

Article

# The Cluster Variation Method: A Primer for Neuroscientists

Alianna J. Maren

Northwestern University School of Professional Studies, Master of Science in Predictive Analytics Program, 405 Church St, Evanston, IL 60201, USA; alianna.maren@northwestern.edu; Tel.: +1-703-618-5847

Academic Editor: Vaibhav Gandhi

Received: 27 July 2016; Accepted: 15 September 2016; Published: 30 September 2016

**Abstract:** Effective Brain–Computer Interfaces (BCIs) require that the time-varying activation patterns of 2-D neural ensembles be modelled. The cluster variation method (CVM) offers a means for the characterization of 2-D local pattern distributions. This paper provides neuroscientists and BCI researchers with a CVM tutorial that will help them to understand how the CVM statistical thermodynamics formulation can model 2-D pattern distributions expressing structural and functional dynamics in the brain. The premise is that local-in-time free energy minimization works alongside neural connectivity adaptation, supporting the development and stabilization of consistent stimulus-specific responsive activation patterns. The equilibrium distribution of local patterns, or *configuration variables*, is defined in terms of a single interaction enthalpy parameter ( $h$ ) for the case of an equiprobable distribution of bistate (neural/neural ensemble) units. Thus, either one enthalpy parameter (or two, for the case of non-equiprobable distribution) yields equilibrium configuration variable values. Modeling 2-D neural activation distribution patterns with the representational layer of a computational engine, we can thus correlate *variational free energy minimization* with specific configuration variable distributions. The CVM triplet configuration variables also map well to the notion of a  $M = 3$  functional motif. This paper addresses the special case of an equiprobable unit distribution, for which an analytic solution can be found.

**Keywords:** brain–computer interfaces; pattern recognition; statistical thermodynamics; Cluster Variation Method; entropy; brain networks; functional motifs; variational free energy; neural activation patterns; deep learning

---

## 1. Introduction

The purpose of this paper is to introduce neuroscientists to cluster variation methods and their potential application in systems neuroscience. This primer focuses on brain computer interfacing; however, the contribution is more wide ranging and can be summarised as follows: If distributed systems (like the cortex) can be characterized as a two-dimensional lattice with local coupling, there exists a set of mathematical tools that enable one to quantify the distribution of local interactions. In brief, these (cluster variation) tools allow one to prescribe the distribution of local configurations in terms of a single parameter; namely, an interaction enthalpy. This plays a similar role to the temperature in statistical thermodynamics, thereby prescribing a distribution over configurational variables. This formulation has a number of potentially exciting applications. For example, under the assumption that biological systems minimize their free energy, the ensuing equilibrium distribution of local configurations can be determined for any given interaction enthalpy.

This offers a principled and parsimonious way to specify prior probability distributions over the distributed states of (neuronal) networks. These prior distributions may play an important role in either generating random networks for statistical comparison in graph theory, or in furnishing observation or generative models for empirical data via prior constraints. Another application of

this technology would be to summarize any empirical distribution of local configurations in terms of a single parameter, enabling one to test for differences between different brain states or, indeed, different subjects and diagnostic categories. Finally, the ability to summarize the statistics of local configurations on lattices paves the way for testing key hypotheses about network organization; for example, do real brain networks minimize free energy for a given interaction enthalpy? We will return to some of these ideas in the discussion.

A major challenge in constructing effective *Brain–Computer Information Interfaces (BCIIs)* is the difficulty in not just accurately classifying time-varying neural activation patterns, but in creating accurate predictions. Bayesian prior probabilities have often been used for this task [1,2]. However, it can sometimes be difficult to compute these prior probabilities, especially considering the growing complexity of BCI data. One means of obtaining the needed Bayesian priors is to use *Bayesian inference*; for example, *approximate Bayesian computation*, which yields the desired values through *variational free energy minimization* [3,4].

Applying this approach to BCIs requires the identification of a computational substrate whose behavior can be described using statistical thermodynamics. In particular, we need to define a system which can be characterized via various free energy-minimized states, so that variational free energy minimization can model perturbations to these states.

Using a classic Ising model to characterize such a computational substrate has various limitations [5]. In particular, the classic approach yields only limited information, which is insufficient to describe the pattern-richness of neural activations. A more interesting approach is the use of the cluster variation method (CVM) [6,7], which specifies the distribution of not only simple on/off (or **A/B**) units, but also the distribution of local patterns (nearest-neighbor, next-nearest-neighbor, and triplets).

The distribution of local triplet patterns (where triplets comprise the *basic cluster* for this formulation) can be of particular interest to brain researchers, as substantial work indicates that groups of three units each (functional and structural motifs which have a Motif number or  $M = 3$ ) are a minimal computational unit for describing fundamental brain connectivity dynamics.

Up until now, the CVM has been largely unknown to brain researchers. This paper attempts to serve the brain research—and particularly the BCI—communities by offering a tutorial-style CVM introduction. The specific values and unique values of this paper to these communities are:

1. Correspondence between CVM triplet configuration variables and the  $M = 3$  functional motif (building block of three neural units) provides a natural means for interpreting CVM-based results in terms of functional motifs in the brain and vice versa,
2. Local-in-time free energy minimization can potentially help stabilize specific neural learned response patterns, and
3. CVM modeling bridges local pattern distributions to other important realms, including brain activity at critical states, the generation of Bayesian belief probabilities, complexity, and other factors.

Structural and functional motifs are a means of expressing the characteristic building blocks in neuroanatomical data. Functional motifs of  $M = 3$  (a set of three completely-connected neural units) represent a local pattern configuration that both maximizes the number of local patterns that can be achieved while simultaneously minimizing the number of structural motifs needed to support the diversity in functional patterns [8].

The CVM ergodic distribution can be specified using either one enthalpy parameter  $h$  (for the case of equiprobable bistate unit distribution) or two parameters (when that constraint is released) [9,10]. Thus, selecting a single parameter value (for equiprobable unit distributions) specifies a distinct point in the trajectory of equilibrium-stabilized (or non-equilibrium steady state for open systems) *distributions* of local patterns.

The interaction enthalpy parameter  $h$  corresponds to an average pairwise coupling strength between neural units. As will be discussed in Section 2.2, several successful models for neural

organization and dynamics are built using predominantly pairwise connections. As an illustration, Tkačik et al. have combined experimental and theoretical studies to propose specific values for the pairwise interaction enthalpy in thermodynamic models of (vertebrate retina) neural systems which closely emulate actual neural dynamics [11,12]. They further note that their model “accurately predicts the correlations among triplets of neurons”, providing a further rationale for applying the CVM with a basic cluster of triplets to neural organization [11].

Characterizing the distribution of local patterns, we also have a new means for interpreting the kind of information present in a 2-D ensemble. When the enthalpy parameter  $h$  is chosen to yield greater diversity in local pattern distributions, more information can be encoded. Additionally, because the CVM approach works with distributions of triplet patterns, this method usefully connects with means for the characterization of functional motifs where  $M = 3$  [8].

Additionally, work by Friston et al. (2012) shows how Bayesian priors can be derived as outcomes of variational free energy minimization [13]. This specifically addresses notions of self-organization within the brain. While Friston et al. postulate how free energy minimization can be intimately tied to a range of important dynamic phenomena in the brain, the next step is to represent the free energy associated with spatio-temporal neural activity *patterns*. Thus, a means for explicitly representing pattern distributions can be most useful.

The CVM also has potential as a computational substrate within deep learning architectures, addressing the needs of generative methods to model diverse patterns over time.

In short, the Cluster Variation Method (which has received substantial attention in areas such as materials science) has potential application for BCIs, both in supporting direct models of neural activity, as well as providing a computational substrate for an adaptive pattern recognition engine.

Because the CVM formalism is not yet widely known, this paper includes both tutorial and review aspects. The following Section 2 describes both the challenges and the overall context for the creation of next-generation BCIs, together with identifying some key related challenges in understanding neural activity as well as neural structure and function. Section 3 briefly overviews how statistical thermodynamics already plays a role in modeling neural activity, from relatively small ensembles up through global networks, and identifies key methods and experimental results (both computational along with in vivo and in vitro studies) for which statistical thermodynamics has already played a major role.

Sections 4–6 introduce the Cluster Variation Method itself. Section 4 introduces and describes the *configuration variables*, which are used to characterize local patterns. It provides a very small but useful illustration of how pattern distributions correlate with the interaction enthalpy parameter  $h$ , using a 1-D CVM to facilitate an immediate intuitive understanding.

Section 5 provides the theoretical framework for the 1-D CVM, again providing an instance within intuitive reach. Section 6 builds on this, providing the theoretical framework for the 2-D CVM, which will typically be more appropriate for modeling neural systems as well as for a BCI computational substrate.

Section 7 continues with a focus on the 2-D CVM, overviewing next steps connecting the 2-D CVM with both anticipated results for the modeling of neural-like systems with a “rich-club” topography and implications for a computational engine that can be used to interpret neural activation patterns. Section 8 provides concluding remarks.

## 2. Background

The rapidly-advancing field of Brain–Computer Interfaces (BCIs) is undergoing a major transition as we move from relatively sparse sensor assemblies (typically external to neural tissue) towards highly dense intra-cortical sensor arrays. As an illustration of this transition, Stevenson and Kording have published summary data showing that the number of neural recording electrodes has doubled approximately every seven years, over the 1960–2010 timeframe [14,15].

Three key areas rapidly hastening the onset of effective intra-cortical BCIs include developments in:

1. Electrodes and other implantable materials with reduced neuro-inflammatory responses, including recent advances in carbon nanotube (CNT) electrodes [16–22],
2. Novel fabrications allowing much greater sensor density [23–26], and
3. Improved wireless communication and power transmission technologies [27–29].

We are thus realizing ever-increasing granularity and refinement in sensed neural activations, in both the spatial and temporal realms. As Tognoli and Kelso state, “From a viewpoint that separates observable and true dynamical properties of neural ensembles, state transitions observed in EEG (electroencephalography) and LFP (local field potential) recordings (STobs) cannot be taken at face value, since they are the indirect manifestation of the dynamical organization of their originating neural ensembles. The issue can be turned the other way around, however, by examining which neural organizations appear as state transitions in neurophysiological recordings” [30].

However, even with externally-placed sensors, increased sensor density is encouraging a greater range of applications, ranging from entertainment to the detection of order–disorder transitions in the brain [31,32]. Maren and Szu describe how a CVM-based approach to modeling EEG signals can be used in a dry-mounted commercially-available EEG recording system, suitable for daily use [33].

Neural activation patterns are related to functional connectivity within the brain, and functional connectivity is in turn related to and dependent upon the underlying structural connectivity. Recent work has established that functionally-connected neural groups are indicated by both search information factors as well as path transitivity along shortest paths—both of these leading to improved communications [34].

In addition to (and potentially complementary to) the methods already advocated, there is another means by which functionally-effective neural patterns can be characterized. This method again fosters a high degree of information storage and communications pathways. This complementary approach uses statistical thermodynamics, particularly the Cluster Variation Method (CVM) [6,9,10,33].

There is an entirely different, although related, role for CVM in modeling brain activation patterns for BCI. This role is that—in addition to directly modeling neural activations themselves—we need a computational system which functions as a pattern repository and classification engine, operating in concert with the time-varying patterns detected in the brain. For this related role, the most common approach thus far has been to construct various categorization engines, interpreting real-time neurally-generated signals [35]. Most significantly, recent advances in deep learning [36] build on earlier neural network architectures [37–39] and use brain architecture-based principles to solve very computationally tough problems.

Thus, this paper introduces the Cluster Variation Method for use in Brain–Computer *Information* Interfaces (BCII), with two purposes: (1) Identifying how the CVM functions in a complementary and supportive manner to existing and ongoing work on functional and structural motifs, dynamics of neural ensembles, and the characterization of neural pattern formation, providing additional input to a BCII system; and (2) Providing the basis for a new BCII computational engine, along the lines of successful deep learning generative architectures, invoking equilibrium processes to stabilize pattern formation in response to specific stimuli. In short, the next demanding challenge deals not so much with the interface itself, but with the *information* interface, the Brain–Computer *Information* Interface (BCII).

The background given in the following three subsections addresses three distinct areas:

1. Means for decoding the information content in neural pattern formation and activation,
2. Pairwise and longer-range neural connectivity, and
3. Functional and structural motifs, neural topographies, and pattern formation: structural and functional organization of small units within the brain (or *motifs*), as they can influence BCII signal classification.

### 2.1. Decoding Information in Neuronal Activation Patterns

The following two subsections address two components of the Brain–Computer Information Interface:

1. Modeling the signal itself, which will continue to be a challenge in what will still be a low signal-to-noise environment, and
2. Use of Bayesian statistics for the classification of neuronal outputs, identifying both current applications and important challenges in the extension of this approach.

#### 2.1.1. The Recurrent Quantum Neural Network

Gandhi et al. have advocated an interesting approach to modeling the time-varying signals produced by a neural ensemble by using a recurring quantum neural network (RQNN) [35], which further develops an approach introduced by Behera et al. [40]. They have applied the RQNN to classifying EEG signals corresponding to subject's motor imagery (MI) of a desired task [41].

A key factor in the RQNN approach is that the average behavior of a neural lattice that estimates the EEG signal is a time-varying probability density function (*pdf*). Each neuron within this computational lattice mediates a spatio-temporal field that aggregates the *pdf* information from the observed noisy input signal. The RQNN approach enables the online estimation of the *pdf*, which then allows the estimation and removal of noise from raw EEG signals, thus addressing a key challenge in working with such low signal-to-noise signals.

The RQNN can be viewed as a Gaussian mixture model estimator of potential energy with fixed centers and variances, with variable weights in the neural network component of the model. Gandhi et al. found significant improvement in RQNN results compared with both the well-known Savitzky–Golay (SG) filter as well as the Kalman filter. They note that one particular advantage of the RQNN method is that, unlike the Kalman filter (for which the statistical behavior of noise and signal are assumed), the RQNN makes no such assumptions. Rather, the RQNN directly estimates the *pdf* without making such assumptions, and can thus preferentially enhance the EEG signal, as the noise *pdf* is naturally non-Gaussian. Thus, the RQNN may be a valuable tool for improving EEG signal enhancement. It is likely that the same method would potentially be extensible to extracting meaningful signals from more granular neural ensembles.

#### 2.1.2. Bayesian Methods

A primary goal with BCI is to infer intention, given neural inputs. These may be relatively stationary tasks (e.g., classification) or smoothly-varying over time (e.g., motor control). Very often, Bayesian methods are crucial to the interpretation of either or both neural signal correlates and/or task intentionality [1,42,43].

Bayesian methods also play a role in the information-processing component of BCIs; that is, within the Brain–Computer Information Interface (BCII). For many applications, classic information theory methods suffice [44–46]. However, these methods typically rely on the ability to model a distribution, or a set of responses to a given stimulus, in terms of prior history. It is not always feasible to obtain these conditional probabilities, or to rely on them once they have been collected. In addition, neural systems are adaptive; they can change their responses over time [47–49]. Another factor is that as we expand to more practical and pervasive interfaces, there are likely to be more extraneous stimuli and internal distractions. These will not always readily fit into a predetermined modeling schema.

Thus, Bayesian models are likely to be essential as we create the next level of BCIIIs. However, we will need to expand on the basic methods. In particular, we will find it useful to approximate Bayesian priors. Work in approximate Bayesian inference will be useful [3,4,50].



## 2.2. Pairwise Correlations Crucial to Neural Pattern Formation

Our understanding of both the functional and structural connectivity of neurons and local neural groups has evolved well beyond the initial ideas proposed by Hebb [51]. An early important influence was the work of Edelman in proposing cortical organization with dynamic re-entry [52,53]. Singer's work led to early appreciation of long-range cortical activity synchronization [54]. A careful review by Honey et al. [55] advances theoretical arguments for how the brain's network topology and spatial embedding should strongly influence network dynamics.

One of the most powerful architectural themes of the neocortex is that of columnar organization, as described by Edelman and Mountcastle in *The Mindful Brain* (1978) [52], and particularly by Hubel and Wiesel regarding the visual cortex [56,57]. As Stevenson et al. note [2]: "One of the primary hypotheses about network structure is that neurons, especially cortical neurons, form groups or assemblies characterized by similar stimulus–response properties... [referring] not only to localized cortical columns, but also to functional assemblies that may be more spatially distributed."

Stevenson et al. show how functional connectivity between neurons can be identified using Bayesian priors in addition to point spike observations, further building a model-based approach to the identification of "effective connectivity", as introduced by Aertsen et al. [58]. Their analysis uses two suppositions about neural interactions (supported by their results, together with work by others): (1) neural connectivity is sparse, even within small functional ensembles where the same neurons may have similar response properties; and (2) the influence of one neuron on another varies smoothly over time.

Fries suggests that neural coherence mechanistically subserves neural communication [59]. With this notion, there would be an advantage to training neural connections as well as firing patterns that unify activations and responses across a set of communicating neurons. Mechanisms of neural adaptation and learning, leading to activation coherence, could serve an information theoretic (or communication) purpose.

Along these lines, neural coherence or mutual activation patterns can occur at different timescales and spatial ranges, contributing to nested or hierarchically-embedded behaviors. Tognoli and Kelso advocate a theory of Coordination Dynamics (CD), by which interactions occurring at different strengths and persisting over different timescales can lead to diverse metastable states [30].

Recent studies of both structural and functional connectivity have uncovered several topological themes, including "rich clubs" in which hubs are densely interconnected in a "hub complex" [60,61]. These hub complexes are a set of hub regions that are more densely interconnected than predicted by chance alone [62]. This extends the previously-known topographies, such as small-world and (to a limited extent) scale-free networks [63]. As Alagapan et al. have noted, hub-like structures promote information transmission within neural assemblies [64]. Such complex topologies can be supported by diverse intrinsic coupling modes, which provide multiscale ongoing connections between neural groups, even in the resting state [65].

Although there are indeed long-range neural connections, functional groups of neurons can be induced via multiple pairwise connections. As Schneidman et al. note, "small correlations among very many pairs could add up to a strong effect on the network as a whole", and they identify an Ising-like minimal model [66] which captures properties of large-scale networks [67]. Cohen and Kohn likewise review measures of correlation for pairwise connections between neurons, focusing on the "measure of the degree to which trial-to-trial fluctuations in response strength are shared by a pair of neurons" [68]. In a related line of work, Poli et al. have used a partial correlation approach to studying how structural connections underlie functional-effective ones in dissociated neuronal cultures, finding that the partial correlation approach is beneficial in determining both structural and topological features [69].

The significant role of pairwise interactions is further supported by Alegapan et al. [64], who have recently identified convergence and divergence properties of neurons, contributing to the

formation of what Sporns et al. have described as a “connectome” [70], noting that the “success of a model that includes only pairwise interactions provides an enormous simplification ...”

These arguments lead us to think that a statistical thermodynamics approach that combines pairwise interactions with the modeling of very local clusters will be useful. One crucial role of such a statistical thermodynamics-based model would be supporting coherent activities within neural ensembles. As Zalesky et al. comment on time-resolved resting state brain networks [71], “This result verifies that nonstationary fluctuations are synchronized across functional brain networks in such a way that multiple connections transition *en masse* ... [so that] dynamic fluctuations in functional connectivity at the pairwise level appear to be coordinated across the brain so as to realize globally coordinated variations in network efficiency over time...”.

Thus, it is reasonable to assume that a statistical thermodynamics model of the Ising-type, that is, invoking pairwise correlations, will be a useful starting point. The crucial step proposed in this work is that modeling cluster behaviors—that is, distributions of local patterns—will also be useful.

### 2.3. Neural Motifs, Topographies, and Pattern-Formation

The question that naturally arises is thus: How extensive should the local pattern models be? That is, even if we restrict our neural interaction models to pairwise interactions, do we need to model large local patterns to effectively model neural organization, or will very small local patterns suffice?

This is an important question, as Sporns and Kötter hypothesize that “... the connection patterns of real brain networks maximize functional motif number and diversity, thus ensuring a large repertoire of functional or effective circuits, while they minimize the number and diversity of structural motifs, thus promoting efficient assembly and encoding.” They further suggest that “brain networks maximize both the number and the diversity of functional motifs, while the repertoire of structural motifs remains small” [8].

An answer comes, at least in part, from studies identifying that neural motifs (or functionally- and structurally-connected units) on the order of  $M = 3$  (where  $M$  is the number of units) is a useful base. As noted, again by Sporns and Kötter: “Particularly interesting is the increased occurrence of a single motif at  $M = 3$  ... and its expanded versions at  $M = 4$  ... this motif type combines two major principles of cortical functional organization, integration and segregation ...”. As corroborated by Papo et al., “Brain graphs also have higher clustering than random graphs. For any graph, clustering can be quantified by counting the proportion of triangular motifs that completely connect a subset of three nodes” [72].

Out of all this work comes a leading role for the clustering of neurons and neural groups within the brain. In particular, Stevenson et al. note that “clustering is not trivial; that is, it is not purely random or a result of bad spike sorting,” and they hypothesize that “neurons in the same clusters may have similar relationships to external covariates, that is, similar common input” [2].

One important area of investigation that will have a strong impact on BCII will be the influence that structural connectivity has on functional organization. Gollo and Breakspear [73] identify two crucial features that shape cortical dynamics: the numbers of resonant pairs (within an ensemble) and frustrated motifs, respectively. They find that the strength of the coupling between pairs plays a crucial role: too strong a coupling, and the neural ensemble becomes globally connected; and too weak a coupling reduces variability, because too few pairs can become synchronized. They note that “spatial clustering leads to a modular architecture, in which nodes are strongly interconnected with regions within the same cluster (SC)—typically nearby in space—than with regions outside their cluster”.

Similarly, Gollo et al. have established the crucial role of resonance pairs within cortical motifs, noting that “the presence of a single reciprocally-connected pair—a *resonance pair*—plays a crucial role in disambiguating those motifs that foster zero-lag synchrony in the presence of conduction delays (such as dynamical relaying) from those that do not (such as the common driving triad). Remarkably,

minor structural changes to the common driving motif that incorporate a reciprocal pair recover robust zerolag synchrony" [74].

We can summarize the discussions of these last two subsections by noting that:

1. Although long-range (inter-cluster) correlations play a significant role, a model focusing on pairwise correlations (including the notion of resonance pairs) is a good starting place, and
2. Relatively small functional units or *motifs* (i.e.,  $M = 3$ , with correspondingly small structural units) play a strong role in neural activation patterns.

We go forward into investigating models with two minimal conditions:

1. Pairwise interaction modeling is essential, and although longer-range correlations are important, we can potentially create a starting point that focuses on local pairwise correlations, and
2. Cluster modeling should be included, and relatively small clusters ( $M = 3$ ) will suffice for correlation with valuable neurophysiological insights.

The following Section 3 overviews how various models—drawn from statistical thermodynamics—are useful for modeling neural activation patterns.

### 3. Statistical Thermodynamics for Neural Ensembles

Over the past several decades, statistical thermodynamics methods have grown in importance for modeling a wide range of neural dynamics and activations, at levels ranging from small ensembles up to large-scale behaviors [75]. Such methods have particularly been useful for modeling critical state phenomena in the brain, although many other applications have also been found.

The most well-known statistical thermodynamics method applied to neural ensemble modeling has been the classic Ising spin glass (Table 1). This has been used with success in several areas. Barton and Cocco, in particular, have used Selective Cluster Expansion (SCE) to model clusters in both cortical and retinal fields [76].

**Table 1.** Glossary of terms.

Term	Meaning
Configuration variable(s)	Nearest neighbor, next-nearest neighbor, and triplet patterns
Degeneracy	Number of ways in which a configuration variable can appear
Free energy	The thermodynamic state function $G$ ; where $G = H-TS$
Enthalpy	Internal energy $H$ results from both per unit and pairwise interactions
Entropy	The entropy $S$ is the distribution over all possible states
Equilibrium point	By definition, the free energy minimum for a closed system
Equilibrium distribution	Configuration variable values when free energy minimized for given $h$
Ergodic distribution	Achieved when a system is allowed to evolve over a long period of time
Interaction enthalpy	Between two unlike units, $\epsilon$ ; influences configuration variables
Interaction enthalpy parameter	Parameter $h$ is defined as $h = e^{\epsilon/4}$
Motif	Fundamental building block; size ( $M$ ) is equal to the number of vertices
Temperature	Temperature $T$ times Boltzmann's constant $k_B$ is set equal to one

However, one of the greatest constraints in using the simple Ising spin glass approach is that the entropy formulation addresses only the distribution of units into one of two states: active and inactive (or **A** and **B**). Recent developments in neural activation modeling address the formation of local patterns (discussed in the previous Section 2), such as are found with functional motifs. Thus, it is appropriate—and by now, necessary—to move beyond the simple Ising formulation.

Fortunately, there is a means for modeling a system's free energy that includes local pattern formation: the *Cluster Variation Method*. This method directly invokes not only whether or not a given neural unit is active or inactive, but also the local configurations of such units. Note that these "neural units" are not size-constrained; they represent the building blocks for pattern-modeling within an



ensemble. As such, they may be cortical columns, or local areas identified by a parcellization method, whether volume or surface-based [71]. It is this process of modeling local patterns via *configuration variables* that gives us a new tool for the characterization of not so much specific patterns, but rather pattern distributions within large ensembles. Presaging a topic that will be addressed more completely in following sections, there is an important role for this kind of advanced entropy modeling in BCI-related pattern recognition.

One of the key findings that will be presented in the following sections is that the distribution of local patterns (described via *configuration variables*) is a function of enthalpy parameters. In particular, for the specific case where  $x_1 = x_2 = 0.5$  (equiprobable distribution of units into states **A** and **B**, which may be thought of as active and inactive neural units), there is an *analytic solution* that gives the equilibrium values for the configuration variables in terms of a single parameter: the interaction enthalpy parameter  $h$ .

Although this paper deals exclusively with the case of equiprobable distribution of units into states **A** and **B**, it is useful to briefly mention how non-equiprobable **A** and **B** distributions would evidence themselves in terms of CVM free energy minima (stable solutions). These remarks are being given in advance of a more detailed manuscript which would have to follow this one, in which the specific effects of non-equiprobable distributions could be examined in detail. In brief, though, the impact would be as follows. First, there would need to be two enthalpy parameters,  $h_1$  and  $h_2$ . One of these ( $h_2$ ) would correspond to the single interaction enthalpy parameter  $h$  as we have been using it in this discussion. The other ( $h_1$ ) would be a measure of the enthalpy difference between the two states **A** and **B**. Second, the free energy minima (stable states) would have to be determined computationally, as the analytic solution presented here would not extend to the nonequiprobable case.

In advance of these computations, though, we have some sense of how the parameter values would impact unit distributions and cluster formation. Specifically, the more that  $h_1$  diverges from zero, the more that the population of neural units will be shifted into non-equiprobable values. That is, there would be considerably more units in one state versus the other.

Nevertheless, even in the case of non-equiprobable unit distributions, the fundamental influence of the  $h_2$  interaction enthalpy parameter would be consistent with what we will observe in this paper; higher values of  $h_2$  will encourage the formation of like-with-like clusters, and lower values will correspond to like-near-unlike nearest-neighbor pairings, which will result in a much lower appearance of like-near-like clusters.

Since the influence of the  $h_2$  interaction enthalpy parameter will be consistent (whether or not  $h_1 = 0$ ), this paper serves the role of introducing cluster formation and characterization (appropriate to modeling triplets of neural units), regardless of the proportionate amounts of active and inactive units.

In short, the ergodic distribution of local patterns (the distribution that would be observed if carried out over a sufficiently large sample, allowed to stabilize over a sufficiently long time) corresponds to either a one- or two-parameter enthalpy specification. This paper focuses on the particular case of equiprobable distributions, for which a single (interaction) enthalpy parameter is sufficient.

The value of this for BCI applications is twofold:

1. Directly modeling neural activation patterns; this characterizes local activation patterns across an ensemble, together with providing a simple means of modeling the time-evolution of local pattern distributions, and
2. As part of a pattern recognition engine interpreting neural activations, it provides a one- (or two-) parameter high-level pattern characterization, as well as a potential predictive component that can be interpreted via Bayesian belief probabilities.

A next step would obviously be to correlate the abstract model in  $h$ -space with specific changes in local neural activation patterns; this invokes correlation with functional networks.

This paper presents only a first step in the direction of what is admittedly a large-scale (and rather ambitious) goal. In order to introduce the method and make clear the mapping from configuration variables to the interaction enthalpy, it focuses exclusively on the case of equiprobable distribution into states **A** and **B**. While this is a substantive constraint, it allows us to work with analytic results and an associated well-defined mapping of equilibrium configuration values to an  $h$ -trajectory.

The remainder of this section sets the stage for addressing the more sophisticated CVM entropy formulation by summarizing significant related work on three important underlying topics:

1. Statistical thermodynamics of small-scale neural assemblies (such as neural columns), together with models of larger-scale neural assemblies,
2. Criticality and phase transitions in neural ensembles, and their associated role in maximizing rich information content, and
3. The role of criticality as it relates to complexity and entropy, providing a rationale for high-entropic neural systems.

### 3.1. Statistical Thermodynamics for Small-Scale and Larger-Scale Neural Assemblies

The most common approach to modeling the statistical thermodynamics of neural collections is the Ising spin glass model, which describes a system composed of bistate units; those which can be either “on” or “off”, or **A** or **B**. This approach leads very naturally to connections with information theory and complexity. The basic Ising model has been applied with some success to collections of neural units ranging from small- to large-scale.

At any scale, models of the statistical thermodynamics of neural systems are a corollary to the more widespread approach of directly modeling neurodynamics, for which the work by Wilson and Cowan is a strong progenitor [77]. Moran et al. have reviewed a set of such models, collectively identified as “dynamical causal modeling” (DCM), which leads to identifiable response states in response to distinct stimuli [78]. These models share the common feature of describing inputs to a given neuron (or neural population) in terms of neural masses and fields. Along these lines, Yang et al. have investigated a four-parameter deterministic Izhikevich neuron model, and have found that it has a wide information transmission range and is generally better at transmitting information than its stochastic counterpart [79].

While dynamic models are valuable, and in some cases lead to stationary or limit-cycle behaviors, the process of free energy minimization can also play an important role. As an example of small-scale modeling, Tkačik et al. have used a maximal entropy approach to model an interacting network of neurons [11,80,81]. They have found that pairwise interactions supplemented by a global interaction controlling synchrony distribution (“K-pairwise” models) extends the more common pairwise Ising models, and provides an excellent account of the data obtained from populations of up to 120 neurons from the salamander retina.

Moving to a larger scale, one of the most valuable contributions of statistical thermodynamics to modeling neuronal systems comes from the investigation of the dynamic properties of activated neuronal ensembles [82]. Indeed, Friston (2010) has proposed that free energy minimization serves as a unifying theory for the description of neural dynamics [83]. Maren has described the phase space for an Ising spin glass system allowing for all-or-nothing phase transitions, coupled with both hysteresis and metastable states, suitable for modeling neural columnar activations [5,84,85].

One of the most important recent findings is that statistical thermodynamics models (most typically the Ising model) can indicate a preferred topology for network organization. Deco et al. used a straightforward Ising spin network to model brain Resting State Networks (RSNs) with coherent and robust spatiotemporal organization [86]. They found that system entropy increases markedly close to a bifurcation point, and that the RSN topology most closely matches a scale-free network topology (with hubs), as compared to other potential topologies, such as random, regular,

or small-world. A key feature in their findings is that the scale-free topology (marked by central hubs) gives rise to a richer dynamic repertoire than is allowed with other topologies.

Deco et al. also found that the scale-free topology is associated with a specific small range of interaction parameters ( $h$ ) for the Ising network with simple entropy formulation. They constructed networks with various topologies and then computed the entropy associated with these networks, in order to find those with both maximal entropy and topologic similarity to connectivity patterns evidenced in imaging studies. Their thesis is that increasing the entropy of the system (with more attractor states) allows for greater richness in the allowable dynamical repertoire. They hypothesize that high-entropic states are thus crucial to allowing the brain to have greater latitude for computation.

Their work introduces a potential connection to a more comprehensive formulation of the entropy, in that a scale-free or hub-based topology may potentially be characterized by systems with a given  $h$ -value, where the  $h$ -value directly indicates certain *configuration variables* associated with specific topographies. This is one of the next steps in an evolving research program.

Thus, the process of free energy minimization can play a role in network behaviors. A neural system comes to a base state of equilibrium. A stimulus causes the system to respond, and it enters into a state where certain patterns become active. It holds this state for a period of time, and potentially even learns to stabilize this as a new equilibrium state; this can happen only if different enthalpy parameters characterize this state vis-a-vis the prior resting state. A system can be perturbed, but it would come to a new (or learned) pattern–response–equilibrium state.

Friston, describing how such perturbations can influence a neural system, notes that “The emerging picture is that endogenous fluctuations are a consequence of dynamics on anatomical connectivity structures with particular scale-invariant and small-world characteristics” and that these models “... also speak to larger questions about how the brain maintains itself near phase-transitions (i.e., self-organized criticality and gain control)” [87].

Movement towards a new state will be influenced by the nature of the stimulus or perturbation. Friston describes this, saying: “In short, the internal states will appear to engage in Bayesian inference, effectively inferring the (external) causes of sensory states. Furthermore, the active states are complicit in this inference, sampling sensory states that maximize model evidence: in other words, selecting sensations that the system expects. This is active inference...” [88].

### 3.2. Statistical Thermodynamics for Phase Transitions and Critical Points in Neural Assemblies

Over the past three decades, various researchers have evolved an understanding that the brain self-organizes to operate at a critical or near-critical threshold. There is a relationship between network topography and the brain’s ability to exhibit spontaneous dynamics (“neural avalanches”), so that different “connectivity rules” drive the brain towards different dynamic states. Massobrio et al. (2015) argue for this understanding with support drawn from a combination of in vitro studies together with theoretical investigations [89].

Kozma, Puljik, and Freeman “treat cortices as dissipative thermodynamic systems that by homeostasis hold themselves near a critical level of activity that is far from equilibrium but steady state, a pseudo-equilibrium” [90]. They use the notation for K (Katchalsky) sets based on Freeman’s notions of a hierarchical brain model, and employ a neuropercolation approach to “[model] brain dynamics ... as a sequence of intermittent phase transitions in an open system”.

Deco et al. similarly advocate the notion that Resting State Networks (RSNs) operate near a critical point, noting that “at rest, the brain operates in a critical regime, characterized by attractor ghosts or weak attractors in the Milnor sense. This leads to chaotic itinerancy—which may be augmented by random fluctuations and an exploration of multistable attractors (in the strong or classical sense). When a stimulus is encountered (or attentional set changes), one attractor becomes stable and is ‘selected’ from a repertoire of latent attractors” [86].

They follow this work with using a dynamic mean-field approximation for neural interactions, which approximates the temporal dynamics of the spiking network [91]. They were able to show that functional connectivity emerges from slow linear explorations in the region of system bifurcation. They suggest that small fluctuations around the spontaneous state better explain the resting state brain dynamics, compared with large excursions shaped by attractor basins. They comment that “... working at the edge of a critical point allows the system to rapidly compute a specific brain function by representing it in an attractor”. Freeman has summed this up by stating that “The cerebral cortex [achieves a critical state] by maintaining itself at the edge of chaos (Nicolis, 1987 [92,93]) in a state of criticality (Kozma et al., 2012 [90])” [94].

One implication for this, as we model brain states with a free energy minimization approach, is that small fluctuations around local pattern configurations can potentially be a means by which these “slow linear explorations” of the state space can occur.

There are, of course, opposing points of view on the nature of critical phenomena in the brain [95]. However, strong support for the notion that retinal circuits, at least, operate within the critical region comes from recent work by Kastner et al., who show that neural circuits operating near a critical point maximize information encoding [96]. With this connection between neural structural and functional architecture and network dynamics, it is reasonable to investigate how the local pattern distributions, as given by configuration variables, correspond with various cluster architectures (random, small-world, and scale-free [89]).

Massobrio et al., studying the self-organized criticality in cortical assemblies which occurs in concurrent scale-free and small-world networks, note that “The obtained results support the hypothesis that the emergence of critical states occurs in specific complex network topologies” [89].

Numerous other studies, most within the past few years, support the existence of critical points within close reach of ongoing neural dynamics. For example, Abum et al. find—by tuning background excitation to induce a Hopf bifurcation approach—that there is a dramatic increase in the autocorrelation length, depending sensitively on which neuronal subpopulation is stochastically perturbed [97].

There is also recent work suggesting that neural systems operate in the subcritical regime. Priesemann et al. find that—in contrast to the nominally suggested self-organizing critical (SOC) systems—neural avalanches observed in vivo imply that neural activity does not reflect a SOC state but a slightly sub-critical regime without a separation of time scales. They note that the potential advantages of operating in the subcritical regime may be faster information processing, together with a safety margin from super-criticality [98]. Tomen et al. find that emerging synchrony in the visual cortex of macaque area V4 yields an increase in selective discriminability. However, this only occurs within a narrow region in the phase space, at the transition from subcritical to supercritical dynamics [99].

Although there are a wide range of dynamic behaviors, the existence of long-term attractor states provides a stabilizing backdrop against which many other activities play out. Braun and Mattia describe various attractor states relevant to this work, including hierarchical nested attractors [100].

Recently, various researchers have introduced significant findings that strongly influence the design of a neuromorphic system that takes advantage of statistical thermodynamics principles. One example is that of Butler et al., introducing a key finding with implications for the design of neuromorphic systems as well as for the direct modeling of neural activation patterns. This finding is that in order to avoid destabilization (e.g., into hallucinogenic states), certain areas of the visual cortex require sparse long-range connections in addition to their local connectivity [101]. Butler et al. interpret this as an evolutionary adaptation, tuning the visual processes in the V1 cortex.

Cowan et al. (2013) examine a simple neural network that exhibits both self-organization and criticality, along with hysteresis. Using renormalization group analysis, they find that the network states undergo the neural analog of a phase transition in the universality class of directed percolation [102].

### 3.3. Criticality, Complexity, and Entropy: A Rationale for High-Entropic Systems

While the previous subsection focused on the role of criticality in brain systems, together with the role of entropy as the crucial mediating process, the notion of complexity is perhaps even more important for information processing within neural systems.

Nicolis introduces how biological systems differ from simpler mechanistic information storage and transmission systems by invoking the notion of complexity versus entropy [92,93,103]. He states regarding the notion of information that “It is rather emerging from the iterated map that the processor applies by way of conjugating externally impinging stimuli with internal activity...” [93]. In short, information is not the storage nor the time series itself, but rather the stable unfolding process; specifically, the “set of (stable) eigenfunctions of such a map—which for a dissipative nonlinear operator is simply the set of the coexisting attractors.”

In a different but related line of thought, Crutchfield and Young (1989) introduced an abstract notion of complexity, and devised a procedure for reconstructing computationally equivalent machines (or  $\epsilon$ -machines) whose properties lead to quantitative estimates of both complexity and entropy [104]. In short, certain lines of research lead to the association of neuronal connectivity (both structural and functional) not only with critical phenomena, but also with potential means for describing complexity and higher-level descriptions of information communication within the brain.

The CVM approach (including 2-D versions) yields critical behavior, and it is thus reasonable to consider using this approach when modeling neural behavior [105]. Connections between the CVM and various topologies need to be investigated, and a potential further linkage to complexity measures would be a follow-on endeavor.

Following lines of thought introduced by Friston [83], from the perspective of the free energy principle (i.e., the minimization of variational free energy), both self-organized criticality and complexity reduction are mandatory. This follows because the variational free energy can always be expressed as complexity minus accuracy, when the enthalpy in turn is interpreted as a log probability density. This means that minimizing free energy necessarily entails a minimization of complexity. In this context, complexity is defined as the Kullback–Leibler (KL) divergence between posterior and prior distributions. In other words, complexity is the degree to which the distribution is pulled away from the prior (equilibrium) distribution upon exposure to the environment [13,106,107].

Interestingly, the variational free energy can also be expressed as entropy minus enthalpy. The entropy term (under some simplifying—Laplacian—assumptions) corresponds to the curvature of the free energy. This means that a minimum free energy solution, in certain situations, occupies a regime of the free energy that has low curvature. In turn, this means that local perturbations induce unstable critical or slow fluctuations. In other words, minimizing free energy necessarily leads to a critical slowing [13].

The immediately following Section 4 introduces the *configuration variables*, and shows visually and intuitively how different sets of local patterns correspond with different values of the interaction enthalpy parameter  $h$ .

The two succeeding sections introduce the theoretical formalism of the CVM; first for a 1-D (Section 5), and then a 2-D array (Section 6).

## 4. Configuration Variables: Describing Local Patterns

This section describes the *configuration variables* used in the *cluster variation method (CVM)*; a method for modeling the free energy that takes into account not only distributions of units into **A** and **B** states, but also the distributions of local patterns given in terms of *configuration variables*.

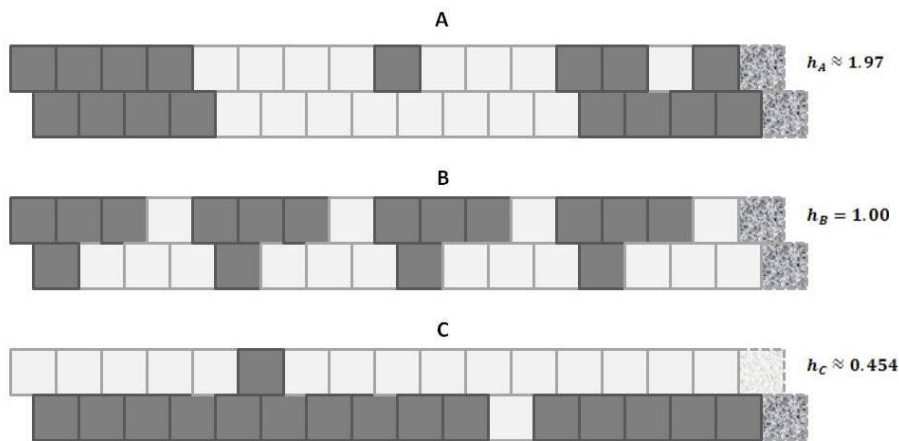
The configuration variables score the frequency with which various local configurations would be sampled from a lattice. In other words, they are the fraction of times they occur under random sampling. This means that they can be interpreted in terms of the probability of finding a particular configuration of one or more states when sampling at random or over long periods of time



(under ergodicity assumptions). We will appeal to their interpretation as probabilities later when linking the configuration variables to entropy.

#### 4.1. Configuration Variables in the Context of Entropy

Before introducing the formalisms, it may be useful to appeal to our intuition about the nature of entropy. The following Figure 1 shows a set of three patterns, each arranged as a 1-D zigzag chain. If we use only our classic entropy expression, which depends only on the distribution of units into *on* and *off* (or **A** and **B**) states, then the entropy for each of these patterns would be the same. In fact, for this illustration, the entropy for each would be at a maximum, since there are equal numbers of **A** and **B** units in each case.



**Figure 1.** Three 1-D zigzag chains of 32 units each. The far right mottled units indicate wrap-around neighboring units from the far left; that is, the two units at the far right of each pattern (on the top and bottom rows) represent the two units at the farthest left; they are shown on the far right in order to make visualization of the nearest-neighbor, next-nearest-neighbor, and triplet clusters easier. In patterns A and B, the two units on the far right are mottled dark grey, corresponding to the two leftmost units that are shown in dark grey. In pattern C, the uppermost unit on the far right is shown in mottled light grey, corresponding to the light grey uppermost unit on the far left of pattern C. The dark grey mottling of the far-right unit on the bottom row of pattern C corresponds to the dark grey unit on the far left bottom row. Each pattern (A–C) has equiprobable *on* (state **A**) and *off* (state **B**) units. The three patterns can be described as: (A) ferromagnetic-like, with one inserted non-like unit in each domain; (B) a set of four eight-unit patterns with an equilibrium distribution of configuration triplets (achieved when the nearest-neighbor interaction energy is 0); and (C) anti-ferromagnetic-like, with one inserted like-near-like (ferromagnetic) unit of each type. In each case, the nearest-neighbor connection is read as the diagonal between two units. The parameter  $h$  (far right) indicates the nature and strength of pairwise interactions. Computation of the  $h$  parameter values for the above figure (and related discussion) is given in the following Section 5. In particular, the  $h$ -values are explained in detail in Figure 5 and in the corresponding text.

Appealing to our intuitive sense about the nature of entropy, we immediately recognize that the top and bottom patterns in this figure are substantially *not* at maximal entropy; each of these two patterns expresses a high degree of order (albeit with a small pattern discontinuity in each). Our understanding of the concept of entropy tells us that neither pattern A (topmost) nor pattern C (bottom-most) is achieving a maximal distribution among all available states. (Obviously, the notion of *states* here has to be broadened to include patterns as well as distribution into **A** and **B** states.)

It is the middle pattern (B) of Figure 1 that intrigues us the most at this moment. As with both the top (A) and bottom-most (C) patterns, it is not entirely a maximal distribution among possible states;

we see that the same eight-unit pattern is repeated four times. However, despite this regularity, we notice that there is a greater diversity of local patterns in this middle example than in either the topmost or bottom-most examples.

The key to expressing this diversity among the types of local patterns, and to noting their resulting impact on the entropy term, is to consider the local *configuration variables*.

We have an analytic solution giving the configuration variables as functions of the interaction enthalpy  $h$  (details to follow in Section 5). This gives us a high-level interpretation for the kinds of patterns that we have within a dataset: are the nearest-neighbors typically like each other (ferromagnetic), or unlike (anti-ferromagnetic)? Using  $h$ , we know the probabilities of what kinds of unit will follow a given instance of a nearest-neighbor pair of units.

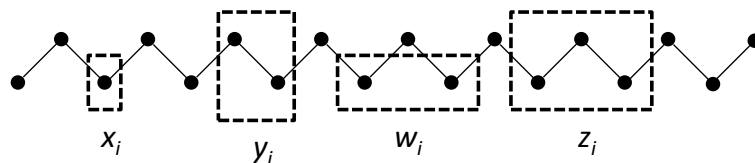
#### 4.2. Introducing the Configuration Variables

This subsection sets the stage for the 1-D CVM method by:

- Precisely articulating the *configuration variables*, and
- Identifying their inter-dependencies.

A following subsection will show how the  $h$ -values given in Figure 1 (on the far right of each pattern) were obtained.

Following [6,7], we create a 1-D lattice as a single zigzag chain composed of two staggered sets of  $M$  units each as shown in Figure 2. A single zigzag chain is used here for two reasons. First, it lends itself readily to constructing and depicting the next-nearest-neighbor and triplet configurations, and was in fact part of the original CVM formulation [6,7]. Second (and perhaps more significantly for our purposes), it provides a means of depicting  $M = 3$  functional and structural units, so it will be feasible in future work to map abstract neurophysiological networks to a CVM representation.



**Figure 2.** The 1-D single zigzag chain is created by arranging two staggered sets of  $M$  units each. The configuration variables shown are  $x_i$  (single units),  $y_i$  (nearest-neighbors),  $w_i$  (next-nearest-neighbors), and  $z_i$  (triplets).

Figure 2 illustrates the three configuration variables:

- $x_i$  - Single units,
- $y_i$  - Nearest-neighbor pairs,
- $w_i$  - Next-nearest-neighbor pairs, and
- $z_i$  - Triplets.

Figure 3 illustrates the *triplet* configuration variables  $z_i$  for a bistate system.

Notice that within this figure, the triplets  $z_2$  and  $z_5$  have two possible configurations each: **A-A-B** and **B-A-A** for  $z_2$ , and **B-B-A** and **A-B-B** for  $z_5$ .

The degeneracy factors  $\beta_i$  and  $\gamma_i$  (number of ways of constructing a given configuration variable) are shown in Figure 4;  $\beta_2 = 2$ , as  $y_2$  and  $w_2$  can be constructed as either **A-B** or as **B-A** for  $y_2$ , or as **B-A** or as **A-B** for  $w_2$ . Similarly,  $\gamma_2 = \gamma_5 = 2$  (for the triplets), as there are two ways each for constructing the triplets  $z_2$  and  $z_5$ . All other degeneracy factors are set to 1.

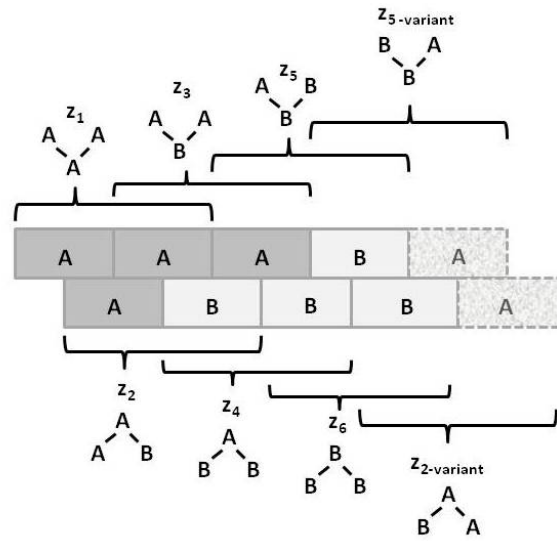


Figure 3. The six ways in which the configurations  $z_i$  can be constructed.

Pattern	Config. Variable <i>(Individual)</i>	Degeneracy
A	$x_1$	1
B	$x_2$	1

Pattern	Config. Variable <i>(Nrst Nbr)</i>	Degeneracy
$\begin{matrix} & A & \\ & / \backslash & \\ A & & A \end{matrix}$	$y_1$	1
$\begin{matrix} & B & \\ & / \backslash & \\ A & & A \\ & B & \end{matrix}$	$y_2$	2
$\begin{matrix} & B & \\ & / \backslash & \\ B & & B \end{matrix}$	$y_3$	1

Pattern	Config. Variable <i>(Next-Nrst Neighbor)</i>	Degeneracy
A -- A	$y_1$	1
A -- B	$y_2$	2
B -- A	$y_2$	2
B -- B	$y_3$	1

Pattern	Config. Variable <i>(Triplet)</i>	Degeneracy
$\begin{matrix} & A & \\ & / \backslash & \\ A & & A \end{matrix}$	$z_1$	1
$\begin{matrix} & A & \\ & / \backslash & \\ A & & B \\ & B & / \backslash \\ & A & & A \end{matrix}$	$z_2$	2
$\begin{matrix} & B & \\ & / \backslash & \\ A & & A \end{matrix}$	$z_3$	1
$\begin{matrix} & A & \\ & / \backslash & \\ B & & B \end{matrix}$	$z_4$	1
$\begin{matrix} & B & \\ & / \backslash & \\ B & & A \\ & A & / \backslash \\ & B & & B \end{matrix}$	$z_5$	2
$\begin{matrix} & B & \\ & / \backslash & \\ B & & B \end{matrix}$	$z_6$	1

Figure 4. The ways in which the configurations variables  $y_i$ ,  $w_i$ , and  $z_i$  can be constructed, together with their degeneracy factors  $\beta_i$  and  $\gamma_i$ .

The relations among the configuration variables are given in Appendix A.

#### 4.3. A Glance at $h$ : The Interaction Enthalpy Corresponds to Observed Configuration Values

In Figure 1, we observed three different illustrative sets of local patterns, each expressed as a 1-D zigzag chain.

Intuitively, we understood that each pattern had a different entropy, dependent on local pattern configurations:


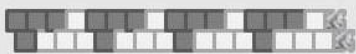

- A: Predominantly like-near-like (“ferromagnetic”), with two small exceptions,
- B: Diverse local patterns, representing the equilibrium distribution of patterns when the interaction enthalpy is set to zero, and
- C: Predominantly like-near-unlike (reading the nearest-neighbor as the *diagonal connection* (“anti-ferromagnetic”), again with two small exceptions.

This figure presented  $h$ -values associated with the three different patterns, A–C. Our intuitive sense of maximizing entropy at equilibrium leads us to think that pattern B would have the highest entropy of the three.

Now we are in a position to express our intuitive understanding in terms of the *configuration variables* and the values which they assume in each of these distinct patterns.

With formalizations for the various configuration variables ( $x_i$ ,  $y_i$ ,  $w_i$ , and  $z_i$ ) defined in the previous subsection, as well as in in Appendix A, we can now determine the  $z_i$  (the triplet fractional values) for each pattern (A–C). Additionally, we can now interpret how  $h$  (the interaction enthalpy parameter) corresponds to the  $z_i$  (triplet distributions).

Note that the values for  $h_{avg}$  (as shown in Figure 5) are computed as the mean of  $h_1$  and  $h_3$ , which are each based on computing  $h$  corresponding to the values for  $z_1$  and  $z_3$ , respectively. The  $h$  computations are done via a look-up table based on results given in Section 5.5, and specifically as shown in Figure 6. The values for  $h_1$  and  $h_3$  differ from each other slightly, because the example systems shown in Figures 1 and 5 are of such limited scale that a perturbation of even a single unit moves the system away from equilibrium. This difference between values would disappear for systems with larger total numbers of units.

Pattern	$Z_1$ [total]	$Z_3$ [total]	$z_1$	$z_3$	$h_1$	$h_3$	$h_{avg}$
A 	11	1	0.344	0.031	2.20	1.73	1.97
B 	4	4	0.125	0.125	1.00	1.00	1.00
C 	1	13	0.031	0.406	0.577	0.331	0.454

**Figure 5.** The previously-shown three patterns (A–C), as shown in Figure 1, can be characterized by their total values for  $Z_1$  and  $Z_3$  and their resulting fractional values,  $z_1$  and  $z_3$ . Specific values for  $h$  correspond to each  $z_i$  value. The resulting  $h_1$  and  $h_3$  values for a given pattern may differ, as they correspond to  $z_i$  values for systems that are not at equilibrium. The  $h_i$  values are averaged to produce  $h_{avg}$ .

Referring to Figure 5 (and also Figure 1), we notice that pattern A is largely like-near-like (or ferromagnetic), and has an  $h_{avg}$  value of 1.97. The bottom-most pattern C is largely composed of unlike pairings (reading nearest-neighbors as the units on the diagonal), and thus could be described

as antiferromagnetic. It has an  $h_{avg}$  of 0.454. The central pattern **B** has an equilibrium distribution of configuration variables, and an  $h_{avg}$  of 1.

As calculations in Sections 5 and 6 will show, the configuration variables are a function of the parameter  $h$ , and we can use a look-up table and extrapolation to obtain the approximate value for  $h$  given a certain set of configuration variables. For the moment, we use the computed  $h$ -values to gain an intuitive understanding of how the interaction enthalpy parameter  $h$  influences the equilibrium distribution of local patterns.

Thus, even before we proceed with the mathematical treatment, we note that  $h_{avg} > 1$  indicates that the overall free energy is minimized (a free energy equilibrium point is found) when nearest neighbors are like each other, and that  $h_{avg} < 1$  indicates that the free energy minimum is found when nearest neighbors are unlike. The central pattern B, where  $h_{avg} = 1$  corresponds to the case where the interaction enthalpy between nearest-neighbor pairs is zero, and thus the maximal distribution among possible local patterns is observed.

We thus have the beginning for characterizing systems in terms of their local pattern distributions using only a single parameter,  $h$ .

The following Section 5 provides the theoretical basis for what we have just observed.

## 5. The Cluster Variation Method for a 1-D Zigzag Chain

The cluster variation method (CVM) was introduced by Kikuchi in 1951 [6], refined by Kikuchi and Brush in 1967 [7], and further refined by Miyata et al. in 1975 [108]. Sanchez et al. have provided a generalized formalism for this method [109], and Mohri [110] has presented substantial theoretical developments, applying the CVM to free energy phase space diagrams for various alloys. The CVM is a means of considering the entropy of a system as being more than a simple distribution among the allowable states for individual units. Rather, it encompasses the local patterns, considering nearest-neighbor, next-nearest-neighbor, and other clusters.

Pelizzola [111] has reviewed the CVM and its many applications, including how it relates to graph theory, as has Yedidia et al., who have identified how the CVM relates to belief propagation [112].

Barton and Cocco used Selective Cluster Expansion—very similar to a CVM approach—to model neural activity in the retina and the prefrontal cortex [76]. They found that their results are closest to those found using a simple mean-field theory with strongest interaction parameters.

The Barton–Cocco result will prove interesting and useful, as we discuss how neural functional and structural organization (specifically, “rich-club” and similar network topographies) will likely correspond with specific  $h$ -values in Section 7.

### 5.1. The Classic Free Energy Formulation

A system reaches free energy equilibrium when the free energy is at a minimum. The free energy itself reflects interaction between two functions: the enthalpy and the entropy.

The free energy of a system is given as

$$G = H - TS \quad (1)$$

where  $G$  is the Gibbs free energy,  $H$  is the enthalpy,  $T$  is the temperature, and  $S$  is the entropy. The Gibbs free energy is one in which the system is closed, with constant pressure. The enthalpy term is a combination of internal energy ( $U$ ) and a pressure–volume term. Naturally, as we anticipate both constant pressure and volume in the brain, the term measuring pressure change goes to zero. For the remainder of this article, the Gibbs free energy will simply be referred to as the free energy.



### 5.1.1. The Enthalpy Term

In the earliest work on the Cluster Variation Method (CVM), Kikuchi found the free energy for his system using an enthalpy term given as:

$$H = N\varepsilon(-y_1 + 2y_2 - y_3) \quad (2)$$

where  $N$  is the total number of units in the system and  $\varepsilon$  is the interaction energy between two nearest-neighbor units.

The physical interpretation of the preceding equation is that the nearest-neighbor interaction between two like units ( $y_1$  and  $y_3$ ) is stabilizing, or has a negative coefficient, and an interaction between unlike units ( $y_2$ ) is destabilizing, or positive.

For this work, it is useful to shift the interaction energy base so that the interaction enthalpy between like units is zero, and the interaction enthalpy between unlike units ( $y_2$ ),  $\varepsilon$ , is constant. This allows us to rephrase the enthalpy equation as:

$$H = 2N\varepsilon y_2. \quad (3)$$

The free energy is then:

$$G = H - TS = 2N\varepsilon y_2 - TS \quad (4)$$

### 5.1.2. The Entropy Term

The classic entropy formulation, using a reduced form that will give a dimensionless expression for the entropy–temperature term, is given as

$$\bar{S} = S/(k_\beta N) = -\sum x_i \ln(x_i) \quad (5)$$

where  $k_\beta$  is Boltzmann's constant. Note that the set of variables  $x_i$  in this equation refers to single units, and does not as yet specify that the units should occur in either of states **A** or **B**. Further, this equation does not express the influence of configuration variables, which will be addressed shortly.

For a typical bistate system, the above gives

$$\bar{S} = -(x_1 \ln(x_1) + x_2 \ln(x_2)) \quad (6)$$

We also note, letting  $x_1 = x$  and  $x_2 = 1 - x$  (which is allowed, because  $x_1 + x_2 = 1$ ), that

$$\bar{S} = -(x \ln(x) + (1 - x) \ln(1 - x)) \quad (7)$$

We insert the entropy formulation into the previous free energy equation, and create an equation for  $\bar{G}$  (the reduced free energy), where  $\bar{G} = G/(Nk_\beta T)$ ; this allows us to work with dimensionless values. This yields

$$\bar{G} = 2\varepsilon y_2 / (k_\beta T) + x \ln(x) + (1 - x) \ln(1 - x) \quad (8)$$

where the term  $k_\beta T$  within the first term on the right hand side of the equation represents a constant that can be set equal to one, as we are dealing with an abstract representation of free energy, and not one tied directly to temperature.

## 5.2. The Entropy for a 1-D Zigzag Chain

To more formally generate the entropy term, we write the entropy of the system as the natural logarithm of the Grand Partition Function  $\Omega$ :

$$S = k_\beta \ln \Omega \quad (9)$$

where  $\Omega$ , the degeneracy factor (Grand Partition Function) is the number of ways of constructing the system in such a way that the fraction variables take on certain values.

Following the approach introduced by Kikuchi [6,7], we begin by considering a one-dimensional system composed of a single zigzag chain, as was shown in Figure 2.

Details for constructing the Grand Partition Function from which we obtain the entropy term are presented in Appendix A. The resulting entropy expression for the 1-D CVM is:

$$S_{1-D} = k_{\beta} \ln \Omega_{double} = 2k_{\beta} \left[ \sum_{i=1}^3 \beta_i Lf(y_i) - \sum_{i=1}^6 \gamma_i Lf(z_i) \right], \quad (10)$$

where  $Lf(x) = x \ln(x) - x$ .

### 5.3. The Free Energy Equation for the 1-D Zigzag Chain

We define the reduced free energy for a 1-D system,  $\bar{G}_{1-D}$ , in terms of the free energy for a 1-D system as

$$\bar{G}_{1-D} = G_{1-D} / (Nk_{\beta}T), \quad (11)$$

where as before, the term  $k_{\beta}T$  will be set equal to one. The total number of units in the 1-D system is  $N = 2M$ , as the system is comprised of two staggered sets of length  $M$  each in order to create a single zigzag chain (see Section 4.2, and also Figure 2). We thus create a reduced equation, normalized with respect to both temperature and the total number of units. The temperature, along with Boltzmann's constant, will be subsumed into the enthalpy parameter, creating a dimensionless term representing interaction strength between two nearest-neighbor units.

We introduce both the entropy for the 1-D chain as well as the pairwise enthalpy into the free energy equation, where the entropy term now not only represents the distribution of units into active/inactive states, but also the distribution of local patterns or configurations. We take the case where the enthalpy-per-active unit is zero, and thus only deal with the interaction enthalpy term.

We also make use of a substitution expressing the configuration variables in terms of each other (identified in Appendix A):

$$y_2 = z_2 + z_4 = z_3 + z_5 \quad (12)$$

$$2y_2 = z_2 + z_4 + z_3 + z_5 \quad (13)$$

This results in a much more complex reduced equation than the standard Ising (Bethe–Peierls) formulation [113,114] (notation drawn from [115,116], see [117]), and where the term  $k_{\beta}T$  has been set equal to one.

$$\begin{aligned} \bar{G}_{1-D} = G_{1-D} / N = & \\ & \varepsilon(z_2 + z_3 + z_4 + z_5) - 2 \sum_{i=1}^3 \beta_i Lf(y_i) + 2 \sum_{i=1}^6 \gamma_i Lf(z_i) \\ & + \mu \left( 1 - \sum_{i=1}^6 \gamma_i z_i \right) + 4\lambda (z_3 + z_5 - z_2 - z_4) \end{aligned} \quad (14)$$

where  $\mu$  and  $\lambda$  are Lagrange multipliers.

Taking the derivative of  $\bar{G}$  with respect to the six configuration variables  $z_i$ , and setting each derivative equal to zero yields the following six equations, presented in detail in Appendix A:

$$\begin{aligned}
z_1q &= y_1 \\
z_2q &= (y_1y_2)^{1/2}e^{-\varepsilon/4}e^\lambda \\
z_3q &= y_2e^{-\varepsilon/2}e^{-2\lambda} \\
z_4q &= y_2e^{-\varepsilon/2}e^{2\lambda} \\
z_5q &= (y_2y_3)^{1/2}e^{-\varepsilon/4}e^{-\lambda} \\
z_6q &= y_3
\end{aligned} \tag{15}$$

where  $q = e^{-\mu/2}$ , and  $\mu$  can be shown to be (for chemical systems) the chemical potential.

#### 5.4. Analytic Solution for the CVM for the 1-D Zigzag Chain

Achieving the analytic solution involves solving a free energy equation for the equilibrium point, where the enthalpy-per-active-unit is given as zero, and the pairwise-interaction enthalpy is given as  $\varepsilon$ . Because the enthalpy-per-active-unit is zero, there is no a priori preference for a unit to be in either the **A** or **B** states; thus,  $x_1 = x_2 = 0.5$ .

This gives a symmetric distribution to the other configuration variables; specifically:

$$\begin{aligned}
y_1 &= y_3 \\
z_1 &= z_6 \\
z_2 &= z_5 \\
z_3 &= z_4
\end{aligned} \tag{16}$$

The resulting solution gives the equilibrium configuration variables  $y_i$  and  $z_i$  in terms of  $h$ , where  $h = \exp(\varepsilon/4)$ .

For the system where  $x_1 = x_2 = 0.5$  and  $\lambda = 0$ , we can solve the previous equations, giving  $z_iq$  in terms of the  $y_i$  and other parameters to obtain the fraction variables  $y_i$  and  $z_i$ . The calculations, briefly summarized in the following paragraphs, are presented in more detail in Appendix A.

Let  $h = e^{\varepsilon/4}$ , and  $s = z_1/z_3$ . Then, as a step in the derivation, we obtain

$$h^2 = s \frac{1 + s^{1/2}}{s + s^{1/2}} \tag{17}$$

or

$$s = h^4 \tag{18}$$

The result is:

$$1/z_3 = 2(s + 1) \pm 4s^{1/2} \tag{19}$$

A slightly expanded summary is found in in Appendix A. A fully-detailed derivation is found in [9].

#### 5.5. Results for the 1-D CVM

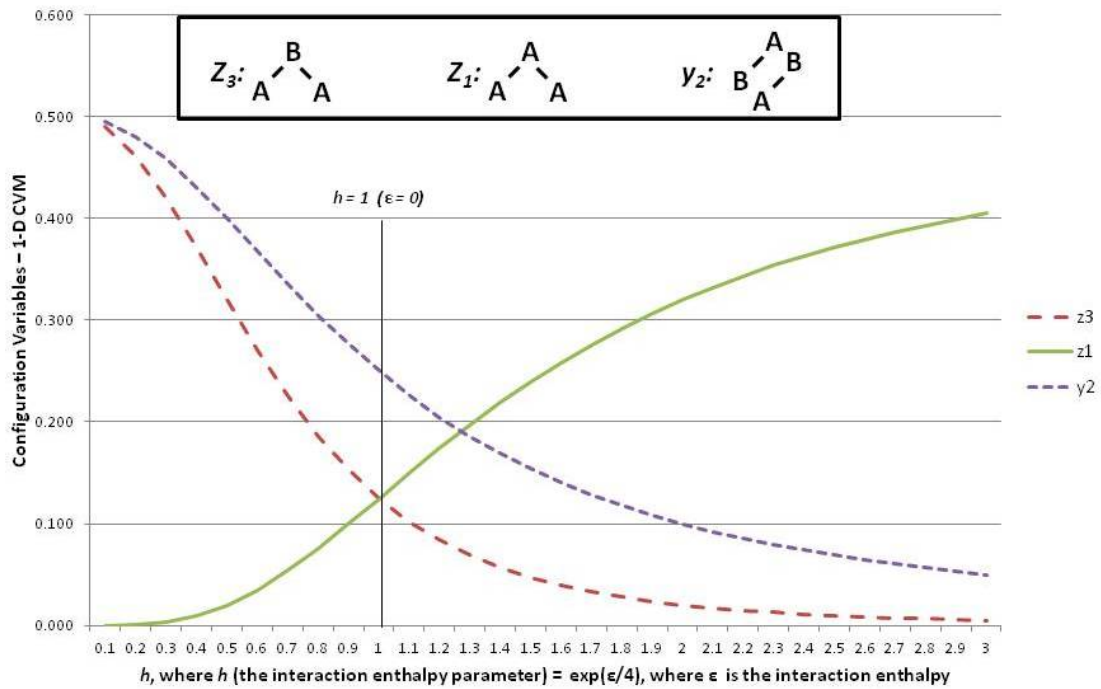
In order to introduce the overall CVM approach, this subsection continues to focus on the 1-D CVM as the simpler and more intuitive result. The following Section 6 introduces the 2-D CVM, for which the equations are somewhat more complex. However, similar lines of thought will apply.

This subsection shows how the configuration variables  $y_i$  and  $z_i$  are mapped to a single  $h$ -value at equilibrium. It also reveals how to interpret the configuration variables as functions of  $h$ .

The analytic solution for the 1-D CVM starts by giving one of the configuration variables ( $z_3$ ) as a function of the reduced interaction energy term  $h$ . From this, the remaining configuration variables are found as functions of  $h$ . The full details of this work were presented in Maren [9].

A similar analytic solution for the 2-D CVM has been presented in Maren [10]. Both analytic results are based on original derivations in Maren [85].

Figure 6 shows three of the configuration variables,  $z_1$ ,  $z_3$ , and  $y_2$ , in terms of the parameter  $h$ .



**Figure 6.** Results for three of the configuration variables,  $z_3$ ,  $z_1$ , and  $y_2$ , used in the cluster variation method. Values for  $h$  are plotted along the x-axis.

The following three subsections assist in interpreting the results shown in Figure 6; these are the cases where the interaction enthalpy:

1.  $\epsilon = 0$  so that  $h = e^{\epsilon/4} = 1$ ,
2.  $\epsilon < 0$  so that  $h = e^{\epsilon/4} < 1$ , and
3.  $\epsilon > 0$  so that  $h = e^{\epsilon/4} > 1$ .

### 5.6. Configuration Variables When the Interaction Enthalpy $\epsilon = 0$ ( $h = 1$ )

We begin by noting (in the slightly left-of-center portion of Figure 6 where  $h = 1$ ) that the observed configuration variable values are exactly as we would expect. Specifically,  $z_1 = z_3 = 0.125$ , and  $y_2 = 0.250$ . This fits precisely with our expectation for the configuration variables at equilibrium when  $\epsilon = 0$  ( $h = e^{\epsilon/4} = 1$ ).

To understand intuitively that this would be the expected result, we reflect on the total number of triplet states available, together with their degeneracies. We use various normalization equations from Appendix A as well [9], and particularly note that

$$1 = x_1 + x_2 = \sum_{i=1}^6 \gamma_i z_i$$

Since we also have  $x_1 = x_2 = 0.5$ , as well as the symmetry results from the end of the last section, we have

$$0.5 = y_1 + y_2 = y_2 + y_3 = \sum_{i=1}^3 \gamma_i z_i$$

We also have that the degeneracy factors are  $\beta_2 = \gamma_2 = \gamma_5 = 2$ , with all other  $\beta_i$  and  $\gamma_i$  set to 1. At equilibrium with  $\epsilon = 0$ , we expect equiprobable distribution of the  $z_i$  among all possible states, so that we expect  $z_1 = z_3 = 0.125$ , as observed in Figure 6. Additionally (not shown in the Figure), we would expect that  $z_2 = 0.250$ , so that  $z_1 + z_2 + z_3 = 0.5$ .

### 5.7. Configuration Variables When $\varepsilon < 0$ ( $h < 1$ )

At the left-hand-side of the preceding Figure 6, we have the case where  $h = e^{\varepsilon/4} < 1$ . These small values for  $h$  means that  $\varepsilon$ , the interaction enthalpy between two unlike units (**A-B**), is negative. This means that we stabilize the system by providing a structure that emphasizes alternate units (e.g., **A-B-A-B**), as was shown in Pattern C of Figure 1. (Recall that the nearest-neighbor connections are read as the diagonal nearest-neighbors in this zigzag chain.)

This is precisely what we observe. The pairwise combination  $y_2$  (**A-B**) increases beyond the nominal expectation (when there is no interaction energy), so that  $y_2 \rightarrow 0.5$ , notably when  $h < 0.1$ .

As a natural consequence of the increase in  $y_2$  when  $h \rightarrow 0$ , we also have  $z_3 \rightarrow 0.5$  (maximizing **A-B-A** triplets) and also  $z_1 \rightarrow 0.0$  (minimizing **A-A-A** triplets).

### 5.8. Configuration Variables When $\varepsilon > 0$ ( $h > 1$ )

Consider the case of a positive interaction energy between unlike units (the **A-B** pairwise combination). This is the case on the right-hand side of Figure 6, where  $\varepsilon > 0$  then yields  $h = e^{\varepsilon/4} > 1$ .

The positive interaction energy ( $\varepsilon > 0$ ) suggests that a preponderance of **A-B** pairs ( $y_2$ ) would destabilize the system, as each **A-B** pair would introduce additional enthalpy to the overall free energy, which is at a minimum in the equilibrium state. We would thus expect that as  $\varepsilon$  increases as a positive value, we would see a decrease in  $y_2$ , and also see smaller values for those triplets that involve non-similar pair combinations. That is, we would expect that the **A-B-A** triplet, or  $z_3$ , would approach zero.

This is exactly what we observe this on the right-hand side of the Figure 6.

Thus, when  $h \gg 1$  (or  $\varepsilon \gg 0$ ), we see that  $z_3$  falls towards zero, and  $y_2$  decreases as well. Note that  $y_2$  can never go towards zero, because there will always be some **A-B** pairs in a system that contains units in a mixture of states **A** and **B**. Correspondingly, this is also the situation in which  $z_1 = z_6$  becomes large; for example, we see that  $z_1 > 0.4$  when  $h > 3.0$ . Notice that  $z_1$  cannot approach 0.5, as  $z_3$  did for the case where  $h \ll 1$ . The reason is that  $z_1$  represents **A-A-A** triplets, and we will always have some **A-A-B** and **B-A-A** triplets, because we have equiprobable distribution of units into states **A** and **B**. Thus, while we could conceivably have constructed a system composed exclusively of **A-B-A** and **B-A-B** triplets, we cannot do so exclusively with **A-A-A** and **B-B-B** triplets.

The realm of  $h \gg 1$  is one in which we observe a highly structured system where large “domains” of like units mass together. These large domains (each comprised of either overlapping **A-A-A** or **B-B-B** triplets) stagger against each other, with relatively few “islands” of unlike units (e.g., the **A-B-A** and **B-A-B** triplets). This is the case observed in Pattern A of Figure 1.

Naturally, this approach—using a “reduced energy term”, where we had set  $\bar{G} = G/(Nk_{\beta}T)$  and the interaction enthalpy term was similarly reduced to  $\varepsilon/(k_{\beta}T)$ , does not tell us whether we are simply increasing the interaction energy or reducing the temperature; they amount to the same thing. Both give the same resulting value for  $h$ , and it is the effect of  $h$  that interests us when we map the CVM variables and (ultimately) the CVM phase space. Thus, because we are working in an abstract space, we have set  $k_{\beta}T = 1$  throughout the various steps. However, the role of temperature has been subsumed into the interaction enthalpy. Kastner et al. [96] identify that “The average noise across the neural population plays the role of temperature in the classic theory of phase transitions”.

## 6. The Cluster Variation Method for a 2-D Zigzag Chain

The preceding section served as a brief introduction and tutorial on how the interaction enthalpy parameter  $h$  influenced the distribution of the configuration variables. For ease of visualization, we examined the configuration variables in the context of a 1-D system, or a single zigzag chain.

For the purposes of modeling neural systems, we are far more interested in a 2-D system. This section presents the results for an analytic solution of the free energy minimum for the 2-D CVM when there is an equiprobable distribution of units into the **A** and **B** states.



### 6.1. Modeling Activation Patterns in 2-D Neural Assemblies

There is ample motivation for modeling patterns in 2-D systems:

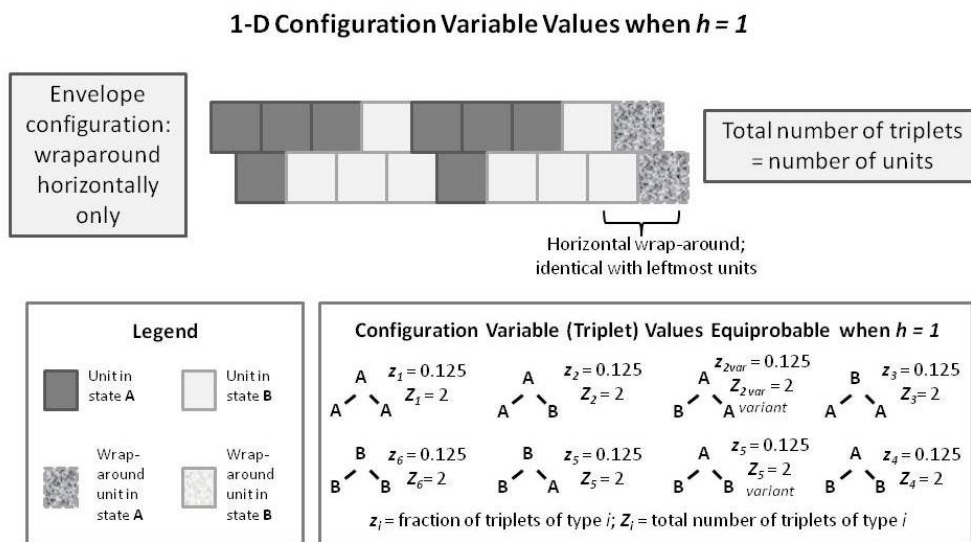
- Many neural patterns of activity are best modeled as dynamics in a 2-D system [118],
- Information is encoded as time-varying patterns in such 2-D arrays,
- The common approach to modeling information content addresses only distribution across two states (“on” and “off”, or **A** and **B**), not as patterns; a more comprehensive pattern characterization may prove to be a useful correlation to information measures.

We have known for some time that the brain makes nearly ubiquitous use of 2-D maps for information encoding. What we are understanding more recently is exactly how different types of neural connectivity operate within the 2-D maps to provide certain kinds of functional effectiveness. For example, Harris and Shepherd have described different neocortical circuit classes, including local and long-range connectivity along with other characteristics [119]. By characterizing these circuit classes, which are widespread in their appearance across different cortical areas, we have a more abstract description for local cortical organization. It may be possible to associate different cortical classes with specific local configuration variable distributions.

We also note that 2-D topographic maps are crucial not only for single-sensor modality representation, but also for sensor fusion, as initially identified in ground-breaking work by Stein and Meredith [120], and more recently comprehensively overviewed by Stein [121]. Thus, we need to model the statistical thermodynamics associated with pattern formation in 2-D systems.

### 6.2. Configuration Variables for the 2-D CVM

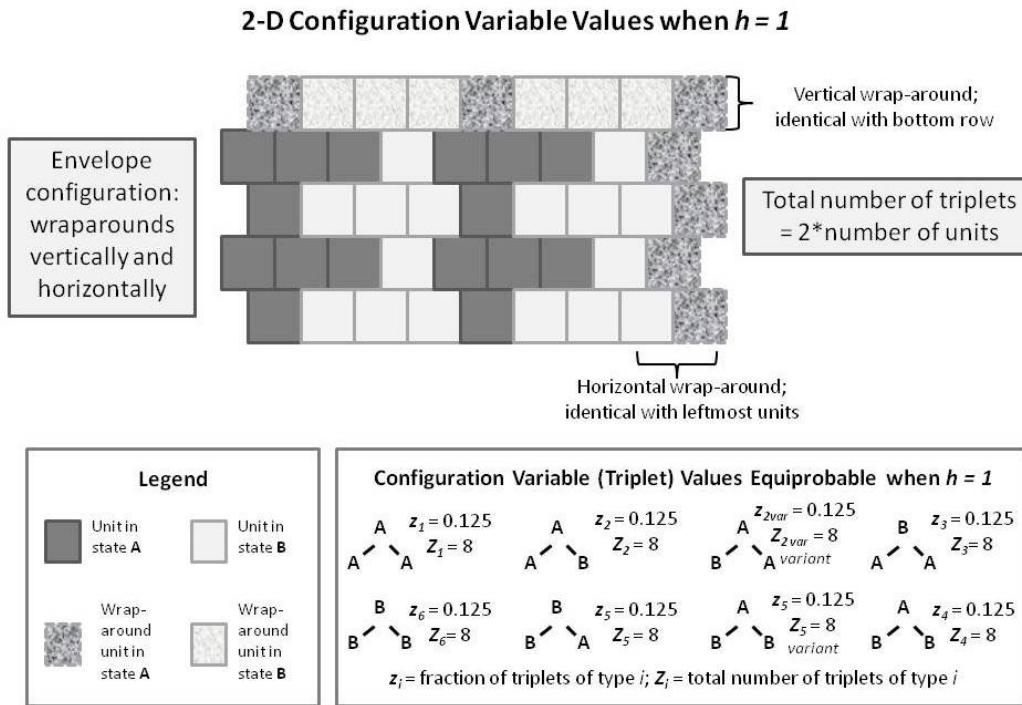
To introduce the 2-D CVM, we first examine a short 1-D CVM configuration, as shown in Figure 7. This figure illustrates (in an abbreviated sense) the same kind of pattern that we used as pattern B in Figures 1 and 5. This figure shows us a pattern in which each of the  $z_i$  triplets appears in their equilibrium values for the case where  $h = 0$ . In particular, note that the total number of triplets (as countable numbers,  $Z_i$ , not just as the fractional values) is equal to the total number of units present. This is true only in the case where we have a horizontal wrap-around, so that there is no discontinuity at the borders.



**Figure 7.** A 1-D CVM configuration (single zigzag chain), together with both the total numbers and fractional amounts for the configuration variables  $z_i$ . Notice that there is a horizontal wrap-around for the computation of the configuration variables. The total number of triplets is equal to the total number of units in the system.

We compare this with a similarly-patterned 2-D CVM system, as shown in the following Figure 8. In this figure, we have built up the 2-D overall configuration by adding two more single rows on top of the original single zigzag chain, creating a set of units that is double the original size. In order to maintain border continuity as we count the various configuration variables, we do an envelope-style wrapping vertically as well as horizontally.

It is worth noting here that the total number of countable triplets  $Z_i$  is now *twice* the number of units in the system. Essentially, we can now count connections in both an up-and-down sense as well as sideways (all expressed on the diagonal, of course), which doubles the number of countable triplets.

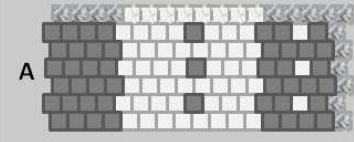




**Figure 8.** A 2-D CVM configuration (built by superimposing two single zigzag chains), together with both the total numbers and fractional amounts for the configuration variables  $z_i$ . Note that the wraparound in this case forms a full envelope; the configuration variables are computed in a smooth manner, wrapping both horizontally and vertically. The total number of triplets is equal to two times the total number of units in the system.

With the previous Figure 8 as a guide to interpreting 2-D CVM patterns, we now examine a 2-D extension of the patterns offered earlier in the 1-D CVM. Previous Figures 1 and 5 showed how the average  $h$ -values were associated with three different types of patterns (A–C). We now examine a similar (albeit 2-D instead of 1-D) representation in Figure 9.

Once again, we used the actual  $z_1$  and  $z_3$  values for each pattern to identify corresponding  $h$ -values, as we did earlier in the 1-D CVM (the following subsection gives highlights of the derivation for expressing configuration variables in terms of  $h$  for the 2-D CVM).

Just as we found in the case of the 1-D CVM, a higher value of  $h_{avg}$  corresponds to the “like-near-like” pattern, whereas the lower value for  $h_{avg}$  corresponds to the “like-near-unlike” pattern. Additionally, as before, when  $h = 1$ , there is no interaction enthalpy between neighboring units, and thus the system falls into an equiprobable distribution of local configurations.

Pattern	$Z_1$ [total] A A	$Z_3$ [total] B A	$z_1$	$z_3$	$h_1$	$h_3$	$h_{avg}$
	66	6	0.344	0.031	2.395 <small>(cf 2.20 for 1-D case)</small>	2.445 <small>(cf 1.73 for 1-D case)</small>	2.42 <small>(cf 1.97 for 1-D case)</small>
	24	24	0.125	0.125	1.00	1.00	1.00
	6	78	0.031	0.406	0.575 <small>(cf 0.577 for 1-D case)</small>	0.332 <small>(cf 0.331 for 1-D case)</small>	0.454 <small>(cf 0.454 for 1-D case)</small>

**Figure 9.** Replicating the kind of patterns shown previously for the 1-D CVM, the same patterns are extended to the 2-D CVM system. The actual fractional values for the various triplets are the same as for those used in the previous 1-D CVM figures.

### 6.3. Free Energy in a 2-D Configuration

The equation for the free energy in a 2-D system, including configuration variables in the entropy term, is

$$\begin{aligned} \bar{G}_{2-D} = G_{2-D}/N = & \varepsilon(z_2 + z_3 + z_4 + z_5) \\ & - \left[ 2 \sum_{i=1}^3 \beta_i Lf(y_i) + \sum_{i=1}^3 \beta_i Lf(w_i) - \sum_{i=1}^2 \beta_i Lf(x_i) - 2 \sum_{i=1}^6 \gamma_i Lf(z_i) \right] \\ & + \mu \left( 1 - \sum_{i=1}^6 \gamma_i z_i \right) + 4\lambda(z_3 + z_5 - z_2 - z_4) \end{aligned} \quad (20)$$

where (as before)  $\mu$  and  $\lambda$  are Lagrange multipliers, and (as we did in the previous Section) we have set  $k_\beta T = 1$  (note: the full derivation of the 2-D CVM free energy is presented in [10], and the preceding equation corresponds to Equations (2)–(14) of that reference).

We note that in comparison with Equation (14), the equation for the 2-D CVM system (Equation (20)) includes terms in  $w_i$  (next-nearest-neighbors) and  $x_i$  (simple distribution into states A and B, which are set to be equiprobable for obtaining the analytic solution for minimum free energy).

Further, the enthalpy term in the preceding equation really should have a factor of two in it; that is,  $2\varepsilon(z_2 + z_3 + z_4 + z_5)$ . This is because when we move from a 1-D to a 2-D CVM, we double the number of configuration variables such as  $y_i$  and  $z_i$ , as can be seen via visual examination of Figure 8 in comparison with Figure 7.

However, the values for parameter  $h$  itself are those that are the best means for comparing behaviors of the 1-D versus 2-D CVM equations, and it is more useful to simply subsume the factor of two into the term  $2\varepsilon$ . As we will typically work within a 2-D CVM framework, the 1-D results are useful more as an introductory pathway than as an actual comparison of system dynamics. Our attention will ultimately be on those values of  $h$  (in the 2-D CVM) that give us

system configurations that bear resemblance to interesting network topologies, such as small world, rich-club, etc.

Following methods very similar to those used to find the analytic solution to the 1-D CVM free energy at equilibrium, and again for the specific case where  $x_1 = x_2 = 0.5$ , we obtain an analytic solution for the configuration variables.

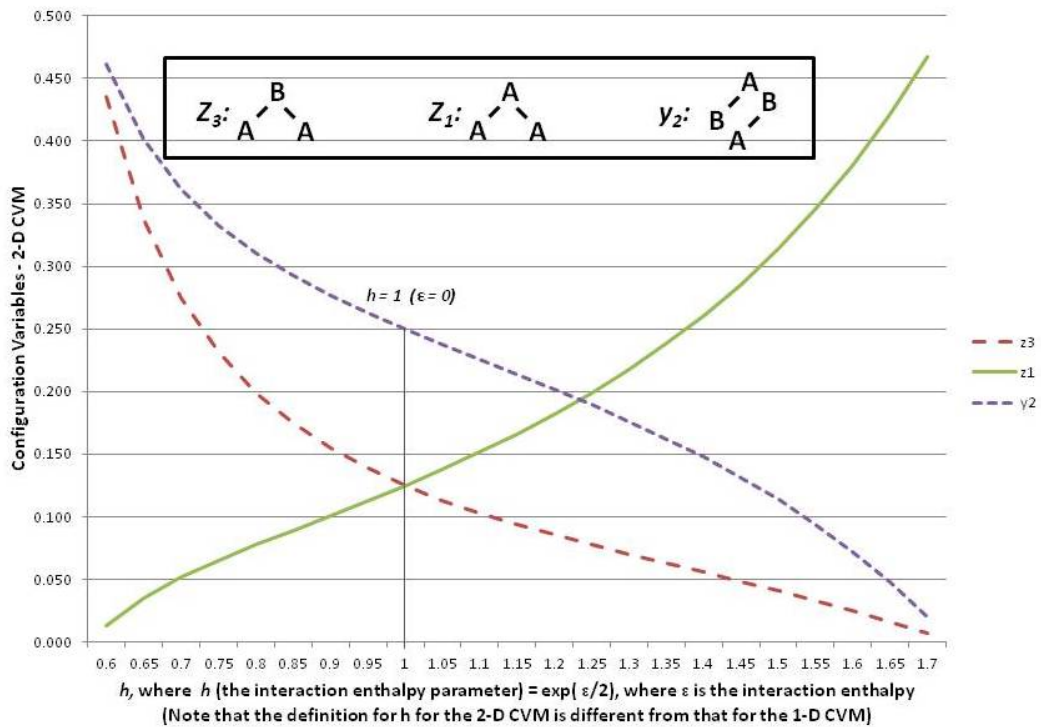
The solution for the specific configuration variable  $z_3$  is given as

$$z_3 = \frac{(h^2 - 3)(h^2 + 1)}{8[h^4 - 6h^2 + 1]} \tag{21}$$

where  $h$  has the same interpretation as in the previous section; that is,  $h = e^{\epsilon/4}$ . Note that the preceding equation corresponds to Equations (2)–(17) of [10].

#### 6.4. Results for the 2-D CVM

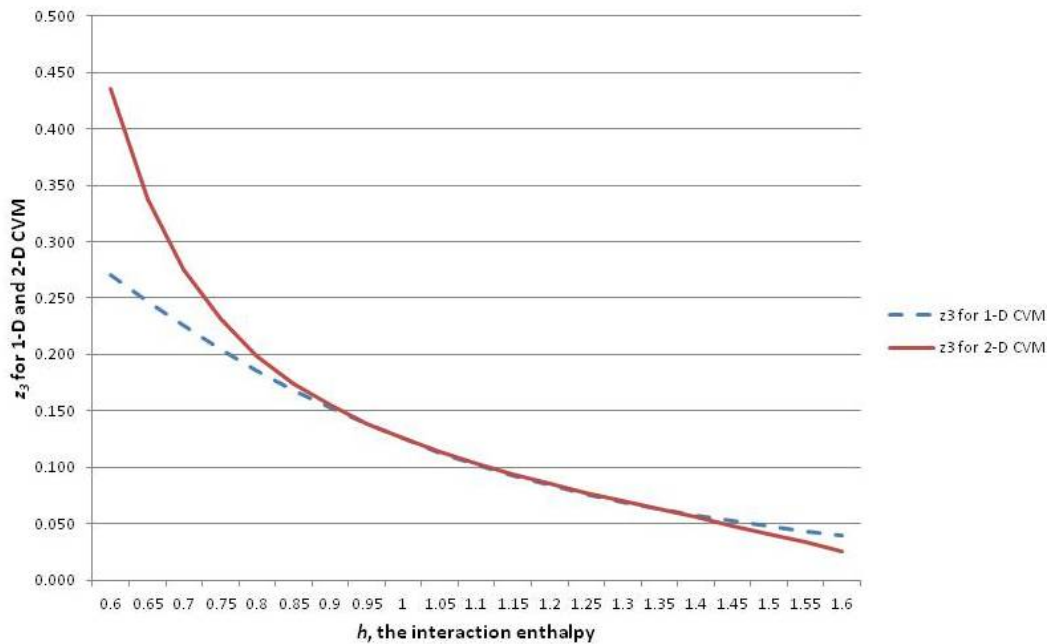
Figure 10 shows the results for the triplet configuration variable  $z_3$  for both the 1-D and 2-D CVM at equilibrium (for the case where distributions into states A and B are equiprobable).



**Figure 10.** Comparison of configuration variable  $z_3$  for the 1-D and 2-D CVM. Both values for  $z_3$  are the expected 0.125 when  $h = 1$ .

The analytic solution for the configuration variables in the 2-D CVM has divergence at two points, due to the presence of a quadratic term in the denominator. However, at values within a reasonable range of  $h = 1$ , we have properties similar to those of the 1-D CVM configuration variables. Most specifically, we obtain the desired values at exactly  $h = 1$ , and the values for the configuration variables trend as would be expected (until divergence causes less-than-welcome behaviors).

The following Figure 11 shows the values for  $z_3$  in the neighborhood of  $h = 1$ , for both the 1-D and 2-D CVM cases.



**Figure 11.** Comparison of configuration variable  $z_3$  for the 1-D and 2-D CVM. Both values for  $z_3$  are the expected 0.125 when  $h = 1$ .

## 7. Discussion

This section briefly highlights important gleanings from the previous results, and overviews an intended research program, examining how a 2-D CVM can be used, and including two aspects:

1. Neural interactions giving rise to specific and changing  $\varepsilon$  values, and
2. Implications for information content in patterns of neural activation.

One of the most interesting findings to emerge in studies of neural architecture is that “rich club” topographies play a strong role, as discussed previously in Section 2.2. As described in Nigam et al. (2016), “Neurons with the highest outgoing and incoming information transfer were more strongly connected to each other than by chance, thus forming a ‘rich club’” [122].

When the rich node coefficient for a network is close to one, it signifies that high degree nodes of the network are well connected with each other. This means that the nodes which have greatest connectivity are mostly connected to other nodes which similarly have high connectivity, thus forming a densely-connected set of nodes, or a “rich club.”

Let us say that the nodes in state **A** are those that participate in a “rich club.” We would then expect those nodes to be strongly connected with each other, and not well-connected with nodes that are not in that state. If we were to suggest that proximal nearest-neighbor placement on a 2-D lattice would be conducive to forming such within-club connections, we would anticipate that we would have a network that largely has a “ferromagnetic-like” character; that is, like-near-like.

We would thus expect a topography in which  $z_1$  (**A-A-A** triplets) is relatively high, and  $z_3$  (**A-B-A**) is relatively small (see Figure 9). Thus, referencing Figure 10, we would expect to have  $h$ -values that are on the right-hand side of the graph; i.e.,  $h > 1$ , or equivalently,  $\varepsilon > 0$ .

To be a bit more specific in our expectations, we refer again to Figure 9, and note the  $h$ -values corresponding to patterns A and B. For pattern B—which has a completely equiprobable distribution of triplets— $h = 1$ . For pattern A—which has two large masses of self-similar units, with just a few singleton-islands of opposing units within the larger masses—the  $h$ -value (average) is 2.42. Clearly, the kind of  $h$ -values that will govern brain networks will be somewhere between these two extremes.



As evidenced by the various works cited in prior sections, there is ample evidence that neural units in the brain organize into clusters that behave in a coherent manner. Thus, we do not expect anything like a completely random or equiprobable distribution of units. In short, we do not expect that  $h$ -values characterizing neural topographies will be near one.

On the other hand, something like pattern A from Figure 9 provides much too little in the way of diversity. If we mass all of our coherently-acting units together into a single, solid mass, then we lose the value provided by having different combinations of active units. Thus,  $h$ -values in the range of two or higher are also not at all likely to characterize neural topographies and activation patterns.

Having thus framed our expectations, the next round of work should yield more particular associations between  $h$ -values and network topographies, which should then correlate with our understanding of pairwise connection strengths. This would also be in keeping with results achieved by Barton and Cocco [76], and it also suggests an immediate task-list for investigation:

1. Identify  $h$ -value correlations with specific, characterizable rich-club, scale-free, and small-world topographies (to the extent that such can be expressed in a 2-D CVM lattice); note that this will likely involve configurations where  $x_1 \neq x_2$ , and thus, computational as well as analytic solutions will be needed,
2. Identify the actual free energies associated with those  $h$ -values and their related topographies, and ensure that any empirical distribution is indeed a minimum of a free energy functional, and
3. Characterize the shape of the free energy landscape in the vicinity of the various topographies, and in particular identify susceptibility to perturbations.

These tasks, particularly that of ensuring that the system is at a free energy minimum, will be essential prior to introducing perturbations and other steps that would invoke a variational principle of free energy minimization. This is simply a starter-set of tasks that will help us understand how the 2-D CVM can be useful in modeling neural structural and functional networks.

In addition to modeling neural systems directly, the CVM has potential as the internal or “hidden” layer in a neural network computational engine. One particular advantage of using the 2-D CVM in this role is that it is a natural means of implementing ideas suggested by Friston [82,83,88] requiring a computational engine whose free energy can be perturbed by input stimulus, and which can either return to its original state or find a new equilibrium state in response to perturbations.

This approach has, in fact, already been introduced [123–125]; the goal is now to move from the earlier instances of a 1-D computational engine to a 2-D, and to study performance under various conditions.

Once the basic performance of a 2-D CVM network as part of a computational engine has been understood, it will be possible to examine how the local structures expressed in such a network and characterized in terms of their enthalpy parameters can lead to a greater understanding of the information qualities of such configurations.

Over the long term, two specific lines of work will be to investigate the case of non-equiprobable distributions of units into states **A** and **B**; that is, where  $x_1 \neq x_2$ . This will involve numerical calculations, as the analytic solution holds only for the case where  $x_1 = x_2$ . However, this will yield rich information that can be applied to the neuronal modeling of various populations of active neural configurations. A second endeavor will be to perform kinetics calculations, using the Path Probability Method (PPM) as was devised by Kikuchi as the natural extension of the CVM to the time domain. The PPM will be a useful means of modeling how a system perturbed by a non-equilibrium process can move towards the final equilibrium state.

As a related theme, Feldman and Crutchfield have introduced information-theoretic measures of spatial structure and pattern in two dimensions, using the notion of excess entropy as a measure of complexity [126]. When the basic CVM network is more understood as a neural computational model, we will then be able to investigate complexity and other measures associated with such a 2-D system.

In relation to the above applications, if it can be shown that empirical neural network configurations minimize free energy under a given interaction enthalpy, then free energy becomes



a Lyapunov function or Lagrangian for neuronal systems. This may have profound implications for elaborating variational principles of neuronal dynamics; e.g., gauge fields [106].

## 8. Conclusions

This paper has presented the 2-D cluster variation method from statistical thermodynamics, invoking the notion of entropy describing distributions among local pattern configurations as well as distributions among two possible states as a potentially useful means of characterizing neural systems.

Previous research has shown that nearest-neighbor interactions are significant in modeling neural systems, and the formulation proposed here relies exclusively on a nearest-neighbor interaction enthalpy. For the case where the distribution into states **A** and **B** is equiprobable ( $x_1 = x_2$ ), an analytic solution of the equilibrium values for configuration variables is possible, where the configuration variables can be expressed in terms of the interaction enthalpy parameter  $h$ .

The broader descriptive power of the 2-D CVM approach provides a means for studying neural network behaviors, and for modeling the information-processing capabilities of these networks. This is crucially important for next-generation Brain–Computer *Information* Interfaces (BCIIs), which will likely invoke densely-configured intracortical electrodes for much higher granularity in neural stimulus and signal recording.

The direction of next-generation Brain–Computer Interfaces will lead us into realms where the mapping of the external system to the neural substrate will not necessarily mimic existing body conditions [127]. One example of such is that of EEG-controlled robot swarms [128].

This will have huge repercussions for human self-perception. As Lakoff and Johnson described in their classic work, *Metaphors We Live By* [129], humans map a great deal of their experience into a somato-sensory reference frame. When we begin to integrate the perception of external systems that do not conform to our body-perception of self, we radically impact our notion of “what is self”. The societal implications will be enormous, ranging from experiential to philosophic, and from theological to legal.

It is into this emerging realm that we envision a closer coupling of modeling neural systems with computational platforms that can interpret neural activity. This will facilitate Brain–Computer Information Interfaces, and be part of an ever-evolving understanding of human experience and self-definition. Statistical thermodynamics has been shown to play a role in modeling neural systems. In the approach afforded by the CVM, we have an opportunity to extend the role of statistical thermodynamics in describing the brain, and thus link our understanding of a wide range of neural behaviors with a useful computational model.

**Acknowledgments:** The author wishes to thank T. Mohri and K. P. Körding for helpful conversations, as well as two Reviewers whose comments were exceptionally insightful and useful.

**Conflicts of Interest:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Appendix A. The 1-D CVM Equation

There are certain relations that exist among the configuration variables  $x_i$ ,  $y_i$ , and  $w_i$ :  
For the  $y_j$ :

$$y_1 = z_1 + z_2 \quad (\text{A1a})$$

$$y_2 = z_2 + z_4 = z_3 + z_5 \quad (\text{A1b})$$

$$y_3 = z_5 + z_6 \quad (\text{A1c})$$

For the  $w_i$ :

$$w_1 = z_1 + z_3 \quad (\text{A2a})$$

$$w_2 = z_2 + z_5 \quad (\text{A2b})$$

$$w_3 = z_4 + z_6 \quad (\text{A2c})$$

For the  $x_i$ :

$$x_1 = y_1 + y_2 = w_1 + w_2 = z_1 + z_2 + z_3 + z_5 \quad (\text{A3a})$$

$$x_2 = y_2 + y_3 = w_2 + w_3 = z_2 + z_4 + z_5 + z_6 \quad (\text{A3b})$$

The normalization is:

$$1 = x_1 + x_2 = \sum_{i=1}^6 \gamma_i z_i \quad (\text{A4})$$

Viewing the zigzag chain as being composed of two horizontal rows, the number of ways of constructing this chain are given as:

$$\Omega_{double} = \frac{\prod_{i=1}^3 (2M_{y_i})!^{\beta_i}}{\prod_{i=1}^6 (2M_{z_i})!^{\gamma_i}} \quad (\text{A5})$$

where  $M$  is the number of lattice points in a row, and  $\Omega_{double}$  refers to the juxtaposition of two rows. The degeneracy factors  $\beta_i$  and  $\gamma_i$  were described in Section 4.2, and illustrated in Figure 4.

When  $M$  is large, Stirling's approximation can be used to express the previous equation as

$$\Omega_{double} = \left[ \frac{\prod_{i=1}^3 (M_{y_i})!^{\beta_i}}{\prod_{i=1}^6 (M_{z_i})!^{\gamma_i}} \right]^2 \quad (\text{A6})$$

where Stirling's approximation is given as:  $N! = N \ln(N) - N$ .

We can now substitute from this equation into the one for the entropy and once again use Stirling's approximation to obtain:

$$S_{1-D} = k_\beta \ln \Omega_{double} = 2k_\beta \left[ \sum_{i=1}^3 \beta_i Lf(y_i) - \sum_{i=1}^6 \gamma_i Lf(z_i) \right], \quad (\text{A7})$$

where  $Lf(x) = x \ln(x) - x$ , and  $k_\beta$  is Boltzmann's constant.

This is the entropy associated with a single zigzag chain.

Achieving the analytic solution involves solving a free energy equation for the equilibrium point, where the enthalpy-per-active-unit is given as zero, and the pairwise-interaction enthalpy is given as  $\varepsilon$ . Because the enthalpy-per-active-unit is zero, there is no a priori preference for a unit to be in either the **A** or **B** states; thus,  $x_1 = x_2 = 0.5$ .

This gives a symmetric distribution to the other configuration variables; specifically:

$$\begin{aligned} y_1 &= y_3 \\ z_1 &= z_6 \\ z_2 &= z_5 \\ z_3 &= z_4 \end{aligned} \quad (\text{A8})$$

The resulting solution gives the equilibrium configuration variables  $y_i$  and  $z_i$  in terms of  $h$ , where (with the 1-D CVM)  $h = \exp(\varepsilon/4)$ . As we are working in an abstract sense,  $k_\beta T$  has been set equal to 1.

For the system where  $x_1 = x_2 = 0.5$  and  $\lambda = 0$ , we can solve the previous equations giving  $z_i q$  in terms of the  $y_i$  and other parameters for the configuration variables  $y_i$  and  $z_i$ . The calculations, briefly summarized below, are presented in careful and tractable detail in [9].

Let  $h = e^{\varepsilon/4}$ , and  $s = z_1/z_3$ . Then

$$h^2 = s \frac{1 + s^{1/2}}{s + s^{1/2}} \quad (\text{A9})$$

or

$$s = h^4 \quad (\text{A10})$$

and

$$z_1 = sz_3 \quad (\text{A11})$$

and

$$z_2 = [1 - 2z_1 - 2z_3]/4 \quad (\text{A12})$$

and

$$y_2 = [1 - 2z_1 + 2z_3]/4 \quad (\text{A13})$$

With this set of equations, it is clear that if we can find an expression for  $z_i q$  in terms of  $s$  (which is itself a function of  $h$ ), then we can also immediately determine the remaining configuration variables.

Such an analytic solution is possible. The details are given in [9]. The result is:

$$1/z_3 = 2(s + 1) \pm 4s^{1/2} \quad (\text{A14})$$

## References

1. Kording, K.P. Bayesian statistics: Relevant for the brain? *Curr. Opin. Neurobiol.* **2014**, doi:10.1016/j.conb.2014.01.003.
2. Stevenson, I.H.; Rebesco, J.M.; Hatsopoulos, N.G.; Haga, Z.; Miller, L.E.; Körding, K.P. Bayesian inference of functional connectivity and network structure from spikes. *IEEE Trans. Neural Syst. Rehab. Eng.* **2009**, *17*, 203–213.
3. Beal, M.J. Variational Algorithms for Approximate Bayesian Inference. Ph.D. Thesis, University College London, London, UK, 2003.
4. Turner, B.M.; Zandt, T.V. A tutorial on approximate Bayesian computation. *J. Math. Psychol.* **2012**, *56*, 69–85.
5. Maren, A. Statistical Thermodynamics: Introduction to Phase Space and Metastable States. Available online: [http://www.aliannajmaren.com/Downloads/Stat\\_Thermo\\_Intro-Phase-Spaces\\_2014-02-27\\_rev.pdf](http://www.aliannajmaren.com/Downloads/Stat_Thermo_Intro-Phase-Spaces_2014-02-27_rev.pdf) (accessed on 25 July 2016)
6. Kikuchi, R. A Theory of Cooperative Phenomena. *Phys. Rev.* **1951**, *81*, 988.
7. Kikuchi, R.; Brush, S. Improvement of the Cluster Variation Method. *J. Chem. Phys.* **1967**, *47*, 195.
8. Sporns, O.; Kötter, R. Motifs in brain networks. *PLoS Biol.* **2004**, *2*, e369.
9. Maren, A. The Cluster Variation Method I: 1-D Single Zigzag Chain: Basic Theory, Analytic Solution and Free Energy Variable Distributions at Midpoint ( $x_1 = x_2 = 0.5$ ). Available online: <http://www.aliannajmaren.com/patents-publications/> (accessed on 25 July 2016).
10. Maren, A. The Cluster Variation Method II: 2-D Grid of Zigzag Chains: Basic Theory, Analytic Solution and Free Energy Variable Distributions at Midpoint ( $x_1 = x_2 = 0.5$ ). Available online: <http://www.aliannajmaren.com/patents-publications/> (accessed on 25 July 2016).
11. Tkačik, G.; Marre, O.; Amodei, D.; Schneidman, E.; Bialek, W.; Berry, M., II. Searching for collective behavior in a large network of sensory neurons. *PLoS Comput. Biol.* **2014**, *10*, e1003408.

12. Tkačik, G.; Mora, T.; Marre, O.; Amodei, D.; Berry, M., II; Bialek, W. Thermodynamics for a network of neurons: Signatures of criticality. Available online: <http://arxiv.org/abs/1407.5946> (accessed on 25 July 2016).
13. Friston, K.; Breakspear, M.; Deco, G. Perception and self-organized instability. *Front. Comput. Neurosci.* **2012**, *6*, 1–19.
14. Stevenson, I.H.; Kording, K.P. How advances in neural recording affect data analysis. *Nat. Neurosci.* **2011**, *14*, 139–142.
15. Stevenson, I.H.; Kording, K.P. Advances in Neural Recording Affect Data Analysis. Available online: <http://www.nature.com/neuro/journal/v14/n2/abs/nn.2731.html> (accessed on 25 July 2016).
16. Bensmaia, S.J.; Miller, L.E. Restoring sensorimotor function through intracortical interfaces: Progress and looming challenges. *Nat. Rev. Neurosci.* **2014**, *15*, 313–325.
17. Cogan, S.F. Neural stimulation and recording electrodes. *Annu. Rev. Biomed. Eng.* **2008**, *10*, 275–309.
18. Fabbro, A.; Scaini, D.; León, V.; Vázquez, E.; Cellot, G.; Privitera, G.; Lombardi, L.; Torrisi, F.; Tomarchio, F.; Bonaccorso, F.; et al. Graphene-based interfaces do not alter target nerve cells. *ACS Nano* **2016**, *10*, 615–623.
19. Jang, J.; Lee, J.; Kang, M.; Song, Y.K. *Smart Sensors for Health and Environment Monitoring*; KAIST Research Series; Springer: Dordrecht, The Netherlands; Heidelberg, Germany; New York, NY, USA, 2015; chapter Implantable Neural Sensors for Brain Machine Interface, pp. 51–74.
20. Jorfi, M.; Skousen, J.L.; Weder, C.; Capadona, J.R. Progress towards biocompatible intracortical microelectrodes for neural interfacing applications. *J. Neural Eng.* **2015**, doi:10.1088/1741-2560/12/1/011001.
21. Shih, J.J.; Krusienski, D.J.; Wolpaw, J.R. Brain-Computer Interfaces in Medicine. *Mayo Clin. Proc.* **2012**, *87*, 268–279.
22. Taylor, A.C.; Vagaska, B.; Edgington, R.; Hébert, C.; Ferretti, P.; Bergonzo, P.; Jackman, R.B. Biocompatibility of nanostructured boron doped diamond for the attachment and proliferation of human neural stem cells. *J. Neural Eng.* **2015**, *12*, 066016.
23. Berényi, A.; Somogyvári, Z.; Nagy, A.J.; Roux, L.; Long, J.D.; Fujisawa, S.; Stark, E.; Leonardo, A.; Harris, T.D.; Buzsáki, G. Large-scale, high-density (up to 512 channels) recording of local circuits in behaving animals. *J. Neurophysiol.* **2014**, *111*, 1132–1149.
24. Jun, S.B. *Smart Sensors for Health and Environment Monitoring*; KAIST Research Series; Springer: Dordrecht, The Netherlands; Heidelberg, Germany; New York, NY, USA, 2015; Chapter: Implantable Brain Interface: High-Density Microelectrode Array for Neural Recording, pp. 51–74.
25. Khalifa, A.; Gao, Z.; Bermak, A.; Wan, Y.; Chan, L.L.H. A novel method for the fabrication of a high-density carbon nanotube microelectrode array. *Sens. Bio-Sens. Res.* **2015**, *5*, 1–7.
26. Norton, J.J.S.; Lee, D.; Lee, J.W.; Lee, W.; Kwon, Q.; Won, I.; Jung, S.Y.; Cheng, H.; Jeong, J.W.; Akce, A.; et al. Soft, curved electrode systems capable of integration on the auricle as a persistent brain-computer interface. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 3920–3925.
27. Okabe, K.; Jeewan, H.P.; Yamagiwa, S.; Kawano, T.; Ishida, M.; Akita, I. Co-design method and wafer-level packaging technique of thin-film flexible antenna and silicon CMOS rectifier chips for wireless-powered neural interface systems. *Sensors* **2015**, *15*, 31821–31832.
28. Lee, S.C.; Shin, Y.H.; Woo, S.; Kim, K.; Lee, H.N. *Brain-Computer Interface Systems—Recent Progress and Future Prospects*; InTech: Rijeka, Croatia, 2013.
29. Vitale, F.; Summerson, S.R.; Aazhang, B.; Kemere, C.; Pasquali, M. Neural stimulation and recording with bidirectional, soft carbon nanotube fiber microelectrodes. *ACS Nano* **2015**, *9*, 4465–4474.
30. Tognoli, E.; Kelso, J.A.S. The metastable brain. *Neuron* **2014**, *81*, 35–48.
31. Szu, H.H.; Lalonde, F.; Tran, B.Q. Noninvasive detection of brain order-disorder transitions using functional EEG. *SPIE Newsroom* **2014**, doi:10.1117/2.1201405.005446.
32. Lalonde, F.; Gogtayand, N.; Giedd, J.; Vydellingum, N.; Brown, D.; Tran, B.; Hsu, C.; Hsu, M.K.; Cha, J.; Jenkins, J.; et al. Brain Order Disorder 2nd Group Report of f-EEG. Available online: <http://spie.org/Publications/Proceedings/Paper/10.1117/12.2051706> (accessed on 25 July 2016).
33. Maren, A.J.; Szu, H.H. A New Eeg Measure Using the 1-D Cluster Variation Method. Available online: <http://www.aliannajmaren.com/patents-publications/> (accessed on 25 July 2016).

34. Goñi, J.; van den Heuvel, M.P.; Avena-Koenigsberger, A.; Velez de Mendizabal, N.; Betzel, R.F.; Griffa, A.; Hagmann, P.; Corominas-Murtra, B.; Thiran, J.P.; Sporns, O. Resting-brain functional connectivity predicted by analytic measures of network communication. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 833–838.
35. Gandhi, V.; Prasad, G.; Coyle, D.; Behera, L.; McGinnity, T.M. Quantum neural network-based EEG filtering for a brain–computer interface. *IEEE Trans. Neural Netw. Learn. Syst.* **2014**, *25*, 278–288.
36. Marblestone, A.H.; Wayne, G.; Kording, K.P. Towards an Integration of Deep Learning And Neuroscience. Available online: <http://arxiv.org/abs/1606.03813> (accessed on 25 July 2016).
37. Hopfield, J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 2554–2558.
38. Kirkpatrick, S.; Gelatt, C.D., Jr.; Vecchi, M.P. Optimization by Simulated Annealing. *Science* **1983**, *220*, 671–680.
39. Szu, H.H.; Miao, L.; Qi, H. Unsupervised learning with mini free energy. In Proceeding of the SPIE Conference: Independent Component Analyses, Wavelets, Unsupervised Nano-Biomimetic Sensors, and Neural Networks V, Orlando, FL, USA, 10–13 April 2007.
40. Behera, L.; Kar, I.; Elitzur, A.C. A recurrent quantum neural network model to describe eye tracking of moving targets. *Found. Phys. Lett.* **2005**, *18*, 357–370.
41. Gandhi, V.; Prasad, G.; Coyle, D.; Behera, L.; McGinnity, T.M. Evaluating Quantum Neural Network filtered motor imagery brain- computer interface using multiple classification techniques. *Neurocomputing* **2015**, *170*, 161–167.
42. Best, M.D.; Takahashi, K.; Suminski, A.J.; Ethier, C.; Miller, L.E.; Hatsopoulos, N.G. Comparing offline decoding performance in physiologically defined neuronal classes. *J. Neural Eng.* **2016**, *13*, 026004.
43. Chen, Z. An overview of Bayesian methods for neural spike train analysis. *Comput. Intell. Neurosci.* **2013**, doi:10.1155/2013/251905.
44. Suminski, A.; Tkach, D.; Hatsopoulos, N. Exploiting multiple sensory modalities in brain-machine interfaces. *Neural Netw.* **2009**, *22*, 1224–1234.
45. Pires, G.; Castelo-Branco, M.; Nunes, U. Visual P300-based BCI to steer a wheelchair: A Bayesian approach. *Conf. Proc. IEEE. Eng. Med. Biol. Soc.* **2008**, *2008*, 658–661.
46. Suk, H.I.; Lee, S.W. A novel Bayesian framework for discriminative feature extraction in brain-computer interfaces. *IEEE Trans. PAMI* **2013**, *35*, 286–298.
47. Hatsopoulos, N.G.; Donoghue, J.P. The science of neural interface systems. *Annu. Rev. Neurosci.* **2009**, *32*, 249–266.
48. Shenoy, P.; Krauledat, M.; Blankertz, B.; Rao, R.P.N.; Müller, K.R. Towards adaptive classification for BCI. *J. Neural Eng.* **2006**, *3*, R13–R23.
49. Sharpee, T.; Calhoun, A.; Chalasani, S. Information theory of adaptation in neurons, behavior, and mood. *Curr. Opin. Neurobiol.* **2014**, *25*, 47–53.
50. Lenormand, M.; Jabot, F.; Deffuant, G. Adaptive approximate Bayesian computation for complex models. *Comput. Stat.* **2013**, *28*, 2777–2798.
51. Hebb, D. *The Organization of Behavior: A Neuropsychological Theory*; Psychology Press: Florence, KY, USA, 1949.
52. Edelman, G.M.; Mountcastle, V.B. *The Mindful Brain: Cortical Organization and the Group-Selective Theory of Higher Brain Function*; MIT Press: Cambridge, MA, USA, 1978.
53. Edelman, G.M. Neural Darwinism: Selection and reentrant signaling in higher brain function. *Neuron* **1993**, *10*, 115–125.
54. Singer, W. Synchronization of cortical activity and its putative role in information processing and learning. *Ann. Rev. Physiol.* **1993**, *55*, 349–374.
55. Honey, C.J.; Thivierge, J.P.; Sporns, O. Can structure predict function in the human brain? *NeuroImage* **2010**, *52*, 766–776.
56. Hubel, D.H.; Wiesel, T.N. Ferrier Lecture: Functional architecture of the macaque monkey visual cortex. *Proc. R. Soc. Lond. B Biol. Sci.* **1977**, *198*, 1–59.
57. Wurtz, R.H. Recounting the impact of Hubel and Wiesel. *J. Physiol.* **2009**, *587*, 2817–2823.
58. Aertsen, A.; Gerstein, G.; Habib, M.; Palm, G. Dynamics of neuronal firing correlation: Modulation of “effective connectivity”. *J. Neurophysiol.* **1989**, *61*, 900–917.

59. Fries, P. A mechanism for cognitive dynamics: Neural communications through neuronal coherence. *Trends Cogn. Neurosci.* **2005**, *9*, 474–480.
60. Schroeter, M.S.; Charlesworth, P.; Kitzbichler, M.G.; Paulsen, O.; Bullmore, E.T. Emergence of rich-club topology and coordinated dynamics in development of hippocampal functional networks in vitro. *J. Neurosci.* **2015**, *35*, 5459–5470.
61. Sporns, O. Structure and function of complex brain networks. *Dialogues Clin. Neurosci.* **2013**, *15*, 247–262.
62. Colizza, V.; Flammini, A.; Serrano, M.A.; Vespignani, A. Detecting rich-club ordering in complex networks. *Nat. Phys.* **2006**, *2*, 110–115.
63. Bullmore, E.; Sporns, O. Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* **2009**, *10*, 186–198.
64. Alagapan, S.; Franca, E.; Pan, L.; Leondopulos, S.; Wheeler, B.; DeMarse, T. Structure, function, and propagation of information across living two, four, and eight node degree topologies. *Front. Bioeng. Biotechnol.* **2016**, *29*, 4–15.
65. Engel, A.K.; Gerloff, C.; Hilgetag, C.C.; Nolte, G. Intrinsic coupling modes: Multiscale interactions in ongoing brain activity. *Neuron* **2013**, *80*, 867–886.
66. Landau, L.; Lifshitz, E. *Statistical Physics*, 3rd ed.; Pergamon: Oxford, UK, 1980.
67. Schneidman, E.; Berry, M.; Segev, R.; Bialek, W. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* **2006**, *440*, 1007–1012.
68. Cohen, M.; Kohn, A. Measuring and interpreting neuronal correlations. *Nat. Neurosci.* **2011**, *14*, 811–819.
69. Poli, D.; Pastore, V.P.; Martinoia, S.; Massobrio, P. From functional to structural connectivity using partial correlation in neuronal assemblies. *J. Neural Eng.* **2016**, *13*, 026023.
70. Sporns, O.; Tononi, G.; Kötter, R. The human connectome: A structural description of the human brain. *PLoS Comput. Biol.* **2005**, doi:10.1371/journal.pcbi.0010042.
71. Zalesky, A.; Fornito, A.; Cocchi, L.; Gollo, L.L.; Breakspear, M. Time-resolved resting-state brain networks. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 10341–10346.
72. Papo, D.; Buldú, J.M.; Boccaletti, S.; Bullmore, E.T. Complex network theory and the brain. *Phil. Trans. R. Soc.* **2014**, doi:10.1098/rstb.2013.0520.
73. Gollo, L.L.; Breakspear, M. The frustrated brain: From dynamics on motifs to communities and networks. *Phil. Trans. R. Soc.* **2014**, doi:10.1098/rstb.2013.0532.
74. Gollo, L.L.; Mirasso, C.; Sporns, O.; Breakspear, M. Mechanisms of zero-lag synchronization in cortical motifs. *PLoS Comput. Biol.* **2014**, *10*, e1003548.
75. Advani, M.; Lahiri, S.; Ganguli, S. Statistical mechanics of complex neural systems and high dimensional data. *J. Stat. Mech.* **2013**, doi:10.1088/1742-5468/2013/03/P03014.
76. Barton, J.; Cocco, S. Ising models for neural activity inferred via selective cluster expansion: Structural and coding properties. *J. Stat. Mech.* **2013**, doi:10.1088/1742-5468/2013/03/P03002.
77. Wilson, H.R.; Cowan, J.D. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.* **1972**, *12*, 1–24.
78. Moran, R.; Pinotsis, D.; Friston, K. Neural masses and fields in dynamic causal modeling. *Front. Comp. Neurosci.* **2013**, doi:10.3389/fncom.2013.00057.
79. Yang, Z.; Gandhi, V.; Karamanoglu, M.; Graham, B. Characterising information correlation in a stochastic Izhikevich neuron. In Proceedings of the 2015 International Joint Conference on Neural Networks (IJCNN), Killamey, Ireland, 12–17 July 2015.
80. Tkačik, G.; Marre, O.; Mora, T.; Amodei, D.; Berry, M., II; Bialek, W. The simplest maximum entropy model for collective behavior in a neural network. *J. Stat. Mech.* **2013**, *2013*, P03011.
81. Berry, M., II; Tkačik, G.; Dubuis, J.; Marre, O.; da Silveira, R.A. A simple method for estimating the entropy of neural activity. *J. Stat. Mech.* **2013**, *04*, P03015.
82. Friston, K.; Levin, M.; Sengupta, B.; Pezzulo, G. Knowing one's place: A free-energy approach to pattern regulation. *J. R. Soc. Interface* **2015**, doi:10.1098/rsif.2014.1383.
83. Friston, K. The free-energy principle: A unified brain theory? *Nat. Rev. Neurosci.* **2010**, *11*, 127–138.
84. Maren, A.J.; Lin, S.H.; Langley, R.H.; Eyring, L. A theoretical model for hysteresis in solid state phase transitions. *J. Solid State Chem.* **1984**, *53*, 329–343.
85. Maren, A. Theoretical Models for Solid State Phase Transitions. Ph.D. Thesis, Arizona State University, Tempe, AZ, USA, 1981.



86. Deco, G.; Senden, M.; Jirsa, V. How Anatomy Shapes Dynamics: A Semi-Analytical Study of the Brain at Rest by a Simple Spin Model. *Front. Comput. Neurosci.* **2012**, *6*, 68.
87. Friston, K. Functional and effective neuroimaging. *Brain Connect.* **2011**, *1*, 13–36.
88. Friston, K. Life as we know it. *J. R. Soc. Interface* **2013**, *10*, 0130475.
89. Massobrio, P.; Pasquale, V.; Martinoia, S. Self-organized criticality in cortical assemblies occurs in concurrent scale-free and small-world networks. *Sci. Rep.* **2015**, *5*, 10578.
90. Kozma, R.; Puljic, M.; Freeman, W.J. Thermodynamic model of criticality in the cortex based on EEG/ECOG Data. In *Criticality in Neural Systems*; Plenz, D., Ed.; Wiley-VCH Verlag GmbH: Weinheim, Germany, 2012; Chapter 1, pp. 1–28.
91. Deco, G.; Ponce-Alvarez, A.; Mantini, D.; Romani, G.L.; Hagmann, P.; Corbetta, M. Resting-state functional connectivity emerges from structurally and dynamically shaped slow linear fluctuations. *J. Neurosci.* **2013**, *33*, 11239–11252.
92. Nicolis, J.S. *Chaotic Dynamics Applied to Biological Information Processing (Mathematical Ecology)*; Akademie-Verlag: Berlin, Germany, 1987.
93. Nicolis, J.S. Chaotic dynamics in biological information processing: A heuristic outline. *Il Nuovo Cimento D* **1987**, *9*, 1359–1388.
94. Freeman, W.J. Thermodynamics of Cerebral Cortex Assayed by Measures of Mass Action. In *Chaos, Information Processing and Paradoxical Games: The Legacy of John S. Nicolis*; Nicolis, G., Basios, V., Eds.; World Scientific Publishing Co.: London, UK, 2014; Chapter 13, pp. 275–298.
95. Beggs, J.M.; Timme, N. Being critical of criticality in the brain. *Front. Physiol.* **2012**, *3*, 163.
96. Kastner, D.B.; Baccus, S.A.; Sharpee, T.O. Critical and maximally informative encoding between neural populations in the retina. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 2533–2538.
97. Abum, M.J.; Holmes, C.A.; Roberts, J.A.; Boonstra, T.W.; Breakspear, M. Critical fluctuations in cortical models near instability. *Front. Physiol.* **2012**, *3*, 1–17.
98. Priesemann, V.; Wibral, M.; Valderrama, M.; Pröpper, R.; Le Van Quyen, M.; Geisel, T.; Triesch, J.; Nikolic, D.; Munk, M.H.J. Spike avalanches in vivo suggest a driven, slightly subcritical brain state. *Front. Syst. Neurosci.* **2014**, *8*, 108.
99. Tomen, N.; Rotermund, D.; Ernst, U. Marginally subcritical dynamics explain enhanced stimulus discriminability under attention. *Front. Syst. Neurosci.* **2014**, *8*, 151.
100. Braun, J.; Mattia, M. Attractors and noise: Twin drivers of decisions and multistability. *Neuroimage* **2010**, *52*, 740–751.
101. Butler, T.C.; Benayoun, M.; Wallace, E.; van Drongelen, W.; Goldenfeld, N.; Cowan, J. Evolutionary constraints on visual cortex architecture from the dynamics of hallucinations. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 606–609.
102. Cowan, J.D.; Neuman, J.; van Drongelen, W. Self-organized criticality in a network of interacting neurons. *J. Stat. Mech.* **2013**, doi:10.1088/1742-5468/2013/04/P04030.
103. Nicolis, J.S. *Chaos and Information Processing: A Heuristic Outline*; World Scientific: Singapore, 1991.
104. Crutchfield, J.P.; Young, K. Inferring statistical complexity. *Phys. Rev. Lett.* **1989**, *63*, 105–108.
105. Pelizzola, A. Cluster Variation Method, Padé approximants and critical behaviour. *Phys. Rev. E Stat. Phys. Plasmas. Fluids Relat. Interdiscip. Top.* **1994**, *49*, R2503–R2506.
106. Sengupta, B.; Tozzi, A.; Cooray, G.K.; Douglas, P.K.; Friston, K.J. Towards a neuronal gauge theory. *PLoS Biol.* **2016**, doi:10.1371/journal.pbio.1002400.
107. Sengupta, B.; Stemmler, M.B.; Friston, K.J. Information and efficiency in the nervous system—A synthesis. *PLoS Comput. Biol.* **2013**, doi:10.1371/journal.pcbi.1003157.
108. Miyata, S.; Motizuki, K. Cluster Approach to magnon raman scattering in cubic antiferromagnet. *J. Phys. Soc. Jpn.* **1975**, *38*, 1020–1026.
109. Sanchez, J.M.; Ducastelle, F.; Gratiyas, D. Generalized Cluster Description of Multicomponent Systems. *Physica* **1984**, *128A*, 334–350.
110. Mohri, T. Cluster Variation Method. *JOM* **2013**, *65*, 1510–1522.
111. Pelizzola, A. Cluster Variation Method in statistical physics and probabilistic graphical models. *J. Phys. A: Math. Gen.* **2005**, *38*, R309.
112. Yedidia, J.; Freeman, W.; Weiss, Y. *Understanding Belief Propagation and Its Generalizations*; Technical Report MERL TR-2001-22; Mitsubishi Electric Research Laboratories: Cambridge, MA, USA, 2002.

113. Stanley, H.E. *Introduction to Phase Transitions and Critical Phenomena*; Oxford University Press: New York, NY, USA, 1971.
114. Feynman, R.P. *Statistical Mechanics: A Set of Lectures*; Benjamin: New York, NY, USA, 1972.
115. Goldenfeld, N. *Lectures on Phase Transitions and the Renormalization Group*; Addison-Wesley: New York, NY, USA, 1992.
116. Hill, T.L. *Statistical Mechanics*; McGraw-Hill: New York, NY, USA, 1956.
117. Maren, A. Statistical Thermodynamics: Basic Theory and Equations. Available online: [http://www.aliannajmaren.com/Downloads/Stat\\_Thermo\\_Basic\\_Theory\\_2013-12-01.pdf](http://www.aliannajmaren.com/Downloads/Stat_Thermo_Basic_Theory_2013-12-01.pdf) (accessed on 25 July 2016).
118. Bednar, J.A.; Williams, C.K. Neural Maps: Their Function and Development. In *From Neuron to Cognition via Computational Neuroscience*; Arbib, M., Bonaiuto, J., Eds.; MIT Press: Cambridge, MA, USA, 2016.
119. Harris, K.D.; Shepherd, G.M.G. The neocortical circuit: Themes and variations. *Nat. Neurosci. Rev.* **2015**, *18*, 170–181, doi:10.1038/nn.3917.
120. Stein, B.E.; Meredith, M.A. *The Merging of the Senses*; MIT Press: Cambridge, MA, USA, 1993.
121. Stein, B.E. (Ed.) *The New Handbook of Multisensory Processing*; MIT Press: Cambridge, MA, USA, 2012.
122. Nigam, S.; Shimono, M.; Ito, S.; Yeh, F.C.; Timme, N.; Myroshnychenko, M.; Lapish, C.C.; Tosi, Z.; Hottowy, P.; Smith, W.C.; et al. Rich-club organization in effective connectivity among cortical neurons. *J. Neurosci.* **2016**, *36*, 670–684.
123. Maren, A.J. Free energy as driving function in neural networks. In Proceedings of the Symposium on Nonlinear Theory and Its Applications, Honolulu, HI, USA, 5–10 December 1993; pp. 89–93.
124. Maren, A.; Schwartz, E.; Seyfried, J. Configurational entropy stabilizes pattern formation in a hetero-associative neural network. In Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Chicago, IL, USA, 8–21 October 1992; pp. 89–93.
125. Schwartz, E.; Maren, A. Domains of interacting neurons: A statistical mechanical model. In Proceedings of the World Congress on Neural Networks (WCNN), San Diego, CA, USA, 5–9 June 1994.
126. Feldman, D.P.; Crutchfield, J.P. *Structural Information in Two-Dimensional Patterns: Entropy Convergence and Excess Entropy*; Technical Report; Santa Fe Institute: Santa Fe, NM, USA, 2002.
127. Kilteni, K.; Maselli, A.; Kording, K.P.; Slater, M. Over my fake body: Body ownership illusions for studying the multisensory basis of own-body perception. *Front. Hum. Neurosci.* **2015**, *9*, 20.
128. Karavas, G.K.; Artemiadis, P. On the effect of swarm collective behavior on human perception: Towards brain-swarm interfaces. In Proceedings of the 2015 IEEE International Conference on Multisensor Fusion and Integration (MFI), San Diego, CA, USA, 14–16 September 2015.
129. Lakoff, G.; Johnson, M. *Metaphors We Live By*; University of Chicago Press: Chicago, IL, USA, 1980.



© 2016 by the author; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).