Statistical Mechanics and Information Theory

Jeremy Gunawardena
BRIMS, Hewlett-Packard Labs
Filton Road, Stoke Gifford
Bristol BS12 6QZ, UK.
jhcg@hplb.hpl.hp.com

1 July 1995

Abstract

A workshop on *"Statistical Mechanics and Information Theory"* was held at Hewlett-Packard's Basic Research Institute in the Mathematical Sciences (BRIMS) in Bristol, England from 5-9 June 1995. This document contains a report on the workshop, the abstracts of the talks and the accompanying bibliography.

1 Report on the workshop

Statistical mechanics and information theory have been linked ever since the latter subject emerged out of the engineering demands of telecommunications. Claude Shannon's original paper on "A mathematical "theory of communication", [71], makes clear the analogy between information entropy and the "H" of Boltzmann's H-theorem. Earlier work of Szilard in attempting to exorcise Maxwell's demon gave a glimpse of the significance of information to physics, [7]. Since that time, the two fields have developed separately for the most part, although both have interacted fruitfully with the field of statistical inference and both have influenced the development of other fields such as neural networks and learning theory. In recent years, this apartheid has been breached by a number of tantalising observations: for instance, of connections between spin-glass models and error-correcting codes, [73].

To explore these developments, a workshop on "Statistical Mechancis and Information Theory" was held at Hewlett-Packard's Basic Research Institute in the Mathematical Sciences (BRIMS) in Bristol, England. The workshop was organised by Jeremy Gunawarden of BRIMS. The purpose was to bring together physicists, information theorists, statisticians and scientists in several application areas to try and demarcate the common ground and to identify the common problems of the two fields. The style of the workshop was deliberately kept informal, with time being set aside for discussions and facilities being provided to enable participants to work together.

Related conferences which have been held recently include the series on "Maximum Entropy and Bayesian Methods"—see, for instance, [41]—the NATO Advanced Study Institute "From Statistical Physics to Statistical Inference and Back", [34], and the IEEE Workshop on "Information Theory and Statistics".

The workshop revealed that the phrase "information" conveyed very different messages to different people. To some, it meant the rigorous study of communication sources and channels represented by the kind of articles that appear in the IEEE Transactions on Information Theory. To others, particularly the physicists influenced by Jaynes' work, [67], it meant a methodological approach to statistical mechanics based on the maximum entropy principle. To others, it was a less rigorous but, nevertheless, stimulating circle of ideas with broad applications to biology and social science. To others, particularly the mathematicians or mathematical physicists, it represented a source of powerful mathematical theorems linking rigorous results in statistical mechanics with probability theory and large deviations. To others still, it was the starting point for a "theory of information", with broad applications outside communication science. The abstracts of the talks, which are collected together in the next section, and the workshop bibliography, which appears after that, give some idea of the range of material that was discussed at the workshop.

Whether the workshop succeeded in "demarcating the common ground" is a moot point. The mathematical insights, particularly the lectures of John Lewis, Anders Martin-Löf and Sergio Verdu, certainly confirmed the existence of a common ground between information theory, probability theory and statistical mechanics, which in many respects is still not properly explored. But to single out this aspect reflects the peculiar prejudices of the organiser. Perhaps the best that can be said is that the workshop revealed to many of the participants the amazing fruitfulness of the information concept and brought home vividly the importance of developing an encompassing "theory of information", upon which biologists, physicists, communication scientists, statisticians and mathematicians can all draw. This still lies in the future.

The bibliography contains a number of references which are not directly cited in the abstracts but are relevant to the overall subject.

The workshop was entirely supported by BRIMS as part of its wider programme of scientific activity. Further information about BRIMS, including a copy of this report, is available on the world wide web at URL:http://www-uk.hpl.hp.com/brims/.

2 Abstracts of talks

A tale of two entropies

Roy Adler, IBM Yorktown Heights
adler@watson.ibm.com

For the purposes of this talk we divide the subject of abstract dynamical systems into two areas: ergodic theory and topological dynamics. In ergodic theory, a dynamical system is endowed with a probability measure so that one can study distribution of orbits in phase space; in topological dynamics, a metric so that one can study, stability, almost periodicity, density of orbits, etc. Ideas originating in Shannon's work in information theory have played a crucial role in isomorphism theories of dynamical systems. Shannon used notions of entropy and channel capacity to determine the amount of information which can be transmitted through channels. Kolmogorov and Sinai made his idea into a powerful invariant for ergodic theory; and then Ornstein proved that entropy completely classifies special systems called Bernouli shifts. Adler, Konheim, and McAndrew introduce into topological dyanamics an analogous invariant called topological entropy which turns out to be a generalization of channel capacity. Then Adler and Marcus developed an isomorphism theory in which this invariant completely classifies symbolic systems called shifts of finite type. The surprise is that this abstract theory has engineering applications: namely their methods of proof lead to construction of encoders and decoders for data transmission and storage. As a result the history of Shannon's ideas have come full circle. He used them to study transmission of information. Mathematicians took up his ideas to classify dynamical systems, first in ergodic theory and then in topological dynamics. Results in thesecond area then lead to methods for optimizing transmission and storage of data. For references, see [1, 2, 61].

(Editor's note: due to personal reasons which arose at the last minute, Roy Adler was unfortunately unable to attend the workshop.)

Non-commutative dynamical entropy

Robert Alicki, University of Gdansk
fizra@univ.gda.pl

The concept of dynamical entropy, called also Kolmogorov-Sinai invariant, proved to be very useful in the classical ergodic theory in particular for studying systems displaying chaos . Since twenty years several attempts to generalize this notion to non-commutative algebraic dynamical systems have been undertaken. In the present lecture a brief survey of the recent investigations based on the new definition of non-commutative dynamical entropy introduced by M.Fannes and the author is presented. Several examples of dynamical systems including classical systems, shift on quantum spin chain, quasi-free Fermion automorphisms, Powers-Price binary shifts, Stormer free shift and quantum Arnold cat map are discussed. The new dynamical entropy is compared with the alternative formalisms of Connes-Narnhofer-Thirring and Voiculescu. For references, see [3, 4, 22].

Non-equilibrium biology

John Collier, University of Newcastle, Australia
pljdc@linga.newcastle.edu.au

Abstract not available, but see [20, 21, 27, 28, 78].

Chaos and detection

Andy Fraser, Portland State University
andy@sysc.pdx.edu

The low likelihood of linear models fit to chaotic signals and the ubiquity of strange attractors in nature suggests that nonlinear modeling techniques can improve performance for some detection problems. We review likelihood ratio detectors and limitations on the performance of linear models implied by the broad Fourier power spectra of chaotic signals. We observe that the KS entropy of a chaotic system establishes an upper limit on the expected log likelihood that any model can attain. We apply variants of the hidden Markov models used in speech research to a synthetic detection problem, and we obtain performance that surpasses the theoretical limits for linear models. KS entropy estimates suggest that still better performance is possible. For references, see [31, 32, 33].

Practical Entropy Computation in Complex Systems

Neil Gershenfeld, MIT Media Lab
neilg@media.mit.edu

I will discuss the interplay between entropy production in complex systems and the estimation of entropy from observed signals. Entropy measurement in lag spaces provides a very general way to determine a system's degrees of freedom, geometrical complexity, predictability, and stochasticity, but it is notoriously difficult to do reliably. I will consider the role of regularized density estimation and sorting on adaptive trees in determining entropy from measurements, and then look at applications in nonlinear instrumentation and the optimization of information processing systems. For references, see [33, 79].

Origin and growth of order in the expanding universe

David Layzer, Harvard
layzer@cfa.harvard.edu

I define order as potential statistical entropy: the amount by which the entropy of a statistical description falls short of its maximum value subject to appropriate constraints. I argue that, as a consequence of a strong version of the postulate of spatial uniformity and isotropy, the universe contains an irreducible quantity of specific statistical entropy; and I postulate that in the initial state the specific statistical entropy was equal to its maximum value: the universe expanded from an initial state of zero order. Chemical order, which shows itself most conspic-

uously as a preponderance of hydrogen in the present-day universe, arose from a competition between the cosmic expansion and nuclear and subnuclear reactions. The origin of structural order is more controversial. I will argue that it might have been produced by a bottom-up process of gravitational clustering. For references, see [44, 45, 46]. See also, [69].

On sequential compression with distortion

Abraham Lempel, HP Israel Science Centre
lempel@hpl.hp.com

Abstract not available.

Large Deviations, Statistical Mechanics and Information Theory

John Lewis, Dublin Institute for Advanced Studies
lewis@stp.dias.ie

In 1965, Ruelle showed how to give a precise meaning to Boltzmann's Formula. Exploring the consequences of Ruelle's definition leads to the theory of large deviations and information theory. For a reference, see [48].

(Editor's note: John Lewis also gave an additional tutorial lecture on large deviations. Among the texts he cited were [15, 29].)

Entropy in States of the Self-Excited System with External Forcing

Grzegorz Litak, University of Lublin
glitak@archimedes.pol.lublin.pl

Self-excited systems with external driving force are characterized by a number of interesting features as synchronization of vibration, and transition to chaos. Using an one dimensional example of the self-excited system–Froude pendulum we examined the influence of parameters on the system itegrity. The motion is described by the following nonlinear equation with Raylaigh damping term

$$\ddot{\phi} - (\alpha - \beta\dot{\phi}^2)\dot{\phi} + \gamma\sin(\phi) = B\cos\omega t \quad .$$

There are two characteristic frequencies in the system: p – self–excited and ω – driving frequency. For the nodal driving force amplitude B the system oscillates with self–excited frequency. For $B \neq 0$, depending on the system parameters, two cases of regular solutions may occurred: mono–frequency solution and the quasi–periodic solution with the modulated amplitude. Except for regular solution the chaotic one appear. The chaotic vibrations correspond with the positive value of Kolmogorov Entropy K. Here the entropy is the measure of chaoticity of the system, indicating how fast the information about the dymamical system state is losing. The definition of Kolmogorov entropy is in analogy with Shannon's in information theory. In our paper the Kolmogorov entropy and other indicators of the dynamical system states

are examined. The regions of system parameters and initial conditions leading to chaotic, synchronized and quasi–periodic are obtained. The condition for the pendulum to escape from the potential well $V(\phi) = \gamma(1 - \cos(\phi))$ was also found. This is joint work with Wojciech Przystupa, Kazimierz Szabelski and Jerzy Warminski.

Sharpening Occam's razor

Seth Lloyd, MIT
slloyd@mit.edu

Abstract not available.

Free Energy Minimization and Binary Decoding Tasks

David MacKay, University of Cambridge
mackay@mrao.cam.ac.uk

I study the task of inferring a binary vector $\vec{s}$ given a noisy measurement of the vector $\vec{t} = A\vec{s}\ mod\ 2$, where A is an $M \times N$ binary matrix. This combinatorial problem can be attacked by replacing the unknown binary vector by a real vector of probabilities which are optimized by variational free energy minimization. The resulting algorithm shows great promise as a decoder for a novel class of error correcting codes. For references, see [51, 36].

The Shannon-McMillan theorem from the point of view of statistical mechanics

Anders Martin-Löf, University of Stockholm
andersml@insanus.matematik.su.se

Abstract not available, but see [53]. Anders provided a set of handwritten notes for this lecture. Copies are available from Jeremy Gunawardena upon request.

Some Properties of the Generalized BFOS Tree Pruning Algorithm

Nader Moayeri, HP Labs Palo Alto
moayeri@alvand.hpl.hp.com

We first show that the generalized BFOS algorithm is not optimal when it is used to design fixed-rate pruned tree-structured vector quantizers (TSVQ). A simple modification is made in the algorithm that makes it optimal. However, this modification has little effect on the (experimental) rate-distortion performance. An asymptotic analysis is presented that justifies the experimental results. It also suggests that in designing fixed-rate, variable-depth TSVQ's, one can get as good a rate-distortion performance with a greedy TSVQ design algorithm as with an optimal pruning of a large tree.

Continuum entropies for fluids

David Montgomery, Dartmouth College
David.C.Montgomery@dartvax.dartmouth.edu

This paper reviews efforts to define useful entropies for fluids and magnetofluids at the level of their continuous macroscopic fields, such as the fluid velocity or magnetic field, rather than at the molecular or kinetic theory level. Several years ago, a mean-field approximation was used to calculate a "most probable state" for an assembly of a large number of parallel interacting ideal line vortices. The result was a nonlinear partial differential equation for the "most probable" stream function, the so-called sinh-Poisson equation, which was then explored in a variety of mathematical contexts. Somewhat unexpectedly, this sinh-Poisson relation turned out much more recently to have quanitative predictive power for the long-time evolution of continuous, two-dimensional, Navier-Stokes turbulence at high Reynolds numbers [54]. It has been our recent effort to define an entropy for non-ideal continuous fluids and magnetofluids that makes no reference to microscopic discrete structures or particles of any kind [55], and then to test its utility in numerical solutions of fluid and magnetofluid equations. This talk will review such efforts and suggest additional possible applications. The work is thought to be a direct but tardy extension of the ideas of Boltzmann and Gibbs from point-particle statistical mechanics. See also W.H. Matthaeus et al, Physica D51, 531 (1991) and references therein.

Statistical mechanics and information theoretic tools for the study of formal neural networks

Jean-Pierre Nadal, ENS Paris
nadal@physique.ens.fr

Abstract not available, but see [14, 34, 57, 58, 59].

Shannon Information and Algorithmic and Stochastic Complexities

Jorma Rissanen, IBM Almaden
rissanen@almaden.ibm.com

This is an introduction to the formal measures of information or, synonymously, description complexity, introduced during the past 70 or so last years, beginning with Hartley. Although all of them are intimately related to ideas of coding theory, I introduce the fundamental Shannon information without it. I also discuss the central role stochastic complexity plays in modeling problems and the problem of inductive inference in general. For references, see [65, 66].

Information and the Fly's Eye

Dan Ruderman, University of Cambridge
dlr1002@cus.cam.ac.uk

The design of compound eyes can be seen as a set of tradeoffs. The most basic of these was discussed by Feynman in his Lectures: smaller facets allow for finer sampling of the world but

reduces acuity due to diffraction. Good "ballpark" estimates of facet size can be made by equating the diffraction limit to resolution with the Nyquist frequency of the sampling lattice. But other factors are involved such as photon shot noise, which fundamentally limits all visual processing. In the 1970's, Snyder *et al*, [72], applied information theory to the compound eye, finding that maximizing information capacity gave reasonable estimates of design parameters. In this talk I extend the information-theoretic treatment by introducing the time domain. Basic questions such as "how fast should a photoreceptor respond as a function of light level and flight speed?" will be addressed, as will the use of moving natural images as a stimulus ensemble. Preliminary results suggest that information theory can provide a useful tool for understanding the tradeoffs and trends in compound eye design, both in the spatial and temporal domains. For references, see [64, 68]. This is joint work with S. B. Laughlin.

Entropy estimates: measuring the information output of a source

Raymond Russell, Dublin Institute for Advanced Studies
russell@stp.dias.ie

Abstract not available.

On Tree Sources, Finite State Machines, and Time Reversal

Gadiel Seroussi, HP Labs Palo Alto
seroussi@hpl.hp.com

We investigate the effect of time reversal on tree models of finite-memory processes. This is motivated in part by the following simple question that arises in some data compression applications: when trying to compress a data string using a universal source modeler, can it make a difference whether we read the string from left to right or from right to left? We characterize the class of finite-memory *two-sided tree processes*, whose time-reversed versions also admit tree models. Given a tree model, we present a construction of the tree model corresponding to the reversed process, and we show that the number of states in the reversed tree might be, in the extreme case, quadratic in the number of states of the original tree. This answers the above motivating question in the affirmative. This is joint work with Marcelo Weinberger. For a reference, see [70].

Statistical physics and replica theory—neural networks; equilibrium and non-equilibrium

David Sherrington, Oxford University
sherr@thphys.ox.ac.uk

Abstract not available, but see [16, 84].

Maximum Quantum Entropy

Richard Silver, Los Alamos National Laboratory rns@loke.lanl.gov

We discuss a generalization of the maximum entropy (ME) principle in which relative Shannon classical entropy may be replaced by relative von Neumann quantum entropy. This yields a broader class of information divergences, or penalty functions, for statistics applications. Constraints on relative quantum entropy enforce prior correlations, such as smoothness, in addition to the convexity, positivity, and extensivity of traditional ME methods. Maximum quantum entropy yields statistical models with a finite number of degrees of freedom, while maximum classical entropy models have an infinite number of degrees of freedom. ME methods may be extended beyond their usual domain of ill-posed inverse problems to new applications such as non-parametric density estimation. The relation of maximum quantum entropy to kernel estimators and the Shannon sampling theorem are discussed. Efficient algorithms for quantum entropy calculations are described. For references, see [34, 35, 38, 60, 65, 87].

Spin glasses and error-correcting codes

Nicolas Sourlas, ENS Paris
sourlas@physique.ens.fr

Abstract not available, but see [73, 74].

Individual open quantum systems

Tim Spiller, HP Labs Bristol
ts@hplb.hpl.hp.com

The reduced density operator is the conventional tool used for describing open quantum systems, that is systems coupled to an environment such as a heat bath. However, this statistical approach is not the only one on the market. It is possible to describe instead the evolution of an individual system, one member of the ensemble. I motivate such an approach and illustrate it with some simple examples. One nice feature is the emergence of characteristic classical behaviour, such as chaos.

Large deviations and conditional limit theorems

Wayne Sullivan, University College Dublin
wsulivan@irlearn.ucd.ie

Abstract not available, but see [49].

Topological Time Series Analysis of a String Experiment and Its Synchronized Model

Nick Tufillaro, Los Alamos National Laboratory
nbt@goshawk.lanl.gov

We show how to construct empirical models directly from experimental low-dimensional time series. Specfically, we examine data from a string experiment and use MDL to contruct an "optimial" empirical model and show how this empirical model can be syncronized to the chaotic experimental data. Such empirical models can then used for process control, monitoring, and non-destructive testing. For a reference, see [75].

Asymptotic equipartition property in source coding

Sergio Verdu, Princeton
verdu@ivy.princeton.edu

Abstract not available.

Sequential Prediction and Ranking in Universal Context Modeling and Data Compression

Marcelo Weinberger, HP Labs Palo Alto
marcelo@hpl.hp.com

We investigate the use of prediction as a means of reducing the model cost in lossless data compression. We provide a formal justification to the combination of this widely accepted tool with a universal code based on context modeling, by showing that a combined scheme may result in faster convergence rate to the source entropy. In deriving the main result, we develop the concept of sequential ranking, which can be seen as a generalization of sequential prediction, and we study its combinatorial and probabilistic properties. For a reference, see [81].

References

[1] R. L. Adler. The torus and the disk. *Contemporary Mathematics*, 135, 1992.

[2] R. L. Adler and B. Marcus. Topological entropy and equivalence of dynamical systems. *Memoirs of the AMS*, 219, 1979.

[3] R. Alicki and M. Fannes. Defining quantum dynamical entropy. *Letters in Mathematical Physics*, 32:75–82, 1994.

[4] R. Alicki and H. Narnhofer. Comparison of dynamical entropies for the noncommutative shifts. *Letters in Mathematical Physics*, 33:241–247, 1995.

[5] J. J. Ashley, B. H. Marcus, and R. M. Roth. Construction of encoders with small decoding look-ahead for input-constrained channels. *IEEE Transactions on Information Theory*, 41:55–76, 1995.

[6] J. J. Atick. Could information theory provide an ecological theory of sensory processsing. *Network*, 3:213–251, 1992.

[7] C. H. Bennett. Demons, engines and the second law. *Scientific American*, pages 88–96, November 1987.

[8] C. H. Bennett. Notes on the history of reversible computation. *IBM Journal of Research and Development*, 32:16–22, 1988.

[9] C. H. Bennett and R. Landauer. The fundamental physical limits of computation. *Scientific American*, pages 38–46, July 1985.

[10] J. Besag and P. J. Green. Spatial statistics and Bayesian computation. *Journal of the Royal Statistical Society*, 55:25–37, 1993.

[11] W. Bialek, F. Rieke, R. de Ruyter van Steveninck, and D. Warland. Reading a neural code. *Science*, 252:1854–1857, 1991.

[12] D. R. Brooks, J. Collier, B. A. Maurer, J. D. H. Smith, and E. O. Wiley. Entropy and information in evolving biological systems. *Biology and Philosophy*, 4:407–432, 1989.

[13] D. R. Brooks, P. H. Leblond, and D. D. Cumming. Information and entropy in a simple evolution model. *Journal of Theoretical Biology*, 109:77–93, 1984.

[14] N. Brunel, J.-P. Nadal, and G. Toulouse. Information capacity of a perceptron. *Journal of Physics A*, 25:5017–5037, 1992.

[15] J. A. Bucklew, editor. *Large Deviation Techniques in Decision, Simulation and Estimation.* Series in Probability and Mathematical Statistics. John Wiley and Sons, 1990.

[16] C. Campbell and A. Robinson. On the storage capacity of neural networks with sign-constrained weights. *Journal of Physics A*, 24:93–95, 1991.

[17] G. J. Chaitin. On the length of the programs for computing finite binary sequences. *Journal of the Association for the Computing Machinery*, 13:547–569, 1966.

[18] G. J. Chaitin. On the length of the programs for computing finite binary sequences: Statistical considerations. *Journal of the Association for the Computing Machinery*, 16:145–159, 1969.

[19] P. Clifford. Discussion on the Meeting on the Gibbs sampler and other Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society*, 55:53–102, 1993.

[20] J. D. Collier. Entropy in evolution. *Biology and Philosophy*, 1:5–26, 1986.

[21] J. D. Collier. Two faces of Maxwell's Demon reveal the nature of irreversibility. *Studies in the History and Philosophy of Science*, 21:57–268, 1990.

[22] A. Connes, H. Narnhofer, and W. Thirring. Dynamical entropy of C^*-algebras and von Neumann algebras. *Communications in Mathematical Physics*, 112:691–719, 1987.

[23] S. Coombes and J. G. Taylor. Using features for the storage of patterns in a fully connected net. Preprint.

[24] R. T. Cox. Probability, frequency, and reasonable expectation. *American Journal of Physics*, 14:1–13, 1946.

[25] P. Cvitanović, I. Percival, and A. Wirzba, editors. *Quantum Chaos-Quantum Measurement*, volume 358 of *NATO ASI Series*. Kluwer Academic, 1992.

[26] A. P. Dawid. Statistical theory: the prequential approach. *Journal of the Royal Statistical Society*, 147:278–292, 1984.

[27] L. Demetrius. Statistical mechanics and population biology. *Journal of Statistical Physics*, 30:709–751, 1983.

[28] L. Demetrius. The thermodynamics of evolution. *Physica A*, 189:416–436, 1992.

[29] R. S. Ellis, editor. *Entropy, Large Deviations and Statistical Mechanics*, volume 271 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, 1985.

[30] R. A. Fisher. Probability, likelihood and quantity of information in the logic of uncertain inference. *Proceedings of the Royal Society*, CXLVI:1–9, 1933.

[31] A. M. Fraser. Information and entropy in strange attractors. *IEEE Transactions on Information Theory*, 35:245–262, 1989.

[32] A. M. Fraser. Reconstructing attractors from scalar time series. *Physica D*, 34:391–404, 1989.

[33] N. Gershenfeld. Information in dynamics. In *Workshop on Physics and Computation: PhysComp92*, pages 276–280. IEEE Computer Society, 1992.

[34] P. Grassberger and J.-P. Nadal. *From Statistical Physics to Statistical Inference and Back*, volume 428 of *NATO ASI Series*. Kluwer Academic, 1994.

[35] J. J. Halliwell, J. Pérez-Mercader, and W. H. Zurek. *Physical Origins of Time Asymmetry*. Cambridge University Press, 1994.

[36] G. E. Hinton and R. S. Zemel. Autoencoders, minimum description length and Helmholtz free energy. In J. D. Cowan, G. Tesauro, and J. Alspector, editors, *Advances in Neural Information Processing Systems 6*. Morgan Kaufmann, 1994.

[37] E. T. Jaynes. *Probability Theory in Science and Engineering*, volume 4 of *Colloquium Lectures in Pure and Applied Science*. Socony Mobil Oil Company, 1959.

[38] E. T. Jaynes. Information theory and statistical mechanics. Proceedings of Brandeis Summer Institute, 1963.

[39] E. T. Jaynes. Prior probabilities. *IEEE Transactions on Systems Science and Cybernetics*, SSC-4:227–241, 1968.

[40] G. Joyce and D. Montgomery. Negative temperature states for the two-dimensional guilding-centre plasma. *Journal of Plasma Physics*, 10:107–121, 1973.

[41] W. T. Grandy Jr and L. H. Schick, editors. *Maximum entropy and Bayesian methods*, volume 43 of *Fundamental Theories of Physics*. Kluwer Academic, 1991.

[42] E. H. Kerner. *Gibbs Ensemble: Biological Ensemble*, volume XII of *International Science Review*. Gordon and Breach, 1972.

[43] A. N. Kolmogorov. Three approaches to the quantitative definition of information. *Problemy Peredachi Informatsii*, 1:3–11, 1965.

[44] D. Layzer. Growth of order in the Universe. Appears in [78].

[45] D. Layzer. The arrow of time. *Scientific American*, 233:56–69, 1975.

[46] D. Layzer. *Cosmogenesis*. Oxford University Press, 1991.

[47] E. Levin, N. Tishby, and S. A. Solla. A statistical approach to learning and generalisation in layered neural networks. *Proceedings of the IEEE*, 78:1568–1574, 1990.

[48] J. T. Lewis and C. E. Pfister. Thermodynamic probability theory: some aspects of large deviations. Technical Report DIAS-STP-93-33, Dublin Institute for Advanced Studies, 1993.

[49] J. T. Lewis, C. E. Pfister, and W. G. Sullivan. Entropy, concentration of probability and conditional limit theorems. Technical report, Dublin Institute for Advavced Studies, 1995.

[50] M. Li and P. Vitányi. *An Introduction to Kolmogorov Complexity and its Applications*. Texts and Monographs in Computer Science. Springer-Verlag, 1989.

[51] D. MacKay. A free energy minimisation framework for inference problems in modulo 2 arithmetic. Preprint available from http://131.111.48.24/mackay/fe.ps.Z.

[52] B. H. Marcus and R. M. Roth. Improved Gilbert-Varshamov bound for constrained systems. *IEEE Transactions on Information Theory*, 38:1213–1221, 1992.

[53] A. Martin-Löf. *Statistical Mechanics and the Foundations of Thermodynamics*. Number 101 in Lecture Notes in Physics. Springer-Verlag, 1979.

[54] D. Montgomery, W. H. Matthaeus, W. T. Stribling, D. Martinez, and S. Oughton. Relaxation in two dimensions and the sinh-Poisson equation. *Physics of Fluids A*, 4:3–7, 1992.

[55] D. Montgomery, X. Shan, and W. H. Matthaeus. Navier-Stokes relaxation to sinh-Poisson states at finite Reynolds numbers. *Physics of Fluids A*, 5:2207–2217, 1993.

[56] D. Montgomery, L. Turner, and G. Vahala. Most probable states in magnetohydrodynamics. *Journal of Plasma Physics*, 21:239–251, 1979.

[57] J.-P. Nadal. Associative memory: on the (puzzling) sparse coding limit. *Journal of Physics A*, 24:1093–1101, 1991.

[58] J.-P. Nadal and N. Parga. Duality between learning machines: a bridge between supervised and unsupervised learning. *Neural Computation*, 6:489–506, 1994.

[59] J.-P. Nadal and N. Parga. Nonlinear neurons in the low-noise limit: a factorial code maximizes information transfer. *Network*, 5:565–581, 1994.

[60] R. M. Neal. Probabilistic inference using Markov Chain Monte Carlo methods. Technical Report CRG-TR-93-1, Department of Computer Science, University of Toronto, 1993.

[61] D. S. Ornstein. *Ergodic theory, randomness and dynamical systems*, volume 5 of *Yale Mathematical Monographs*. Yale University Press, 1974.

[62] P. M. Pardalos, D. Shalloway, and G. Xue. Optimization methods for computing global minima of nonconvex potential energy functions. *Journal of Global Optimisation*, 4:117–133, 1994.

[63] I. Prigogine. *From Being to Becoming*. W. H. Freeman, 1980.

[64] F. Rieke, D. Warland, and W. Bialek. Coding efficiency and information rates in sensory neurons. *Europhysics Letters*, 22:151–157, 1993.

[65] J. Rissanen. *Stochastic Complexity in Statistical Enquiry*. World Scientific, 1989.

[66] J. Rissanen. Shannon-Wiener information and stochastic complexity. In *Proceedings of Norbert Wiener Centenary Congress*, 1995.

[67] R. D. Rosenkrantz. *E. T. Jaynes: papers on Probability, Statistics and Statistical Physics*, volume 158 of *Synthese Library*. D. Reidel, 1983.

[68] D. L. Ruderman. The statistics of natural images. *Network*, 5:517–549, 1994.

[69] R. Russell. A footnote to David Layzer's lecture. Preprint available from russell@stp.dias.ie.

[70] G. Seroussi and M. J. Weinberger. On tree sources, finite state machines and time reversal; or, how does the tree look from the other side? to be presented at the 1995 IEEE International Symposium on Information Theory.

[71] C. E. Shannon. A mathematical theory of communication. *Bell Systems Technical Journal*, 27:379–423, 1948.

[72] A. W. Snyder, D. G. Stavenga, and S. B. Laughlin. Spatial information capacity of compound eyes. *Journal of Comparative Physiology*, 116:182–207, 1977.

[73] N. Sourlas. Statistical mechanics and error-correcting codes. Appears in [34].

[74] D. L. Stein. Spin glasses. *Scientific American*, pages 52–59, 1989.

[75] N. B. Tufillaro, P. Wyckoff, R. Brown, T. Schreiber, and T. Molteno. Topological time series analysis of a string experiment and its synchronized model. *Physical Review E*, 51:164–174, 1995.

[76] J. M. van Campenhout and T. M. Cover. Maximum entropy and conditional probability. *IEEE Transactions on Information Theory*, 4:483–489, 1981.

[77] A. Vourdas. Information in quantum optical binary communications. *IEEE Transactions on Information Theory*, 36, 1990.

[78] B. H. Weber, D. J. Depew, and J. D. Smith. *Entropy, Information and Evolution.* MIT Press, 1988.

[79] A. S. Weigend and N. A. Gershenfeld, editors. *Time Series Prediction.* Santa Fe Institute Studies in the Sciences of Complexity. Addison-Wesley, 1993.

[80] M. J. Weinberger, J. Rissanen, and M. Feder. A universal finite memory source. *IEEE Transactions on Information Theory*, IT-41, 1995.

[81] M. J. Weinberger and G. Seroussi. Sequential prediction and ranking in universal context modeling and data compression. Technical Report HPL-94-111, Hewlett-Packard Labs, 1994.

[82] E. O. Wiley and D. R. Brooks. Victims of history—a nonequilibrium approach to evolution. *Systematic Zoology*, 31:1–24, 1982. See also subsequent discussion in Syst. Zoo.

[83] K. Y. M. Wong and C. Campbell. Competitive attraction in neural networks with sign-constrained weights. *Journal of Physics A*, 25:2227–2243, 1992.

[84] K. Y. M. Wong, C. Campbell, and D. Sherrington. Order parameter evolution in a feedforward neutral network. *Journal of Physics A*, 28:1603–1615, 1995.

[85] W. H. Zurek. Algorithmic randomness and physical entropy. *Physical Review A*, 40:4731–4751, 1989.

[86] W. H. Zurek. Thermodynamic cost of computation, algorithmic complexity and the information metric. *Nature*, 341:119–125, 1989.

[87] W. H. Zurek, editor. *Complexity, Entropy and the Physics of Information.* Santa Fe Institute Studies in the Sciences of Complexity. Addison-Wesley, 1993.