a Neural Network with Time-delayed Connections. *IEEE Tr. ASSP.* 39: 698-713.
- [14] Vogels TP, Rajan K and Abbott LF, 2005 Neural Network Dynamics. *Annu. Rev. Neurosci.* 28: 357.
- [15] White BL and Fiser J, 2008 The Relationship Between Awake and Anesthetized Neural Responses in the Primary Visual Cortex of the Rat. *Soc. Neurosci. Abstracts* 38: 459.12.