

# The Origins of Computational Mechanics: A Brief Intellectual History and Several Clarifications

James P. Crutchfield

SFI WORKING PAPER: 2017-10-035

SFI Working Papers contain accounts of scientific work of the author(s) and do not necessarily represent the views of the Santa Fe Institute. We accept papers intended for publication in peer-reviewed journals or proceedings volumes, but not papers that have already appeared in print. Except for papers by our external faculty, papers must be based on work done at SFI, inspired by an invited visit to or collaboration at SFI, or funded by an SFI grant.

©NOTICE: This working paper is included by permission of the contributing author(s) as a means to ensure timely distribution of the scholarly and technical work on a non-commercial basis. Copyright and all rights therein are maintained by the author(s). It is understood that all persons copying this information will adhere to the terms and constraints invoked by each author's copyright. These works may be reposted only with the explicit permission of the copyright holder.

[www.santafe.edu](http://www.santafe.edu)



**SANTA FE INSTITUTE**

# The Origins of Computational Mechanics: A Brief Intellectual History and Several Clarifications

James P. Crutchfield\*

*Complexity Sciences Center and Department of Physics,  
University of California at Davis, One Shields Avenue, Davis, CA 95616*

(Dated: October 18, 2017)

The principle goal of computational mechanics is to define pattern and structure so that the organization of complex systems can be detected and quantified. Computational mechanics developed from efforts in the 1970s and early 1980s to identify strange attractors as the mechanism driving weak fluid turbulence via the method of reconstructing attractor geometry from measurement time series and in the mid-1980s to estimate equations of motion directly from complex time series. In providing a mathematical and operational definition of structure it addressed weaknesses of these early approaches to discovering patterns in natural systems.

Since then, computational mechanics has led to a range of results from theoretical physics and nonlinear mathematics to diverse applications. The former include closed-form analysis of finite- and infinite-state Markov and non-Markov stochastic processes that are ergodic or nonergodic and their measures of information and intrinsic computation. The applications range from complex materials and deterministic chaos and intelligence in Maxwellian demons to quantum compression of classical processes and the evolution of computation and language.

This brief review clarifies several misunderstandings and addresses concerns recently raised regarding early works in the field (1980s). We show that misguided evaluations of the contributions of computational mechanics are groundless and stem from a lack of familiarity with its basic goals and from a failure to consider its historical context. For all practical purposes, its modern methods and results largely supersede the early works. This not only renders recent criticism moot and shows the solid ground on which computational mechanics stands but, most importantly, shows the significant progress achieved over three decades and points to the many intriguing and outstanding challenges in understanding the computational nature of complex dynamic systems.

PACS numbers: 02.50.-r 89.70.+c 05.45.Tp 02.50.Ey 02.50.Ga

Keywords: information theory, computational mechanics, rate-distortion theory, statistical physics, symbolic dynamics, stochastic process, dynamical system

## I. GOALS

The rise of dynamical systems theory and the maturation of the statistical physics of critical phenomena in the 1960s and 1970s led to a new optimism that complicated and unpredictable phenomena in the natural world were, in fact, governed by simple, but nonlinearly interacting systems. Moreover, new mathematical concepts and increasingly powerful computers provided an entrée to understanding how such phenomena emerged over time and space. The overarching lesson was that intricate structures in a system's state space amplify microscopic uncertainties, guiding and eventually attenuating them to form complex spatiotemporal patterns. In short order, though, this new perspective on complex systems raised the question of how to quantify their unpredictability and organization.

By themselves, qualitative dynamics and statistical mechanics were mute to this challenge. The first hints at addressing it lay in Kolmogorov's (and contemporaries')

introduction of computation theory [1–3] and Shannon's information theory [4] into continuum-state dynamical systems [5–11]. This demonstrated that information had an essential role to play in physical theories of complex phenomena—a role as important as energy, but complementary. Specifically, it led to a new algorithmic foundation to randomness generated by physical systems—behavior that cannot be compressed is random—and so a bona fide measure of unpredictability of complex systems was established.

Generating information, though, is only one aspect of complex systems. How do they store and process that information? Practically, the introduction of information and algorithmic concepts side-stepped questions about how the internal mechanisms of complex systems are structured and organized. Delineating their informational architecture was not addressed, for good reason. The task is subtle.

Even if we know their governing mechanisms, complex systems (worth the label) generate patterns over long temporal and spatial scales. For example, the Navier-Stokes partial differential equations describe the local in time and space balance of forces in fluid flows. A static pres-

---

\* chaos@ucdavis.edu

sure difference leads to material flow. However, despite the fact that any flow field is governed instantaneously by these equations of motion, the equations themselves do not directly describe fluid structures such as vortices, vortex pairs, vortex streets, and vortex shedding, let alone turbulence [12]. When structures are generated at spatiotemporal scales far beyond those directly specified by the equations of motion, we say that the patterns are *emergent*.

Two questions immediately come to the fore about emergent patterns. And, this is where the subtlety arises [13]. We see that something new has emerged, but how do we objectively describe its structure and organization? And, more prosaically, how do we discover patterns in the first place?

Refining the reconstruction methods developed to identify chaotic dynamics in fluid turbulence [14, 15], computational mechanics [16, 17] provided an answer that was as simple as it was complete: a complex system’s architecture lies in its causal states. A *causal state* is a set of histories, all of which lead to the same set of futures. It’s a simple dictum: Do not distinguish histories that lead to the same predictions of the future.

The causal states and the transition dynamic over them give a canonical representation—the  $\epsilon$ -*machine*. A system’s  $\epsilon$ -machine is its unique optimal predictor of minimal size [16, 18, 19]. The historical information stored in the causal states of a process quantifies how structured the process is. A process’  $\epsilon$ -machine is its effective theory—its equations of motions. One notable aspect of the  $\epsilon$ -machine construction is that focusing on how to optimally predict a process leads to a notion of structure. Predictability and organization are inextricably intertwined.

With a system’s  $\epsilon$ -machine minimal representation in hand, the challenge of quantifying emergent organization is solved. The answer lies in a complex system’s *intrinsic computation* [16] which answers three simple questions:

1. How much of the past does a process store?
2. In what architecture is that information stored?
3. How is that stored information used to produce future behavior?

The answers are direct: the stored information is that in the causal states; the process’ architecture is laid out explicitly by the  $\epsilon$ -machine’s states and transitions; and the production of information is the process’ Shannon entropy rate.

At first blush it may not be apparent, but in this, computational mechanics parallels basic physics. Physics tracks various kinds of energy and monitors how they can be

transformed into one another. Computational mechanics asks, What kinds of information are in a system and how are they transformed into one another? Although the  $\epsilon$ -machine describes a mechanism that generates a system’s statistical properties, computational mechanics captures more than mere generation. And, this is how it was named: I wished to emphasize that it was an extension of statistical mechanics that went beyond analyzing a systems’ statistical properties to capturing its computation-theoretic properties—how a system stores and processes information, how it intrinsically computes.

## II. PROGRESS

One might be concerned that this view of complex systems is either not well grounded, on the one hand, or not practical, on the other. Over the last three decades, however, computational mechanics led to a number of novel results from theoretical physics and nonlinear mathematics that solidified its foundations to applications that attest to its utility as a way to discover new science. Discoveries from over the last decade or so give a sense of the power of the ideas and methods, both their breadth and technical depth.

Recent theoretical physics and nonlinear mathematics contributions include the following:

- Continuum and nonergodic processes [20–27];
- Analytical complexity [28–33];
- Causal rate distortion theory [34–36];
- Synchronization and control [37–40];
- Enumerating memoryful processes [16, 41, 42];
- Crypticity and causal irreversibility [43–49];
- Bayesian structural inference [50–52];
- Input-output systems [53];
- Complexity of prediction versus generation [54];
- Predictive features and their dimensions [55];
- Sufficient statistics from effective channel states [56];
- Equivalence of history and generator  $\epsilon$ -machines [57];
- Informational anatomy [58]; and
- Automated pattern detection [59].

Recent applications of computational mechanics include the following:

- Complex materials [32, 33, 60–62];
- Stochastic thermodynamics [63–71];
- Information fluctuations [72, 73];
- Information creation, destruction, and storage [74];
- Spatiotemporal computational mechanics [75, 76];
- Quantum mechanics [77–81]; and

- Evolution [82–86].

Staying true to our present needs, this must leave out detailed mention of a substantial body of computational mechanics research by others—a body that ranges from quantum theory and experiment [87–89] and stochastic dynamics [90–95] to spatial [96–101] and social systems [102].

### III. HISTORY

What’s lost in listing results is the intellectual history of computational mechanics. Where did the ideas come from? What is their historical context? What problems drove their invention? Revisiting the conditions from which computational mechanics emerged shows that aspects of this history resonate with the science that followed.

My interests started as a fascination with mainframe computers in the 1960s and with information theory in the 1970s. I worked for a number of years in Silicon Valley, for IBM at what was to become its Almaden Research Center on information storage technology—magnetic bubble devices—and at Xerox’s Palo Alto Research Center—which at the time was busily inventing our current computing environment of packet-based networks (ethernet), internet protocols, graphical user interfaces, file servers, bitmap displays, mice, and personal workstations. An active member of the Homebrew Computer Club, I built a series of microcomputers—4-bit, 8-bit, and eventually 16-bit machines. There, I met many technology buffs, several who later become titans of modern information technology. I suggested and then helped code up the first cellular automaton simulator on a prototype 6502 (8-bit) microcomputer, which would become the Apple I.

As a college student at the University of California, Santa Cruz (UCSC), I learned about the mathematics of computers and communication theory directly from the information theory pioneer David Huffman. Huffman, in particular, was well known for his 1950s work on minimal machines—on what was called *machine synthesis*. His pioneering work was an integral part of his discrete mathematics and information theory courses. Harry Huskey, one of the engineers on the first US digital computers (ENIAC and EDVAC) also taught at UCSC and I learned computer architecture from him. In short, thinking about computing and its physical substrates went hand in hand with my physics training in statistical mechanics and mathematics training in dynamical systems theory. This theme drove the bulk of my research on chaotic dynamics.

With this background in mind, let me turn to address what were the immediate concerns of nonlinear physics

in the 1980s. As computers reduced in size and cost, they became an increasingly accessible research tool. In the late 1970s and early 1980s it was this revolution that led to the burgeoning field of nonlinear dynamics. In contrast with abstract existence proofs, through computer simulations we could simply look at and interact with the solutions of complex nonlinear systems. In this way, the new tools revealed, what had been relatively abstract mathematics through most of the 20<sup>th</sup> century, a new universe of exquisitely complex, highly ramified structures and unpredictable behaviors.

Randomness emerged spontaneously, though paradoxically we knew (and had programmed) the underlying equations of motion. This presented deep challenges. What is randomness? Can we quantify it? Can we extract the underlying equations of motion from observations? Soberingly, was each and every nonlinear system, in the vast space of all systems, going to require its own “theory”? The challenge, in essence, was to describe the qualitative properties of complex systems without getting bogged down in irrelevant explicit detail and microscopic analysis. How to see the structural forest for the chaotic trees?

In the 1970s a target problem to probe these questions was identified by the nonlinear physics community—fluid turbulence—and a testable hypothesis—the Ruelle-Takens conjecture that strange attractors were the internal mechanism driving it [103]. This formalized an earlier proposal—“deterministic nonperiodic flow”—by the meteorologist Lorenz [104]: nonlinear instability was responsible for the unpredictability of weather and fluid turbulence generally.

There was a confounding problem, though. On the one hand, we had time series of measurements of the fluid velocity at a point in a flow. On the other, we had the abstract mathematics of strange attractors—complicated manifolds that circumscribed a system’s instability. How to connect them? This was solved by the proposals to use the measured time series to “reconstruct” the system’s effective state space. This was the concept of extracting the attractor’s “geometry from a time series” (1980-81) [14, 15]. These reconstruction methods created an effective state space in which to look at the chaotic attractors and to quantitatively measure their degree of instability (Kolmogorov-Sinai entropy and Lyapunov characteristic exponents) and their attendant complicatedness (embedding and fractal dimensions). This was finally verified experimentally in 1983 [105], overthrowing the decades-old Landau-Lifshitz multiple incommensurate-oscillator view of turbulence.

Reconstructing a chaotic attractor from a time series became a widely used technique for identifying and quan-

tifying deterministic chaotic behavior, leading to the field of nonlinear time series modeling [106].

Reconstruction, however, fell short of concisely expressing a system’s internal structure. Could we extend reconstruction to extract the system’s very equations of motion? A substantial benefit would be a robust way to predict chaotic behavior. The answer was provided in a method to reconstruct “Equations of Motion from a Data Series” [107, 108].

This worked quite well, when one happened to choose a mathematical representation that matched the class of nonlinear dynamics generating the behavior. But as Ref. [107] demonstrated in 1987, if you did not have the correct representational “basis” it not only failed miserably, it also did not tell you how and where to look for a better basis. Thus, even this approach to modeling complex systems had an inherent subjectivity in the choice of representation. Structural complexity remained elusive.

How to remove this subjectivity? The answer was provided by pursuing a metaphor to the classification scheme for automata developed in discrete computation theory [3, 109, 110]. There, the mathematics of formal languages and automata had led in the 1950s and 1960s to a structural hierarchy of representations that went from devices that used finite memory to infinite memories organized in different architectures—tapes, stacks, queues, counters, and the like.

Could we do this, not for discrete bit strings, but continuous chaotic systems? Answering this question led directly to computational mechanics as laid out in 1989 by Ref. [16]. The answer turned on a predictive equivalence relation developed from the geometry-of-a-time-series concept of reconstructed state [14] and adapted to an automata-theoretic setting. The equivalence relation gave a new kind of state that was a distribution of futures conditioned on past trajectories in the reconstructed state space. These were the causal states and the resulting probabilistic automata were  $\epsilon$ -machines. In this way, many of the notions of information processing and computing could be applied to nonlinear physics.

#### IV. MISDIRECTION

The preceding history introduced the goals of computational mechanics, showed its recent progress, and put its origins in the historical context of nonlinear dynamics of complex systems, such as fluid turbulence. As we will now see, the original history and recent progress form a necessary backdrop for some distracting, but pressing business.

It is abundantly clear at this point that the preceding overview is *not a literature review* on intrinsic computation embedded in complex systems. Such a review would be redundant since reviews and extensive bibliographies that cite dozens of active researchers have been provided elsewhere and at semi-regular intervals since Ref. [16] (1989); see, e.g., Refs. [17–19, 111, 112]. Rather, the preceding is provided as a narrative synopsis of its motivations, goals, and historical setting. After three decades of extensive work by many researchers in computational mechanics, why is this necessary? The reason is that critiques appeared recently that concern computational mechanics publications from the 1980s and 1990s—that is, works that are two and three decades old. And so, the early history and recent progress is a necessary backdrop.

The following addresses the issues raised and explains that, aside from several interesting, detailed mathematical issues, they are in large measure misguided. They are based on arguments that selectively pick details, either quoting them out of context or applying inappropriate contexts of interpretation. As presented, they are obscured technically so that expertise is required to evaluate the arguments. In other cases, the issues raised are not criticisms at all—they are already well known. The following (A) reviews the issues and offers a broad response that shows they are misguided at best and (B) highlights the rhetorical style of argumentation, which shows that the nontechnical (in some cases, ad hominem) arguments rely on fundamental errors of understanding. After reviewing all of them carefully, we cannot find any concern that would lead one to question the very solid and firm grounding of computational mechanics.

##### A. Technical Contentions

As analysis tools,  $\epsilon$ -machines are defined and used in two different ways. In the first they are defined via the predictive equivalence relation over sequences, as already discussed and as will be detailed shortly; these are *history  $\epsilon$ -machines*. In the second,  $\epsilon$ -machines are defined as predictive generators of processes; these are *generator  $\epsilon$ -machines*. (Mathematically, they are unifilar hidden Markov models with probabilistically distinct states that generate a given process.) The definitions are complementary. In the first, one goes from a given process to its  $\epsilon$ -machine; in the second, one specifies an  $\epsilon$ -machine to generate a given process. Importantly, the definitions are equivalent and this requires a nontrivial proof [57]. The criticisms concern history  $\epsilon$ -machines and so we need focus only on them. The computational mechanics of  $\epsilon$ -machine generators is not at issue.

Reference [113] raises technical concerns regarding sta-

tistical estimation of finite-state and probabilistic finite state machines, as discussed in several-decades-old computational mechanics publications; principally two from 1989 and 1990: Refs. [16] and [114], respectively.

The simplest response is that almost all of the concerns have been superseded by modern computational mechanics: mixed-state spectral decomposition [28, 29] and Bayesian structural inference and  $\epsilon$ -machine enumeration methods [41, 52]. The view from the present is that the issues are moot.

That said, when taken at face value, the bulk of the issues arise from technical misinterpretations. Largely, these stem from a failure to take into account that computational mechanics introduced and regularly uses a host of different equivalence relations to identify related, but different kinds of state. Ignoring this causes confusion. Specifically, it leads to Ref. [113]’s misinterpretations of covers and partitions of sequence space, transient versus recurrent causal states, the vanishing measure of nonsynchronizing sequences, and an  $\epsilon$ -machine’s start state. It also leads to a second confusion over various machine reconstruction methods. Let’s take these two kinds of misunderstanding in turn.

*Effective states and equivalence relations* One of computational mechanics’ primary starting points is to identify a stochastic process’ effective states as those determined by an equivalence relation. Said most simply, group pasts that lead to the same distribution of futures. Colloquially: do not make distinctions that do not help in prediction. The equivalence relation  $\sim$  connects two pasts  $x_{-K:0} = x_{-K} \dots x_{-1}$  and  $x_{-K':0} = x_{-K'} \dots x_{-1}$ , if the future  $X_{0:L} = X_0 \dots X_{L-1}$  after having seen each looks the same:

$$x_{-K:0} \sim x_{-K':0} \Leftrightarrow \Pr(X_{0:L}|x_{-K:0}) = \Pr(X_{0:L}|x_{-K':0}) .$$

Taking finite or infinite pasts and futures and those of equal or unequal lengths defines a family of equivalence relations and so of different kinds of causal state.

“Inferring Statistical Complexity” (1989) focused on determining a process’ long-term memory and so used  $K, K' \rightarrow \infty$  and  $L \rightarrow \infty$  [16]. That is, it worked with a process’ *recurrent causal states*, defining the process’ *statistical complexity* as the amount of information they store. This and later works also used finite pasts ( $K, K' \in \{0, 1, 2, 3, \dots\}$ ) and infinite futures ( $L \rightarrow \infty$ ) to define causal states more broadly. This introduced the notion of *transient causal states*. In turn, they suggested the more general notion of *mixed states* that monitor how an observer comes to know a process’ effective states—how the observer *synchronizes* to a process. And, finally, in this regime one has the *subtree reconstruction method*

that merges candidate states with different-length pasts. The mixed states are critical to obtaining closed-form expressions for a process’ information measures [29–31]. This setting also introduces the notion of an  $\epsilon$ -machine’s *start state*—the effective state the process is in, having a correct model in hand, but having made *no measurements*:  $K, K' = 0$ . Similarly, later works used infinite pasts and finite-length futures. Finally, using pasts and futures of equal length, but increasing them incrementally from zero leads to the class of *causal-state splitting reconstruction* methods [115].

Why all these alternatives? The answer is simple: each equivalence relation in the family poses a different question to which the resulting set of states is the answer or, at least, is an aid in answering. For example and somewhat surprisingly, Upper showed that even with infinite pasts and futures and the induced recurrent causal states, there are *elusive* and *unreachable* states that are never observed [116]. More to the point, defining other kinds of state has been helpful in other ways, too. For example, to define and then calculate a process’ *Markov* and *cryptic orders* requires a different kind of transient state [37]. Analogously, very general convergence properties of stochastic processes are proved by constructing the states of a process’ *possibility machine* [38, 39, 117].

With this flexibility in defining states, the mathematical foundations of computational mechanics give a broad set of analytical tools that tell one how a given process is organized, how it generates and transforms its information. Insisting on and using only one definition of causal state gives a greatly impoverished view of the structure of stochastic processes. Each kind is an answer to a different question. Apparently, this richness and flexibility is a source of confusion. No surprise, therefore, that if a question of interest is misunderstood, then a given representation may appear wrong, when it is in fact correct for the task at hand.

*Reconstruction methods* Reference [113] is unequivocal in its interpretation of machine reconstruction. It turns out there is little need to go into a detailed rebuttal of its statements, as they arise from a kind of misinterpretation similar to the misinterpretations discussed above. In short, Ref. [113] confuses a set of related, but distinct machine reconstruction methods.

For example, sometimes one is interested in a representation of the state machine that simply describes a process’ set of allowed realizations; that is, we are not interested in their probabilities, only which strings occur and which do not. This is the class of *topological machine reconstruction* methods; the origins of which go back to the earliest days of the theory of computation—to David Huffman’s

work. One can also, as a quick approximation, take a topologically reconstructed machine and have it read over a process’ sequence data and accumulate transition and state probabilities. This is a mixture of topological reconstruction and empirical estimation. And, finally, one can directly estimate fully probabilistic  $\epsilon$ -machines via algorithms that implement the equivalence relation of interest.

One can then use this range of reconstruction methods—topological, topological plus empirical, and probabilistic—with one or the other of the above equivalence relations.

It is important to point out that these statistical methods all have their weaknesses. That is, for a given reconstruction algorithm implementation, one can design a process sample for which the implementation will behave misleadingly. For example, it has been known for some time that causal-state splitting reconstruction methods [115] often give machines with a diverging set of states, if one presents them with increasingly more data. This occurs due to its “determinization” step, which has an exponential state-set blow-up when converting an intermediate, approximate nondeterministic presentation to a deterministic (or unifilar) one. Analogously, the subtree reconstruction method suffers from “dangling states” in which inadequate data leads to improperly estimated future conditional distributions from which there is no consistent transition. This is not surprising in the least. Many arenas of statistical inference are familiar with such problems, especially when tasked to do out-of-class modeling. The theoretical sleight of hand one finds in mathematical statistics is to *assume* data samples come from a known model class. For those interested in pattern discovery, this begs the question of what are patterns in the first place.

Now, many such problems can be overcome in a theoretical or computational research setting by presenting the algorithms with a sufficient amount data. However, in a truly empirical setting with finite data, one must take care in their use.

To address the truly empirical setting, these problems led us to introduce *Bayesian Structure Inference* for  $\epsilon$ -machines [52]. It relies on an exact enumeration of a set of candidate  $\epsilon$ -machines and related models [41]. It does not suffer from the above estimation problems in that it does not directly convert data to states and transitions as the above reconstruction algorithms do. Rather, it uses well-defined candidate models ( $\epsilon$ -machines) to estimate the probability that each produced the given data. It works well and is robust, even for very small data sets. That is, it is data parsimonious and relatively computationally efficient. And, if one has extra knowledge (from theoretical or symmetry considerations) one needs only

use a set of candidate models consistent with that knowledge. In many settings, this leads to markedly increased computational efficiency.

To close this section, it is clear that one could spend an inordinate amount of time arguing which combination of the above equivalence relations and reconstruction methods is “correct” and which is “incorrect”. This strikes me as unnecessarily narrow. The options form a toolset and those methods that produce consistent results, strengthened by testing against known cases, yield important process properties. Practically, I recommend Bayesian Structural Inference [52]. If I know a source will have low entropy rate and I want to see if it is structurally complex, though, I use probabilistic subtree reconstruction. I avoid causal-state splitting reconstruction.

## B. Rhetorical Diversions

The preceding text offers a concise rebuttal to Ref. [113]’s claims by identifying their common flaws. The latter’s technical discussion, though, is embedded in a misleading rhetorical style. The import of this misdirection may be conveyed by analyzing two less technical points that are also presented with distracting emotion.

The first is a misreading of the 1989 computational mechanics publication, “Inferring Statistical Complexity”. The claim is that the title is grossly misleading since the article is not about statistical inference. This is an oddly anachronistic view of work published 30 years ago, which seems to require looking through the lens of our present Big Data era and the current language of machine learning.

Read dispassionately, the title does allude to “inferring”, which the dictionary says is “deducing or concluding (information) from evidence and reasoning rather than from explicit statements”. And that, indeed, is how the article approaches statistical complexity—discovering patterns of intrinsic computation via the causal equivalence relation. It not only defines statistical complexity, but also introduces the mathematics to extract it. Yes, the article is not statistical inference. The topic of statistical inference as it is understood today was addressed in a number of later works; the most recent of which was mentioned above—“Bayesian Structural Inference for Hidden Processes” [52]. In short, the criticism is as specious as the rhetoric is distracting: the claim attributes anachronistic and inaccurate meanings to the article.

The second nontechnical issue is developed following a similar strategy, and it also reveals a deep misunderstanding. Packard and I had studied the convergence properties of Shannon’s entropy rate [118–120] and along with Rob

Shaw [121] had realized there was an important complexity measure—the past-future mutual information or *excess entropy*—that not only controlled convergence, but was on its own a global measure of process correlation. As those articles and Packard’s 1982 PhD dissertation [122] point out this quantity was already used to classify processes in ergodic theory [123].

Given that excess entropy’s basic properties and alternative definitions had been explored by then, Packard and I moved on to develop a more detailed scaling theory for entropy convergence, as one of the articles noted in its title “Noise Scaling of Symbolic Dynamics Entropies”. In this we defined the *normalized excess entropy*, which was normalized to its exact infinite-history, zero-noise value. This followed standard methods in phase transition theory to use “reduced” parameters. (A familiar example is the *reduced temperature*  $\hat{t} = (T - T_c)/T_c$  normalized to vanish at the critical temperature  $T_c$  at which the phase transition of interest occurs.)

The complaint is that this definition is intentionally misleading since it is not the excess entropy. Indeed, it is not. The normalized excess entropy is a proxy for a single term in the excess entropy. And, the article is absolutely clear about its focus on scaling and the tools it employs. Once one does have a theory of how entropy convergence scales, in particular the convergence rate, then it is easy to back-out the excess entropy. A simple formula expresses the excess entropy in terms of that rate and the single-symbol entropy.

So, this too is a toothless criticism, but it exemplifies the emotion and rhetorical style employed throughout Ref. [113]. The nontechnical and ad hominem criticisms intertwined with the technical faults are evidence of the consistent projection of irrelevant meanings onto the material. Once such an intellectually unproductive strategy is revealed, further rebuttal is unnecessary.

## V. FINAL REMARKS

To summarize, computational mechanics rests on firm foundations—a solidity that led to many results over the last three decades, ranging from theoretical physics and nonlinear mathematics to diverse applications. It is a direct intellectual descendant of many researchers’ efforts,

including my own, in the 1970s and early 1980s to describe the complex behaviors found in fluid turbulence.

Reference [113]’s technical claims arise from a misunderstanding of computational mechanics’ goals, methods, successes, and history. Its rhetoric reveals a strategy of quoting out of context and reinterpreting decades-old work either without benefit of modern results or projecting arbitrary assumptions onto the early work. Any dogmatic conclusions on what is “correct” that follow from such a strategy are flawed. Moreover, Ref. [113]’s claims to precedence are based on false memories, are unsubstantiated, and, in light of the history of events, are unsubstantiatable.

Current work simply eclipses the questions raised in distant retrospect, rendering the criticisms moot. Time passes. We should let it move on.

Over the years, computational mechanics has been broadly extended and applied, far beyond its initial conception 30 years ago. That said, its hope to lay the foundations of a fully automated “artificial science” [16]—in which theories are built automatically from raw data—remains a challenge. Though the benefits are tantalizing, it was and remains an ambitious goal.

## ACKNOWLEDGMENTS

I thank many colleagues who, over the years, improved, corrected, and extended computational mechanics using constructive debate and civil correspondence and who made the effort to understand its motivations, its contributions, and its limitations. Most immediately, I am grateful to Cina Aghamohammadi, Korana Burke, Chris Ellison, Jeff Emenheiser, Dave Feldman, David Gier, Mile Gu, Martin Hilbert, Ryan James, John Mahoney, Sarah Marzen, Norman Packard, Geoff Pryde, Paul Riechers, Dawn Sumner, Susanne Still, Meredith Tromble, Dorman Varn, Greg Wimsatt, Howard Wiseman, and Karl Young for helpful comments. I thank the Santa Fe Institute for its hospitality during visits, where I have been a faculty member for three decades. This material is based upon work supported by, or in part by, John Templeton Foundation grant 52095, Foundational Questions Institute grant FQXi-RFP-1609, the U.S. Army Research Laboratory and the U. S. Army Research Office under contracts W911NF-13-1-0390, W911NF-13-1-0340, and W911NF-12-1-0288.

---

[1] A. M. Turing. On computable numbers, with an application to the entscheidungsproblem. *Proc. Lond. Math.*

*Soc. Ser. 2*, 42:230, 1936. 1

- [2] C. E. Shannon. A universal Turing machine with two internal states. In C. E. Shannon and J. McCarthy, editors, *Automata Studies*, number 34 in *Annals of Mathematical Studies*, pages 157–165. Princeton University Press, Princeton, New Jersey, 1956.
- [3] M. Minsky. *Computation: Finite and Infinite Machines*. Prentice-Hall, Englewood Cliffs, New Jersey, 1967. 1, 4
- [4] C. E. Shannon. A mathematical theory of communication. *Bell Sys. Tech. J.*, 27:379–423, 623–656, 1948. 1
- [5] A. N. Kolmogorov. *Foundations of the Theory of Probability*. Chelsea Publishing Company, New York, second edition, 1956. 1
- [6] A. N. Kolmogorov. Three approaches to the concept of the amount of information. *Prob. Info. Trans.*, 1:1, 1965.
- [7] A. N. Kolmogorov. Combinatorial foundations of information theory and the calculus of probabilities. *Russ. Math. Surveys*, 38:29–40, 1983.
- [8] A. N. Kolmogorov. A new metric invariant of transient dynamical systems and automorphisms in Lebesgue spaces. *Dokl. Akad. Nauk. SSSR*, 119:861, 1958. (Russian) *Math. Rev.* vol. 21, no. 2035a.
- [9] A. N. Kolmogorov. Entropy per unit time as a metric invariant of automorphisms. *Dokl. Akad. Nauk. SSSR*, 124:754, 1959. (Russian) *Math. Rev.* vol. 21, no. 2035b.
- [10] Ja. G. Sinai. On the notion of entropy of a dynamical system. *Dokl. Akad. Nauk. SSSR*, 124:768, 1959.
- [11] G. Chaitin. On the length of programs for computing finite binary sequences. *J. ACM*, 13:145, 1966. 1
- [12] W. Heisenberg. Nonlinear problems in physics. *Physics Today*, 20:23–33, 1967. 2
- [13] J. P. Crutchfield. Is anything ever new? Considering emergence. In G. Cowan, D. Pines, and D. Melzner, editors, *Complexity: Metaphors, Models, and Reality*, volume XIX of *Santa Fe Institute Studies in the Sciences of Complexity*, pages 479–497, Reading, MA, 1994. Addison-Wesley. Santa Fe Institute Technical Report 94-03-011; reprinted in *Emergence: Contemporary Readings in Philosophy and Science*, M. A. Bedau and P. Humphreys, editors, Bradford Book, MIT Press, Cambridge, MA (2008) 269–286. 2
- [14] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw. Geometry from a time series. *Phys. Rev. Lett.*, 45:712, 1980. 2, 3, 4
- [15] F. Takens. Detecting strange attractors in fluid turbulence. In D. A. Rand and L. S. Young, editors, *Symposium on Dynamical Systems and Turbulence*, volume 898, page 366, Berlin, 1981. Springer-Verlag. 2, 3
- [16] J. P. Crutchfield and K. Young. Inferring statistical complexity. *Phys. Rev. Lett.*, 63:105–108, 1989. 2, 4, 5, 7
- [17] J. P. Crutchfield. Between order and chaos. *Nature Physics*, 8(January):17–24, 2012. 2, 4
- [18] J. P. Crutchfield and C. R. Shalizi. Thermodynamic depth of causal states: Objective complexity via minimal representations. *Phys. Rev. E*, 59(1):275–283, 1999. 2
- [19] C. R. Shalizi and J. P. Crutchfield. Computational mechanics: Pattern and prediction, structure and simplicity. *J. Stat. Phys.*, 104:817–879, 2001. 2, 4
- [20] S. Marzen and J. P. Crutchfield. Informational and causal architecture of continuous-time renewal processes. *J. Stat. Phys.*, 168(a):109–127, 2017. 2
- [21] S. Marzen and J. P. Crutchfield. Structure and randomness of continuous-time discrete-event processes. *J. Stat. Physics*, in press, 2016.
- [22] S. Marzen and J. P. Crutchfield. Informational and causal architecture of discrete-time renewal processes. *Entropy*, 17(7):4891–4917, 2015.
- [23] S. Marzen and J. P. Crutchfield. Statistical signatures of structural organization: The case of long memory in renewal processes. *Phys. Lett. A*, 380(17):1517–1525, 2016.
- [24] S. Marzen, M. R. DeWeese, and J. P. Crutchfield. Time resolution dependence of information measures for spiking neurons: Scaling and universality. *Front. Comput. Neurosci.*, 9:109, 2015.
- [25] S. Marzen and J. P. Crutchfield. Information anatomy of stochastic equilibria. *Entropy*, 16(9):4713–4748, 2014.
- [26] J. P. Crutchfield and S. Marzen. Signatures of infinity: Nonergodicity and resource scaling in prediction, complexity, and learning. *Phys. Rev. E*, 91(5):050106(R), 2015.
- [27] N. Travers and J. P. Crutchfield. Infinite excess entropy processes with countable-state generators. *Entropy*, 16:1396–1413, 2014. 2
- [28] P. M. Riechers and J. P. Crutchfield. Beyond the spectral theorem: Decomposing arbitrary functions of nondiagonalizable operators. 2016. Santa Fe Institute Working Paper 16-07-015; arxiv.org:1607.06526 [math-ph]. 2, 5
- [29] J. P. Crutchfield, P. Riechers, and C. J. Ellison. Exact complexity: Spectral decomposition of intrinsic computation. *Phys. Lett. A*, 380(9-10):998–1002, 2016. 5
- [30] P. Riechers and J. P. Crutchfield. Spectral simplicity of apparent complexity, Part I: The nondiagonalizable metadynamics of prediction. submitted. Santa Fe Institute Working Paper 2017-05-018; arxiv.org:1705.08042.
- [31] P. Riechers and J. P. Crutchfield. Spectral simplicity of apparent complexity, Part II: Exact complexities and complexity spectra. submitted. Santa Fe Institute Working Paper 17-06-019; arxiv.org:1706.00883. 5
- [32] P. M. Riechers, D. P. Varn, and J. P. Crutchfield. Pairwise correlations in layered close-packed structures. *Acta Cryst. A*, 71(4):423–443, 2015. 2
- [33] P. M. Riechers, D. P. Varn, and J. P. Crutchfield. Diffraction patterns of layered close-packed structures from hidden markov models. 2014. Santa Fe Institute Working Paper 2014-10-038; arxiv.org:1410.5028 [cond-mat.mtrl-sci]. 2
- [34] S. Still and J. P. Crutchfield. Structure or noise? 2007. Santa Fe Institute Working Paper 2007-08-020; arxiv.org physics.gen-ph/0708.0654. 2
- [35] S. Still, J. P. Crutchfield, and C. J. Ellison. Optimal causal inference: Estimating stored information and approximating causal architecture. *CHAOS*, 20(3):037111, 2010.

- [36] S. Marzen and J. P. Crutchfield. Predictive rate-distortion for infinite-order Markov processes. *J. Stat. Phys.*, 163(6):1312–1338, 2014. 2
- [37] R. G. James, J. R. Mahoney, C. J. Ellison, and J. P. Crutchfield. Many roads to synchrony: Natural time scales and their algorithms. *Phys. Rev. E*, 89:042135, 2014. 2, 5
- [38] N. Travers and J. P. Crutchfield. Exact synchronization for finite-state sources. *J. Stat. Phys.*, 145(5):1181–1201, 2011. 5
- [39] N. Travers and J. P. Crutchfield. Asymptotic synchronization for finite-state sources. *J. Stat. Phys.*, 145(5):1202–1223, 2011. 5
- [40] J. P. Crutchfield, C. J. Ellison, J. R. Mahoney, and R. G. James. Synchronization and control in intrinsic and designed computation: An information-theoretic analysis of competing models of stochastic computation. *CHAOS*, 20(3):037105, 2010. 2
- [41] B. D. Johnson, J. P. Crutchfield, C. J. Ellison, and C. S. McTague. Enumerating finitary processes. 2012. SFI Working Paper 10-11-027; arxiv.org:1011.0036 [cs.FL]. 2, 5, 6
- [42] D. P. Feldman, C. S. McTague, and J. P. Crutchfield. The organization of intrinsic computation: Complexity-entropy diagrams and the diversity of natural information processing. *CHAOS*, 18(4):043106, 2008. 2
- [43] J. P. Crutchfield, C. J. Ellison, and J. R. Mahoney. Time’s barbed arrow: Irreversibility, crypticity, and stored information. *Phys. Rev. Lett.*, 103(9):094101, 2009. 2
- [44] C. J. Ellison, J. R. Mahoney, and J. P. Crutchfield. Prediction, retrodiction, and the amount of information stored in the present. *J. Stat. Phys.*, 136(6):1005–1034, 2009.
- [45] J. R. Mahoney, C. J. Ellison, and J. P. Crutchfield. Information accessibility and cryptic processes: Linear combinations of causal states. 2009. arxiv.org:0906.5099 [cond-mat].
- [46] J. R. Mahoney, C. J. Ellison, and J. P. Crutchfield. Information accessibility and cryptic processes. *J. Phys. A: Math. Theo.*, 42:362002, 2009.
- [47] J. P. Crutchfield and C. J. Ellison. The past and the future in the present. 2014. SFI Working Paper 10-12-034; arxiv.org:1012.0356 [nlin.CD].
- [48] J. R. Mahoney, C. J. Ellison, R. G. James, and J. P. Crutchfield. How hidden are hidden processes? A primer on crypticity and entropy convergence. *CHAOS*, 21(3):037112, 2011.
- [49] C. J. Ellison, J. R. Mahoney, R. G. James, J. P. Crutchfield, and J. Reichardt. Information symmetries in irreversible processes. *CHAOS*, 21(3):037107, 2011. 2
- [50] C. C. Strelhoff and J. P. Crutchfield. Optimal instruments and models for noisy chaos. *CHAOS*, 17:043127, 2007. 2
- [51] C. C. Strelhoff, J. P. Crutchfield, and A. Hübler. Inferring Markov chains: Bayesian estimation, model comparison, entropy rate, and out-of-class modeling. *Phys. Rev. E*, 76(1):011106, 2007.
- [52] C. C. Strelhoff and J. P. Crutchfield. Bayesian structural inference for hidden processes. *Phys. Rev. E*, 89:042119, 2014. 2, 5, 6
- [53] N. Barnett and J. P. Crutchfield. Computational mechanics of input-output processes: Structured transformations and the  $\epsilon$ -transducer. *J. Stat. Phys.*, 161(2):404–451, 2015. 2
- [54] J. Ruebeck, R. G. James, John R. Mahoney, and J. P. Crutchfield. Prediction and generation of binary markov processes: Can a finite-state fox catch a markov mouse? 2017. SFI Working Paper 2017-08-027; arxiv.org:1708.00113 [cond-mat.stat-mech]. 2
- [55] S. E. Marzen and J. P. Crutchfield. Nearly maximally predictive features and their dimensions. *Phys. Rev. E*, 95(5):051301(R), 2017. 2
- [56] R. G. James, J. R. Mahoney, and J. P. Crutchfield. Information trimming: Sufficient statistics, mutual information, and predictability from effective channel states. *Phys. Rev. E*, 95(6):060102, 2017. 2
- [57] N. Travers and J. P. Crutchfield. Equivalence of history and generator  $\epsilon$ -machines. 2014. SFI Working Paper 11-11-051; arxiv.org:1111.4500 [math.PR]. 2, 4
- [58] R. G. James, C. J. Ellison, and J. P. Crutchfield. Anatomy of a bit: Information in a time series observation. *CHAOS*, 21(3):037109, 2011. 2
- [59] C. S. McTague and J. P. Crutchfield. Automated pattern discovery—An algorithm for constructing optimally synchronizing multi-regular language filters. *Theoretical Computer Science*, 359(1-3):306–328, 2006. 2
- [60] D. P. Varn and J. P. Crutchfield. What did Erwin mean? The physics of information from the materials genomics of aperiodic crystals and water to molecular information catalysts and life. *Phil. Trans. Roy. Soc. A*, 374:20150067, 2016. 2
- [61] D. P. Varn and J. P. Crutchfield. Chaotic crystallography: How the physics of information reveals structural order in materials. *Curr. Opin. Chem. Eng.*, 7:47–56, 2015.
- [62] X. Lei, D. P. Varn, and J. P. Crutchfield. Islands in the gap: Intertwined transport and localized in structurally complex materials. 2017. SFI Working Paper 17-07-025; arxiv.org:1707.05894 [cond-mat.mtrl-sci]. 2
- [63] S. E. Marzen and J. P. Crutchfield. Prediction and power in molecular sensors: Uncertainty and dissipation when conditionally markovian channels are driven by semi-markov environments. 2017. SFI Working Paper 2017-07-020; arxiv.org:1702.08565 [cond-mat.stat-mech]. 2
- [64] C. Aghamohammadi and J. P. Crutchfield. Thermodynamics of random number generation. *Phys. Rev. E*, 95(6):062139, 2017.
- [65] P. M. Riechers and J. P. Crutchfield. Fluctuations when driving between nonequilibrium steady states. *J. Stat. Phys.*, 168(4):873–918, 2017.
- [66] A. B. Boyd, D. Mandal, and J. P. Crutchfield. Correlation-powered information engines and the thermodynamics of self-correction. *Phys. Rev. E*, 95(1):012152, 2017.

- [67] A. B. Boyd and J. P. Crutchfield. Maxwell demon dynamics: Deterministic chaos, the Szilard map, and the intelligence of thermodynamic systems. *Phys. Rev. Lett.*, 116:190601, 2016.
- [68] A. B. Boyd, D. Mandal, and J. P. Crutchfield. Above and beyond the Landauer bound: Thermodynamics of modularity. 2017. SFI Working Paper 2017-08-030; arxiv.org:1708.03030 [cond-mat.stat-mech].
- [69] A. B. Boyd, D. Mandal, and J. P. Crutchfield. Identifying functional thermodynamics in autonomous Maxwellian ratchets. *New J. Physics*, 18:023049, 2016.
- [70] A. B. Boyd, D. Mandal, P. M. Riechers, and J. P. Crutchfield. Transient dissipation and structural costs of physical information transduction. *Phys. Rev. Lett.*, 118:220602, 2017.
- [71] A. B. Boyd, D. Mandal, and J. P. Crutchfield. Leveraging environmental correlations: The thermodynamics of requisite variety. *J. Stat. Phys.*, 167(6):1555–1585, 2016. 2
- [72] C. Aghamohammadi and J. P. Crutchfield. Minimum memory for generating rare events. *Phys. Rev. E*, 95(3):032101, 2017. 2
- [73] J. P. Crutchfield and C. Aghamohammadi. Not all fluctuations are created equal: Spontaneous variations in thermodynamic function. 2016. Santa Fe Institute Working Paper 16-09-018; arxiv.org:1609.02519 [math-ph]. 2
- [74] R. G. James, K. Burke, and J. P. Crutchfield. Chaos forgets and remembers: Measuring information creation, destruction, and storage. *Phys. Lett. A*, 378:2124–2127, 2014. 2
- [75] A. Rupe, J. P. Crutchfield, K. Kashinath, and Prabhath. A physics-based approach to unsupervised discovery of coherent structures in spatiotemporal systems. Santa Fe Institute Working Paper 2017-09-033; arXiv.org:1709.03184 [physics.flu-dyn]. 2
- [76] A. Rupe and J. P. Crutchfield. Local causal states and discrete coherent structures. in preparation. 2
- [77] C. Aghamohammadi, S. P. Loomis, J. R. Mahoney, and J. P. Crutchfield. Extreme quantum advantage for rare-event sampling. 2017. Santa Fe Institute Working Paper 2017-08-029; arxiv.org:1707.09553 [quant-ph]. 2
- [78] C. Aghamohammadi, J. R. Mahoney, and J. P. Crutchfield. Extreme quantum advantage when simulating strongly coupled classical systems. *Sci. Reports*, 7(6735):1–11, 2017.
- [79] P. M. Riechers, J. R. Mahoney, C. Aghamohammadi, and J. P. Crutchfield. Minimized state-complexity of quantum-encoded cryptic processes. *Phys. Rev. A*, 93(5):052317, 2016.
- [80] J. R. Mahoney, C. Aghamohammadi, and J. P. Crutchfield. Occam’s quantum stop: Synchronizing and compressing classical cryptic processes via a quantum channel. *Scientific Reports*, 6:20495, 2016.
- [81] C. Aghamohammadi, J. R. Mahoney, and J. P. Crutchfield. The ambiguity of simplicity. *Phys. Lett. A*, 381(14):1223–1227, 2017. 2
- [82] J. P. Crutchfield and S. Whalen. Structural drift: The population dynamics of sequential learning. *PLoS Computational Biology*, 8(6):e1002510, 2010. 3
- [83] J. P. Crutchfield and M. Mitchell. The evolution of emergent computation. *Proc. Natl. Acad. Sci.*, 92:10742–10746, 1995.
- [84] J. P. Crutchfield and O. Gernerup. Objects that make objects: The population dynamics of structural complexity. *J. Roy. Soc. Interface*, 3:345–349, 2006.
- [85] O. Gernerup and J. P. Crutchfield. Hierarchical self-organization in the finitary process soup. *Artificial Life Journal*, 14(3):245–254, 2008. Special Issue on the Evolution of Complexity; arxiv.org nlin.AO/0603001; Proceedings of Artificial Life X.
- [86] O. Gernerup and J. P. Crutchfield. Primordial evolution in the finitary process soup. In I. Licata and A. Sakaji, editors, *Physics of Emergence and Organization*, pages 297–311. World Scientific Publishers, New Jersey, 2008. 3
- [87] M. Gu, K. Wiesner, E. Rieper, and V. Vedral. Quantum mechanics can reduce the complexity of classical models. *Nature Comm.*, 3(762):1–5, 2012. 3
- [88] M. S. Palsson, M. Gu, H. Ho, H. M. Wiseman, and G. J. Pryde. Experimental quantum processing enhancement in modelling stochastic processes. *arXiv:1602.05683*.
- [89] J. Thompson, A. J. P. Garner, V. Vedral, and M. Gu. Using quantum theory to simplify input-output processes. *Nature Quant. Info.*, 3:6, 2017. 3
- [90] V. Ryabov and D. Nerukh. Computational mechanics of molecular systems: Quantifying high-dimensional dynamics by distribution of Poincare recurrence times. *CHAOS*, 21(3):037113, 2011. 3
- [91] D. Nerukh, C. H. Jensen, and R. C. Glen. Identifying and correcting non-Markov states in peptide conformational dynamics. *J. Chem. Physics*, 132:084104, 2010.
- [92] D. Nerukh. Non-Markov state model of peptide dynamics. *J. Molec. Liquids*, 176:65–70, 2012.
- [93] D. Kelly, M. Dillingham, A. Hudson, and K. Wiesner. A new method for inferring hidden Markov models from noisy time sequences. *PLoS One*, 7(1):e29703, 01 2012.
- [94] C.-B. Li, H. Yang, and T. Komatsuzaki. Multiscale complex network of protein conformational fluctuations in single-molecule time series. *Proc. Natl. Acad. Sci. USA*, 105:536–541, 2008.
- [95] A. Witt, A. Neiman, and J. Kurths. Characterizing the dynamics of stochastic bistable systems by measures of complexity. *Phys. Rev. E*, 55:5050–5059, 1997. 3
- [96] J. Delgado and R. V. Solé. Collective-induced computation. *Phys. Rev. E*, 55:2338–2344, 1997. 3
- [97] R. Das. *The Evolution of Emergent Computation in Cellular Automata*. PhD thesis, Colorado State University, 1996.
- [98] W. Hordijk. *Dynamics, Emergent Computation, and Evolution in Cellular Automata*. PhD thesis, University of New Mexico, Albuquerque, 1999.
- [99] W. M. Gonçalves, R. D. Pinto, J. C. Sartorelli, and M. J. de Oliveira. Inferring statistical complexity in the dripping faucet experiment. *Physica A*, 257(1-4):385–389, 1998.

- [100] A. J. Palmer, C. W. Fairall, and W. A. Brewer. Complexity in the atmosphere. *IEEE Trans. Geosci. Remote Sens.*, 38:2056–2063, 2000.
- [101] R. W. Clarke, M. P. Freeman, and N. W. Watkins. The application of computational mechanics to the analysis of geomagnetic data. *Phys. Rev. E*, 67:160–203, 2003. 3
- [102] D. Darmon, J. Sylvester, M. Girvan, and W. Rand. Predictability of user behavior in social media: Bottom-up versus top-down modeling. [arxiv.org:1306.6111](https://arxiv.org/abs/1306.6111). 3
- [103] D. Ruelle and F. Takens. On the nature of turbulence. *Comm. Math. Phys.*, 20:167–192, 1971. 3
- [104] E. N. Lorenz. Deterministic nonperiodic flow. *J. Atmos. Sci.*, 20:130, 1963. 3
- [105] A. Brandstater, J. Swift, Harry L. Swinney, A. Wolf, J. D. Farmer, E. Jen, and J. P. Crutchfield. Low-dimensional chaos in a hydrodynamic system. *Phys. Rev. Lett.*, 51:1442, 1983. 3
- [106] M. Casdagli and S. Eubank, editors. *Nonlinear Modeling*, SFI Studies in the Sciences of Complexity, Reading, Massachusetts, 1992. Addison-Wesley. 4
- [107] J. P. Crutchfield and B. S. McNamara. Equations of motion from a data series. *Complex Systems*, 1:417 – 452, 1987. 4
- [108] J. D. Farmer and J. Sidorowitch. Predicting chaotic time series. *Phys. Rev. Lett.*, 59:366, 1987. 4
- [109] N. Chomsky. Three models for the description of language. *IRE Trans. Info. Th.*, 2:113–124, 1956. 4
- [110] J. E. Hopcroft and J. D. Ullman. *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley, Reading, 1979. 4
- [111] J. P. Crutchfield. The calculi of emergence: Computation, dynamics, and induction. *Physica D*, 75:11–54, 1994. 4
- [112] J. P. Crutchfield and D. P. Feldman. Regularities unseen, randomness observed: Levels of entropy convergence. *CHAOS*, 13(1):25–54, 2003. 4
- [113] P. Grassberger. [arXiv:1708.04190](https://arxiv.org/abs/1708.04190), [1708.03197](https://arxiv.org/abs/1708.03197). 4, 5, 6, 7
- [114] J. P. Crutchfield and K. Young. Computation at the onset of chaos. In W. Zurek, editor, *Entropy, Complexity, and the Physics of Information*, volume VIII of *SFI Studies in the Sciences of Complexity*, pages 223 – 269, Reading, Massachusetts, 1990. Addison-Wesley. 5
- [115] C. R. Shalizi, K. L. Shalizi, and J. P. Crutchfield. Pattern discovery in time series, Part I: Theory, algorithm, analysis, and convergence. [arXiv.org/abs/cs.LG/0210025](https://arxiv.org/abs/cs.LG/0210025). 5, 6
- [116] D. R. Upper. *Theory and Algorithms for Hidden Markov Models and Generalized Hidden Markov Models*. PhD thesis, University of California, Berkeley, 1997. Published by University Microfilms Intl, Ann Arbor, Michigan. 5
- [117] N. F. Travers. Exponential bounds for convergence of entropy rate approximations in hidden Markov models satisfying a path-mergeability condition. *Stochastic Proc. Appln.*, 124(12):4149–4170, 2014. 5
- [118] 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. 6
- [119] J. P. Crutchfield and N. H. Packard. Noise scaling of symbolic dynamics entropies. In H. Haken, editor, *Evolution of Order and Chaos*, pages 215–227, Berlin, 1982. Springer-Verlag.
- [120] J. P. Crutchfield and N. H. Packard. Symbolic dynamics of noisy chaos. *Physica*, 7D(1-3):201 – 223, 1983. 6
- [121] R. Shaw. *The Dripping Faucet as a Model Chaotic System*. Aerial Press, Santa Cruz, California, 1984. 7
- [122] N. H. Packard. *Measurements of Chaos in the Presence of Noise*. PhD thesis, University of California, Santa Cruz, 1982. 7
- [123] A. del Junco and M. Rahe. Finitary codings and weak Bernoulli partitions. *Proc. AMS*, 75:259, 1979. 7