REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
10-07-2015	Final Report	11-Jul-2012 - 10-Jan-2015

4. TITLE AND SUBTITLE
Final Report: Structural Complexity in Linguistic Systems (Research Topic 3: Mathematical Sciences)

5a. CONTRACT NUMBER
W911NF-12-1-0288

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHORS
James P. Crutchfield

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
University of California - Davis
1850 Research Park Drive
Suite 300
Davis, CA 95618 -6153

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
62169-MA.14

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
We adapted computational mechanics to determine the role of structural complexity in linguistic systems, including language systems that adapt and evolve. The latter pose substantial challenges to any conventional communication-channel view of language. Our approach promised to overcome these difficulties, since it gives a clear and constructive view of structure in memoryful stochastic processes. In principle, this allows on to track how intrinsic structure evolves over time and to explore the tendency of language systems to develop in ways that balance structure (expressed in vocabulary and syntax) and entropy (the need to convey novelty and express meaning

15. SUBJECT TERMS
statistical complexity, self-organization, structure, memory, information, hidden Markov model, epsilon-machine, entropy, quantitative linguistics, computational linguistics

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
UU	UU	UU	UU		James Crutchfield
					19b. TELEPHONE NUMBER: 530-752-0060

Standard Form 298 (Rev 8/98)
Prescribed by ANSI Std. Z39.18

Report Title

Final Report: Structural Complexity in Linguistic Systems (Research Topic 3: Mathematical Sciences)

ABSTRACT

We adapted computational mechanics to determine the role of structural complexity in linguistic systems, including language systems that adapt and evolve. The latter pose substantial challenges to any conventional communication-channel view of language. Our approach promised to overcome these difficulties, since it gives a clear and constructive view of structure in memoryful stochastic processes. In principle, this allows on to track how intrinsic structure evolves over time and to explore the tendency of language systems to develop in ways that balance structure (expressed in vocabulary and syntax) and entropy (the need to convey novelty and express meaning unpredicted by the receiver).

Given the project's short duration and episodic funding, we focused on developing several measures of intrinsic structural complexity applicable to probe structure in linguistic models and language corpora. Using these, we analyzed the role of stored, hidden, acausal, and nonpredictive informations. We analyzed how a receiver comes to understand an information source—via what we call synchronization. Finally, we extended computational mechanics techniques to address the same questions in infinite-memory communication systems. However, the project ended just as the tools were refined and could be applied to actual language corpora.

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Received		Paper
07/10/2015	13.00	James P. Crutchfield, Sarah Marzen. Signatures of infinity: Nonergodicity and resource scaling in prediction, complexity, and learning, Physical Review E, (05 2015): 0. doi: 10.1103/PhysRevE.91.050106
08/26/2014	3.00	John R. Mahoney, Christopher J. Ellison, James P. Crutchfield, Ryan G. James. Many roads to synchrony: Natural time scales and their algorithms, Physical Review E, (04 2014): 0. doi: 10.1103/PhysRevE.89.042135
08/26/2014	4.00	Ryan G. James, Korana Burke, James P. Crutchfield. Chaos forgets and remembers: Measuring information creation, destruction, and storage, Physics Letters A, (06 2014): 0. doi: 10.1016/j.physleta.2014.05.014
08/26/2014	5.00	Christopher C. Strelioff, James P. Crutchfield. Bayesian structural inference for hidden processes, Physical Review E, (04 2014): 0. doi: 10.1103/PhysRevE.89.042119
08/26/2014	6.00	James Crutchfield, Nicholas Travers. Infinite Excess Entropy Processes with Countable-State Generators, Entropy, (03 2014): 0. doi: 10.3390/e16031396

TOTAL: **5**

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

Received	Paper

TOTAL:

Number of Papers published in non peer-reviewed journals:

(c) Presentations

-- Sarah Marzen (UCB) “Information Theoretic Methods of Time Series Analysis”, March 2014, Machine Learning seminar, Computer Science Department, University of Hawaii, Manoa.
-- Sarah Marzen (UCB), “Causal Structure of Neural Spikes Trains”, April 2014, Redwood Center for Theoretical Neuroscience, UC Berkeley.
-- Sarah Marzen, “Causal Structure of Neural Spikes Trains”, April 2014, Joint Redwood-Ganguli-Theunissen Lab Meeting, UC Berkeley.
-- Sarah Marzen, Mike DeWeese, Jim Crutchfield, “Statistical Complexity of Neural Spike Trains”, Poster presentation, March 2014, Computational and Systems Neuroscience (Cosyne 2014), Salt Lake City, Utah.
-- Training: JPC taught two graduate courses Winter and Spring 2014 on the physics of information and the physics of computation, respectively.
-- PI Crutchfield gave John von Neumann Complexity and Computation Public Lecture “Demonology: The Curious Role of Intelligence in Physics & Biology”, Wisconsin Institutes for Discovery. (6 November 2013.)
-- Nix Barnett, “ε-Transducers: Parts I, II, and III”, 27 May and 9 and 16 July 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
-- Alexandra Nilles, “Applying Predicted Information Gain to Physically Embodied Agents”, 13 August 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis; 14 August 2014, REU Presentations, Physics, UC Davis.
-- Aleksander Zujev “Coupled Systems”, 25 June 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
-- Benny Brown, review of D. Little and F. Sommer “Learning and exploration in action-perception loops”, 21 May and 18 June 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
-- Fritz Sommer (Redwood Ctr, UCB), “Predictive Information Gain for Adaptive Learning”, 11 June 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
-- Sarah Marzen, “Information Anatomy of Continuous Stochastic Processes”, 16 April 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
-- Ryan James (U Colorado, Boulder), “Intersection Information”, 2 April 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
-- Korana Burke and Greg Wimsatt, review of Physical Review Letter “Measurement of Stochastic Entropy Product”, 12 March 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
-- Katie Amrine (Dept Viticulture and Enology) on “Measuring ‘Functional’ Information in transfer-RNAs to detect and resolve inconsistencies in deep roots of the Tree of Life”, 12 February 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
-- Pooneh Mohammadiara on “Elusive Information”, 5 February 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
-- Sarah Marzen on “Structural Complexity of Stochastic Renewal Processes”, 29 January 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
-- J. P. Crutchfield, “Information in Infinitary Processes”, 3 December 2013, Workshop on “Information in Infinitary Processes”, 3-4 December 2013, Complexity Sciences Center, UC Davis.
-- Dave Feldman (College of the Atlantic), “2D Information Theory”, 3 December 2013, Workshop on “Information in Infinitary Processes”, 3-4 December 2013, Complexity Sciences Center, UC Davis.
-- Chris Ellison (WID, Wisconsin), “Countably Infinite Processes”, 3 December 2013, Workshop on “Information in Infinitary Processes”, 3-4 December 2013, Complexity Sciences Center, UC Davis.
-- Sarah Marzen (UCB), “Continuous-Time Processes”, 3 December 2013, Workshop on “Information in Infinitary Processes”, 3-4 December 2013, Complexity Sciences Center, UC Davis.
-- Chris Strelioff, “Inferring Infinite-State Processes”, 3 December 2013, Workshop on “Information in Infinitary Processes”, 3-4 December 2013, Complexity Sciences Center, UC Davis.
-- Paul Riechers, “Spectral Decomposition”, 3 December 2013, Workshop on “Information in Infinitary Processes”, 3-4 December 2013, Complexity Sciences Center, UC Davis.
-- Korana Burke, reviewed Physical Review E “Information Content in Turbulence”, 20 November 2013, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
-- J. P. Crutchfield, “Information in Infinitary Processes”, 13 November 2013, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
-- J. P. Crutchfield, review of “Learning in embodied action-perception loops through exploration”, 30 October 2013, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
-- Dowman Varn, “MultiAgent Dynamical Systems”, 23 October 2013, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
-- John Mahoney, “Crypticity in Stochastic Processes: Parts I and II”, 4 and 11 September 2013, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
-- Brenden Roberts, “Study of Time Reversal in Physical Processes”, 14 August 2013, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.

Number of Presentations: 29.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received	Paper

TOTAL:

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received	Paper

TOTAL:

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

Received		Paper
07/10/2015	11.00	Sarah Marzen, James P. Crutchfield. Informational and Causal Architecture of Discrete-Time Renewal Processes, Entropy (03 2015)
07/10/2015	12.00	James P. Crutchfield, Ryan G. James, Sarah Marzen, Dowman P. Varn. Understanding and Designing Complex Systems: Response to `A framework for optimal high-level descriptions in science and engineering—preliminary report', arXiv.org:1412.8520 (12 2014)
08/26/2014	9.00	Nicholas F. Travers, James P. Crutchfield. Equivalence of History and Generator Epsilon-Machines, arxiv.org:1111.4500 [math.PR] (11 2011)
08/26/2014	10.00	P. M. Riechers, D. P. Varn, J. P. Crutchfield. Pairwise Correlations in Layered Close-Packed Structures, arxiv.org:1407.7159 [cond-mat] (07 2014)
08/26/2014	8.00	James Crutchfield, Christopher Ellison. The Past and the Future in the Present, arXiv:1012.0356 (12 2010)
08/30/2013	1.00	Nicholas F. Travers, James P. Crutchfield. Infinite Excess Entropy Processes with Countable-State Generators, Electronic Journal of Probability (08 2013)
08/30/2013	2.00	Benjamin D. Johnson, James P. Crutchfield, Christopher J. Ellison, Carl S. McTague. Enumerating Finitary Processes, Theoretical Computer Science (12 2012)

TOTAL: **7**

Number of Manuscripts:

Books

Received	Book

TOTAL:

Received Book Chapter

TOTAL:

Patents Submitted

Patents Awarded

Awards

-- PI Crutchfield's publication “Regularities Unseen, Randomness Observed” was selected by American Institute of Physics journal Chaos—Journal of Nonlinear Science as one of the top 10 publications in the journal's 25 year history.
-- Kevin Taylor, “Extending Reinforcement Learning Methods with a Simple Robot”, Undergraduate Senior Thesis (June 2013), awarded Highest Honors by Physics Department.
-- PI Crutchfield's publication “Chaos Forgets and Remembers” highlighted in Nature: P.-M. Binder and R. M. Pipes, "How chaos forgets and remembers", Nature 510 (June 2015) 343-344.
-- PI Crutchfield's publication “Signatures of Infinity” was selected as a Physical Review E Editor's Choice: Link.
-- Ryan G. James: Assumed post-doctoral research position in Computer Science Department, University of Colorado, Boulder (September 2013-August 2014).
-- Ryan G. James: Assumed post-doctoral research position in Complexity Sciences Center, Physics, University of California, Davis (September 2014-present).
-- Nicholas F. Travers: Assumed post-doctoral research position in Mathematics Department, Technion University, Israel (September 2013-August 2015).
-- Nicholas F. Travers: Assistant Professor, Mathematics Department, University of Indiana, Bloomington. (September 2015-).
-- Korana Burke: UC Davis Chancellor Postdoctoral Fellowship, Complexity Sciences Center, Physics Department (2012-2014).
-- Sarah Marzen (Physics, UCB) UC Berkeley Chancellor’s Graduate Fellow (Fall 2013-present).
-- Sarah Marzen (Physics, UCB) National Science Foundation Graduate Fellow (Fall 2013-present).

Graduate Students

NAME	PERCENT_SUPPORTED	Discipline
Nix Barnett	0.10	
Sarah Marzen	0.00	
Paul Riechers	0.10	
Alec Boyd	0.10	
Pooneh Ara	0.10	
Greg Wimsatt	0.20	
FTE Equivalent:	**0.60**	
Total Number:	**6**	

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
Korana Burke	0.00
Ryan James	1.00
FTE Equivalent:	**1.00**
Total Number:	**2**

Names of Faculty Supported

NAME	PERCENT_SUPPORTED	National Academy Member
James P. Crutchfield	0.10	
FTE Equivalent:	**0.10**	
Total Number:	**1**	

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED	Discipline
Alexandra Nilles	0.00	Physics
Greg Wimsatt	0.20	Physics
Kevin Taylor	0.00	Physics
Brenden Roberts	0.00	Physics
FTE Equivalent:	**0.20**	
Total Number:	**4**	

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period: 4.00

The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:...... 4.00

The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:...... 4.00

Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):...... 3.00

Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:...... 0.00

The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense 1.00

The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields:...... 3.00

Names of Personnel receiving masters degrees

NAME

Total Number:

Names of personnel receiving PHDs

NAME

Ryan James

Nicholas Travers

Total Number: **2**

Names of other research staff

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Sub Contractors (DD882)

Inventions (DD882)

Scientific Progress

Final Progress Report - Contract # W911NF-13-1-0288 (Linguistics)
Proposal Number 62169MA
(Reporting Period: September 2012 – January 2015)

Structural Complexity in Linguistic Systems

James P. Crutchfield, Principal Investigator
Complexity Sciences Center, Department of Physics
University of California, Davis, California 95616

Statement of the Problem Studied

We adapted computational mechanics to determine the role of structural complexity in linguistic systems, including language systems that adapt and evolve. The latter pose substantial challenges to any conventional communication-channel view of language. Our approach promised to overcome these difficulties, since it gives a clear and constructive view of structure in memoryful stochastic processes. In principle, this allows one to track how intrinsic structure evolves over time and to explore the tendency of language systems to develop in ways that balance structure (expressed in vocabulary and syntax) and entropy (the need to convey novelty and express meaning unpredicted by the receiver).

Given the project's short duration and episodic funding, we focused on developing several measures of intrinsic structural complexity applicable to probe structure in linguistic models and language corpora. Using these, we analyzed the role of stored, hidden, acausal, and nonpredictive informations. We analyzed how a receiver comes to understand an information source—via what we call synchronization. Finally, we extended computational mechanics techniques to address the same questions in infinite-memory communication systems. However, the project ended just as the tools were refined and could be applied to actual language corpora.

ApproachThe scientific approach consists of
-- Stochastic Process Theory: Hidden Markov models and stochastic grammars give a detailed view of the mechanisms that guide and constrain behavior.
Information Theory: Provides the basic metrics that our approach extends.
-- Computational Mechanics: Allows for the exact characterization of the intrinsic structure in complex systems.
-- Quantitative Linguistics: Characterizes the basic statistical properties of language systems that we are modeling and predicting.
-- Statistical Mechanics: Provides renormalization group and regularization methods to analyze long-range correlations and infinite-memory processes.

Several technical approaches are used to identify relevant structures in linguistic systems—such as, determining mutual information, scaling, effective states, ε-machines, and measures of intrinsic computation. The approaches include:
-- Analytical methods to solve in closed-form for the simplest prototype language system behaviors and their emergent properties.
-- Symbolic computing to extend the range of closed-form analysis to complex languages.
-- Large-scale simulation using high-performance computing (HPC) to go beyond the limitations of analytical and empirical methods.
-- Unsupervised learning for structural inference, applied to language corpora.

Relevance to Army

Success in the overall research program will enhance a number of capabilities, including:
-- Fundamental science of structural complexity: This is key to understanding organization in existing language systems of interest to Army. It includes identifying the range of possible distinct kinds of information, including predictive and semantic informations.
-- Fundamental science of acausality in intentional communication: This is key to properly identifying the distinct kinds of temporally nonlocal and retrodictive information that human communication often embodies.
-- Fundamental science of synchronization in communication: This is key to properly identifying the mechanisms by which information is revealed in a communication system.
-- Fundamental science of infinite memory in communication: This is key to properly analyzing natural languages, whose information content diverges as a function of message length.

Summary of the Most Important Results

The project's research strategy was to synthesize the scientific areas via the following tasks.
-- Task A: Structural Complexity Adapt computational mechanics' analysis of complex systems to finite-memory linguistic

processes, focusing on the informational properties hidden Markov models (HMMs) and transducers (HMTs).
-- Task B: Generalized HMMs To circumvent the limitations of finite-state HMMs, extend com- putational mechanics to Generalized HMMs (GHMMs).
-- Task C: Acausal Information Develop measures of acausality in finite-memory linguistic pro- cesses by adapting causal irreversibility, gauge information, and oracular information.
-- Task D: Source-Observer Synchronization Analyze the mechanisms by which an observer comes to understand the messages embedded in language, by estimating the synchronization rates and transient information.
-- Task E: Infinitary Linguistic Processes Develop countably infinite-state HMMs for modeling infinite-range correlations and infinite-memory in linguistic systems. Use these to predict their scaling properties.

The original proposal was solicited as exploratory, with all parties involved acknowledging that the PI had no publication record in quantitative or computational linguistics. (This was seen as a strength; a way to inject novelty. One, not surprisingly, under-appreciated by the original reviewers.) As such, the tasks above were intentionally ambitious and rather broad. Nonetheless, there was a central and unifying focus: Does computational mechanics' approach to intrinsic computation bear on our understanding of structure (pattern, correlation, memory, semantics, ...) observed in linguistic systems and process?

On this score we believe the project very successfully provided a resounding “yes” answer. This positive answer means we identified a several insightful ways in which (statistical and computational) linguistics can reframe the basic questions it pursues in terms of computational mechanics and so overcome several of its currently acknowledged roadblocks. These slightly self-critical remarks aside, it should be emphasized that the project nonetheless was materially and scientifically very productive, as we recount below. Note that two of the publications had immediate and important impact: One highlighted in a Nature News & Views column and the other selected as an “Editor's Choice” by Physical Review E for promotion in American Physical Society publications.

The success and accomplishments must be understood in the context of the project's boundary conditions, though: the original exploratory project was to be 18 months in duration. At the very end of that period, it was generously extended for another 12 months. One consequence was that this intermittent funding strategy made it difficult to plan ahead, to tackle large difficult problems, and to recruit graduate students and train them. The net result was that all five tasks, above, could not be pursued in depth. In particular, Task B on Generalized Hidden Markov Models still remains untouched. The success described above, though, means that it and the other tasks deserve a major future effort to develop. In short, the PI was able to bridge his perspective to several (statistical and computational) linguistics challenges and came away convinced that future efforts, if properly supported, would not only be successful, but lead to significant new insights into the structure and functioning of human language.

The Publications (P#), manuscripts Under Review (UR#), and Manuscript in Preparation (IP#) document the task accomplishments and scientific progress. Their publication data and associated reprints and preprints have been uploaded to https://extranet.aro.army.mil. They are listed again below in the Bibliography to provide references for the narrative summary of the accomplishments.

-- Information Measures: We successfully extended existing information measures to joint and input-output processes using optimal minimal transducers (ε-transducers), opening up a much wider range of linguistics applications. (IP5)
Nonergodicity: We successfully extended the computational mechanics of ε-machines to nonergodic processes, opening up a much wider range of applications to linguistic processes. (IP4)
-- Chaos Forgets and Remembers: This project's most historically surprising result showed how chaotic systems create, destroy, and store information. For over a half century chaotic dynamical systems were characterized by a single “system invariant”, the rate at which they generated information, called the Kolmogorov-Sinai entropy rate. We showed that this rate consists of two radically distinct kinds of information processing: (i) a component that is created and dissipated and (ii) a component that is created but actively stored by the system. Preliminary quantitative studies indicated that these play a key role in written English. (P2)
-- Nature Highlight: The foundational and insightful aspect of these results were acknowledged in a Nature News and Views commentary on (P2): P. M. Binder and R. M. Pipes, “How Chaos Forgets and Remembers”, Nature 510 (19 June 2014) 343-344.
-- Information in Continuous Processes: Computational mechanics has been successfully extended to processes over continuous-time and continuous-state. Using these results, we demonstrated how apparent complexity is not the structural complexity of the underlying generative process. We critically applied the results to power laws often observed in nominally “complex”, “self-organizing” processes such as human language (e.g., Zipf's, Giraud's, and Hilberg's Laws) (P5 and IP1). In concert with our constructions of countable-state generators with infinite excess entropy, we now have a substantially more structural and mechanistic view of how infinite complexity arises (P4).
-- The Elusive Present: We developed methods to analytically calculate the amount of hidden correlation between the past and future, relating this to a process's Markov order. For the first time, this provided closed-form expressions in terms of a process's ε–machine spectral decomposition. (IP2)
-- Mixed-States as Efficient Representations: A key part of these advances was our extending the underlying representation, the

so-called mixed states, beyond simple discrete-alphabet and symbolic-dynamic processes to vector-valued and high-dimensional dynamical systems. One practical consequence is that the computational efficiencies we now enjoy for analyzing the intrinsic computation in low-complexity processes applies to arbitrarily highly complex linguistic processes. (IP3)
-- Horizons of Prediction and Retrodiction: We introduced a theory and algorithms for determining the prediction and contingency horizons—past and future window lengths—required for successful structural inference and for properly determining the range of syntactic and semantic dependence in written texts. (IP6)
-- Structural Inference: We developed a complete Bayesian inference method to infer the structure of hidden processes from time series data. This goes substantially beyond statistical parameter estimation methods typical of modern machine learning approaches to estimation and modeling in that it makes minimal assumptions about the relevant model class. The latter, it should be emphasized, is the key issue when attempting to discover structure and not merely verify a structural modeling assumption. That is, our methods determine the appropriate model class from data directly. (P3, UR1, and IP7)
-- Signatures of Infinity: We delineated a hierarchy of informational divergences, each of which distinguishes distinct structural classes (P6). This hierarchy is critical to understanding the organization of linguistic systems.
-- Editor's Choice: (P6) was selected as a Physical Review E Editor's Choice: Link.
-- Synchronization: A key property of communication via natural language is how a listener comes to know the speaker's semantic intention. We modeled this in terms of controlling nonlinear dynamical systems, in which the problem reduces to how the driven system becomes correlated, or not, to the imposed driving signal. To understand this and frame it information-theoretically, we analyzed in some detail and then provided exact calculational algorithms to determine the amount and kind of information and time required for an observer to synchronize to a structured process. (P1)
-- We responded to a recently promulgated proposal for structural inference that was, in fact, highly derivative of earlier work. (UR2)

♀Bibliography

Resulting Journal Publications During Reporting Period
P1. R. G. James, J. R. Mahoney, C. J. Ellison, and J. P. Crutchfield, "Many Roads to Synchrony: Natural Time Scales and Their Algorithms", Physical Review E 89 (2014) 042135.
P2. R. G. James, K. Burke, and J. P. Crutchfield, "Chaos Forgets and Remembers: Measuring Information Creation, Destruction, and Storage", Physics Letters A 378 (2014) 2124-2127.
P3. C. C. Strelioff and J. P. Crutchfield, "Bayesian Structural Inference for Hidden Processes", Physical Review E 89 (2014) 042119.
P4. N. Travers and J. P. Crutchfield, "Infinite Excess Entropy Processes with Countable-State Generators", Entropy 16 (2014) 1396-1413.
P5. S. Marzen and J. P. Crutchfield, "Informational and Causal Architecture of Discrete-Time Renewal Processes", Entropy (2015) to appear. Santa Fe Institute Working Paper 14-08-032. arXiv:1408.6876 [cond-mat].
P6. J. P. Crutchfield and S. Marzen, "Signatures of Infinity: Nonergodicity and Resource Scaling in Prediction, Complexity, and Learning", Physical Review E 91 (2015) 050106(R).

Manuscripts under Review and Technical Reports
UR1. B. D. Johnson, J. P. Crutchfield, C. J. Ellison, C. S. McTague. "Enumerating Finitary Processes", Journal of Experimental Mathematics (2014) submitted. arxiv.org:1011.0036 [cs.FL]
UR2. J. P. Crutchfield, R. G. James, S. Marzen and D. P. Varn, "Understanding and Designing Complex Systems: Response to `A framework for optimal high-level descriptions in science and engineering—preliminary report'" (2015). Santa Fe Institute Working Paper 14-12-048. arXiv:1412.8520 [cond-mat].

Manuscripts in Preparation
IP1. S. Marzen and J. P. Crutchfield, "When is Complicatedness Complexity? Long-Range Memory in Stationary Renewal Processes", arxiv.org and Physical Review E, submission expected August 2015.
IP2. P. Mohammadiara, P. Riechers, and J. P. Crutchfield, "The Convergence to Markov Order in Hidden Processes", arxiv.org and Physica D, submission expected August 2015.
IP3. C. J. Ellison and J. P. Crutchfield, "On States of States of Uncertainty: Observer-Process Synchronization", arxiv.org and Journal of Statistical Physics, submission expected September 2015.
IP4. J. P. Crutchfield and C. J. Ellison, "Way More Than the Sum of Their Parts: From Statistical to Structural Mixtures", arxiv.org and Physical Review Letters, submission expected October 2015.
IP5. N. Barnett and J. P. Crutchfield, "Computational Mechanics of Input-Output Processes:
Shannon Information Measures and Decompositions", arxiv.org and Physica D, submission expected September 2015.
IP6. C. J. Ellison and J. P. Crutchfield, "Prediction and Contingency Horizons for Causal Discovery", arxiv.org and Physical Review E, submission expected November 2015.
IP7. C. C. Strelioff and J. P. Crutchfield, "Bayesian Structural Inference for ε-Machines", arxiv.org and Physical Review E, submission expected November 2015.
♀Appendices

The bulk of the following information has been uploaded to the webforms at https://extranet.aro.army.mil. However, the following provides more detail.

Collaborations and Technology Transfer
Due to the large number of new results, we have increased our efforts to develop the open source library “Computational Mechanics in Python” (CMPy) that make available the bulk of the project’s constructive mathematical results and methods.

Ph.D. Degrees Earned During Reporting Period
-- Ryan G. James (Physics) completed Ph.D. August 2013.
-- Nicholas F. Travers (Mathematics) completed Ph.D. June 2013.

Postdoctoral Researchers Involved During Reporting Period
-- Korana Burke: (Physics) September 2012-2014.
-- Ryan G. James (Physics) completed Ph.D. August 2013.

Graduate Students Involved During Reporting Period
-- Jason (Nix) Barnett (Mathematics) Ph.D. Fall 2015, expected.
-- Sarah Marzen (Physics, UCB) Ph.D. Fall 2016, expected.
-- Paul M. Riechers (Physics) Ph.D. Winter 2016, expected.
-- Alec Boyd (Physics) Ph.D. Spring 2016, expected.
-- Pooneh Mohammediari (Physics), M.S. Fall 2015, expected.
-- Greg Wimsatt (Physics) Ph.D. Spring 2016, expected.

Undergraduate Students Involved During Reporting Period
-- Alexandra Nilles (NSF REU undergraduate intern, Physics, 2014).
-- Greg Wimsatt (Undergraduate intern, Physics, 2013).
-- Kevin Taylor (Undergraduate intern, Physics, 2012-2013).
-- Brenden Roberts (NSF REU undergraduate intern, Physics, 2013).

Awards, Honors and Appointments
-- PI Crutchfield's publication “Regularities Unseen, Randomness Observed” was selected by American Institute of Physics journal Chaos—Journal of Nonlinear Science as one of the top 10 publications in the journal's 25 year history.
-- Kevin Taylor, “Extending Reinforcement Learning Methods with a Simple Robot”, Undergraduate Senior Thesis (June 2013), awarded Highest Honors by Physics Department.
-- PI Crutchfield's publication “Chaos Forgets and Remembers” highlighted in Nature: P.-M. Binder and R. M. Pipes, "How chaos forgets and remembers", Nature 510 (June 2015) 343-344.
-- PI Crutchfield's publication “Signatures of Infinity” was selected as a Physical Review E Editor's Choice: Link.
-- Ryan G. James: Assumed post-doctoral research position in Computer Science Department, University of Colorado, Boulder (September 2013-August 2014).
-- Ryan G. James: Assumed post-doctoral research position in Complexity Sciences Center, Physics, University of California, Davis (September 2014-present).
-- Nicholas F. Travers: Assumed post-doctoral research position in Mathematics Department, Technion University, Israel (September 2013-August 2015).
-- Nicholas F. Travers: Assistant Professor, Mathematics Department, University of Indiana, Bloomington. (September 2015-).
-- Korana Burke: UC Davis Chancellor Postdoctoral Fellowship, Complexity Sciences Center, Physics Department (2012-2014).
-- Sarah Marzen (Physics, UCB) UC Berkeley Chancellor’s Graduate Fellow (Fall 2013-present).
-- Sarah Marzen (Physics, UCB) National Science Foundation Graduate Fellow (Fall 2013-present).

Additional information on programs, students, and faculty of the Complexity Sciences Center is available online: http://csc.ucdavis.edu/.

Talks and Presentations: For those by group members see Presentations section filed at https://extranet.aro.army.mil.
-- Sarah Marzen (UCB) “Information Theoretic Methods of Time Series Analysis”, March 2014, Machine Learning seminar, Computer Science Department, University of Hawaii, Manoa.
-- Sarah Marzen (UCB), “Causal Structure of Neural Spikes Trains”, April 2014, Redwood Center for Theoretical Neuroscience, UC Berkeley.
-- Sarah Marzen, “Causal Structure of Neural Spikes Trains”, April 2014, Joint Redwood-Ganguli-Theunissen Lab Meeting, UC Berkeley.
-- Sarah Marzen, Mike DeWeese, Jim Crutchfield, “Statistical Complexity of Neural Spike Trains”, Poster presentation, March 2014, Computational and Systems Neuroscience (Cosyne 2014), Salt Lake City, Utah.
-- Training: JPC taught two graduate courses Winter and Spring 2014 on the physics of information and the physics of computation, respectively.

-- PI Crutchfield gave John von Neumann Complexity and Computation Public Lecture “Demonology: The Curious Role of Intelligence in Physics & Biology”, Wisconsin Institutes for Discovery. (6 November 2013.)
-- Nix Barnett, “ε-Transducers: Parts I, II, and III”, 27 May and 9 and 16 July 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
-- Alexandra Nilles, “Applying Predicted Information Gain to Physically Embodied Agents”, 13 August 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis; 14 August 2014, REU Presentations, Physics, UC Davis.
-- Aleksander Zujev “Coupled Systems”, 25 June 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
-- Benny Brown, review of D. Little and F. Sommer “Learning and exploration in action-perception loops”, 21 May and 18 June 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
-- Fritz Sommer (Redwood Ctr, UCB), “Predictive Information Gain for Adaptive Learning”, 11 June 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
-- Sarah Marzen, “Information Anatomy of Continuous Stochastic Processes”, 16 April 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
-- Ryan James (U Colorado, Boulder), “Intersection Information”, 2 April 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
-- Korana Burke and Greg Wimsatt, review of Physical Review Letter “Measurement of Stochastic Entropy Product”, 12 March 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
-- Katie Amrine (Dept Viticulture and Enology) on “Measuring ‘Functional’ Information in transfer-RNAs to detect and resolve inconsistencies in deep roots of the Tree of Life”, 12 February 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
-- Pooneh Mohammadiara on “Elusive Information”, 5 February 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
-- Sarah Marzen on “Structural Complexity of Stochastic Renewal Processes”, 29 January 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
-- J. P. Crutchfield, “Information in Infinitary Processes”, 3 December 2013, Workshop on “Information in Infinitary Processes”, 3-4 December 2013, Complexity Sciences Center, UC Davis.
-- Dave Feldman (College of the Atlantic), “2D Information Theory”, 3 December 2013, Workshop on “Information in Infinitary Processes”, 3-4 December 2013, Complexity Sciences Center, UC Davis.
-- Chris Ellison (WID, Wisconsin), “Countably Infinite Processes”, 3 December 2013, Workshop on “Information in Infinitary Processes”, 3-4 December 2013, Complexity Sciences Center, UC Davis.
-- Sarah Marzen (UCB), “Continuous-Time Processes”, 3 December 2013, Workshop on “Information in Infinitary Processes”, 3-4 December 2013, Complexity Sciences Center, UC Davis.
-- Chris Strelioff, “Inferring Infinite-State Processes”, 3 December 2013, Workshop on “Information in Infinitary Processes”, 3-4 December 2013, Complexity Sciences Center, UC Davis.
-- Paul Riechers, “Spectral Decomposition”, 3 December 2013, Workshop on “Information in Infinitary Processes”, 3-4 December 2013, Complexity Sciences Center, UC Davis.
-- Korana Burke, reviewed Physical Review E “Information Content in Turbulence”, 20 November 2013, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
-- J. P. Crutchfield, “Information in Infinitary Processes”, 13 November 2013, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
-- J. P. Crutchfield, review of “Learning in embodied action-perception loops through exploration”, 30 October 2013, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
-- Dowman Varn, “MultiAgent Dynamical Systems”, 23 October 2013, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
-- John Mahoney, “Crypticity in Stochastic Processes: Parts I and II”, 4 and 11 September 2013, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
-- Brenden Roberts, “Study of Time Reversal in Physical Processes”, 14 August 2013, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.

Meetings
-- Postdoc Korana Burke organized a symposium on computational mechanics at the Annual Meeting of the Society for Industrial and Applied Mathematics, Chicago, Illionis, (July 2014).
-- UCD group (Crutchfield, Boyd, Riechers) weekly visits to Redwood Center for Theoretical Neuroscience, UC Berkeley. Crutchfield organizer, CSC Workshop on “Information in Infinitary Processes”, 3-4 December 2013, Complexity Sciences Center, UC Davis. See: http://csc.ucdavis.edu/Infinitary_Meeting.html.

Technology Transfer

Due to the large number of new results, we have increased our efforts to develop the open source library “Computational Mechanics in Python” (CMPy) that make available the bulk of the project’s constructive mathematical results and methods.

Final Progress Report - Contract # W911NF-13-1-0288 (Linguistics)

Proposal Number 62169MA

(Reporting Period: September 2012 – January 2015)

Structural Complexity in Linguistic Systems

James P. Crutchfield, Principal Investigator
Complexity Sciences Center, Department of Physics
University of California, Davis, California 95616

Statement of the Problem Studied

We adapted computational mechanics to determine the role of structural complexity in linguistic systems, including language systems that adapt and evolve. The latter pose substantial challenges to any conventional communication-channel view of language. Our approach promised to overcome these difficulties, since it gives a clear and constructive view of structure in memoryful stochastic processes. In principle, this allows one to track how intrinsic structure evolves over time and to explore the tendency of language systems to develop in ways that balance structure (expressed in vocabulary and syntax) and entropy (the need to convey novelty and express meaning unpredicted by the receiver).

Given the project's short duration and episodic funding, we focused on developing several measures of intrinsic structural complexity applicable to probe structure in linguistic models and language corpora. Using these, we analyzed the role of stored, hidden, acausal, and nonpredictive informations. We analyzed how a receiver comes to understand an information source—via what we call synchronization. Finally, we extended computational mechanics techniques to address the same questions in infinite-memory communication systems. However, the project ended just as the tools were refined and could be applied to actual language corpora.

Approach

The scientific approach consists of

1. Stochastic Process Theory: Hidden Markov models and stochastic grammars give a detailed view of the mechanisms that guide and constrain behavior.
2. Information Theory: Provides the basic metrics that our approach extends.
3. Computational Mechanics: Allows for the exact characterization of the intrinsic structure in complex systems.
4. Quantitative Linguistics: Characterizes the basic statistical properties of language systems that we are modeling and predicting.
5. Statistical Mechanics: Provides renormalization group and regularization methods to analyze long-range correlations and infinite-memory processes.

Several technical approaches are used to identify relevant structures in linguistic systems—such as, determining mutual information, scaling, effective states, ε-machines, and measures of intrinsic computation. The approaches include:

1. Analytical methods to solve in closed-form for the simplest prototype language system behaviors and their emergent properties.
2. Symbolic computing to extend the range of closed-form analysis to complex languages.
3. Large-scale simulation using high-performance computing (HPC) to go beyond the limitations of analytical and empirical methods.
4. Unsupervised learning for structural inference, applied to language corpora.

Relevance to Army

Success in the overall research program will enhance a number of capabilities, including:

1. *Fundamental science of structural complexity*: This is key to understanding organization in existing language systems of interest to Army. It includes identifying the range of possible distinct kinds of information, including predictive and semantic informations.
2. *Fundamental science of acausality in intentional communication*: This is key to properly identifying the distinct kinds of temporally nonlocal and retrodictive information that human communication often embodies.
3. *Fundamental science of synchronization in communication*: This is key to properly identifying the mechanisms by which information is revealed in a communication system.
4. *Fundamental science of infinite memory in communication*: This is key to properly analyzing natural languages, whose information content diverges as a function of message length.

Summary of the Most Important Results

The project's research strategy was to synthesize the scientific areas via the following tasks.
Task A: Structural Complexity Adapt computational mechanics' analysis of complex systems to finite-memory linguistic processes, focusing on the informational properties hidden Markov models (HMMs) and transducers (HMTs).
Task B: Generalized HMMs To circumvent the limitations of finite-state HMMs, extend computational mechanics to Generalized HMMs (GHMMs).
Task C: Acausal Information Develop measures of acausality in finite-memory linguistic processes by adapting causal irreversibility, gauge information, and oracular information.
Task D: Source-Observer Synchronization Analyze the mechanisms by which an observer comes to understand the messages embedded in language, by estimating the synchronization rates and transient information.
Task E: Infinitary Linguistic Processes Develop countably infinite-state HMMs for modeling infinite-range correlations and infinite-memory in linguistic systems. Use these to predict their scaling properties.

The original proposal was solicited as exploratory, with all parties involved acknowledging that the PI had no publication record in quantitative or computational linguistics. (This was seen as a strength; a way to inject novelty. One, not surprisingly, under-appreciated by the original reviewers.) As such, the tasks above were intentionally ambitious and rather broad. Nonetheless, there was a central and unifying focus: Does computational mechanics' approach to intrinsic computation bear on our understanding of structure (pattern, correlation, memory, semantics, ...) observed in linguistic systems and process?

On this score we believe the project very successfully provided a resounding "yes" answer. This positive answer means we identified a several insightful ways in which (statistical and computational) linguistics can reframe the basic questions it pursues in terms of computational mechanics and so overcome several of its currently acknowledged roadblocks. These slightly self-critical remarks aside, it should be emphasized that the project nonetheless was materially and scientifically very productive, as we recount below. Note that two of the publications had immediate and important impact: One highlighted in a *Nature* News & Views column and the other selected as an "Editor's Choice" by *Physical Review E* for promotion in American Physical Society publications.

The success and accomplishments must be understood in the context of the project's boundary conditions, though: the original exploratory project was to be 18 months in duration. At the very end of that period, it was generously extended for another 12 months. One consequence was that this intermittent funding strategy made it difficult to plan ahead, to tackle large difficult problems, and to recruit graduate students and train them. The net result was that all five tasks, above, could not be pursued in depth. In particular, Task B on Generalized Hidden Markov Models still remains untouched. The success described above, though, means that it and the other tasks deserve a major future effort to develop. In short, the PI was able to bridge his perspective to several (statistical and computational) linguistics challenges and came away convinced that future efforts, if properly supported, would not only be successful, but lead to significant new insights into the structure and functioning of human language.

The Publications (P#), manuscripts Under Review (UR#), and Manuscript in Preparation (IP#) document the task accomplishments and scientific progress. Their publication data and associated reprints and preprints have been uploaded to https://extranet.aro.army.mil. They are listed again below in the Bibliography to provide references for the narrative summary of the accomplishments.

1. *Information Measures*: We successfully extended existing information measures to joint and input-output processes using optimal minimal transducers (ε-transducers), opening up a much wider range of linguistics applications. (IP5)
2. *Nonergodicity*: We successfully extended the computational mechanics of ε-machines to nonergodic processes, opening up a much wider range of applications to linguistic processes. (IP4)
3. *Chaos Forgets and Remembers*: This project's most historically surprising result showed how chaotic systems create, destroy, and store information. For over a half century chaotic dynamical systems were characterized by a single "system invariant", the rate at which they generated information, called the *Kolmogorov-Sinai entropy rate*. We showed

that this rate consists of two radically distinct kinds of information processing: (i) a component that is created and dissipated and (ii) a component that is created but actively stored by the system. Preliminary quantitative studies indicated that these play a key role in written English. (P2)

4. *Nature Highlight*: The foundational and insightful aspect of these results were acknowledged in a *Nature News and Views commentary* on (P2): P. M. Binder and R. M. Pipes, "How Chaos Forgets and Remembers", Nature **510** (19 June 2014) 343-344.
5. *Information in Continuous Processes*: Computational mechanics has been successfully extended to processes over continuous-time and continuous-state. Using these results, we demonstrated how apparent complexity is not the structural complexity of the underlying generative process. We critically applied the results to power laws often observed in nominally "complex", "self-organizing" processes such as human language (e.g., Zipf's, Giraud's, and Hilberg's Laws) (P5 and IP1). In concert with our constructions of countable-state generators with infinite excess entropy, we now have a substantially more structural and mechanistic view of how infinite complexity arises (P4).
6. *The Elusive Present*: We developed methods to analytically calculate the amount of hidden correlation between the past and future, relating this to a process's Markov order. For the first time, this provided closed-form expressions in terms of a process's ε–machine spectral decomposition. (IP2)
7. *Mixed-States as Efficient Representations*: A key part of these advances was our extending the underlying representation, the so-called *mixed states*, beyond simple discrete-alphabet and symbolic-dynamic processes to vector-valued and high-dimensional dynamical systems. One practical consequence is that the computational efficiencies we now enjoy for analyzing the intrinsic computation in low-complexity processes applies to arbitrarily highly complex linguistic processes. (IP3)
8. *Horizons of Prediction and Retrodiction*: We introduced a theory and algorithms for determining the prediction and contingency horizons—past and future window lengths—required for successful structural inference and for properly determining the range of syntactic and semantic dependence in written texts. (IP6)
9. *Structural Inference*: We developed a complete Bayesian inference method to infer the structure of hidden processes from time series data. This goes substantially beyond statistical parameter estimation methods typical of modern machine learning approaches to estimation and modeling in that it makes minimal assumptions about the relevant model class. The latter, it should be emphasized, is the key issue when attempting to discover structure and not merely verify a structural modeling assumption. That is, our methods determine the appropriate model class from data directly. (P3, UR1, and IP7)
10. *Signatures of Infinity*: We delineated a hierarchy of informational divergences, each of which distinguishes distinct structural classes (P6). This hierarchy is critical to understanding the organization of linguistic systems.
11. *Editor's Choice*: (P6) was selected as a Physical Review E Editor's Choice: Link.
12. *Synchronization*: A key property of communication via natural language is how a listener comes to know the speaker's semantic intention. We modeled this in terms of controlling nonlinear dynamical systems, in which the problem reduces to how the driven system

becomes correlated, or not, to the imposed driving signal. To understand this and frame it information-theoretically, we analyzed in some detail and then provided exact calculational algorithms to determine the amount and kind of information and time required for an observer to synchronize to a structured process. (P1)

13. We responded to a recently promulgated proposal for structural inference that was, in fact, highly derivative of earlier work. (UR2)

Bibliography

Resulting Journal Publications During Reporting Period
P1. R. G. James, J. R. Mahoney, C. J. Ellison, and J. P. Crutchfield, “Many Roads to Synchrony: Natural Time Scales and Their Algorithms”, Physical Review E **89** (2014) 042135.
P2. R. G. James, K. Burke, and J. P. Crutchfield, “Chaos Forgets and Remembers: Measuring Information Creation, Destruction, and Storage”, Physics Letters A 37**8** (2014) 2124-2127.
P3. C. C. Strelioff and J. P. Crutchfield, “Bayesian Structural Inference for Hidden Processes”, Physical Review E **89** (2014) 042119.
P4. N. Travers and J. P. Crutchfield, “Infinite Excess Entropy Processes with Countable-State Generators”, Entropy **16** (2014) 1396-1413.
P5. S. Marzen and J. P. Crutchfield, “Informational and Causal Architecture of Discrete-Time Renewal Processes”, Entropy (2015) to appear. Santa Fe Institute Working Paper 14-08-032. arXiv:1408.6876 [cond-mat].
P6. J. P. Crutchfield and S. Marzen, “Signatures of Infinity: Nonergodicity and Resource Scaling in Prediction, Complexity, and Learning”, Physical Review E 91 (2015) 050106(R).

Manuscripts under Review and Technical Reports
UR1. B. D. Johnson, J. P. Crutchfield, C. J. Ellison, C. S. McTague. “Enumerating Finitary Processes”, Journal of Experimental Mathematics (2014) submitted. arxiv.org:1011.0036 [cs.FL]
UR2. J. P. Crutchfield, R. G. James, S. Marzen and D. P. Varn, “Understanding and Designing Complex Systems: Response to `A framework for optimal high-level descriptions in science and engineering—preliminary report'” (2015). Santa Fe Institute Working Paper 14-12-048. arXiv: 1412.8520 [cond-mat].

Manuscripts in Preparation
IP1. S. Marzen and J. P. Crutchfield, “When is Complicatedness Complexity? Long-Range Memory in Stationary Renewal Processes”, arxiv.org and Physical Review E, submission expected August 2015.
IP2. P. Mohammadiara, P. Riechers, and J. P. Crutchfield, “The Convergence to Markov Order in Hidden Processes”, arxiv.org and Physica D, submission expected August 2015.
IP3. C. J. Ellison and J. P. Crutchfield, “On States of States of Uncertainty: Observer-Process Synchronization”, arxiv.org and Journal of Statistical Physics, submission expected September 2015.
IP4. J. P. Crutchfield and C. J. Ellison, “Way More Than the Sum of Their Parts: From Statistical to Structural Mixtures”, arxiv.org and Physical Review Letters, submission expected October 2015.
IP5. N. Barnett and J. P. Crutchfield, “Computational Mechanics of Input-Output Processes: Shannon Information Measures and Decompositions”, arxiv.org and Physica D, submission expected September 2015.
IP6. C. J. Ellison and J. P. Crutchfield, “Prediction and Contingency Horizons for Causal Discovery”, arxiv.org and Physical Review E, submission expected November 2015.

IP7. C. C. Strelioff and J. P. Crutchfield, “Bayesian Structural Inference for ε-Machines”, arxiv.org and Physical Review E, submission expected November 2015.

Appendices

The bulk of the following information has been uploaded to the webforms at https://extranet.aro.army.mil. However, the following provides more detail.

Collaborations and Technology Transfer

- Due to the large number of new results, we have increased our efforts to develop the open source library "Computational Mechanics in Python" (CMPy) that make available the bulk of the project's constructive mathematical results and methods.

Ph.D. Degrees Earned During Reporting Period

- Ryan G. James (Physics) completed Ph.D. August 2013.
- Nicholas F. Travers (Mathematics) completed Ph.D. June 2013.

Postdoctoral Researchers Involved During Reporting Period

- Korana Burke: (Physics) September 2012-2014.
- Ryan G. James (Physics) completed Ph.D. August 2013.

Graduate Students Involved During Reporting Period

- Jason (Nix) Barnett (Mathematics) Ph.D. Fall 2015, expected.
- Sarah Marzen (Physics, UCB) Ph.D. Fall 2016, expected.
- Paul M. Riechers (Physics) Ph.D. Winter 2016, expected.
- Alec Boyd (Physics) Ph.D. Spring 2016, expected.
- Pooneh Mohammediari (Physics), M.S. Fall 2015, expected.
- Greg Wimsatt (Physics) Ph.D. Spring 2016, expected.

Undergraduate Students Involved During Reporting Period

- Alexandra Nilles (NSF REU undergraduate intern, Physics, 2014).
- Greg Wimsatt (Undergraduate intern, Physics, 2013).
- Kevin Taylor (Undergraduate intern, Physics, 2012-2013).
- Brenden Roberts (NSF REU undergraduate intern, Physics, 2013).

Awards, Honors and Appointments

- PI Crutchfield's publication "Regularities Unseen, Randomness Observed" was selected by American Institute of Physics journal *Chaos—Journal of Nonlinear Science* as one of the top 10 publications in the journal's 25 year history.
- Kevin Taylor, "Extending Reinforcement Learning Methods with a Simple Robot", Undergraduate Senior Thesis (June 2013), awarded Highest Honors by Physics Department.
- PI Crutchfield's publication "Chaos Forgets and Remembers" highlighted in *Nature*: P.-M. Binder and R. M. Pipes, "How chaos forgets and remembers", Nature **510** (June 2015) 343-344.
- PI Crutchfield's publication "Signatures of Infinity" was selected as a Physical Review E Editor's Choice: Link.

- Ryan G. James: Assumed post-doctoral research position in Computer Science Department, University of Colorado, Boulder (September 2013-August 2014).
- Ryan G. James: Assumed post-doctoral research position in Complexity Sciences Center, Physics, University of California, Davis (September 2014-present).
- Nicholas F. Travers: Assumed post-doctoral research position in Mathematics Department, Technion University, Israel (September 2013-August 2015).
- Nicholas F. Travers: Assistant Professor, Mathematics Department, University of Indiana, Bloomington. (September 2015-).
- Korana Burke: UC Davis Chancellor Postdoctoral Fellowship, Complexity Sciences Center, Physics Department (2012-2014).
- Sarah Marzen (Physics, UCB) UC Berkeley Chancellor's Graduate Fellow (Fall 2013-present).
- Sarah Marzen (Physics, UCB) National Science Foundation Graduate Fellow (Fall 2013-present).

Additional information on programs, students, and faculty of the Complexity Sciences Center is available online: http://csc.ucdavis.edu/.

Talks and Presentations: For those by group members see Presentations section filed at https://extranet.aro.army.mil.

1. Sarah Marzen (UCB) "Information Theoretic Methods of Time Series Analysis", March 2014, Machine Learning seminar, Computer Science Department, University of Hawaii, Manoa.
2. Sarah Marzen (UCB), "Causal Structure of Neural Spikes Trains", April 2014, Redwood Center for Theoretical Neuroscience, UC Berkeley.
3. Sarah Marzen, "Causal Structure of Neural Spikes Trains", April 2014, Joint Redwood-Ganguli-Theunissen Lab Meeting, UC Berkeley.
4. Sarah Marzen, Mike DeWeese, Jim Crutchfield, "Statistical Complexity of Neural Spike Trains", Poster presentation, March 2014, Computational and Systems Neuroscience (Cosyne 2014), Salt Lake City, Utah.
5. Training: JPC taught two graduate courses Winter and Spring 2014 on the physics of information and the physics of computation, respectively.
6. PI Crutchfield gave John von Neumann Complexity and Computation Public Lecture "Demonology: The Curious Role of Intelligence in Physics & Biology", Wisconsin Institutes for Discovery. (6 November 2013.)
7. Nix Barnett, "ε-Transducers: Parts I, II, and III", 27 May and 9 and 16 July 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
8. Alexandra Nilles, "Applying Predicted Information Gain to Physically Embodied Agents", 13 August 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis; 14 August 2014, REU Presentations, Physics, UC Davis.
9. Aleksander Zujev "Coupled Systems", 25 June 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.

10. Benny Brown, review of D. Little and F. Sommer "Learning and exploration in action-perception loops", 21 May and 18 June 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
11. Fritz Sommer (Redwood Ctr, UCB), "Predictive Information Gain for Adaptive Learning", 11 June 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
12. Sarah Marzen, "Information Anatomy of Continuous Stochastic Processes", 16 April 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
13. Ryan James (U Colorado, Boulder), "Intersection Information", 2 April 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
14. Korana Burke and Greg Wimsatt, review of Physical Review Letter "Measurement of Stochastic Entropy Product", 12 March 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
15. Katie Amrine (Dept Viticulture and Enology) on "Measuring 'Functional' Information in transfer-RNAs to detect and resolve inconsistencies in deep roots of the Tree of Life", 12 February 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
16. Pooneh Mohammadiara on "Elusive Information", 5 February 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
17. Sarah Marzen on "Structural Complexity of Stochastic Renewal Processes", 29 January 2014, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
18. J. P. Crutchfield, "Information in Infinitary Processes", 3 December 2013, Workshop on "Information in Infinitary Processes", 3-4 December 2013, Complexity Sciences Center, UC Davis.
19. Dave Feldman (College of the Atlantic), "2D Information Theory", 3 December 2013, Workshop on "Information in Infinitary Processes", 3-4 December 2013, Complexity Sciences Center, UC Davis.
20. Chris Ellison (WID, Wisconsin), "Countably Infinite Processes", 3 December 2013, Workshop on "Information in Infinitary Processes", 3-4 December 2013, Complexity Sciences Center, UC Davis.
21. Sarah Marzen (UCB), "Continuous-Time Processes", 3 December 2013, Workshop on "Information in Infinitary Processes", 3-4 December 2013, Complexity Sciences Center, UC Davis.
22. Chris Strelioff, "Inferring Infinite-State Processes", 3 December 2013, Workshop on "Information in Infinitary Processes", 3-4 December 2013, Complexity Sciences Center, UC Davis.
23. Paul Riechers, "Spectral Decomposition", 3 December 2013, Workshop on "Information in Infinitary Processes", 3-4 December 2013, Complexity Sciences Center, UC Davis.
24. Korana Burke, reviewed Physical Review E "Information Content in Turbulence", 20 November 2013, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
25. J. P. Crutchfield, "Information in Infinitary Processes", 13 November 2013, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.

26. J. P. Crutchfield, review of “Learning in embodied action-perception loops through exploration”, 30 October 2013, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
27. Dowman Varn, “MultiAgent Dynamical Systems”, 23 October 2013, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
28. John Mahoney, “Crypticity in Stochastic Processes: Parts I and II”, 4 and 11 September 2013, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.
29. Brenden Roberts, “Study of Time Reversal in Physical Processes”, 14 August 2013, Dynamics of Learning Group Seminar, Complexity Sciences Center, UC Davis.

Meetings

1. Postdoc Korana Burke organized a symposium on computational mechanics at the Annual Meeting of the Society for Industrial and Applied Mathematics, Chicago, Illionis, (July 2014).
2. UCD group (Crutchfield, Boyd, Riechers) weekly visits to Redwood Center for Theoretical Neuroscience, UC Berkeley.
3. Crutchfield organizer, CSC Workshop on “Information in Infinitary Processes”, 3-4 December 2013, Complexity Sciences Center, UC Davis. See: http://csc.ucdavis.edu/Infinitary_Meeting.html.