

Measuring Complexity

Lecture 14 I400/I590

Artificial Life as an approach to Artificial Intelligence

Larry Yaeger

Professor of Informatics, Indiana University

Is It Alive? (Part 2)

- Farmer & Belin's list is about as accurate and complete a "laundry list" of life as exists
- Offers some valuable insights
- But ultimately fails
 - Purely qualitative nature makes it largely untestable and easy to attack
 - Would disallow an intelligent robot that had no internal self-representation
 - Would allow Polyworld organisms

But Who Cares?

- If we can obtain (by design or evolution) greater and greater levels of machine intelligence, it really doesn't matter if the artificial organisms are considered to be alive (for a while)
- If there is ever evidence of a subjective sensation of pain (at whatever level of intelligence), then it may influence our testing
 - For the same ethical reasons animal cognition is often framed in terms of pain perception
- Above a certain threshold of intelligence "human rights" become an issue, but that remains safely in the realm of science fiction for a good while yet

Quantifying Life & Intelligence

- By contrast, what if we had a quantitative measure of life and/or intelligence?
- Might a thermostat be 0.00003 alive, and the organisms of Polyworld a whopping 0.002 alive?
- If the measure could be applied to biological systems, it would give a formal meaning to the intuitive sense of a growth in complexity over evolutionary timescales
- If normalized so average adult human intelligence computed out to be 100.0, might it lend more meaning to the measurement of IQ (or point out its deficiencies)?
- Regardless, its application to artificial systems should permit a quantitative assessment of progress towards machine intelligence

Measuring Complexity

- Complexity is one possible candidate for providing our quantitative assessment of life and intelligence
- But how do we actually measure it?
- There are many metrics that claim to measure complexity
- Information theory seems to provide the best approach
- But not *just* Shannon Information (entropy, uncertainty), nor Kolmogorov, nor Chaitin, nor any measure that responds maximally to surprise
- Let's make some measurements in some well understood systems: Cellular Automata

A Capacity for Computation

- Chris Langton (1990) observes:
 - Most ALife papers assume the existence of a physical system with the capacity to support computation
 - He seeks to understand the conditions under which this "capacity to support computation" might emerge in a physical system

Problem Statement

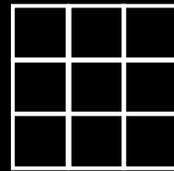
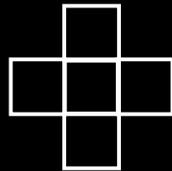
- "Under what conditions will physical systems support the basic operations of information transmission, storage, and modification constituting the capacity to support computation?"
- Difficult to address directly, reformulate the question in terms of a formal abstraction of physical systems:
- "Under what conditions will cellular automata support the basic operations of information transmission, storage, and modification?"
- This turns out to be a tractable problem, with an answer that leads to a hypothesis about the conditions under which computation might emerge spontaneously in nature

Formal Definition of Cellular Automata

- Lattice of dimension D with a finite automaton at each lattice site
- Each automaton takes as input the states of automata within a local neighborhood N
- $|N|$, the size of N , is just the number of lattice sites (cells) covered by N
 - Let $N = |N|$
- By convention, an automaton is considered to be a member of its own neighborhood

Formal Definition of Cellular Automata

- Two typical two-dimensional neighborhoods are:



- Each automaton may take on a number of states, S
- $|S|$, the size of S , is just the number of possible states
 - Let $K = |S|$

Formal Definition of Cellular Automata

- Input to an automaton is the set of states (from S) at all cells in the neighborhood (N)
 - The set of all possible input states at all possible neighbors is called the *input alphabet*, I , and may be written as $I = S^N$
- Output of an automaton is its own next state, determined from one of a set of rules, the *transition function*, R
$$R: S^N \rightarrow S$$

Formal Definition of Cellular Automata

- The size of the input alphabet is
$$|\mathcal{I}| = |\mathcal{R}| = |\mathcal{S}^N| = K^N$$
- To define a transition function R , you must define a single output state in \mathcal{S} for each possible input state in \mathcal{I}
- Since there are $K = |\mathcal{S}|$ possible output states that could be assigned to each of the $|\mathcal{I}|$ possible input states, there are $K^{(K^N)}$ possible transition functions R that can be defined
- Call this set of all possible transition functions that can be defined with K states and N neighbors \mathcal{D}_N^K

Examples

- 1-D binary CA with neighborhood size $N = 3$ and number of states $K = 2$ yields

$$|D_N^K| = 2^{(2^3)} = 256 \text{ possible rule sets}$$

- 2-D Conway Game of Life with neighborhood size $N = 9$ and number of states $K = 2$ yields

$$|D_N^K| = 2^{(2^9)} \approx 10^{154} \text{ possible rule sets}$$

- 2-D lattice with neighborhood size $N = 5$ and number of states $K = 8$ yields

$$|D_N^K| = 8^{(8^5)} \approx 10^{30,000} \text{ possible rule sets}$$

Parameterizing CA Rule Space

- Once K and N are chosen, the complete space of possible rule sets, D_N^K , is fixed
- However, there is no intrinsic order within D_N^K , so there is no way to characterize or select amongst the possible rule sets R
- Any method for imposing some structure on this rule space D_N^K , and associating that structure with the resulting dynamical behaviors, may help us characterize the conditions under which computation is likely to emerge

The Lambda Parameter

- Consider a subset of D_N^K characterized by a parameter, λ , defined as follows:
 - Pick an arbitrary state $s \in S$, and call it the *quiescent state* s_q
 - Count the number of rules that produce this particular quiescent state, and call it n
 - The other $K^N - n$ transitions must produce the remaining $K - 1$, *non-quiescent* states of $S - s_q$, but may otherwise be chosen at random

Define: $\lambda = (K^N - n) / K^N$

Boundary Conditions on $\lambda = (K^N - n) / K^N$

- If $n = K^N$, so all rules lead to the quiescent state, then no rules lead to any non-quiescent states and $\lambda = 0.0$
- If $n = 0$, so no rules lead to the quiescent state, then all K^N rules lead to non-quiescent states and $\lambda = 1.0$
- When all K states are represented equally in the rule set, then $\lambda = 1.0 - 1/K$
 - $n = K^N / K$
 - $\lambda = (K^N - K^N/K) / K^N = (1 - 1/K) / 1 = 1 - 1/K$
- $\lambda = 0$ corresponds to the most homogeneous rule set, and $\lambda = 1 - 1/K$ corresponds to the most heterogeneous rule set, so most experiments are in this range

Parametric Study

- λ is used to sample the total rule space D_N^K as follows:
 - Step incrementally from $\lambda = 0$ to $\lambda = 1 - 1/K$
 - Randomly construct a rule set R for each λ
 - Run the resulting CA, gathering data on its dynamical behavior
 - Plot and examine these data as a function of λ
- There are two methods used for the construction of the rule set

Random-Table Method

- For each new value of λ , start with an empty rule table
- Stepping through each of the rules in the table
 - Select a random number, r , between 0.0 and 1.0
 - If $r > \lambda$, then make output of the current rule s_q
 - Else randomly select one of the non-quiescent rules
- This is "flipping a λ -biased coin" for each neighborhood state (each rule)

Table-Walk-Through Method

- Start with a table in which all rules lead to s_q ($\lambda = 0$)
- Generate tables with larger values of λ by randomly replacing a few of the transitions to s_q with transitions to other states (also randomly selected)
- Generate tables with smaller values of λ by randomly replacing a few of the transitions to other states with transitions to s_q

The Two Methods

- With the random-table method, each new table is generated from scratch
- With the table-walk-through method, we progressively perturb "the same table"
- The first method is good for randomly sampling state space
- The second method is good for following trajectories through state space

Observations About Lambda

- Other parameterizations of CA rule space exist, but the simplicity and single-dimensionality of λ make it an attractive first cut
- λ discriminates well between dynamical regimes for "large" values of K and N , but not for small dimensional spaces
 - For example, λ is only roughly correlated with dynamical behavior for 1-D CAs with $K=2$ and $N=3$
 - This may be why some previous studies failed to observe the relationships Langton presents
 - Langton sticks to CAs with $K \geq 4$ and $N \geq 5$, which results in transition tables of size $4^5 = 1024$ or larger

Elements of Computation

- All proofs of universal computation in CAs rely on three fundamental features of the CA dynamics:
 - *Storage of information*—the system must be able to preserve local state for arbitrarily long times
 - *Transmission of information*—the system must be able to propagate signals over arbitrarily long distances
 - *Modification of information*—stored and transmitted signals must be able to interact with one another, resulting in a possible modification of one or the other

Requirements for Computation

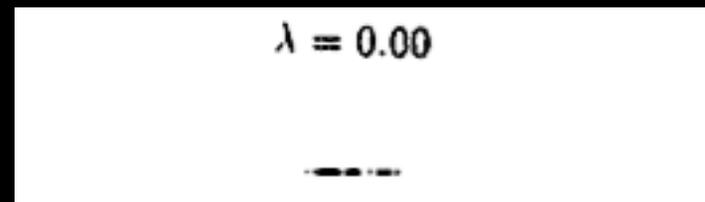
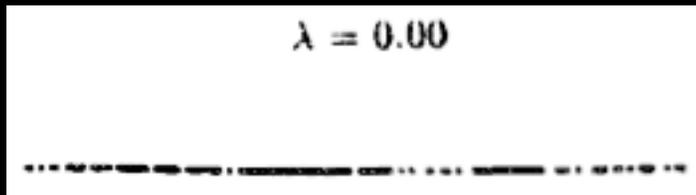
- Taken together, the elements of computation require that any dynamical system capable of computation *"must exhibit arbitrarily large correlation lengths in space and time"*
- These correlation lengths must be *potentially infinite*, but not *necessarily so*
- E.F. Codd says this propagation of information must be *unbounded in principle*, but *bounded in practice*

Characterizing CA Dynamics

- Step through λ , using table-walk-through method, for a CA with $K = 4$, $N = 5$
- Width of CA is 128 cells, with wrap-around
- We will look at two series of tests:
 - The first, on the left, always starts from the same pattern created by randomly setting the state over all 128 sites
 - The second, on the right, always starts from the same pattern created by randomly setting the state of the central 20 sites (with all others being initialized to zero)

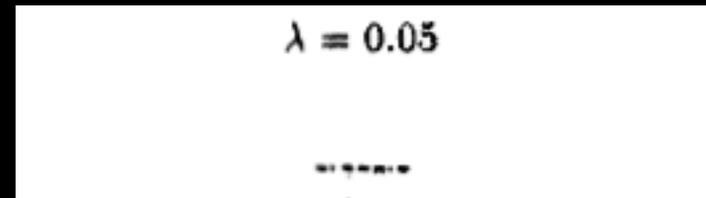
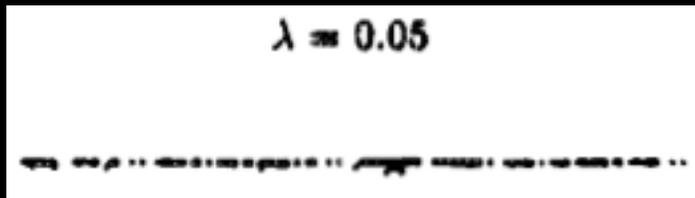
CA Dynamics, $\lambda = 0.0$

- Activity dies out in one time step, all cells in state s_q



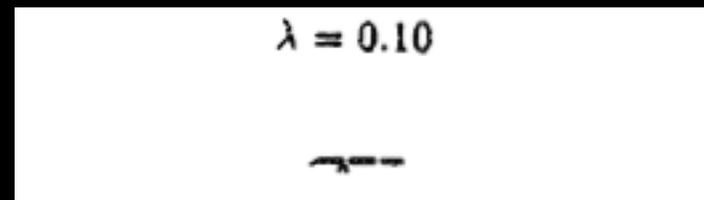
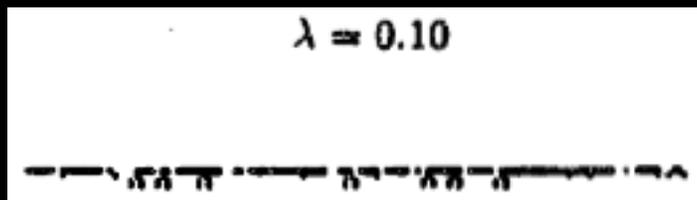
CA Dynamics, $\lambda = 0.05$

- Reaches uniform s_q fixed point after about 2 time steps



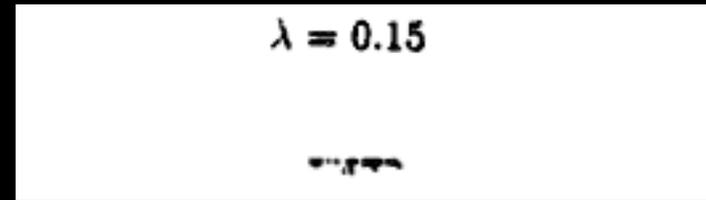
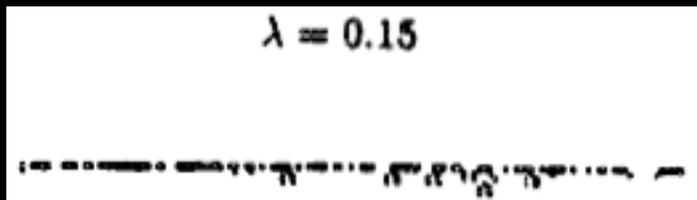
CA Dynamics, $\lambda = 0.10$

- Reaches uniform s_q fixed point after 3 or 4 time steps



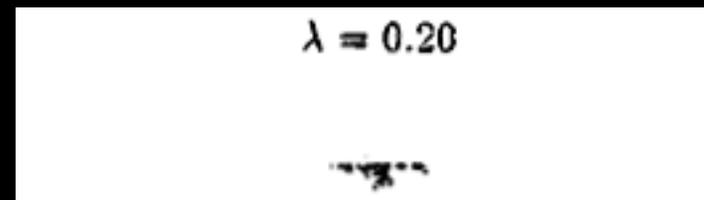
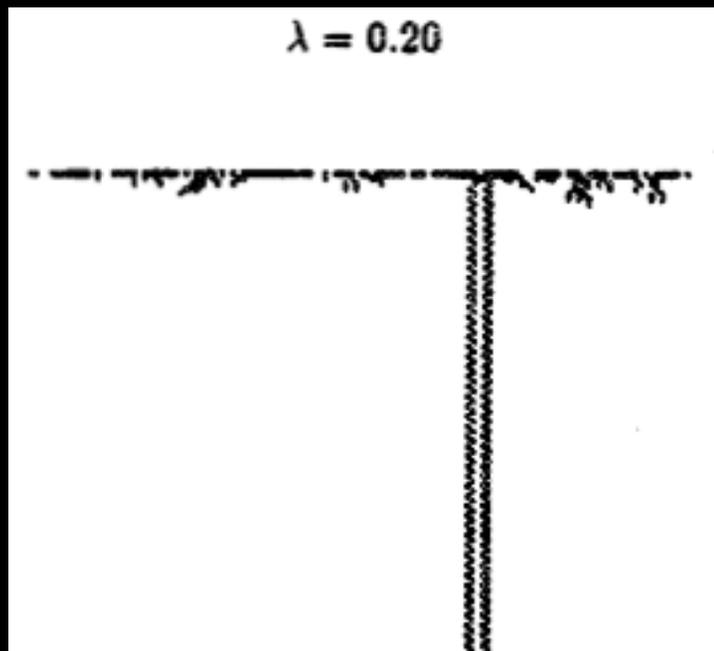
CA Dynamics, $\lambda = 0.15$

- Reaches uniform s_q fixed point after 4 or 5 time steps



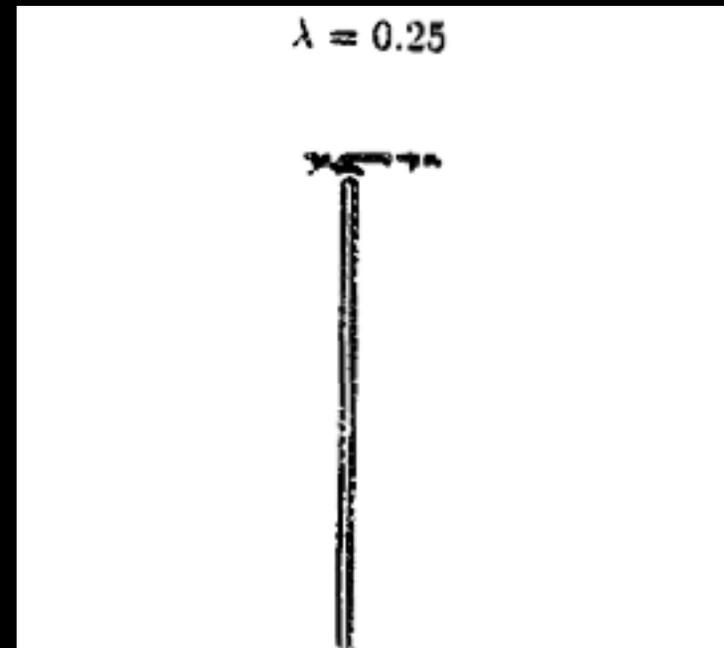
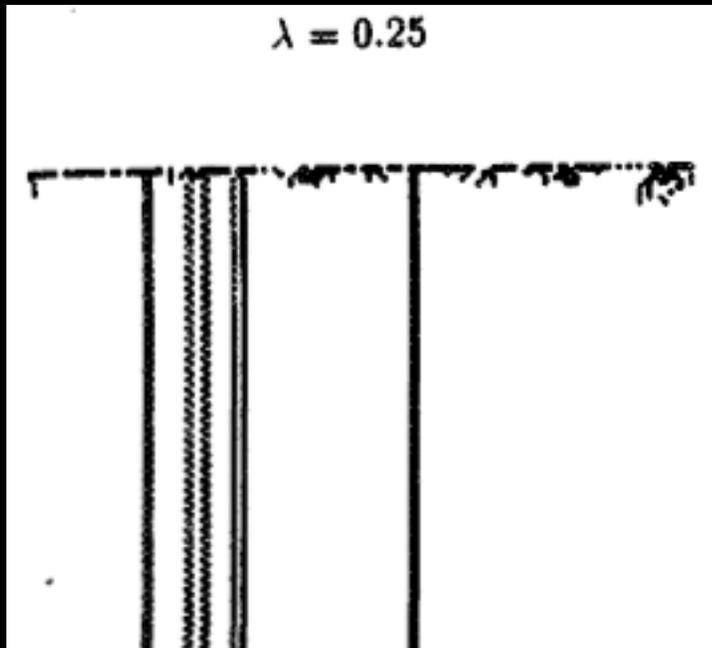
CA Dynamics, $\lambda = 0.20$

- Attains a periodic structure; transients last 7 to 10 steps



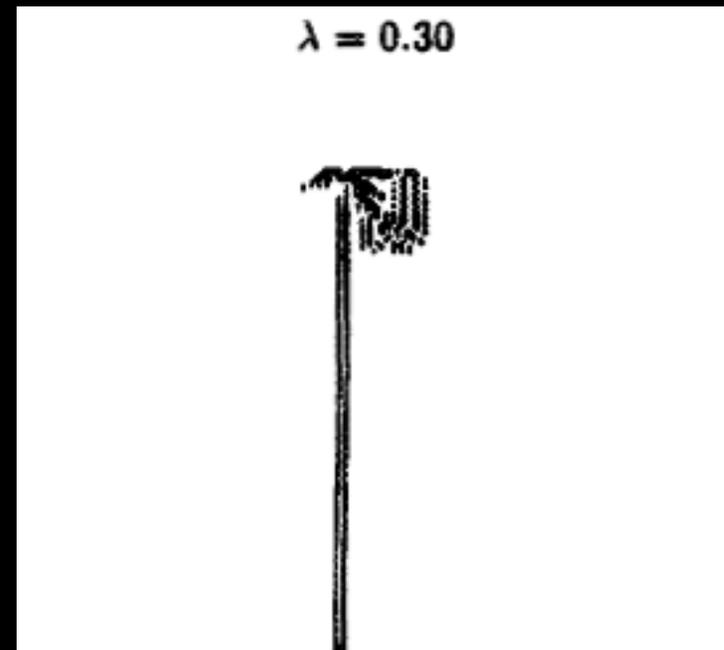
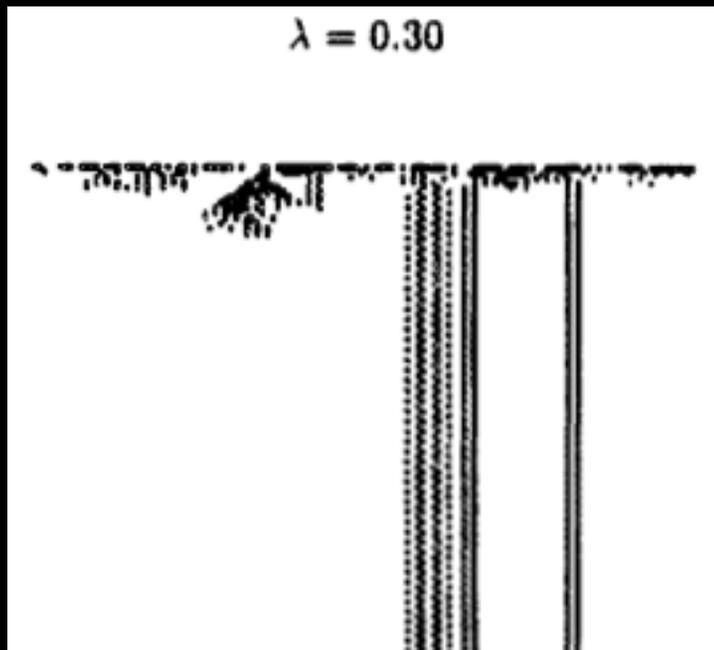
CA Dynamics, $\lambda = 0.25$

- Structure of period 1 appears; 3 final states possible - fixed s_q , fixed s_q and other, or fixed s_q and periodic; transients have grown in length



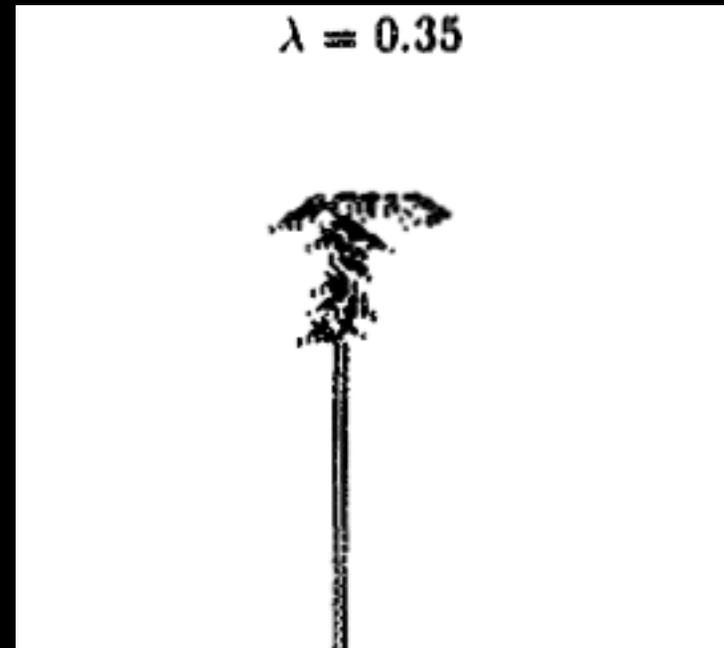
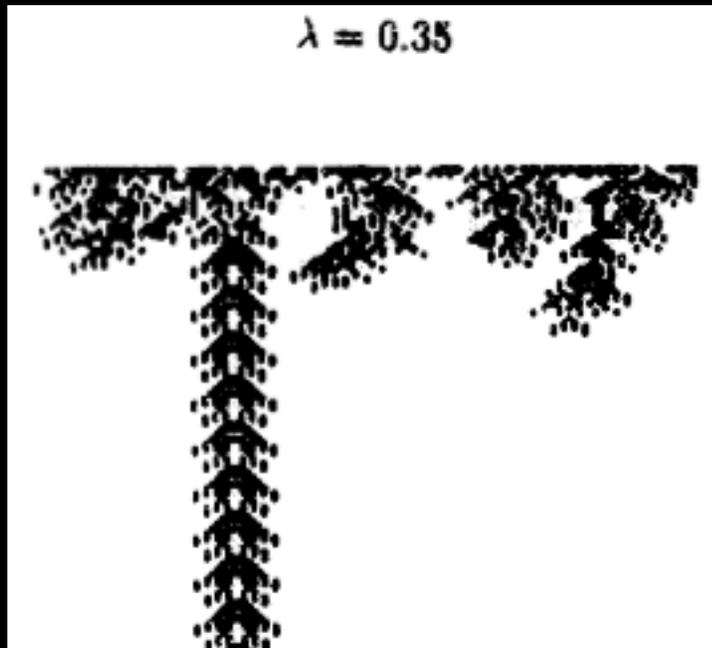
CA Dynamics, $\lambda = 0.30$

- Transients have lengthened again



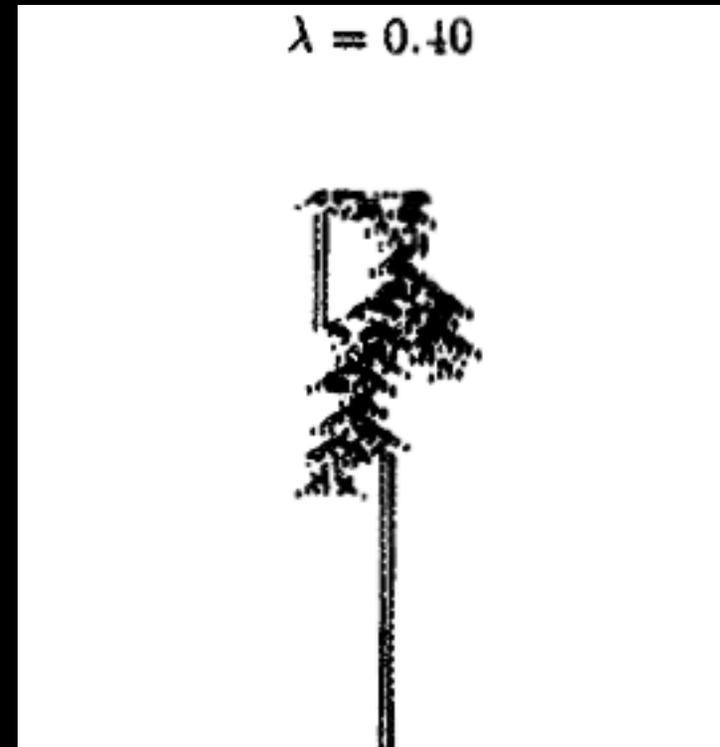
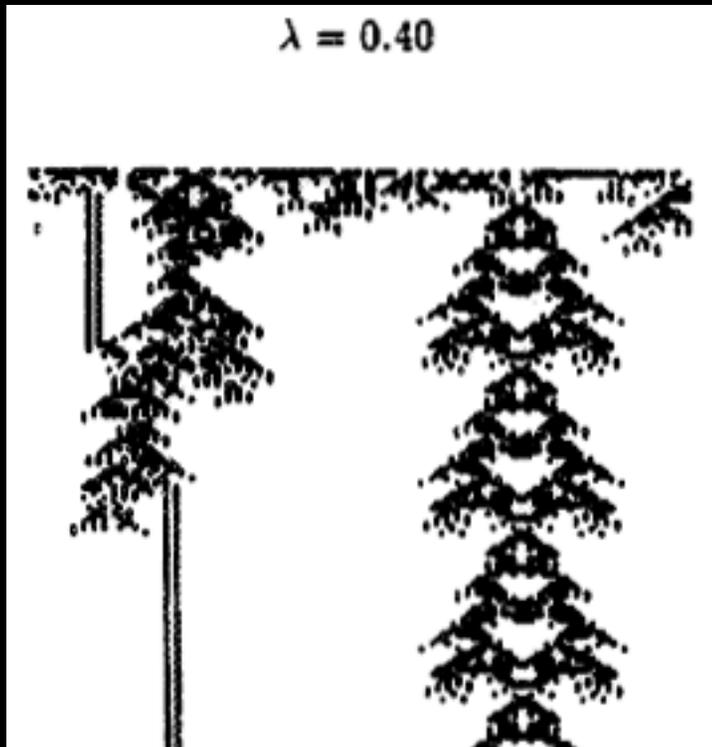
CA Dynamics, $\lambda = 0.35$

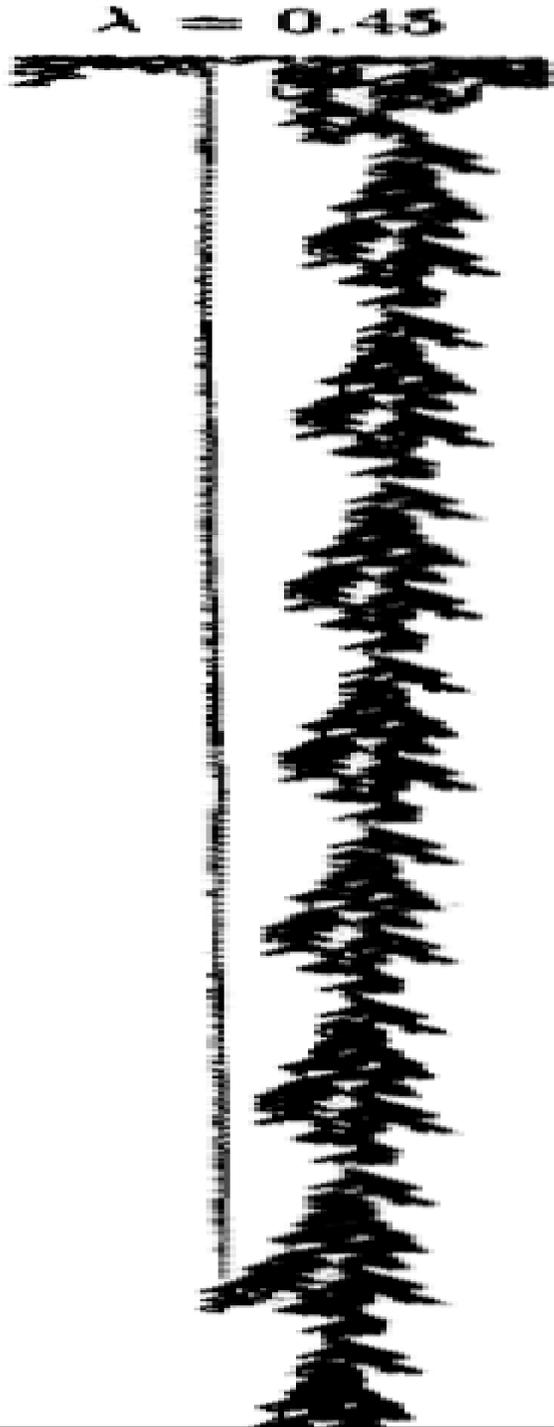
- Transients have lengthened; longer period structure has appeared; number of dynamical solutions is growing



CA Dynamics, $\lambda = 0.40$

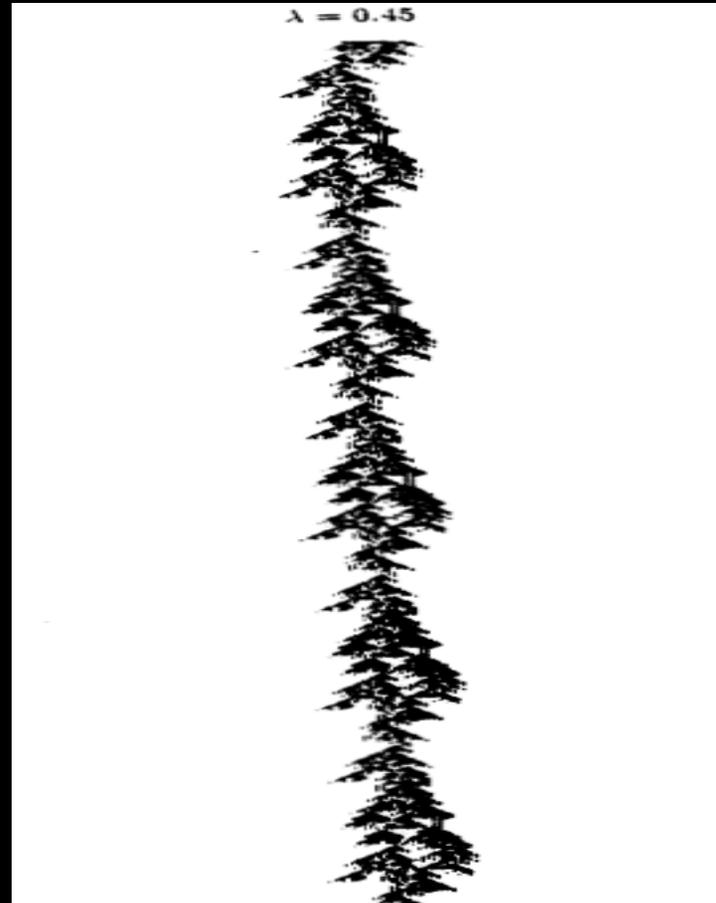
- Transients have lengthened to about 60 steps; structure of period 40 has appeared; dynamical activity still converges to periodic structures





CA Dynamics, $\lambda = 0.45$

- Transients have lengthened to almost 1,000 steps; true period of structure on left is 14,848 steps; dynamic area balanced between expansion and contraction



$\lambda = 0.50$



10,000 time steps



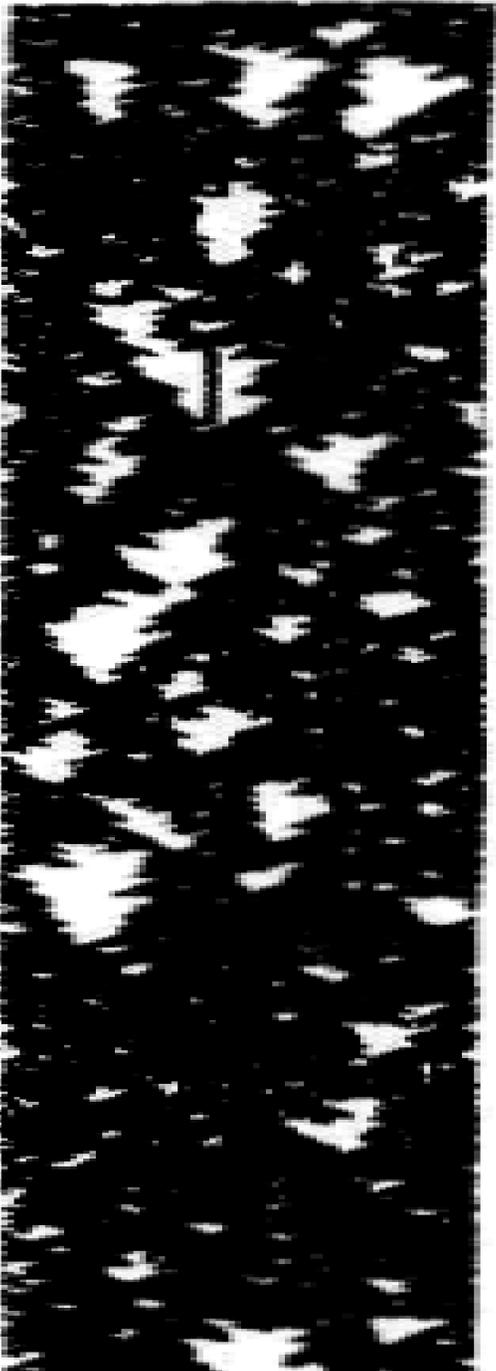
CA Dynamics, $\lambda = 0.50$

- Typical transient on order of 12,000 steps; dynamics may settle into periodic structure, but tendency is to expand

$\lambda = 0.50$



$\lambda = 0.55$



CA Dynamics, $\lambda = 0.55$

- Transients begin to shorten; dynamics become chaotic; arrow: site-occupation density = 1% of long-term average

$\lambda = 0.55$



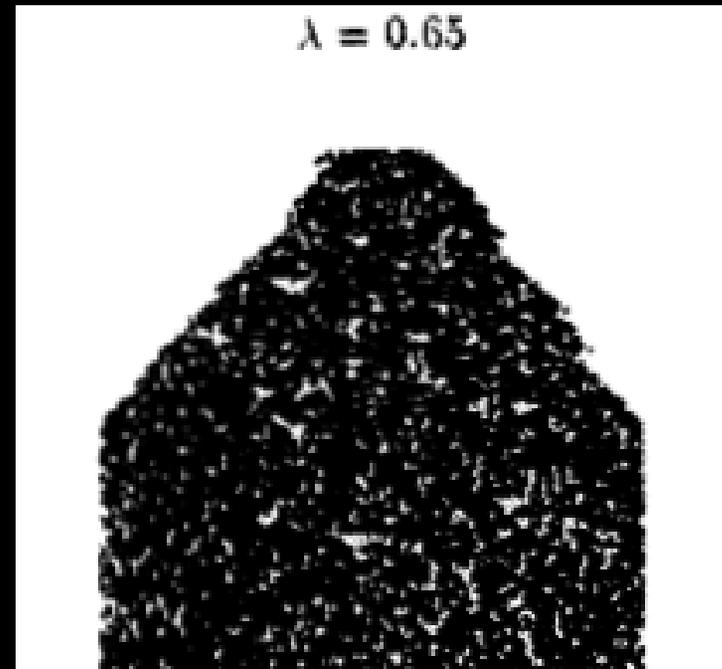
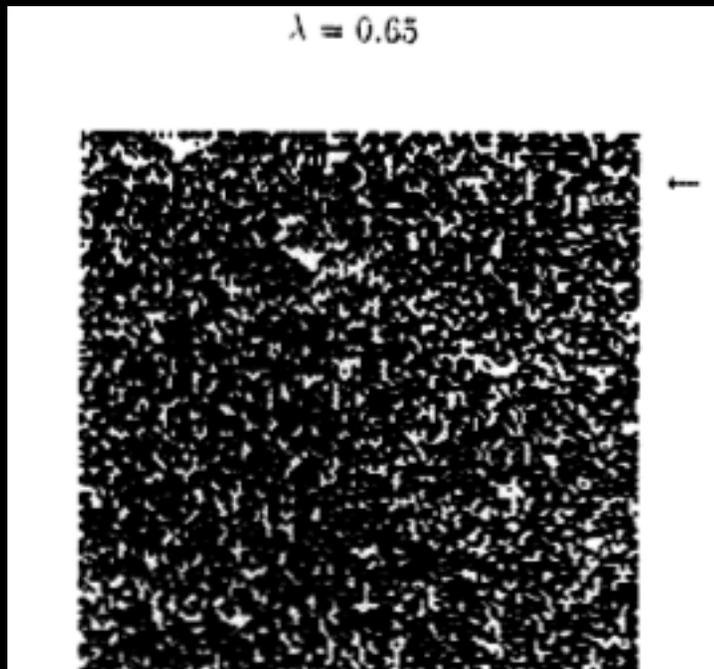
CA Dynamics, $\lambda = 0.60$

- Transient time to onset of chaos reduced; dynamics always chaotic; dynamical area expands more rapidly



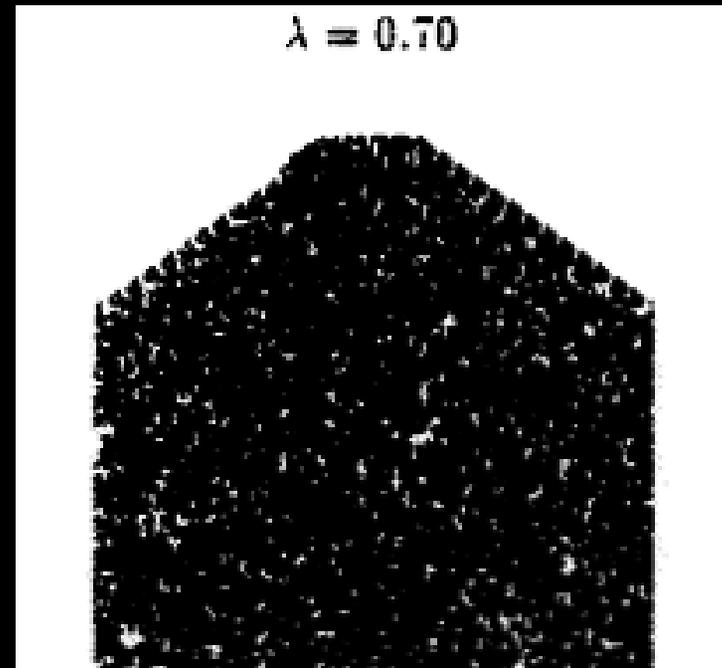
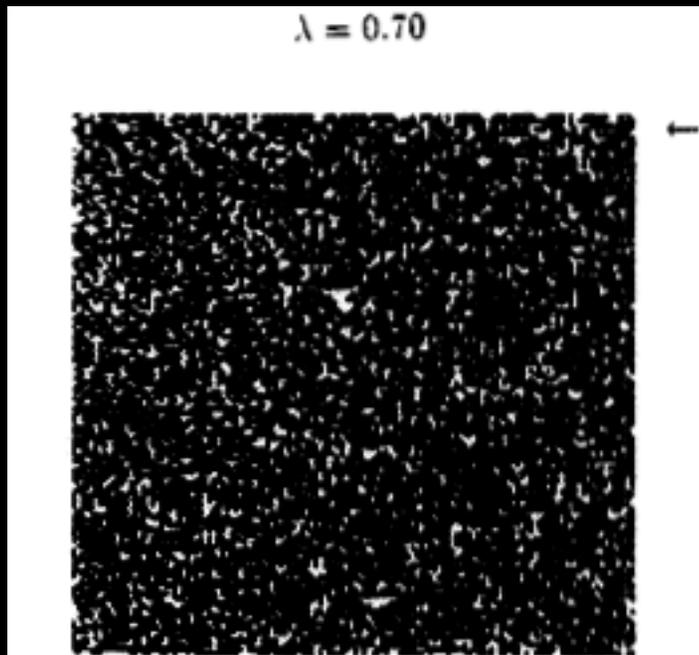
CA Dynamics, $\lambda = 0.65$

- Chaotic behavior achieved in only 10 steps or so; dynamical area expands 1 cell/step (1/2 max rate)



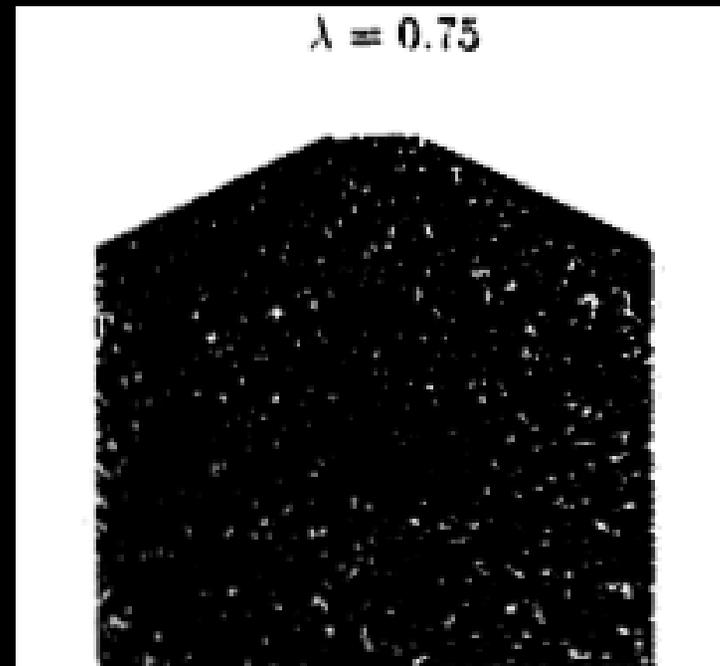
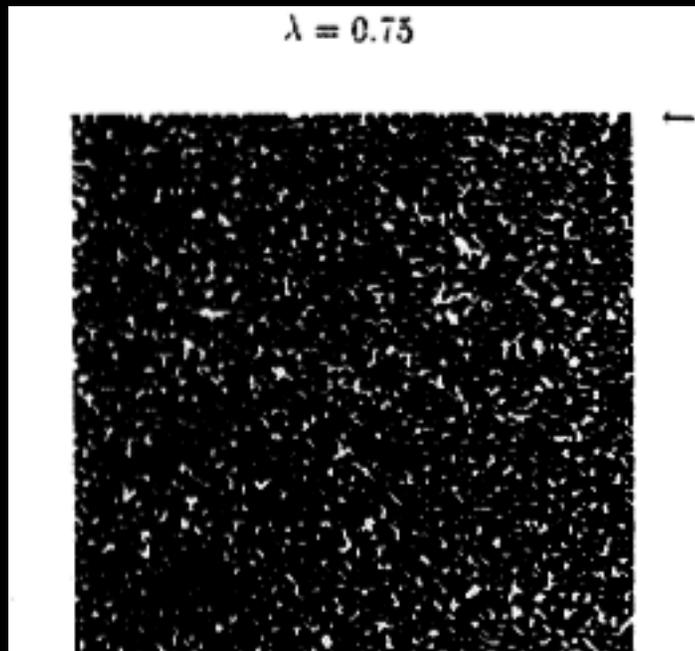
CA Dynamics, $\lambda = 0.70$

- Chaotic behavior achieved in just 2 steps; dynamical area expands even more rapidly



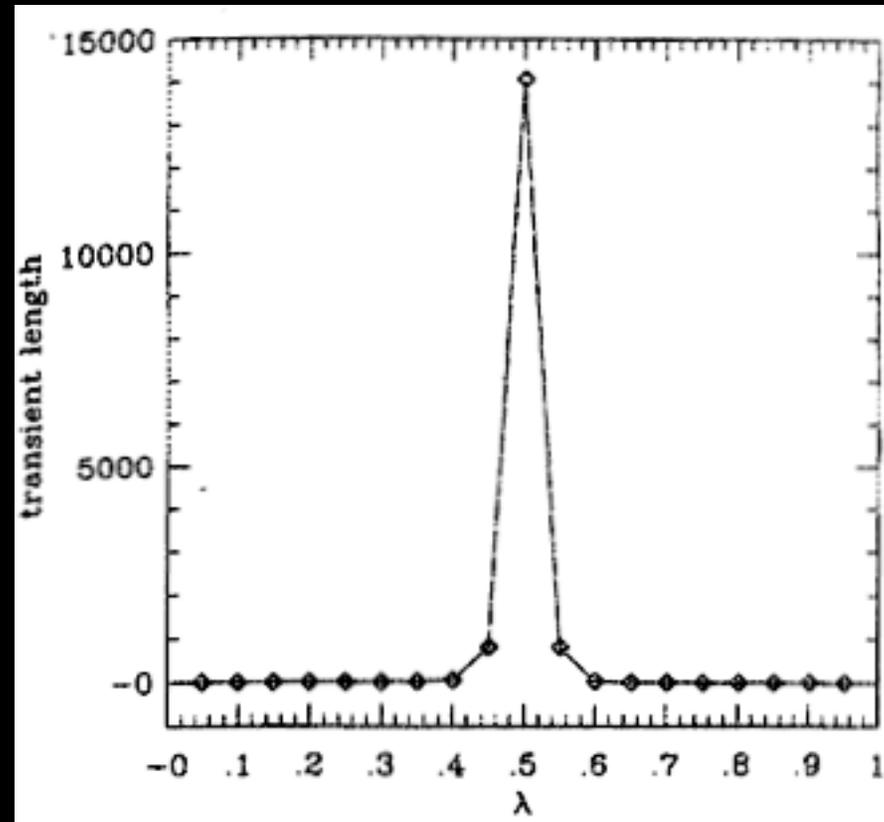
CA Dynamics, $\lambda = 0.75$

- Chaotic after a single step; dynamical area expands at maximum rate



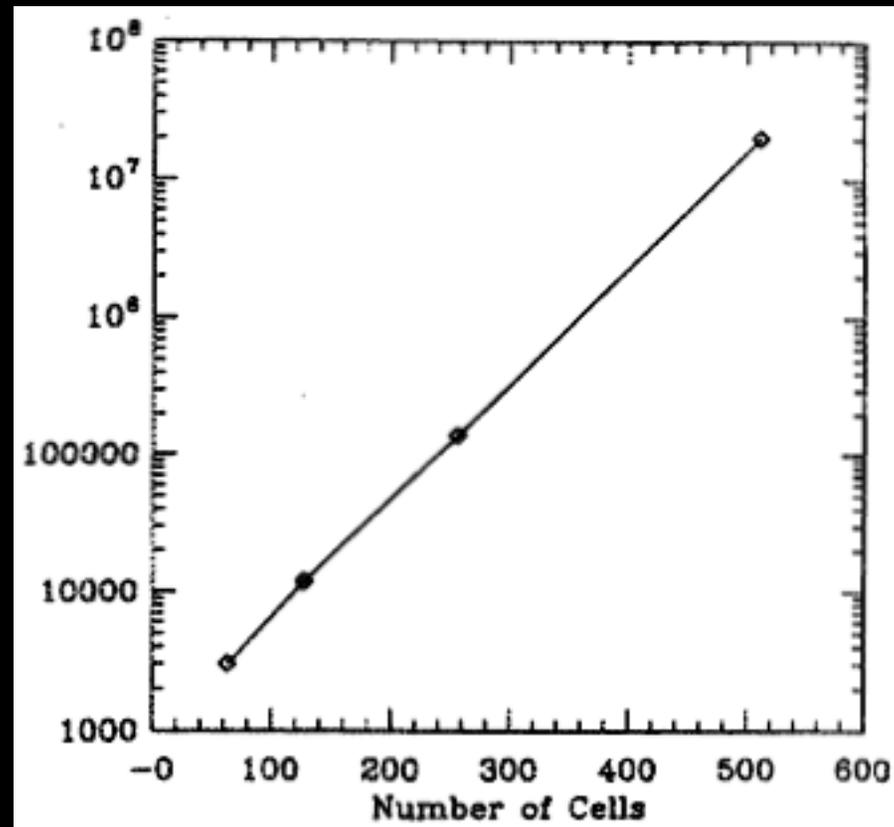
Transient Growth Rate

- Transients grow rapidly near the transition between ordered and disordered dynamics ("critical slowing down" phenomenon in study of phase transitions)



Dependency on Array Size

- No dependency for low or high λ , but at $\lambda = 0.5$ length of transients is exponential in array size:



Transition Region

- This critical transition region supports both static and propagating structures
- Langton likens the propagating, particle-like structures to solitary waves
 - Crutchfield and others have studied CA dynamics in terms of "particles" propagating along boundaries between dynamical regions
 - Somewhat like "gliders" in Conway Game of Life
 - $\lambda = 0.273$ for Life, which is in the transition region for 2D CAs with $K=2$, $N=9$

Complications and Observations

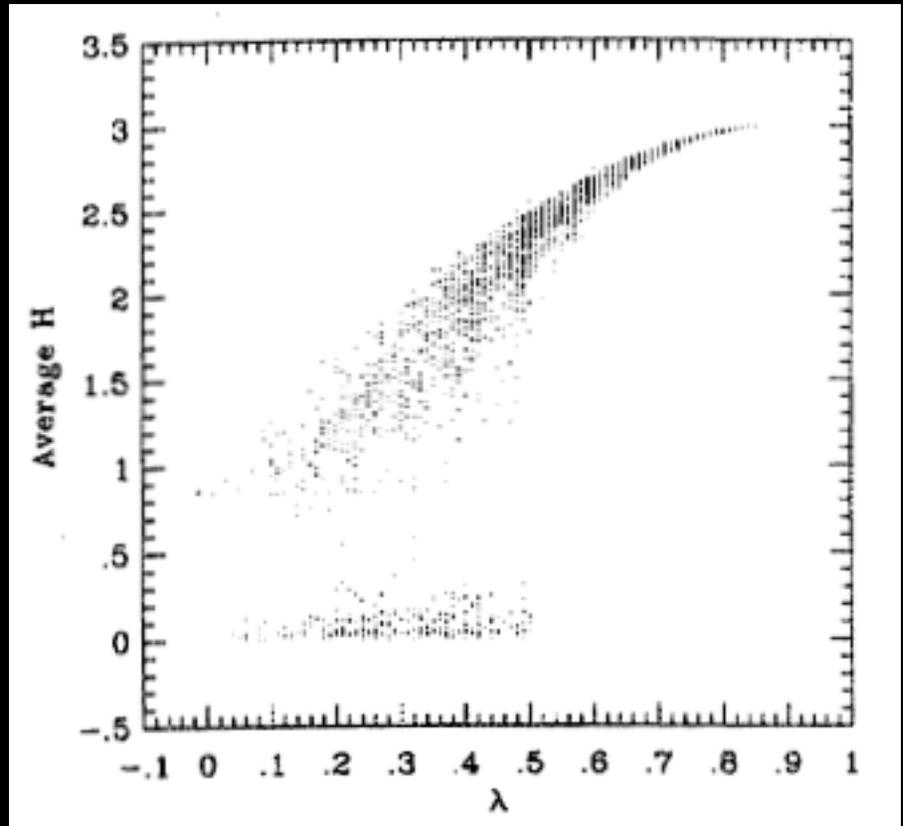
- Different traversals of λ space using the table-walk-through method make the transition to chaotic behavior at different λ values
 - However, there is a well defined distribution around a mean value
- Sometimes the transition from order to disorder is abrupt, suggesting both first- and second-order phase transitions are possible
- There is a clear phase transition between periodic and chaotic behavior, and the most complex behavior is found in the vicinity of the transition

Quantifying Dynamics

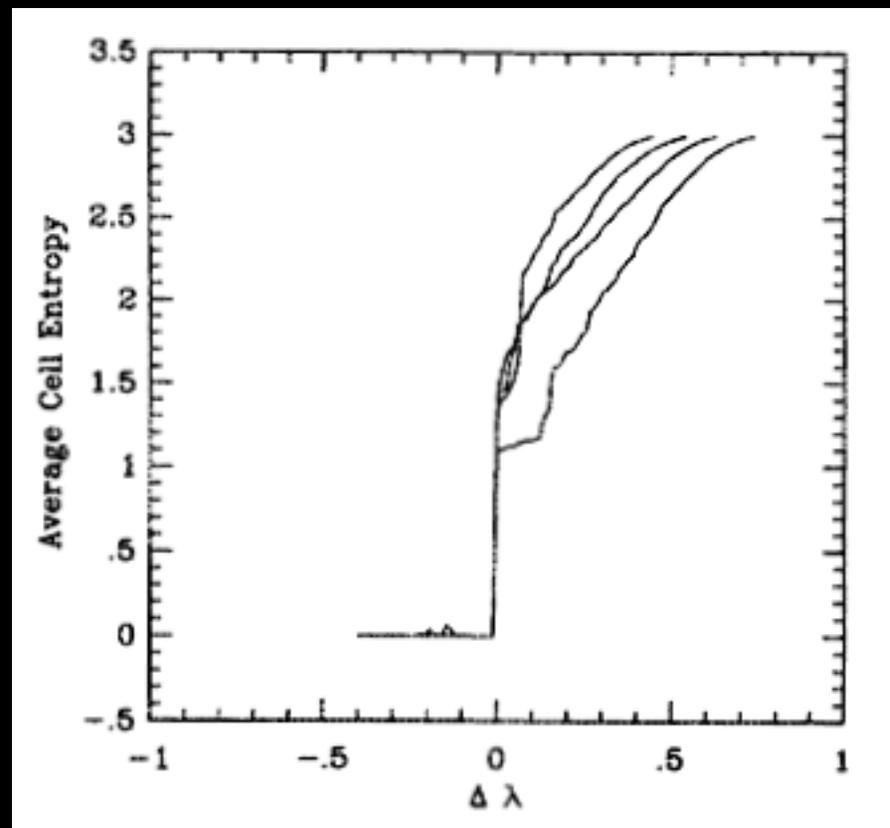
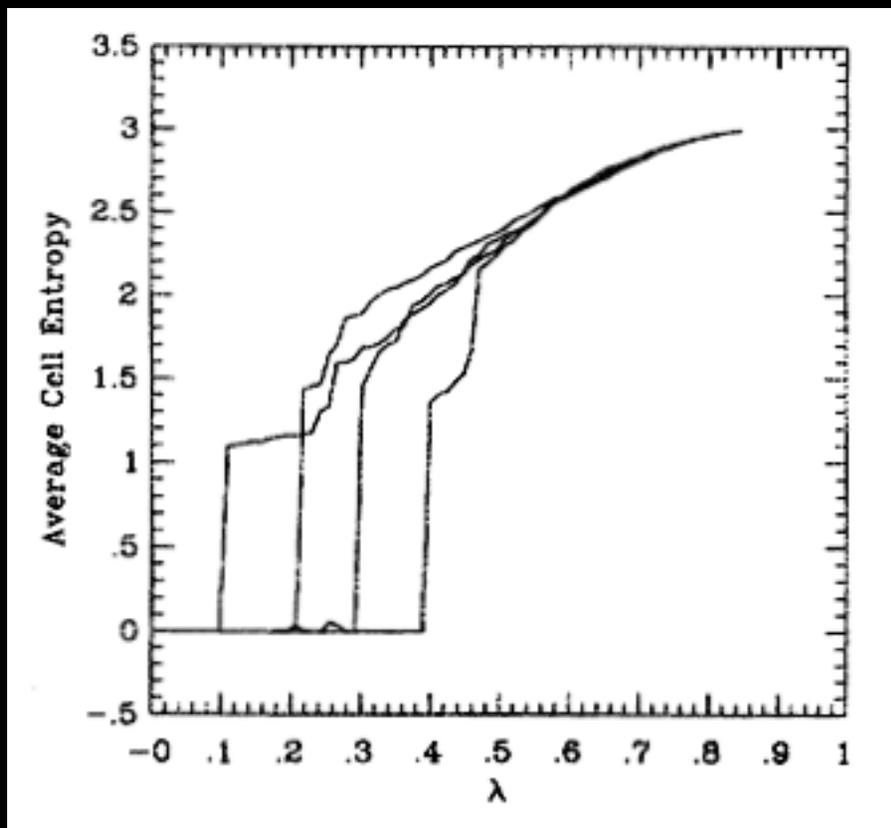
- Langton ran another series of experiments using 2D CAs with $K=8$ and $N=5$, on arrays of size 64×64 , with periodic boundary conditions
- He looks at average entropy per cell, as a measure of basic information capacity
- And average mutual information per cell, ultimately, as a measure of complexity

Information Capacity

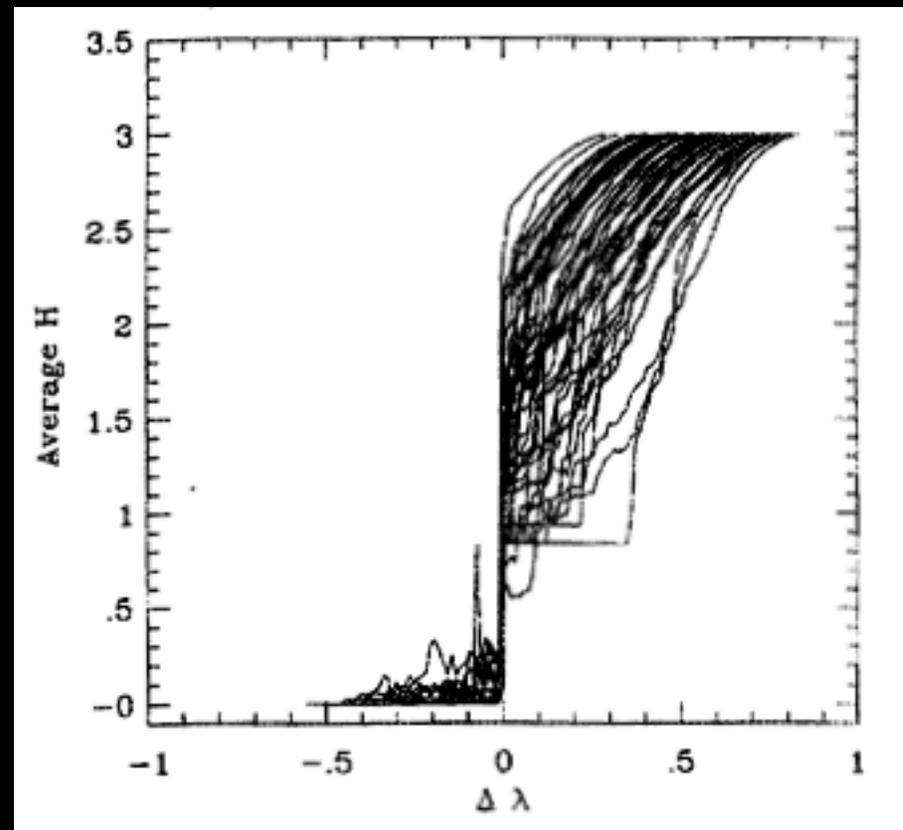
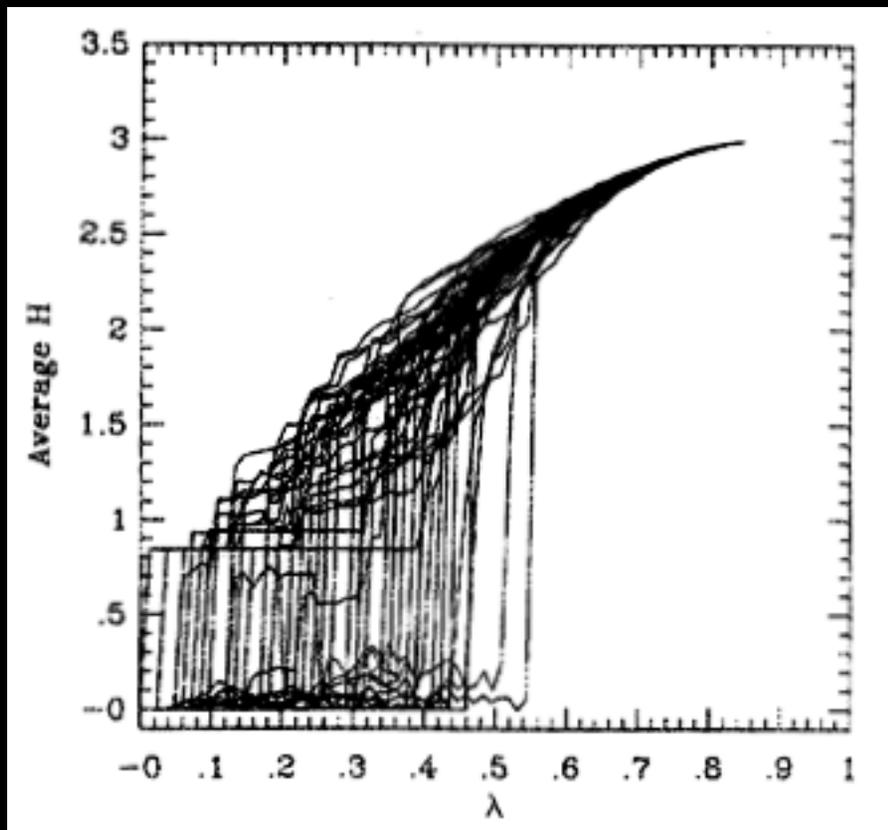
- Data gathered using random-table method
- Bimodal distribution suggests a phase transition
- Low H values stop at site-percolation threshold of neighborhood template
- (Other interesting characteristics in paper)



Entropy with Table-Walk-Through



Entropy with Table-Walk-Through

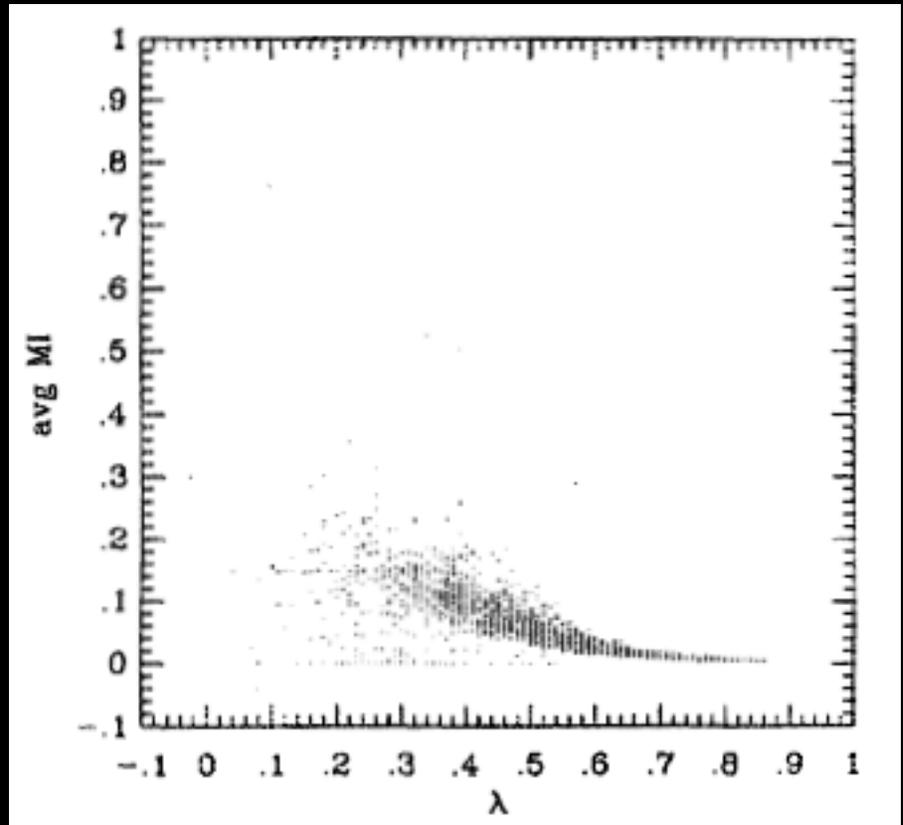


Mutual Information

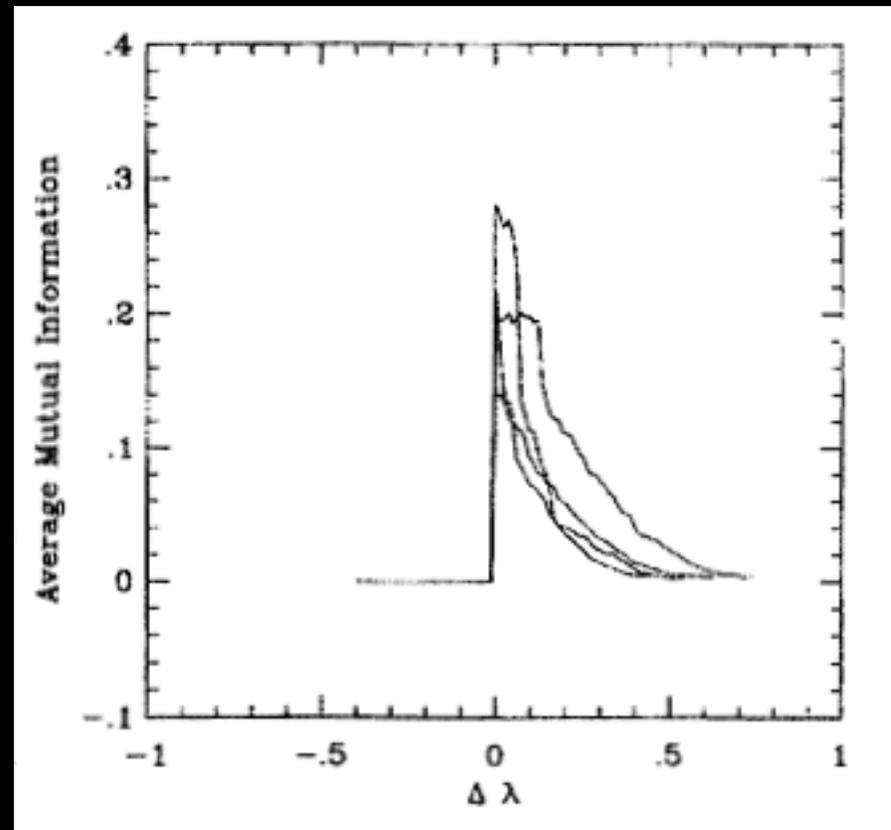
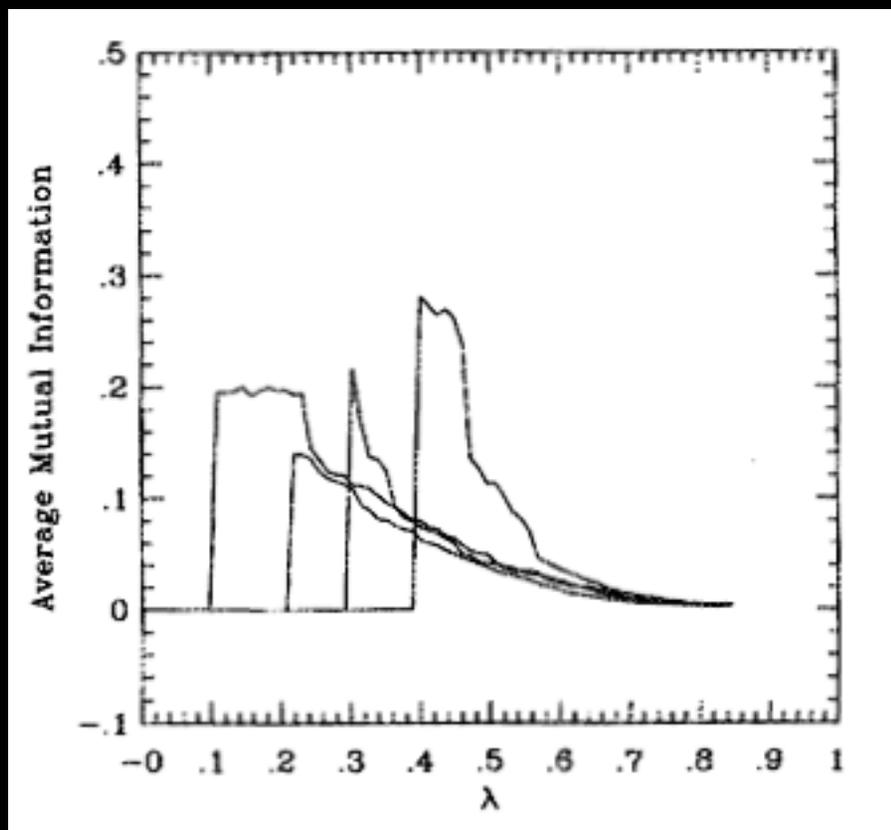
- In order for two cells to cooperate in support of computation, they must be able to affect each other's behavior
- Therefore we should be able to find correlations in activity between pairs of cells
- Mutual Information captures such correlations quantitatively

Mutual Information

- This is average MI between a cell and itself at the next time step
- Note that MI is low for both low and high λ values
- Growth of MI at intermediate λ values is evidence of increased correlation length (and a phase transition)

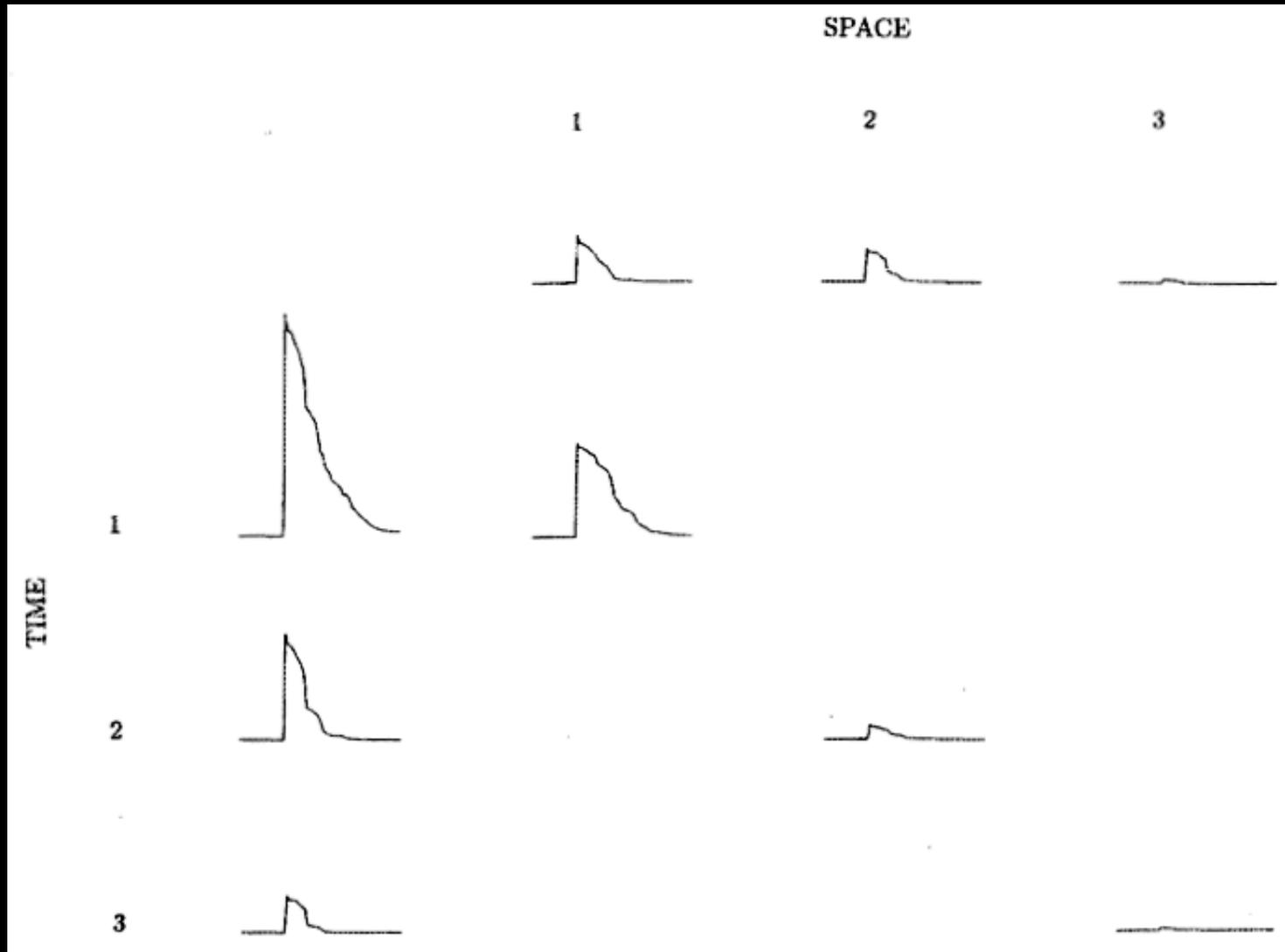


MI with Table-Walk-Through



Jump corresponds to onset of chaotic region
Decaying tail indicates transition to fully chaotic behavior

MI Decay Over Space and Time

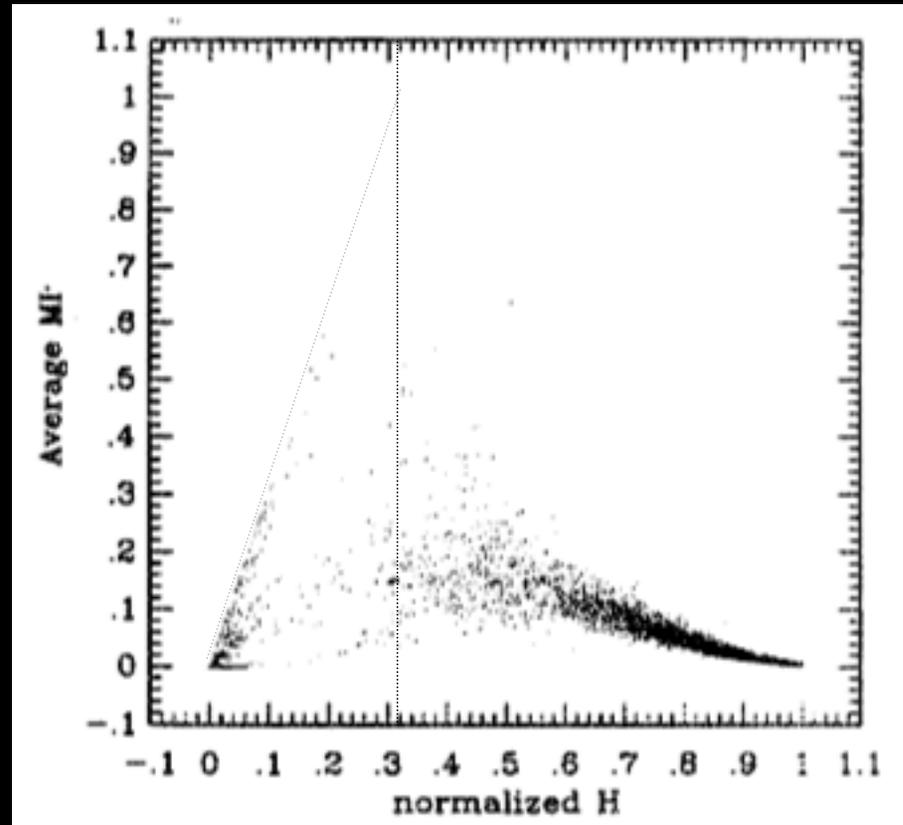


Cooperative Computation

- For cells to cooperate in computation, they must exhibit some, but not too much, correlation
- Too much and they merely mimic each other
- Too little and they behave completely independent of each other
- Just right correlation implies a kind of "common code, or protocol, by which changes of state in one cell can be recognized and understood by the other as a *meaningful signal*"

Mutual Information and Entropy

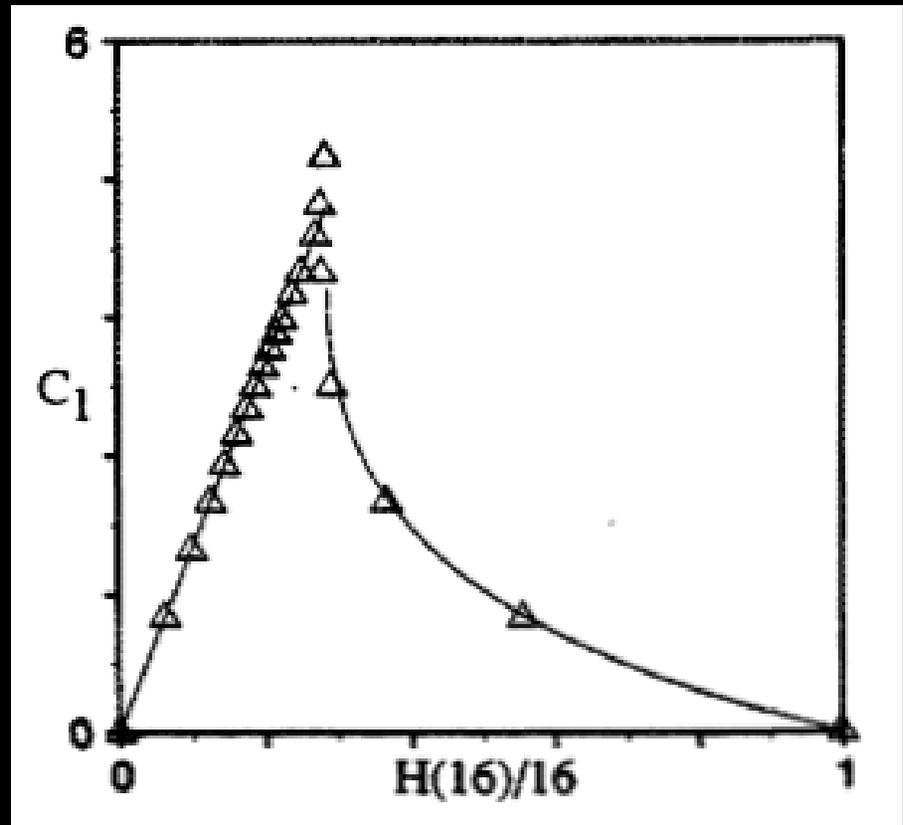
- There is a sharply defined maximum value of MI at a specific value of λ
- MI falls off rapidly on either side
- Optimal working entropy derives from tradeoff between information storage (lower entropy) and information transmission (higher entropy)



Note: Graphic was printed poorly; text makes it clear where transition occurs, so I added lines

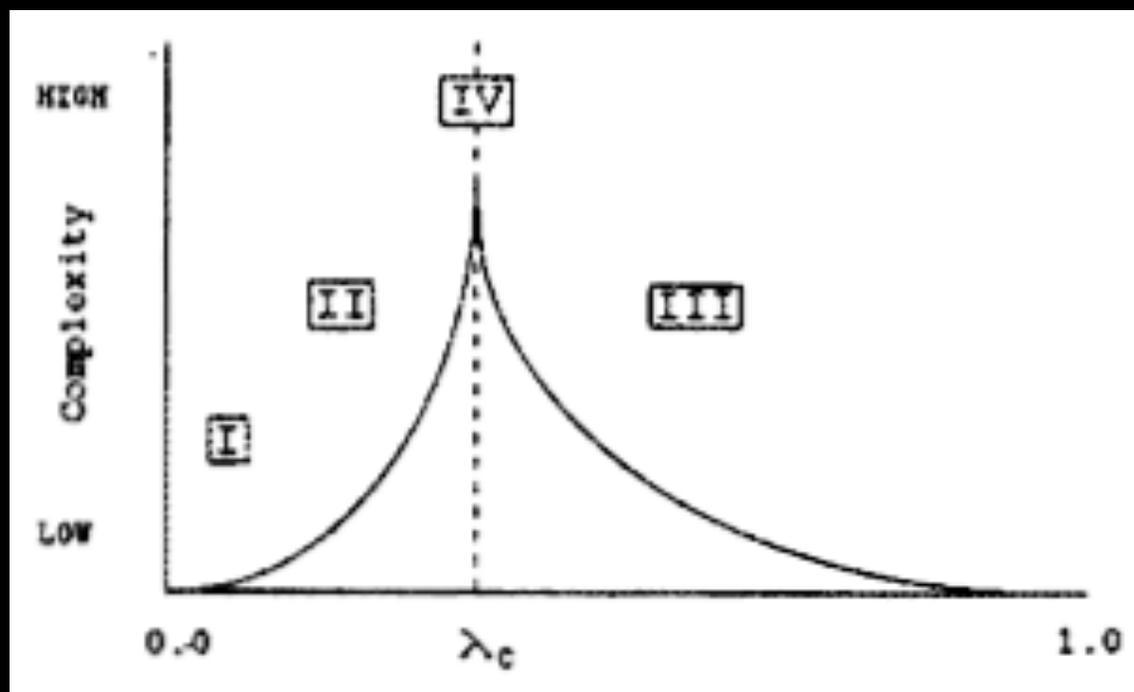
Mutual Information and Entropy

- Jim Crutchfield, then at Berkeley, produced similar results measuring complexity (machine size) versus entropy for finite state machines predicting binary strings



Locating the Wolfram Classes

- I = fixed
- II = periodic
- III = chaotic
- IV = complex



Dynamical Classes

- Langton suggests that the solid and fluid phases of matter represent "two fundamental universality classes of dynamical behavior"
- Until now we've only had common experience with these classes of behavior in matter, but in computers we are able to look at dynamics abstracted from any particular material substrate
- These two classes are separated by a phase transition, in the vicinity of which mechanisms exist for information storage, transmission, and modification, thus providing a capacity for emergent computation

The Edge of Chaos

- Langton draws on other research to suggest that dynamical systems must constantly balance a need for homeostasis with a need for dynamic variation
- Langton observes that life itself may have its origin in the kind of extended transients seen at the phase transition in dynamical systems, and that we may be "examples of the kind of 'computation' that can emerge in the vicinity of a phase transition given enough time"
- Thus computation and life itself exist at, and because of, the edge of chaos



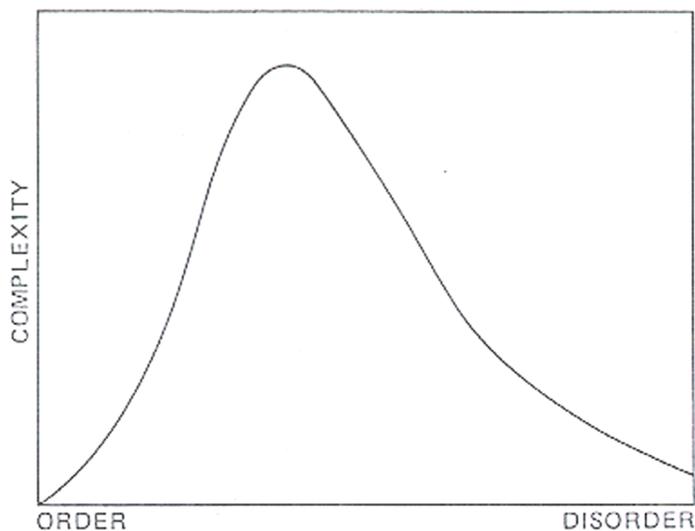
“What clashes here of wills gen wonts,
oystrygods gaggin fishygods! Brékkek
Kékkek Kékkek Kékkek! Kóax Kóax
Kóax! Ualu Ualu Ualu! Quáouauh!”



“Happy families are all alike; every unhappy
family is unhappy in its own way.”



“All work and no play makes Jack a dull boy.
All work and no play makes Jack a dull boy.
All work and no play makes Jack a dull boy.”

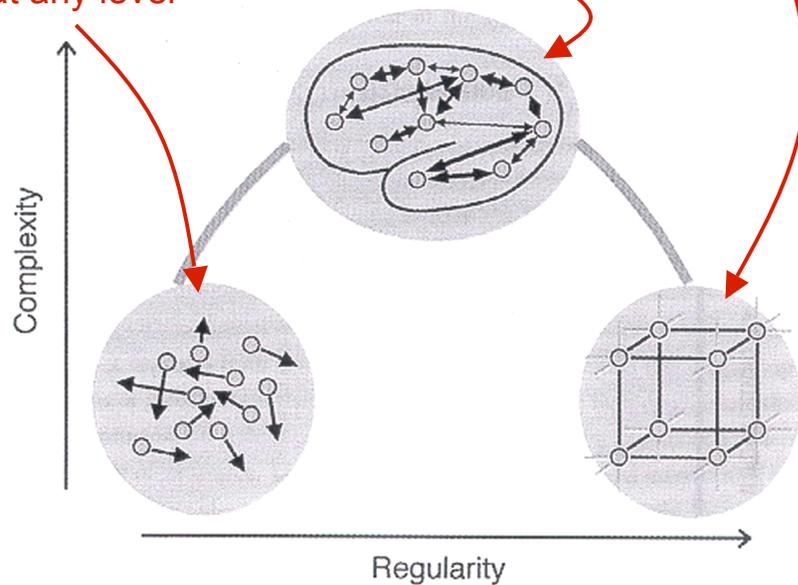


Reference:
B.A. Huberman and T.Hogg (1986) Physica 22D, 376.

randomness,
no structure at any level

non-repeating structure
at multiple levels

identical structure
at all levels



Reference:
G. Tononi, G.M. Edelman, O. Sporns (1998) TICS 2, 474.

Integration

Integration measures the statistical dependence among all elements $\{x_i\}$ of a system X .

$$I(X) = \sum_{i=1}^n H\{x_i\} - H(X) \quad MI(x_1, x_2) = H(x_1) + H(x_2) - H(x_1 x_2)$$

$H\{x_i\}$ is the entropy of the i^{th} individual element x_i
 $H(X)$ is the joint entropy of the entire system X

Note, $I(X) \geq 0$.

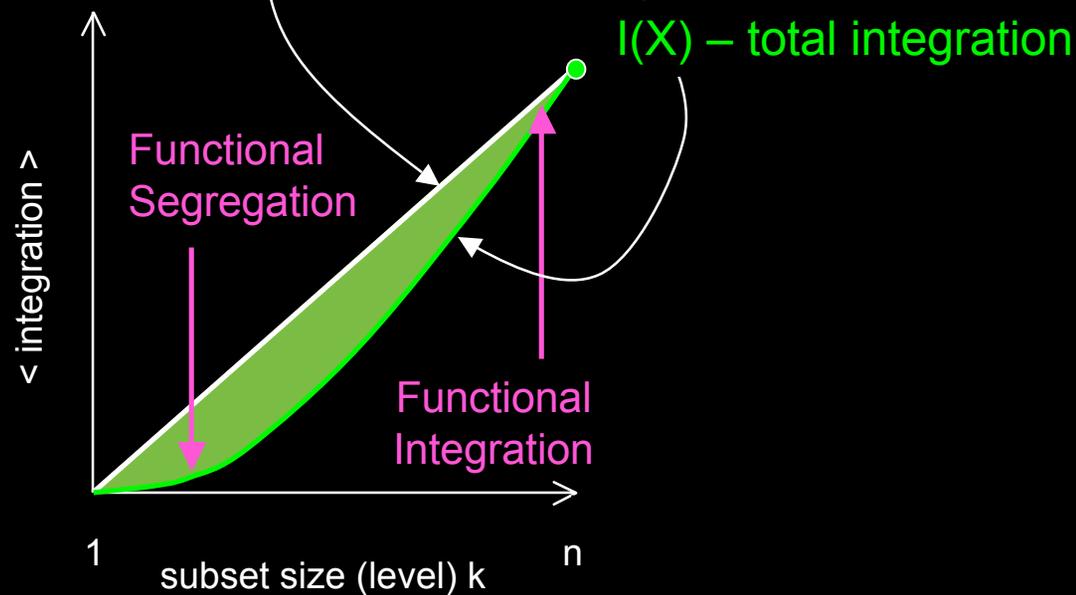
Note, $I(X) = 0$ if all elements are statistically independent

Any amount of structure (i.e. connections) within the system will reduce the joint entropy $H(X)$ and thus yield positive integration.

Information and Complexity

- **Complexity**, as expressed in terms of the ensemble average of integration (structure) at all levels:

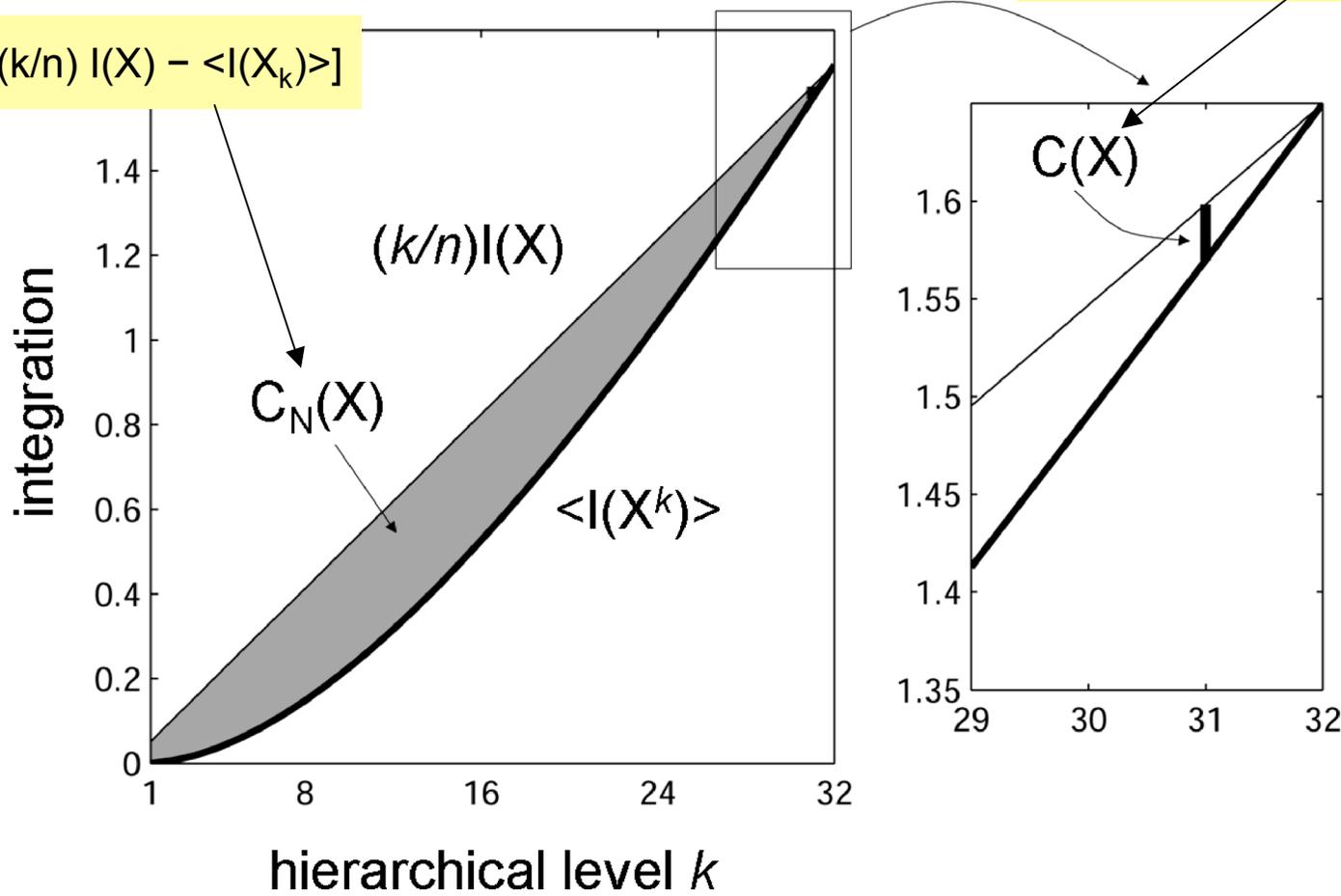
$$C_N(X) = \sum_{k=1}^n [(k/n) I(X) - \langle I(X_k) \rangle] = \sum_{k=1}^{n/2} \langle MI(X_k; X - X_k) \rangle$$



Simpler Complexity

$$\begin{aligned}
 C(X) &= H(X) - \sum_i H(x_i | X - x_i) \\
 &= \sum_i MI(x_i, X - x_i) - I(X) \\
 &= (n-1)I(X) - n \langle I(X - x_i) \rangle
 \end{aligned}$$

$$C_N(X) = \sum_{k=1}^n [(k/n) I(X) - \langle I(X^k) \rangle]$$



More Complicated Complexity

- Tononi's Phi (see extras) partitions all subsets of a network, treating each partition in turn as a source of noise, and computing essentially the complexity we just looked at (for each partition)
- Tononi suggests that Phi provides a formal theory explaining many empirical observations with regard to perception and consciousness
 - "The theory entails that consciousness is a fundamental quantity, that it is graded, that it is present in infants and animals, and that it should be possible to build conscious artifacts."

Credits

- Lecture based largely on the reading assignment:
Langton, C. G., *Computation at the Edge of Chaos: Phase Transitions and Emergent Computation*, p. 12-37, *Emergent Computation, Proceedings of the Ninth Annual International Conference of the Center for Nonlinear Studies on Self-organizing, Collective, and Cooperative Phenomena in Natural and Artificial Computing Networks*, Los Alamos, NM, 1989, ed. Stephanie Forrest, North Holland, 1990

References

- Site-percolation thresholds clearly defined in Appendix 6 of Peter Meyer's *Computational Studies of Pure and Dilute Spin Models* at <http://www.hermetic.ch/compsci/thesis/app6.htm>
Entire thesis at <http://www.hermetic.ch/compsci/thesis/>