University of Milano-Bicocca
Department of Informatics, Systems and Communications

Open problems in the design of cryptographic applications based on Cellular Automata

Luca Mariot

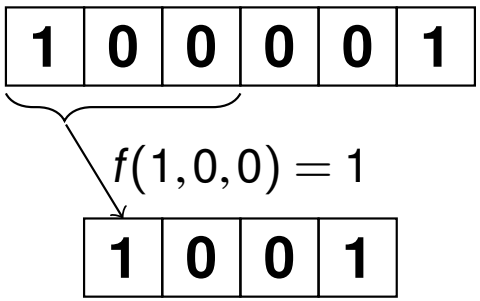
luca.mariot@disco.unimib.it

Delft – June 19, 2018

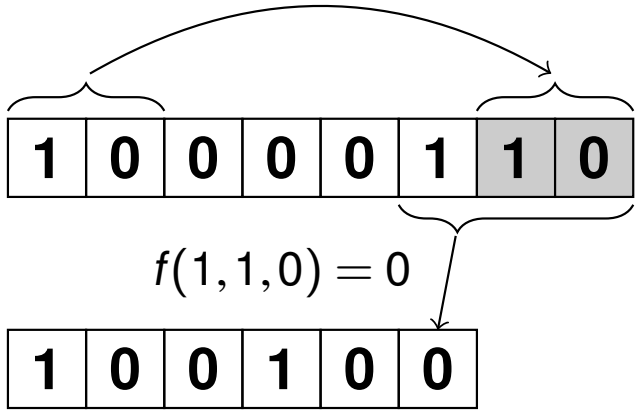
Context (1/2): Cellular Automata

- One-dimensional Cellular Automaton (CA): a discrete parallel computation model composed of a finite array of n cells
- Each cell updates its state $s \in \{0,1\}$ by applying a local rule $f : \{0,1\}^d \rightarrow \{0,1\}$ to itself and the $d-1$ cells to its right

Example: $n = 6$, $d = 3$, $f(s_i, s_{i+1}, s_{i+2}) = s_i \oplus s_{i+1} \oplus s_{i+2}$



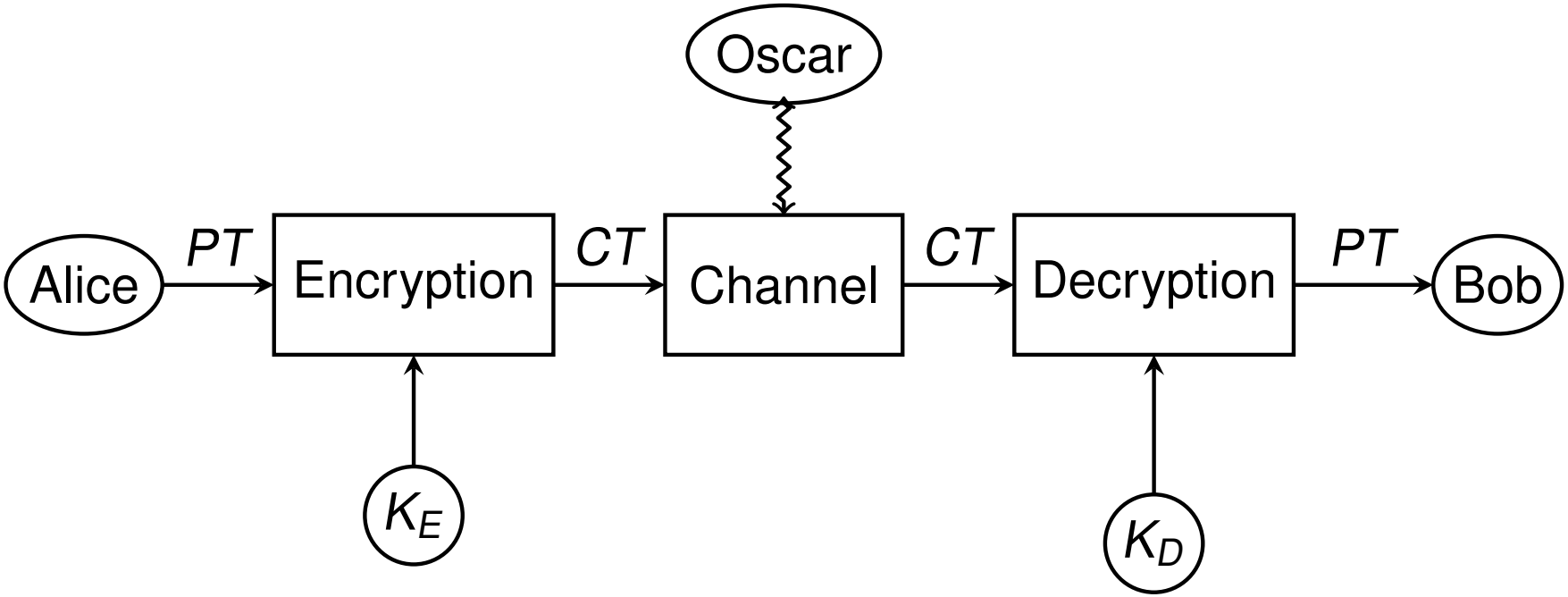
No Boundary CA – NBCA



Periodic Boundary CA – PBCA

Context (2/2): Cryptography

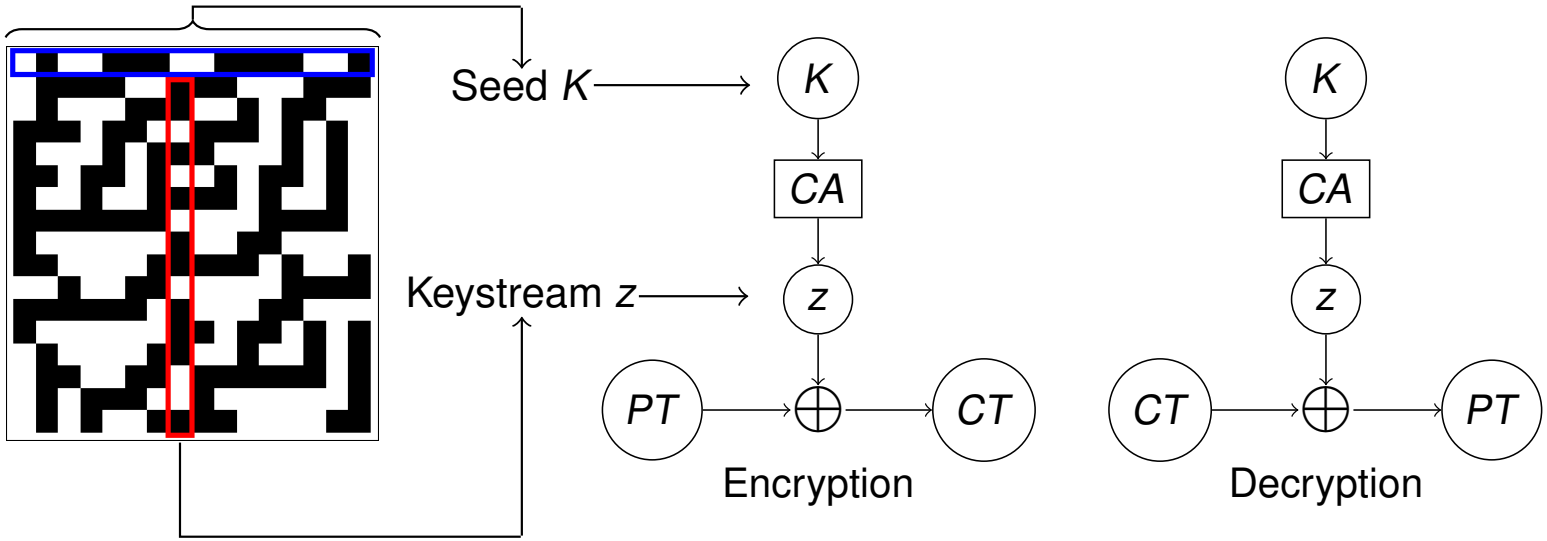
Basic Goal of Cryptography: Enable two parties (Alice and Bob, A and B) to securely communicate over an insecure channel, even in presence of an opponent (Oscar, O)



- PT: plaintext
- CT: ciphertext
- K_E: encryption key
- K_D: decryption key

CA-based Crypto History: Wolfram's PRNG

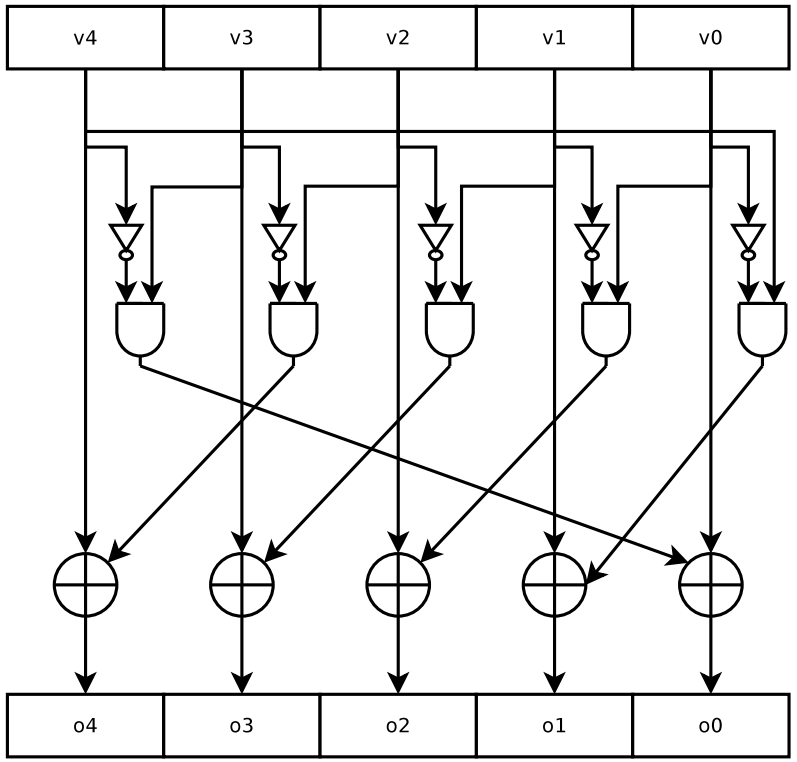
- CA-based Pseudorandom Generator (PRG) [W86]: central cell of rule 30 CA used as a stream cipher keystream



- This CA-based PRNG was later shown to be vulnerable [MS91]

CA-Based Crypto History: Keccak χ S-box

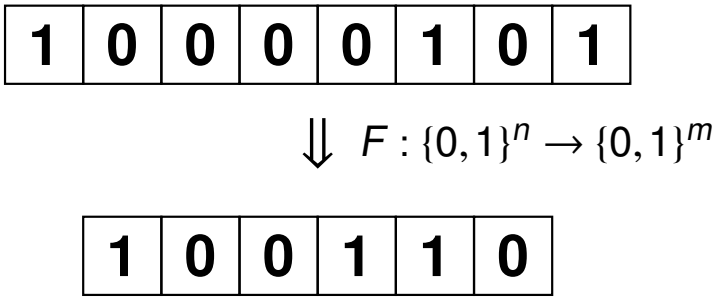
- Local rule: $\chi(x_1, x_2, x_3) = x_1 \oplus (1 \oplus (x_2 \cdot x_3))$ (rule 210)
- Invertible for every odd size n of the CA [DGV94]



- Used as a PBCA with $n = 5$ in the Keccak specification of SHA-3 standard [BDPV11]

Research Goal & Motivations

Research Goal: Investigate S-boxes $F : \{0,1\}^n \rightarrow \{0,1\}^m$ induced by CA to be used in block ciphers



Why CA, anyway?

1. **Security from Complexity**: Simple local rules can lead to very complex global behaviour in CA $\Rightarrow$ useful to provide confusion and diffusion in block ciphers
2. **Efficient implementation**: Leverage CA parallelism and locality for lightweight cryptography

State of the art in CA-based S-boxes

Nonlinearity of Boolean Functions

- Boolean function: a mapping $f : \{0,1\}^n \rightarrow \{0,1\}$

Truth table representation:

(x_1, x_2, x_3)	000	100	010	110	001	101	011	111
$f(x_1, x_2, x_3)$	0	1	1	1	1	0	0	0

$$\Downarrow$$

$$\Omega_f = (0,1,1,1,1,0,0,0)$$

- Nonlinearity of f: minimum Hamming distance of f from the set of all linear functions $L_\omega(x) = \omega \cdot x = \omega_1 x_1 \oplus \cdots \oplus \omega_n x_n$:

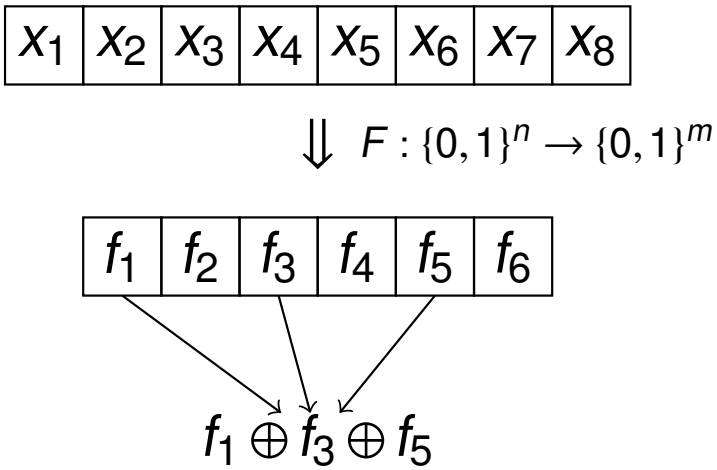
$$N_f = 2^{n-1} - \frac{1}{2}(|W_{max}(f)|)$$

where $W_{max}(f)$ is the maximum of the Walsh transform of f:

$$W_f(\omega) = \sum_{x \in \{0,1\}^n} (-1)^{f(x) \oplus \omega \cdot x}$$

Nonlinearity of S-boxes

- Substitution Box (S-box): mapping $F : \{0,1\}^n \to \{0,1\}^m$
- Component functions $v \cdot F : \{0,1\}^n \to \{0,1\}$ for $v \in \{0,1\}^m$: linear combinations of coordinate functions $f_i : \{0,1\}^n \to \{0,1\}$



- Nonlinearity of $F \Leftrightarrow$ minimum nonlinearity among all its component functions
- S-boxes with high nonlinearity allow to resist to linear cryptanalysis attacks

Upper Bound on Nonlinearity of CA S-Boxes

- We proved the following upper bound for S-boxes based on both NBCA and PBCA [MPLD18]:

Theorem

The nonlinearity of the S-box F of an n-cell NBCA or PBCA with local rule $f : \{0,1\}^d \to \{0,1\}$ satisfy

$$N_F \leq 2^{n-d} \cdot N_f$$

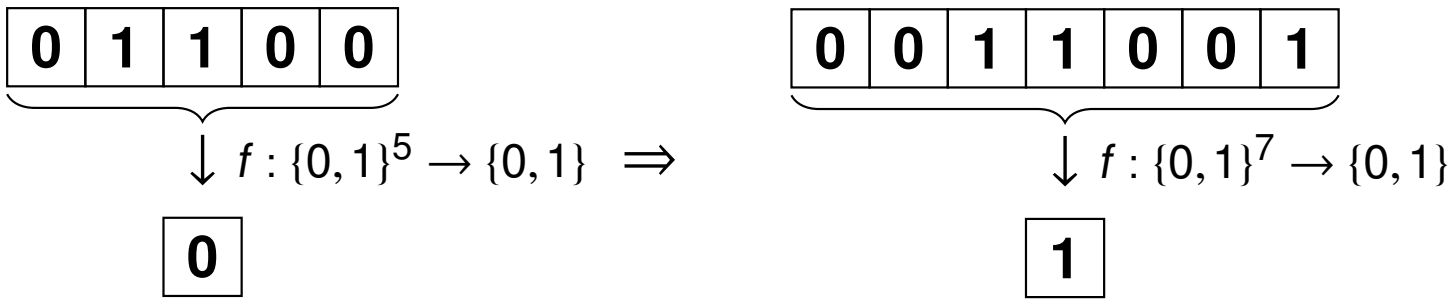
- **Remark**: This explains why adding cells to a CA makes the cryptographic properties of the S-box worse (see e.g. KECCAK)

Open Problems

Lower Bounds

- Up to now, we only know how good the nonlinearity of a CA-based S-box can be
- Necessity to characterize the nonlinearity of CA component functions more precisely for a lower bound
- Interesting byproduct: Secondary construction of Boolean functions based on CA

Secondary construction: generate a family of (larger) Boolean functions with specific nonlinearity starting from known ones



Original function $f : \{0,1\}^5 \to \{0,1\}$ Extended function $f' : \{0,1\}^7 \to \{0,1\}$

Plateaued Boolean Functions & CA

- $f : \{0,1\}^n \to \{0,1\}$ is plateaued iff:

$$W_f(\omega) = \sum_{x \in \{0,1\}^n} (-1)^{f(x) \oplus \omega \cdot x} \in \{-2^r, 0, +2^r\}$$

- Plateaued functions achieves maximal nonlinearity, and satisfy other interesting crypto properties (e.g., resiliency)
- Example: Keccak rule χ is a plateaued function of 3 variables

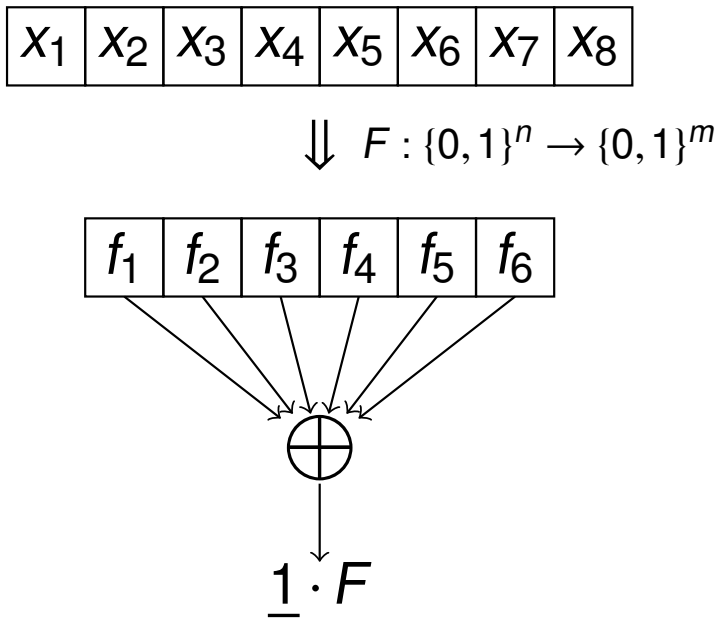
Question: Find plateaued functions via a secondary construction based on cellular automata

Component Functions

- We focused on the following component function of a n-cell NBCA F with local rule $f : \{0,1\}^d \to \{0,1\}$:

$$\underline{1} \cdot F = \bigoplus_{i=1}^{m} f_i(x_1, \cdots, x_n) = \bigoplus_{i=1}^{m} f(x_i, \cdots, x_{i+d-1})$$

- In other words, we take the component which XORs *all* coordinate functions of the CA:



Preliminary Observations

Table : Nonlinearities and numbers of plateaued local rules of $d = 3$ variables whose $\underline{1} \cdot F$ components are plateaued with index $r = \lceil \frac{n+1}{2} \rceil$

n	$NI(\underline{1} \cdot F)$	#RULES
3	2	112
4	4	48
5	12	112
6	24	80
7	56	96
8	112	80
9	240	96
10	480	64
11	992	96

Remark: plateauedness of local rule is not a sufficient condition for plateauedness of $\underline{1} \cdot F$

Hypothesis Formulation and Testing

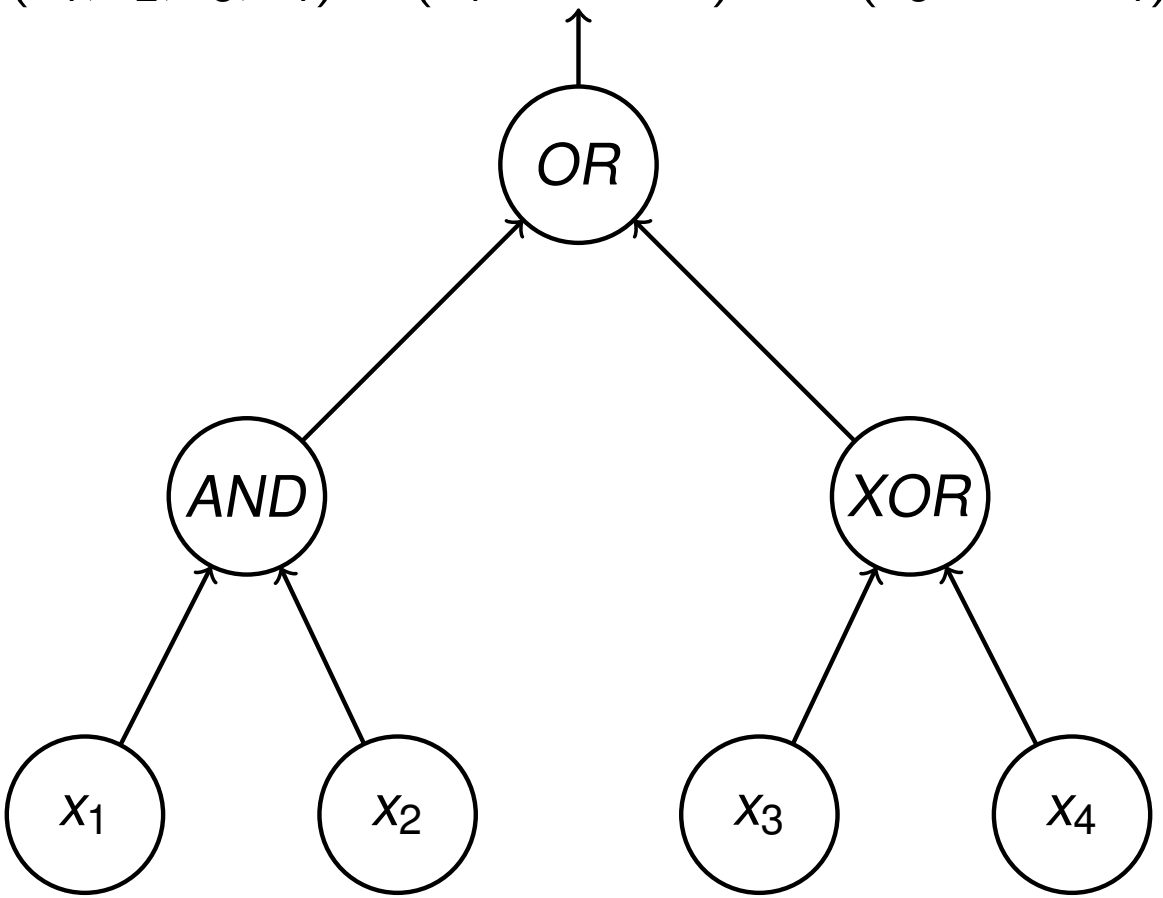
- **Conjecture**: for a certain subset of plateaued local rules of index $r = \lceil \frac{d+1}{2} \rceil$, the component $\underline{1} \cdot F$ of the n-cell NBCA is plateaued with index $r = \lceil \frac{n+1}{2} \rceil$
- **Question**: how to characterize such subset?
- Up to $d = 5$, the conjecture can be exhaustively checked (since there are 2^{2^d} d-variable Boolean functions)
- for $d > 5$, there is the necessity to use *heuristic methods* – such as Genetic Programming (GP)

Genetic Programming (GP)

- Optimization method inspired by evolutionary principles, introduced by Koza [K93]
- Each candidate solution (individual) is represented by a tree
  - Terminal nodes: input variables
  - Internal nodes: Boolean operators (AND, OR, NOT, XOR, ...)
- New solutions are created through genetic operators like tree crossover and subtree mutation applied to a population of candidate solutions
- Optimization is performed by evaluating the new candidate solutions wrt a fitness function

GP Tree Encoding – Example

$$f(x_1, x_2, x_3, x_4) = (x_1 \ AND \ x2) \ OR \ (x_3 \ XOR \ x_4)$$



Wrapping up – The Roadmap

A possible way to go to solve this conjecture:

1. Apply exhaustive search up to $d = 5$ to construct the subset of plateaued local rules yielding plateaued $\underline{1} \cdot F$ CA components
2. Formulate a hypothesis on the mathematical structure of this subset
3. Apply GP to test this hypothesis on local rules with $d > 5$
4. If GP finds a counterexample, then reformulate the structure of the subset and go back to 3. Otherwise, attempt to formally prove the conjecture

Further extension: Use this method to investigate construction of bent functions

Thank you!

References

- [BDPV11] Bertoni, G., Daemen, J., Peeters, M., Van Assche, G. 2011. The Keccak reference. `http://keccak.noekeon.org/` (2011)
- [DGV94] Daemen, J., Govaerts, R., Vandewalle, J. An efficient nonlinear shift-invariant transformation. In Proceedings of the 15th Symposium on Information Theory in the Benelux, pp. 108-115 (1994)
- [K93] J. R. Koza: Genetic programming – on the programming of computers by means of natural selection. Complex adaptive systems, MIT Press 1993
- [MPLD18] Mariot, L. Picek, S., Leporati, A., Jakobovic, D.: Cellular Automata Based S-Boxes. *Cryptography and Communications*, DOI: 10.1007/s12095-018-0311-8
- [MS91] Meier, W., Staffelbach, O. Analysis of Pseudo Random Sequence Generated by Cellular Automata. In EUROCRYPT, Vol. 91, pp. 186-200 (1991)
- [Wolfram86] Wolfram, S.: Random Sequence Generation by Cellular Automata. Adv. Appl. Math. 7(2), 123–169 (1986)