Partial information decomposition as a spatiotemporal filter

Benjamin Flecker,[1] Wesley Alford,[1] John M. Beggs,[2] Paul L. Williams,[3] and Randall D. Beer[3,4]
[1]*Department of Psychological and Brain Sciences, Indiana University, Bloomington, Indiana 47406, USA*
[2]*Department of Physics, Indiana University, Bloomington, Indiana 47406, USA*
[3]*Cognitive Science Program, Indiana University, Bloomington, Indiana 47406, USA*
[4]*School of Informatics and Computing, Indiana University, Bloomington, Indiana 47406, USA*

(Received 1 June 2011; accepted 24 August 2011; published online 30 September 2011)

Understanding the mechanisms of distributed computation in cellular automata requires techniques for characterizing the emergent structures that underlie information processing in such systems. Recently, techniques from information theory have been brought to bear on this problem. Building on this work, we utilize the new technique of partial information decomposition to show that previous information-theoretic measures can confound distinct sources of information. We then propose a new set of filters and demonstrate that they more cleanly separate out the background domains, particles, and collisions that are typically associated with information storage, transfer, and modification in cellular automata.
[doi:10.1063/1.3638449]

The mechanisms underlying information processing in spatially distributed systems have been a central interest in the field of complex systems since its very beginning. In particular, identifying the fundamental mechanisms of computation (particles) in cellular automata (CAs), a simple model of such distributed processing, has been a major focus of study. Traditionally, stationary, moving, and colliding particles are associated with information storage, transfer, and modification in CAs, respectively. Recently, techniques from information theory have been used to quantify the information content of such emergent structures. We build on this work in three ways. First, we apply the new technique of *partial information decomposition* to previous measures, demonstrating that they can confound distinct sources of information. Second, we use the insights gained from this decomposition to propose new filters for stationary configurations, moving particles, and particle collisions in cellular automata. Third, we apply our partial information filters to an elementary cellular automata in order to demonstrate that it can more clearly identify particle collisions than can previous information-theoretic measures.

I. INTRODUCTION

Information processing in the natural world is often spatially distributed across many interacting components. Cellular automata (CAs), in which a lattice of discrete states is updated according to some local rule, are a common model of such phenomena. For example, a well-studied instance of distributed computation in CAs is the density classification task, in which a CA must evolve to a state of all zeros or all ones depending upon whether its initial state has a higher proportion of zeros or ones, respectively.[1] A particularly simple CA model, known as elementary cellular automata (ECAs), consists of a 1-dimensional lattice of 2-state cells, whose update rules depend only on the two nearest neighbors of each cell and its own state. There are 256 such automata and, despite their simplicity, ECAs are capable of a wide range of behavior including universal computation.[2]

A central problem in the theory of cellular automata is how to understand the mechanisms of emergent computation in such systems. Foundational work in this area was carried out by Hanson and Crutchfield,[3] who applied the techniques of computational mechanics to CAs. By filtering out the regular background of a CA space-time diagram, computational mechanics reveals the sets of emergent structures (called particles) and the laws of interaction between these structures that underlie the distributed computation being carried out by the CA. For example, in a study of CAs evolved for the density classification task,[4] a computational mechanics analysis showed how various emergent particles carry information about local initial density from widely-separated regions and how this information is combined through particle collisions into an eventual decision about the global density of the entire initial state.

The density classification task has also been used by others to investigate CAs. For example, Marques-Pita and Rocha found that two qualitatively different rules which solve this task are actually quite similar if you perform a particular rediscription of their lookup tables (called schema redescription). This similarity was previously unknown; further, what differences that did exist between the two rules were clearly and easily explained by such a redescription as an unnecessary ambiguity in one of the rules' input states that was not present in the other.[5] Other methods of grouping CAs have been proposed, such as Wuensche's "automatic classification" method, whereby automata are classified by plotting the mean entropy against the standard deviation of the entropy for a given span of time. These plots revealed interesting results which seem to allow CAs to be grouped into ordered, complex, and chaotic classes.[6]

Local statistical properties of CAs have also been studied for purposes of filtering for structure. Filters have been constructed based on the local Lyapunov exponents and local statistical complexity, revealing autonomous structures and their interactions.[7] Interestingly, a recent class of filters is developing based on the local information-theoretic properties of CAs. An early example, the aptly-named "local information" filter, highlights symbols that are unexpected given the already observed history of symbols at that particular point. Such a filtering method successfully picks out moving particles and their interactions in CAs.[8]

Lizier *et al.* have recently introduced a new information-theoretic approach to the characterization of distributed computation in CAs.[9] By associating distinct local measures with each computational operation, they propose a set of filters that can decompose the information-processing in a CA into its essential components. In particular, they hypothesize that (1) information storage is characterized by local active information, (2) information transfer is characterized by local apparent transfer entropy, and (3) information modification is characterized by local separable information. Based on studies of ECA rules 18, 54, and 110, Lizier *et al.* also propose a framework for interpreting the results of these measures which shows that sites of information storage are associated with background domains, sites of information transfer are associated with particles, and sites of information modification are associated with collisions. However, their framework also identifies instances of information transfer and modification beyond the particles and collisions that are usually associated with these processes, and it requires an arbitrary thresholding of the measures in order to recover the standard interpretation.

In this paper, we extend Lizier *et al.*'s work by applying the new technique of partial information (PI) decomposition.[10] PI decomposition exhaustively partitions the Shannon information that a set of sources provides about a target into nonoverlapping information "atoms" consisting of various combinations of redundant, unique, and synergistic terms. After a brief review of PI decomposition, we apply it to the information measures proposed by Lizier *et al.*, demonstrating that their measures can confound distinct sources of information. Next, we propose a new PI-based ECA filter that associates 3-way redundancy with the regular domain, unique information with moving particles, and 3-way synergy with particle collisions. We then demonstrate that, although our PI filter gives the same results as Lizier *et al.* for ECAs in which their confounds play no role, it more clearly separates out background, particles, and collisions in ECAs such as Rule 184 that involve a mixture of redundant, unique, and synergistic interactions. We conclude with a discussion of the interpretational challenges involved in extending our measures to more complex ECA rules.

II. PARTIAL INFORMATION DECOMPOSITION

PI decomposition provides a general method of decomposing the information that a set of random variables $\mathbf{S} = \{S_1,\ldots,S_n\}$ provides about another random variable T.[10] The central idea underlying this method is a measure of redundancy, or the common information that two or more sets of variables share about a target. In particular, given $\mathbf{A_1},\ldots,\mathbf{A_k}$ nonempty and potentially overlapping subsets of $\mathbf{S}$, called sources, the redundant information that all sources share about T is defined by

$$I_{\min}(T;\mathbf{A}_1,\ldots,\mathbf{A}_k) = \sum_t p(t) \min_{\mathbf{A}_i} I(T=t;\mathbf{A}_i), \tag{1}$$

where

$$I(T=t;\mathbf{A}) = \sum_{\mathbf{a}} p(\mathbf{a}|t)\left[\log\frac{1}{p(t)} - \log\frac{1}{p(t|\mathbf{a})}\right], \tag{2}$$

is the specific information that $\mathbf{A}$ provides about each state t of T. Thus, redundancy is defined as the minimum information that any source provides about each state of T, averaged over all possible states. This definition captures the idea that redundancy is the information shared by all sources (the minimum that any one provides) while taking into account that sources may provide information about different states of T.

Using $I_{\min}$ and a form of inclusion-exclusion, the total information that $\mathbf{S}$ provides about T can then be decomposed into a collection of PI terms, given by the PI function Π. In the simplest case of $\mathbf{S} = \{S_1, S_2\}$, this method partitions the total information $I(T; S_1, S_2)$ into three distinct kinds of informational contributions: redundancy, unique information, and synergy. The redundancy for S_1 and S_2 is given by $\Pi(T;\{S_1\}\{S_2\}) = I_{\min}(T; S_1,S_2)$. The unique information from S_1 is given by $\Pi(T;\{S_1\}) = I(T; S_1) \quad I_{\min}(T; S_1, S_2)$, or the total information from S_1 minus the redundancy with S_2, and likewise for S_2. Thus, unique information captures the information available from S_1 but not S_2, or vice versa. Finally, the synergy for S_1 and S_2 is given by the inclusion-exclusion formula

$$\begin{aligned}\Pi(T;\{S_1,S_2\}) = I(T;S_1,S_2) - I(T;S_1) - I(T;S_2)\\ + I_{\min}(T;S_1,S_2),\end{aligned} \tag{3}$$

or more simply by

$$\Pi(T;\{S_1,S_2\}) = I(T;S_1,S_2) - I_{\max}(T;S_1,S_2), \tag{4}$$

where $I_{\max}$ is defined exactly the same as $I_{\min}$ except substituting max for min. Synergy captures the information provided by the combination of S_1 and S_2 that is not provided by either variable alone. The relationships between synergy, redundancy, and unique information can be represented using a PI diagram (Fig. 1(a)), which shows the set-theoretic breakdown of total information into PI terms.

The PI diagram for $\mathbf{S} = \{S_1, S_2, S_3\}$ is shown in Fig. 1(b), and from this, the general structure of PI decomposition can be seen. First, for each S_i, there is a region corresponding to $I(T; S_i)$. Then, for every subset $\mathbf{A}$ of $\mathbf{S}$ with two or more elements, $I(T; \mathbf{A})$ is represented as a region containing $I(T; A)$ for all $A \in \mathbf{A}$ but not coextensive with $\cup_{A\in\mathbf{A}} I(T;A)$. The difference between $I(T; \mathbf{A})$ and $\cup_{A\in\mathbf{A}} I(T;A)$ represents the synergy for $\mathbf{A}$, the information provided by the combination of all elements in $\mathbf{A}$ that is not provided by any subset. Also, regions of the diagram intersect generically, representing all

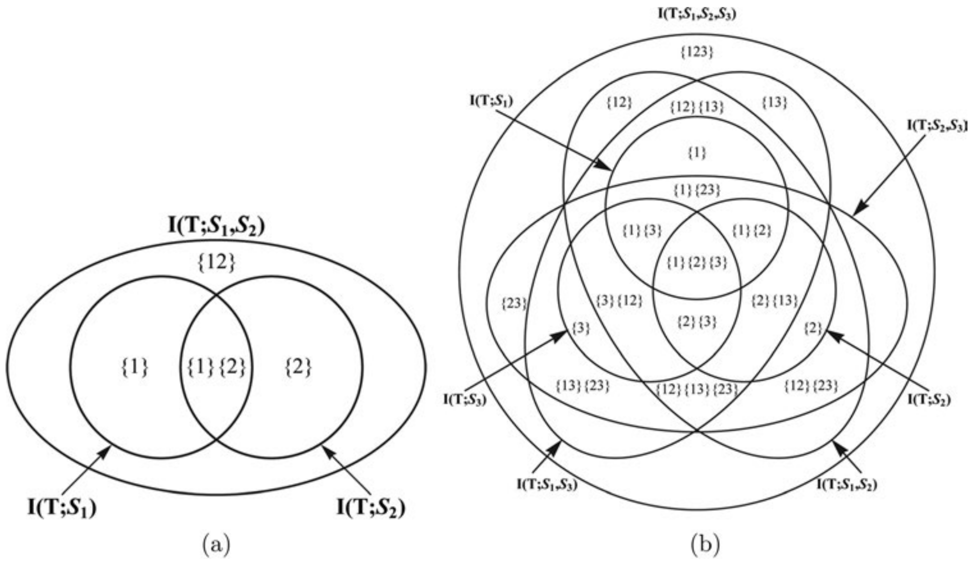


FIG. 1. PI diagrams for (a) 3 and (b) 4 variables. Each region labelled with a bracketed expression represents the PI term for a particular collection of sources, with the bracketed expression indicating the indices for each source. For instance, $\{12\}$ represents $\Pi(T;\{S_1,\ S_2\})$, $\{1\}\{2\}$ represents $\Pi(T;\{S_1\}\{S_2\})$, and so forth.

possibilities for redundancy. Thus, in general, the PI term for a collection of sources represents the information provided redundantly by the synergies of all sources in the collection, corresponding to one distinct way for the elements of **S** to contribute information about T.

Here, we will be interested in calculating only a few of these PI terms; namely, the unique information provided a single S_i, the n-way synergy for all elements in **S**, and the n-way redundancy for all elements in **S**. Fortunately, each of these terms can be calculated quite easily. Their calculation utilizes the fact that, in the context of PI diagrams, $I_{\min}$ and $I_{\max}$ are formally analogous to set intersection and set union, respectively. That is, $I_{\min}(T;\mathbf{A}_1,\ldots,\mathbf{A}_k)$ corresponds to the region $\cap_i I(S;\ \mathbf{A}_i)$, and likewise $I_{\max}(T;\mathbf{A}_1,\ldots,\mathbf{A}_k)$ corresponds to the region $\cup_i I(S;\mathbf{A}_i)$. This correspondence between $I_{\min}/I_{\max}$ and $\cap/\cup$ means that expressions for PI regions can be calculated using the operations of set theory. Thus, the unique information for a single S_i is conveniently given by

$$\Pi(T;\{S_i\}) = I(T;S_i) - I_{\min}(T;S_i,\mathbf{S}\setminus S_i). \tag{5}$$

Similarly, the n-way synergy for all variables in **S** is given simply by

$$\Pi(T;\{S_1,\ldots,S_n\}) = I(T;\mathbf{S}) - I_{\max}(T;\{\mathbf{A}\subset\mathbf{S}:|\mathbf{A}|=n-1\}). \tag{6}$$

Finally, the n-way redundancy for all variables in **S** is given by

$$\Pi(T;\{S_1\}\cdots\{S_n\}) = I_{\min}(T;S_1,\ldots,S_n). \tag{7}$$

III. PI DECOMPOSITION OF EXISTING MEASURES

We begin our investigation by analyzing the measures proposed by Lizier *et al.*[9] in light of PI decomposition. We consider, in order, the active information, apparent transfer entropy, and separable information, which were proposed to quantify the information storage, transfer, and modification, respectively.[11] In the first two cases, our analysis shows that the measures actually quantify several different kinds of information, suggesting that a clearer picture may be obtained by teasing apart their constituent elements. In the third case, PI decomposition suggests an entirely different measure that appears better suited to capture what separable information was designed to quantify.

A. Active information

The active information for a stochastic process X at time step $n+1$ is defined by

$$A_X(n+1) = I(X_{n+1};X_n^{(k)}), \tag{8}$$

where X_{n+1} denotes the state of X at the current time step and $X_n^{(k)}$ denotes the states of X in the preceding k time steps, i.e., $X_n^{(k)} = \{X_{n\ \ k+1},\ldots,X_n\}$. Thus, A_X measures the information that the previous k states of X provide about its own next state. Lizier *et al.* describe this measure as the information stored by the process X that is currently being used in computing its own next state hence, the idea of active information storage. In particular, in the context of CAs, the authors propose that A_X should specifically highlight the background domains of a space-time diagram, with these being the emergent structures primarily associated with information storage. However, we propose that background domains can be identified with a measure that is considerably simpler than the active information. To support this proposal, we first note that A_X is composed of several different kinds of information, as revealed by its PI decomposition (Fig. 2). In particular, A_X includes both the information that X obtains uniquely from its own previous states as well as the information provided redundantly by other potential influences. Defining $\mathbf{V}_X$ to be the set of causal information sources to X (e.g., if X is a cell in an ECA, then $\mathbf{V}_X$ consists of X and its two immediate neighbors), the active information can be written as

$$A_X(n+1) = \Pi(X_{n+1};\{X_n^{(k)}\}) + I_{\min}(X_{n+1};X_n^{(k)},\mathbf{V}_X\setminus X), \tag{9}$$

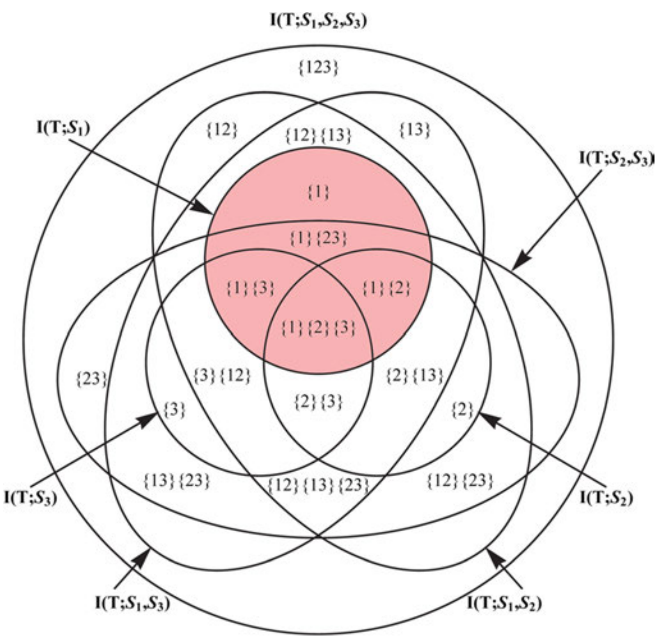


FIG. 2. (Color online) PI decomposition of active information for target variable T X_{n+1} and causal information sources S_1 $X_n^{(k)}$, S_2 $W_n^{(l)}$, and S_3 $Y_n^{(m)}$.

where the first term is the unique information that $X_n^{(k)}$ provides about X_{n+1} and the second term is the redundancy between $X_n^{(k)}$ and all other causal sources. Moreover, the latter term can be decomposed further into the redundancy with one other source, with both other sources, and with the synergy from both other sources. Thus, for causal contributors $\mathbf{V}_X = \{\mathrm{W, X, Y}\}$, the full expansion of A_X is given by

$$\begin{aligned} A_X(n+1) = \; & \Pi(X_{n+1}; \{X_n^{(k)}\}) \\ & + \Pi(X_{n+1}; \{X_n^{(k)}\}\{W_n^{(l)}\}) \\ & + \Pi(X_{n+1}; \{X_n^{(k)}\}\{Y_n^{(m)}\}) \\ & + \Pi(X_{n+1}; \{W_n^{(l)}\}\{X_n^{(k)}\}\{Y_n^{(m)}\}) \\ & + \Pi(X_{n+1}; \{X_n^{(k)}\}\{W_n^{(l)}, Y_n^{(m)}\}). \end{aligned} \quad (10)$$

However, only one of the terms in this expansion captures the essential property of background domains that they are *spatially and temporally homogeneous*.[3] This homogeneity means that each cell in a background domain is highly predictive of all other cells, or, in informational terms, that all cells are *maximally redundant*. Consequently, we hypothesize that background domains can be identified by the 3-way redundancy

$$\Pi(X_{n+1}; \{W_n^{(l)}\}\{X_n^{(k)}\}\{Y_n^{(m)}\}), \quad (11)$$

which is the only term in the expansion of A_X that captures maximal redundancy between a target cell and each of its parents.

B. Apparent transfer entropy

The apparent transfer entropy from one process Y to another process X at time step $n+1$ is defined by

$$T_{Y\to X}(n+1) = I(X_{n+1}; Y_n^{(m)} | X_n^{(k)}). \quad (12)$$

Thus, $T_{Y\to X}$ measures the information that the m previous states of Y provide about the current state of X when conditioned on the k previous states of X.[9,12] The idea behind this measure, as described by Lizier *et al.*, is that conditioning on the previous states of X eliminates or "conditions out" X's self-influence (i.e., the information that previous states of X transfer to its current state), thereby isolating the information that is transferred from Y to X. However, as has previously been noted for transfer entropy,[13] this notion of "conditioning out" is misleading, since conditioning on $X_n^{(k)}$ does not actually remove its influence. Rather, while conditioning on $X_n^{(k)}$ does remove the redundant information provided by both $X_n^{(k)}$ and $Y_n^{(m)}$, it also factors in the synergistic information associated with the interaction of $X_n^{(k)}$ and $Y_n^{(m)}$. In particular, for $\mathbf{S} = \{X_n^{(k)}, Y_n^{(m)}\}$, a PI decomposition of apparent transfer entropy yields

$$T_{Y\to X}(n+1) = \Pi(X_{n+1}; \{Y_n^{(m)}\}) + \Pi(X_{n+1}; \{X_n^{(k)}, Y_n^{(m)}\}), \quad (13)$$

where the first term quantifies the unique information that Y transfers to X and the second term quantifies the synergistic information transferred from Y and X's own past (Fig. 3). Therefore, we find that local transfer entropy incorporates the information that Y provides but does not isolate it. Consequently, we hypothesize that by quantifying the unique information $\Pi(X_{n+1}; \{Y_n^{(m)}\})$ we can completely eliminate X's self-influence and identify only those interactions that involve transfer from Y to X.

C. Separable information

The separable information for a process X at time step $n+1$ is defined by

$$S_X(n+1) = A_X(n+1) + \sum_{Y \in \mathbf{V}_X \setminus X} T_{Y\to X}(n+1). \quad (14)$$

As described by Lizier *et al.*, this measure is intended to identify interactions where informationally "the whole is greater than the sum of its parts," corresponding to nontrivial information processing or information modification. Here "the whole" refers to the information contributed by all causal sources together, while "the parts" refers to the contributions from each individual source. Thus, it is somewhat

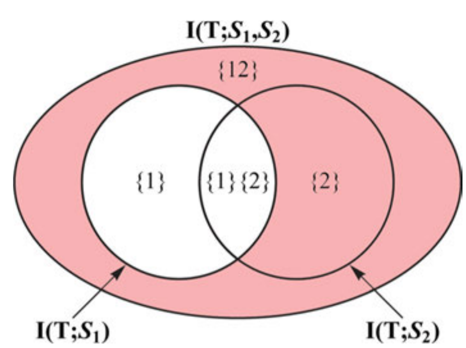


FIG. 3. (Color online) PI decomposition of apparent transfer entropy for target variable T X_{n+1} and causal information sources S_1 $X_n^{(k)}$ and S_2 $Y_n^{(m)}$.

odd that separable information does not actually include a measure of "the whole" as such, i.e., there is no measure of the contribution from all information sources. Rather, as Lizier *et al*. note, the separable information is explicitly a calculation of the sum of the parts i.e., the self-influence A_X and the transfer $T_{Y\rightarrow X}$ from each neighbor while the comparison with the whole is made only implicitly by a somewhat arbitrary thresholding of S_X. In contrast, PI decomposition suggests a natural and direct way of quantifying the extent to which the whole contributes information beyond the sum of the parts; namely, the 3-way synergy

$$\Pi(X_{n+1};\{W_n^{(l)},X_n^{(k)},Y_n^{(m)}\}). \tag{15}$$

Consequently, we hypothesize that this measure, which is not included at all in the separable information (Fig. 4), will be better able to detect information modification events.

IV. APPLICATION TO RULE 184

In this section, we apply a filtering process based on PI decomposition to ECA Rule 184. We have selected this rule because it clearly illustrates situations where our approach differs from the approach proposed by Lizier *et al*., affording us the opportunity to explain in detail, and often mechanistically, why these two methods produce different results. Our procedure for filtering CA runs is as follows. First, a CA consisting of 200 cells with periodic boundary conditions is iterated for 200 time steps from a random initial state. The first 30 time steps are discarded to reduce the effects of initial transients, while the remaining time steps are retained for our analysis. Next, for each cell in the CA, a joint probability distribution is estimated for that cell and its three causal sources, where for each causal source we used a history length of 6 (i.e., $k=l=m=6$). This probability distribution is estimated

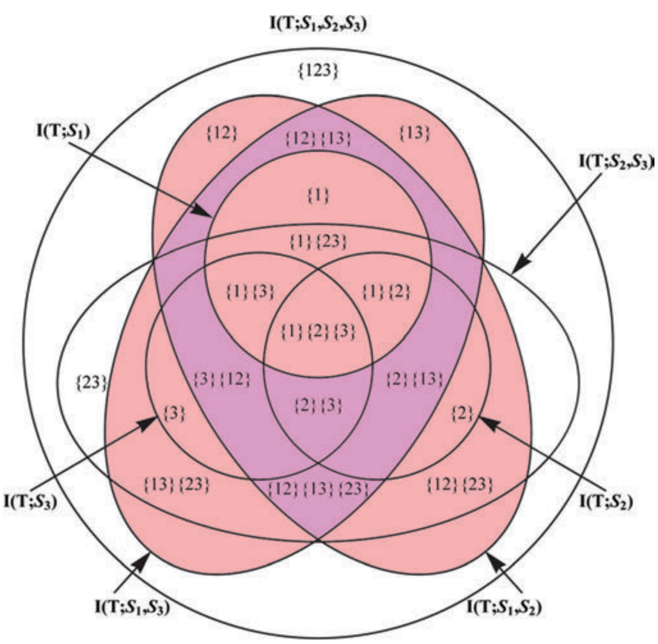


FIG. 4. (Color) PI decomposition of separable information for target variable T X_{n+1} and causal information sources S_1 $X_n^{(k)}$, S_2 $W_n^{(l)}$, and S_3 $Y_n^{(m)}$. Purple regions denote PI terms that are double counted.

from samples of the cell and its causal sources in the current and 5 preceding time steps, for a total of 6 samples. This small number of samples was used in order to capture the local statistics of each cell, thus allowing our measures to quantify the local information dynamics. Finally, PI measures are calculated from these local probability distributions, and filtered plots of the run are produced by assigning the values of each PI measure to the location of the corresponding target cell (Fig. 5). Our comparison plots were produced using exactly the method described by Lizier *et al*.[9] using $k=6$, and all results were confirmed with at least 50 runs from different random initial states.

Before proceeding, it is important to understand how our measure is "local" and how this differs from the version of localization used by Lizier *et al*.'s measures. We consider the PI decomposition filter to be local in that it only considers state histories in a localized spatiotemporal region. In contrast, Lizier *et al*.'s measures are local in the sense that they consider the unaveraged information a particular state history provides. This difference is important to note, as it explains why their measures may provide negative values, but the PI decomposition filter does not.

A. Regular domain

Recall that, in ECAs, information storage is often associated with the regular domain, as this domain retains uniform or repeating configurations of cells over time until disrupted. We would therefore expect that both Lizier *et al*.'s approach to characterizing information storage and the approach described in this paper to characterizing stationary particles would clearly identify the regular domain in Rule 184. Fig. 6(a) shows the raw output of the ECA Rule 184; note the ample regions of regular domain.

As expected, active information, as proposed by Lizier *et al*. to measure information storage, does identify regular domains. This is shown by the positive values of active information (red) in Fig. 6(c). Note, however, that active information does not only assign value to the regular domain; it also assigns negative values (orange and yellow) to particles and their collisions, as well as various other values to all other areas of the Rule 184 output. This broad assignment of value can be adjusted so that only the regular domain is given a positive value, but this requires the adjustment of a threshold. The level of this threshold will, in general, be different for other ECA rules to produce this effect, so it is not broadly applicable. From this, it is not obvious that the active information can be used to identify only information storage across all ECA rules.

We now turn to how a partial information decomposition can be used to identify the regular domain, associated with information storage. Here, we propose that 3-way redundancy can be used for this purpose. Recall that this is the minimum information that three source cells share about the target that does not overlap with unique or synergistic information. From Fig. 6(b), it is clear that 3-way redundancy assigns positive value to the regular domain in the output of Rule 184. Note that it does not assign value to particles (it does, however, assign nonzero values to the edges of particles due to

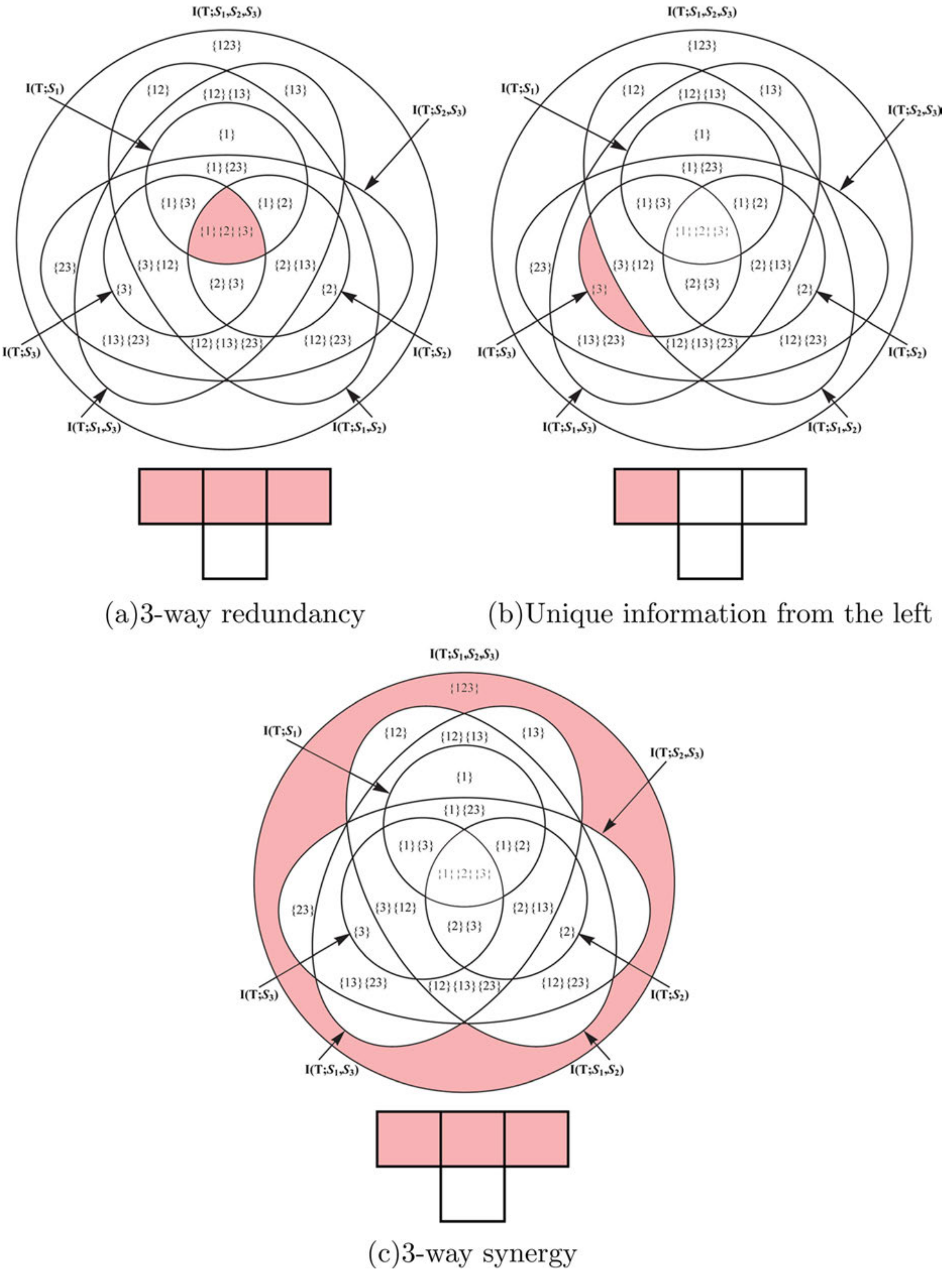


FIG. 5. (Color online) A schematic representation of how the PI decomposition relates to ECA. (a) 3 way redundancy looks at the information that all three parent cells provide redundantly about the target cell. (b) Unique information from the left looks at the information that the leftmost cell provides uniquely about the state of the target cell. (c) 3 way synergy looks at the information that all three parent cells provide synergistically about the target cell.

how the state histories were constructed). We obtain very similar results when 3-way redundancy is applied to ECA rules 19 and 50 (data not shown), which are simple automata that produce almost entirely regular domains.

A potential anomaly is highlighted by the circle in Fig. 6(a). Here, 3-way redundancy produces a nonzero value at the collision point between two particles. If 3-way redundancy is to be measuring only the regular domain, why would this

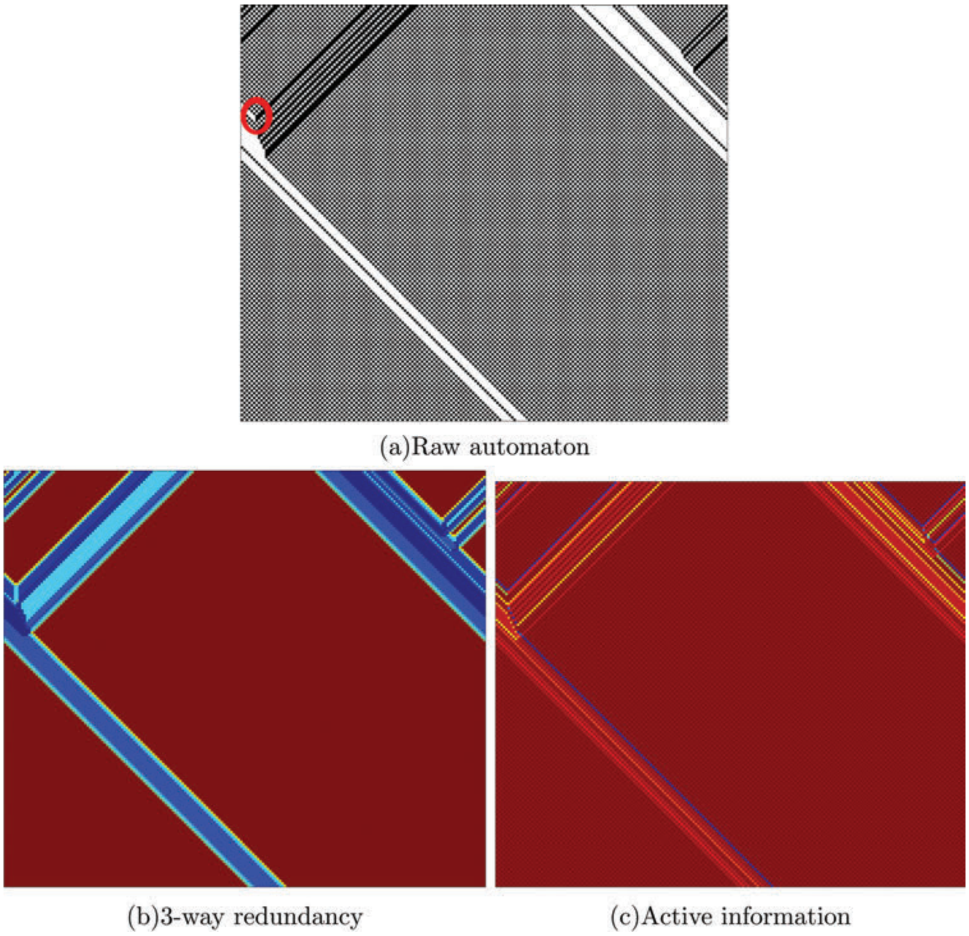


FIG. 6. (Color) Regular domain in Rule 184. (a) Raw output of the CA. (b) 3 way redundancy, max. 16 bits (dark red), min. 0 bits (dark blue). (c) Active information, max. 1.04 bits (dark red), min. 5.44 bits (dark blue). Dark orange is approximately 0 bits.

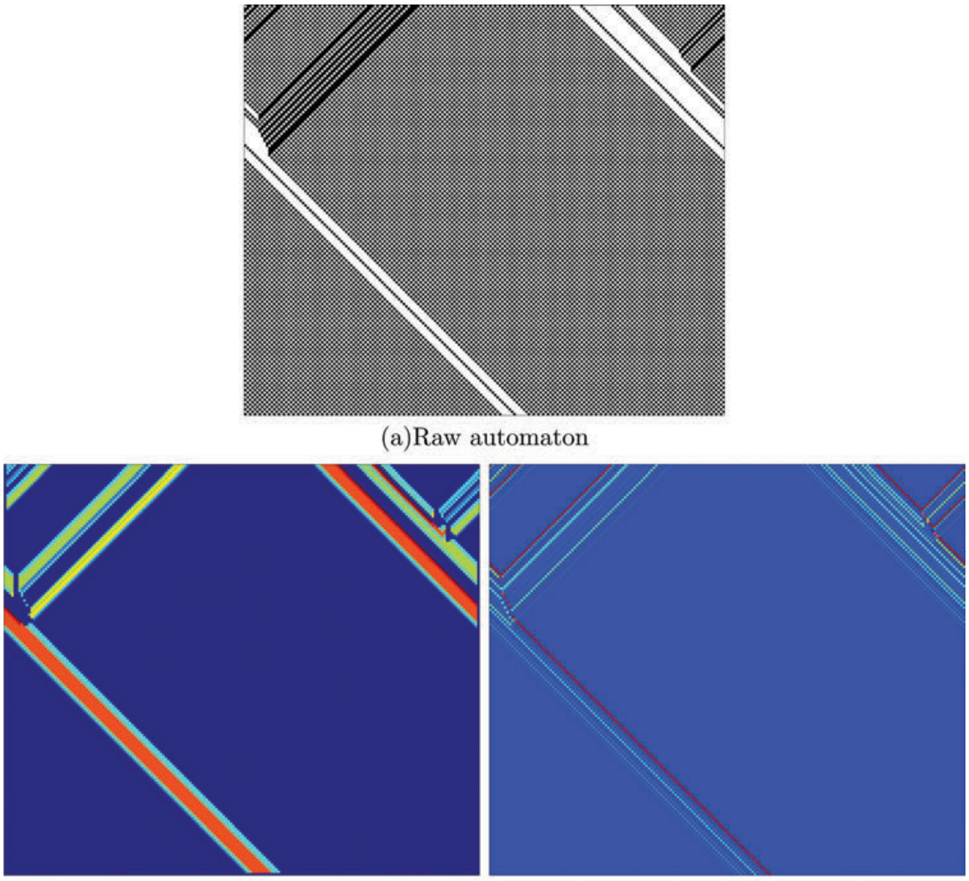
occur? In order to adequately explain this, we must first understand what is happening at the collision sites mechanistically. In Rule 184, there are two types of collisions: collisions that result in the modification of one of the particles and collisions that result in the destruction of both particles. Note from Fig. 6(b) collisions that result in the modification of a particle are not measured to be redundant, which is consistent with our hypothesis that 3-way redundancy will identify stationary particles and not particle collisions. In Rule 184, we find that only totally destructive collisions are measured to be redundant. Mechanistically, this is because when a white particle and a black particle of equal width meet, they form the pattern of the regular domain and thus become part of the background. Thus, we can see that destructive collision sites are measured as redundant because they match the pattern of the regular domain, which is used for storing information in ECA. From this perspective, the results of the partial information decomposition are still consistent with our hypothesis, and we can conclude that 3-way redundancy appropriately quantifies the regular domain in ECA Rule 184.

Why then does active information not produce results similar to those of 3-way redundancy, instead highlighting particles and collision points, as well as the regular domain? Recall how we used partial information to decompose active information into non-overlapping parts, given in Eq. (10) and depicted in Fig. 2. From this, it was seen that active information is actually a mixture of unique, synergistic, and redundant terms. Thus, it should not be surprising that it assigns value to particles and collisions in addition to the regular domain in Rule 184. Clearly, active information and 3-way redundancy do not filter out the same things. Only 3-way redundancy uniquely identifies the regular domain in Rule 184.

B. Traveling particles

We now turn to our approach for identifying particles, which are commonly associated with information transfer. Information transfer is thought to be accomplished by moving particles, as they can communicate the state of one or more cells over space and time in ECAs. We would therefore expect that both Lizier *et al.*'s approach and the approach described in this paper to characterizing these structures would clearly identify particles in the output of Rule 184. Fortunately, there are many particles clearly visible in Fig. 7(a) of the raw output of ECA Rule 184.

Lizier and colleagues propose that local apparent transfer entropy can identify information transfer. Because this is a directional measure (as is unique information, given below), we measure it from both the left and from the right to quantify the total information transfer from both directions. When we do this to the output of Rule 184 in Fig. 6(c), we see that it does highlight particles and not the regular domain. However, it also weakly highlights collisions, sites that are not thought to be involved in information transfer. This is not what we would expect if apparent transfer entropy is measuring information transfer alone.



(a)Raw automaton

(b)Unique information

(c)Local apparent transfer entropy

FIG. 7. (Color) Traveling particles in Rule 184. (a) Raw output of the CA. (b) Unique information from the left and right summed, max. 22.78 bits (dark red), min. 0 bits (dark blue). (c) Local apparent transfer entropy from the left and right summed, max. 6.40 bits (dark red), min. 1.12 bits (dark blue). Blue is approximately 0 bits.

In contrast, we propose that unique information can be used to identify only the traveling particles. When unique information is measured (from both the left and the right) in the output of Rule 184, only particles are highlighted (Fig. 7(b)). The regular domain, as well as collision sites, is assigned values of zero. We obtain similar results when unique information is extracted from the output of Rules 2 and 48, simple automata that consist almost entirely of traveling particles (data not shown). This is as we would expect for a measure of moving particles.

Why does local apparent transfer entropy highlight collisions in addition to the particles it is expected to identify? Again, we may refer back to the partial information decomposition that was performed in Eq. (13) and is depicted in Fig. 3. Local apparent transfer entropy is made up of both unique and synergistic terms. From this, we would expect it to identify particles as well as collisions. We wish to note that local apparent transfer entropy and unique information do not measure the same things. Only unique information identifies the traveling particles, and only the traveling particles, in the output of Rule 184.

C. Particle collisions

Here, we will examine both the separable information and PI approaches to quantifying particle collisions. It is widely believed that collisions between particles are sites of information modification, as the results from a collision can often be predicted only by jointly considering the particles involved. Collisions are neither merely uninterrupted transfer of information nor are they stored information in the form of a continuation of the regular domain. We expect that both approaches to characterizing information modification would clearly identify collisions in the output of Rule 184. There are several collisions clearly visible in Fig. 8 of the raw output of ECA Rule 184.

Lizier *et al*. propose that information modification will be identified by negative values of separable information. This is plausible and deserves some explanation so the reader can appreciate their argument. They reason that the past histories of both particles involved in a collision, as well as the past history of the cell where the collision occurs, when considered independently, are insufficient to correctly predict what will come after the collision. From their perspective, information modification is an event produced by joint interactions and is unexpected from the perspective of any single contributor to the collision. Thus, the sum of active information and the two apparent transfer entropies, which they define as the separable information, will be negative after a collision, indicating misinformation about the results of the collision. Accordingly, they have characterized information modification as an event where the whole is more than merely the sum of its parts.[9]

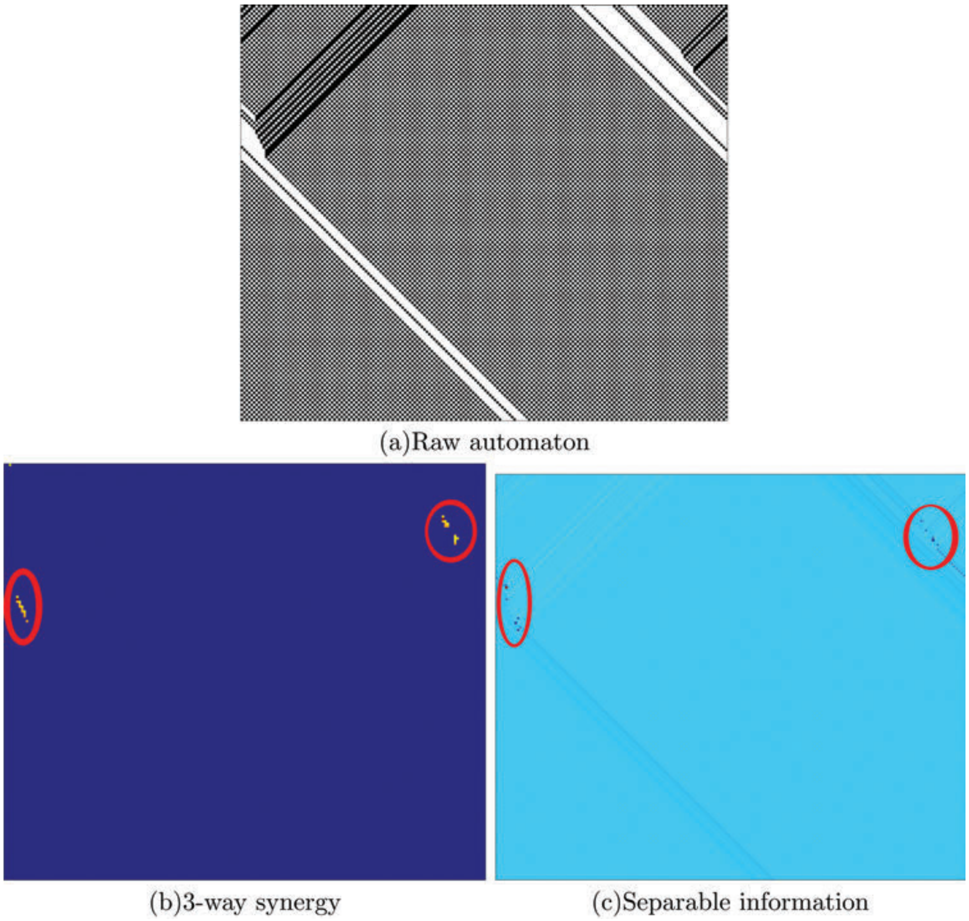


FIG. 8. (Color) Particle collisions in Rule 184. (a) Raw output of the CA. (b) 3 way synergy, max. 8.68 bits (dark red), min. 0 bits (dark blue). (c) Separable information, max. 3.72 (dark red), min. 0.33 bits (dark blue). Light blue is approximately 1 bit.

When we examine the local separable information produced by Rule 184 in Fig. 8(c), we see that it does not highlight all collisions with negative values, but most collision points are in fact mostly positive. Although the largest negative values occur at collision points, not all collisions produce negative values. This indicates that negative local separable information cannot, by itself, be used to identify collisions in Rule 184. As in the case of active information, it may be possible to introduce a threshold of some sort so that only collision points are highlighted, but it is not clear if such a threshold would generalize to other ECA rules, and it is not clear how such a threshold would be determined from first principles.

We propose that instead 3-way synergy can be used as an indicator of particle collisions. Recall that this is the information that the three source cells provide jointly about the target that is not unique or redundant. When we examine the values of 3-way synergy produced by Rule 184 in Fig. 8(b), we see that it produces positive values only at the collisions; the particles and the regular domain are all zero. Interestingly, the locations where 3-way synergy is positive do not overlap with the regions that were highlighted by unique information (compare Figs. 8(b) and 7(b)). In fact, they fill the gaps left in the plot of unique information, which further fill the gaps left in the plot of 3-way redundancy (Fig. 6(b)). This is as we would expect if information is decomposed into non-overlapping terms, as the partial information approach claims to do.

Why do the different approaches produce such different results? Again, we refer back to the decomposition that we performed in Eq. (14), recalling that separable information is the sum of active information and apparent transfer entropies. This decomposition showed that the local separable information is in fact a combination of redundancy, unique information, and synergy. From this perspective, one might expect that it would highlight particles and the domain as we see in Rule 184. Again we are forced to conclude that the partial information decomposition and the approach proposed by Lizier and colleagues have substantial differences. In the case of Rule 184, it is clear that only 3-way synergy uniquely highlights collisions.

V. CONCLUSION

In this paper, we have suggested some new directions for information-theoretic approaches to the characterization of distributed computation in CAs. First, we have demonstrated issues with the specific measures of active information, apparent transfer entropy and separable information proposed by Lizier *et al.* to identify information storage, transfer, and modification. In particular, using partial information decomposition, we have shown that both active information and apparent transfer entropy can confound multiple sources of information. In addition, we have argued that separable information does not really capture locations where "the whole is greater than the sum of the parts" as Lizier

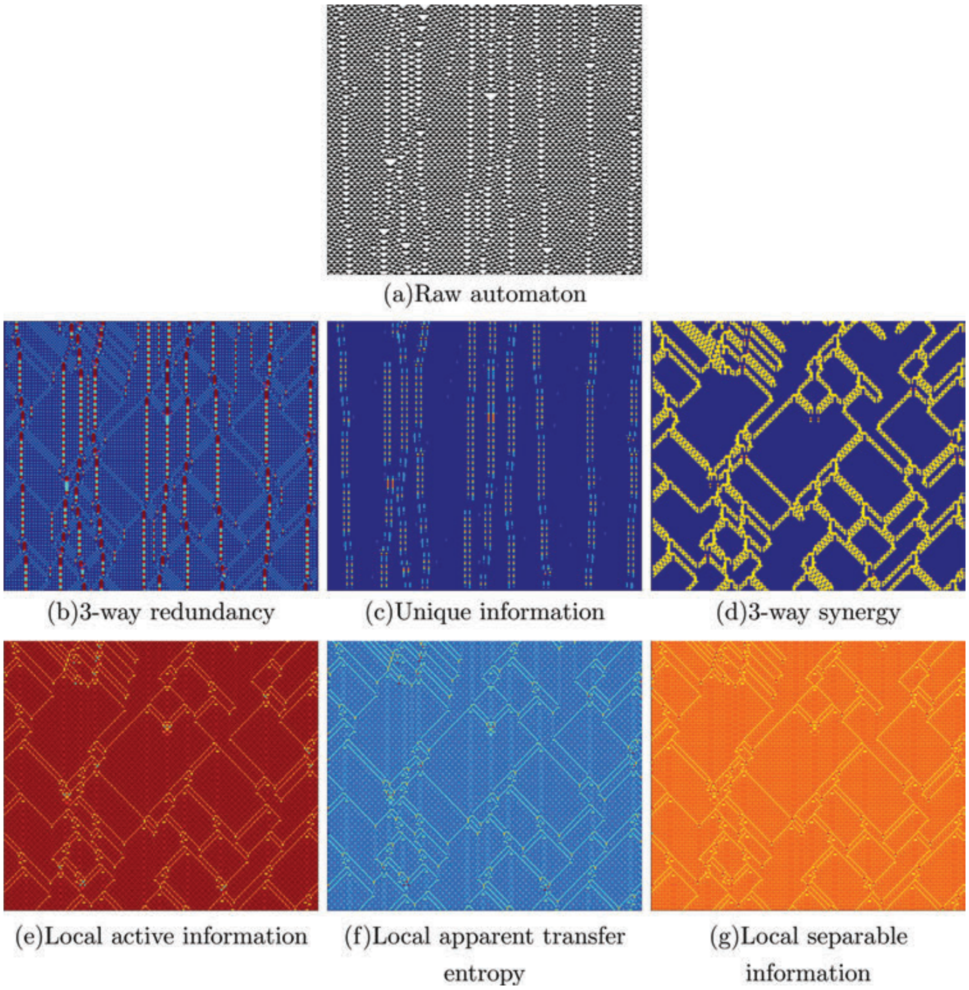


FIG. 9. (Color) Information filters applied to Rule 54.

et al. intended. Second, we have proposed new filters in which 3-way redundancy is associated with the regular domain, unique information is associated with traveling particles, and 3-way synergy is associated with particle collisions. Finally, we have demonstrated the utility of our filter using ECA Rule 184. On this ECA, Lizier *et al*.'s measure identify regions of information transfer in collisions and information modification in both collisions and particles. In general, it appears that these measures require potentially arbitrary thresholds to distinguish collisions, particles, and background. In contrast, our measures cleanly separate the regular domain, traveling particles, and particle collisions.

A key future direction for our approach is its extension to more complex ECAs such as Rules 18, 54, and 110 analyzed by Lizier *et al*.[9] Applying the PI filter itself to these rules is straightforward; the challenge lies in interpreting the results. For example, an application of our filter to Rule 54 is shown in Fig. 9. As would be expected from the analysis of Rule 184 in this paper, our results for Rule 54 and those discussed by Lizier *et al*. are quite different. However, unlike for Rule 184, our filter does not cleanly separate background, particles, and collisions in Rule 54. We believe that the central problem here is one of scale since, in contrast to Rule 184, particles in Rule 54 can consist of multiple adjacent cells. We note that identifying such larger-scale particles and their collisions is a problem for *any* local information-theoretic measure. Thus, if we wish to quantify the emergent information processing in distributed systems, an important direction for future research is determining the appropriate scale at which to apply an information-theoretic analysis. One possibility would be a hybrid approach that utilizes computational mechanics to identify the scale of the relevant emergent structures and then applies information theory to quantify the flow of information through them. In any case, the different results produced by our measures and Lizier *et al*.'s suggest that partial information decomposition can provide insights into distributed computation in CAs in addition to those provided by the innovative work of Lizier *et al*.

ACKNOWLEDGMENTS

The authors would like to thank Joseph Lizier for his comments. This work was supported in part by NSF Grant No. IIS-0916409 (to R.D.B.), an IGERT fellowship (to P.L.W.), and NSF Grant No. 0904912 (to J.M.B.).

[1]M. Mitchell, J. Crutchfield, and P. Hraber, Physica D **75**, 361 (1994).
[2]S. Wolfram, Physica D **10**, 1 (1984); M. Cook, Complex Syst. **15**, 1 (2004).
[3]J. Hanson and J. Crutchfield, Physica D **103**, 169 (1997).
[4]J. Crutchfield and M. Mitchell, Proc. Natl. Acad. Sci. **93**, 10742 (1995).
[5]M. Marques Pita and L. Rocha, The 2011 IEEE Symposium on Artificial Life, IEEE Press, 233 (2011).

[6]A. Wuensche, Complexity **4**, 47 (1999).
[7]C. Shalizi, R. Haslinger, J. Rouquier, K. Klinkner, and C. Moore, Phys. Rev. E **73**, 036104 (2006).
[8]T. Helvik, K. Lindgren, and M. Nordahl, Proc. Int. Conf. Cell. Automata Res. Ind. **3305**, 121 (2004).
[9]J. Lizier, M. Prokopenko, and A. Zomaya, Chaos **20**, 037109 (2010); Phys. Rev. E **77**, 026110 (2008).
[10]P. Williams and R. Beer, e print http://arxiv.org/abs/1004.2515 (2010).
[11]Although Lizier *et al.* focus on local versions of these measures, defined for individual states rather than variables, here we consider their averaged or global forms, since these provide for the most natural comparison with PI measures.
[12]T. Schreiber, Phys. Rev. Lett. **85**, 461 (2000).
[13]P. Williams and R. Beer, e print http://arxiv.org/abs/1102.1507 (2011).