Neural Statics and Dynamics

Robert L. Fry

The Johns Hopkins University Applied Physics Laboratory
11100 Johns Hopkins Road
Laurel, MD 20723-6099
robert.fry@jhuapl.edu

Abstract. A theory of cybernetic systems was previously proposed as a quantitative basis for neural computation. This theory dictates the architectural aspects of a single-neuron system including its operation, adaptation, and most importantly, its computational objective. The present paper completes prior work by formalizing the Hamiltonian for the single-neuron system and by providing an expression for its partition function. New findings suggest the presence of a computational temperature T above which the system must operate to avoid "freezing," upon which useful computation becomes impossible. T serves at least two important functions: (1) it provides a computational degree of freedom to the neuron enabling the realization of probabilistic Bayesian decisioning, and (2) it can be varied by the neuron so as to maximize its throughput capacity in the presence of measurement noise.

Keywords: Boolean algebra, logical questions, Hamiltonian, information theory, partition function

OVERVIEW

Communications systems [1] and general systems, which we simply call *systems*, have been described from a common perspective [2–5] that embraces the quantification of the relativity of information and action through the formulation of *logical questions* and *assertions* [6]. A communications system generates and inserts symbols into a channel, while a receiver attempts to acquire and reconstruct the information at the other side of the channel. Similarly, a system acquires information from its inputs and then uses that information to guide its output decisions. The neuron, having well-defined inputs and outputs, comprises a system and therefore should abide by a theory of systems that quantifies the computational objective and then renders a computational architecture.

The design objective of a communications system is the reliable reconstruction of the source input information at an output receiver. Similarly, the computational objective of a single-neuron system may be the reliable reconstruction of its input information at its output. While a communications system may seek the lossless transmission of information, this represents an impossible goal for a neural system owing to the relative sizes of the input codebook (2^n) and output codebook (2) per channel usage. Regardless, the computational goal of the neuron may remain the maximal preservation of input information at its output, subject to its architectural constraints.

We first summarize previously results [7–14] and then segue into new results including the proposed neural system Hamiltonian, partition function Z. We conclude with the definition and interpretation of the computational temperature $T = 1/b$, which serves to regulate neural decisioning and computation in the presence of measurement noise.

SUMMARY OF PREVIOUS RESULTS

Figure 1 summarizes the previously described model of neural computation, including regulatory mechanisms for maximizing its information transfer capacity when noise is not present. This is achieved through the homeostatic balancing of its information acquisition rate to its decision rate through the coordinated adaptation of $2n + 1$ parameters, where n is the number of neural inputs. Central to Figure 1 is an information or I-diagram [15] described by the ellipse labeled Y lying inside an outside ellipse labeled X. Here X and Y denote *logical questions*. The I-diagram shown in Figure 1 depicts the system output decision rate or entropy Y being maximized by the enlargement of Y. The acquired information X should have maximal relevancy to Y as shown by the matching of Y and X. Finally, the redundant information in X is eliminated as indicated by the shrinking of X.

Y represents the output information of the system and is defined by $Y = \{y, \sim y\}$, where y and $\sim y$ are complementary assertions. Y is ostensibly the system "output" codebook. The system input information X is formally the conjunction of the questions X_i, $i = 1, 2, ..., n$ which describe the individual input information sources and therefore $X = X_1 \wedge X_2 \wedge ... \wedge X_n$. The *conjunction* [6] of the logical questions X_i means that X is defined by *all* possible 2^n input codes that can be rendered to the system and therefore represents its "input" codebook

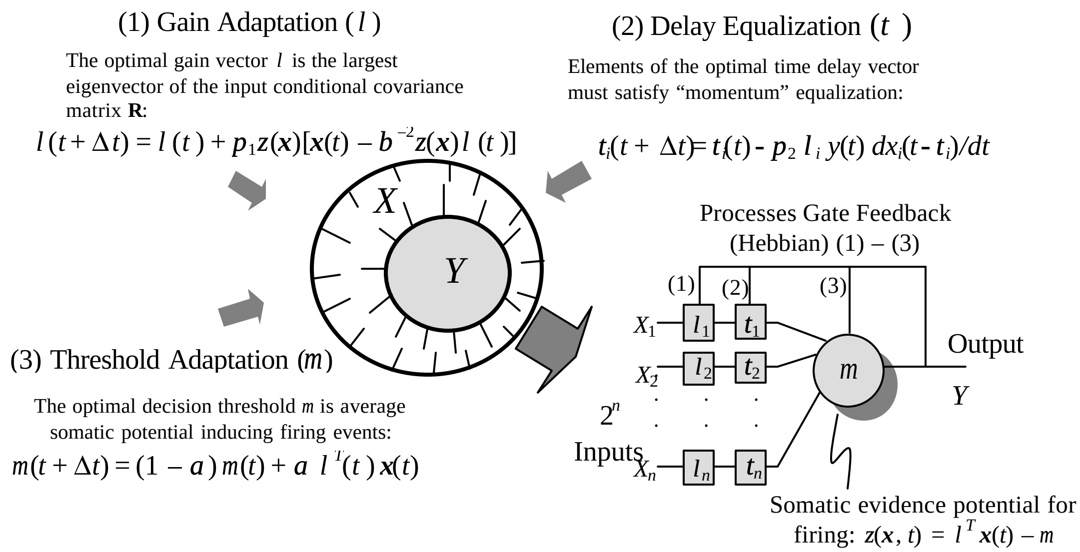

Figure 1: Overview of noiseless model of single-neuron computation.

The ellipse labeled X in Figure 1 is shrinking and becoming congruent to the ellipse labeled Y. Computationally, this is realized through the cooperative adaptation of two n-vectors consisting of a gain vector l and a time-delay vector t. At equilibrium, the gain vector l should be aligned with the largest eigenvector of the conditional input covariance matrix [10] defined by $\mathbf{R} = \left\langle \left[\boldsymbol{x} - \langle \boldsymbol{x} | y = 1 \rangle \right] \left[\boldsymbol{x} - \langle \boldsymbol{x} | y = 1 \rangle \right]^T \middle| \, y = 1 \right\rangle$. The condition $y = 1$ corresponds to the explicit Hebbian gating [16] of this and the other system adaptation rules. One "online" way for a neuron to sequentially compute l is through the modification of an equation originally proposed by Oja [10, 17]. This equation appears in its modified form in Figure 1 and is labeled *(1) Gain Adaptation*. The form of this equation enforces the normalization constraint $|l|^2 = b^2$, where b will be seen to correspond to the inverse computational temperature of the neural system.

The vector t defines a dendritic delay parameters that adapt cooperatively and lie in 1-to-1 those of l. The ability to modify individual dendritic delays is essential if the neuron is to mitigate timing skews between information-bearing inputs and its evidence potential z computed at the soma. The adaptation of the individual delays t_i comprising t are driven by the equation in Figure 1 labeled *(2) Delay Equalization*. Delay equalization seeks to achieve an equilibrium condition whereby there is zero "momentum" transfer between the environment and the neuron, where "momentum" is defined by $l_i y(t) dx_i(t - t_i)/dt$. Thus, the gain l_i acts like mass, the input time derivative $dx_i(t - t_i)/dt$ like velocity, with the output $y(t)$ serving as an explicit Hebbian gate. While the adaptation of the two n-vectors l and t serves to determine *what* and *when* the neuron selectively acquires information from its environment (the question it poses), a single parameter m determines *what* and *when* decisions are generated. m is the system decision threshold against which the measurement innovations $n = l^T \boldsymbol{x}$ are compared. The scalar m is predicted to undergo modest variations [10] that drive it to the equilibrium value $m = E\{n | y = 1\}$. This condition can easily be realized through the long-term temporal averaging of n, contingent on a positive firing event $y = 1$. An especially simple algorithm for achieving this is a convex sum such as that in Figure 1 labeled *(3) Threshold Adaptation* with $a \in (0, 1)$. In summary, the mutually cooperative adaptation of the $2n + 1$ spatiotemporal parameters l, t, and m serves to achieve the computational objective and functionality described graphically by the I-diagram in Figure 1.

Regarding neural decisioning, the terms that define the *evidence* function $z = l^T\boldsymbol{x} - m$ are in 1-to-1 correspondence with Bayes' theorem expressed logarithmically [8] such that Eq. (1) holds:

$$
\begin{aligned}
b \log \frac{p(y=1 \mid \boldsymbol{x})}{p(y=0 \mid \boldsymbol{x})} &= b \log \frac{p(\boldsymbol{x} \mid y=1)}{p(\boldsymbol{x} \mid y=0)} + b \log \frac{p(y=1)}{p(y=0)} \\
&= b\,(?^T \boldsymbol{x} - m) = b\,z \;.
\end{aligned}
\qquad (1)
$$

Therefore, m can be understood to be the logarithm of the prior odds of firing while $n = l^T\boldsymbol{x}$ is the neural log-likelihood measurement statistic which is sufficient. The coefficient b is a scale factor having no preferred value if no noise is present. The optimal adaptation of m drives the system to have maximum output entropy [10] thereby making the last term in Eq. (1) approach zero since $p(y=0) = p(y=1) = ½$.

Simulations have validated the described neural model and determined it to be exact [10] in the sense that the resulting neural capacity is maximized at 1 bit. However, three outstanding issues remain with this model. First, one would expect that the firing rule itself should be probabilistic, abiding by Eq. (1) as opposed to being deterministic as achieved through a simple threshold test. Second, the inverse temperature b is a free parameter with no apparent preferred scale, although it is known from simulations that if b exceeds a critical value of approximately 2, then the neuron ceases to function [10]. Third, the role and effect of measurement noise are not captured in this model. The remainder of this paper describes how these issues can be resolved in a mutually consistent manner that gives deeper insights into the previous results, including the role of measurement noise in neural decisioning and the use of b by the neuron to mitigate the effects of its measurement noise.

NEW RESULTS

The core element of the current model is the single-neuron probability distribution $p(\boldsymbol{x},y)$. Its derivation [7–14] is based on the principle of maximum entropy (ME) [18, 19]. The impetus for this approach is that the neuron invokes inductive reasoning principles that include Bayes' theorem and maximized entropy in defining its architecture, regulating its adaptation, and guiding its operation. Furthermore, through axioms of logical consistency and universality, all systems must use these principles. The assumption made in deriving $p(\boldsymbol{x}, y)$ was that the neuron was capable of computing average statistics on n+1 observables that include the joint input-output moments $\langle x_i y\rangle$ for $i = 1, 2, \ldots, n$ and the output moment $\langle y\rangle$. If the fluctuations in these averages are sufficiently small, then the law of large numbers guarantees that these averages approach the corresponding statistical moments, thereby enabling the application of the ME principle. Therefore, just as all statistical mechanics can be built from information theory [20], the mechanics of neural computation may be based on similar principles. The ME functional maximized over the joint input-output distribution $p(\boldsymbol{x}, y)$ is given by $J = -\sum_{X,Y} p(\boldsymbol{x},y)\log p(\boldsymbol{x},y) + l_0 \sum_{X,Y} p(\boldsymbol{x},y) + \sum_{i=1}^{n} b\, l_i E\{x_i y\} - b\, m E\{y\}$, with the optimal distribution determined by solving $dJ/dp = 0$. This gives $p(\boldsymbol{x}, y) = \exp(-b\,H)/Z$, where $H = -z\,y = -(l^T\boldsymbol{x} - m)y$ is the system Hamiltonian and partition function Z. It can be seen that bl and bm correspond to n+1 Lagrange multipliers in J and where the normalization condition $|l|^2 = b^2$ is enforced. Consequently, b becomes a scale factor multiplying both sides of Eq. (1). The partition function Z is given by

$$
Z = \sum_{\boldsymbol{x} \in B^n} \sum_{y \in B} \exp\left[b \left(?^T \boldsymbol{x} y - m y \right) \right] = \exp(1 - l_0) \;, \qquad (2)
$$

where B^n is the binary n-cube and $B \equiv \{0, 1\}$. Therefore the resulting neural model is analogous to an "Ising" model for a magnetic system with varying coupling strengths l_i and mean field m.

The partition function in Eq. (2) can be evaluated by summing first over the outputs $y \in B$ to obtain $Z = 2^n\,[1 + e^{-bm}\,\Sigma\,\exp(b\ l^T\mathbf{x})]$, with the summation over B^n. This sum can be evaluated in many ways including standard transfer matrix techniques used in statistical physics [21] or even by direct inspection and seen to be $\sum_{x \in X} \exp(b\ ?^T \mathbf{x}) = \prod_{i=1}^{n} (1 + e^{b l_i})$. It can be verified using $p(\mathbf{x}, y)$ in conjunction with the known equilibrium value for the specified threshold $m = E\{l^T\mathbf{x}|y=1\}$, that $m = \Sigma l_i \exp(b l_i)/[1+\exp(b l_i)]$ or $m = \Sigma l_i p(x_i=1|y=1)$. After finding the Taylor expansion of $\exp(b l_i)/[1+\exp(b l_i)]$ about $b l_i = ½$, keeping the first two terms and then simplifying, Z becomes

$$Z = 2^n + 2^n \prod_{i=1}^{n} \cosh\left(\frac{b l_i}{2}\right) e^{-b^4/4}. \qquad (3)$$

Results summarized at the beginning of this paper are contingent on Z being approximately equal to 2^{n+1}. This can only be true if the second term in Eq. (3) is also approximately 2^n. Noting that $\Sigma l_i^2 = b^2$ by constraint and assuming that the gains l_i^2 have a nominal fixed value l^2, then $l_i = l = b/n^{1/2}$ in Eq. (3). Variations in l_i from l have been observed to have little impact on Z as compared to the factor $\exp(-b^4/4)$ in Eq. (3). Now, $Z = 2^n + 2^n \cosh^n(b^2/2n^{1/2}) \exp(-b^4/4)$, which we write as $Z = 2^n + Z_1$. The function $\log_{10} Z_1/2^n$ is plotted in Figure 2 as a two-dimensional grayscale image with log-intensity. The nominal operational region is where $\log_{10} Z_1/2^n \approx 0$. One can see criticality effects in b as it increases or $T = 1/b$ decreases. It can be seen that Z_1 and Z are relatively independent of n; however, there is a modest increase in the allowable lower range of T with increasing n.

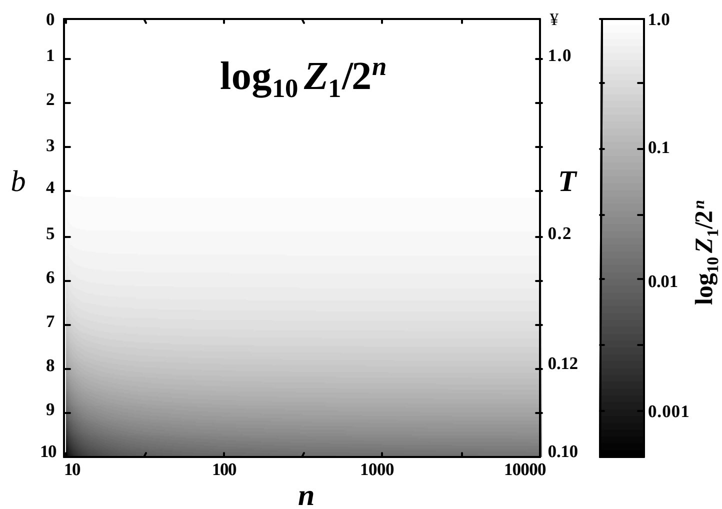

Figure 2: Critical region of neural operation over b and n.

Upon using the Taylor expansion for $\cosh(b^2/2n^{1/2})$ in b, ignoring higher than cubic terms, and then noting that limit $(1+a/n)^n = e^a$ as $n \to \infty$, one obtains $Z \approx 2^n[1+\exp(-b^4/8)]$. Hence as b exceeds $2 \approx 8^{1/4}$, Z is rapidly diminished from 2^{n+1} to 2^n, thus violating the fundamental premise [8, 10] that $Z \approx 2^{n+1}$. As $Z \to 2^n$, the neural system ostensibly freezes and the number of system states is reduced by ½, thereby denying it the ability to perform useful computation. The effects on the described computational model as $Z \to 2^n$ have been observed in numerical simulations, but not previously explained [10].

If the soma serves as a spatiotemporal integrative structure by computing the innovation n and then comparing it against the threshold m, then neural decisions are deterministic. If this is true, then a neuron can achieve an optimal transduction rate of 1 bit per transaction or decision. However, this decision scheme obviates the flow of *a posteriori* probabilistic information per Eq. (1). If one admits the presence of neural inputs that are non-information bearing relative to n, then these inputs can collectively induce a somatic noise potential $h(t)$ that, owing to a large number of nonspecific input sites and the central limit theorem, is posited as having a normal distribution $N(m_n,\ s_n^2)$ in the soma. Noise effects, being additive and independent by assumption relative to $n(t)$, give rise to a total somatic potential $z(t) = bn(t)+h(t)$.

Because of independence, the distribution of z is the convolution of the distributions of $n(t)$ and $h(t)$. This has a surprising result regarding neural decisioning. The probability of firing obeys almost exactly Bayes' theorem in Eq. (1) because the sigmoid function $p(y=1|\boldsymbol{x}) = 1/[1+\exp(-bz)]$ as derivable from Eq. (1), and the noise-induced probability of firing as derivable from the independence of n and h and given by $p(y=1|z+h) = ½\ \mathrm{erfc}[-z/(2^{1/2}s_n)]$, are approximately equal if $b \approx (2p)^{1/2}\ \ln 2/s_n$. Thus noise can provide an enabling mechanism for probabilistic decisioning using Bayes' Theorem. Regarding the mean induced noise potential m_n, the threshold adaptation rule, left unchanged, ensures that the mean level of noise activity m_n will be removed in adjustments to the threshold m and hence the mean level of noise activity can effectively be eliminated through system habituation.

Noise has one deleterious effect, however, if the neuron uses probabilistic decisioning. Although in one regard noise serves to realize Bayesian decisioning, a side effect is the reduction in the throughput capacity C of the neuron owing to the corruptive effects of additive noise. As $s_n^2 \to 0$, the noiseless model holds and C approaches 1 bit per decision event and the noiseless results hold.

Since $z(t)=bn(t)+h(t)$, the inverse temperature b, although irrelevant in the noiseless model, offers a computational degree of freedom wherein it serves to mitigate the effects of noise by providing gain prior its addition. However, criticality limits the upper bound of the gain b.

As stated earlier, the innovation $n(t)$ is a log-likelihood sufficient statistic. Therefore, the neuron can compute using the marginal distributions $p(\boldsymbol{x}|y=0)$ and $p(\boldsymbol{x}|y=1)$, or equivalently, the probabilities $p(n|y=0)$ and $p(n|y=1)$ if known. One can determine $p(n|y=0)$ and $p(n|y=1)$ directly from $p(\boldsymbol{x},y)$ under the assumption that the number of inputs n is large and n consequently Gaussian. One can easily determine the mutual information between the input $\boldsymbol{x}$ rendered as the measurement statistic n at the soma and the now nondeterministic output y. Sample results are shown in Figure 3, where noise is varied over a "signal-to-noise ratio" SNR=b/s_n of 0.01 to 100, b is fixed at $b = 0.1$, and n is varied from 10^1 to 10^4. One can see that system capacity improves systematically with increasing n and that even for modest SNRs, useful information transfer by the neural system is possible.

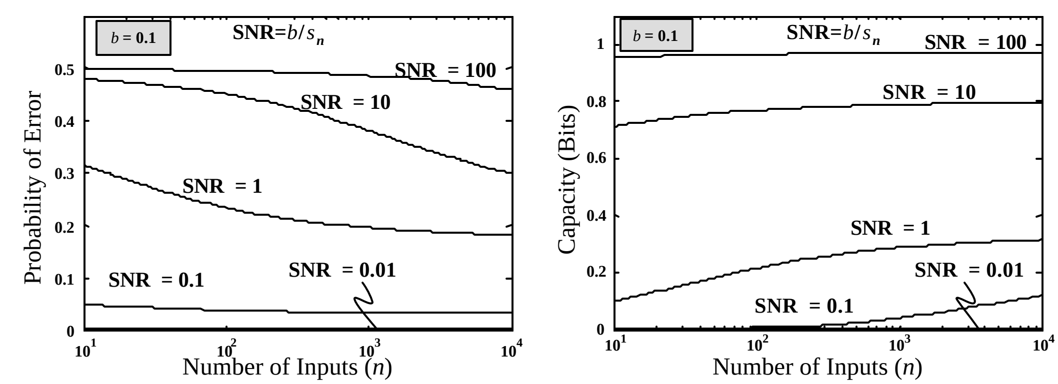

Figure 3: Probability of decision error (left) and neural system capacity (right) vs. SNR and n for $b = 0.1$.

SUMMARY

This paper summarizes previous and new results relating to a theory of neural computation based on logic, probability, and information theory and its extension to systems that operate cybernetically. Previous results describe a neural model defined by its $2n + 1$ parameters that interactively adapt upon the feedback of a positive firing event so as to maximize its information throughput capacity. New results substantiate and explain previous results and further predict the presence of a computational degree of freedom b that we call *computational inverse temperature*. This brings the total number of possible regulatory parameters to $2n + 2$. The inverse temperature b provides the neuron with the ability to maximize its information throughput in the presence of measurement noise and to perform probabilistic decisioning. Space does not permit describing other findings and so these details will be reported elsewhere.

REFERENCES

1. C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical J*, **27**, 379–423 and 623–656, July and Oct., 1948.
2. R. I. Joseph, R. L. Fry, and V. K. Dogra, "Logical and Geometric Inquiry," in *Proc. 22nd Workshop on Maximum Entropy and Bayesian Methods*, American Institute of Physics, New York, 2003.
3. R. L. Fry, "The Engineering of Cybernetic Systems," in *Proc. 21st Workshop on Maximum Entropy and Bayesian Methods*, Baltimore, MD, R. L. Fry (ed.), American Institute of Physics, New York, 2002.
4. R. L. Fry, "Cybernetic Systems Based on Inductive Logic" in *Proc. of the 20th Workshop on Maximum Entropy and Bayesian Methods*, American Institute of Physics, New York, 2001.
5. R. L. Fry, "Cybernetic Aspects of Neural Computation," poster presentation at the *6th International Conference on Cognitive and Neural Systems*, Center for Adaptive Systems and Department of Cognitive and Neural Systems, Boston University, 2001 (available from the author).
6. R. T., Cox, "Of Inference and Inquiry, An Essay in Inductive Logic" in *Proc. Maximum Entropy Formalism*, first maximized entropy workshop held at MIT, D. Levine and M. Tribus (eds.), MIT Press, Boston, pp. 119–168, 1979, .
7. R. L. Fry, "A Theory of Neural Computation," *Neurocomputing*, 52-54, pp. 255-263, available at http://www.compscipreprints.com/comp/Preprint/show/index.htt.
8. R. L. Fry, and Sova, R. M., "A Logical Basis for Neural Network Design," in *Techniques and Applications of Artificial Neural Networks* **3**, Academic Press, 1998.
9. R. L. Fry, "Neural Mechanics," in *Proc. Int. Conf. Neur. Info. Proc. (ICONIP)*, Hong Kong, 1996.
10. R. L. Fry, "Observer-Participant Models of Neural Processing," *IEEE Trans. Neural Networks*, **6**, 918–928, 1995.
11. R. L. Fry, "Rational Neural Models Based on Information Theory," *Proc. 14th Workshop Maximum Entropy and Bayesian Methods*, Santa Fe, NM, 1995.
12. R. L. Fry, "Neural processing of information," in *Proc. IEEE Int. Symp. on Info. Theory*, Norway, 1994.
13. R. L. Fry, "Rational Neural Models Based On Information Theory," in *Proc. Neural Information Processing Systems - Natural and Synthetic*, Denver, CO, 1995.
14. R. L. Fry, "Maximized Mutual Information Using Macrocanonical Probability Distributions," in *Proc. IEEE/IMS Workshop Information Theory and Statistics*, Arlington, VA, 1994.
15. R. W. Yeung, "A New Outlook on Shannon's Information Measures," in *IEEE Trans. Info. Th*eory, 466–475, 1991.
16. D. O. Hebb, *Organization of Behavior*, Wiley, New York, 1949.
17. E. Oja, "A Simplified Neuron Model as a Principal Component Analyzer," *J. Math. Biol.*, **15**, 267–273, 1982.
18. E. T. Jaynes, "Information Theory and Statistical Mechanics," *Phys. Rev.*, Part I, **106**, 620-630, 1957 and Part II, **108**, 171–190, 1957.
19. J. Skilling, "The Axioms of Maximum Entropy," in *Maximum Entropy and Bayesian Methods in Science and Engineering*, G. J. Erickson and C. R. Smith (eds.), Kluwer, Dordrecht, 1988.
20. M. Tribus, *Thermostatics and Thermodynamics*, van Nostrand, Princeton, NJ, 1961.
21. H. E. Stanley, *Introduction to Phase Transitions and Critical Phenomena*, Oxford Univ. Press, UK, 1971.