Toward a Physics of Pattern

Focus Issue on
"Randomness, Structure, and Causality:
Measures of Complexity from Theory to Applications"

James P. Crutchfield[1, 2, *] and Jon Machta[3, 2, †]

¹Complexity Sciences Center & Physics Department,
University of California at Davis,
One Shields Avenue, Davis, CA 95616
²Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501
³Physics Department
University of Massachusetts, Amherst, MA 01003
(Dated: August 9, 2011)

We introduce the contributions to this Focus Issue and describe their origin in a recent Santa Fe Institute workshop.

The workshop on "Randomness, Structure, and Causality: Measures of Complexity from Theory to Applications" was held at the Santa Fe Institute in January 2011. This Focus Issue records work that was presented at and stimulated by this workshop. The contributions treat both fundamental questions in the theory of complex systems and information theory and their application to a wide range of disciplines including biology, linguistics, computation, and dynamical systems.

I. INTRODUCTION

In 1989, the Santa Fe Institute (SFI) hosted a workshop—*Complexity, Entropy, and the Physics of Information*—on fundamental definitions of complexity. This workshop and the proceedings that resulted [1] stimulated a great deal of thinking about how to define complexity. In many ways—some direct, many indirect—the foundational theme of the workshop colored much of the evolution of complex systems science since then. Complex systems science has considerably matured as a field in the intervening decades. As a result, it struck us that it was time to revisit fundamental aspects of this nascent field in a workshop. Partly, this was to take stock; but it was also to ask what innovations are needed for the coming decades, as complex systems ideas continue to extend their influence in the sciences, engineering, and humanities.

The workshop's goal was to bring together researchers from a variety of fields to discuss structural and dynamical measures of complexity appropriate for their field and the commonality between these measures. Some of the questions addressed were:

1. Are there fundamental measures of complexity that can be applied across disciplines or are measures of complexity necessarily tied to particular domains?
2. How is a system's causal organization, reflected in models of its dynamics, related to its complexity?
3. Are there universal mechanisms at work that lead to increases in complexity or does complexity arise for qualitatively different reasons in different settings?
4. Can we reach agreement on general properties that all measures of complexity must have?
5. How would the scientific community benefit from a consensus on the properties that measures of complexity should possess?
6. Some proposed measures of complexity are difficult to effectively compute. Is this problem inherent in measures of complexity generally or an indication of an unsuitable measure?

The Santa Fe Institute hosted 20 workshop participants in mid-January 2011. It turned out to be a stimulating and highly interdisciplinary group with representation from physics, biology, computer science, social science, and mathematics. An important goal was to understand the successes and difficulties in deploying complexity measures in practice. And so, participants came from both theory and experiment, with a particular emphasis on those who have constructively bridged the two.

Since the 1989 SFI workshop, a number of distinct strands have developed in the effort to define and measure complexity. Several of the well developed strands are based on:

―――――――

*Electronic address: chaos@csc.ucdavis.edu
†Electronic address: machta@physics.umass.edu

- Predictive information and excess entropy [2–7],
- Statistical complexity and causal structure [8–10],
- Logical depth and computational complexity [11–15], and
- Effective complexity [16, 17].

While these measures are broadly based on information theory or the theory of computation, the full set of connections and contrasts between them has not been sufficiently well fleshed out. Some have sought to clarify the relationship among these measures [7, 17–21] and one goal of the workshop was to foster this kind of comparative work by bringing together researchers developing various measures.

A number of lessons devolved from these early efforts, though. Several come immediately to mind:

1. There are two quite different but complementary meanings of the term "complexity". The term is used both to indicate randomness and structure. As workshop discussions repeatedly demonstrated, these are diametrically opposed concepts. Conflating them has led to much confusion.

2. Moreover, a correct understanding of complexity reveals that both are required elements of complex system. In particular, we now have a large number of cases demonstrating that structural complexity arises from the dynamical interplay of tendencies to order and tendencies to randomness. Organization in critical phenomena arising at continuous phase transitions is only one example—and a static example at that. Much of the work described in the Focus Issue addresses the interplay of randomness and structure in complex systems.

3. Even if one concentrates only on detecting and measuring randomness, one must know how the underlying system is organized. The flip side is that missing underlying structure leads one to conclude that a process is more random than it really is.

4. Elaborating on the original concept of extracting "Geometry from a Times Series" [22], we now know that processes do tell us how they are best represented. This, in turn, lends credence to the original call for "artificial science" [8]—a science that automatically builds theories of natural systems.

The lessons echoed throughout the workshop and can be seen operating in the Focus Issue contributions.

A second motivation for the workshop was to bring together workers interested in foundational questions—who were mainly from the physics, mathematics, and computer science communities—with complex systems scientists in experimental, data-driven fields who have developed quantitative measures of complexity, organization, and emergence that are useful in their fields. The range of data-driven fields using complexity measures is impressively broad: ranging from molecular excitation dynamics [23] and spectroscopic observations of the conformational dynamics of single molecules [24] through modeling subgrid structure in turbulent fluid flows [25] and new visualization methods for emergent flow patterns [26] to monitoring market efficiency [27] and the organization of animal social structure [28]. In this light, the intention was to find relations between the practically motivated measures and the more general and fundamentally motivated measures. Can the practically motivated measures be improved by an appreciation of fundamental principles? Can fundamental definitions be sharpened by considering how they interact with real-world data?

The workshop's goal was to re-ignite the efforts that began with *Complexity, Entropy, and the Physics of Information* workshop. A new level of rigor, in concepts and in analysis, is now apparent in how statistical mechanics, information theory, and computation theory can be applied to complex systems. The meteoric rise of both computer power and machine learning has led to new algorithms that address many of the computational difficulties in managing data from complex systems and in estimating various complexity measures. Given progress on all these fronts, the time was ripe to develop a much closer connection between fundamental theory and applications in many areas of complex systems science.

II. OVERVIEW OF CONTRIBUTIONS TO THE FOCUS ISSUE

The Focus Issue reflects the work of a highly interdisciplinary group of contributors representing engineering, physics, chemistry, biology, neuroscience, cognition, computer science, and mathematics. An important goal was to understand the successes and difficulties in deploying these concepts in practice. Here is a brief preview of those contributions.

A Geometric Approach to Complexity by Nihat Ay: Ay develops a thorough-going mathematical treatment of the complexity question, from the point of view of the relationship of whole versus the parts. This builds a bridge to the differential geometric approach to statistical inference pioneered by Amari in artificial neural networks. New here, Ay connects the approach to Markov processes, going beyond merely thinking of in terms of "nodes on a graph".

Partial Information Decomposition as a Spatiotemporal Filter by Benjamin Flecker, Wesley Alford, John Beggs, Paul Williams, and Randall Beer: The authors apply their new method of partial information decomposition to analyze the spacetime information dynamics generated by elementary cellular automata. This reframes recent efforts to develop a theory of local information storage, transfer, and modification. They show that prior approaches can be reinterpreted and recast into a clearer form using partial information decomposition. In particular, one that does not require arbitrarily selected threshold for detection. Importantly, the decomposition suggests a new level of semantic analysis of what would

otherwise be mere syntactic information. The authors compare alternative approaches to capturing the emergent structures in several elementary cellular automata. The contribution gives a concrete example of why and how the partial information decomposition is a valuable addition to the field of multivariate information theory.

Excess Entropy in Natural Language: Present State and Perspectives by Lukasz Debowski: Debowski explains various empirically observed phenomena in quantitative linguistics, using information theory and a mathematical model of human communication in which texts are assumed to describe an infinite set of facts in a highly repetitive fashion. He develops a relation between Herdan's law on the power-law growth of the number of distinct words used in a text of length n and Hilberg's conjecture on the power-law growth of n-symbol block entropies and mutual informations. Debowski introduces a stationary time series, the "Santa Fe process" that illustrates these ideas. Finally, Debowski proposes that maximizing mutual information (excess entropy) between adjacent text blocks leads to the emergence of random hierarchical structures. Overall, the result is a macroscopic view of the human communication system.

Effective Theories for Circuits and Automata by Simon DeDeo: DeDeo proposes a method to construct effective, coarse-grained theories from an underlying detailed mechanistic theory. This is developed in the setting of finite-state automata, with a given automaton standing in as the detailed "theory" for a phenomenon and the words recognized by the automaton being the possible "behaviors". The main technical tools come from semigroup theory and, in particular, one of its main results—the Krohn-Rhodes decomposition. The overall result is a pragmatic demonstration of the utility for nonlinear dynamics and the physical sciences of a hitherto little known but powerful mathematical formalism.

Information Symmetries in Irreversible Processes by Chris Ellison, John Mahoney, Ryan James, James Crutchfield, and Jörg Reichardt: The authors introduce the notion of dynamical irreversibility that, as they note, extends thermodynamically irreversible processes to nonequilibrium steady states. They find that most processes are dynamically irreversible: Most finite-memory processes appear statistically and structurally different in forward and reverse time. The methods introduced are constructive and lead directly to algorithms that efficiently estimate a number of complexity measures.

Local Information Measures for Spin Glasses by Matthew Robinson, David Feldman, and Susan McKay: The authors consider a disordered spin model from the perspective of local entropy densities—a novel way of characterizing disordered systems. They demonstrate the power of information-theoretic methods in the setting of quenched randomness. These methods have been proved valid for homogeneous systems, but their correctness had not been verified for diluted systems. The authors provide numerical evidence for just this and explore the consequences.

Challenges for Empirical Complexity Measures: A Perspective from Social Dynamics and Collective Social Computation by Jessica Flack and David Krakauer: The authors seek to understand social dynamics in animal systems and construct useful measures of complexity for these systems. The context for this work is the behavior of a well studied group of macaque monkeys for which there is a data set consisting of a series of fights and the individuals involved in the fights. The authors consider the predictive power of various rules that might be used by individuals to decide whether to enter a fight and how different classes of rules lead to differing collective behavior such as the fraction of fights of various sizes. They then discuss the general question of appropriate measures of complexity at different levels of description—individual, small groups of interacting individuals, and societal—and how these relate to the computational power of the individual organisms.

Anatomy of a Bit: Information in a Time Series Measurement by Ryan James, Chris Ellison, and Jim Crutchfield: The authors deconstruct the information contained in a single measurement, showing that there is a variety of different kinds of information, even in a single bit. There is information that is fleeting, created in the moment and forgotten forever. There is information that is created, but then becomes stored in the system, affecting its future behavior. They do this by surveying multivariate information measures, showing how to extend familiar measures of uncertainty and storage. Most surprisingly, their purely informational analysis identifies one kind of information that, strictly speaking, falls outside of information theory proper. This is information that is not carried in the current observation, but that is somehow stored by the system. It's existence indicates why we must build models.

Darwinian Demons, Evolutionary Complexity and Information Maximization by David Krakauer: Krakauer draws parallels between natural selection in evolutionary biology and Maxwell's demon in thermodynamics and how both take a system away from thermodynamic equilibrium at the expense of dissipation—mortality in the case of natural selection. He demonstrates bounds on the information encoded in the genome based on both the richness of the environment and the error rate of reproduction. The paper then discusses how these bounds can be circumvented by learning and plasticity and how these mechanisms can also be understood in the general framework on selective demons. Finally, he discusses the mutual information between the organism and its environment and how this form of complexity can increase indefinitely via niche construction.

Natural Complexity, Computational Complexity and Depth by Jon Machta: Machta surveys the notion of parallel computational depth as a measure of complexity. After reviewing concepts from statistical physics and computational complexity theory, he compares depth with two prominent statistical measures of complexity. In

the context of several examples, he concludes that there is structure, related to embedded computation, that is picked up by the parallel depth, but not by entropy-based measures.

How Hidden is a Hidden Process? A Primer on Crypticity and Entropy Convergence by John Mahoney, Chris Ellison, Ryan James, and Jim Crutchfield: The authors explore the difference between a process's internal stored information and the information that is directly observable. They show that this difference, the crypticity, is related to a number of other, more familiar, system properties, such as synchronizability, statistical complexity, excess entropy, and entropy convergence. They provide a detailed analysis of spin systems from this perspective.

Ergodic Parameters and Dynamical Complexity by Rui Villela-Mendes: Vilela-Mendes develops ties between the fields of dynamical systems, ergodic theory, and information theory. In particular, he extends the mathematical framework of the well known spectrum of Lyapunov exponents to include other information measures—most notably, the excess entropy, a measure of system memory. The author begins by noting that the Lyapunov spectra of a dynamical system throws away much information about the invariant measure when time averages are taken. He then proposes several ways to extend "cocycle" statistics to capture more of the dynamical behavior. A number of interesting examples and methods are presented. He introduces a "structure index," which might indicate when a system is temporally structured, containing subsystems with long-time correlations. The paper also presents a method for characterizing a "phase transition" between the random regime and the "small worlds" regime of a network, as coupling strengths are changed. Furthermore, self-organized criticality is examined from the point of view of exponent spectra.

Quantum Computation of Structural Complexity by Janet Anders and Karoline Wiesner: The authors review several of the distinctions between quantum and classical systems and then point out, via simple examples, how ideas in complexity theory should be modified to include quantum effects. They define a quantum statistical complexity and show that it less than or equal to the standard, classical statistical complexity. They also demonstrate a simple example for which entanglement permits a computation that is classically impossible.

III. CLOSING REMARKS

The workshop hosted a number of additional talks, not represented in this collection. Nevertheless, their contributions to the workshop itself were invaluable. The speakers and titles are as follows:

- Tony Bell (UC Berkeley): "Learning Out of Equilibrium";

- Luis Bettencourt (LANL): "Information Aggregation in Correlated Complex Systems and Optimal Estimation";

- Gregory Chaitin (IBM): "To a Mathematical Theory of Evolution and Biological Creativity";

- James Crutchfield (UC Davis): "Framing Complexity";

- Melanie Mitchell (PSU): "Automatic Identification of Information-Processing Structures in Cellular Automata";

- Cris Moore (UNM): "Phase Transitions and Computational Complexity";

- Rob Shaw (Protolife, Inc.): "Dominoes, Ergodic Flows"; and

- Susanne Still (U Hawaii): "Statistical Mechanics of Interactive Learning".

This Focus Issue is a permanent record of an otherwise ephemeral event. (See also `http://csc.ucdavis.edu/~chaos/share/rsc/RSC/Home.html`.) We hope that some of the workshop's spirit of creativity and forward-looking comes through. We would be particularly honored if the contributions here stimulate further research along these lines. In no sense are we at an end-point, despite the long history of these research themes. While there are a number of concrete new results presented, there is much more to do. The challenge has not diminished; rather it's grown in sophistication and clarity. We believe that a physics of pattern and how pattern arises from the interplay of structure and randomness is closer than ever.

[1] W. Zurek, editor. *Entropy, Complexity, and the Physics of Information*, volume VIII of *SFI Studies in the Sciences of Complexity*. Addison-Wesley, Reading, Massachusetts, 1990.

[2] A. del Junco and M. Rahe. Finitary codings and weak bernoulli partitions. *Proc. AMS*, 75:259, 1979.

[3] J. P. Crutchfield and N. H. Packard. Symbolic dynamics of one-dimensional maps: Entropies, finite precision, and noise. *Intl. J. Theo. Phys.*, 21:433, 1982.

[4] K-E. Eriksson and K. Lindgren. Structural information in self-organizing systems. *Physica Scripta*, 1987.

[5] P. Grassberger. Toward a quantitative theory of self-generated complexity. *Intl. J. Theo. Phys.*, 25:907, 1986.

[6] W. Ebeling, L. Molgedey, J. Kurths, and U. Schwarz. Entropy, complexity, predictability and data analysis of time series and letter sequences. `http://summa.physik.hu-berlin.de/tsd/`, 1999.

[7] J. P. Crutchfield and D. P. Feldman. Regularities unseen, randomness observed: Levels of entropy convergence. *CHAOS*, 13(1):25–54, 2003.

[8] J. P. Crutchfield and K. Young. Inferring statistical complexity. *Phys. Rev. Let.*, 63:105–108, 1989.

[9] J. P. Crutchfield and D. P. Feldman. Statistical complexity of simple one-dimensional spin systems. *Phys. Rev. E*, 55(2):1239R–1243R, 1997.

[10] C. R. Shalizi and J. P. Crutchfield. Computational mechanics: Pattern and prediction, structure and simplicity. *J. Stat. Phys.*, 104:817–879, 2001.

[11] C. H. Bennett. How to define complexity in physics, and why. In W. H. Zurek, editor, *Complexity, Entropy and the Physics of Information*, page 137. SFI Studies in the Sciences of Complexity, Vol. 7, Addison-Wesley, 1990.

[12] M. Li and P. M. B. Vitanyi. *An Introduction to Kolmogorov Complexity and its Applications*. Springer-Verlag, New York, 1993.

[13] C. H. Bennett. Universal computation and physical dynamics. *Physica D*, 86:268, 1995.

[14] J. Machta and R. Greenlaw. The computational complexity of generating random fractals. 82:1299, 1996.

[15] J. Machta. Complexity, parallel computation and statistical physics. *Complexity Journal*, 11(5):46–64, 2006.

[16] Murray Gell-Mann and Seth Lloyd. Information measures, effective complexity, and total information. *Complexity*, 2(1):44–52, 1996.

[17] Nihat Ay, Markus Mueller, and Arleta Szkola. Effective complexity and its relation to logical depth, 2008.

[18] K. Lindgren and M. G. Norhdal. Complexity measures and cellular automata. *Complex Systems*, 2(4):409–440, 1988.

[19] J. P. Crutchfield. The calculi of emergence: Computation, dynamics, and induction. *Physica D*, 75:11–54, 1994.

[20] D. P. Feldman and J. P. Crutchfield. Discovering noncritical organization: Statistical mechanical, information theoretic, and computational views of patterns in simple one-dimensional spin systems. 1998. Santa Fe Institute Working Paper 98-04-026.

[21] J. P. Crutchfield, W. Ditto, and S. Sinha. Intrinsic and designed computation: Information processing in dynamical systems—Beyond the digital hegemony. *CHAOS*, 20(3):037101, 2010.

[22] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw. Geometry from a time series. *Phys. Rev. Let.*, 45:712, 1980.

[23] D. Nerukh, V. Ryabov, and R.C. Glen. Complex temporal patterns in molecular dynamics: A direct measure of the phase-space exploration by the trajectory at macroscopic time scales. *Physical Review E*, 77(3):036225, 2008.

[24] C.-B. Li, H. Yang, and T. Komatsuzaki. Multiscale complex network of protein conformational fluctuations in single-molecule time series. *Proceedings of the National Academy of Sciences USA*, 105:536–541, 2008.

[25] A. J. Palmer, C. W. Fairall, and W. A. Brewer. Complexity in the atmosphere. *IEEE Transactions on Geoscience and Remote Sensing*, 38(4):2056–2063, 2000.

[26] H. Janicke, A. Wiebel, G. Scheuermann, and W. Kollmann. Multifield visualization using local statistical complexity. *IEEE Transactions on In Visualization and Computer Graphics*, 13(6):1384–1391, 2007.

[27] J.-S. Yang, W. Kwak, T. Kaizoji, and I.-M. Kim. Increasing market efficiency in the stock markets. *European Physical Journal B*, 61(2):241–246, 2008.

[28] N. Ay, J.C. Flack, and D.C. Krakauer. Robustness and complexity co-constructed in multimodal signalling networks. *Philos. Trans. Roy. Soc. London B*, 362:441–447, 2007.