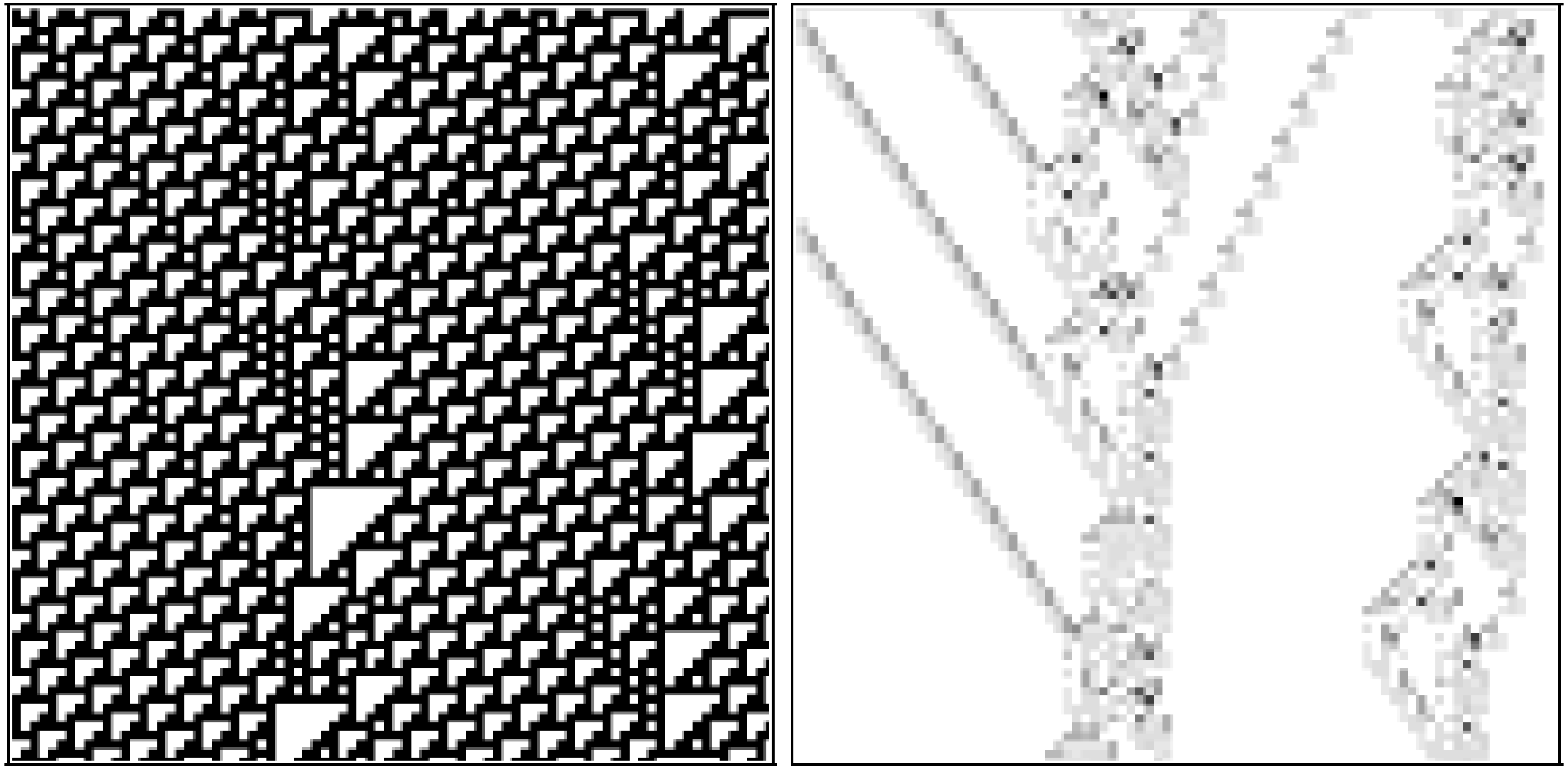


Detecting non-trivial computation in complex dynamics

Joseph T. Lizier[1,2], Mikhail Prokopenko[1], Albert Y. Zomaya[2]

1. CSIRO ICT Centre; 2. School of IT, The University of Sydney

Entropy Discussion Group, 5th March 2008

CSIRO

Detecting non-trivial computation in complex dynamics

Joseph T. Lizier[1,2], Mikhail Prokopenko[1], and Albert Y. Zomaya[2]

[1] CSIRO Information and Communications Technology Centre, Locked Bag 17, North Ryde, NSW 1670, Australia

[2] School of Information Technologies, The University of Sydney, NSW 2006, Australia

jlizier@it.usyd.edu.au

Abstract. We quantify the local information dynamics at each spatiotemporal point in a complex system in terms of each element of computation: information storage, transfer and modification. Our formulation demonstrates that information modification (or non-trivial information processing) events can be locally identified where "the whole is greater than the sum of the parts". We apply these measures to cellular automata, providing the first quantitative evidence that collisions between particles therein are the dominant information modification events.

- *Aim*: Demonstrate how to quantify distributed computation in terms of local information dynamics.
- *Application*: Prove long held conjectures about distributed computation in cellular automata

Contents

1. Background:
    1. Information dynamics in complex systems
    2. Computation in CAs
    3. Information theory
2. Derive metrics for local information dynamics
3. Apply to CAs, proving conjectures about computation therein.

Information dynamics in complex systems

- Information storage:
  - Maximising → fast locomotion in modular robotic system
- Information transfer:
  - Conjectured to be of importance in the vicinity of order-chaos phase transitions (maximised or intermediate?)
  - Empowerment (Klyubin et al): maximising channel capacity of perception-action loop → necessary structure.
  - Wavefront propagation in Belousov-Zhabotinsky media
- Information modification
  - Said to be maximised at phase transitions (Kinouchi)
  - Complexity said to be equivalent to capability for universal computation.

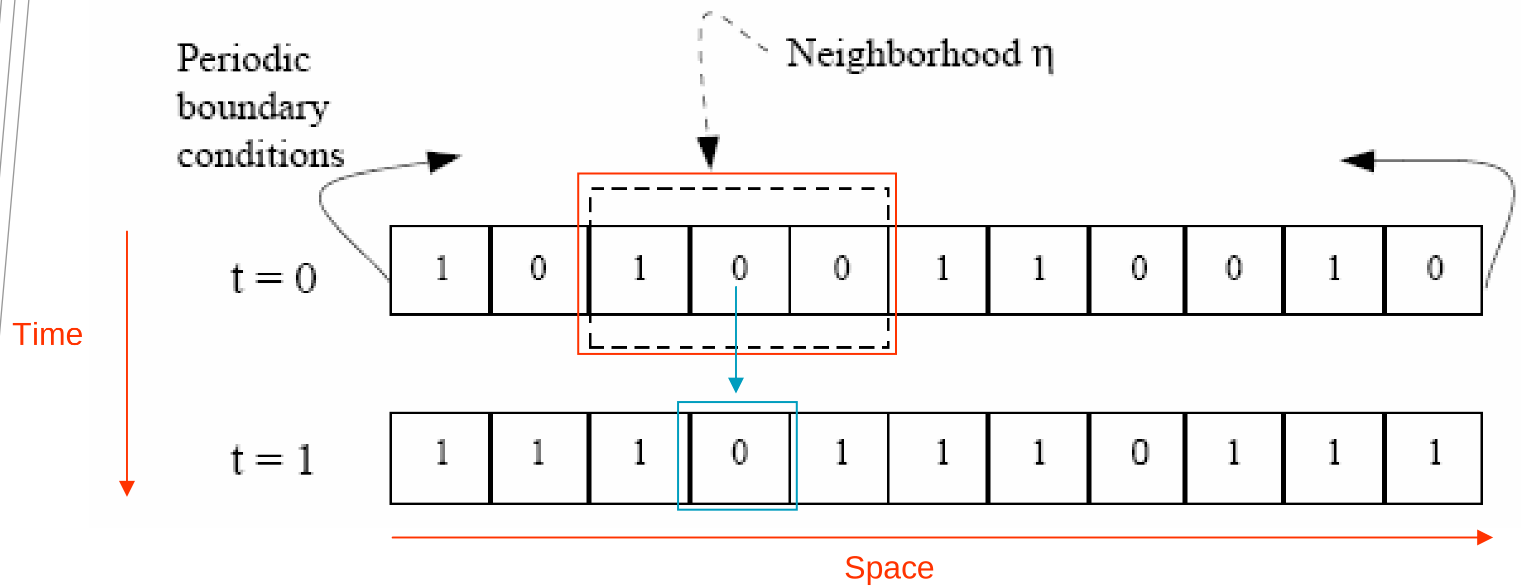
Cellular Automata – micro-level rules

Rule table ϕ:

neighborhood:	000	001	010	011	100	101	110	111
output bit:	0	1	1	1	0	1	1	0

= Rule 0x6e = Rule 110

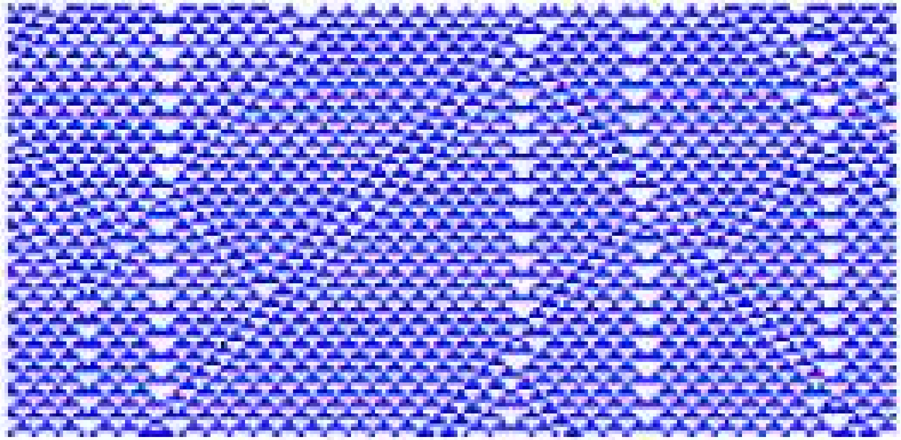
Lattice:



- "Computation in Cellular Automata: A selected review", Mitchell, 1998

Cellular Automata – emergent structure

- “Classifying Cellular Automata Automatically …”, Wuensche, 1999

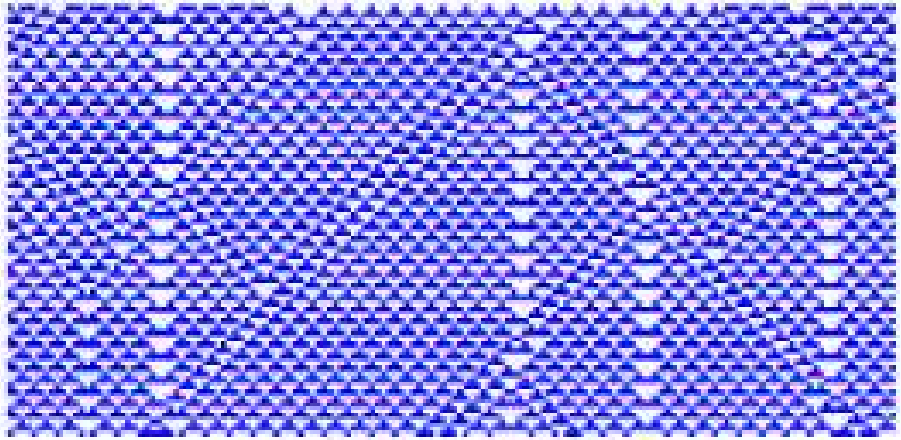


cells by value

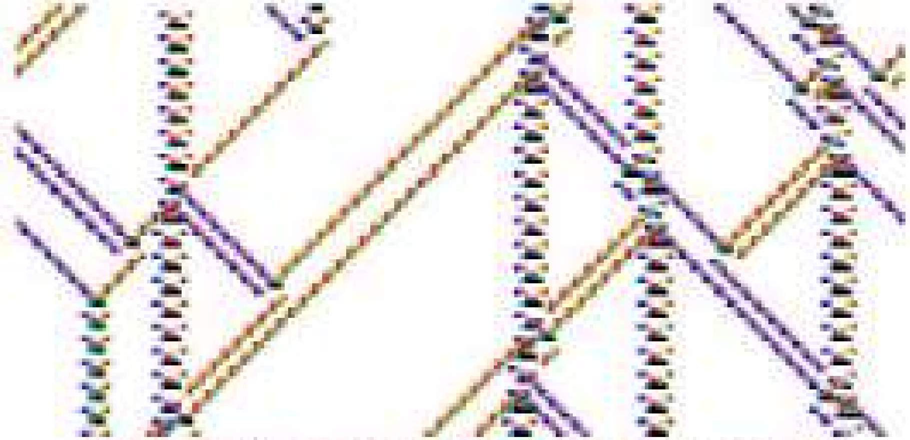
- Existing filtering methods to highlight emergent structure:
  - Computational dynamics (finite state transducers) (Crutchfield and Hanson)
  - Frequency of rule execution (Wuesnche)
  - Local statistical complexity and sensitivity (Shalizi et al)
  - Local information (really local spatial entropy rate) (Helvik et al)

Cellular Automata – emergent structure

- "Classifying Cellular Automata Automatically …", Wuensche, 1999



cells by value



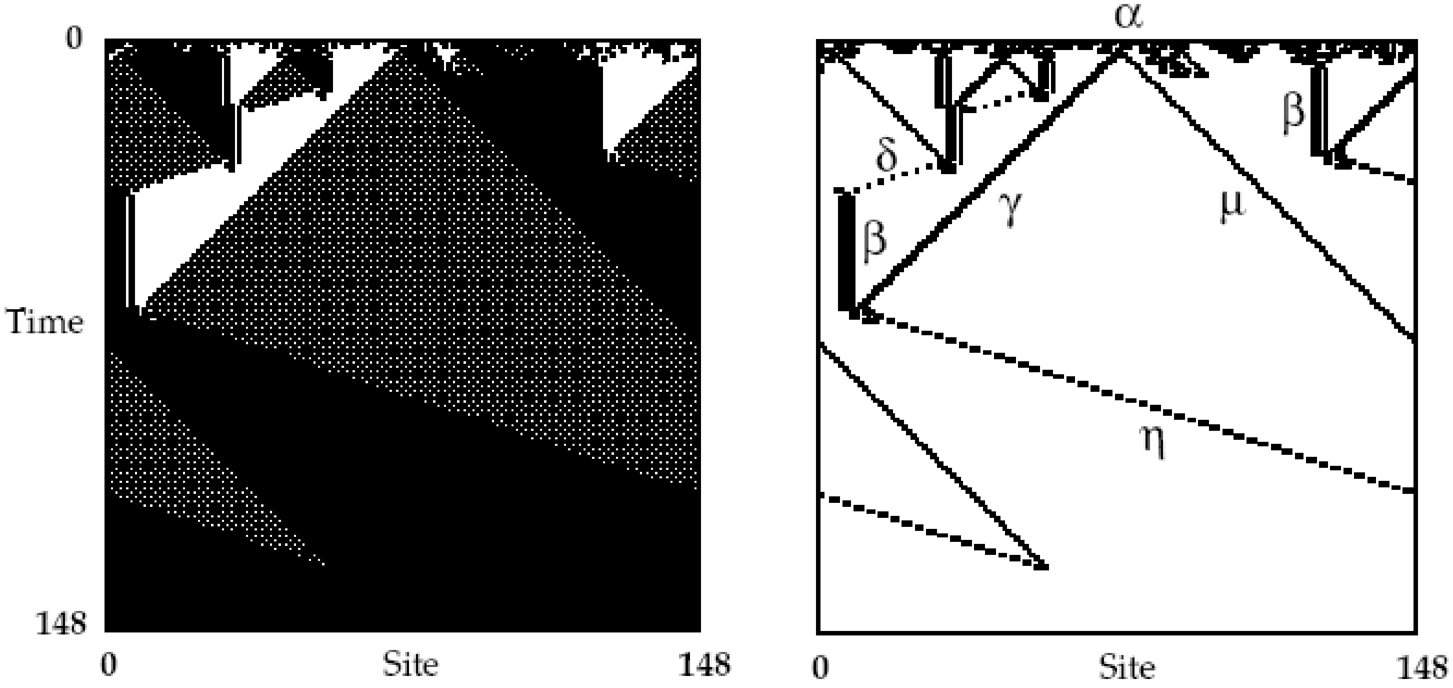
cells by look-up and filtered

- Emergent structure:
  - Domain — Blinkers
  - Particles
    - Gliders, Domai
  - Collisions

- Conjectured to represent:
  - Information storage
  - Information transfer
    - "
  - Information modification

No quantified evidence !!

Understanding computation in CAs



- Mitchell et al (1994,1996) used GAs to evolve CAs to solve specific computational tasks.
- In attempting the density classification task (above), the CA uses:
  - blinkers (β) to store information
  - gliders (γ,η) to transmit information
  - glider collisions (γ+ β →η) to modify/process information

Information-theoretical preliminaries

- Shannon entropy

$$H(X) = -\sum_{x \in A_X} p(x) \log p(x)$$

- Joint entropy

$$H(X,Y) = -\sum_{xy \in A_X A_Y} p(x,y) \log p(x,y)$$

- Conditional entropy

$$H(X|Y) = -\sum_{xy \in A_X A_Y} p(x,y) \log p(x|y)$$

- Mutual information

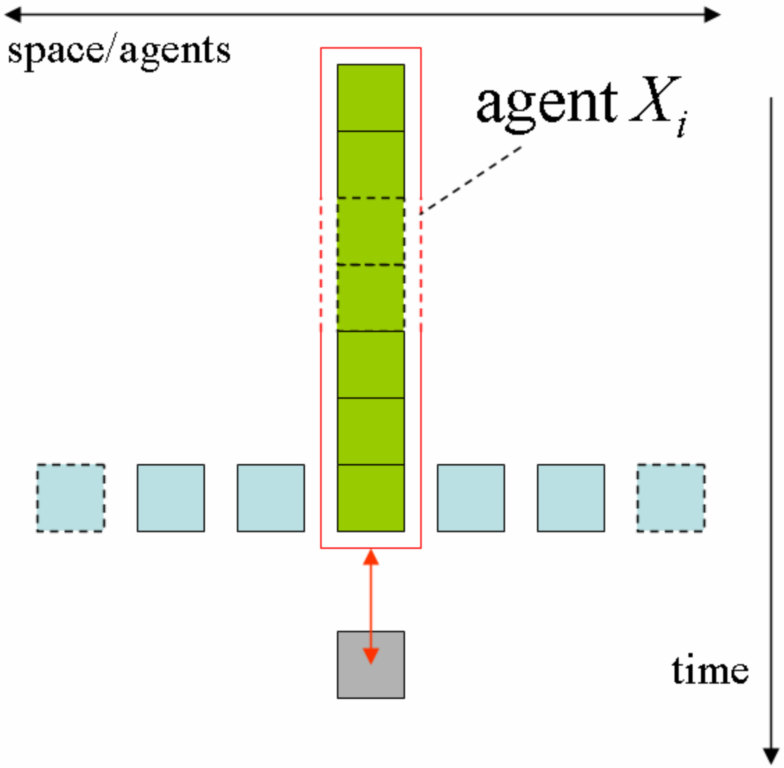
$$I(X;Y) = \sum_{xy \in A_X A_Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

- Conditional mutual information

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z) = H(Y|Z) - H(Y|X,Z)$$

Information storage



- Excess entropy captures average *total* information storage.
- Introduce active information storage to quantify average storage *currently in use*.
- *Local* active information storage = storage currently in use *at a given space-time point*:

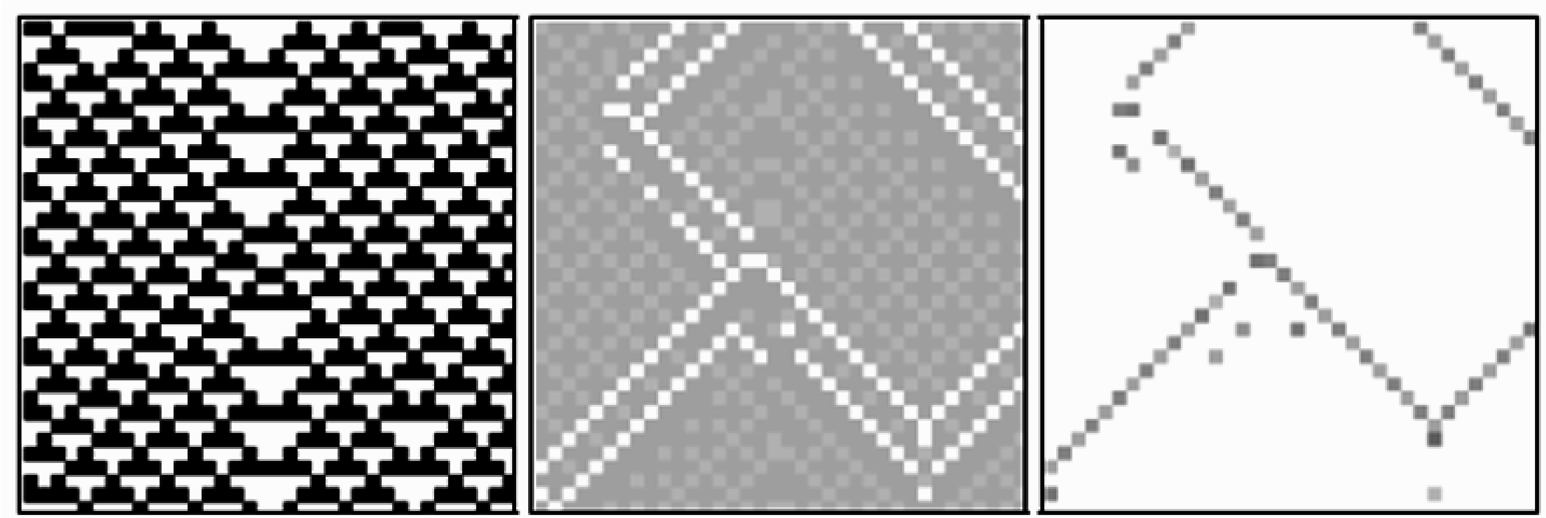
$$a(i, n+1) = \lim_{k \to \infty} \log \frac{p(x_{i,n}^{(k)}, x_{i,n+1})}{p(x_{i,n}^{(k)})p(x_{i,n+1})}$$

- a(i,n+1) > 0: past informs an observer about next state = strong information storage
- a(i,n+1) < 0: past *misinforms* an observer about next state (outcome was relatively unlikely)

Application to CAs

- Using Elementary CAs (ECAs)
- 10 000 cells, periodic boundary conditions
- First 30 time steps eliminated to allow CA to settle.
- Next 600 time steps kept for estimate of probability distribution functions for $a(i,n,k=16)$, and other measures.
- Local information dynamics measured at each space-time point
- All results confirmed by several CA runs.

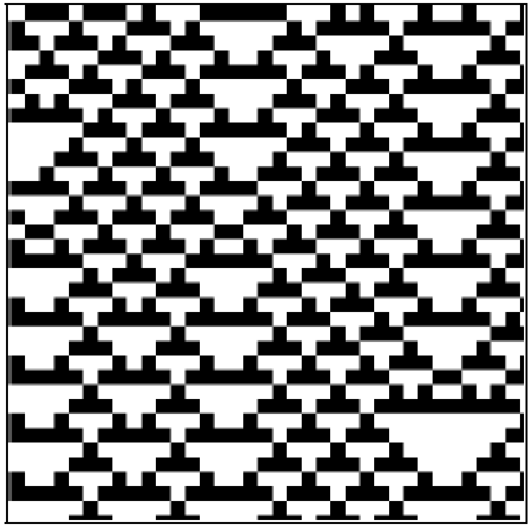
Results: Active information storage



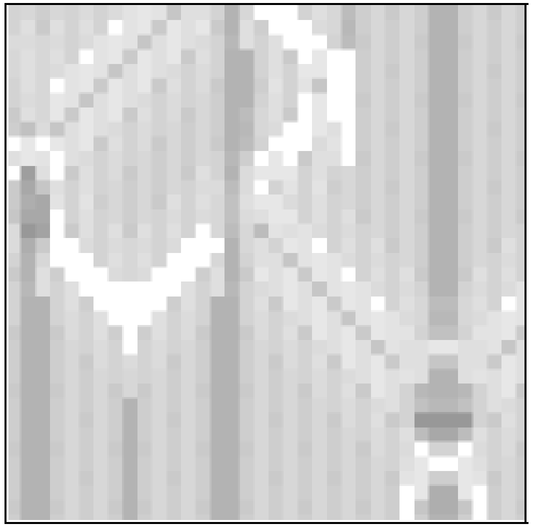
ECA rule 54 | a(i,n,k=16) > 0 | a(i,n,k=16) < 0

- Blinkers and domains are information storage elements.
- Local active information storage is misinformative at gliders (would be a good filter).

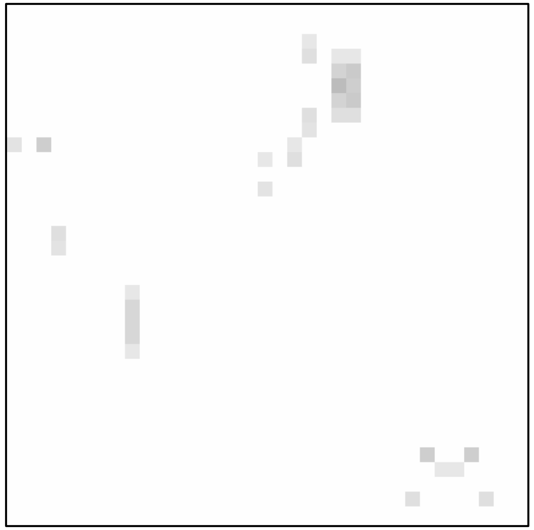
Results: Excess entropy



ECA rule 54



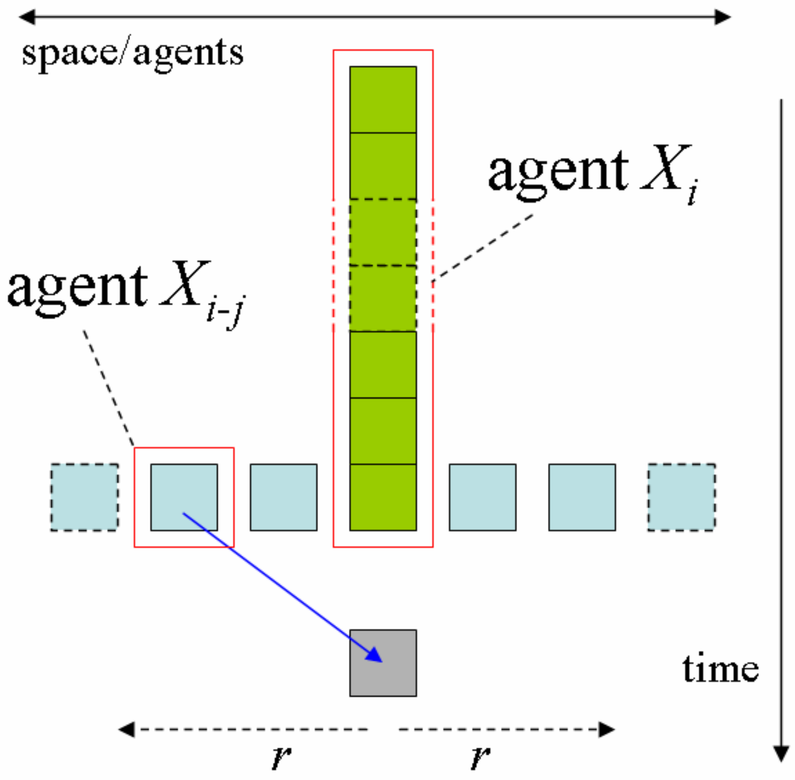
e(i,n,k=8) > 0



e(i,n,k=8) < 0

- Reveals blinkers as storing more information in total than domain.
- More loosely tied to the dynamics at the local space-time point than active information.

Information transfer

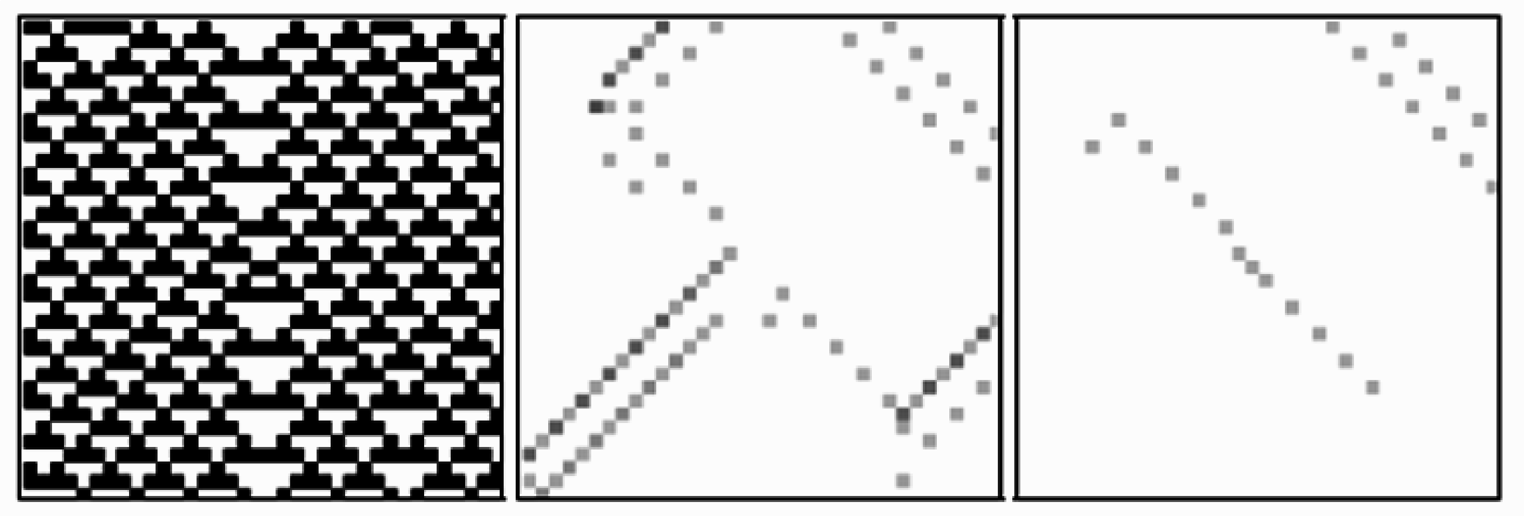


- Transfer entropy captures information transfer between a source and destination: info added by source about destination that was not contained in destination's past.
- *Local* transfer entropy = information transfer *at a given space-time point*

$$t(i,j,n+1) = \lim_{k\to\infty} \log \frac{p(x_{i,n+1}|x_{i,n}^{(k)}, x_{i-j,n})}{p(x_{i,n+1}|x_{i,n}^{(k)})}$$

- t(i,j,n+1) > 0: source *i-j* is informative about next state of *i* = strong information transfer
- t(i,j,n+1) < 0: source *misleads* an observer about next state in context of past (outcome was relatively unlikely)

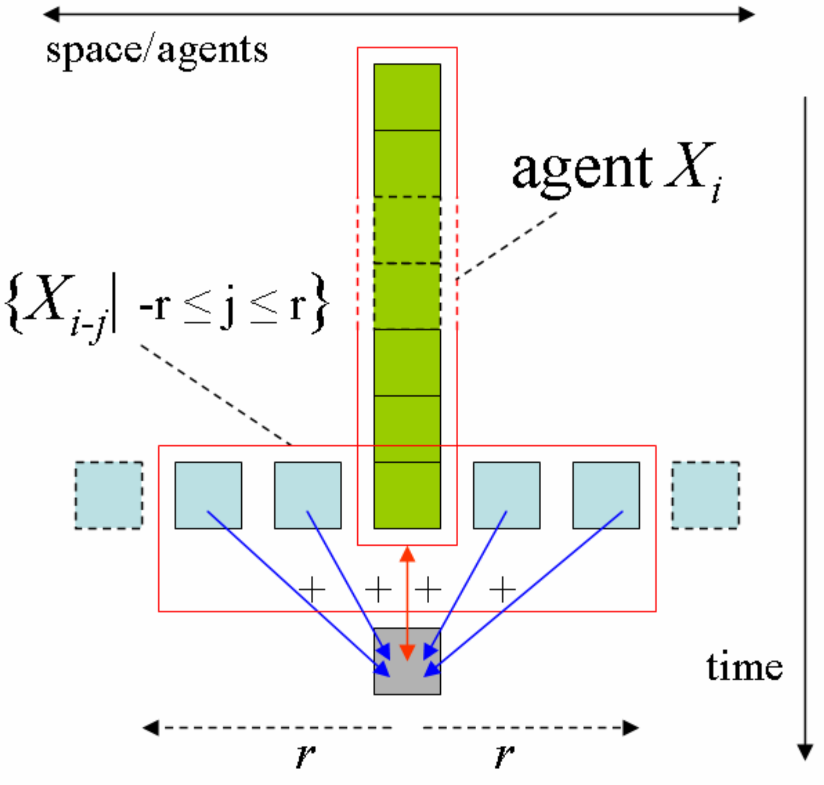
Results: Information transfer



ECA rule 54 | t(i,j=-1,n,k=16) > 0 | t(i,j=-1,n,k=16) < 0

- Gliders are dominant information transfer agents in their direction of motion.
- Local transfer entropy is mis-informative in reverse direction to gliders.

Total information



- Total information to predict the next state:

$$h(x_{i,n+1}) = a(i,n+1,k) + h(x_{i,n+1}|x_{i,n}^{(k)}).$$

Active information

Entropy rate

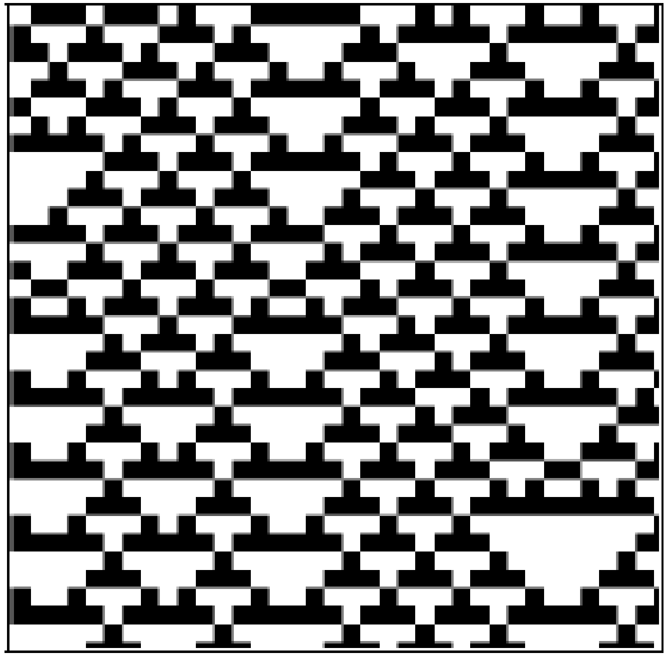
$$h(x_{i,n+1}|x_{i,n}^{(k)}) = t(i,n+1,k) + h(x_{i,n+1}|v_{i,n}^{-r,r}, x_{i,n}^{(k)})$$

Collective information transfer

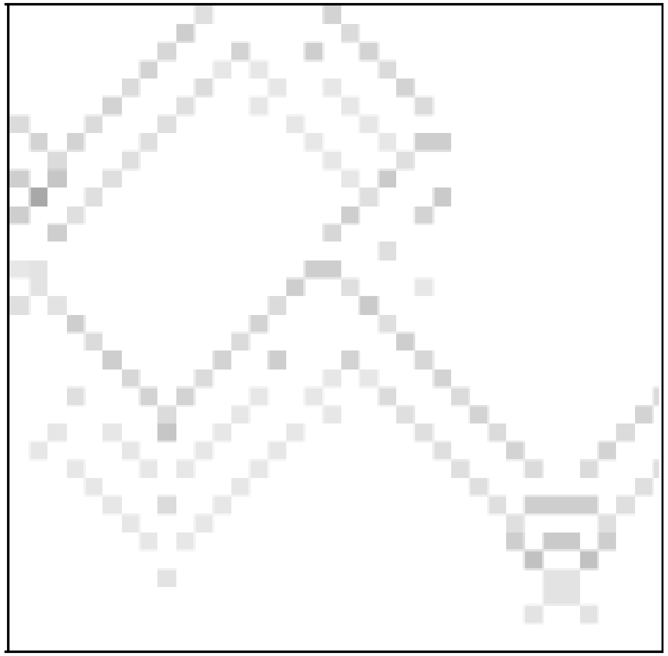
Intrinsic uncertainty

- Entropy rate useful for showing collective information transfer in a deterministic system.
- Total information is not the measure for collisions.

Results: Collective information transfer



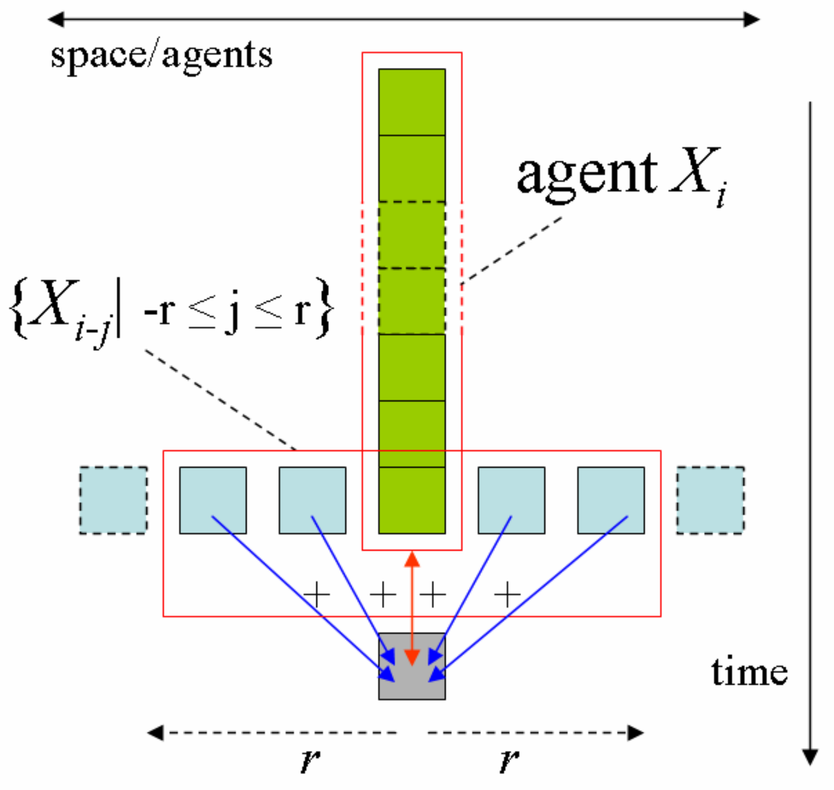
ECA rule 54



t(i,n,k=16) > 0

- Collective information transfer (entropy rate for a deterministic system) highlights gliders travelling in all directions.

Information modification



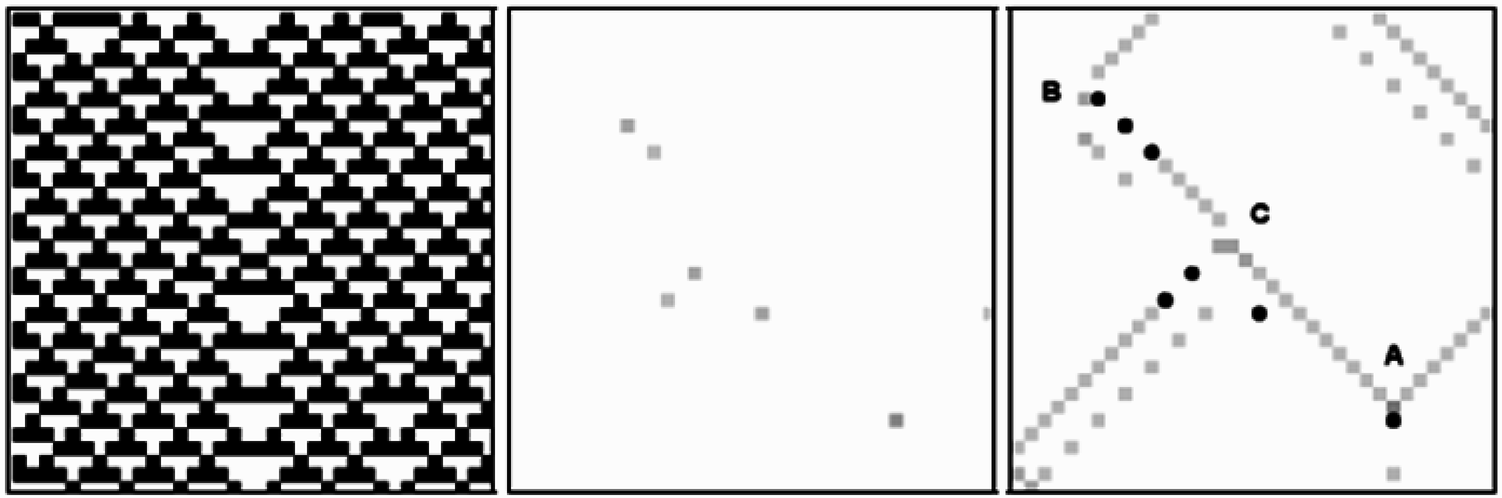
- Local separable information = information gained about next state from observing each causal source independently (in context of past) *at a given space-time point*:

$$s(i,n) = a(i,n) + \sum_{j=-r, j \neq 0}^{+r} t(i,j,n)$$

  - s(i,n) > 0: independent observations are informative overall; highly separable sources.

- s(i,n) < 0: independent observations are *misinformative* overall; sources are interacting and so are not separable.
- s(i,n) < 0 suggested to detect collisions, where "*the whole is greater than the sum of the parts*": non-trivial information modification.

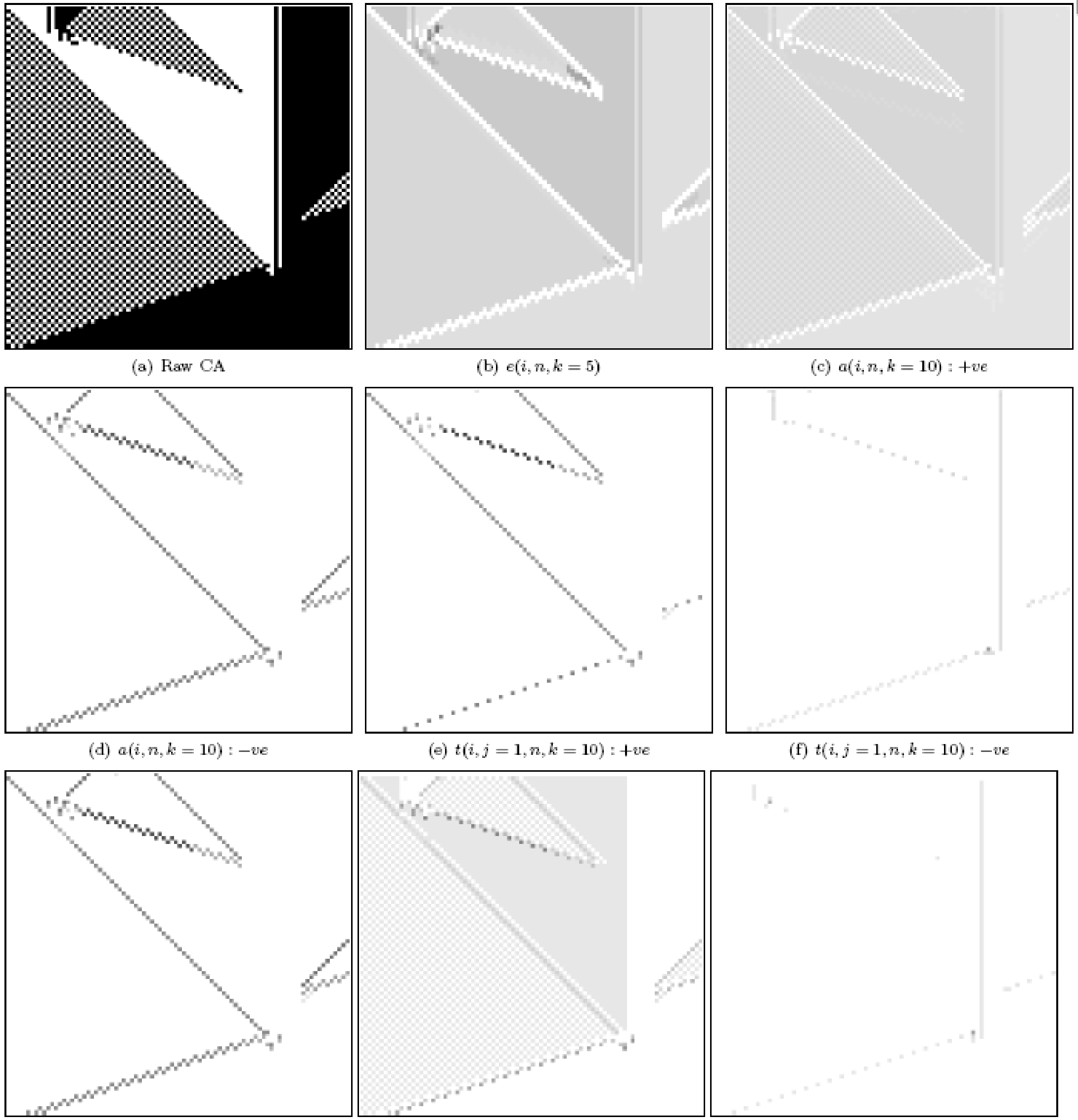
Results: Information modification



ECA rule 54 | s(i,n) < 0 | Locations of s(i,n) < 0

- Glider collisions are dominant information modification events (non-trivial information processing).
- Collision points are not where one would trivially identify them.

Local information dynamics – density class'n

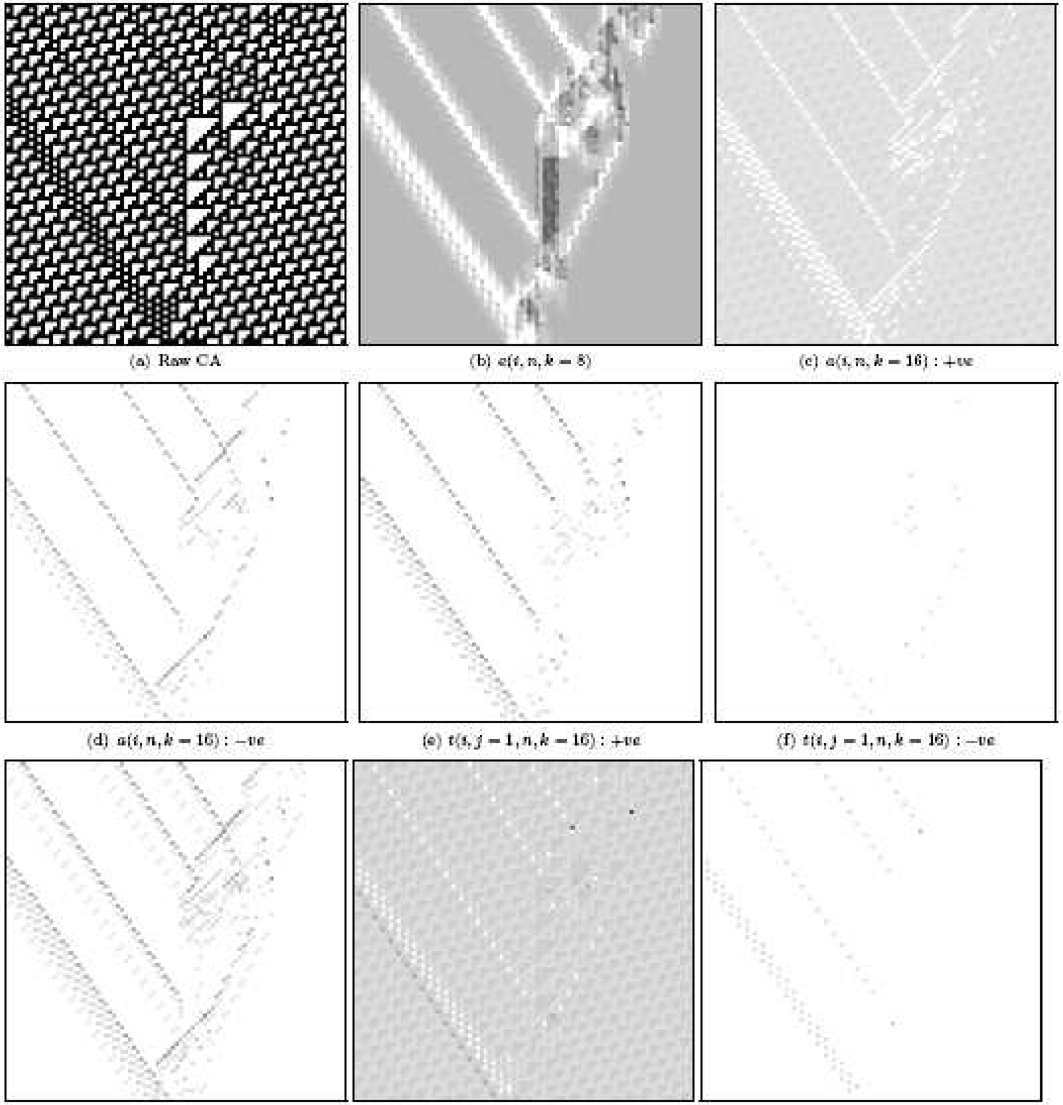


(a) Raw CA (b) $e(i, n, k = 5)$ (c) $a(i, n, k = 10) : +ve$

(d) $a(i, n, k = 10) : -ve$ (e) $t(i, j = 1, n, k = 10) : +ve$ (f) $t(i, j = 1, n, k = 10) : -ve$

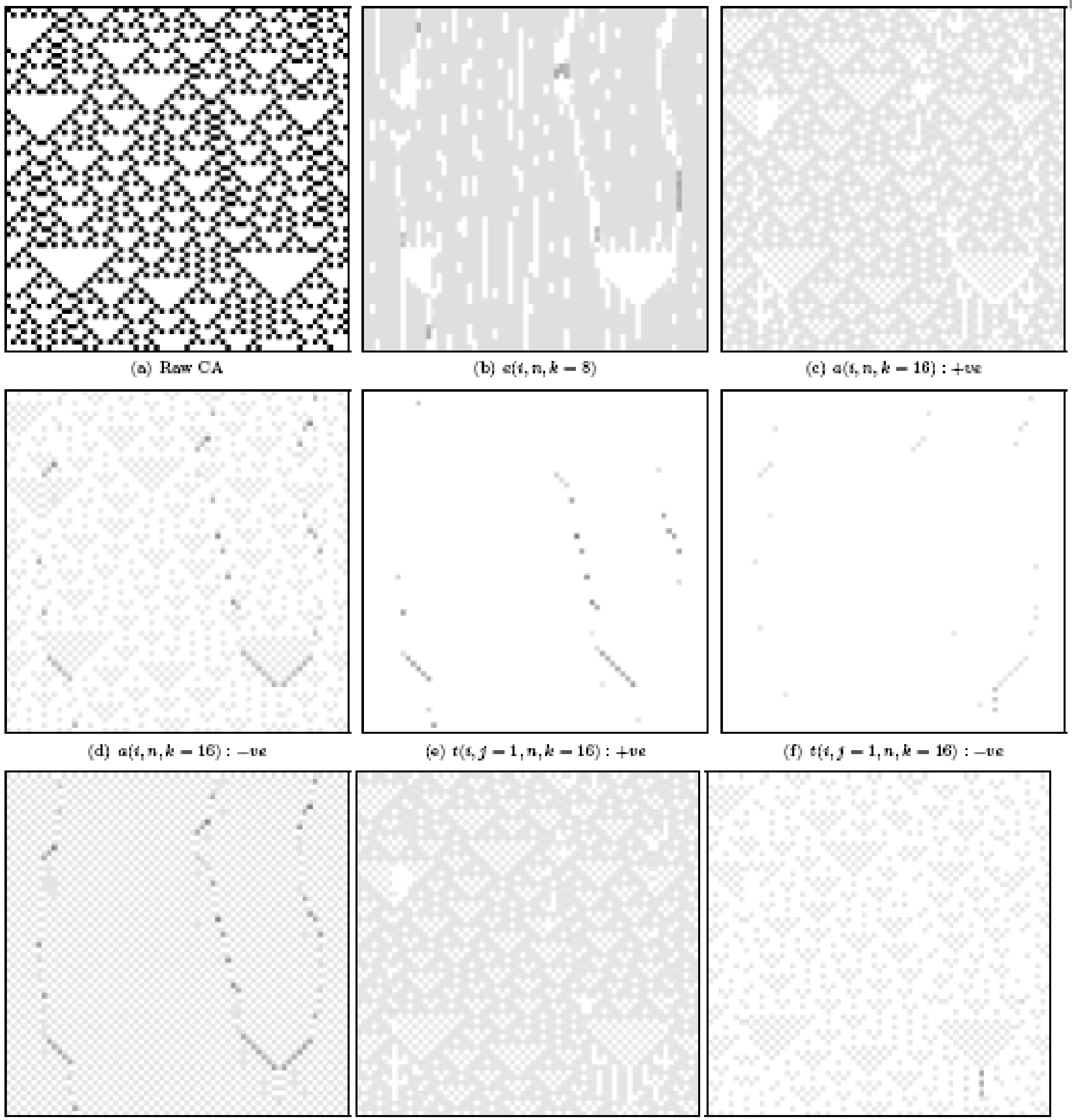
(g) $t(i, n, k = 10)$ (h) $s(i, n, k = 10) : +ve$ (i) $s(i, n, k = 10) : -ve$

Local information dynamics – rule 110



(a) Raw CA
(b) $a(i, n, k = 8)$
(c) $a(i, n, k = 16)$: $+ve$
(d) $a(i, n, k = 16)$: $-ve$
(e) $t(i, j = 1, n, k = 16)$: $+ve$
(f) $t(i, j = 1, n, k = 16)$: $-ve$
(g) $t(i, n, k = 16)$
(h) $s(i, n, k = 16)$: $+ve$
(i) $s(i, n, k = 16)$: $-ve$

Local information dynamics – rule 18



(a) Raw CA

(b) $a(i, n, k = 8)$

(c) $a(i, n, k = 16) : +ve$

(d) $a(i, n, k = 16) : -ve$

(e) $t(i, j = 1, n, k = 16) : +ve$

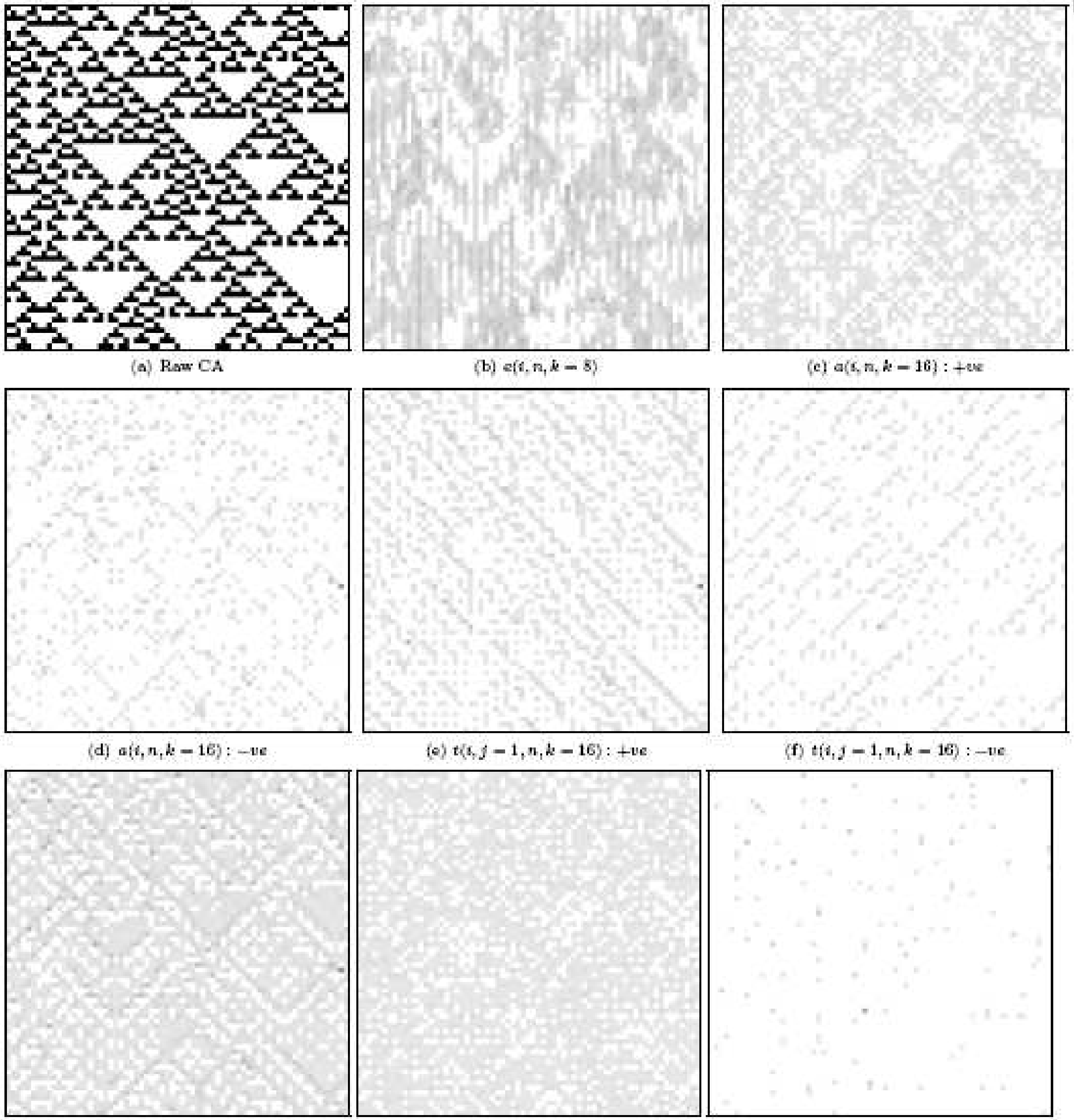
(f) $t(i, j = 1, n, k = 16) : -ve$

(g) $t(i, n, k = 16)$

(h) $s(i, n, k = 16) : +ve$

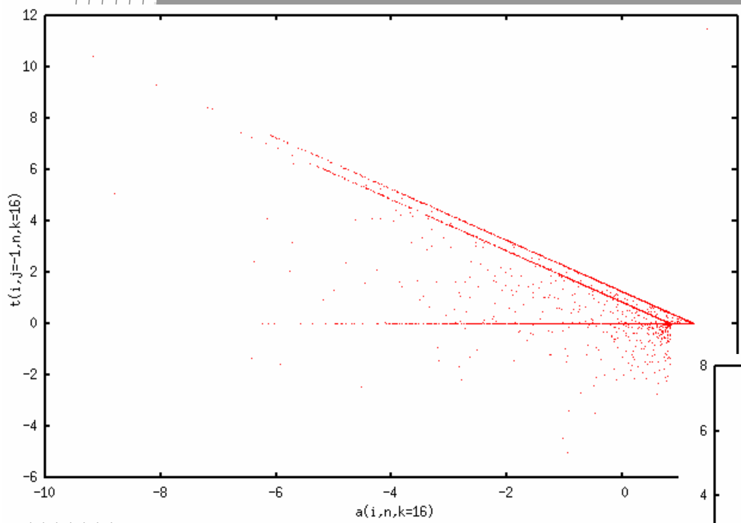
(i) $s(i, n, k = 16) : -ve$

Local information dynamics – rule 22

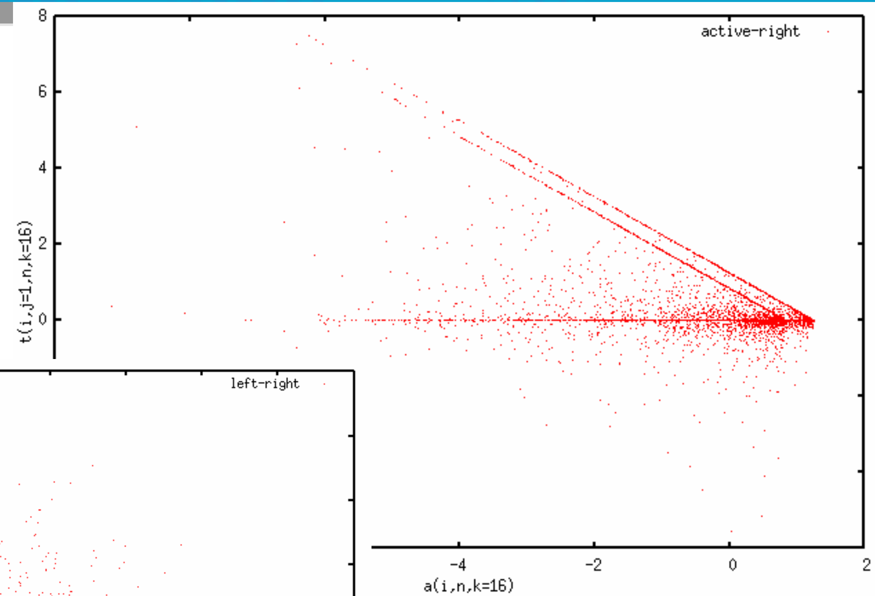


(a) Raw CA
(b) $e(i, n, k = 8)$
(c) $a(i, n, k = 16)$: +ve
(d) $a(i, n, k = 16)$: −ve
(e) $t(i, j = 1, n, k = 16)$: +ve
(f) $t(i, j = 1, n, k = 16)$: −ve
(g) $t(i, n, k = 16)$
(h) $s(i, n, k = 16)$: +ve
(i) $s(i, n, k = 16)$: −ve

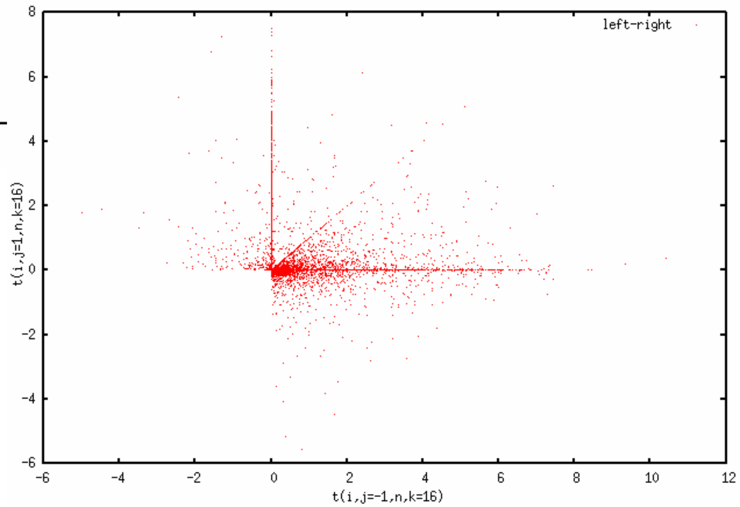
Coherence 1 – State space of local measures



t(i,j=-1,n) vs a(i,n)



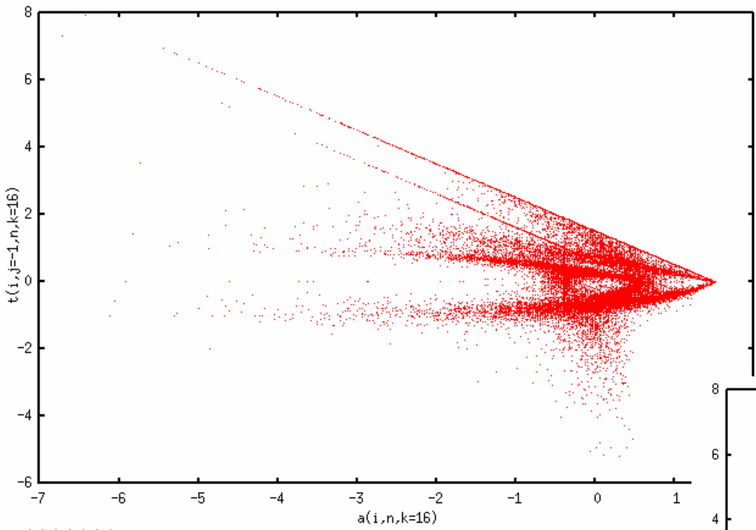
t(i,j=+1,n) vs a(i,n)



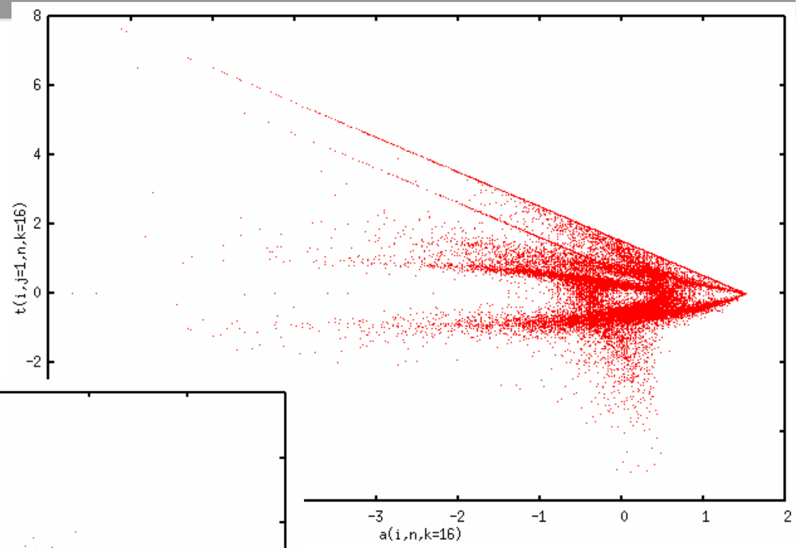
t(i,j=+1,n) vs t(i,j=-1,n)

- Rule 110 – clear structure seen in local information dynamics.
  - Structure appears to imply coherence of computation.
- What do we expect for rule 22?

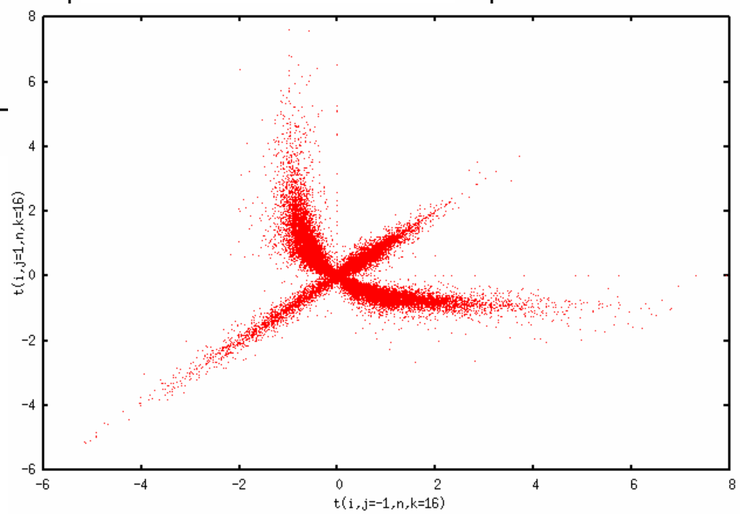
Coherence 1 – State space of local measures



t(i,j=-1,n) vs a(i,n)



t(i,j=+1,n) vs a(i,n)



t(i,j=+1,n) vs t(i,j=-1,n)

- Rule 22 – also has structure in state space of local dynamics.
- Suggests state space is not an appropriate indicator of coherent computation, though there is structure to rule 22 …

Coherence 2 - Average information dynamics

Rule	H	A	T(j=1)	T_c(j=1)	S	S>0	S<0	Count of S<0
110	0.985	0.806	0.065	0.071	0.979	0.981	-0.002	16724
54	0.998	0.725	0.080	0.193	0.885	0.895	-0.010	147141
22	0.934	0.187	0.187	0.559	0.562	0.615	-0.052	547969
18	0.818	0.286	0.014	0.517	0.315	0.458	-0.143	1449480
30	1	0.008	0.733	0.984	0.749	0.819	-0.069	487146

- In complex rules:
  - Apparent TE is a high proportion of complete TE for more than one channel. Suggests propagation of coherent effects from distinct sources.
  - Very low proportion of negative separable information. Suggests few collisions, allowing coherent computation, though with high impact each.

Local information dynamics: Conclusions

- New set of analytic tools and filters for spatiotemporal structure and local information dynamics.
- Long-held conjectures that particles are information transfer agents, and collisions are information modification proven quantitatively.
  - Evidence that these metrics are appropriate measures of relevant elements of computation.
- Future work:
  - Apply to other complex systems and conjectures about computation therein, e.g. microtubules, swarm behaviour, network dynamics.
  - Relate to collective computation and more traditional measures of computation
  - Relativistic information dynamics.

CSIRO

ICT Centre
Joseph Lizier
PhD Student

Phone: +61 2 9325 3167
Email: joseph.lizier at csiro.au
Web: www.ict.csiro.au

www.csiro.au

Thank you