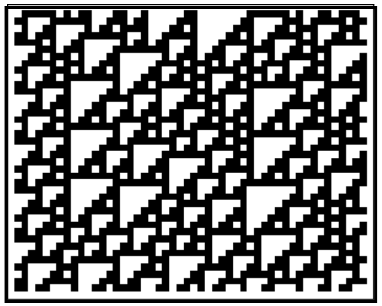
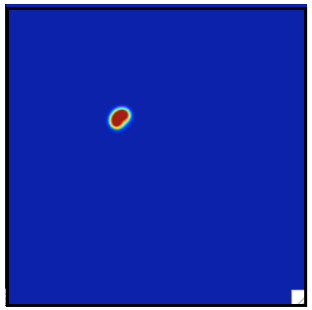
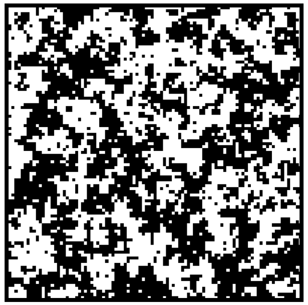
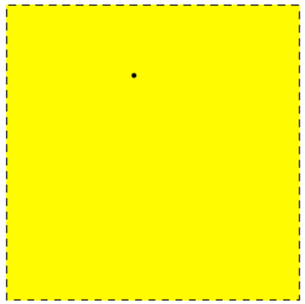
Information theory for complex systems









1: Cellular automata 2: Pattern formation 3: Spinn systems and Baker's map

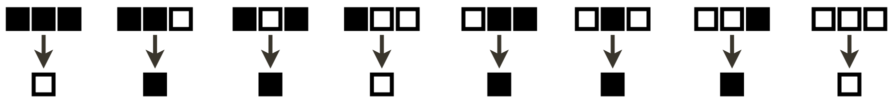
Kristian Lindgren
Complex systems group, Department of Energy and Environment
Chalmers University of Technology, Gothenburg, Sweden

Cellular automata and information

- 1-dimensional CA
- Elementary CA rules
  - Binary state (0 or 1, ”white” or ”black”)
  - Nearest neighbour interaction
  - CA state: bi-infinite sequence (... 0 1 1 1 0 0 1 1 ...)
- Dynamics given by deterministic local rule, updating all cells in parallel

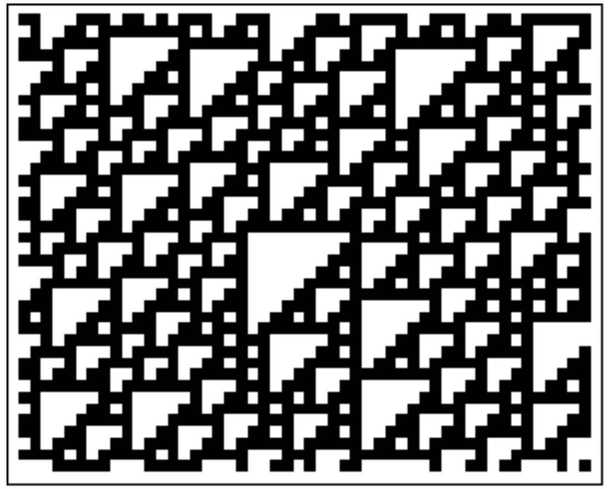
Rule table

- Example: Rule 110



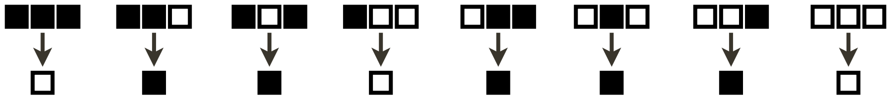
111	110	101	100	011	010	001	000
0	1	1	0	1	1	1	0

$$(01101110)_2=110$$



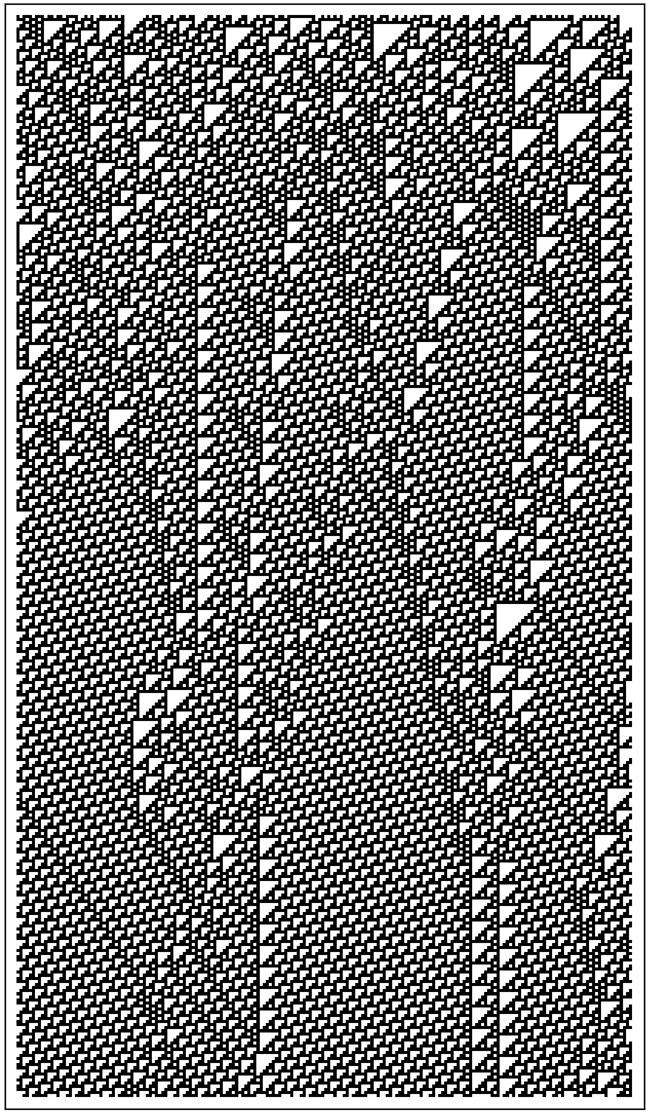
Rule table

- Example: Rule 110



111	110	101	100	011	010	001	000
0	1	1	0	1	1	1	0

$$(01101110)_8 = 110$$



CA classes

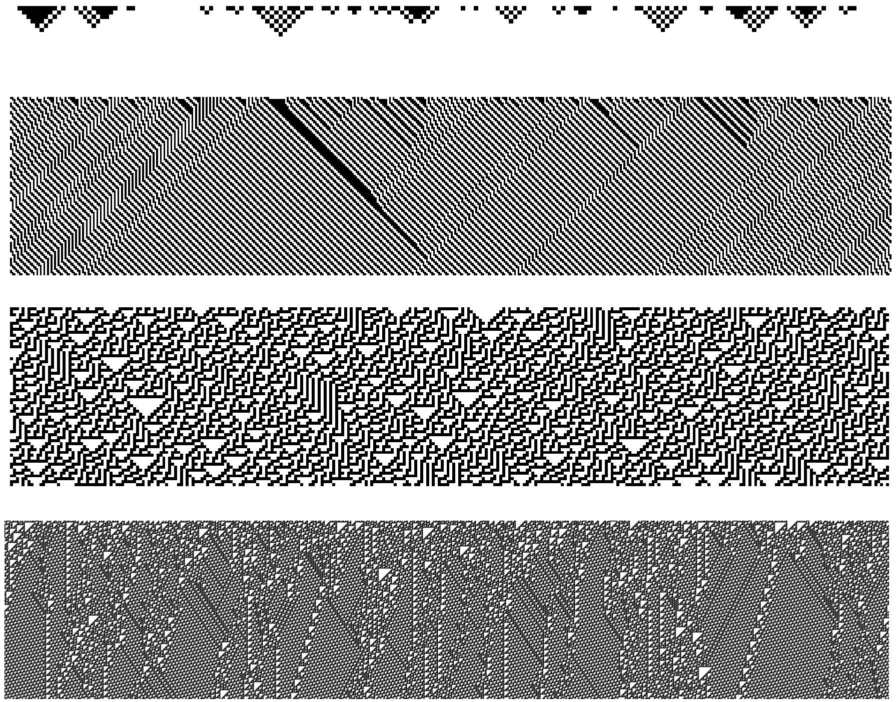
Four classes of dynamics:

(I) Towards homogenous fixed point.

(II) Towards inhomogenous fixed point, shift, and/or periodic behavior.

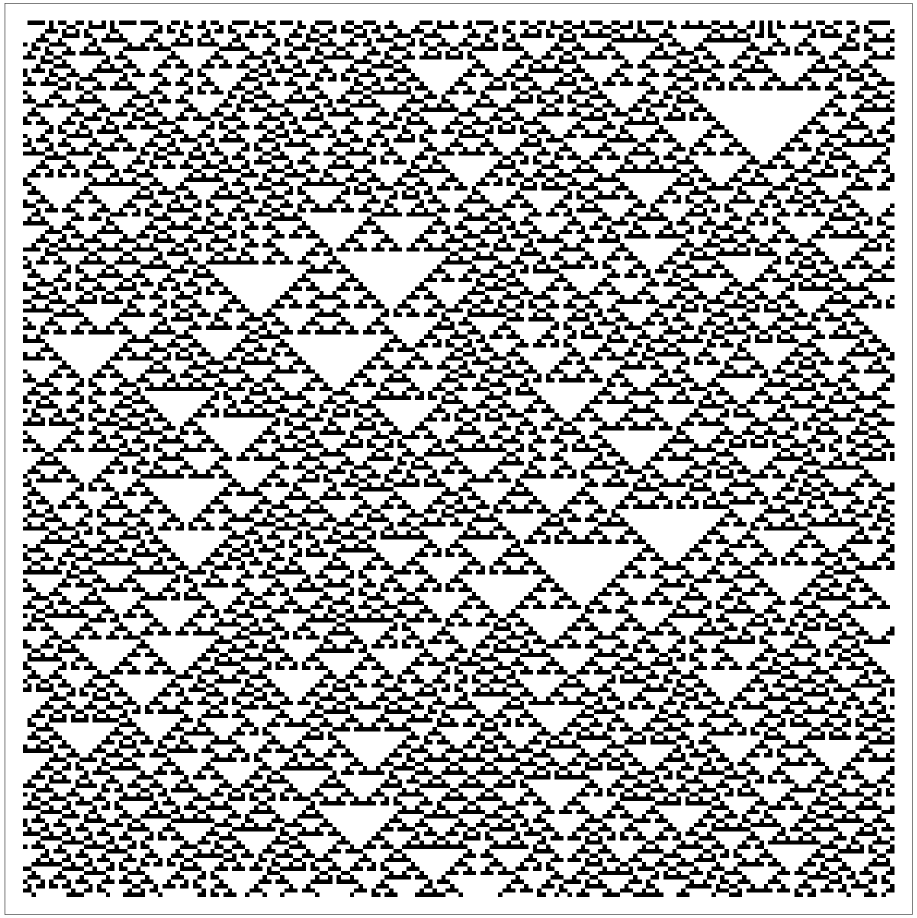
(III) Irregular behavior – ”chaotic”

(IV) In between (II) and (III); long transients, ”complex”.



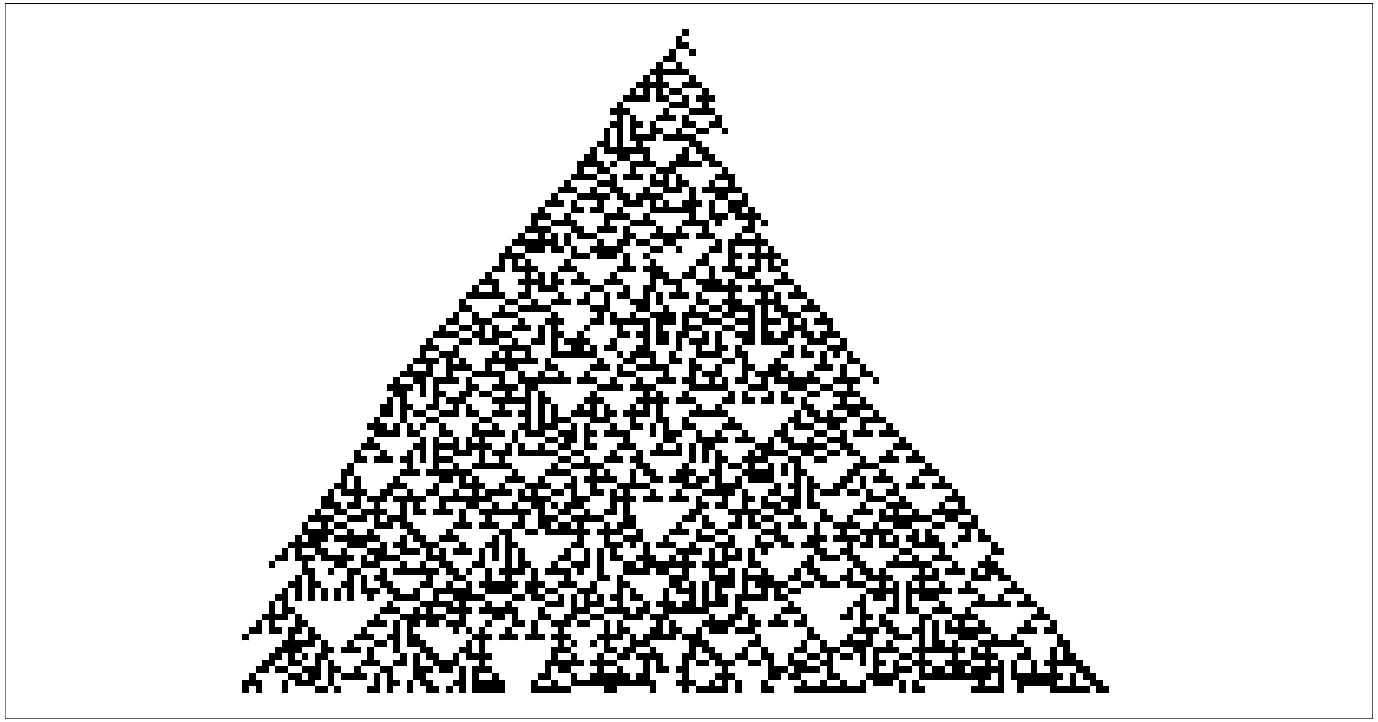
Class III example: R22

"Chaotic"



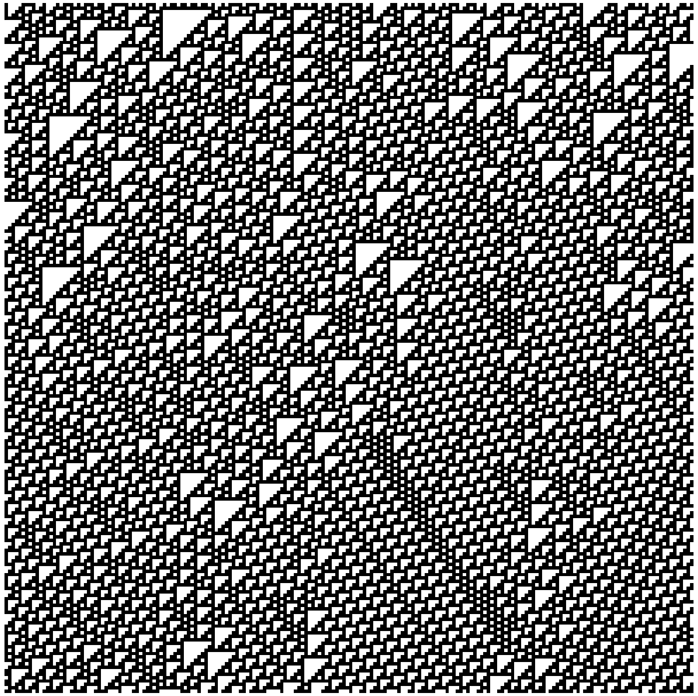
Difference pattern

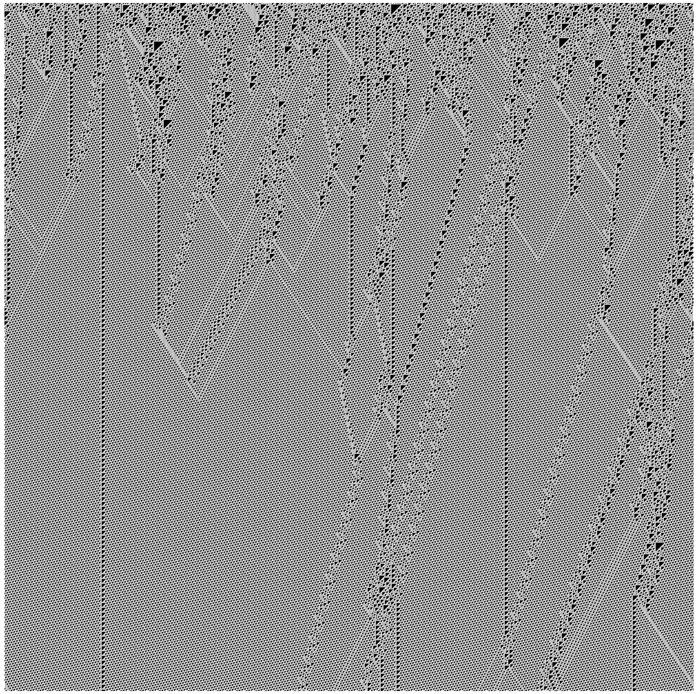
A single cell state at the centered is changed, and the difference pattern illustrates how the disturbance is spread.



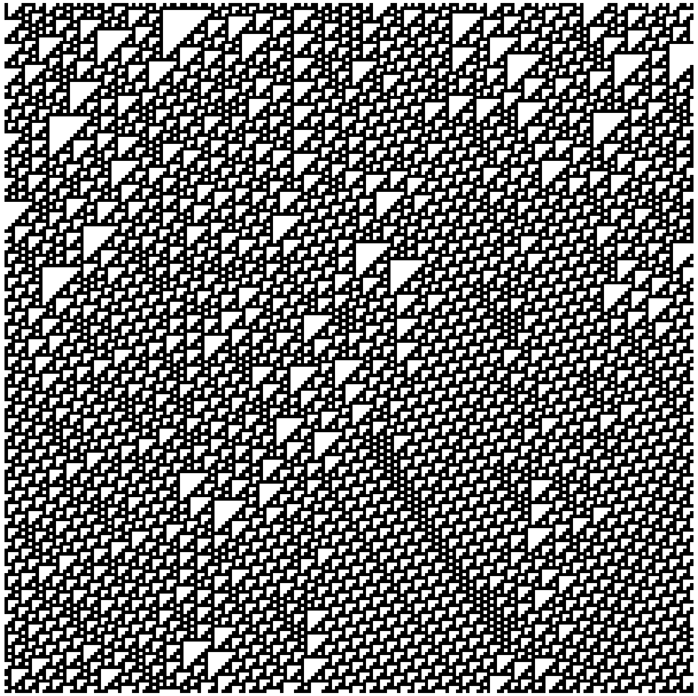
Class IV example: R110

Computationally universal





Information characteristics?



Information in a symbol sequence

. . . 0 1 0 0 1 0 1 1 0 1 ?

Basic information $I_0 = \log n = \log 2 = 1$ (bit)

Statistics of the sequence give probabilities...

With probability p of the event, this is generalized to

$$I = \log \frac{1}{p}$$

Information in a symbol sequence

. . . 0 1 0 0 1 0 1 1 0 1 ?

Probability of x_m given x_1, x_2, ..., x_{m-1}

$$p(x_m | x_1 \ldots x_{m-1}) = \frac{p(x_1 \ldots x_{m-1} x_m)}{p(x_1 \ldots x_{m-1})}$$

Information gained when observing a symbol — local information

$$I = \log \frac{1}{p} = \log \frac{1}{p(x_m | x_1 \ldots x_{m-1})}$$

Symmetric local information

. . . 0 1 0 0 1 0 1 1 0 1 ? 1 0 1 1 1 0 0 0 0 1 0 . . .

Use probabilities that depend on either m–1 symbols to the left or to the right, p_{L} or p_{R},

$$p_L(x_k|x_{k-m+1}...x_{k-1}) = \frac{p(x_{k-m+1}...x_{k-1}x_k)}{p(x_{k-m+1}...x_{k-1})}$$

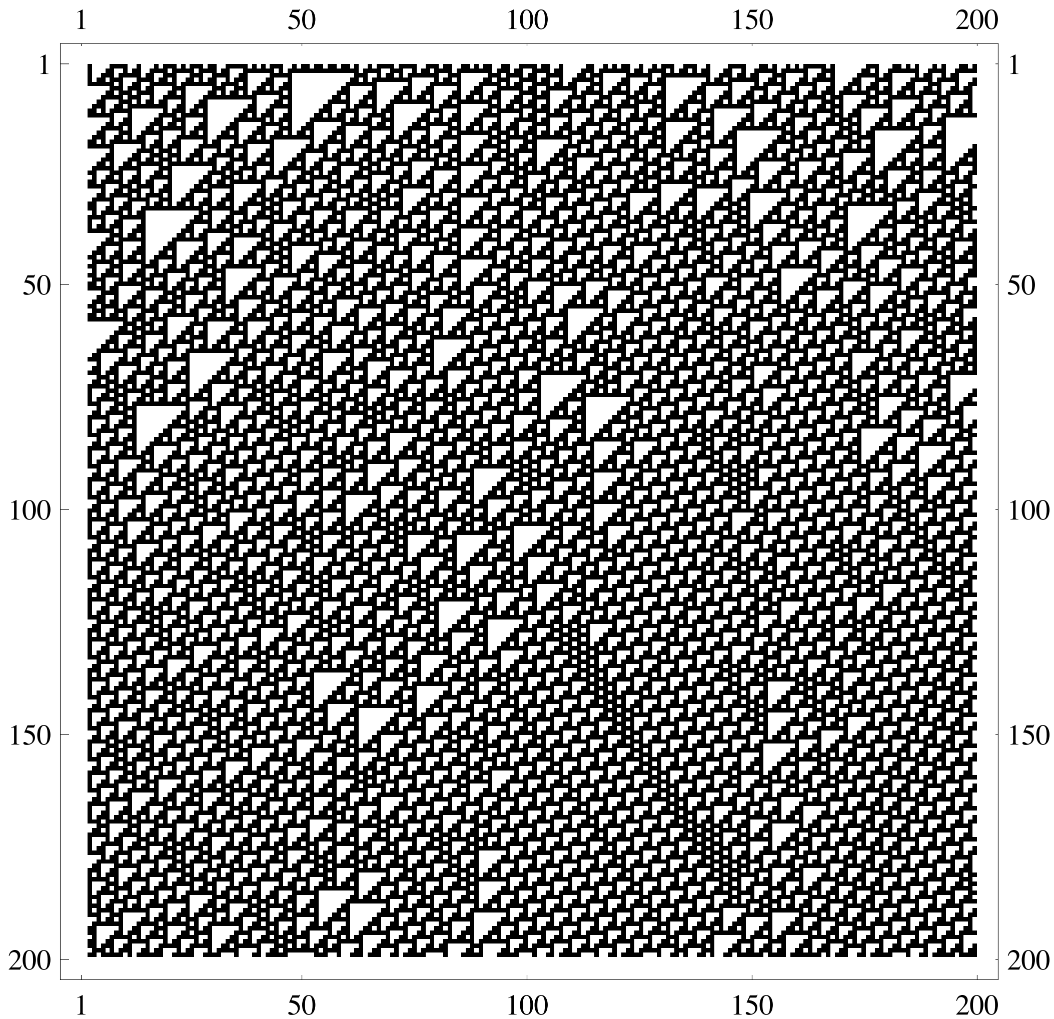
$$p_R(x_k|x_{k+1}...x_{k+m-1}) = \frac{p(x_k x_{k+1}...x_{k+m-1})}{p(x_{k+1}...x_{k+m-1})}$$

Local symmetric information combines ”left” and ”right”

$$I = \frac{1}{2}\left(\log\frac{1}{p_L(x_k|x_{k-m+1}...x_{k-1})} + \log\frac{1}{p_R(x_k|x_{k+1}...x_{k+m-1})}\right)$$

Regularity filter

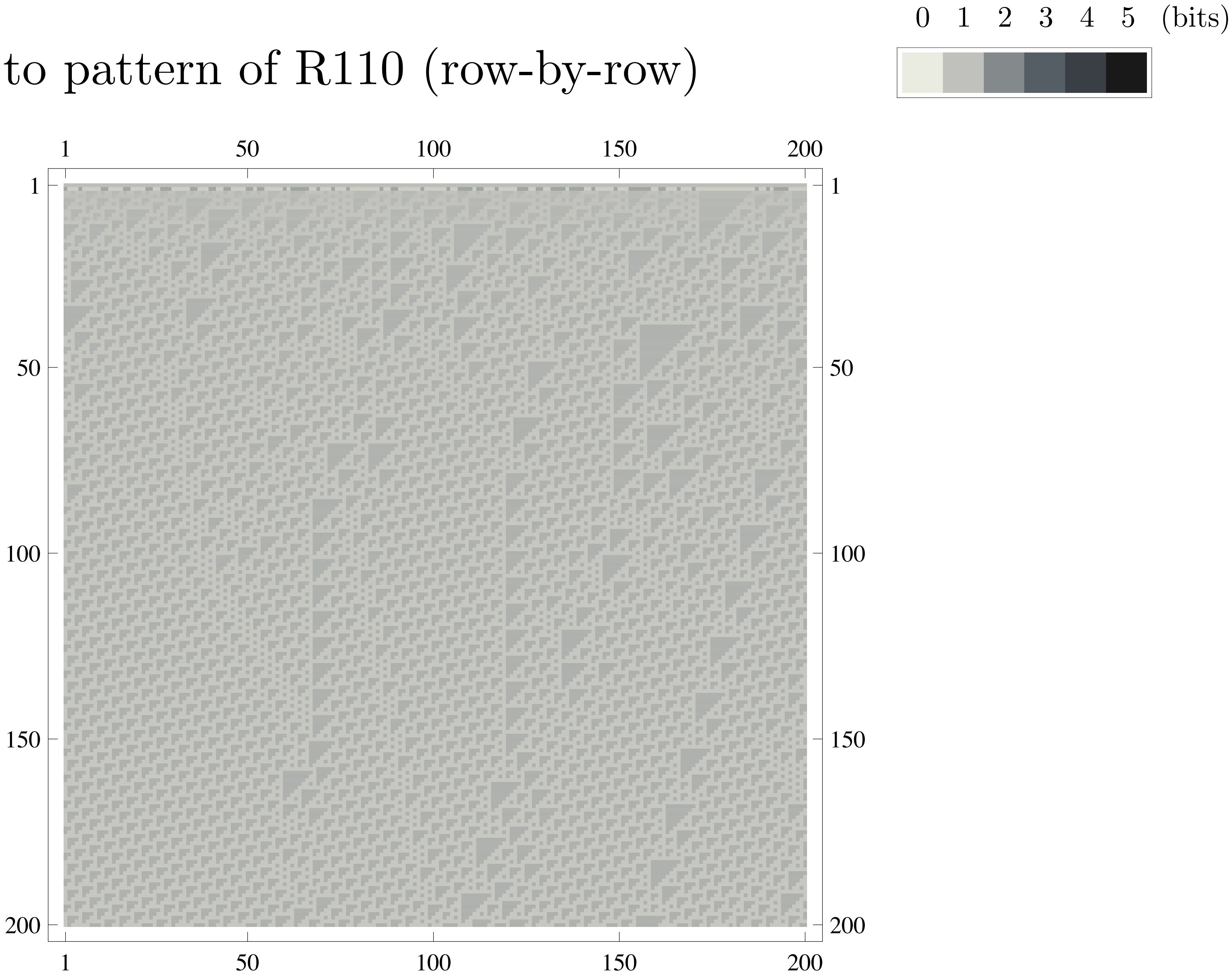
- "Local" information I applied to pattern of R110 (row-by-row)



Regularity filter

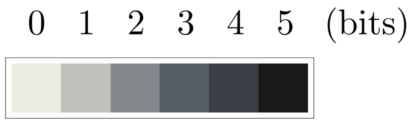
- Applied to pattern of R110 (row-by-row)

$m = 1$

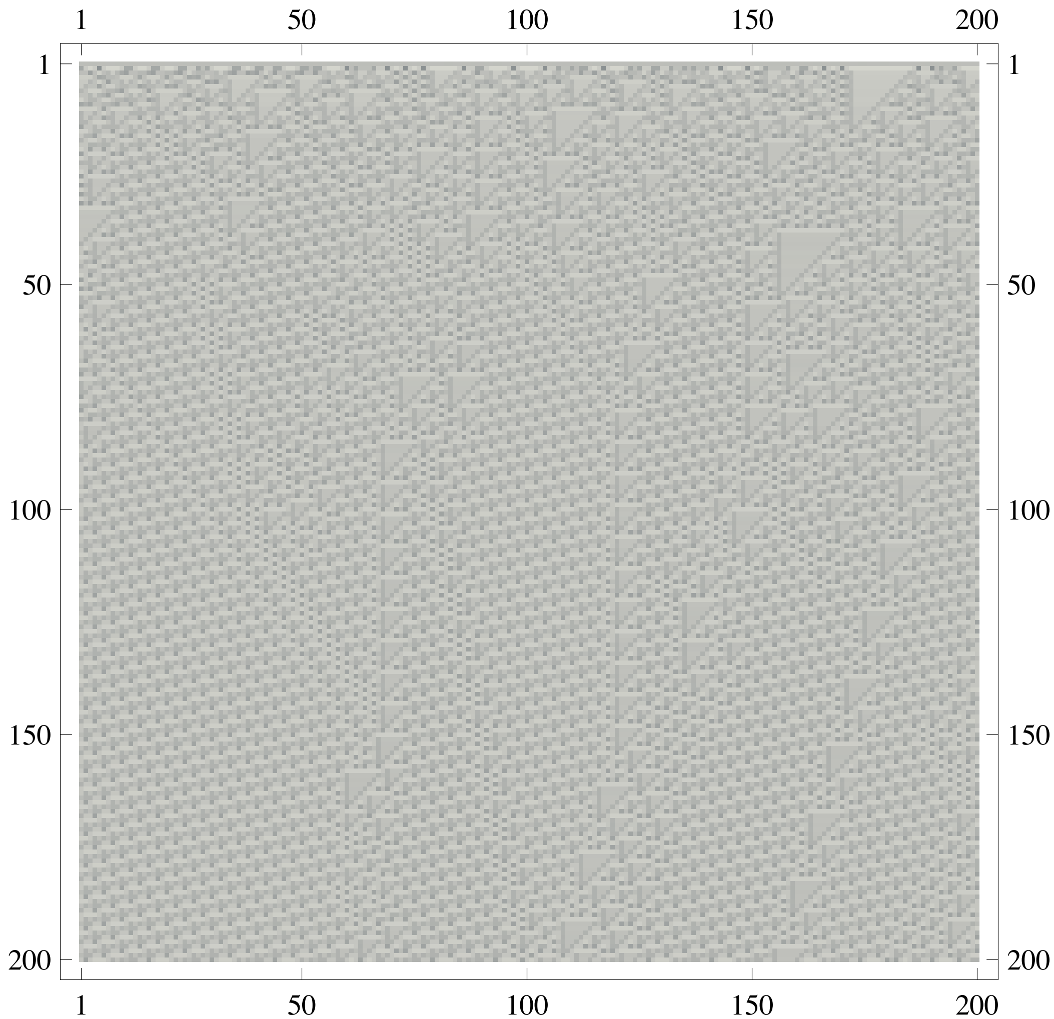


Regularity filter

- Applied to pattern of R110 (row-by-row)

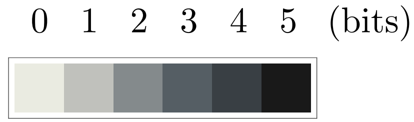


$m = 2$

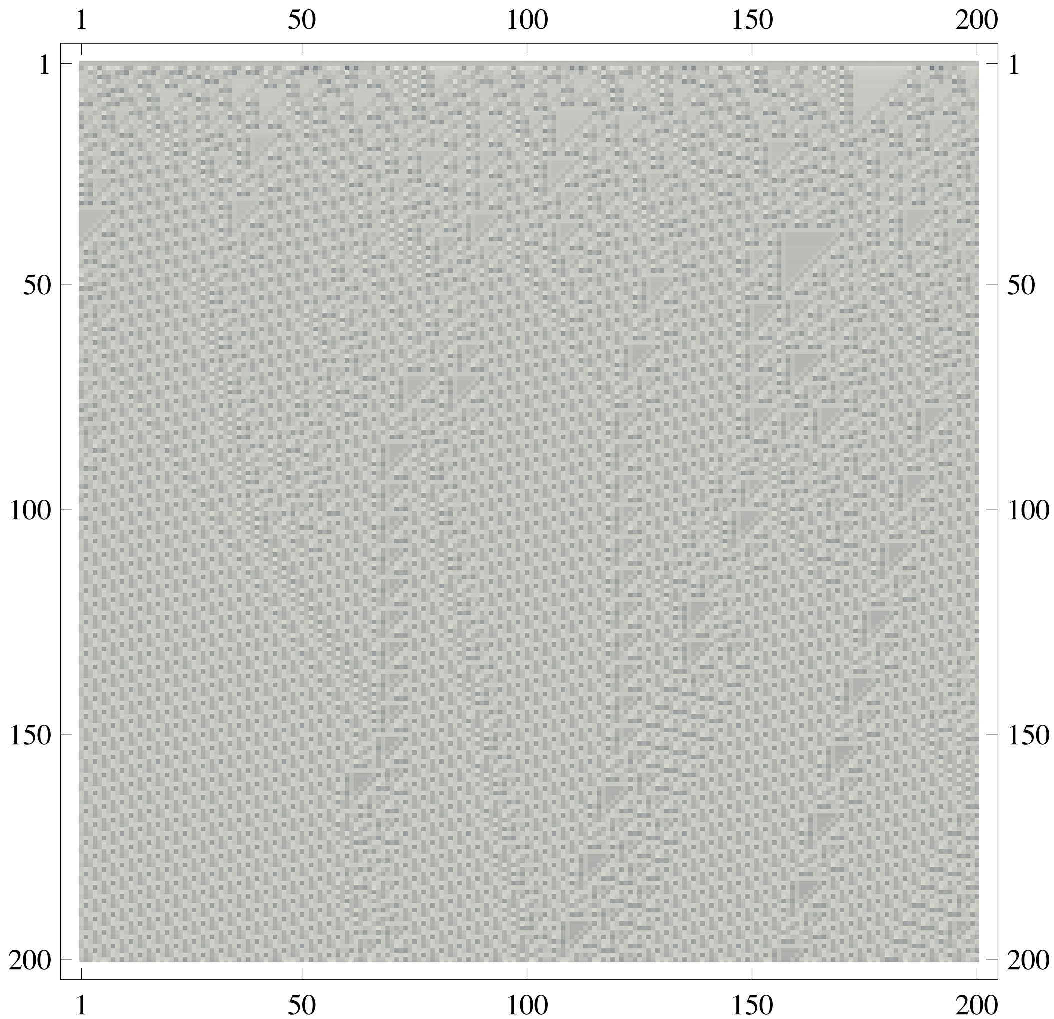


Regularity filter

- Applied to pattern of R110 (row-by-row)

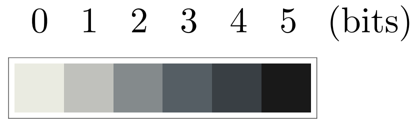


$m = 3$

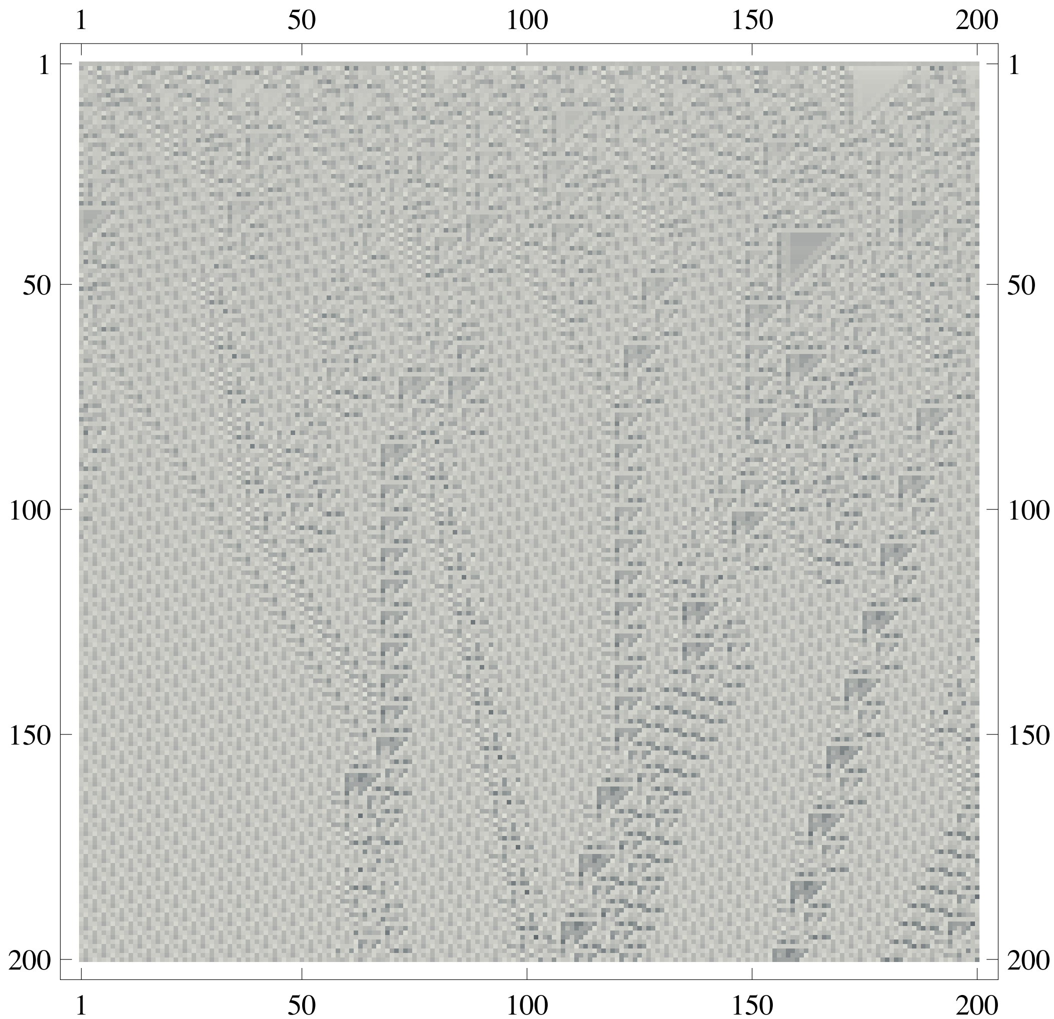


Regularity filter

- Applied to pattern of R110 (row-by-row)



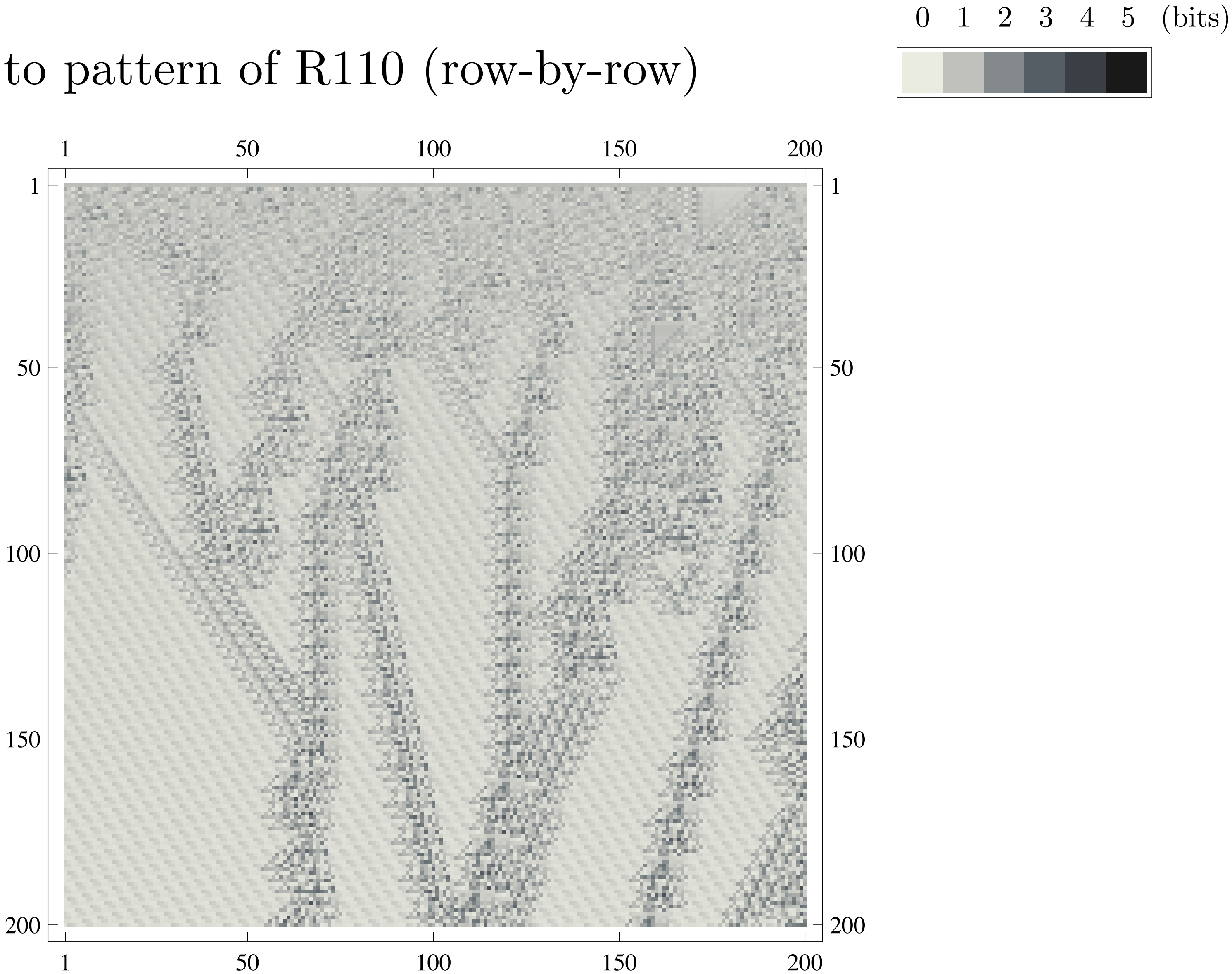
$m = 4$



Regularity filter

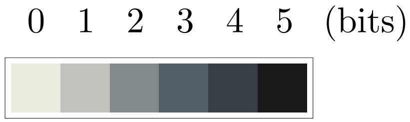
- Applied to pattern of R110 (row-by-row)

$m = 5$

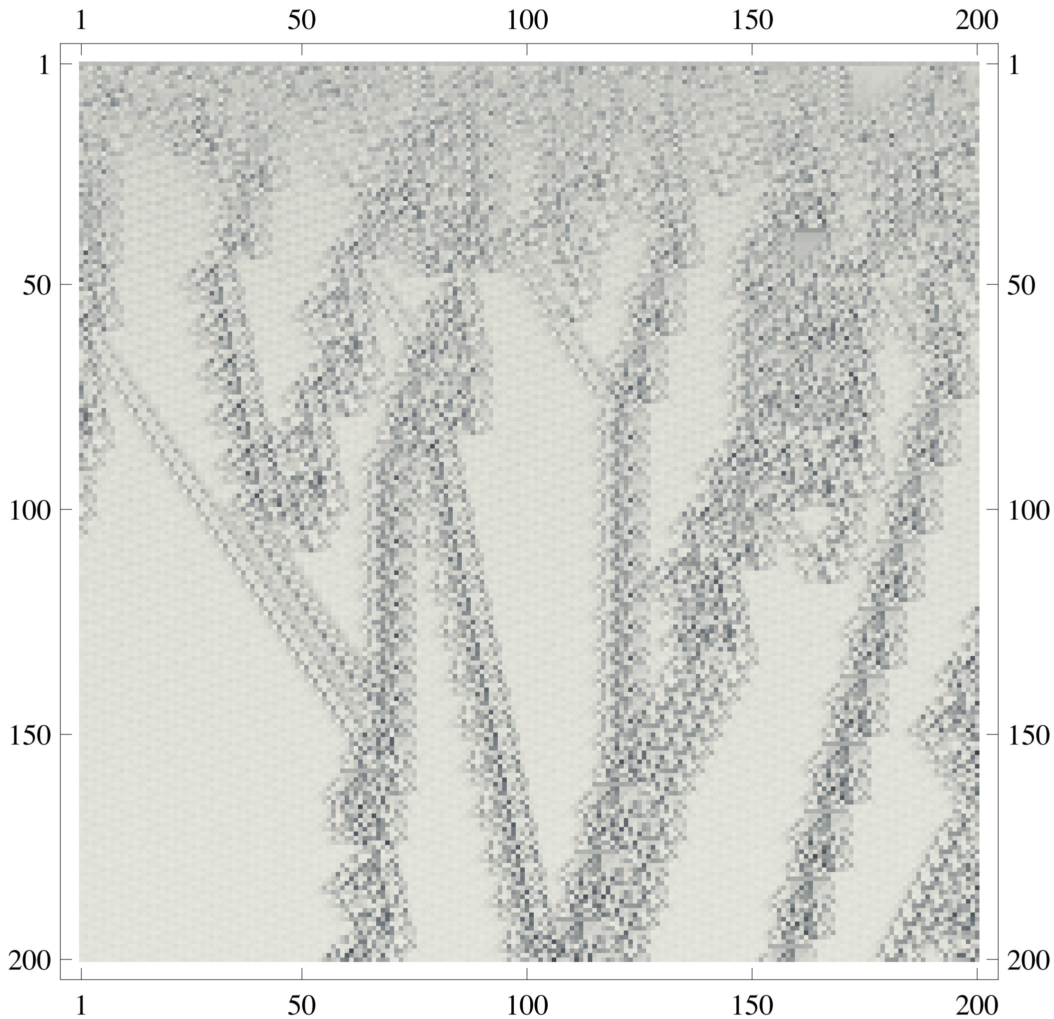


Regularity filter

- Applied to pattern of R110 (row-by-row)



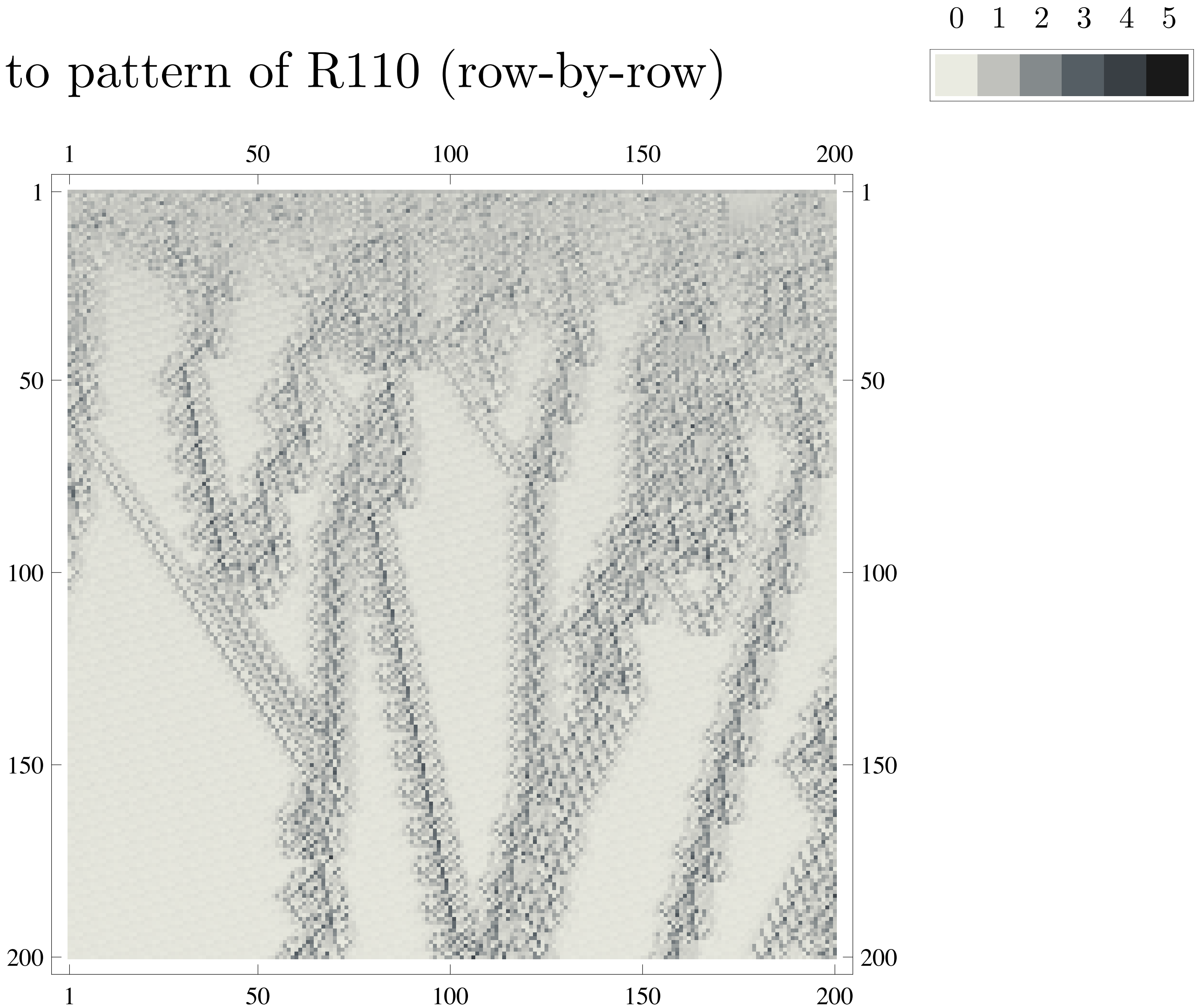
$m = 6$



Regularity filter

- Applied to pattern of R110 (row-by-row)

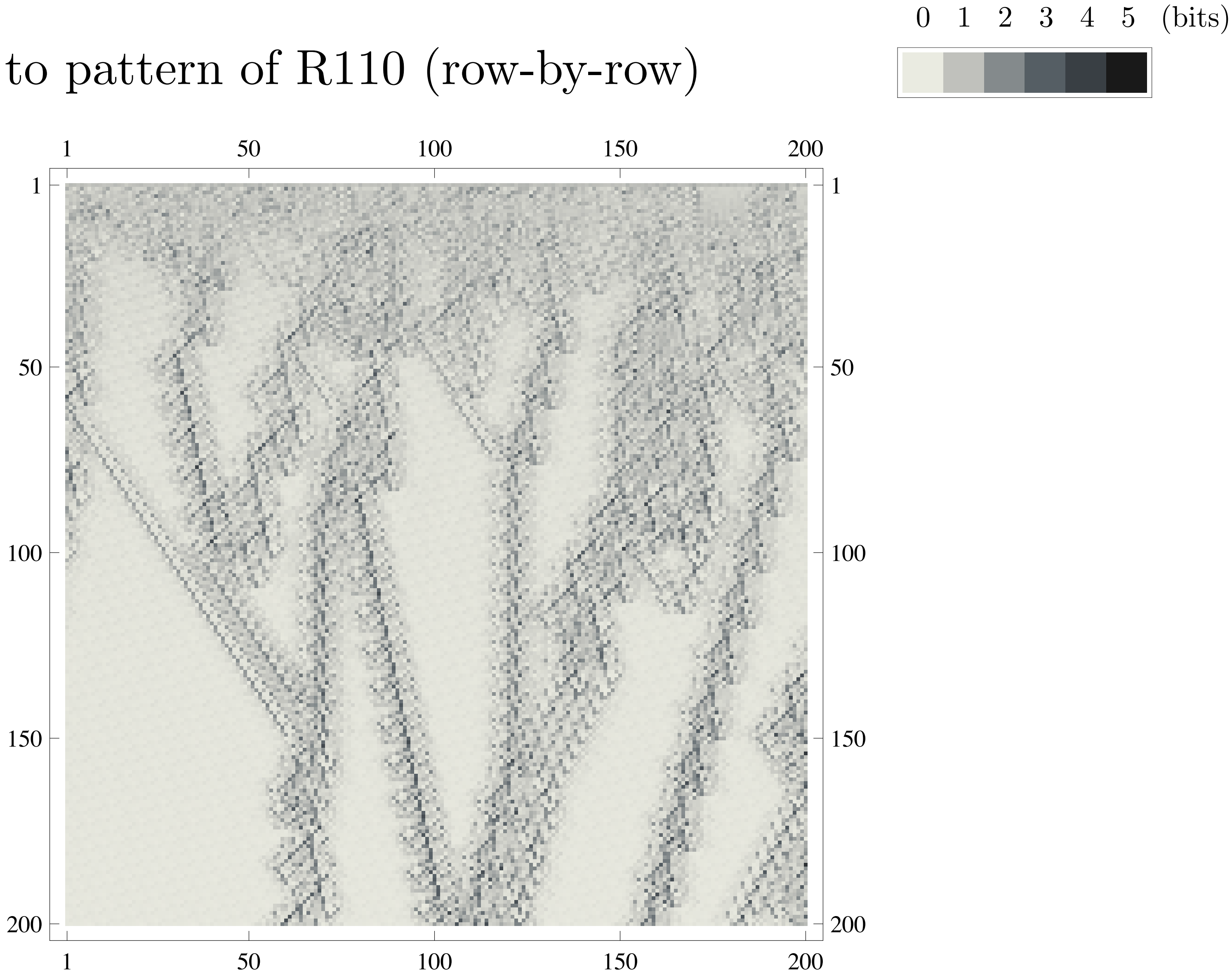
$m = 7$



Regularity filter

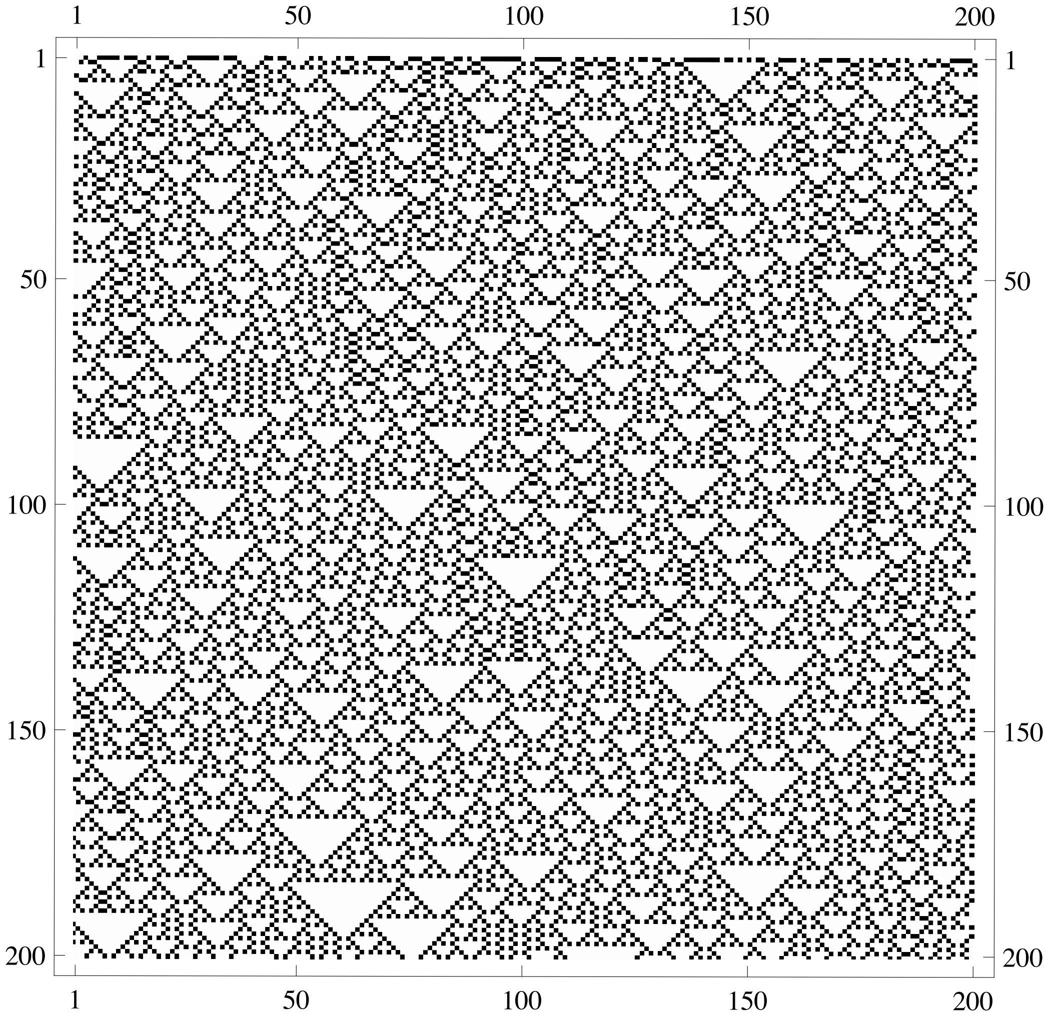
- Applied to pattern of R110 (row-by-row)

$m = 8$



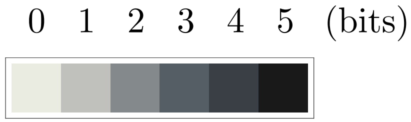
Regularity filter

- Applied to space-time pattern of R18 (row-by-row)

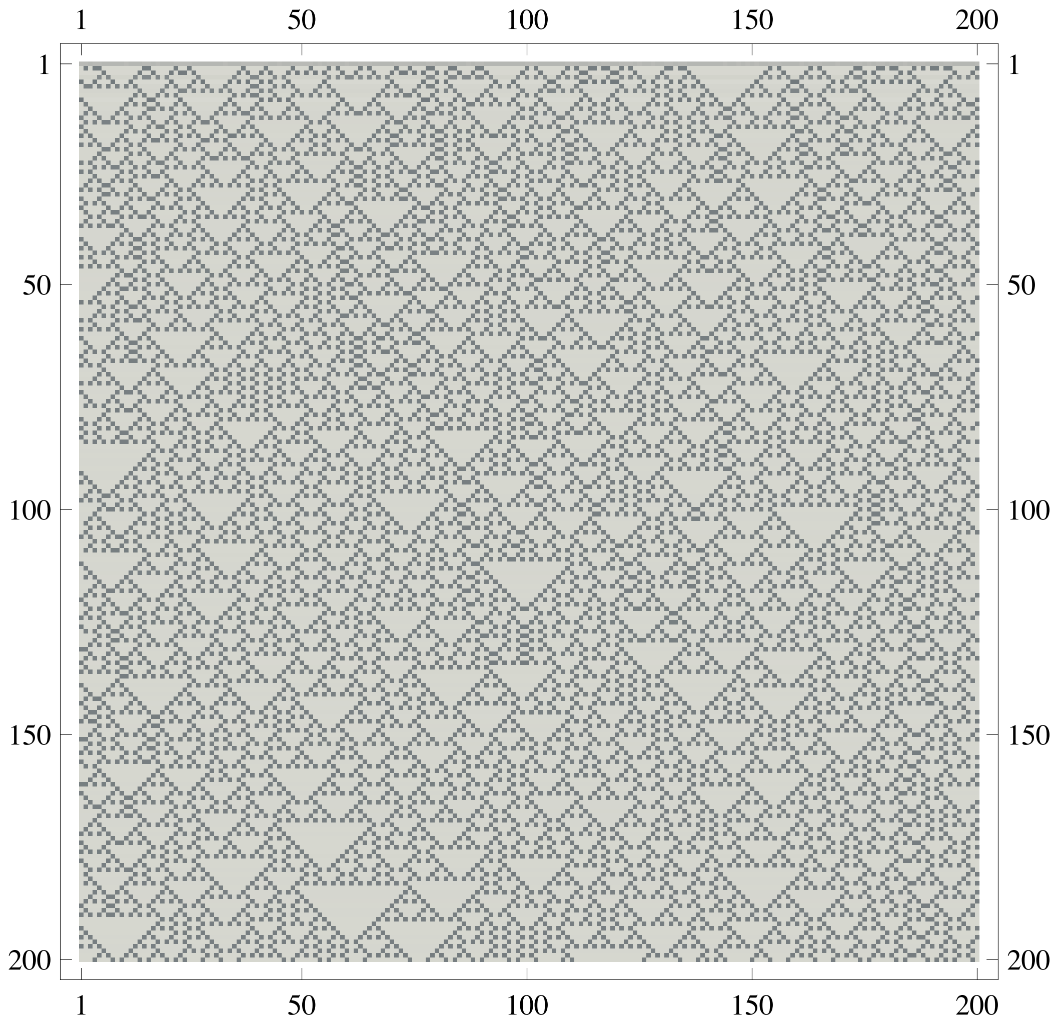


Regularity filter

- Applied to pattern of R18 (row-by-row)



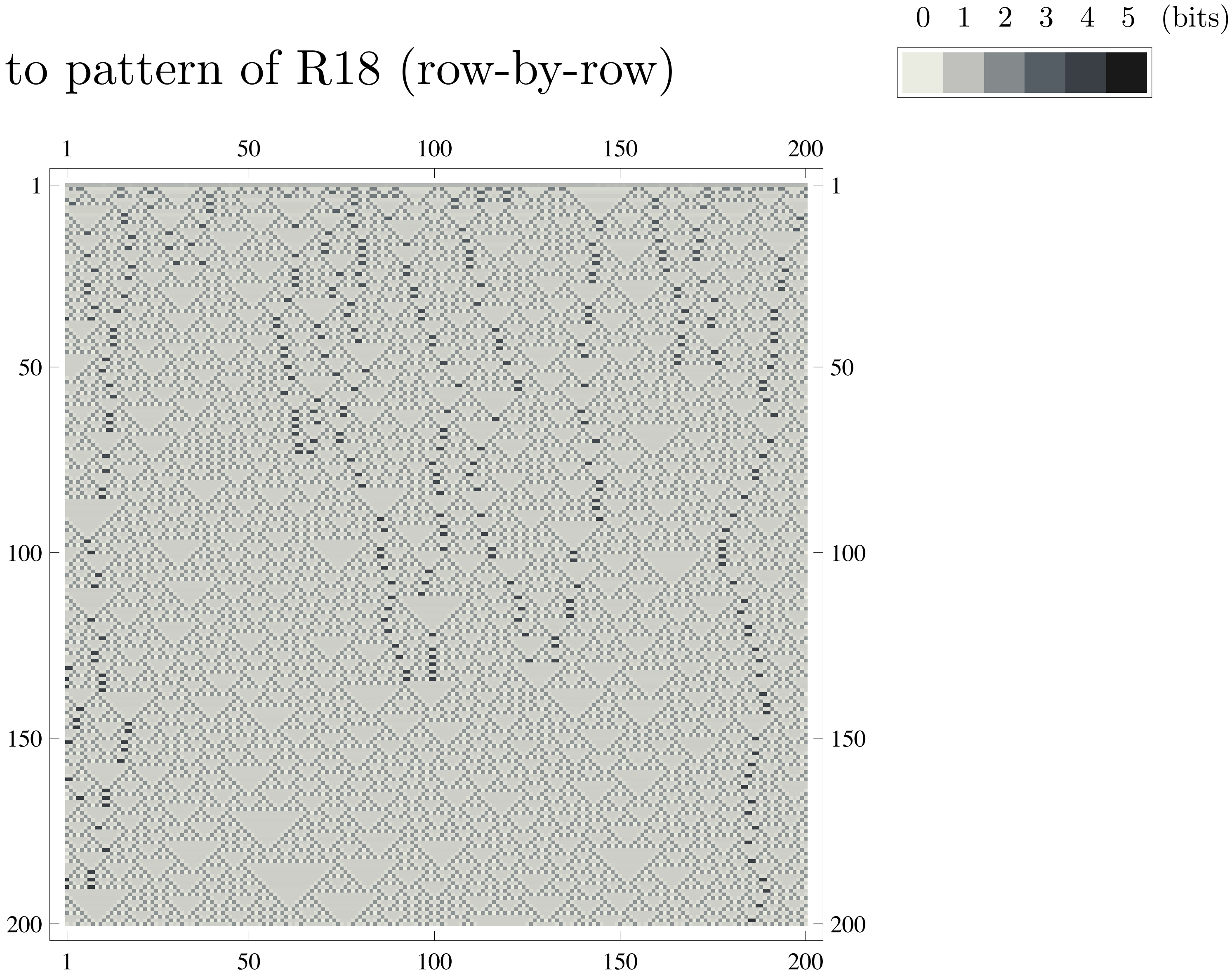
$m = 1$



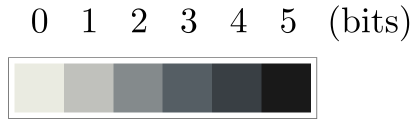
Regularity filter

- Applied to pattern of R18 (row-by-row)

$m = 2$

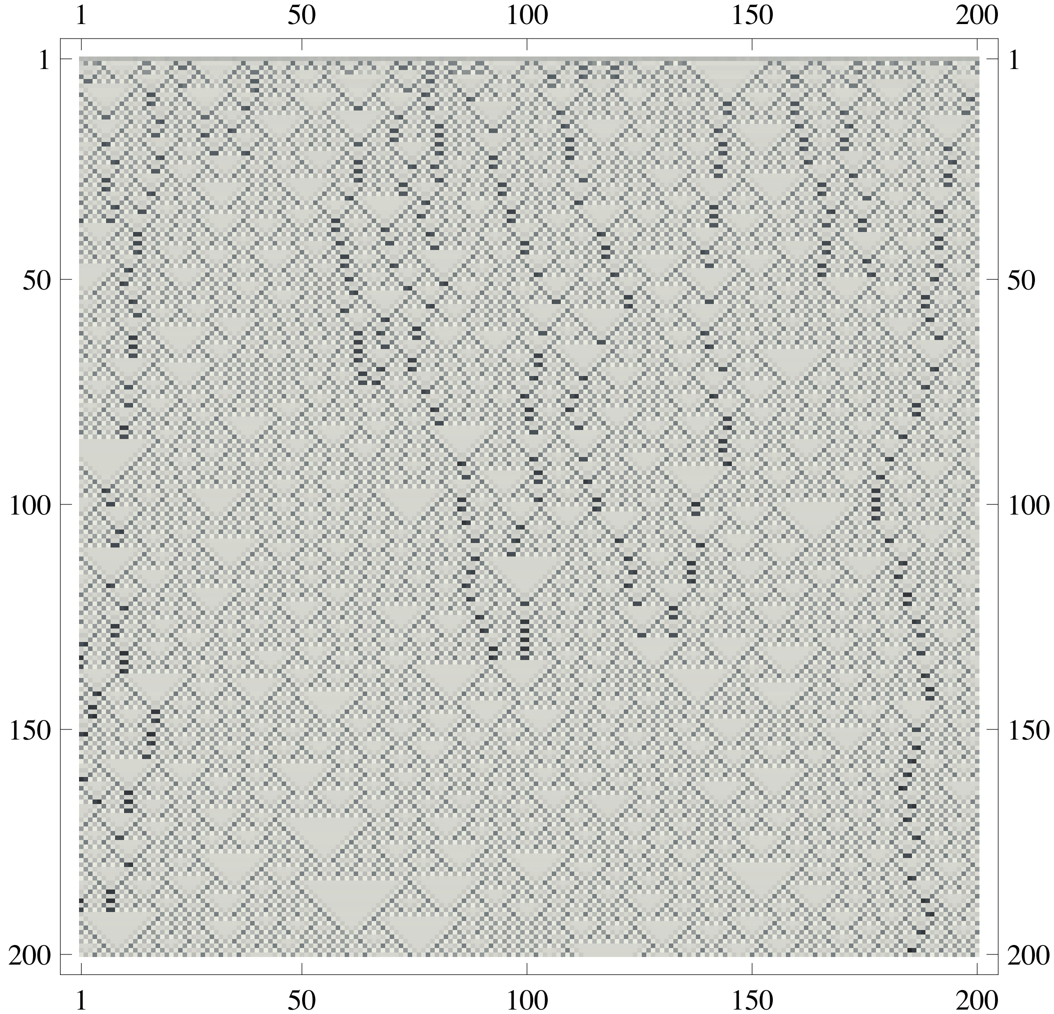


Regularity filter



- Applied to pattern of R18 (row-by-row)

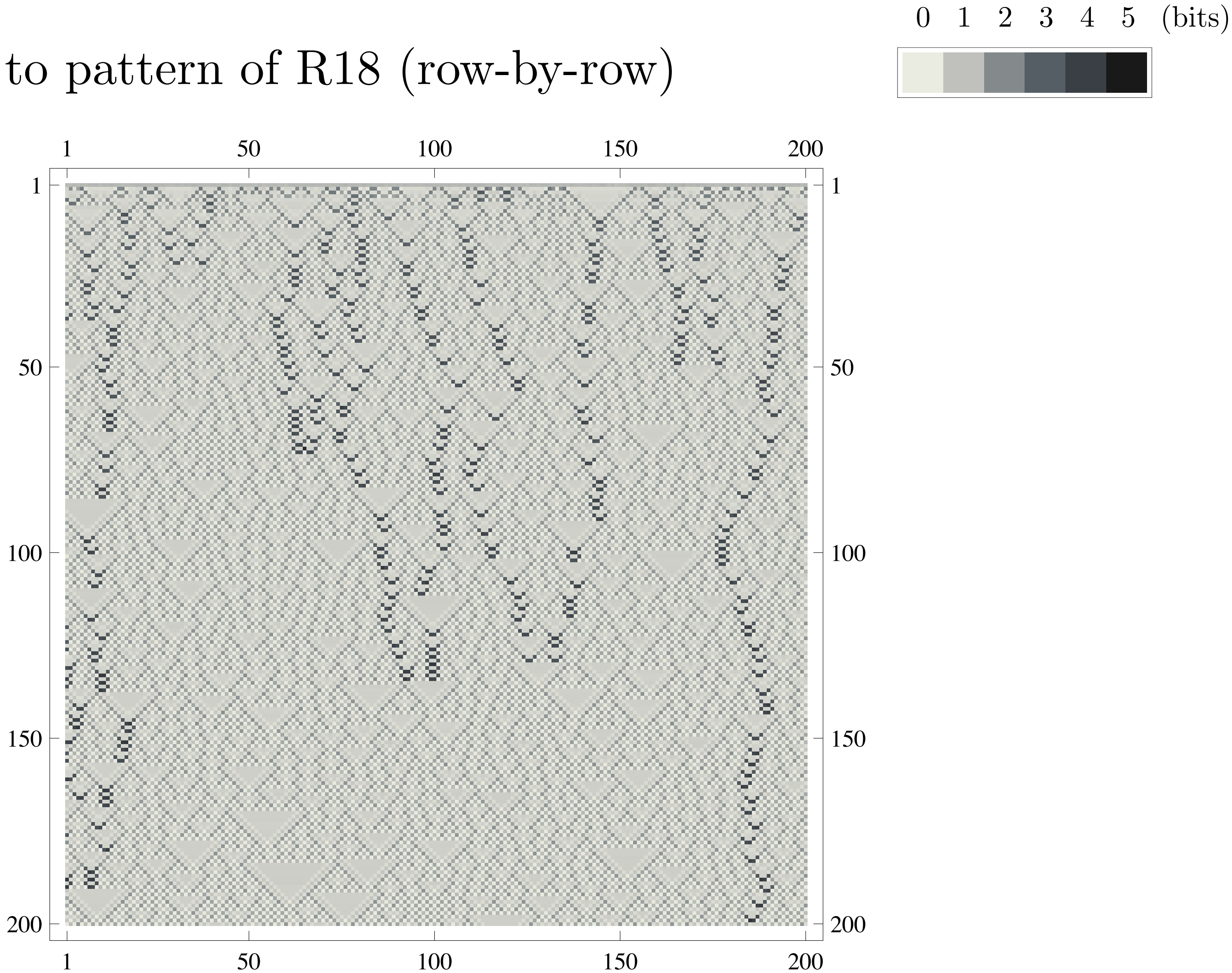
$m = 3$



Regularity filter

- Applied to pattern of R18 (row-by-row)

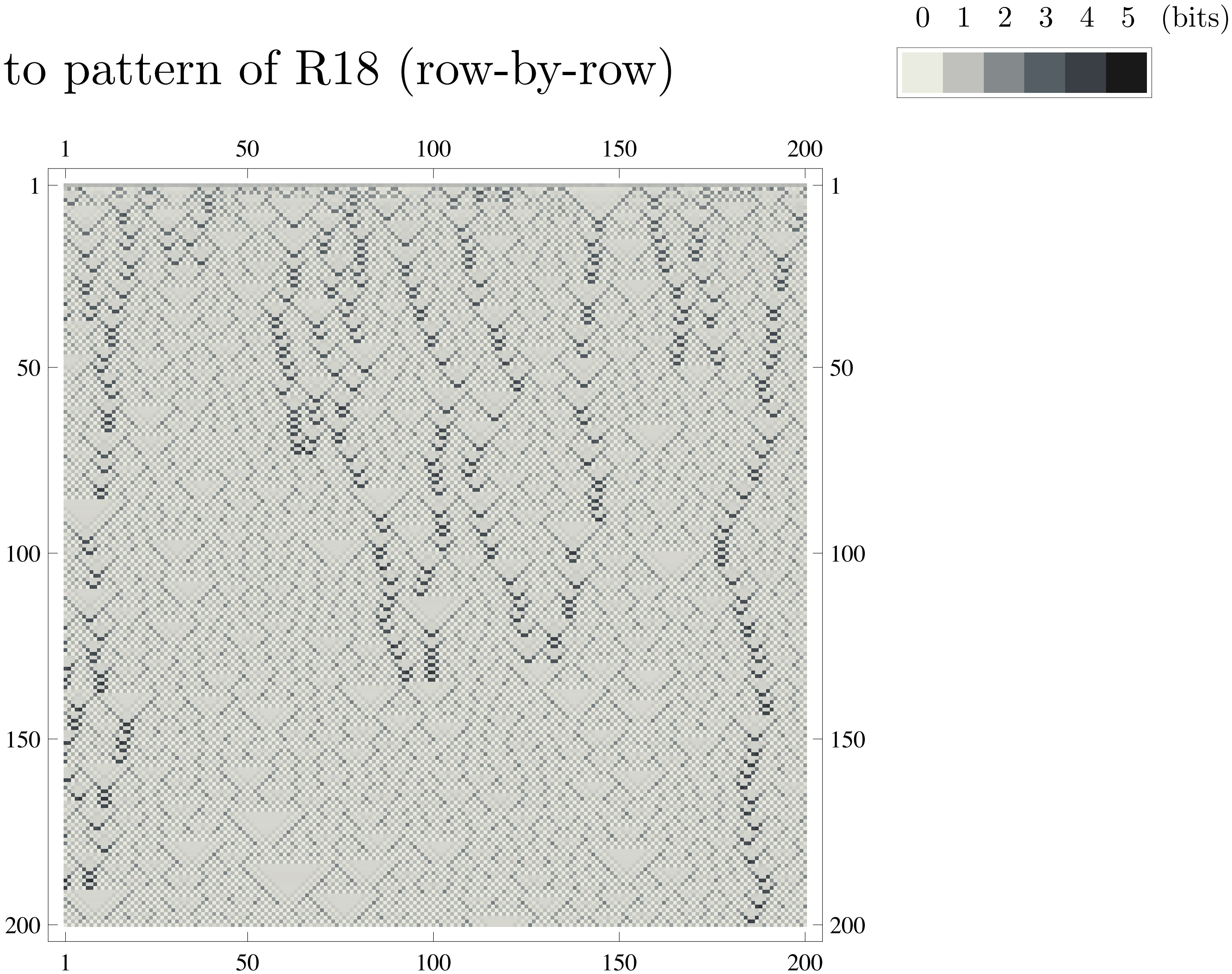
$m = 4$



Regularity filter

- Applied to pattern of R18 (row-by-row)

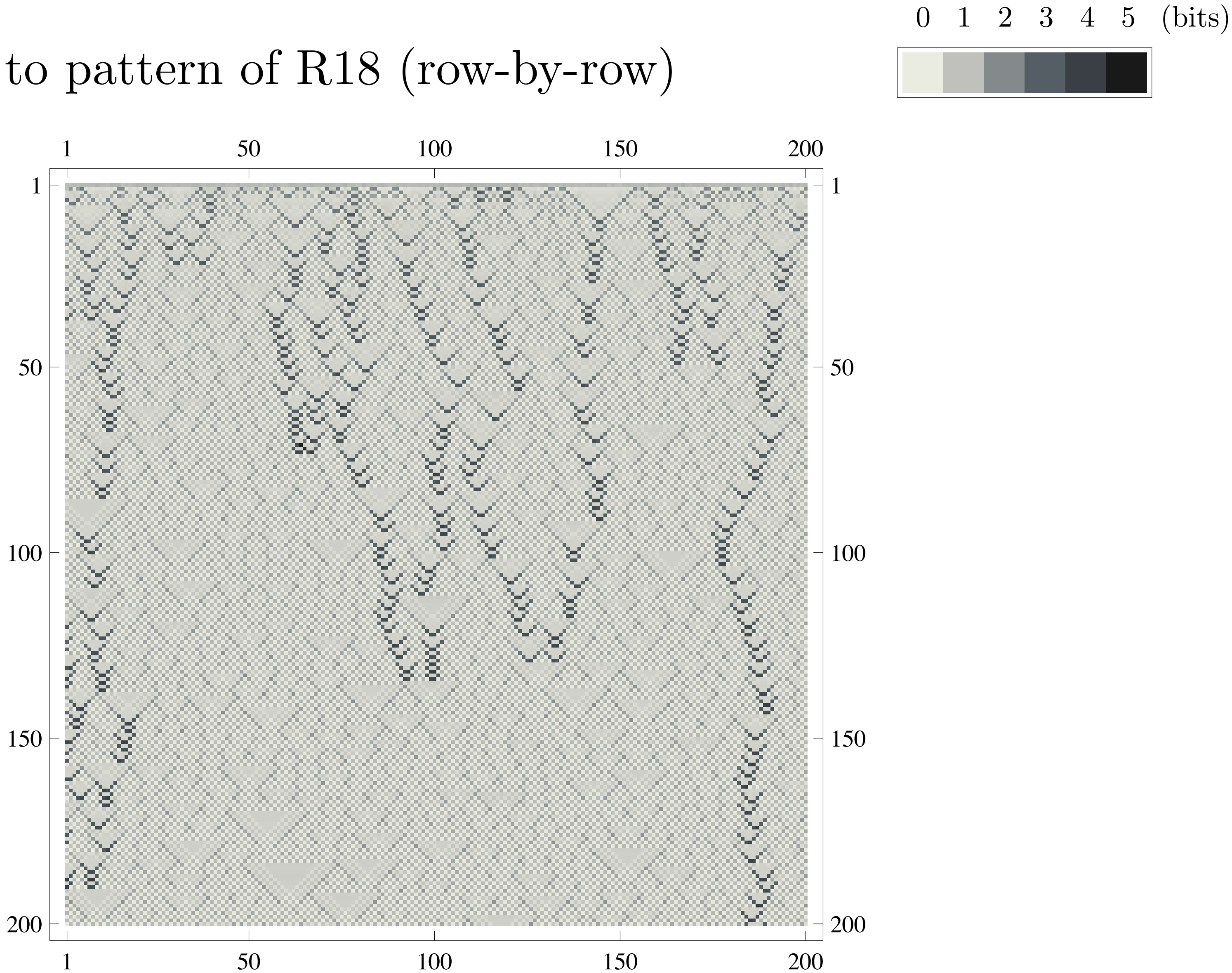
$m = 5$



Regularity filter

- Applied to pattern of R18 (row-by-row)

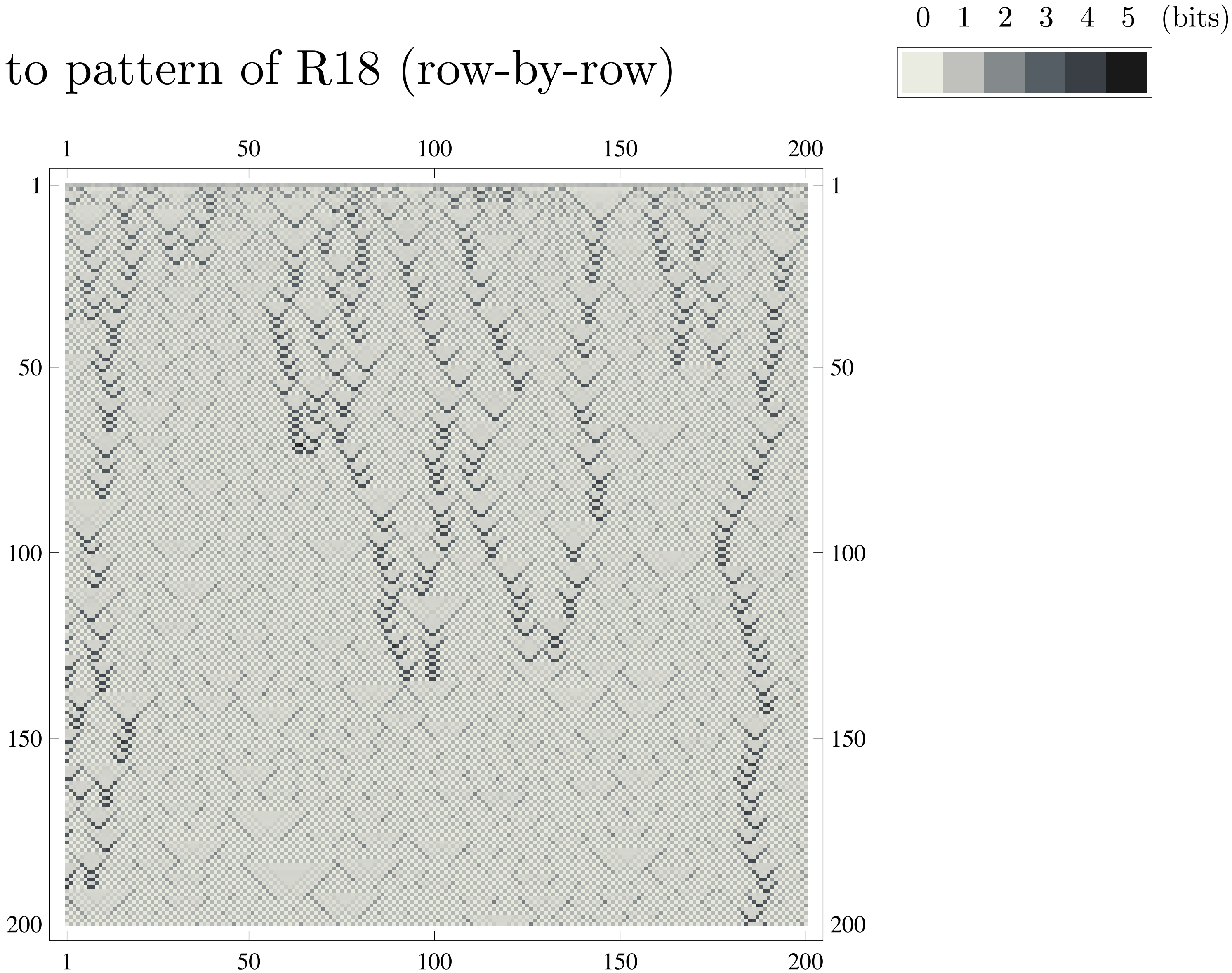
$m = 6$



Regularity filter

- Applied to pattern of R18 (row-by-row)

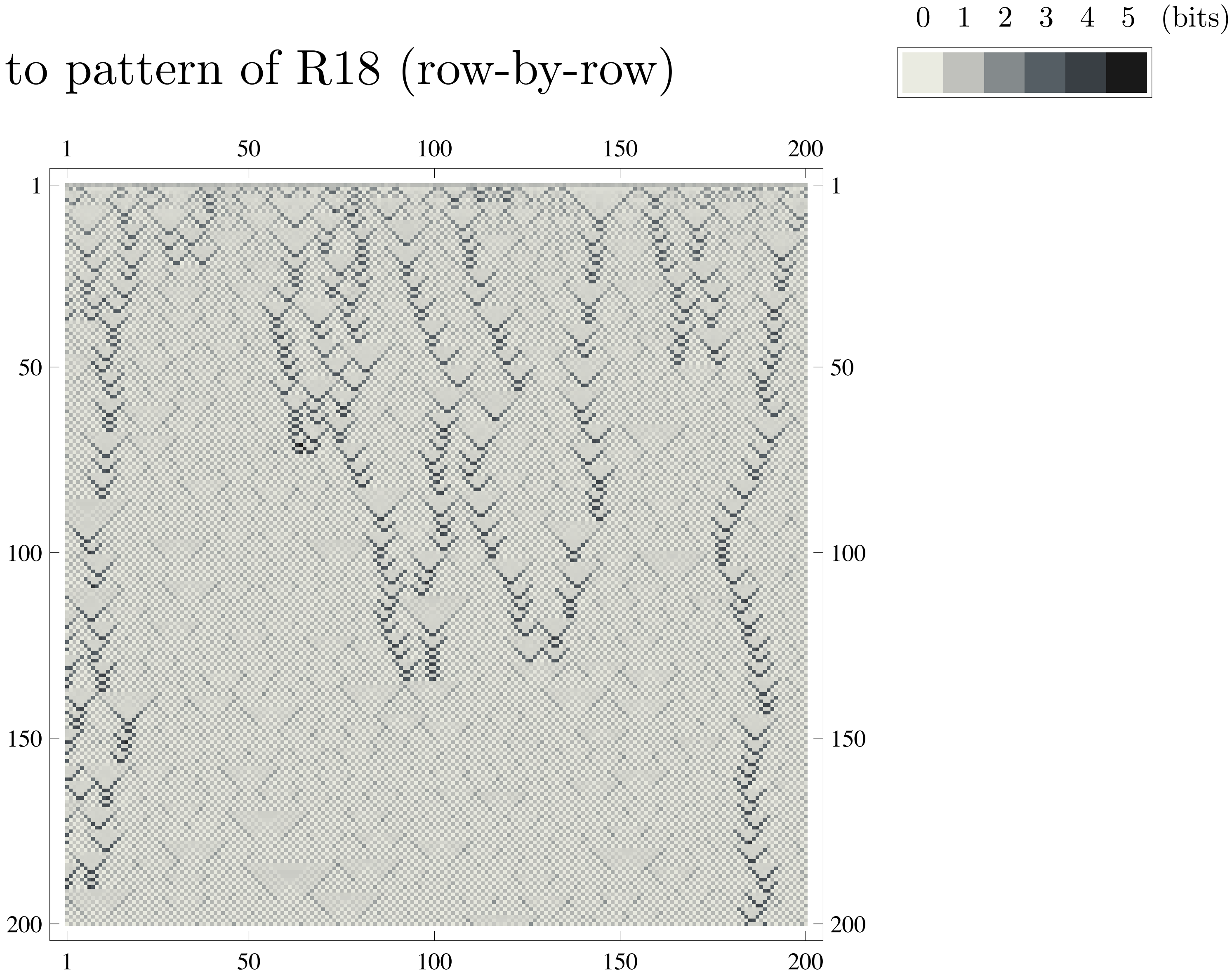
$m = 7$



Regularity filter

- Applied to pattern of R18 (row-by-row)

$m = 8$



Average of the information quantity

- Average of the information quantity (here the "Left" one):

$$h_m = \sum_{x_1 \ldots x_m} p(x_1 \ldots x_m) \log \frac{1}{p(x_m|x_1 \ldots x_{m-1})}$$
$$= \sum_{x_1 \ldots x_{m-1}} p(x_1 \ldots x_{m-1}) \sum_{x_m} p(x_m|x_1 \ldots x_{m-1}) \log \frac{1}{p(x_m|x_1 \ldots x_{m-1})}$$

- Two interpretations...

Entropy

For a probability distribution $P = \{p(k)\}_{k=1,\ldots,n}$

$$S[P] = \sum_{k=1}^{n} p(k) \log \frac{1}{p(k)}$$

quantifies

- the expected gain of information, or

- the lack of information — uncertainty about the state

Entropy of a stochastic process

- The entropy per symbol, s, is the average uncertainty about the next symbol x_m given the previously read ones $x_1 \ldots x_{m-1}$ in the limit of infinite m

$$s = \lim_{m \to \infty} h_m =$$

$$= \lim_{m \to \infty} \sum_{x_1 \ldots x_{m-1}} p(x_1 \ldots x_{m-1}) \sum_{x_m} p(x_m | x_1 \ldots x_{m-1}) \log \frac{1}{p(x_m | x_1 \ldots x_{m-1})}$$

- The entropy s quantifies the degree of "randomness" of the sequence.

Change of entropy in CA time evolution

- How does the entropy s change from one time step to the next in a CA?

  In general, for deterministic rules, as entropy characterizes "randomness", entropy cannot increase,

$$\Delta_t s(t) \leq 0$$

Relative information

- How does the information about the next symbol change when we extend the number of preceding symbols step-by-step?

$$\left\langle \log \frac{1}{p(x_m|x_2 \ldots x_{m-1})} - \log \frac{1}{p(x_m|x_1 x_2 \ldots x_{m-1})} \right\rangle$$

- Correlation information

$$k_m = \sum_{x_1 \ldots x_{m-1}} p(x_1 \ldots x_{m-1}) \sum_{x_m} p(x_m|x_1 x_2 \ldots x_{m-1}) \log \frac{p(x_m|x_1 x_2 \ldots x_{m-1})}{p(x_m|x_2 \ldots x_{m-1})}$$

Relative information

- Relative information or Kullback-Liebler information — quantifies how much information is gained when one distribution $P_0 = \{p_0(k)\}$ is replaced by a new one $P = \{p(k)\}$,

$$K[P_0; P] = \sum_k p(k) \log \frac{p(k)}{p_0(k)} \geq 0$$

Density information

- Without looking at preceding symbols, how does the information about the next symbol change when we learn the frequencies?

$$\left\langle \log 2 - \log \frac{1}{p(x)} \right\rangle$$

- Density information

$$P_0 = \{1/2, 1/2\},\ P = \{p(0), p(1)\}$$

$$k_1 = K[P_0; P] = \sum_x p(x) \log \frac{p(x)}{1/2}$$

Decomposition of information

- The total information of 1 bit per cell can be decomposed into the entropy s and the redundant information k_{corr},

$$1 = s + k_{\text{corr}}$$

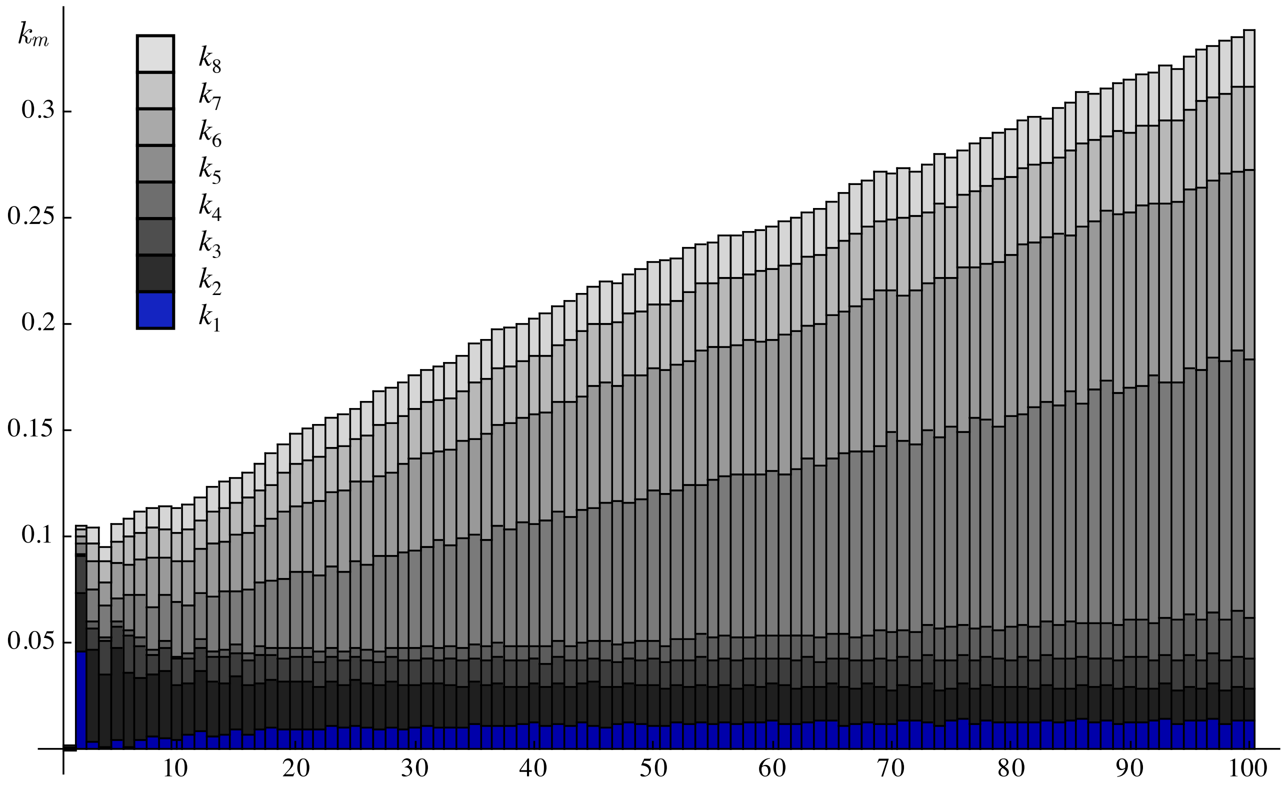
- and the redundant information further into density information k_1 and correlation information k_m (m=2, 3, ...)

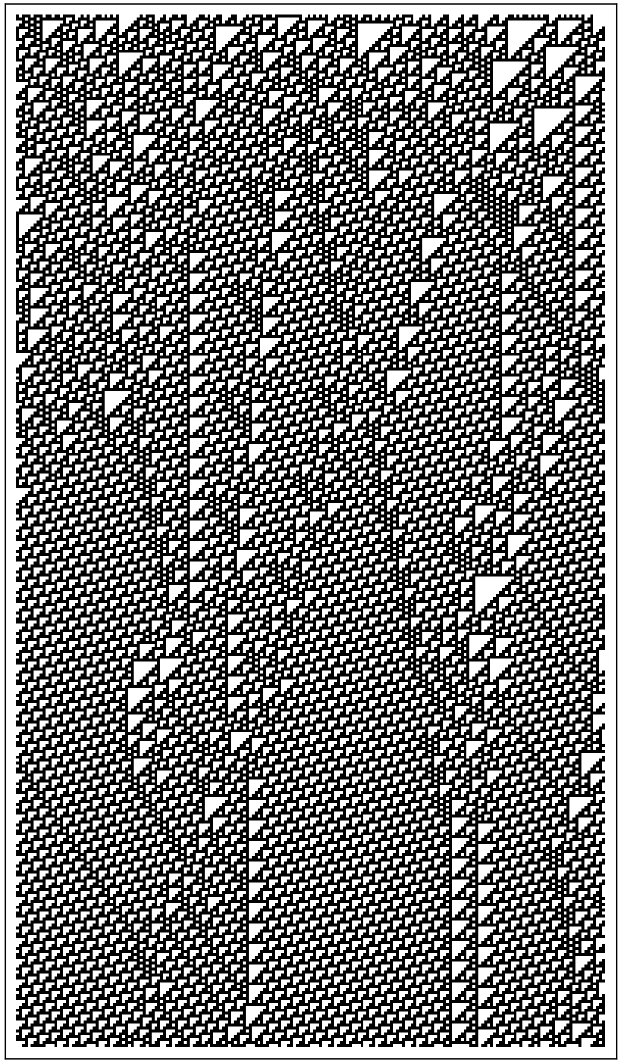
$$k_{\text{corr}} = \sum_{k=1}^{\infty} k_m$$

Information characteristics of CA time evolution

- Example: rule R110

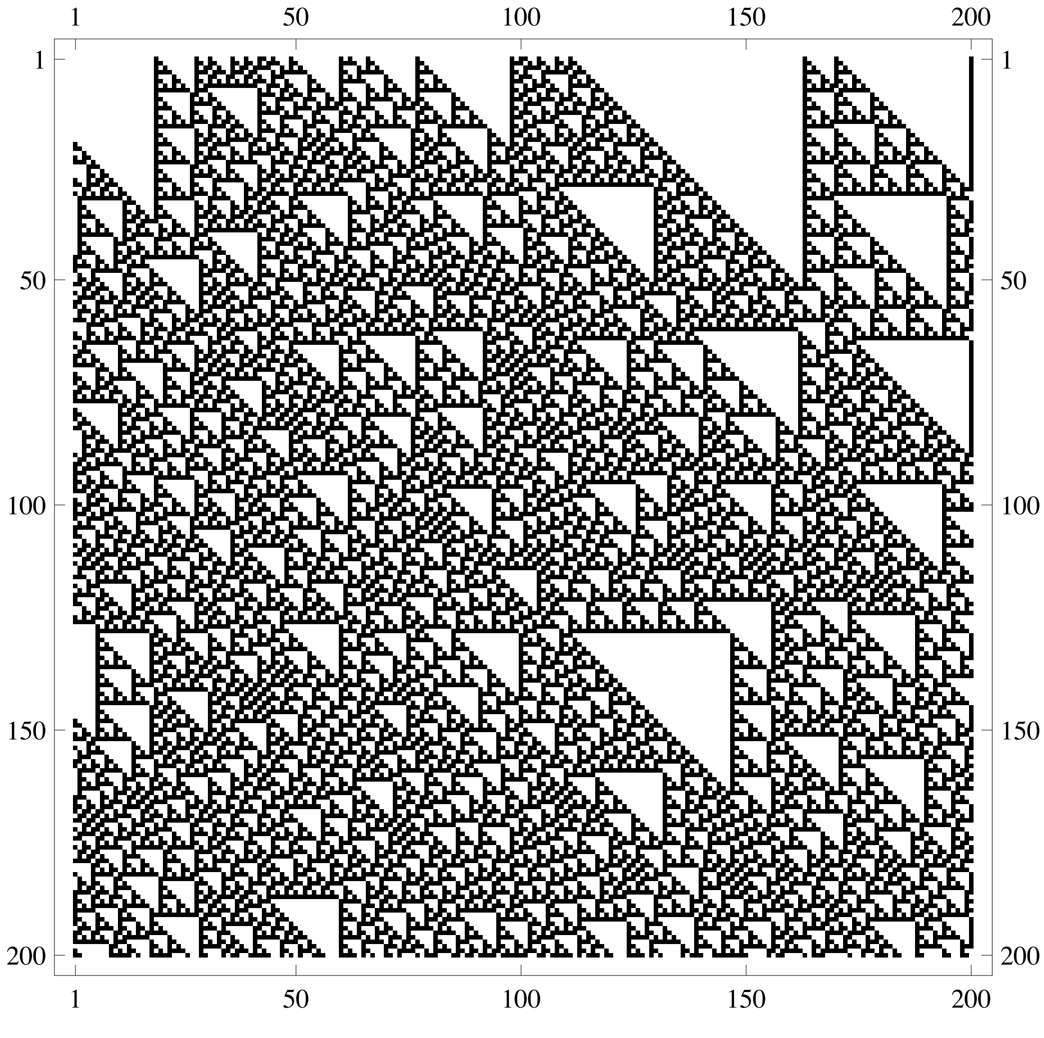
Correlation information





Regularity filter

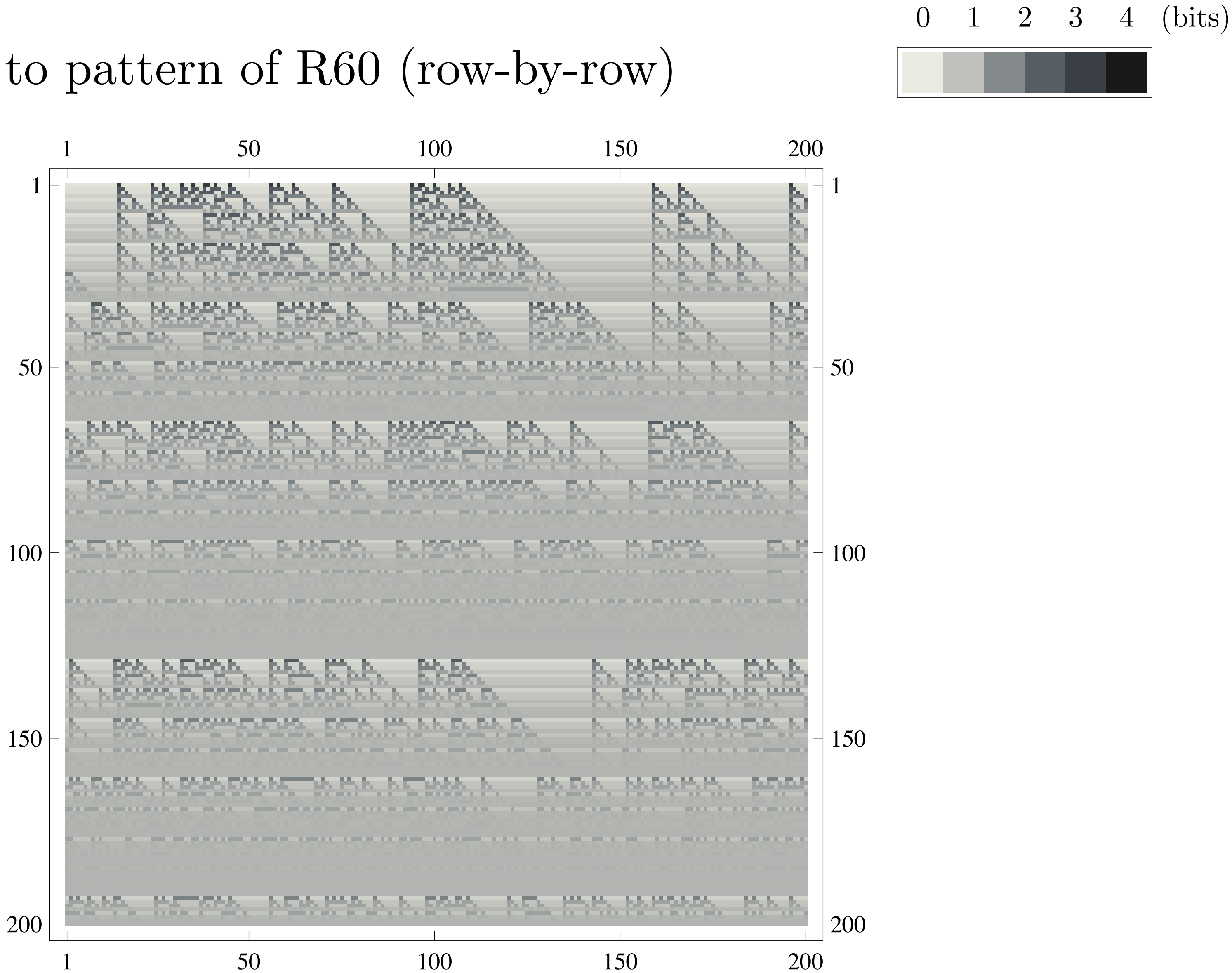
- "Local" information I applied to pattern of R60 (row-by-row)



Regularity filter

- Applied to pattern of R60 (row-by-row)

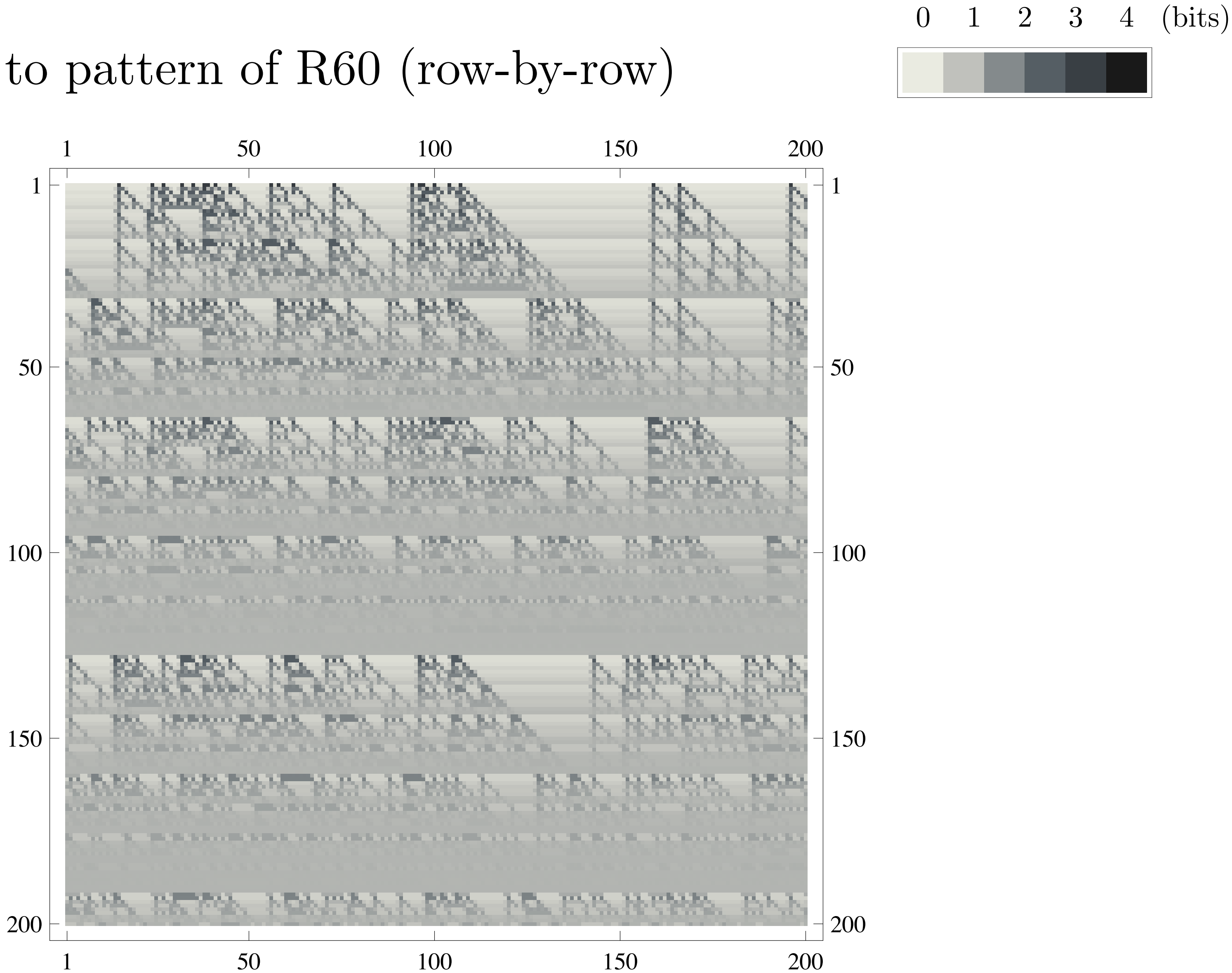
$m = 1$



Regularity filter

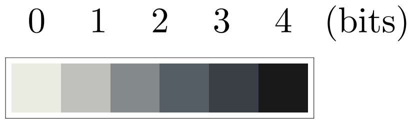
- Applied to pattern of R60 (row-by-row)

$m = 2$

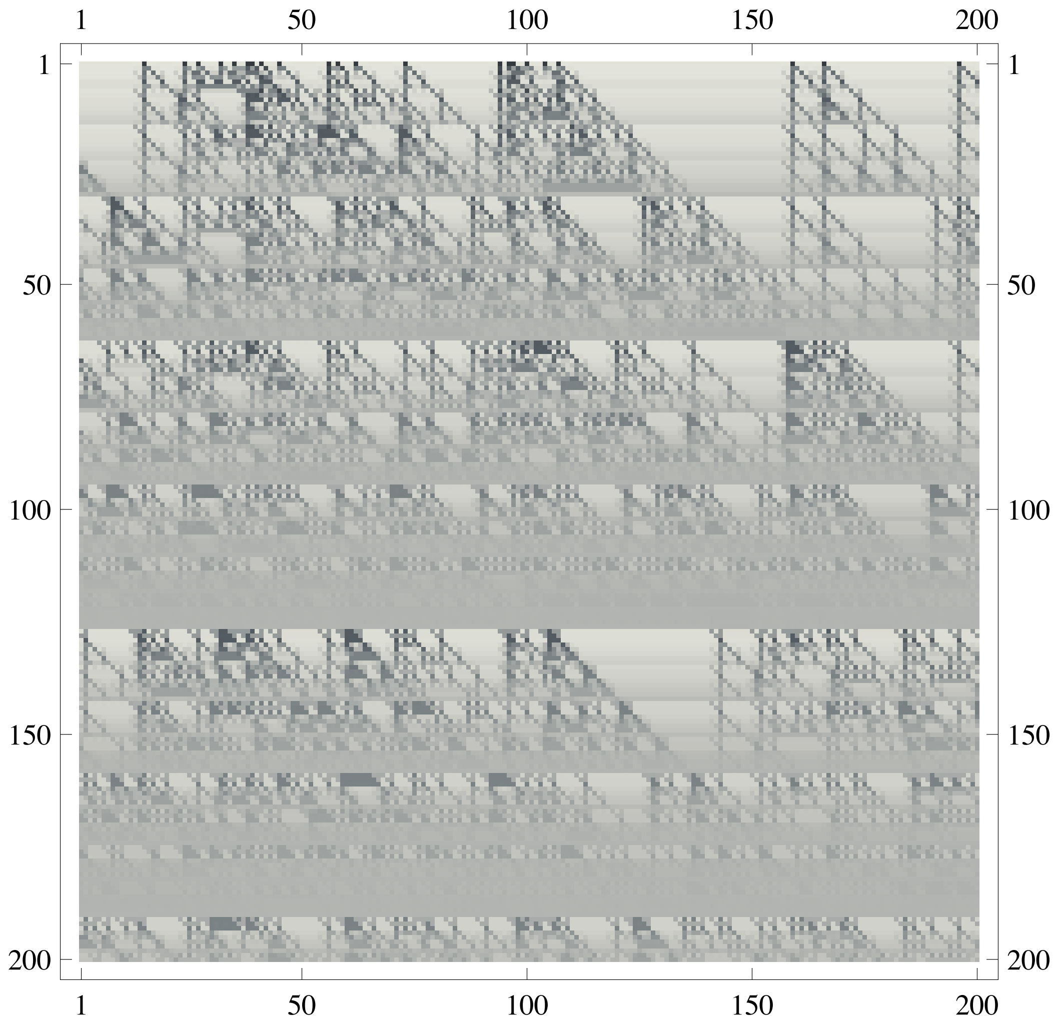


Regularity filter

- Applied to pattern of R60 (row-by-row)

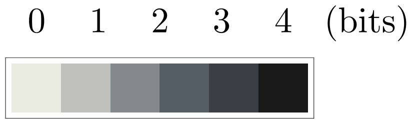


$m = 3$

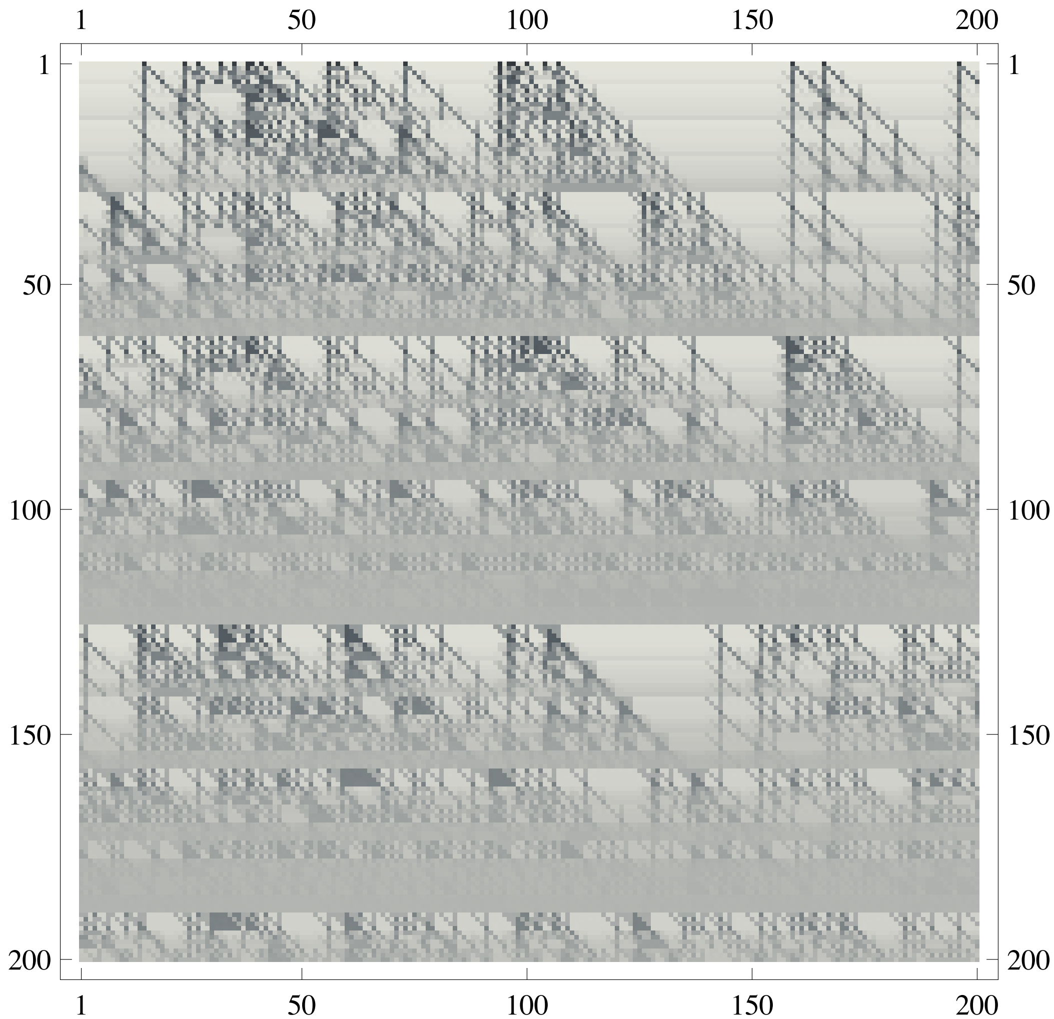


Regularity filter

- Applied to pattern of R60 (row-by-row)

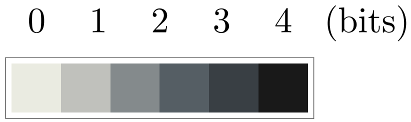


$m = 4$

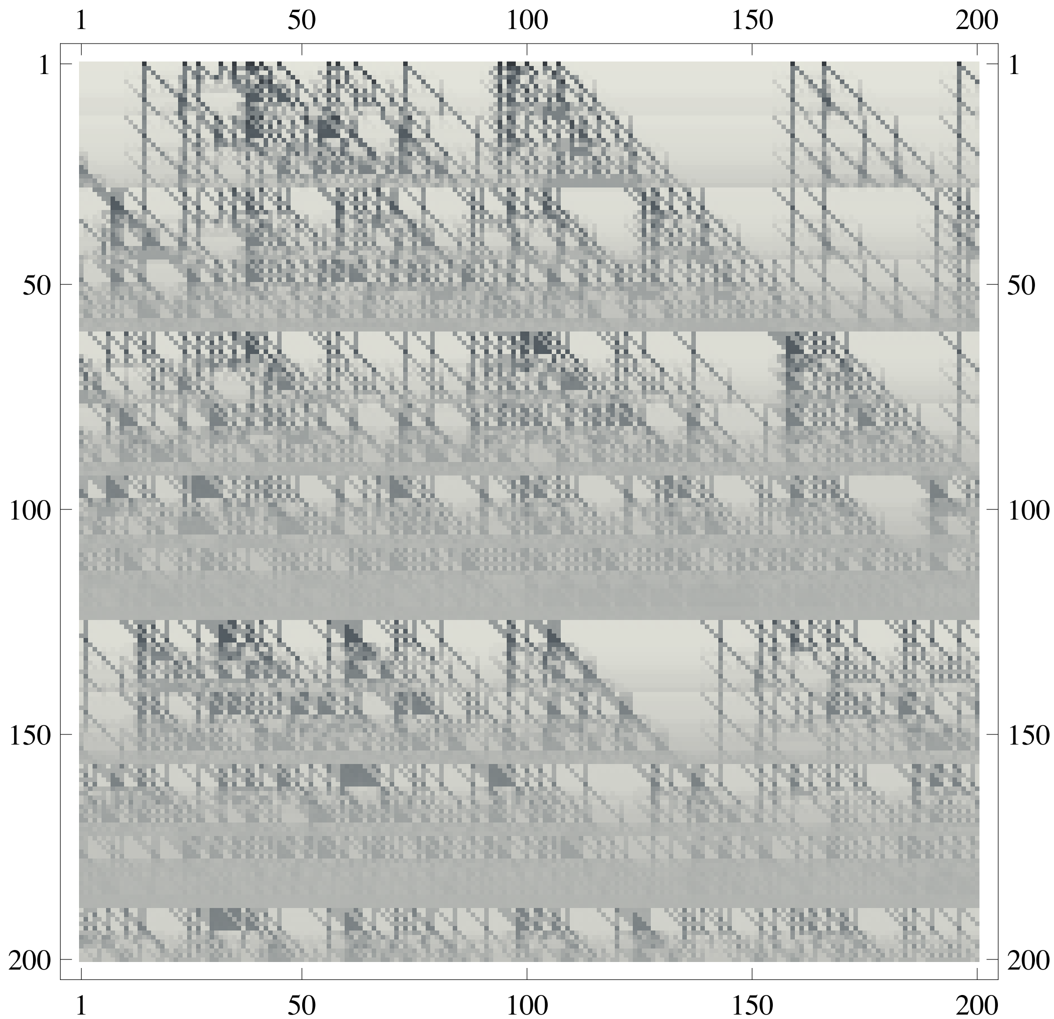


Regularity filter

- Applied to pattern of R60 (row-by-row)



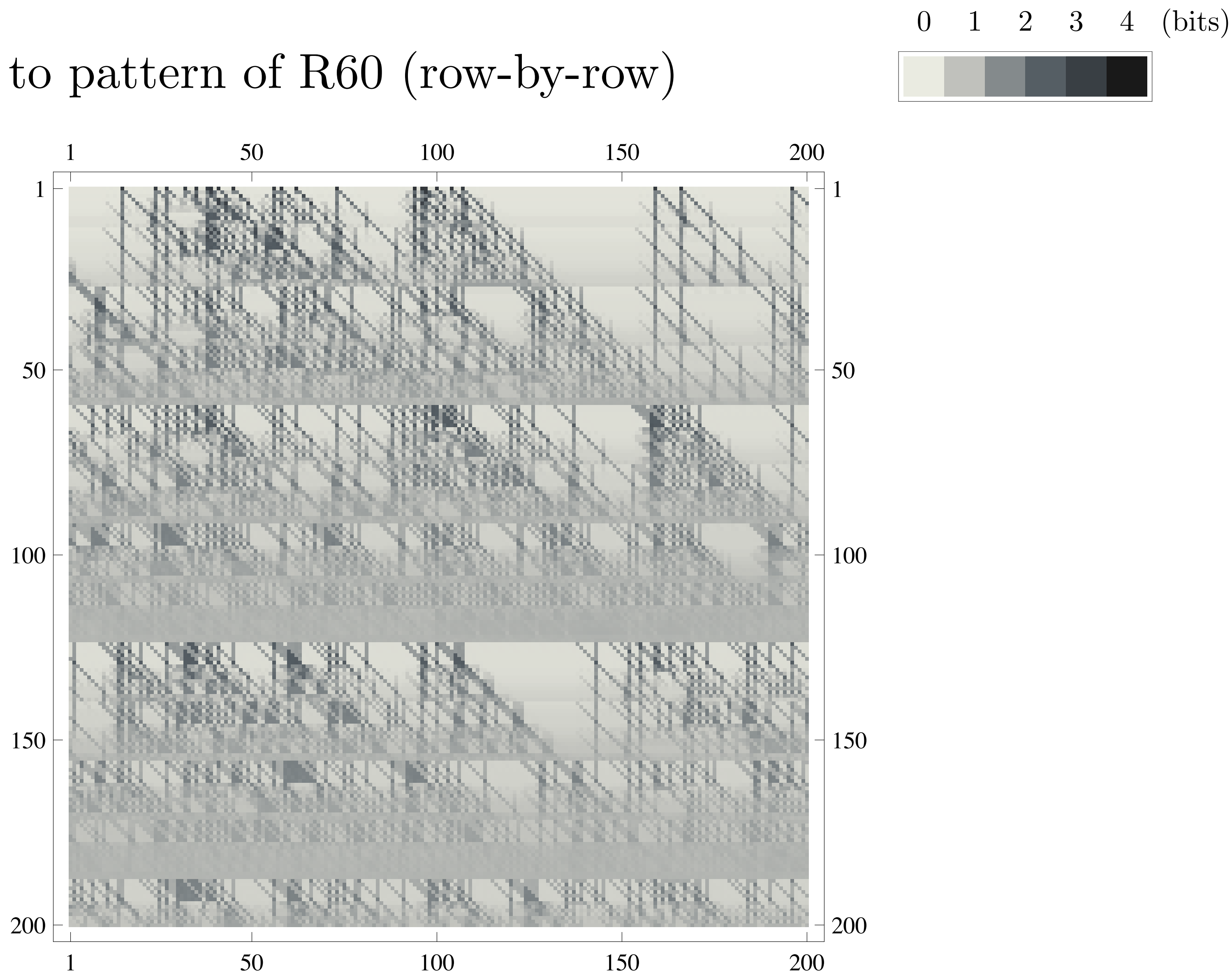
$m = 5$



Regularity filter

- Applied to pattern of R60 (row-by-row)

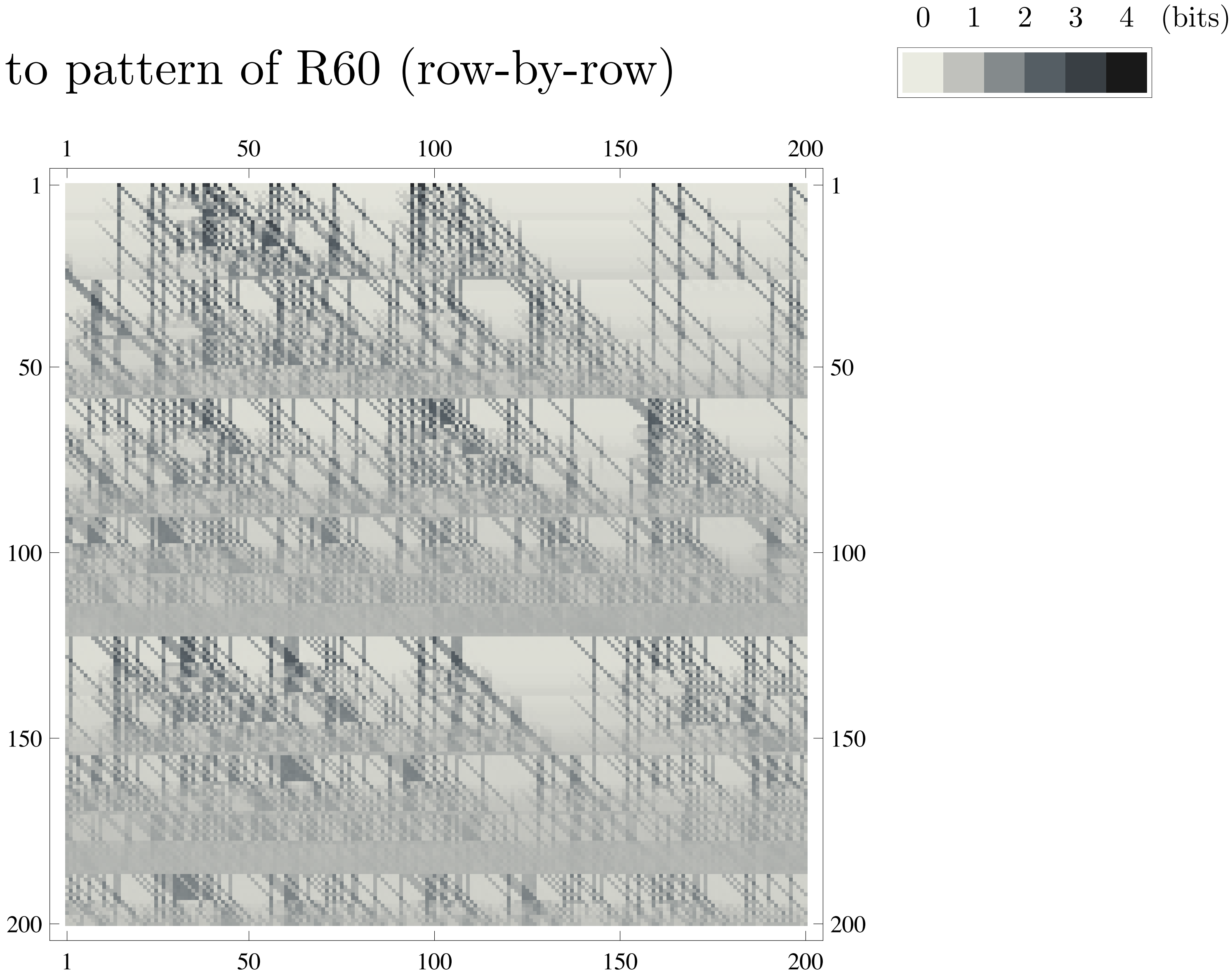
$m = 6$



Regularity filter

- Applied to pattern of R60 (row-by-row)

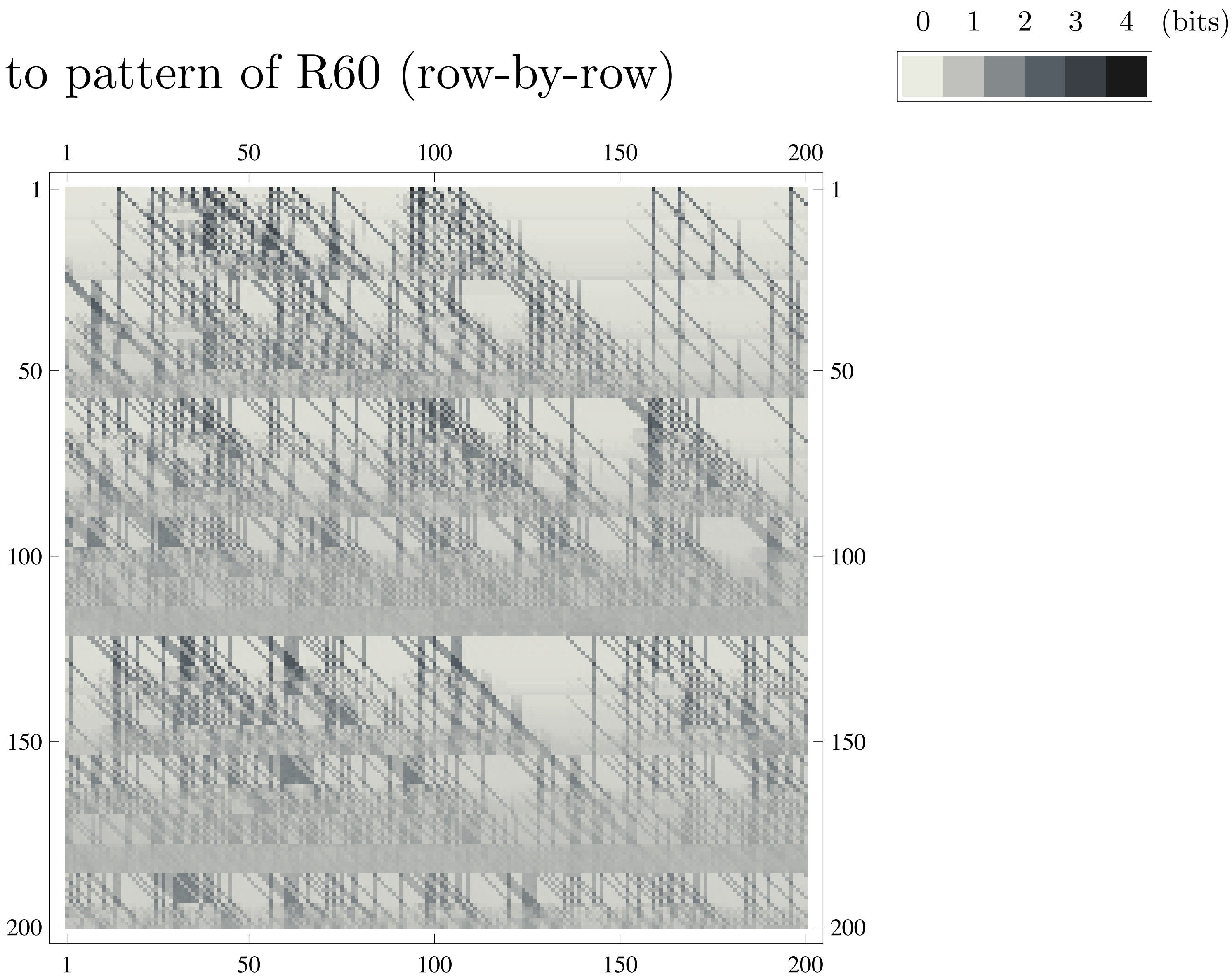
$m = 7$



Regularity filter

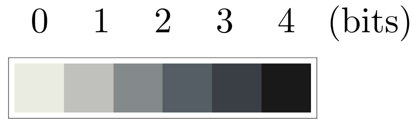
- Applied to pattern of R60 (row-by-row)

$m = 8$

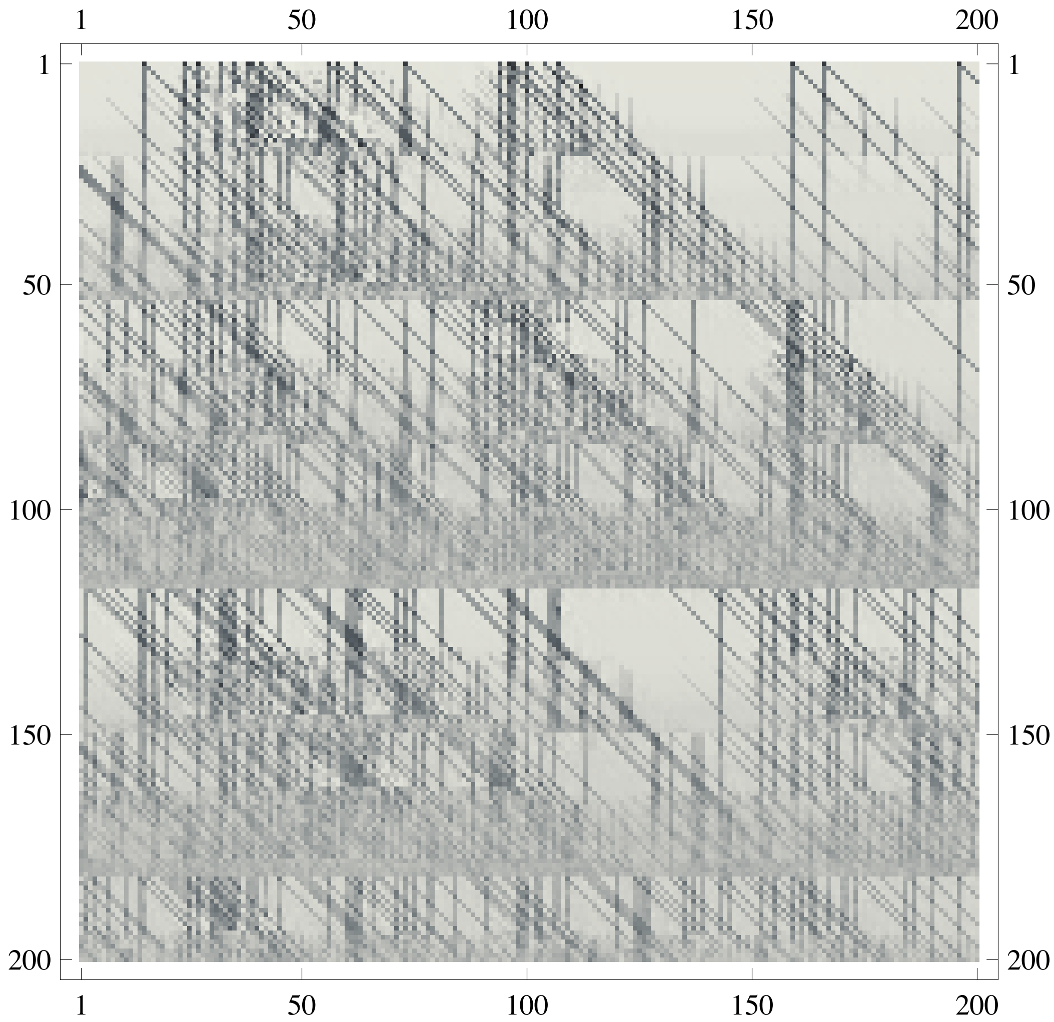


Regularity filter

- Applied to pattern of R60 (row-by-row)

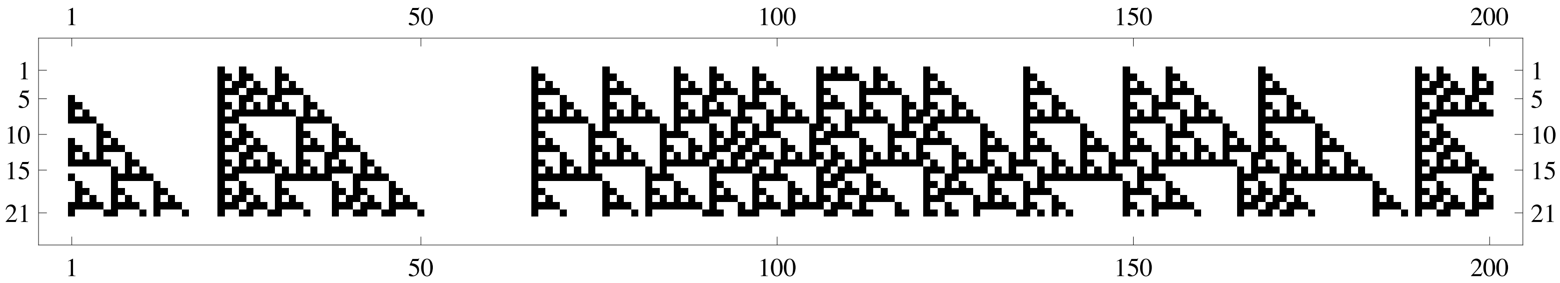


$m = 12$



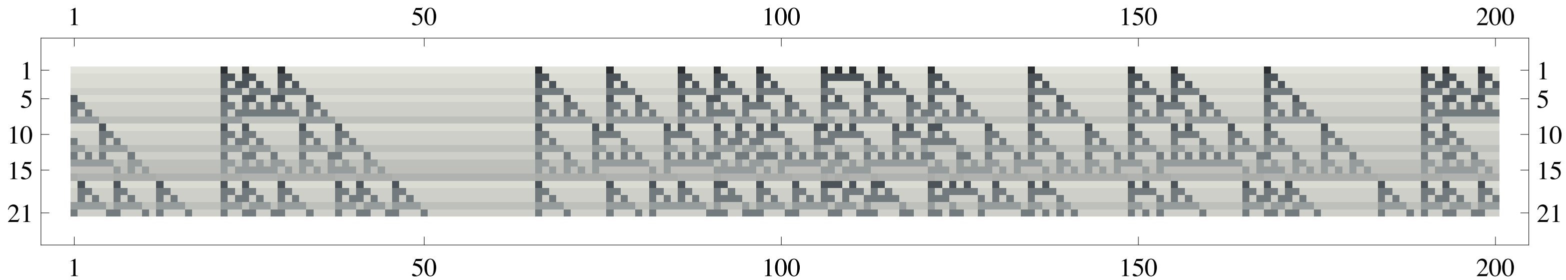
Regularity filter

- R60 up to $t = 21$.



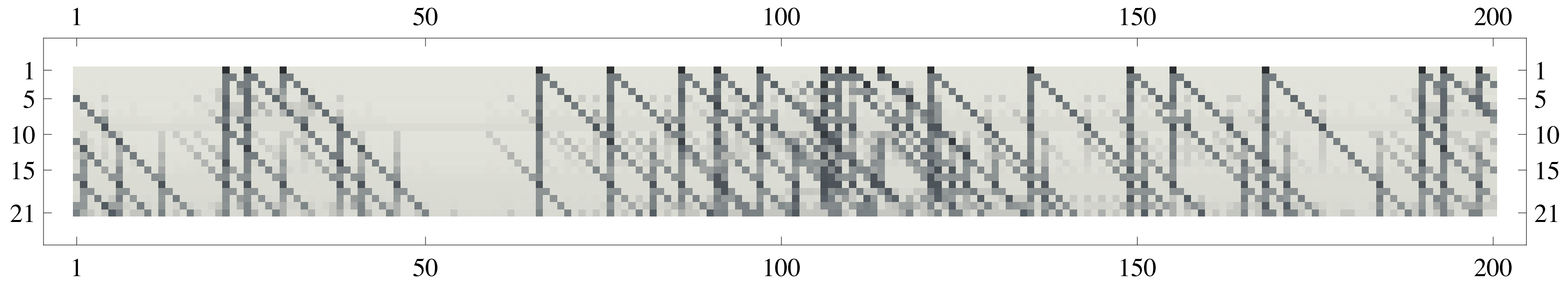
Regularity filter

- R60 up to length 15 blocks $m = 1$



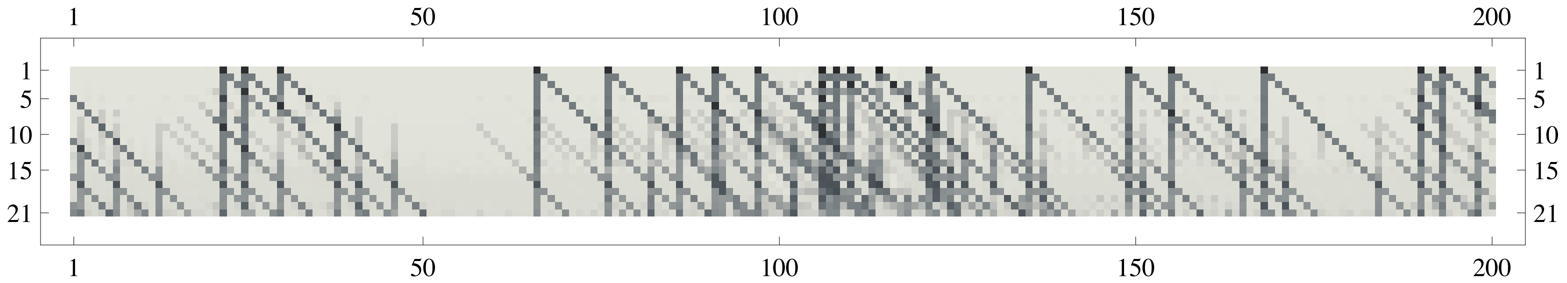
Regularity filter

- R60 up to length 15 blocks $m = 8$



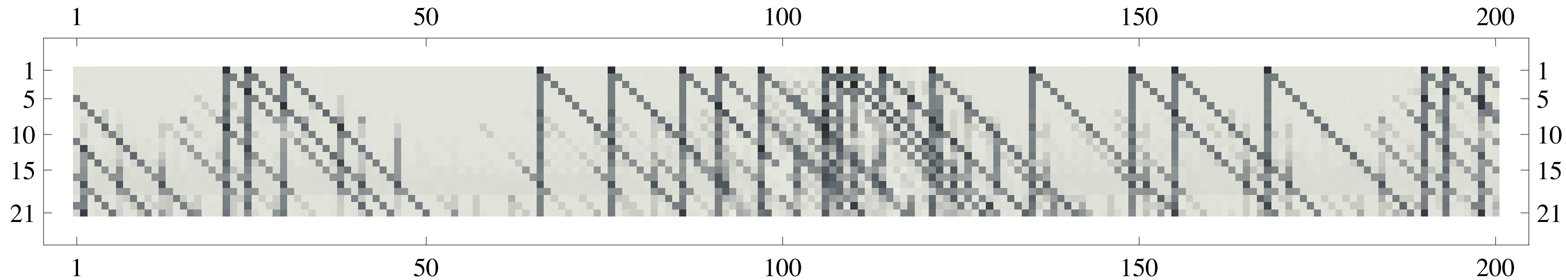
Regularity filter

- R60 up to length 15 blocks $m = 12$



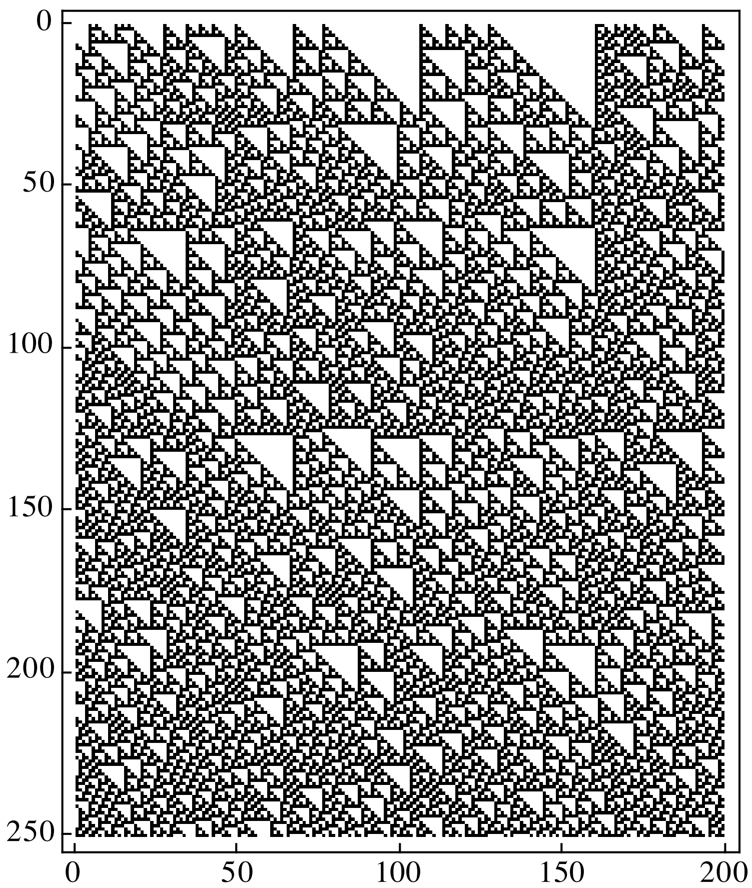
Regularity filter

- R60 up to length 15 blocks $m = 15$

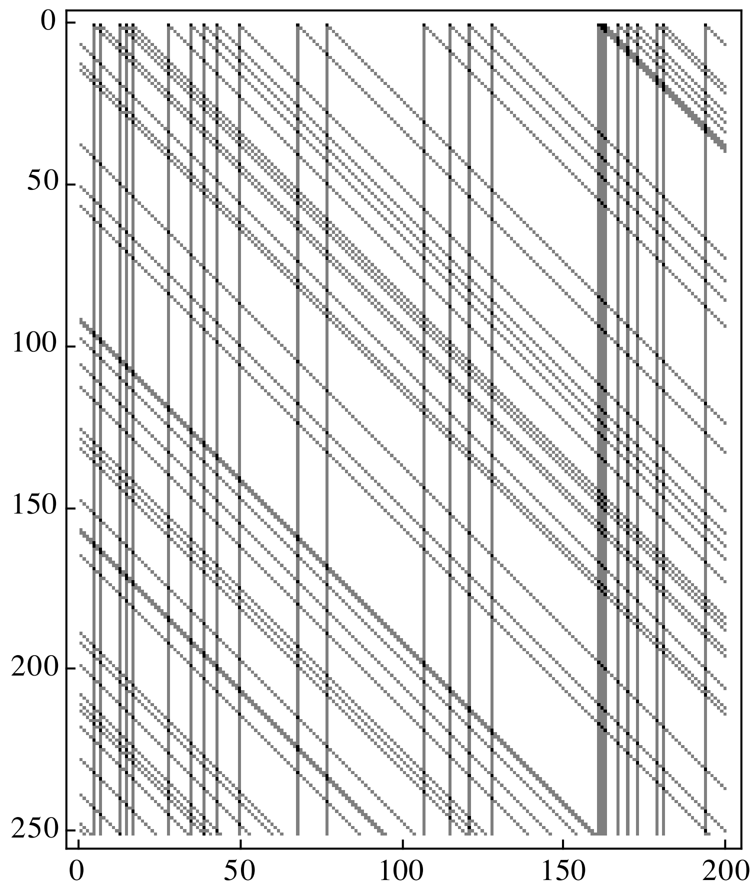


Analytic solution

Space-time diagram for rule R60



Local information, infinite m limit



"Additive" CA rule R60

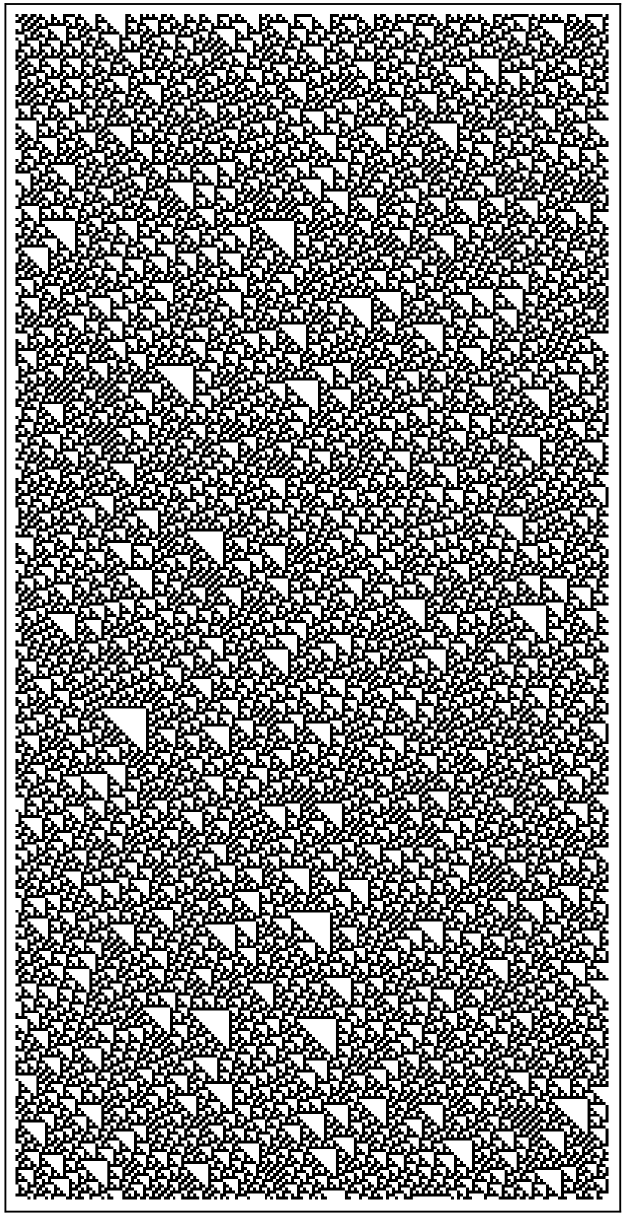
The rule that adds two neighbouring states mod 2 (XOR operation) has a certain degree of reversibility.

$$x_i' = f(x_{i-1}, x_i) = x_{i-1} + x_i \quad (\text{mod } 2)$$

An additive CA has a finite number of preimages to any state, and they define a class of "almost reversible" CA. For these CA one can show that entropy is conserved in time,

$$\Delta_t s(t) = 0$$

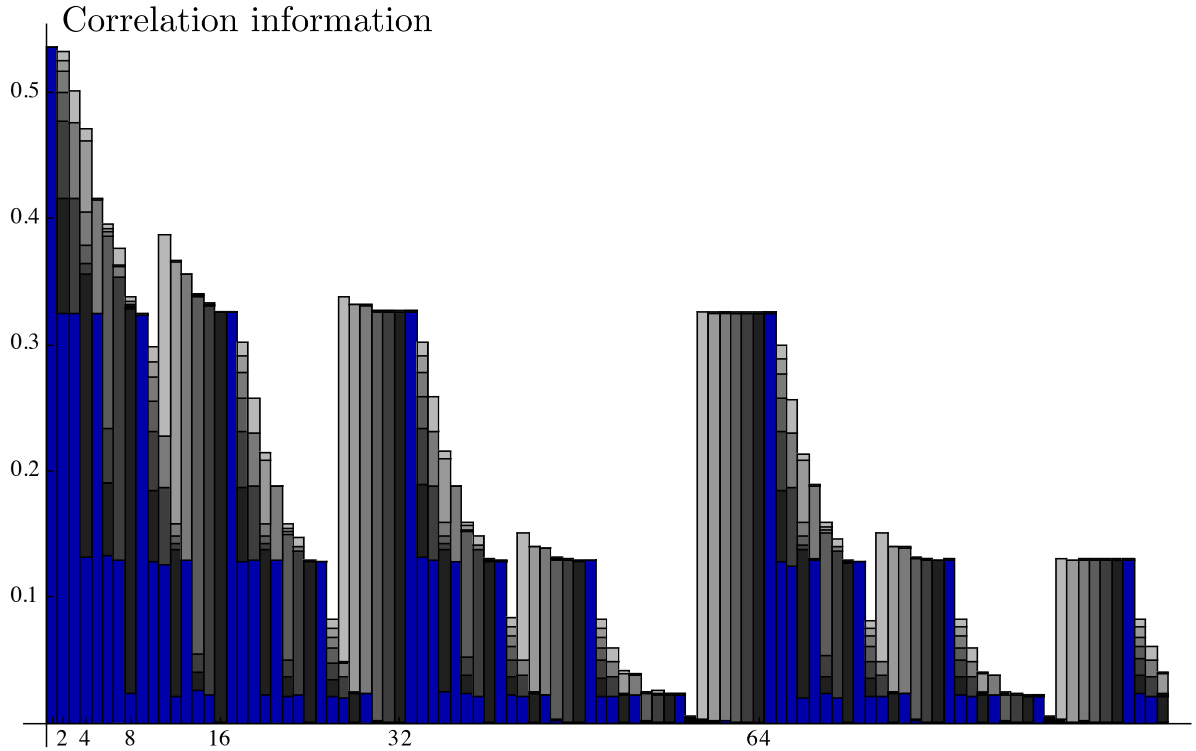
This means that if one starts with a completely random state (with maximum entropy $s = 1$), the state at any time will also be completely random.

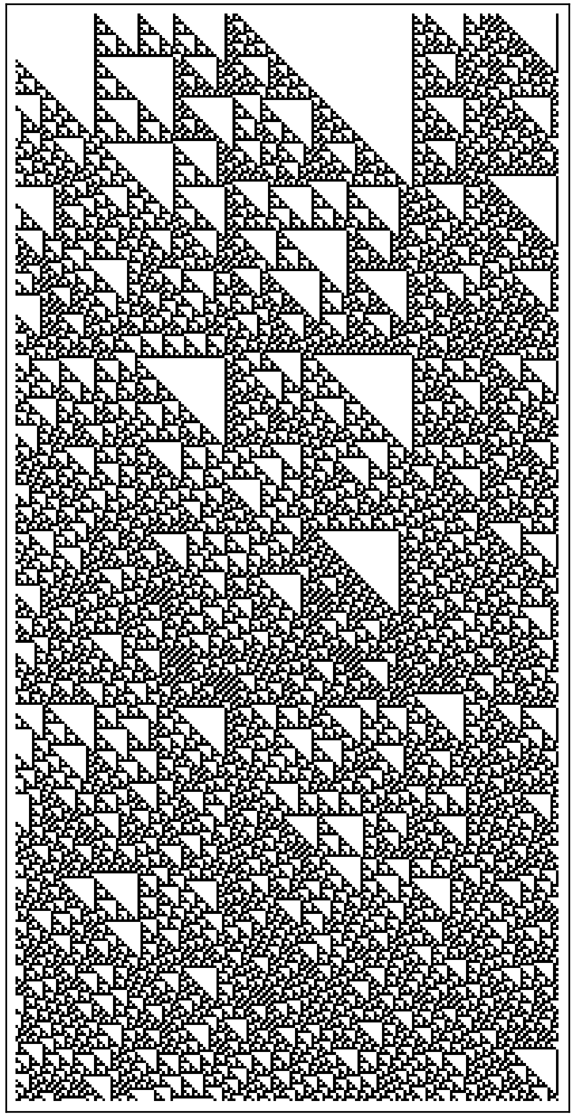


"Additive" CA rule R60

Slightly ordered initial state: low density of 1's, $p(1) = 0.1$, results in

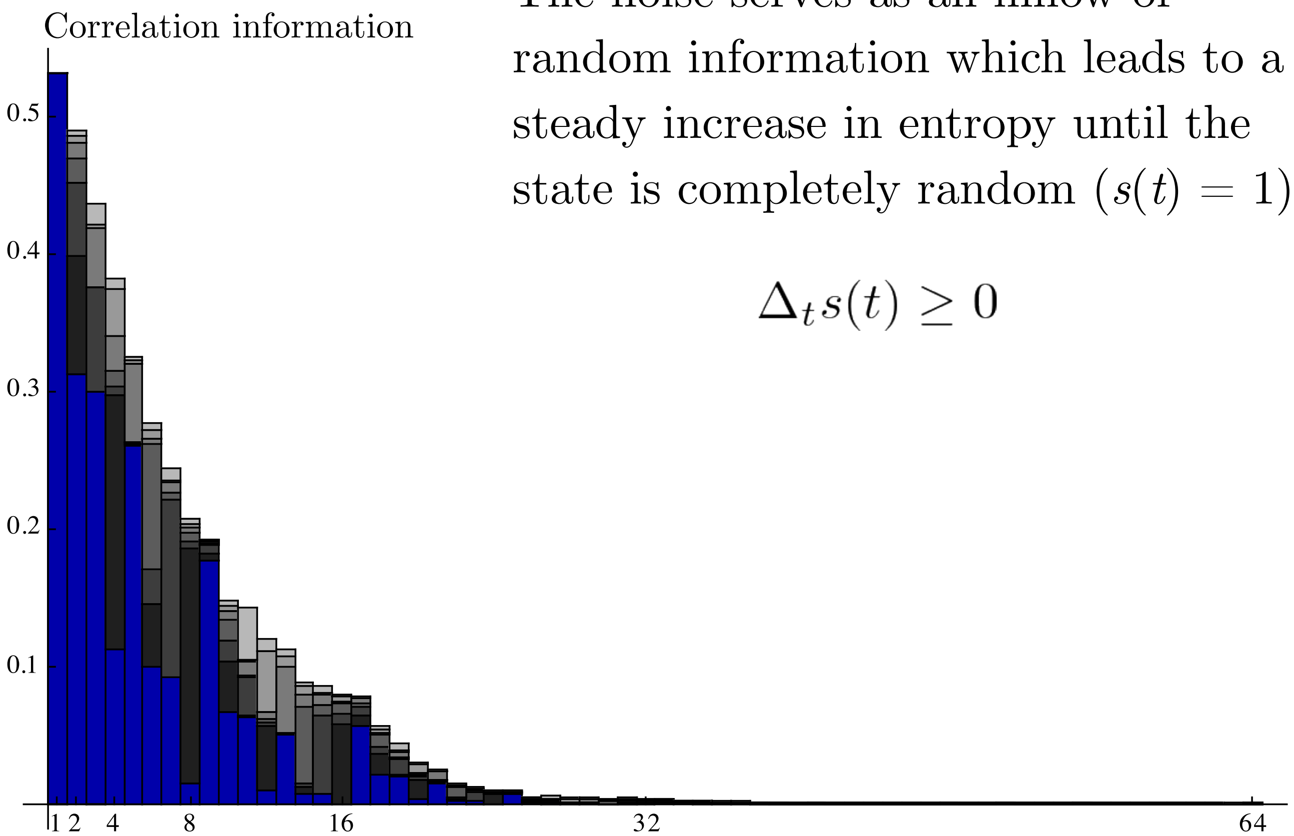
$$k_1 \approx 0.53,\ s \approx 0.47,\ k_m = 0 \text{ (for all } m > 0).$$





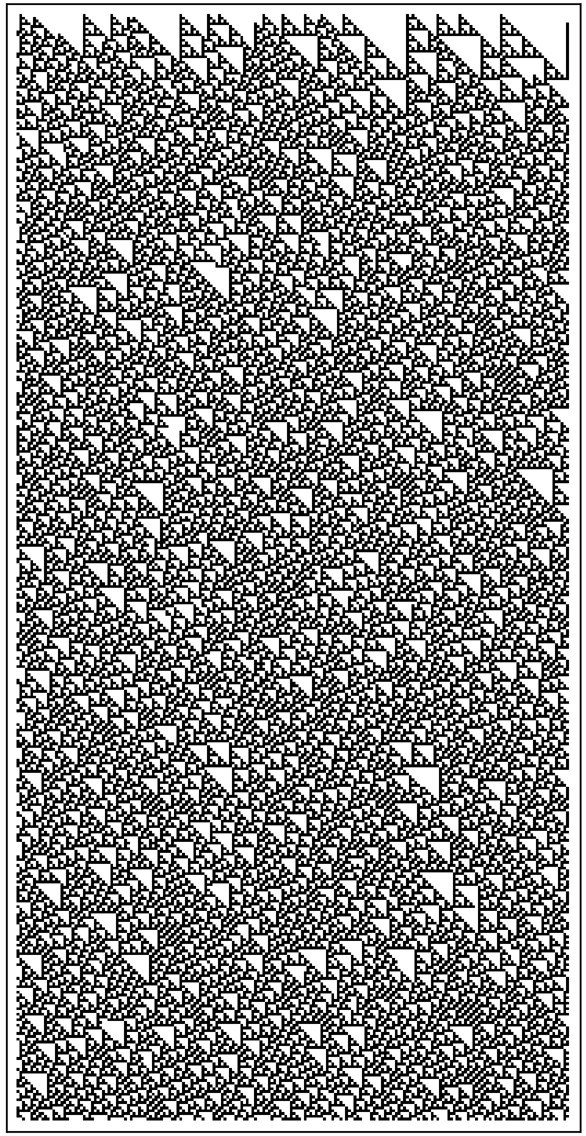
”Additive” CA rule R60 with noise

Noise added to the CA rule: with probability q a cell state is flipped (here $q = 0.005$). The noise destroys all long-range correlations.



The noise serves as an inflow of random information which leads to a steady increase in entropy until the state is completely random ($s(t) = 1$),

$$\Delta_t s(t) \geq 0$$



Complexity quantities

- How to quantify ”complexity” in a symbol sequence?
  - Entropy?
  - How correlation information is distributed?
  - How much information is there in the preceding symbols about the ones not yet read?

Effective measure complexity or Excess entropy

- Information in the past (σ_m) about the future (τ_n), expressed by the relative information,

$$\eta = \lim_{m\to\infty} \lim_{n\to\infty} \sum_{\sigma_m} p(\sigma_m) \sum_{\tau_n} p(\tau_n|\sigma_m) \log \frac{p(\tau_n|\sigma_m)}{p(\tau_n)}$$

- The distribution of correlation information over block lengths,

$$\eta = \sum_{k=2}^{\infty} (m-1) k_m =$$

$$= k_{\text{corr}} \sum_{k=2}^{\infty} (m-1) \frac{k_m}{k_{\text{corr}}} = k_{\text{corr}}\, l_{\text{corr}}$$

Excess entropy for rule R60

Assume an initial state (at $t = 0$) without correlations

- If $s = 1$ (maximum; equal densities of 0's and 1'), then $k_m = 0$ all $m \geq 2$ and

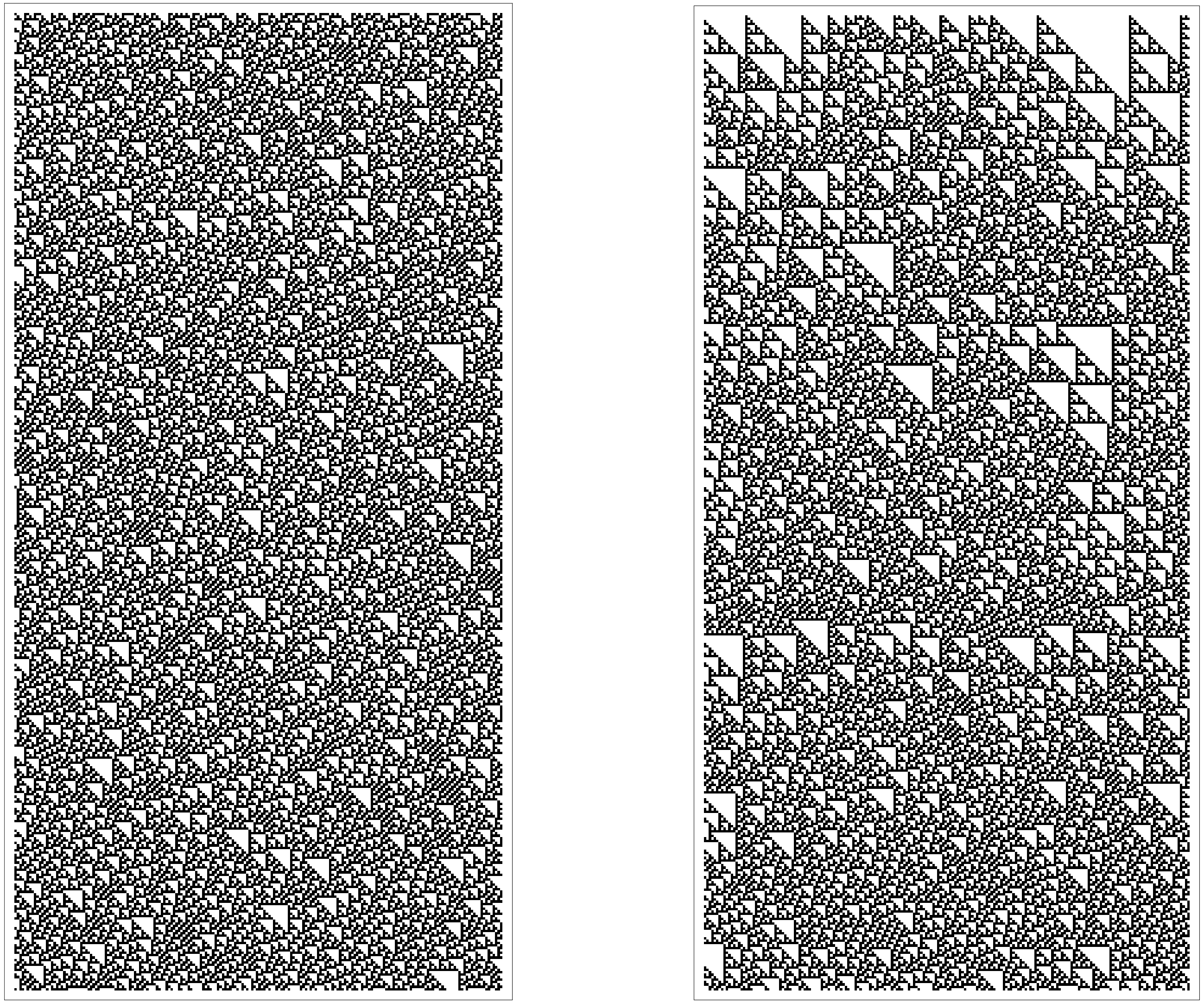
$$\eta(t) = 0$$

- If s < 1 (unequal densities of 0's and 1'), then

$$\eta(t) = t\ s$$

Excess entropy for R60

$s = 1$ $\quad$ $s < 1$



$\eta(t) = 0$ $\quad$ $\eta(t) = t\ s$

More information...

- Lecture notes (draft) available on course web site:

  http://studycas.com/node/114

  (Several papers can be provided on request.)