Words generated by cellular automata

Eric Rowland

University of Waterloo

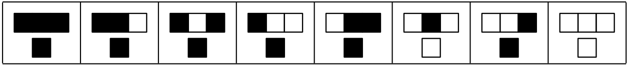
(soon to be LaCIM)

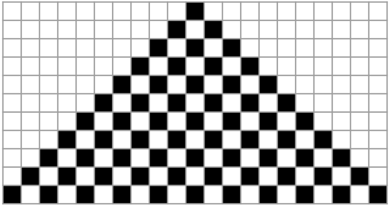
November 25, 2011

Outline

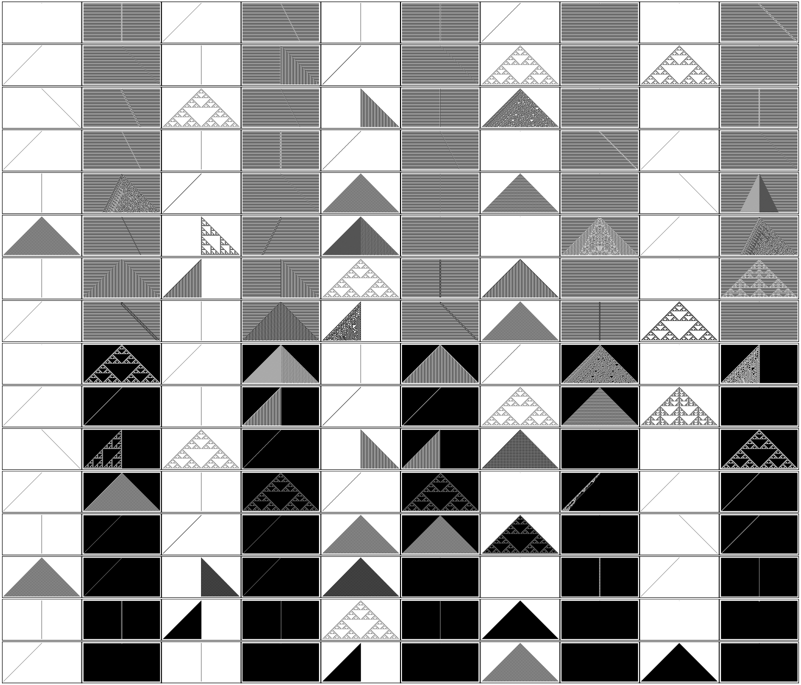
One-dimensional cellular automata

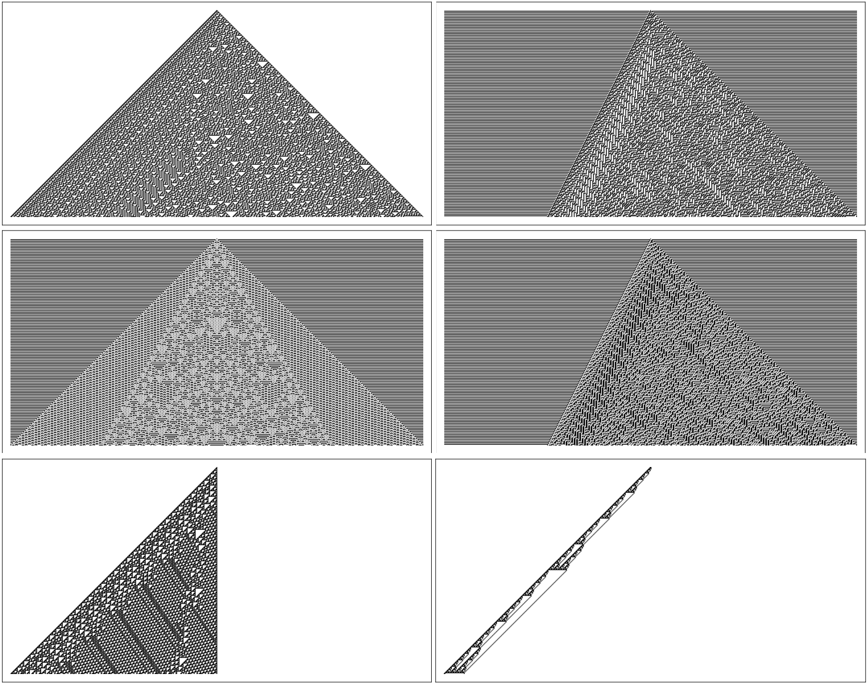
- alphabet Σ of size k (for example $\{0, 1, \ldots, k-1\}$)
- function $i : \mathbb{Z} \to \Sigma$ (the initial condition)
- function $f : \Sigma^d \to \Sigma$ (the update rule)





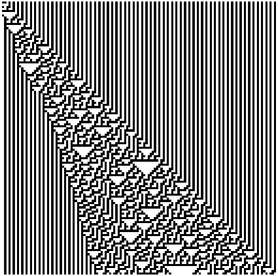
Naming scheme: $11111010_2 = 250$.
Wolfram: Look at all k^{k^d} k-color rules depending on d cells.

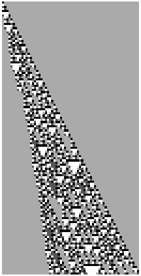




All but finitely many cells in the initial condition have the same color.

We could also allow periodic backgrounds,
but coarse-graining reduces to constant background.





Most of our examples will use $k = 2$ colors and the initial condition

$\cdots$ □□□■□□□ $\cdots$.

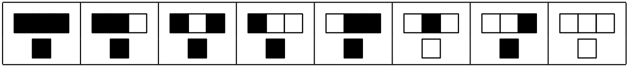
Outline

Row words

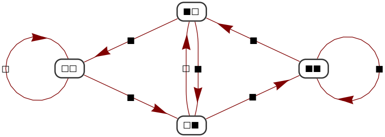
Wolfram 1984:
Given a cellular automaton rule, the set of finite words obtainable at step n is a regular language.

Step 0: All words are obtainable.

Step 1: To obtain $w_1 w_2 \cdots w_\ell$ we must overlap d-tuples appropriately.

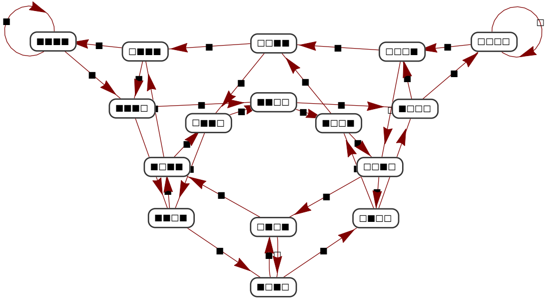


Construct the rule's "overlap graph" for words of length $d - 1$:



Row words (continued)

Step 2:
A subword of the initial condition of length $2(d-1)+1$ determines one letter on step 2. Construct overlap graph for words of length $2(d-1)$:



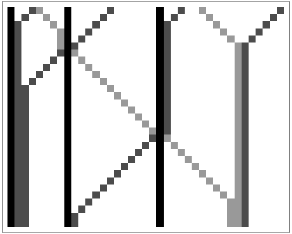
Step n:
A subword of length $n \cdot (d-1)+1$ determines one letter on step n. Construct the overlap graph for words of length $n \cdot (d-1)$.

Limit language

The set of words obtainable at every step is the limit language.

Hurd 1987:
There exist rules for which the limit language is . . .

- not regular.
- not context-free.
- not recursive.



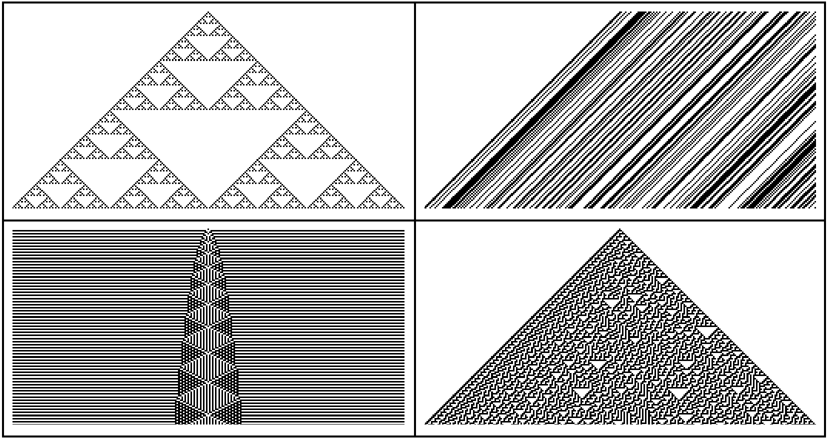
$$\text{limit language} \cap \square\square\blacksquare\square^{*}\blacksquare\square^{*}\blacksquare\square\square = \{\square\square\blacksquare\square^{n}\blacksquare\square^{n}\blacksquare\square\square : n \geq 0\}.$$

Open question: Which languages occur as limit languages?

Outline

Infinite column words

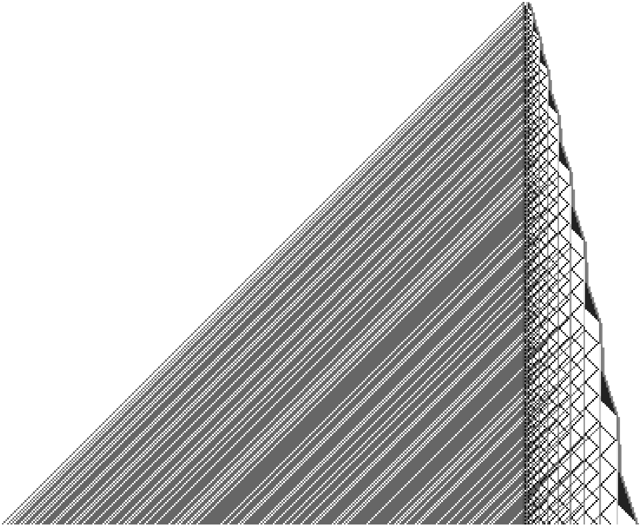
characteristic sequence of 2^n | bits of π



characteristic sequence of n^2 | statistically random sequences

Characteristic sequence of primes

A 16-color rule depending on 3 cells that computes the primes:



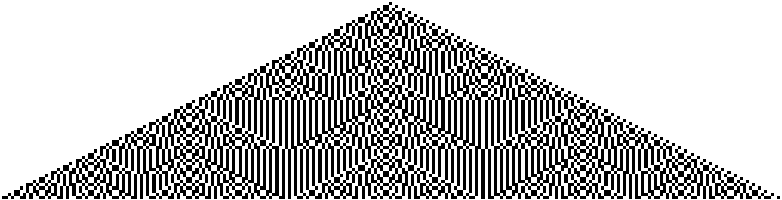
The Thue–Morse sequence

$$a(n) = \begin{cases} 0 & \text{if the binary representation of } n \text{ has an even number of 1s} \\ 1 & \text{if the binary representation of } n \text{ has an odd number of 1s.} \end{cases}$$

For $n \geq 0$, the Thue–Morse sequence is

$$0\,1\,1\,0\,1\,0\,0\,1\,1\,0\,0\,1\,0\,1\,1\,0\,1\,0\,0\,1\,0\,1\,1\,0\,0\,1\,1\,0\,1\,0\,0\,1 \cdots .$$

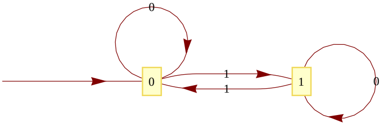
$a(n)$ occurs as a column of this $d = 5$ automaton:



The Thue–Morse sequence

We can construct a different automaton containing $a(n)$…

The Thue–Morse sequence is 2-automatic:



The generating function $f(x) = \sum_{n \geq 0} a(n) x^n$ is algebraic over $\mathbb{F}_2(x)$:

$$(x+1)^3 f(x)^2 + (x^2+1) f(x) + x = 0.$$

Automatic sequences

Furstenberg 1967:
A power series $f(x)$ over $\mathbb{F}_{p^\alpha}$ is algebraic if and only if it is the diagonal of a rational series $g(x, y)$ over $\mathbb{F}_{p^\alpha}$.

Litow–Dumas 1993:
Write $g(x/y, y) = P(x, y)/Q(x, y) = \sum_{n \geq 0} r_n(y) x^n$.
Then $Q(x, y)$ encodes a linear recurrence satisfied by $r_n(y)$.
This gives a cellular automaton with memory.

If $a(n)$ is p-automatic, then there exists a cellular automaton with column $a(n)$.

Corollary:
Every periodic sequence occurs.

Open questions

- Does every periodic sequence on an alphabet of size k occur in a k-color cellular automaton?
- Does every k-automatic sequence occur in a cellular automaton (if k is not prime)?
- Does the Fibonacci word

  $$abaababaabaababaababaabaababaabaab\cdots$$

  (the fixed point of $\varphi(a) = ab, \varphi(b) = a$) occur in a cellular automaton?
- Exhibit some sequence that does not occur as the column of a cellular automaton.

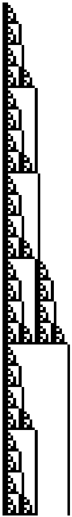
Outline

Rule 106

Rule 106 grows like $\sqrt{n}$.

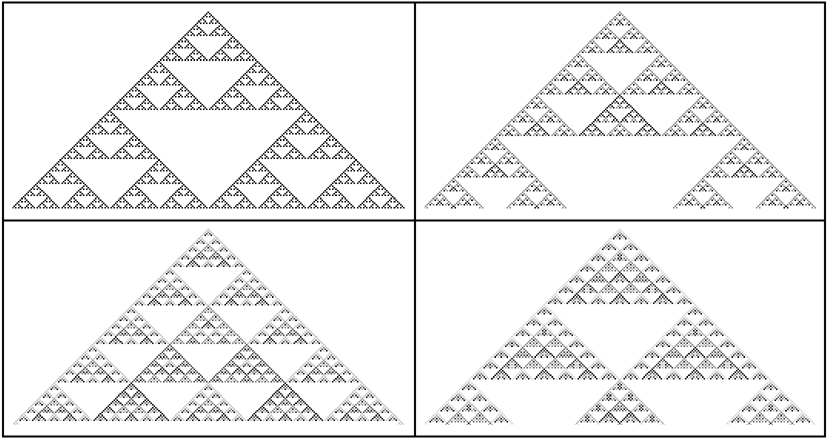
The number $a(n)$ of black cells on row n is 2-regular:

$$\begin{aligned}
a(4n+0) &= a(n) \\
a(4n+1) &= a(4n+2) \\
a(8n+2) &= a(2n+1) \\
a(8n+3) &= 2a(2n+1) - a(2n) \\
a(8n+6) &= 2a(2n+1) - a(2n) \\
a(8n+7) &= 4a(2n+1) - 3a(2n)
\end{aligned}$$



Binomial coefficients

Binomial coefficients modulo k are produced by cellular automata.



Nonzero binomial coefficients

Let $a_{p^\alpha}(n) = |\{0 \leq m \leq n : \binom{n}{m} \not\equiv 0 \mod p^\alpha\}|$.

Write $n = n_\ell \cdots n_1 n_0$ in base p.

Let $|n|_w$ be the number of occurrences of w in $n_\ell \cdots n_1 n_0$.

- Glaisher 1899:

$$a_2(n) = 2^{|n|_1}.$$

- Fine 1947:

$$a_p(n) = \prod_{i=0}^{\ell} (n_i + 1).$$

For example, $a_5(n) = 2^{|n|_1} 3^{|n|_2} 4^{|n|_3} 5^{|n|_4}$.

It follows that $a_p(n)$ is p-regular.

Nonzero binomial coefficients

Rowland 2011:
Algorithm for obtaining a symbolic expression in $|n|_w$ for $a_{p^\alpha}(n)$.
It follows that $a_{p^\alpha}(n)$ is p-regular for each $\alpha \geq 0$.

For example:

$$a_{p^2}(n) = \left(\prod_{i=0}^{\ell} (n_i + 1)\right) \cdot \left(1 + \sum_{i=0}^{\ell-1} \frac{p - (n_i + 1)}{n_i + 1} \cdot \frac{n_{i+1}}{n_{i+1} + 1}\right).$$

Expressions for $p = 2$ and $p = 3$:

$$a_4(n) = 2^{|n|_1} \left(1 + \frac{1}{2}|n|_{10}\right)$$

$$a_9(n) = 2^{|n|_1} 3^{|n|_2} \left(1 + |n|_{10} + \frac{1}{4}|n|_{11} + \frac{4}{3}|n|_{20} + \frac{1}{3}|n|_{21}\right)$$

Nonzero binomial coefficients

Higher powers of 2:

$$a_8(n) = 2^{|n|_1} \left(1 + \frac{1}{8}|n|_{10}^2 + \frac{3}{8}|n|_{10} + |n|_{100} + \frac{1}{4}|n|_{110}\right)$$

$$\begin{aligned}\frac{a_{16}(n)}{2^{|n|_1}} = 1 &+ \frac{5}{12}|n|_{10} + \frac{1}{2}|n|_{100} + \frac{1}{8}|n|_{110} \\ &+ 2|n|_{1000} + \frac{1}{2}|n|_{1010} + \frac{1}{2}|n|_{1100} + \frac{1}{8}|n|_{1110} + \frac{1}{16}|n|_{10}^2 \\ &+ \frac{1}{2}|n|_{10}|n|_{100} + \frac{1}{8}|n|_{10}|n|_{110} + \frac{1}{48}|n|_{10}^3\end{aligned}$$

Additive automata

Binomial coefficients are the coefficients of $(1 + y)^n$:

$$(1 + y)(1 + 3y + 3y^2 + y^3)$$
$$= 1 + (3 + 1)y + (3 + 3)y^2 + (1 + 3)y^3 + y^4$$

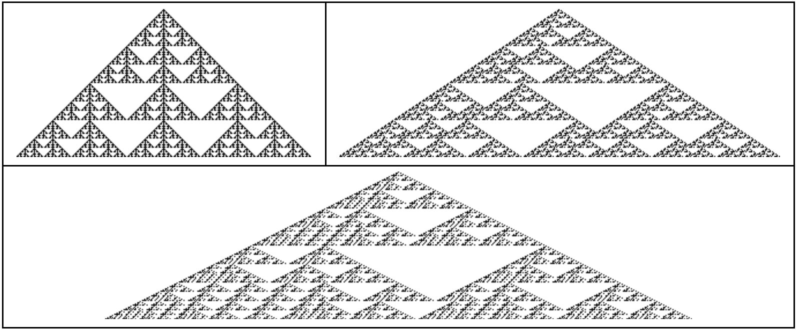
Martin–Odlyzko–Wolfram 1984:
Let $q(y), r_0(y)$ be polynomials with coefficients in some finite ring.
There is a cellular automaton whose nth row consists of the coefficients of $q(y)^n r_0(y)$.

The entire evolution of the automaton is encoded in

$$\sum_{n \geq 0} r_n(y) x^n = \sum_{n \geq 0} q(y)^n r_0(y) x^n = \frac{r_0(y)}{1 - xq(y)}.$$

Additive automata

Here is $(1 + y + y^{d-1})^n$ over $\mathbb{F}_2$ for $d = 3, 4, 5$:



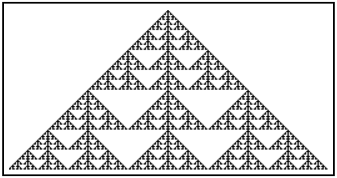
Amdeberhan–Stanley $\sim$2008:
Let $f(x_1, \ldots, x_m) \in \mathbb{F}_{p^\alpha}[x_1, \ldots, x_m]$. The number $a(n)$ of nonzero terms in the expanded form of $f(x_1, \ldots, x_m)^n$ is p-regular.

Outline

Row lengths

Joint work with Charles Brummitt (UC Davis) . . .

$\ell(n)$ = width of region on row n that differs from the background

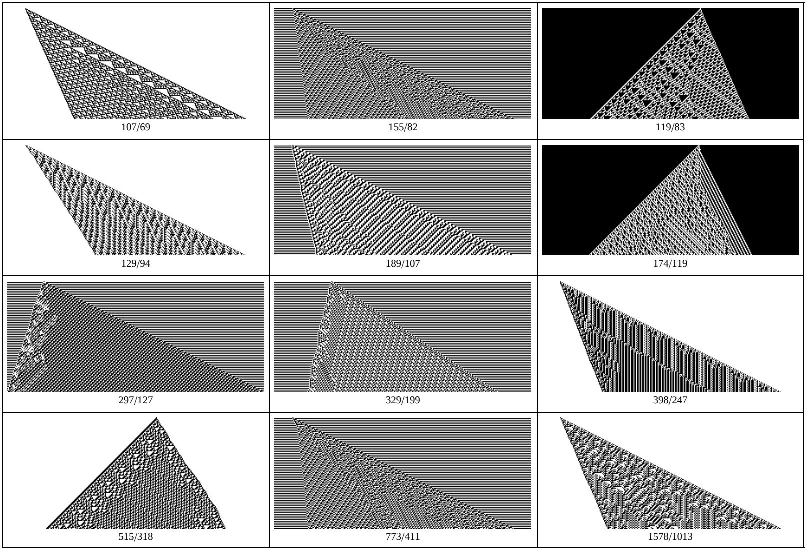


For example, $\ell(n) = 2n + 1$.

Upper bound: $\ell(n) \leq (d - 1)n + c$.

For many automata, $\ell(n)$ is linear.
For $k = 2$ and $d \leq 3$, the only slopes that occur are $0, 1, 3/2, 2$.

Interesting slopes for $d = 4$



Boundary word

For these automata, $\ell(n+1) - \ell(n)$ is eventually periodic.

Definition

The boundary word of an automaton is the infinite word $(\ell(n+1) - \ell(n))_{n \geq 0}$.

The boundary word is not necessarily a word on a finite subset of $\mathbb{Z}$. But often it is.

Properties of the automaton are reflected in the boundary word.

Rule 106 again

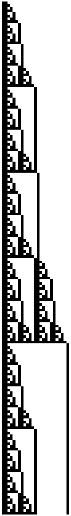
Boundary word:

$$\mathbf{w}_{106} = 11010011000000010000000011010011\cdots$$
$$= 1^2 0^1 1^1 0^2 1^2 0^7 1^1 0^8 1^2 0^1 1^1 0^2 1^2 0^{31} 1^1 0^{32} \cdots .$$

Let

$$\varphi : A \to ABCD,\ B \to CCAB,\ C \to CCCC,\ D \to CCCD$$
$$\psi : A \to 1,\ B \to 1,\ C \to 0,\ D \to 1$$

Then $\mathbf{w}_{106} = \psi(\varphi^\omega(A))$.
In particular, $\mathbf{w}_{106}$ is morphic.



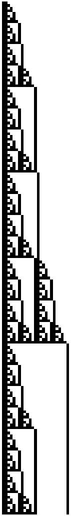
Square-root growth rate can be derived from φ.

The morphism φ is 4-uniform.

Row lengths in rule 106

The length $\ell(n)$ is 2-regular:

$$\ell(4n+1) = 1/2\ell(4n) + 1/2\ell(4n+2)$$
$$\ell(8n+2) = -2\ell(2n) + \ell(8n) + 2\ell(2n+1)$$
$$\ell(8n+3) = -2\ell(2n) + \ell(8n) + 2\ell(2n+1)$$
$$\ell(8n+4) = -3\ell(2n) + \ell(8n) + 3\ell(2n+1)$$
$$\ell(8n+6) = -3\ell(2n) + \ell(8n) + 3\ell(2n+1)$$
$$\ell(8n+7) = -4\ell(2n) + \ell(8n) + 4\ell(2n+1)$$
$$\ell(16n+0) = -2\ell(n) + 3\ell(4n) + \ell(4n+2) - \ell(4n+3)$$
$$\ell(16n+8) = -2\ell(n) + 1/2\ell(4n) + 7/2\ell(4n+2) - \ell(4n+3)$$

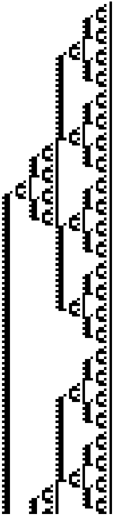


$d = 4$ rule 39780

Rule 39780 also grows like $\sqrt{n}$.

Its boundary word is $\psi(\varphi^\omega(A))$, where

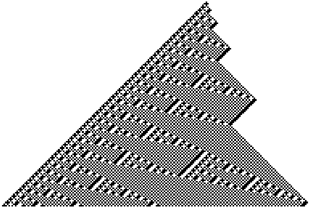
$$\varphi : A \to ABC,\ B \to DAB,$$
$$C \to CECE,\ D \to CECD,\ E \to CECE$$
$$\psi : A \to 2,\ B \to 2,\ C \to 1,\ D \to 0,\ E \to -1$$



φ is not uniform.

$\ell(n)$ is evidently not 2-regular.

$d = 4$ rule 2230



Growth is linear: $\ell(n) \in \Theta(n)$.

But $\lim_{n\to\infty} \frac{\ell(n)}{n}$ does not exist.

- $\liminf \ell(n)/n = 6/5$
- $\limsup \ell(n)/n = 3/2$

Boundary word is $\psi(\varphi^\omega(A))$, where

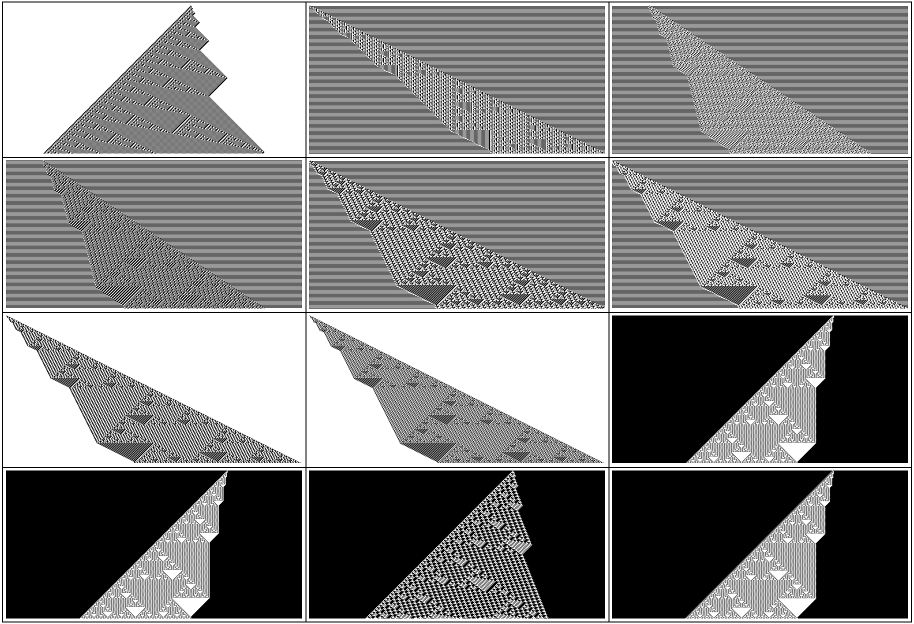
$$\varphi : A \to ABCB,\ B \to BB,\ C \to CC$$
$$\psi : A \to \epsilon,\ B \to 2,\ C \to 0.$$

The fixed point of φ is

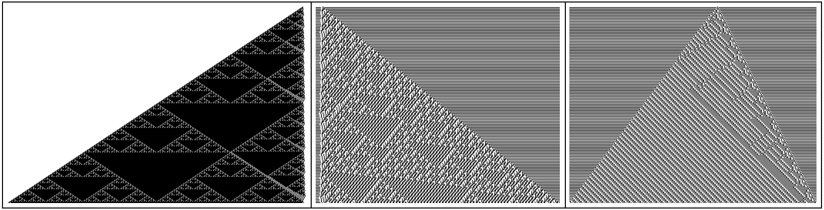
$$\varphi^\omega(A) = ABCBBBCCBBBBBBCCCCBBBBBBBBBBBB\cdots.$$

The frequencies of B and C don't exist!

Automata with the same morphism φ



Morphic boundaries where the slope exists



For the first, the boundary word is (basically) the fixed point

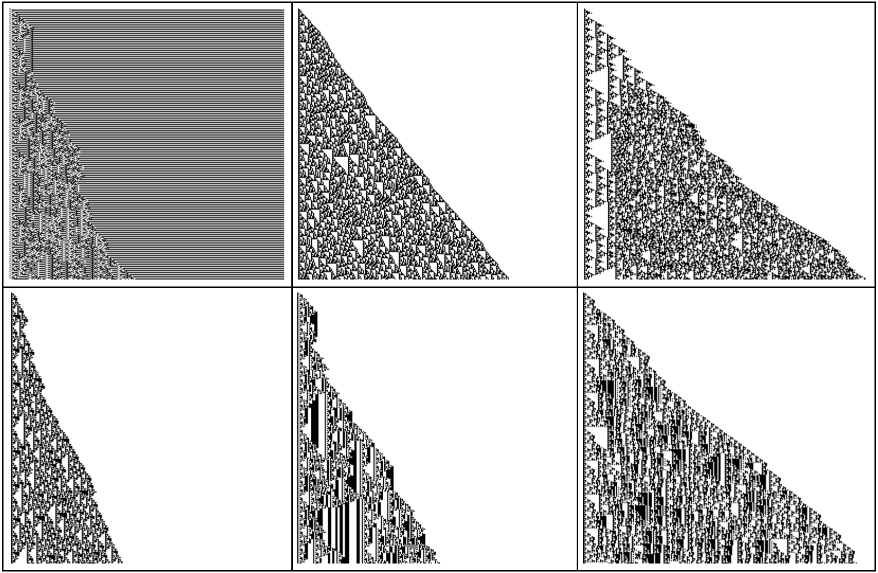
$$\varphi^{\omega}(2) = 221221122122111221221122122111 \cdots$$

of the morphism $\varphi(1) = 1, \varphi(2) = 221$.

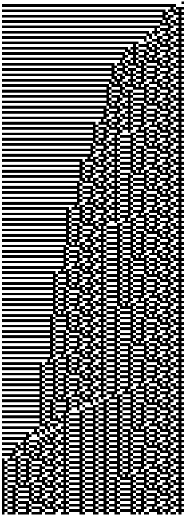
Questions

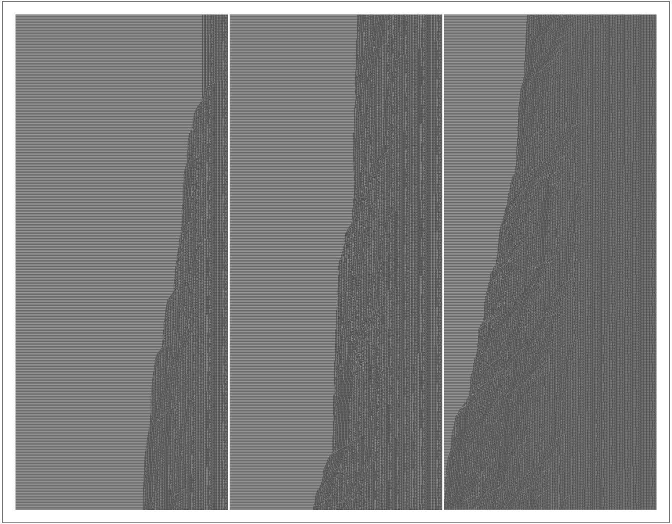
- Does every morphic word occur as the boundary word of a cellular automaton?
- Vague conjecture.
  If $\ell(n)$ is computable faster than the automaton computes it, then the boundary word is morphic.
- Vague open question.
  Make the vague conjecture precise!

Chaotic boundaries



A misleading example





Around step 524500, growth increases rapidly (10000 steps shown).