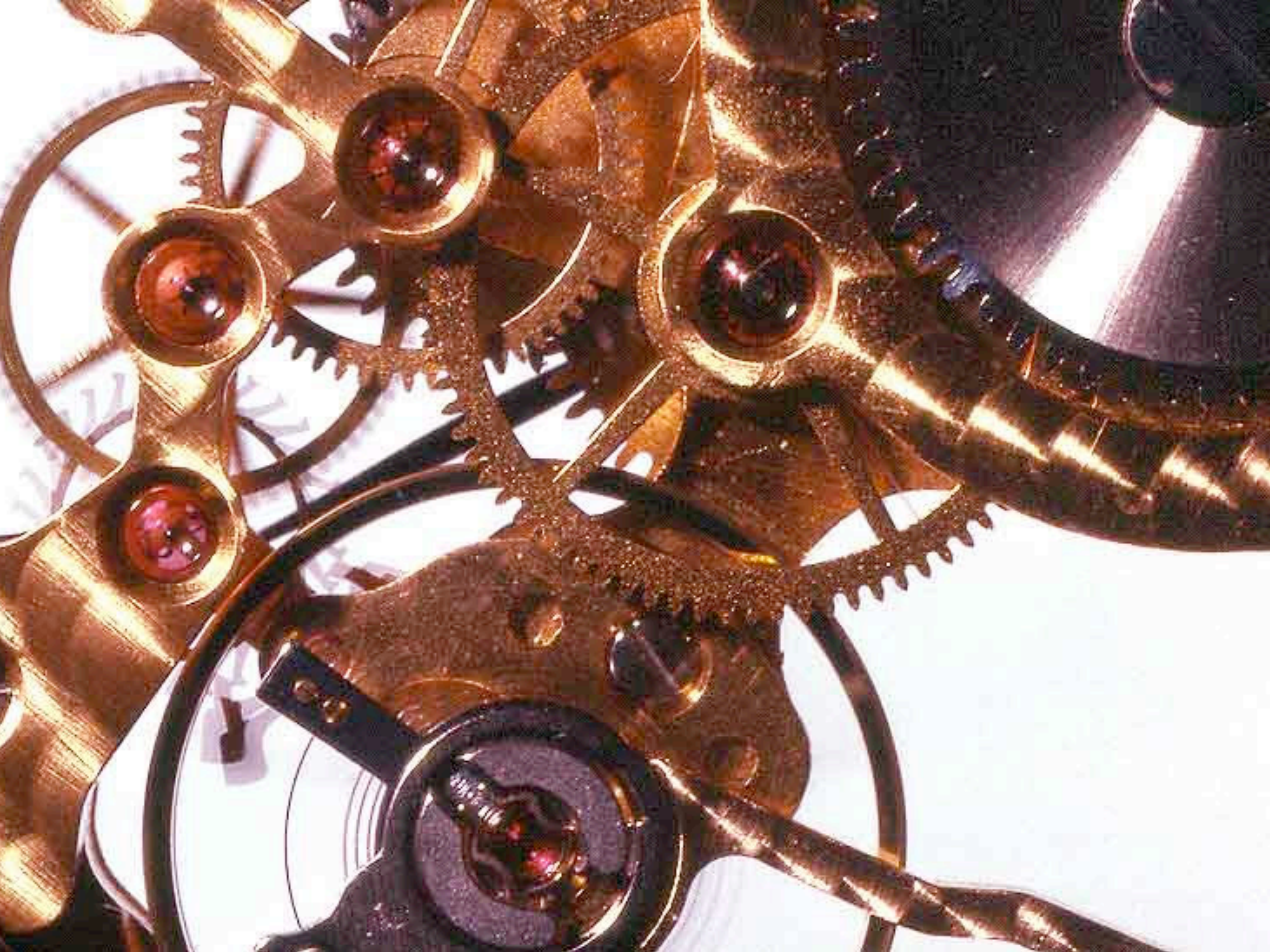
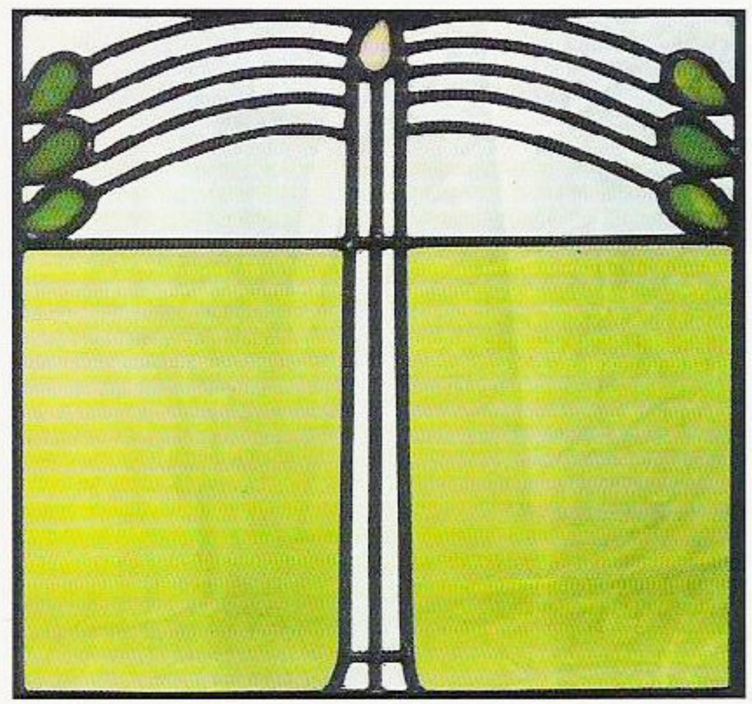


Classifying Streams

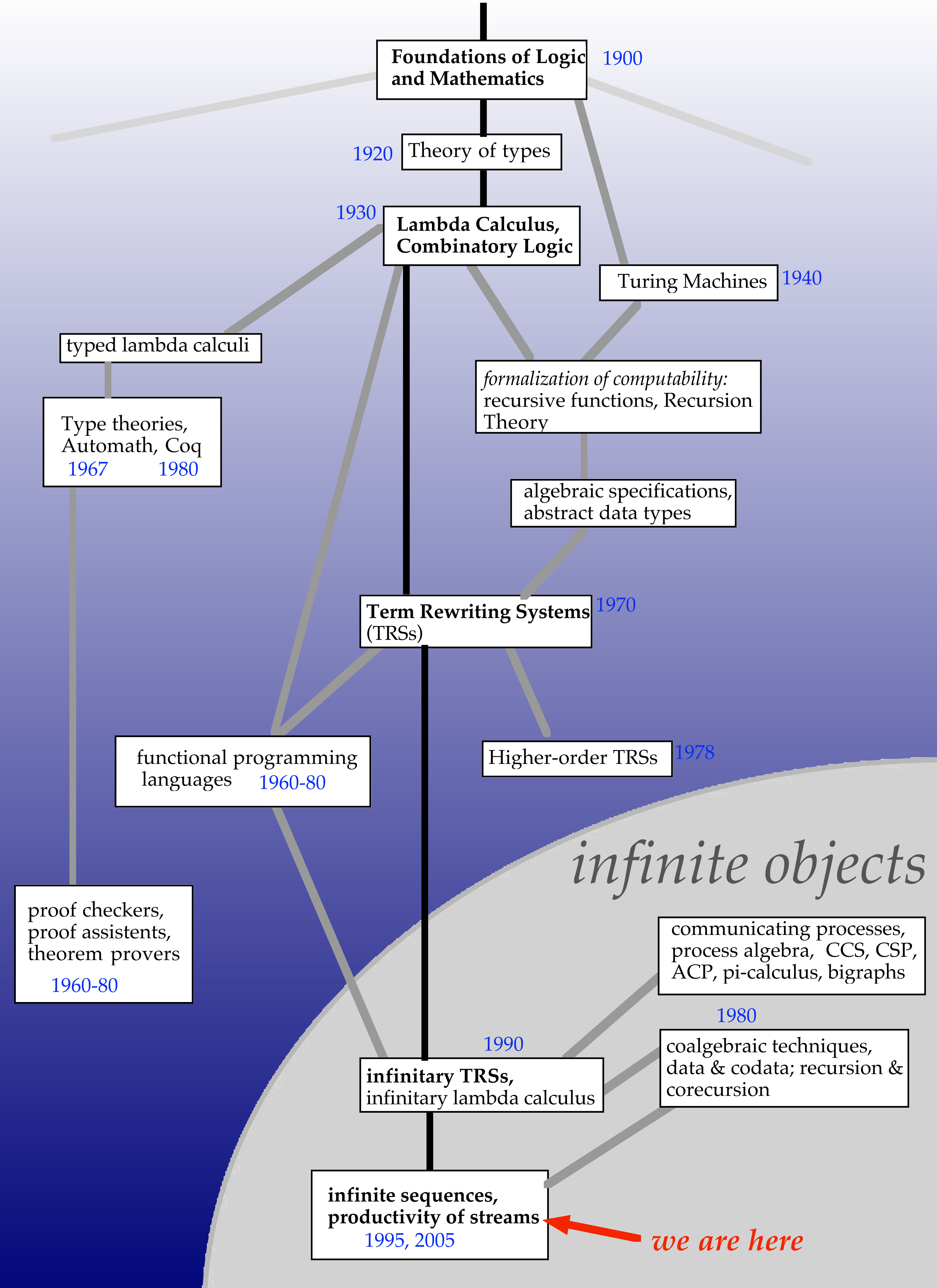
Stream Seminar



20 April 2010 Nijmegen

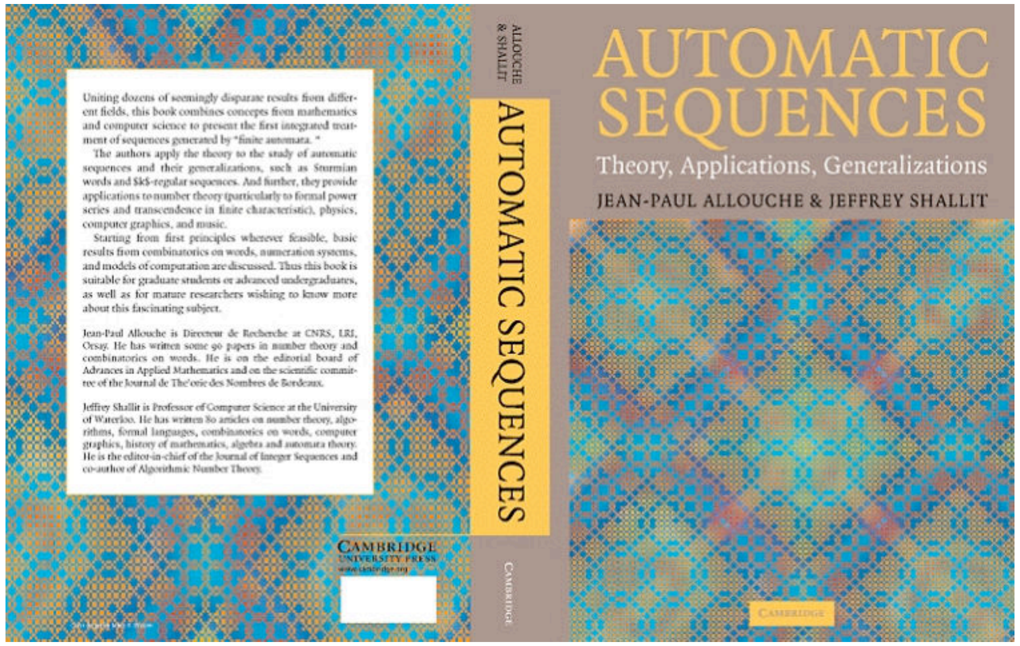
Jörg Endrullis
Dimitri Hendriks
Jan Willem Klop

history and context of our interest in streams



Classifying Streams

The Landscape of Streams	jan willem
Spot the difference	jan willem
Prime-generated streams	dimitri
Degrees of streams	joerg



(i) *Automatic sequences* [3]. A stream σ is called k-automatic if there exists a finite automaton computing the letter $\sigma(i)$ at index i when fed the digits of i in base k. E.g. the Thue–Morse sequence M:

$$\mathsf{M} = 0110100101101001\ldots$$

is 2-automatic, obtainable by the automaton in Figure 2, where $\mathsf{M}(i)$ is the number of the final state of the automaton.

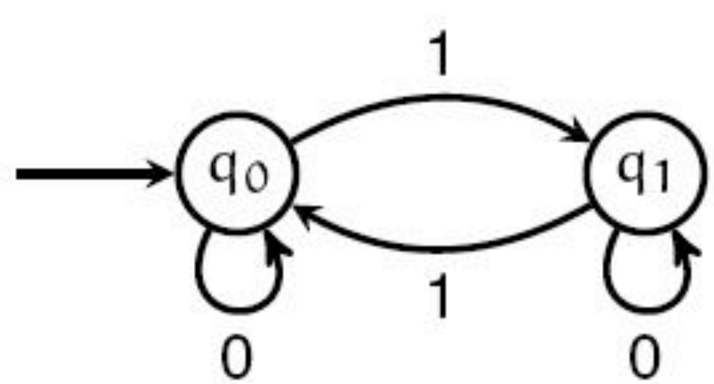
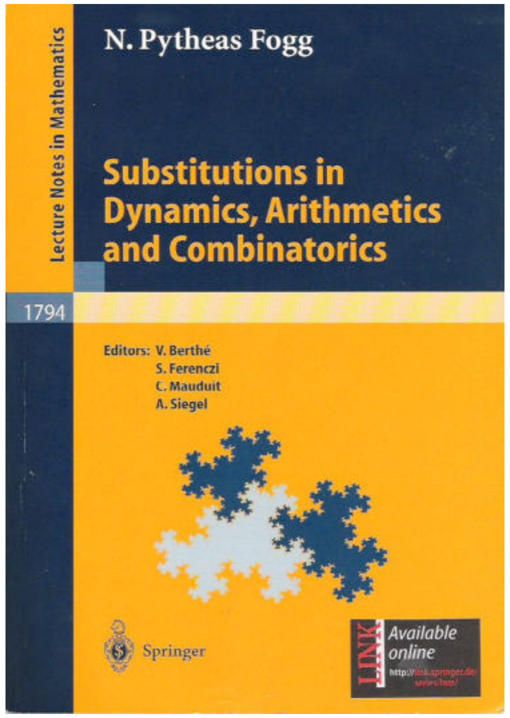


Figure 2: *Thue–Morse is 2-automatic.*

(ii) *Morphic sequences* [3, 53] subsume the automatic sequences. A morphic sequence is generated from a start word by a morphism (a letter to word substitution), followed by a letter to letter substitution applied to the limit word. E.g. M is also obtained by:

$$0 \to 01 \qquad\qquad 1 \to 10$$

from start word 0.



(iii) *Toeplitz words* [31, 8]. Let x be a finite word in $(\Sigma \cup \{?\})^*$. The Toeplitz word T_x is x^ω where the places of '?' are filled in with T_x itself. For example for $x = \mathsf{101}?$ we obtain

$$x^\omega = \mathsf{101}?\mathsf{101}?\mathsf{101}?\mathsf{101}?\mathsf{101}?\mathsf{101}?\mathsf{101}?\dots \quad \text{and}$$
$$\mathsf{T}_x = \mathsf{101}\underline{\mathsf{1}}\mathsf{101}\underline{\mathsf{0}}\mathsf{101}\underline{\mathsf{1}}\mathsf{101}\underline{\mathsf{1}}\mathsf{101}\underline{\mathsf{1}}\mathsf{101}\underline{\mathsf{0}}\mathsf{101}\underline{\mathsf{1}}\dots$$

This stream, henceforth denoted by T, is called the 'period doubling sequence'. Also T is morphic.

(iv) *Sturmian sequences* [3, 6, 51]. Sturmian sequences can be viewed as discretization of straight lines $\nearrow$ in the plane with a unit grid. Write a 0 whenever $\nearrow$ crosses a vertical line of the grid and 1 for every horizontal line.

Thus the Fibonacci sequence:

$$\mathsf{F} = \mathsf{0100101001001010010100100101001001010}\ldots$$

is obtained by the straight line from the origin with slope $\frac{1}{\phi}$ with the golden ratio $\phi \approx 1.618$.

F is also morphic, obtained by

$$\mathsf{0} \to \mathsf{01} \qquad\qquad \mathsf{1} \to \mathsf{0}$$

from start word $\mathsf{0}$.

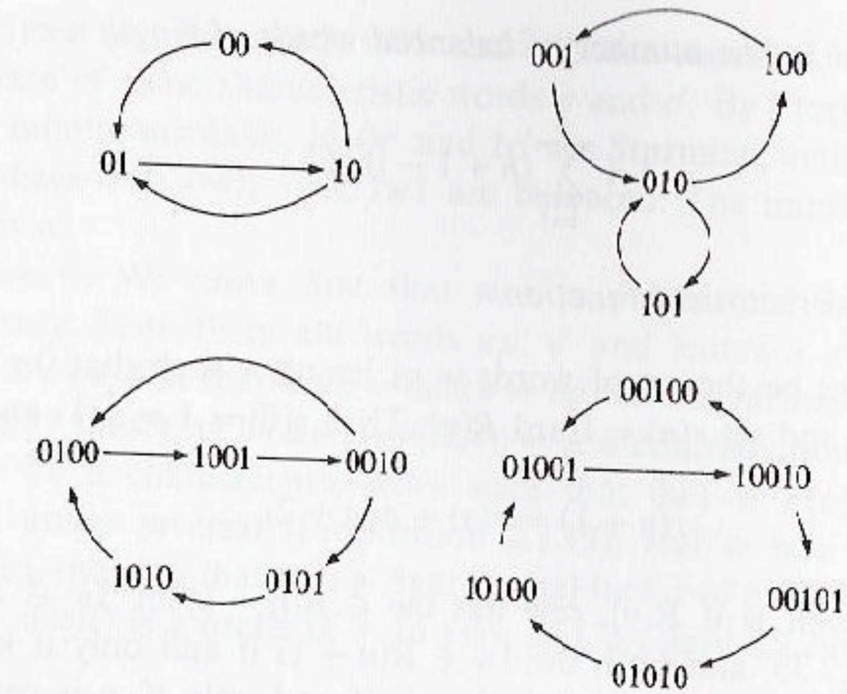


Figure 2.6. Factor graphs for the Fibonacci word.

EXAMPLE 1.3.11. Let t be the infinite Thue–Morse
1.2.9). It is easily checked that 000 and 111 are no
factor graph $G_3(t)$ is given in Figure 1.9.

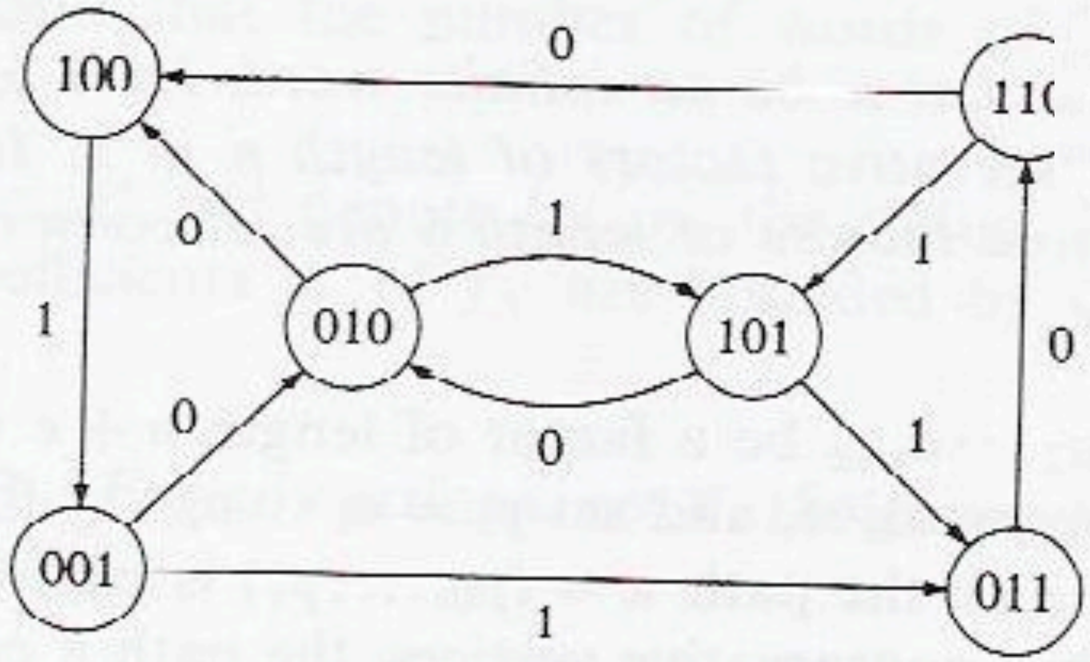


Figure 1.9. Factor graph of order 3 for the Thue–Morse word.

Encyclopedia of Mathematics and its Applications 90

ALGEBRAIC COMBINATORICS ON WORDS

M. Lothaire

CAMBRIDGE

CHAPTER 2

Sturmian Words

2.0. Introduction

Sturmian words are infinite words over a binary alphabet that have exactly $n+1$ factors of length n for each $n \geq 0$. It appears that these words admit several equivalent definitions, and can even be described explicitly in arithmetic form. This arithmetic description is a bridge between combinatorics and number theory. Moreover, the definition by factors makes Sturmian words define symbolic dynamical systems. The first detailed investigations of these words were done from this point of view. Their numerous properties and equivalent definitions, and also the fact that the Fibonacci word is Sturmian, have led to a great development, under various terminologies, of the research.

The aim of this chapter is to present basic properties of Sturmian words and of their transformation by morphisms. The style of exposition relies basically on combinatorial arguments.

The first section is devoted to the proof of the Morse–Hedlund theorem stating the equivalence of Sturmian words with the set of balanced aperiodic words and the set of mechanical words of irrational slope. We also mention several other formulations of mechanical words, such as rotations and cutting sequences. We next give properties of the set of factors of one Sturmian word, such as closure under reversal, the minimality of the associated dynamical system, the fact that the set depends only on the slope, and we give the description of special words.

In the second section, we give a systematic exposition of standard pairs and standard words. We prove the characterization by the double palindrome property, describe the connection with Fine and Wilf's theorem. Then, standard sequences are introduced to connect standard words to characteristic Sturmian words. The relation to Beatty sequences is in the Problems. This section also contains the enumeration formula for finite Sturmian words. It ends with a short description of frequencies.

sturmian streams
two-distance sequences
Beatty sequences
characteristic sequences
spectra
digitized straight lines
mechanical sequences
cutting sequences
musical sequences

4. Much studied in mathematics, for its implications for number theory, there is the family of automatic sequences known as *Sturmian sequences*. The most famous example here is the Fibonacci sequence F, which is also a morphic sequence, obtainable by $\mathsf{0} \to \mathsf{01}$, $\mathsf{1} \to \mathsf{0}$ from start word $\mathsf{0}$, resulting in $\mathsf{01001010}\ldots$. This sequence F can be rendered in PSF as:

$$\begin{aligned} \mathsf{F} &= \mathsf{0} : \mathsf{1} : \mathsf{g}(\mathsf{tail}(\mathsf{F})) \\ \mathsf{g}(\mathsf{0} : \sigma) &= \mathsf{0} : \mathsf{1} : \mathsf{g}(\sigma) \\ \mathsf{g}(\mathsf{1} : \sigma) &= \mathsf{0} : \mathsf{g}(\sigma) \end{aligned}$$

Quite wonderfully, this sequence can also be obtained in a well-known direct geometrical way ('rotation sequences' or 'cutting sequences'), namely by intersections with the unit grid in the plane and the straight line from the origin with slope ϕ, with ϕ the golden ratio.

Sturmian Sequence

If a sequence has the property that the block growth function $B(n) = n + 1$ for all n, then it is said to have minimal block growth, and the sequence is called a Sturmian sequence. An example of this is the sequence arising from the substitution map

$$0 \to 01 \quad (1)$$
$$1 \to 0, \quad (2)$$

yielding $0 \to 01 \to 010 \to 01001 \to 01001010 \to \ldots$, which gives us the Sturmian sequence 01001010....

Sturm functions are sometimes also said to form a Sturmian sequence.

Jacques Charles François Sturm
Born
September 29, 1803
Geneva
Died
December 15, 1855
Paris
Nationality
French
Fields
Mathematics
Institutions
École Polytechnique
Known for
Sturm–Liouville theory
Sturm's theorem
Speed of sound
Notable awards
Légion d'Honneur (1837)
Copley Medal (1840)

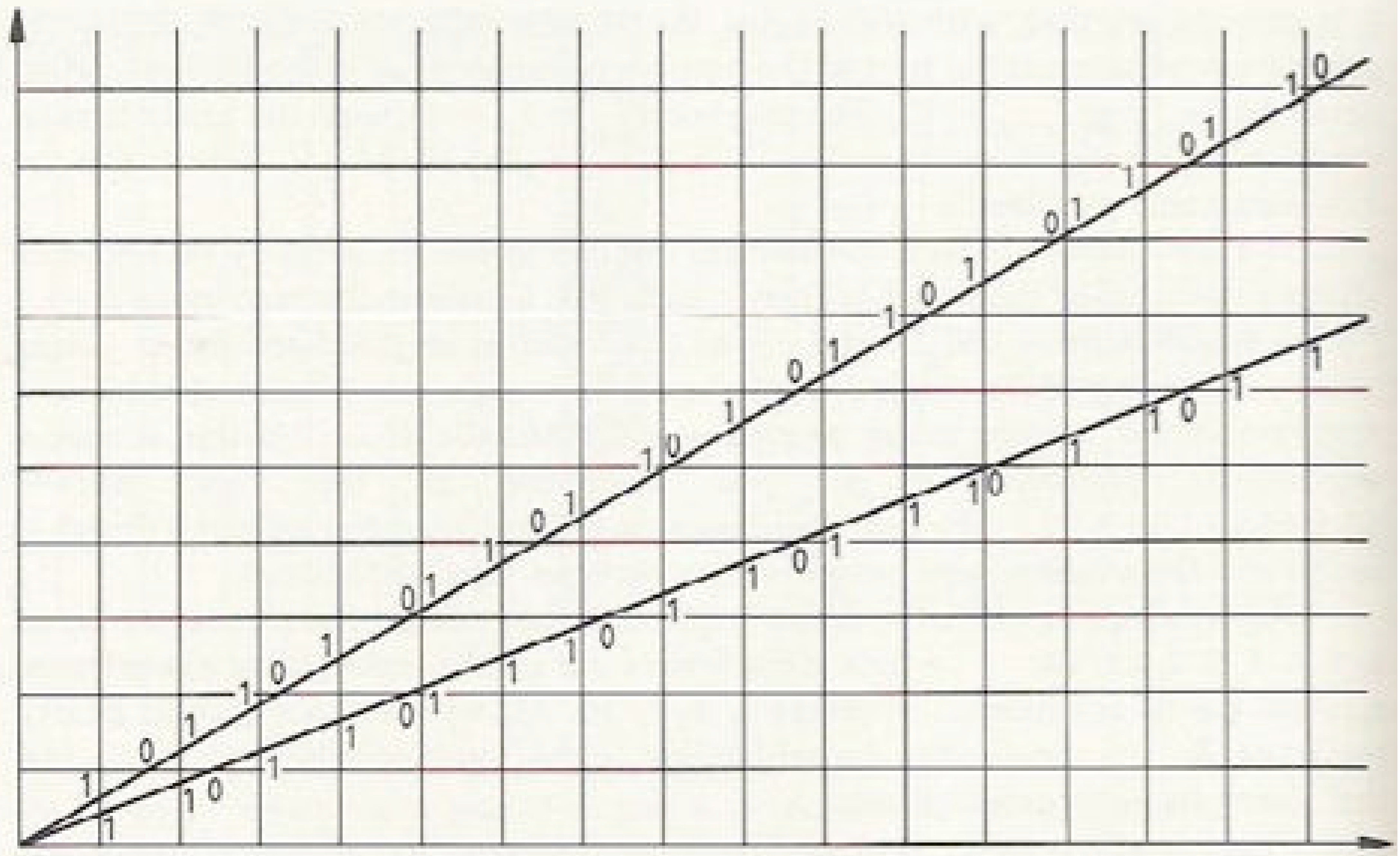


Figure 27 Square lattice and straight line with golden-mean slope γ generates the rabbit sequence 10110 . . . . The lower straight line has the silver-mean slope $\sqrt{2} - 1$ and generates another self-similar binary sequence.

0,1 stream is balanced if subwords of the same length have a number of 1's that differs at most 1.

stream is sturmian if it is aperiodic and balanced

morse stream 0110100110010110... is not balanced

fibonacci stream (obtained by morphism 0 -> 01, 1 -> 0 0100101001001010010100 1... is balanced

there are uncountably many sturmian streams.
They are not closed under diff; e.g. diff fibonacci is not balanced, being 0110001101100011000 11

for every finite balanced 0,1 word w there is a sturmian sequence containing w as factor (subword)

the structure of sturmian morphisms is exactly known, they form the monoid of Sturm and are composed of 3 basic ones.

Question: is every sturmian FST morphic?

sturmian streams can be defined as streams on {0,1} with subword complexity n+1

EXAMPLE 2.1.6. The height of $x = 0100101$ is 3, and its slope is 3/7. The word x can be drawn on a grid by representing a 0 (resp. a 1) as a horizontal (resp. a diagonal) unit segment. This gives a polygonal line from the origin to the point $(|x|, h(x))$, and the line from the origin to this point has slope $\pi(x)$. See Figure 2.1.

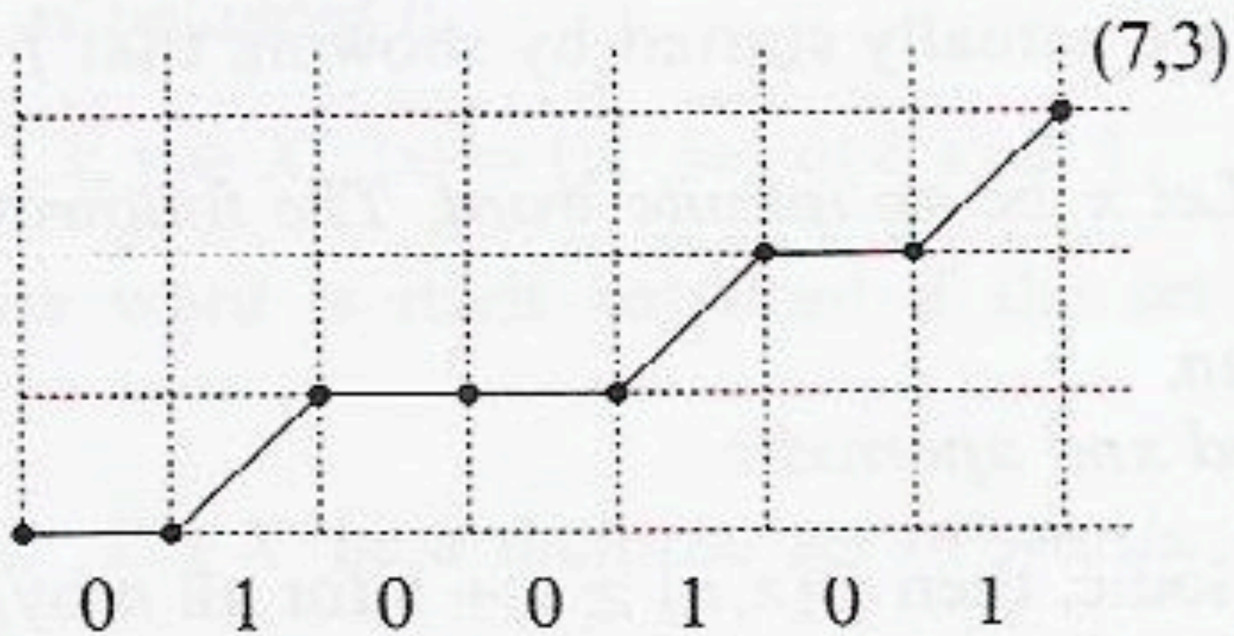


Figure 2.1. Height and slope of the word 0100101.

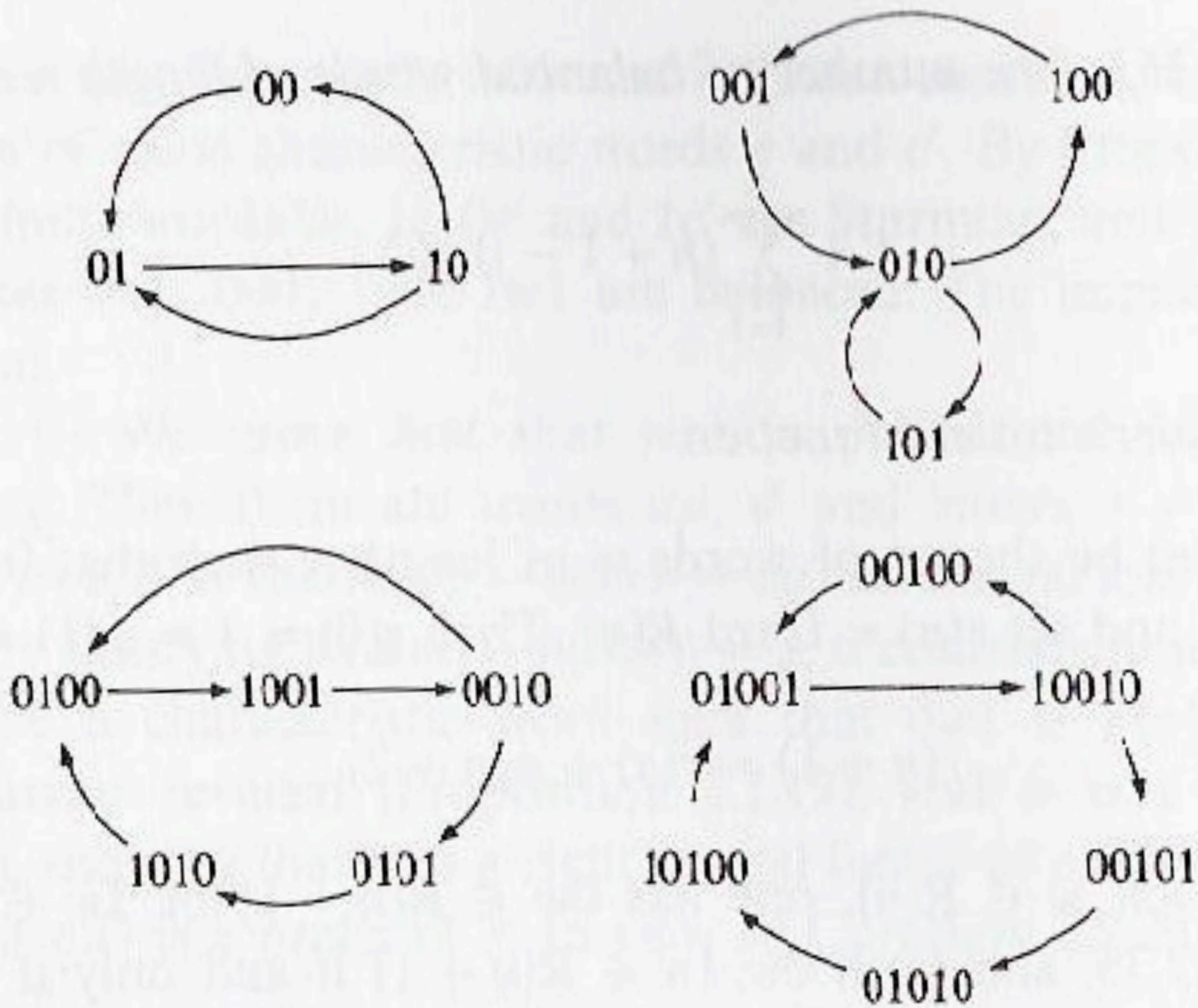


Figure 2.6. Factor graphs for the Fibonacci word.

An alternative way of specifying T (actually $\mathsf{inv}(\mathsf{T})$) is:

$$\mathsf{T}(n) = (\text{number of 2's in the prime factorization of } n) \mod 2$$

We recently considered a generalization of this scheme as follows. Let $\mathbf{P}$ be the set of prime numbers, and $A \subseteq \mathbf{P}$. We define the stream σ_A by: $\sigma_A(n)$ is the number of occurrences of factors in A in the prime factorization of n (modulo 2). E.g. for $A = \{2, 3\}$ we obtain $\sigma_{\{2,3\}}(90) = \sigma_{\{2,3\}}(\mathbf{2} \cdot \mathbf{3} \cdot \mathbf{3} \cdot 5) = 3 \mod 2 = 1$. We call such streams *prime-generated*. Thus $\sigma_{\{2\}}$ is T, and for $A = \{p \in \mathbf{P} \mid p \equiv 3 \pmod 4\}$ we obtain D.

(v) *Self-generating words* [7, 10, 16, 13, 36]. Self-generating words such as the Kolakoski word over the alphabet $\{1, 2\}$ which is identical to the sequence of its 'run-lengths', that is, the length of blocks of equal consecutive symbols:

$$\mathsf{K} = \mathsf{1221121221221121122121121221}\ldots$$

Kolakoski is the source of many open problems, e.g.: is the density of $1'$s in this sequence $\frac{1}{2}$?

(vi) *Recurrence equations* [54, 11]. Often the sequences in the families above are expressible by a set of recurrence equations; e.g. for M we have

$$\begin{aligned}
\mathsf{M}(0) &= 0 \\
\mathsf{M}(2n) &= \mathsf{M}(n) \\
\mathsf{M}(2n+1) &= 1 - \mathsf{M}(n)
\end{aligned}$$

An interesting infinite stream given by Conway is:

$$\begin{aligned}
\mathsf{C}(1) &= \mathsf{C}(2) = 1 \\
\mathsf{C}(n) &= \mathsf{C}(\mathsf{C}(n-1)) + \mathsf{C}(n - \mathsf{C}(n-1))
\end{aligned}$$

Even more general definitions, involving dependencies of stream entries on previous and even future entries, are the subject of the NWO project *LaPro*, mentioned in Section 2.

(vii) *Generalized Morse sequences*, introduced by Keane [33], an uncountable family of streams, containing morphic words over $\{0, 1\}$ where the morphism is of the form $0 \to w$, $1 \to \overline{w}$ for some word w, where $\overline{w}$ is obtained from w by flipping zeros and ones. An example is the Mephisto Waltz W, see Table 1.

Thue–Morse	$M = 0 : \mathsf{zip}_{1,1}(\mathsf{inv}(M), \mathsf{tail}(M))$ $\mathsf{zip}_{n,m}(\sigma, \tau) = \sigma(0) : \ldots : \sigma(n-1) : \mathsf{zip}_{m,n}(\tau, \mathsf{tail}^n(\sigma))$ $\mathsf{tail}(x : \sigma) = \sigma$ $\mathsf{inv}(0 : \sigma) = 1 : \mathsf{inv}(\sigma)$ $\mathsf{inv}(1 : \sigma) = 0 : \mathsf{inv}(\sigma)$
Period doubling	$T = \mathsf{zip}_{3,1}(w_T, T)$ $w_T = 1 : 0 : 1 : w_T$
Mephisto Waltz	$W = h_W(0 : \mathsf{tail}(W))$ $h_W(0 : \sigma) = 0 : 0 : 1 : h_W(\sigma)$ $h_W(1 : \sigma) = 1 : 1 : 0 : h_W(\sigma)$
Kolakoski	$K = 2 : 2 : f_1(\mathsf{tail}(K))$ $f_1(1 : \sigma) = 1 : f_2(\sigma)$ $f_1(2 : \sigma) = 1 : 1 : f_2(\sigma)$ $f_2(1 : \sigma) = 2 : f_1(\sigma)$ $f_2(2 : \sigma) = 2 : 2 : f_1(\sigma)$
Fibonacci	$F = h_F(1 : \mathsf{tail}(F))$ $h_F(1 : \sigma) = 1 : 0 : h_F(\sigma)$ $h_F(0 : \sigma) = 1 : h_F(\sigma)$
Dragon curve	$D = \mathsf{zip}_{1,1}(A, D)$ [14, 1] $A = 0 : 1 : A$

Generalized Morse sequences (Keane)

Using the notation, we may define the well-known Morse sequence x (see e.g. [4], [7], [8]) as an infinite "product" of blocks: set $b = (01)$ and $x = b \times b \times b \times \cdots$. In words, this rule says: first write down **01**, and then at each succeeding step write the mirror image of the complete previous production to the right of the same. The first 32 members of x are

$$01 \,|\, 10 \,|\, 1001 \,|\, 10010110 \,|\, 1001011001101001 \,|\, \ldots .$$

Let us denote by Ω the space of two-sided sequences of zeroes and ones, and by T the shift transformation on Ω. The following results were announced by S. Kakutani in [4]. If the Morse sequence x is continued to the left in a suitable manner to produce a point of Ω, then the orbit closure $\mathcal{O}_x$ of this point under T is a strictly ergodic subsystem of (Ω, T). Furthermore, T possesses partly continuous and partly discrete spectrum on $\mathcal{O}_x$ with respect to the uniquely determined probability measure on $\mathcal{O}_x$, and the group $\mathcal{G}_x$ of eigenvalues of T on $\mathcal{O}_x$ coincides with the group of all 2^k-th roots of unity.

In this paper we consider the infinite sequences which can be produced by the above-mentioned method of generating new sequences from old ones. For instance, if we set $b = (001)$, then $x = b \times b \times b \times \ldots$ defines a "ternary" sequence

$$x = (001\ \ 001\ \ 110\ \ 001\ \ 001\ \ 110\ \ 110\ \ 110\ \ 001\ \ \ldots),$$

[Kea68] M. Keane. Generalized Morse Sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 10(4):335–353, 1968.

We investigate Keane words, introduced as *recurrent sequences* in [Kea68].

Definition 1. Let $\mathbf{2} = \{0, 1\}$. For a word $x \in \mathbf{2}^\infty$, we write $\overline{x}$ for its *inverse*, obtained by changing 0s into 1s and vice versa. The operation $\times : \mathbf{2}^* \times \mathbf{2} \to \mathbf{2}^*$ is defined by:

$$u \times 0 = u \qquad\qquad u \times 1 = \overline{u}$$

for all $u \in \mathbf{2}^*$, and is extended to $\mathbf{2}^* \times \mathbf{2}^\infty \to \mathbf{2}^*$ by:

$$u \times \varepsilon = \varepsilon \qquad\qquad u \times a\sigma = (u \times a)(u \times \sigma)$$

for all $u \in \mathbf{2}^*$, $\sigma \in \mathbf{2}^\infty$ and $a \in \mathbf{2}$.

A word $u \in \mathbf{2}^*$ is called a *block* if $u = 0v$ for some non-empty word $v \in \mathbf{2}^+$. Let $\beta = u_0, u_1, u_2, \ldots$ be an infinite sequence of blocks u_i. Then, the *Keane word generated by* β, denoted by κ_β, is defined as the infinite product:

$$\kappa_\beta = u_0 \times u_1 \times u_2 \times \cdots \tag{1}$$

So, the product $u \times v$ is formed by concatenation of $|v|$ copies of either u or its inverse $\overline{u}$, so that $|u \times v| = |u| \cdot |v|$, taking the i-th copy as u if $v(i) = 0$, and as $\overline{u}$ if $v(i) = 1$, for $0 \leq i \leq |v| - 1$. Note that 0 is the identity element with respect to the $\times$-operation: $0 \times u = u = u \times 0$. Hence, if u is a block, then $u \times v$ is a proper extension of u.

For Keane words defined by an infinite product $u \times u \times u \times \cdots$ of a single block u — *uniform* Keane words as we call them — there is a simple recursive definition:

$$\kappa_u = u \times \kappa_u \tag{2}$$

Now, in order to see that this equation indeed defines an infinite word, define $u^{(n)}$ by $u^{(0)} = 0$ and $u^{(n+1)} = u \times u^{(n)}$. Then, by definition of u being a block (begins with a 0, has length $|u| \geq 2$), larger and larger prefixes stabilize. Indeed, we observe that $u^{(n)}$ is a proper prefix of $u^{(n+1)}$, for all $n \in \mathbb{N}$. Hence, the solution of the equation (2) is a unique and infinite sequence.

However, orienting the equation (2), and the defining rules for $\times$, from left to right, does not yield a *productive* rewrite system. Explain why! But, using a simple trick, we can find a productive rewrite system for uniform Keane words. To turn (2) into a productive specification we replace the occurrence of κ_u in the right-hand side with $(0{:}\mathsf{tail}(\kappa_u))$, which is justified because all Keane words begin with a 0, and obtain:

$$\kappa_u \to u \times (0 : \mathsf{tail}(\kappa_u)) \tag{3}$$

The operation $\times$ is left-distributive over word concatenation:

Lemma 2. $u \times vw = (u \times v)(u \times w)$, *for all* $u, v \in 2^*$ *and* $w \in 2^\infty$. $\square$

Lemma 3. *On finite words, the operation* $\times$ *is associative.*

Proof. Let $u, v, w \in 2^*$. We prove $(u \times v) \times w = u \times (v \times w)$ by induction on w. The base case $w = \varepsilon$ is trivial. So let $w = aw'$ for some $a \in 2$ and $w' \in 2^*$. Then we get $(u \times v) \times w = ((u \times v) \times a)((u \times v) \times w')$, and $u \times (v \times w) = u \times ((v \times a)(v \times w')) = (u \times (v \times a))(u \times (v \times w'))$ by Lemma 2, and we conclude by the induction hypothesis for w'. $\square$

Crazy syntax in the following definition.

Definition 4. Let $u = \mathsf{cons}_*(a_0, \mathsf{cons}_*(a_1, \ldots))$ be a block. Then the iTRS $\mathcal{T}_u$ is defined by $\mathcal{T}_u = \langle \Sigma_u, R_u \rangle$ where $\Sigma_u = \{\mathsf{K}_u, \mathsf{nil}, \mathsf{cons}_*, \mathsf{cons}_\infty, \mathsf{mult}_1, \mathsf{mult}_\infty,$

Classifying Streams

The Landscape of Streams	jan willem
Spot the difference	jan willem
Prime-generated streams	dimitri
Degrees of streams	joerg

THE MORSE AND TOEPLITZ SEQUENCE

morse 1001011001101001

toeplitz 101110101011101110

toeplitz = diff morse

morse: DOL system $1 \rightarrow 10$, $0 \rightarrow 01$, start 1

toeplitz: DOL system $1 \rightarrow 10$, $0 \rightarrow 11$, start 1

toeplitz T is defined by

T = 1: zip(inv(T), ones)

1 0 1 1 1 0 1 0 1 0 1 1 1 0 1 1 1 0
1 0 1 1 1 0 1 0 1

Term rewriting does not only allow to specify individual streams, but moreover facilitates defining operations on streams. One well-known transformation, to which we return later, is the 'first difference operator' δ, for 01-streams σ:

$$\delta(x : y : \sigma) \to (x + y) : \delta(y : \sigma)$$

where $+$ is addition modulo 2. For example, we have $\mathsf{T} = \delta(\mathsf{M})$, that is, the period doubling sequence T is the first difference of Thue–Morse M.

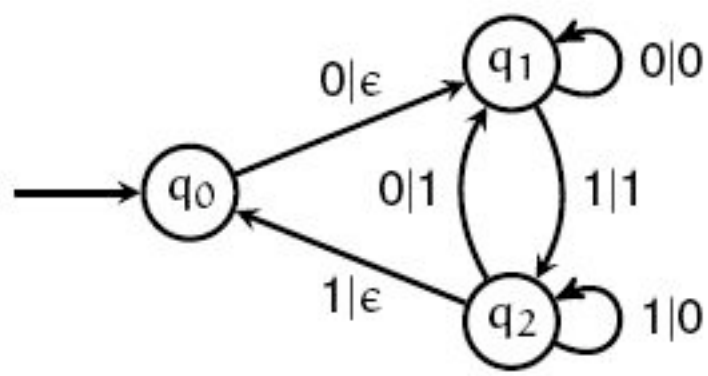
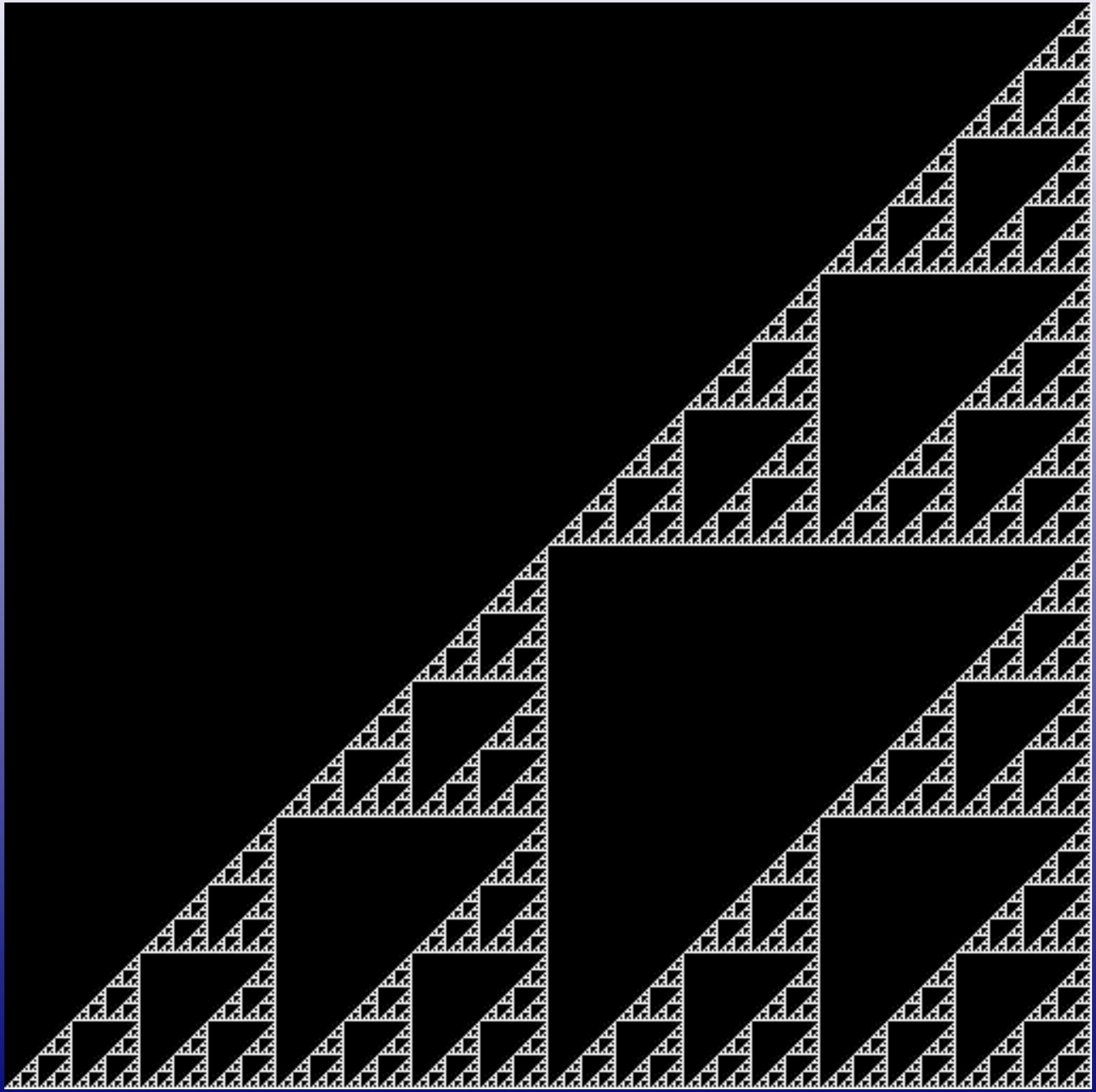


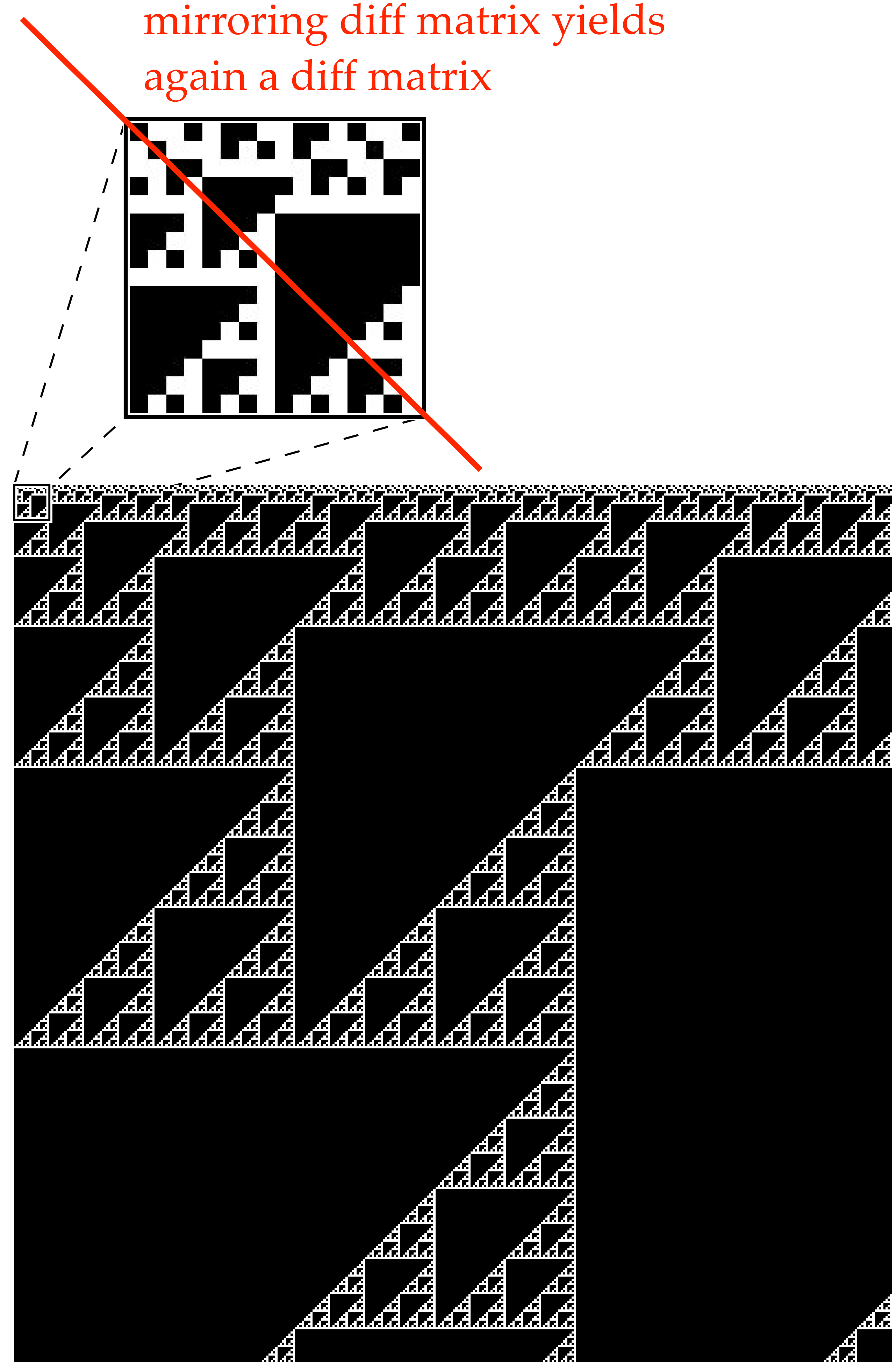
Figure 3: *From Thue–Morse (*M*) to the period doubling sequence (*T*).*

eventually periodic diff matrix of

$0^{511}\,1\,0^{\omega}$

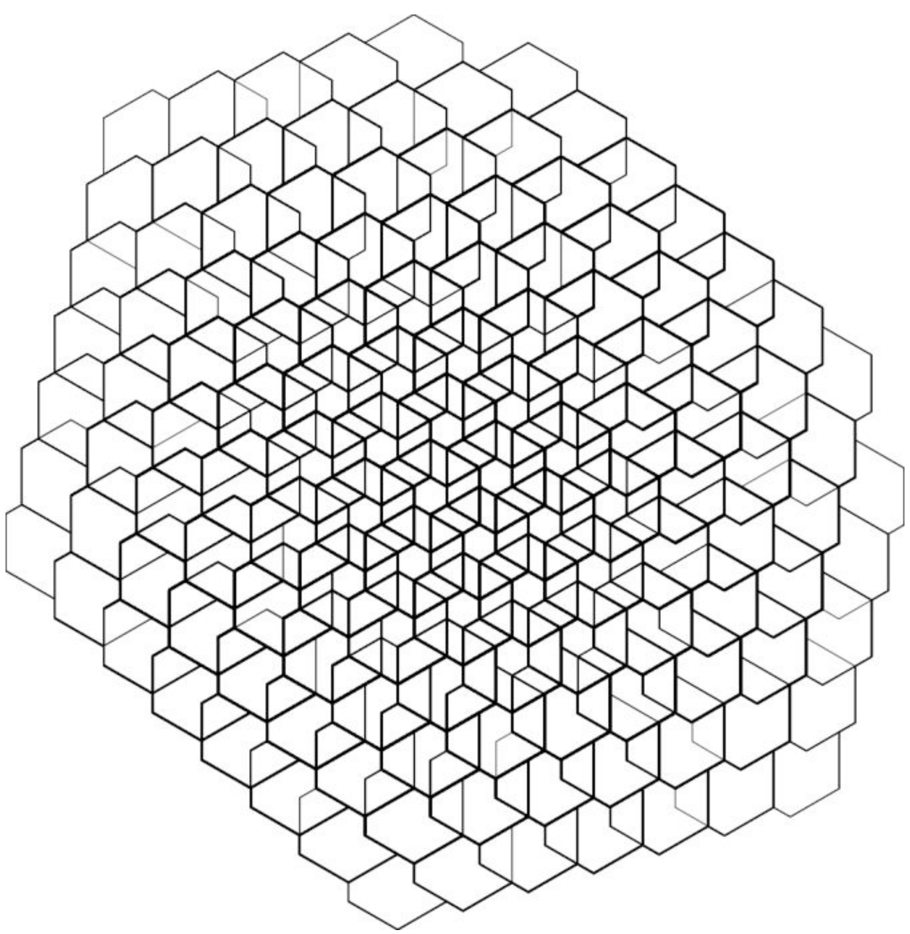


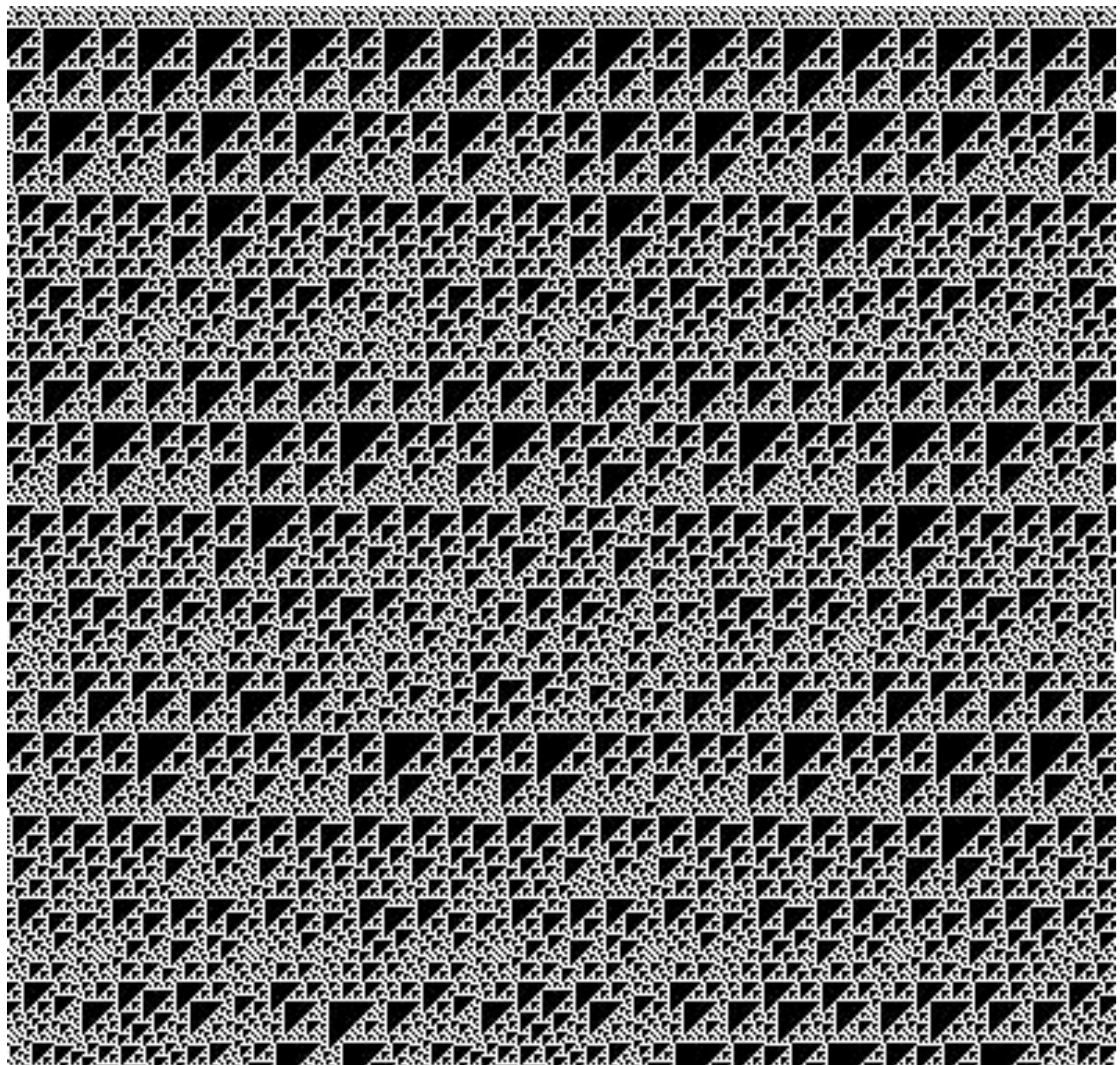
not eventually periodic diff matrix of morse stream



Fibonacci

$$\mathsf{F} = \mathsf{h_F}(1 : \mathsf{tail}(\mathsf{F}))$$
$$\mathsf{h_F}(1 : \sigma) = 1 : 0 : \mathsf{h_F}(\sigma)$$
$$\mathsf{h_F}(0 : \sigma) = 1 : \mathsf{h_F}(\sigma)$$





Fingerprint of Kolakoski stream

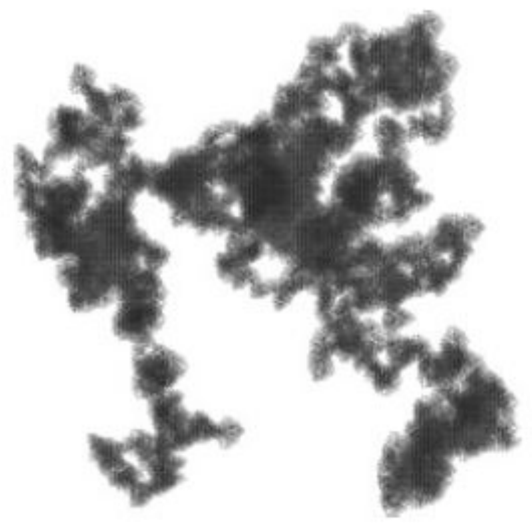
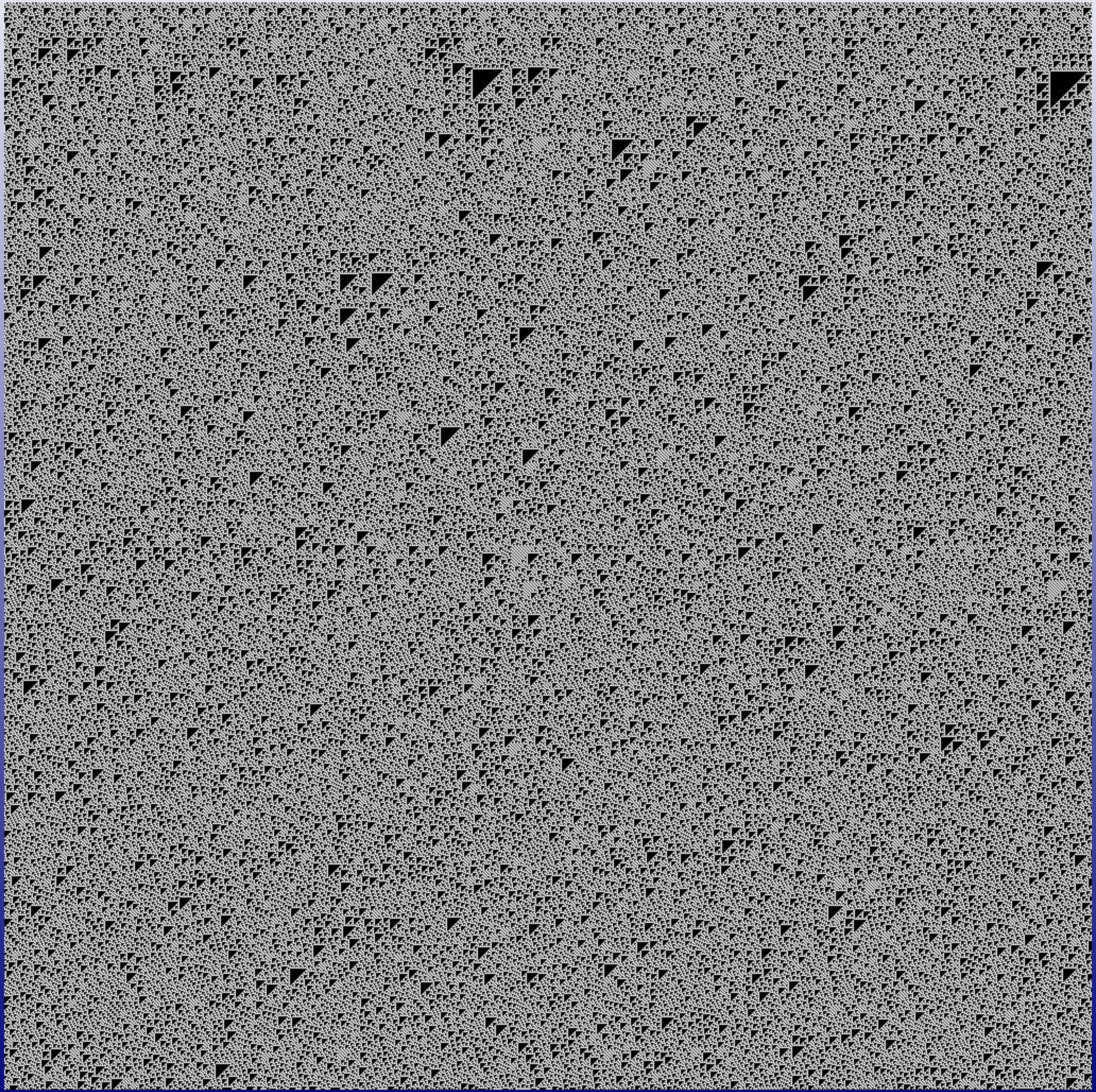


Figure 6: *A turtle trajectory for the Kolakoski sequence* K *for a prefix of* $2 \cdot 10^6$ *entries.*



... infamous Kolakoski stream, that verbally can be defined as the sequence that is equal to its own ‘runlength’ sequence. This mysterious description needs explanation.

The alphabet in which the sequence is written, is {1,2}. We could have used 0,1 but for pychological reasons 1,2 is more appropriate. If 111 22 1111 2 111 2222 ... is some 1,2-sequence, the sequence of its runlengths is 3,2,4,1,3,4,...; a run is the length of a maximal subword consisting of identical letters. So the Kolakoski sequence, described in 1965, is

22 11 2 1 22 1 22 11 2...

The Sierpinsky stream S

$$\mathsf{S} = \mathsf{zip}_{8,1}(\mathsf{w_S}, \mathsf{S})$$

$$\mathsf{w_S} = 1 : 1 : 0 : 0 : 0 : 0 : 1 : 1 : \mathsf{inv}(\mathsf{w_S})$$

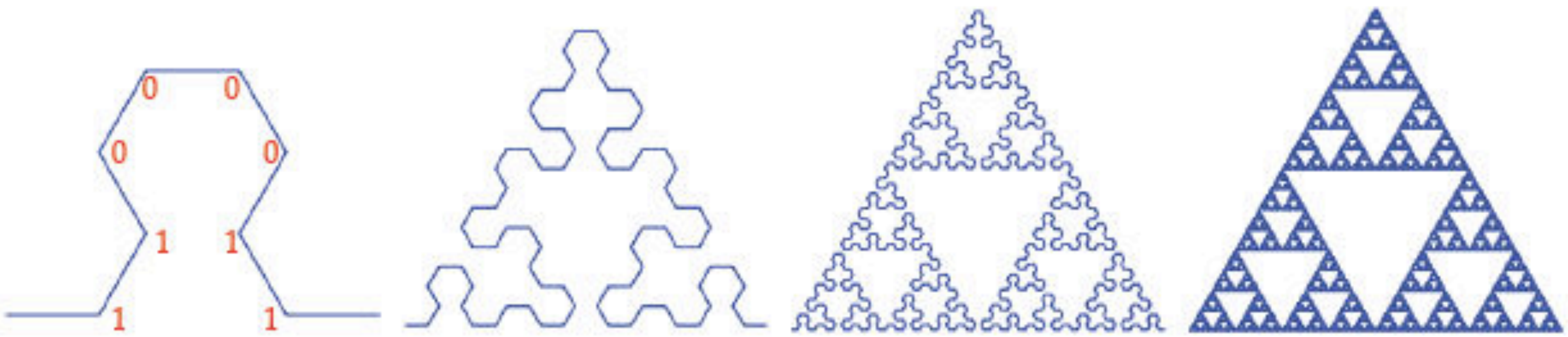


Figure 8: *Construction of the Sierpiński triangle.*

As far as we know, the Sierpiński stream S does not occur in the literature:

$$\mathsf{S} = 110000111\ 001111001\ 110000110\ 001111000\ 110000110\ \ldots$$

The Sierpiński arrowhead curve is obtained back from S by interpreting its entries as turtle drawing instructions (turning angle $0 \mapsto -\pi/3$, $1 \mapsto \pi/3$).

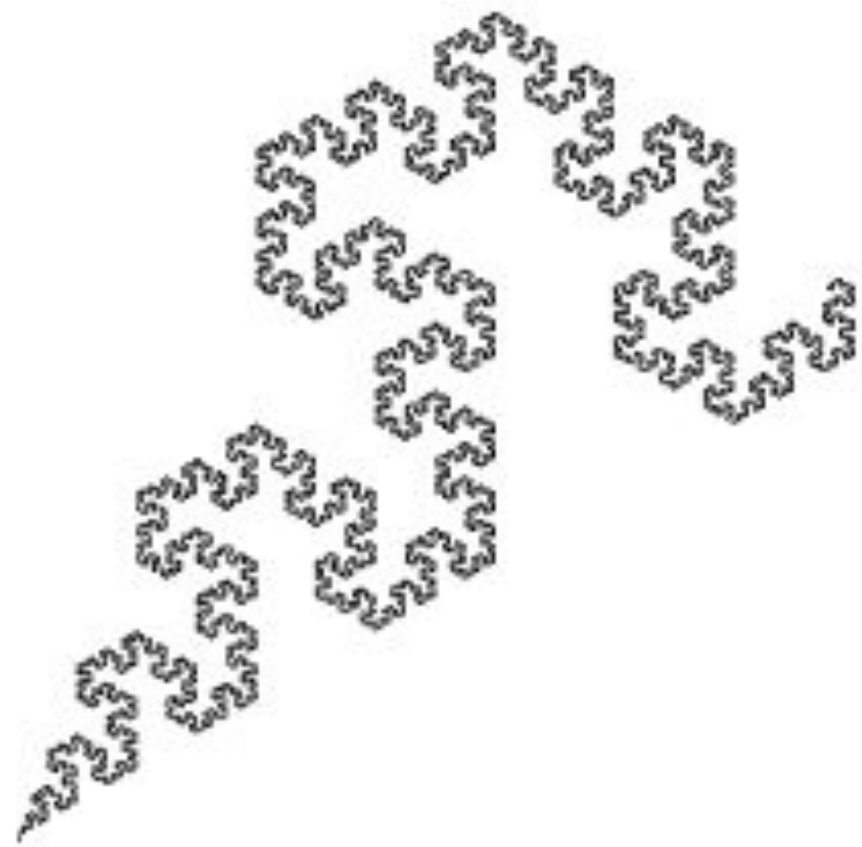
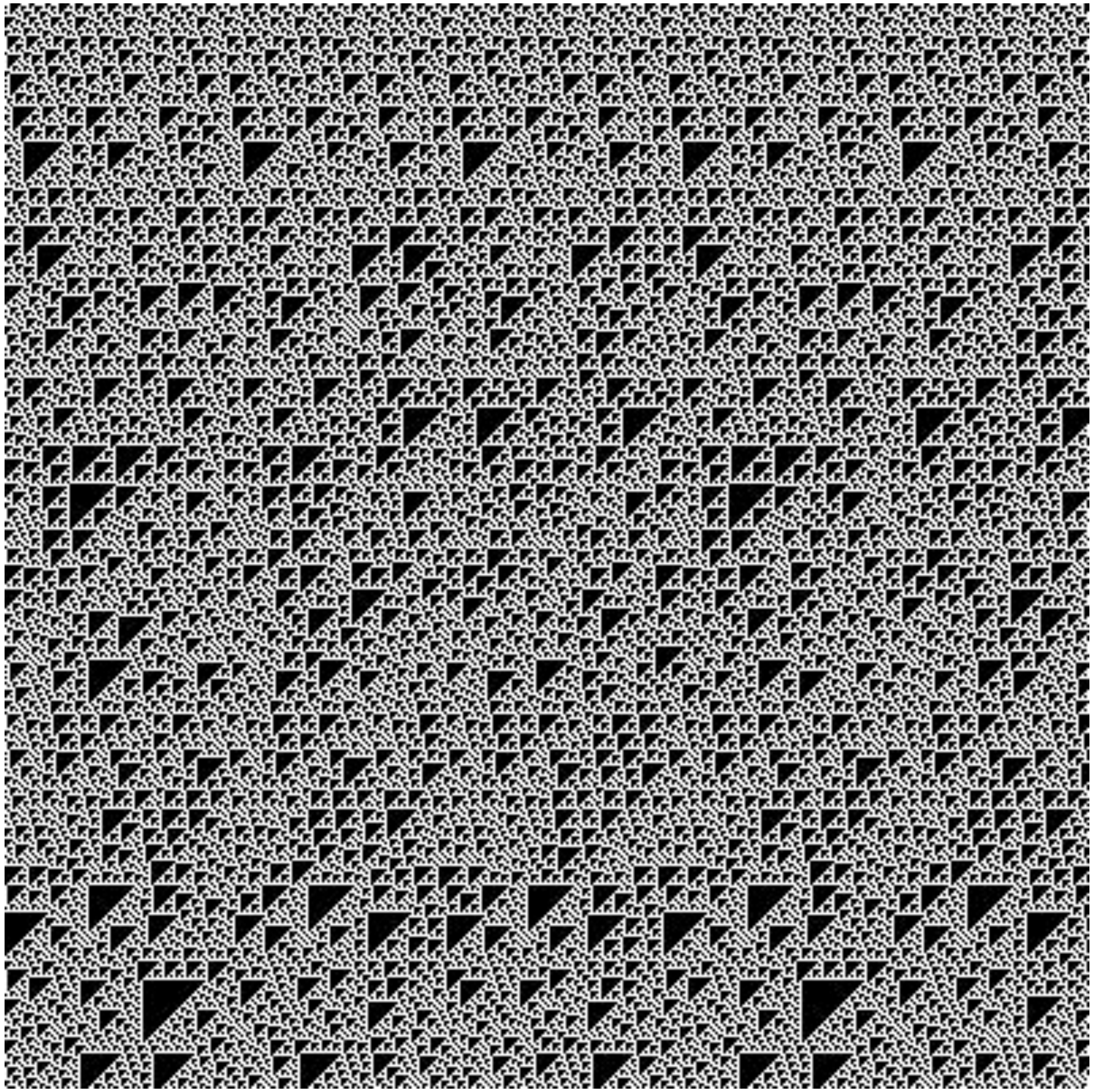
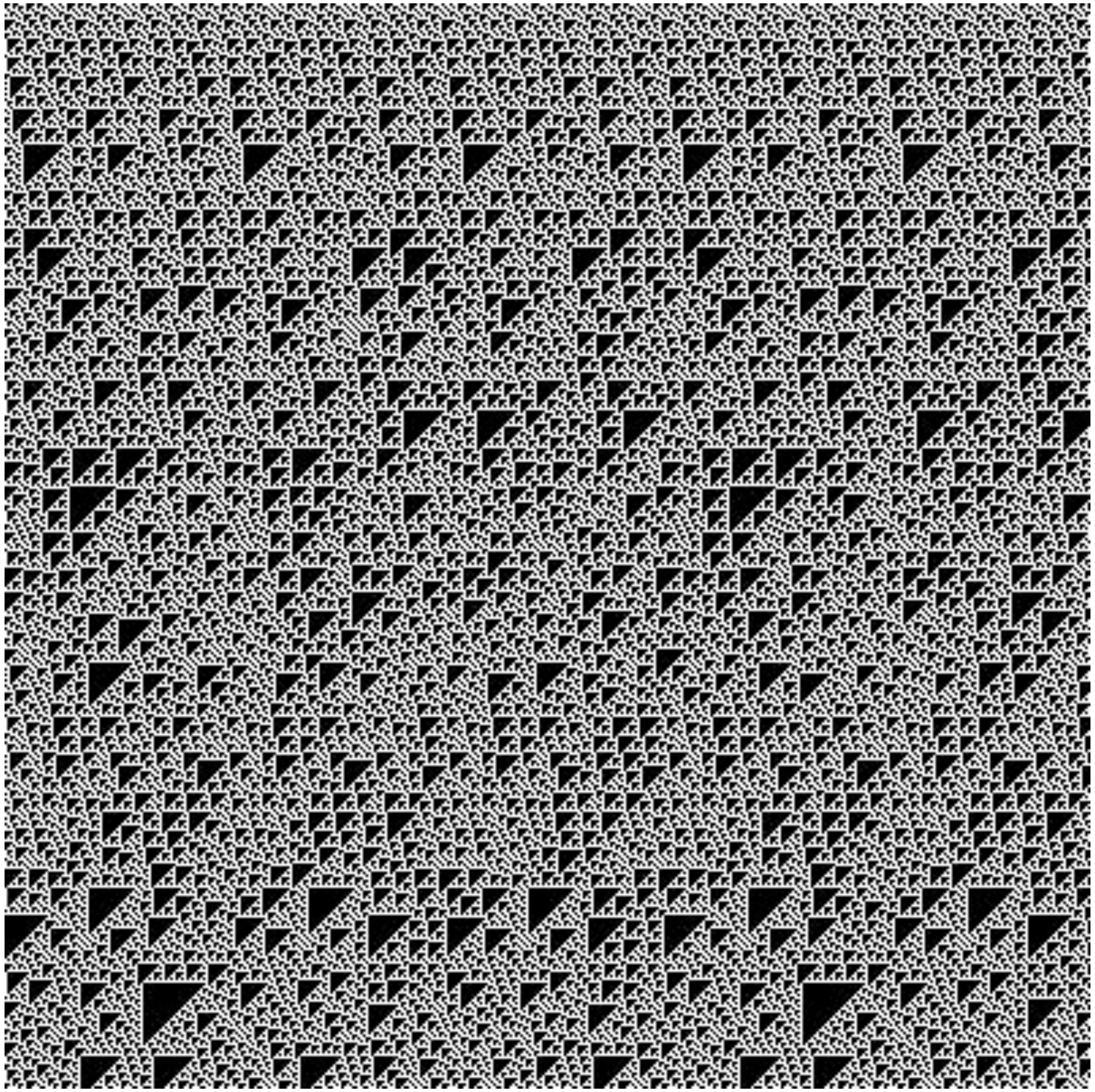


Figure 8: *A turtle trajectory for the Mephisto Waltz, the stream which can be obtained from the morphism* $0 \rightarrow 001, 1 \rightarrow 110$ *on the initial word* 0.

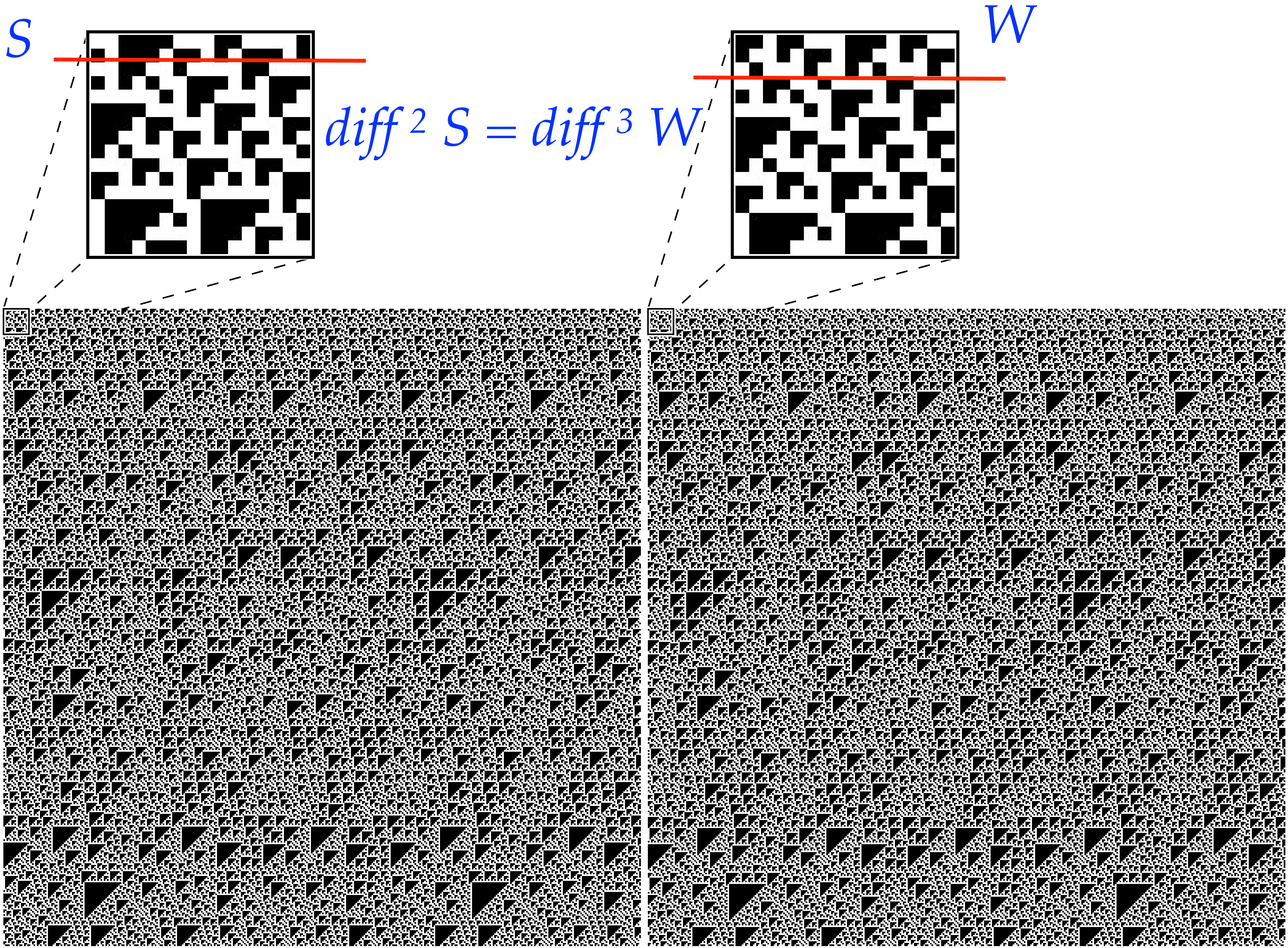
spot the difference



Sierpinski S



Mephisto W

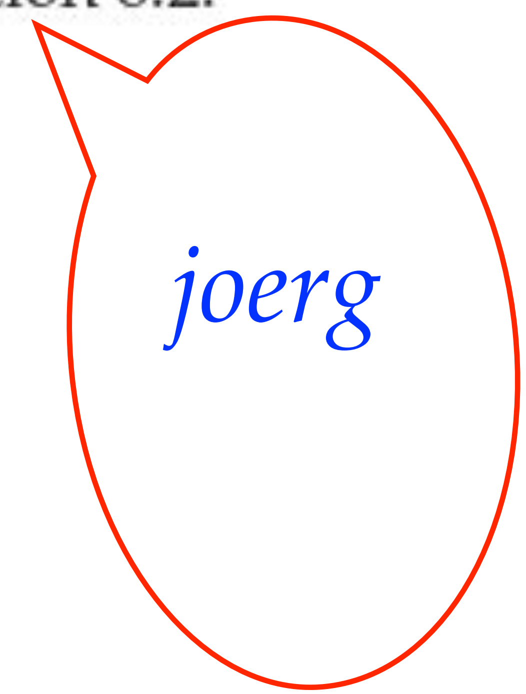

S
W
$diff^2\, S = diff^3\, W$

We also give an example showing how significant information can be detected from a consideration of these 'fingerprint' patterns exhibited by the δ-orbits, displayed as matrices as in Figures 9, and 10.

Namely, in an experiment it turned out (see Figure 10) that the δ-matrix of the Sierpiński stream S and the Mephisto Waltz W of Keane [33] are after the first couple of rows exactly the same. In this way we find that

$$\delta^2(\mathsf{S}) = \delta^3(\mathsf{W})$$

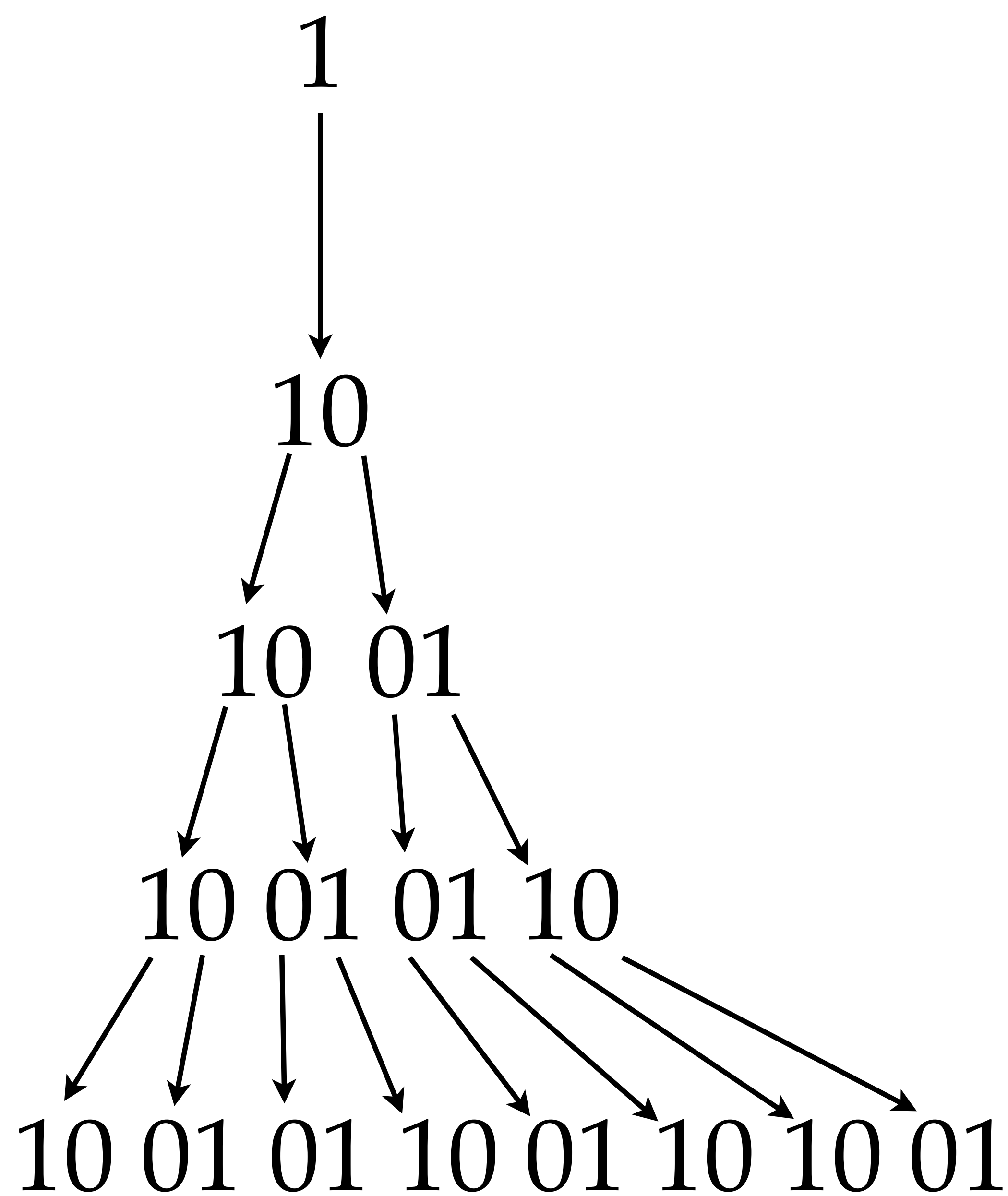
a curious fact that seems hard to find or guess otherwise, because S and W seem totally unrelated in their definition. So $\mathsf{S} \diamond \mathsf{W}$: thus the graphical analysis yields information about the degree hierarchy of Section 6.2.

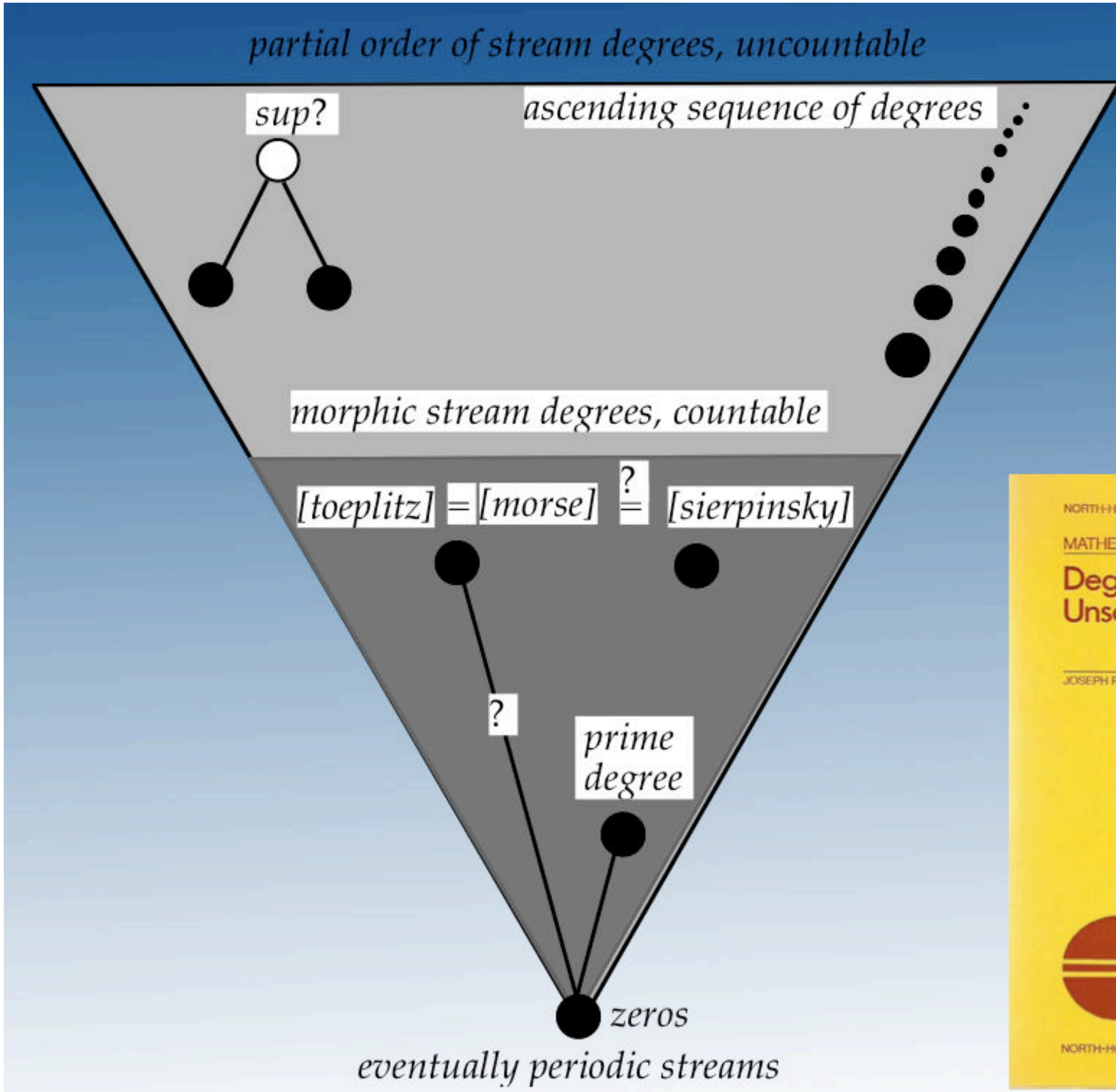


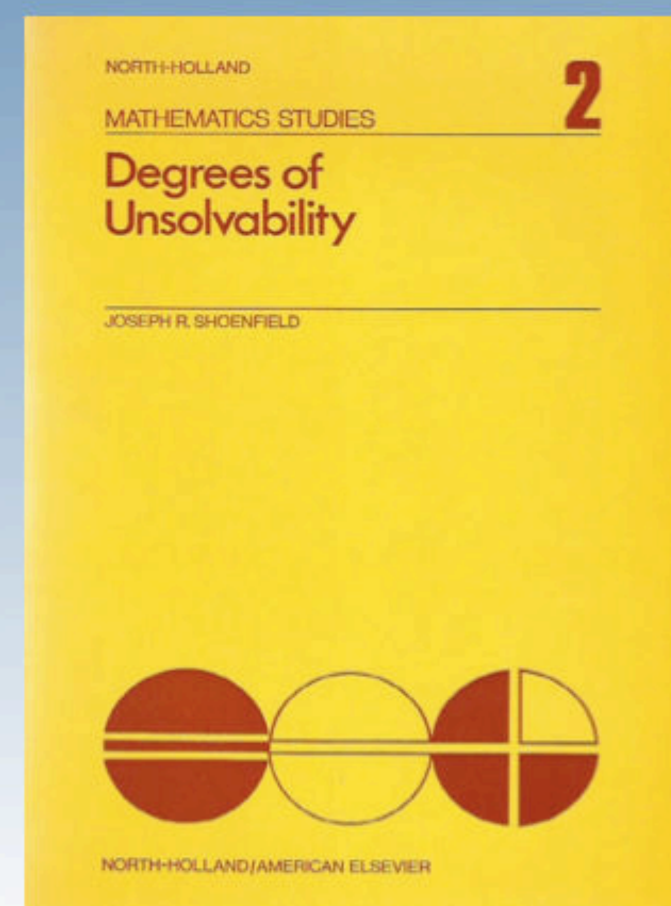
morse als D0L systeem

regels $1 \rightarrow 10$, $0 \rightarrow 01$,
startwoord 1

morphic sequence




partial order of stream degrees, uncountable
sup?
ascending sequence of degrees
morphic stream degrees, countable
[toeplitz] = [morse] ?= [sierpinsky]
?
prime degree
zeros
eventually periodic streams


NORTH-HOLLAND
MATHEMATICS STUDIES
2
Degrees of Unsolvability
JOSEPH R. SHOENFIELD
NORTH-HOLLAND/AMERICAN ELSEVIER

This shows $\mathsf{M} \rhd \mathsf{T}$. The reverse direction $\mathsf{T} \rhd \mathsf{M}$ holds as well. So M and T are 'twin brothers', belonging to the same degree in our hierarchy of streams.

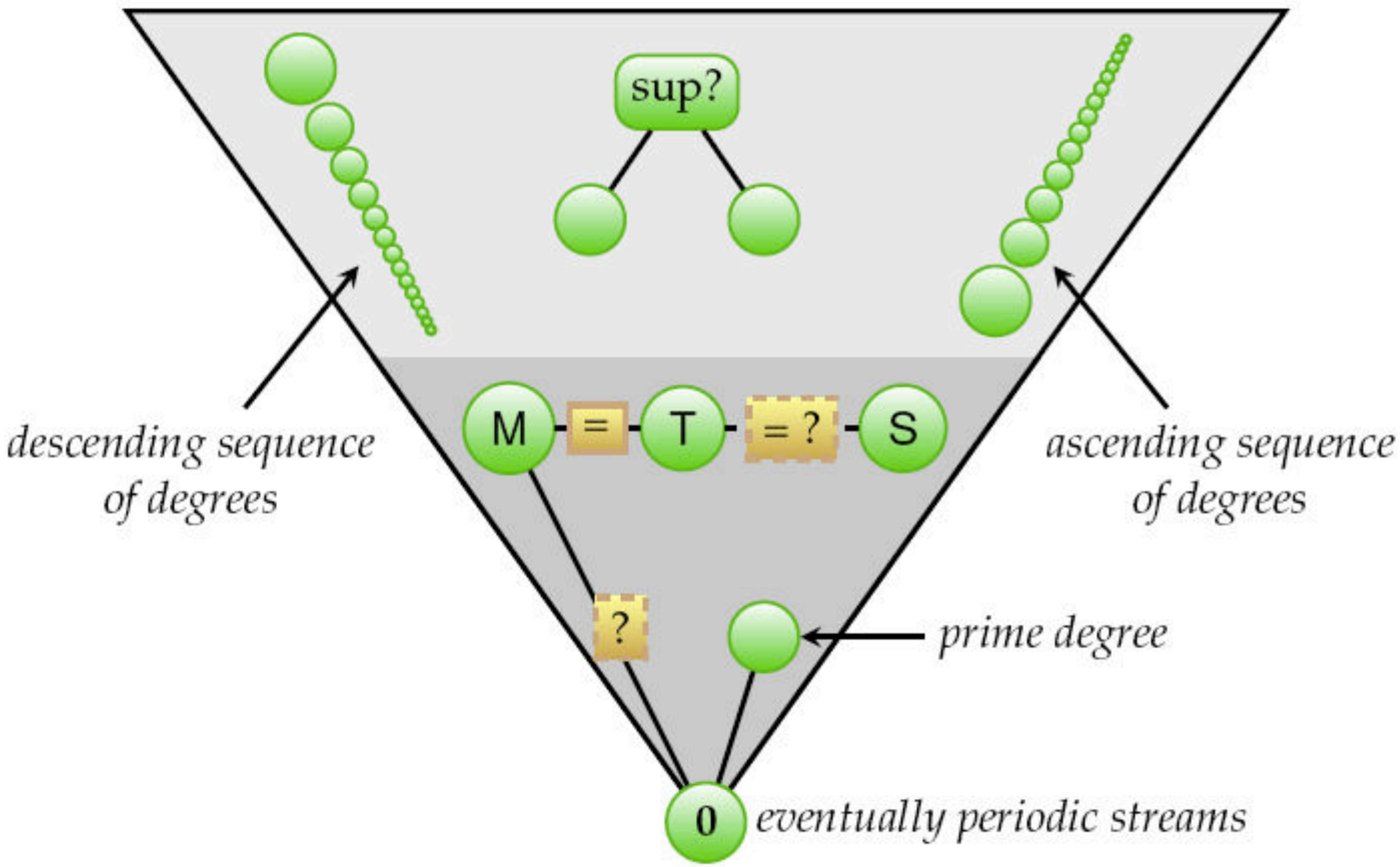
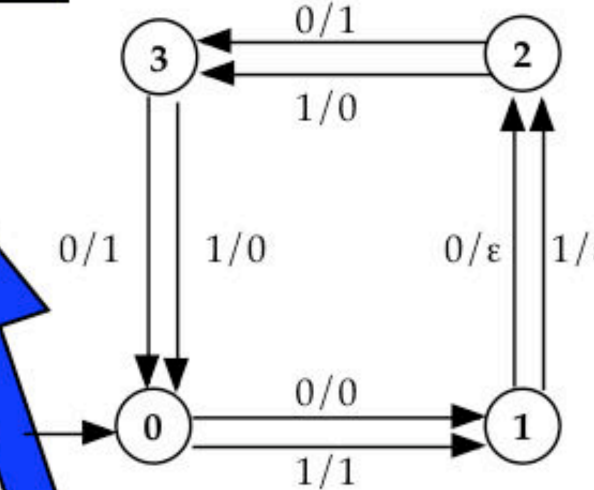


Figure 4: *Uncountable partial order of stream degrees. The darker, countable part consists of morphic degrees.*

two failing experiments to cut the morse diamond into smaller nontrivial (i.e. not evt periodic) pieces

morse = 1001011001101001...

3-morse = 100 101 100 110 100 1...



morse/3 = 100 101 100 110 100 1...
= 11111110111111011111110...

Speakers: Dimitri Hendriks, Joerg Endrullis, Jan Willem Klop
Title: Classifying streams

Abstract:
In the first part we will give a survey on the landscape of infinite streams, presenting next to the well-known families such as morphic streams, automatic streams, sturmian streams also some less well-known families, such as generalized morse streams, streams defined by recurrence relations, and prime-generated streams. In the second part we present a definition of stream reducibility using FSTs, finite state transducers, leading to an interesting hierarchy of stream degrees whose structure is largely unexplored. We present some initial observations, and state our favourite conjecture about our favourite stream.

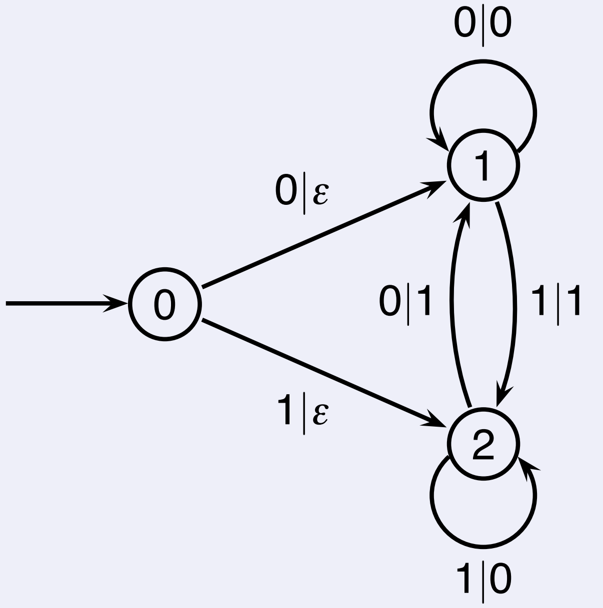
Transducing streams

We transduce streams using deterministic Mealy automata (DMA).

- output words $\in \Sigma^*$ along the edges

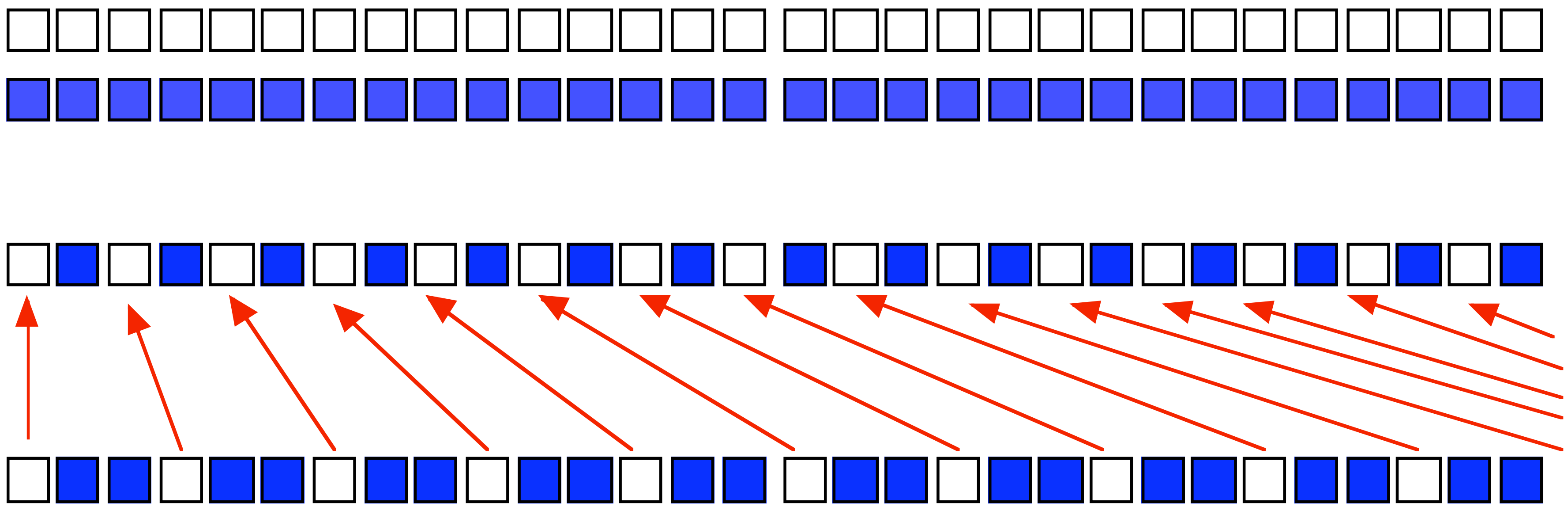
Example

The following automaton computes the diff of a stream:



Thus it reduces Morse to Toeplitz.

$$01101001\ldots \to 1011101\ldots$$



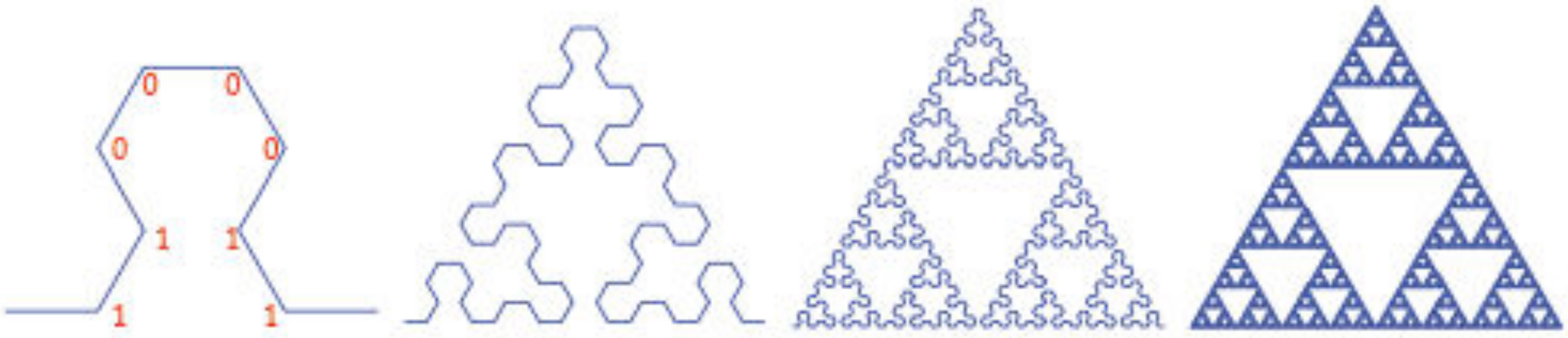
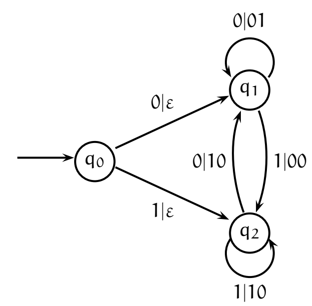


Figure 8: *Construction of the Sierpiński triangle.*



en heb voor de lol de FST getekend, zie attached.

Als pure stream functie is het ook makkelijk:

mfm(0:0:s) = 0:1:mfm(0:s)
mfm(0:1:s) = 0:0:mfm(1:s)
mfm(1:0:s) = 1:0:mfm(0:s)
mfm(1:1:s) = 1:0:mfm(1:s)

of, overkomend de FST:

mfm(0:s) = q1(s)
mfm(1:s) = q2(s)
q1(0:s) = 0:1:q1(s)
q1(1:s) = 0:0:q2(s)
q2(0:s) = 1:0:q1(s)
q2(1:s) = 1:0:q2(s)

time between the pulses. Every time a 1 is read, the clock is synchronized. However, for a long sequence of 0s, clock error accumulates, which may cause the data to read incorrectly. To counteract this effect the encoded sequence is required to have no long stretches of 0s.

A common coding scheme called *modified frequency modulation* (MFM) inserts a 0 between each two symbols unless they are both 0s, in which case it inserts a 1. For example, the sequence

$$10100110001$$

is encoded for storage as

$$100010010010100101001.$$

This requires twice the length of the track, but results in fewer read/write errors. The set of sequences produced by the MFM coding is a sofic system (Exercise 3.8.3).

There are other considerations for storage devices that impose additional conditions on the sequences used to encode data. For example, the total

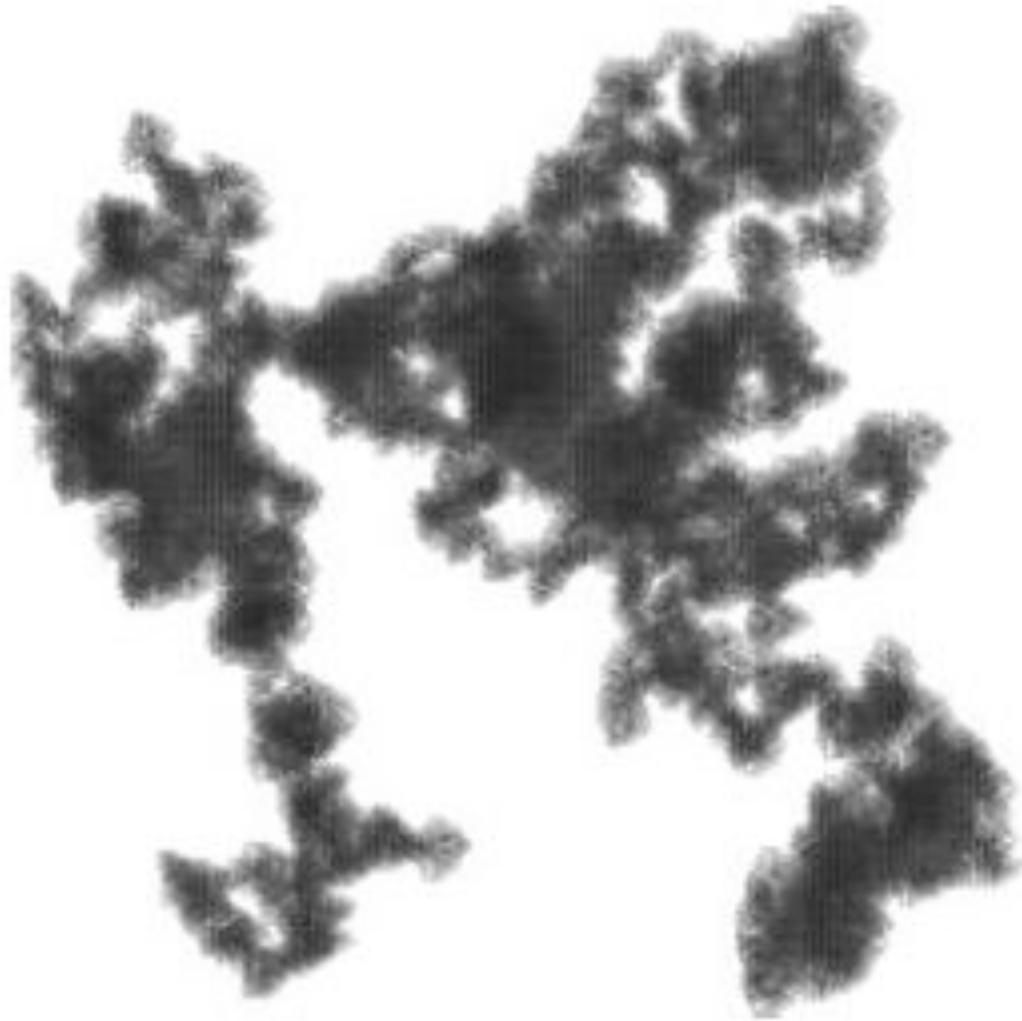


Figure 6: *A turtle trajectory for the Kolakoski sequence* K *for a prefix of* $2 \cdot 10^6$ *entries.*